

Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific

 Anthony Stigliani,¹  Kevin S. Weiner,¹ and  Kalanit Grill-Spector^{1,2}

¹Department of Psychology and ²Stanford Neurosciences Institute, Stanford University, Stanford, California 94305

Prevailing hierarchical models propose that temporal processing capacity—the amount of information that a brain region processes in a unit time—decreases at higher stages in the ventral stream regardless of domain. However, it is unknown if temporal processing capacities are domain general or domain specific in human high-level visual cortex. Using a novel fMRI paradigm, we measured temporal capacities of functional regions in high-level visual cortex. Contrary to hierarchical models, our data reveal domain-specific processing capacities as follows: (1) regions processing information from different domains have differential temporal capacities within each stage of the visual hierarchy and (2) domain-specific regions display the same temporal capacity regardless of their position in the processing hierarchy. In general, character-selective regions have the lowest capacity, face- and place-selective regions have an intermediate capacity, and body-selective regions have the highest capacity. Notably, domain-specific temporal processing capacities are not apparent in V1 and have perceptual implications. Behavioral testing revealed that the encoding capacity of body images is higher than that of characters, faces, and places, and there is a correspondence between peak encoding rates and cortical capacities for characters and bodies. The present evidence supports a model in which the natural statistics of temporal information in the visual world may affect domain-specific temporal processing and encoding capacities. These findings suggest that the functional organization of high-level visual cortex may be constrained by temporal characteristics of stimuli in the natural world, and this temporal capacity is a characteristic of domain-specific networks in high-level visual cortex.

Key words: domain specificity; extrastriate body area; fMRI; fusiform face area; parahippocampal place area; visual word form area

Significance Statement

Visual stimuli bombard us at different rates every day. For example, words and scenes are typically stationary and vary at slow rates. In contrast, bodies are dynamic and typically change at faster rates. Using a novel fMRI paradigm, we measured temporal processing capacities of functional regions in human high-level visual cortex. Contrary to prevailing theories, we find that different regions have different processing capacities, which have behavioral implications. In general, character-selective regions have the lowest capacity, face- and place-selective regions have an intermediate capacity, and body-selective regions have the highest capacity. These results suggest that temporal processing capacity is a characteristic of domain-specific networks in high-level visual cortex and contributes to the segregation of cortical regions.

Introduction

Visual stimuli bombard us at different rates in our daily life. For example, words and scenes are typically stationary and vary at slow rates. In contrast, bodies are dynamic and typically change at faster rates. While these effects may seem intuitive, little is known about

how temporal processing capacity is implemented in high-level visual cortex and how it contributes to perception.

The prevailing hypothesis suggests that temporal processing capacity, or the amount of information that a particular brain region can process in a unit of time, is domain general and decreases at progressively higher stages of the ventral stream hierarchy (Singh et al., 2000; Mukamel et al., 2004; McKeef et al., 2007; Hasson et al., 2008; Gauthier et al., 2012). This hypothesis is derived from investigations of the temporal characteristics of neurons in nonhuman primates providing evidence for lower capacity in higher stages of the ventral stream processing hierarchy (i.e., in inferotemporal cortex, IT) than primary visual cortex, namely V1. Three types of empirical evidence support this hierarchy hypothesis. First, the firing rate of V1 neurons decreases for

Received Nov. 21, 2014; revised July 1, 2015; accepted July 26, 2015.

Author contributions: A.S. and K.G.-S. designed research; A.S., K.S.W., and K.G.-S. performed research; K.S.W. contributed unpublished reagents/analytic tools; A.S. and K.G.-S. analyzed data; A.S., K.S.W., and K.G.-S. wrote the paper.

This research was funded by National Institutes of Health Grant 1R01EY02391501A1. We thank Samantha Guiry for help generating stimuli.

The authors declare no competing financial interests.

Correspondence should be addressed to Anthony Stigliani, Department of Psychology, Jordan Hall, Stanford University, Stanford, CA 94305. E-mail: astiglia@stanford.edu.

DOI:10.1523/JNEUROSCI.4822-14.2015

Copyright © 2015 the authors 0270-6474/15/3512412-13\$15.00/0

presentation rates >10 Hz (Foster et al., 1985; Chance et al., 1998), while neuronal firing rates in IT decrease for slower rates >4 Hz (Keysers et al., 2001). Second, information accumulation in V1 is at least twice as fast as in IT (Optican and Richmond, 1987; Richmond and Optican, 1990; Heller et al., 1995). Third, response latencies increase across the visual hierarchy from V1 to IT (Schmolsky et al., 1998), as does the dispersion of latencies across anatomically adjacent neurons (Vogels and Orban, 1994). fMRI studies in humans also support this hypothesis. Several studies reported that early visual areas (V1–hV4) process 10–30 items per second, whereas high-level regions in ventral temporal cortex (VTC) process only 4–6 items per second (Mukamel et al., 2004; Liu and Wandell, 2005; McKeef et al., 2007; Gauthier et al., 2012).

Contrary to the prevailing hierarchy hypothesis, recent studies found similar temporal processing in lateral occipital temporal cortex (LOTc) and VTC (Mukamel et al., 2004; Gentile and Rossion, 2014), even though LOTc is considered to precede VTC along this hierarchy (Haxby et al., 2000; Peelen and Downing, 2007; Schwarzlose et al., 2008; Dilks et al., 2013). These conflicting findings suggest a new unconsidered hypothesis that the temporal processing capacity of high-level visual regions may be domain specific and related to the natural temporal regularities of the stimuli that they process rather than determined by their stage in the visual hierarchy. This temporal hypothesis predicts that regions processing stimuli with fast dynamics in the natural world (such as bodies) will have a higher temporal processing capacity than regions processing stimuli with slower dynamics (such as words).

To test these hypotheses, we scanned subjects with fMRI while they viewed stimuli from multiple domains (Fig. 1A, characters, bodies, faces, and places) at various rates (Fig. 1B, 1–8 Hz). Crucially, the amount of information in a trial was maintained across rates by keeping the number of stimuli in a trial constant. We reasoned that selectivity in regions processing stimuli of a particular domain would be highest when the presentation rate is aligned with their temporal processing capacity and would substantially decrease when that capacity is surpassed. We considered two hypotheses: (1) a domain-general hierarchy hypothesis that predicts decreasing capacities from LOTc to VTC independent of category preference and (2) a domain-specific temporal hypothesis that predicts domain-varying temporal capacities independent of processing stage in LOTc or VTC.

Materials and Methods

Participants

Twelve right-handed subjects (five female, ages 19–44) with normal or corrected-to-normal vision were recruited from Stanford University for the main experiment (Experiment 1). Eight of these subjects (four female) returned for both a control study (Experiment 2) and a retinotopic mapping experiment to functionally define primary visual cortex (V1) using a population receptive field model (Dumoulin and Wandell, 2008). Participants gave their written informed consent. The Stanford Internal Review Board on Human Subjects Research approved all procedures.

fMRI

Experiment 1: fMRI temporal processing capacity experiment. Subjects viewed gray-level stimuli, which consisted of two different types of five different domains: characters, bodies, faces, places, and objects (Fig. 1A, example stimuli in first row). Characters consisted of pseudowords (Glezer et al., 2009) and numbers. Bodies consisted of either whole bodies with obscured heads or limbs. Faces were either those of adults or children. Places were either indoor corridors or houses. Finally, objects

were either cars or guitars. The view, size, and retinal position of the images varied. Each item was overlaid on a 10.5° phase-scrambled background generated from a randomly selected image from the entire set to minimize low-level differences across categories.

We introduce a novel paradigm in which the number of stimuli (information) in a trial is constant (eight stimuli per trial; Fig. 1B) and the presentation rate, and consequently trial duration, varies across conditions. Prior studies (Mukamel et al., 2004; McKeef et al., 2007; Gentile and Rossion, 2014) used constant block duration, confounding the number of stimuli with rate: blocks with faster rates had more stimuli compared with blocks with slower rates. Consequently, with this design, it is impossible to disentangle two competing effects on fMRI responses: stimulus rate versus the number of stimuli. In contrast, our new design measures the effect of rate alone on cortical responses. In each trial, stimuli of one category were shown in rapid serial visual presentation (RSVP) at a single rate of 1, 2, 4, or 8 items per second (Hertz). Stimulus trials were counterbalanced with blank baseline trials. In each run, stimuli were shown at one rate. The order of runs at different rates was counterbalanced across subjects. The total duration of visual presentation at a particular rate was 11 min, which equated the statistical power across rates. The same set of 1440 images was shown across all rates, but images did not repeat within a rate.

Subjects fixated on a central dot and pressed a button when a phase-scrambled oddball image appeared randomly in a trial. Trials contained 0, 1, or 2 phase-scrambled images with probability of 0.25, 0.5, or 0.25, respectively.

Experiment 2: control study. The design of the second experiment was identical to the main experiment, except that we modified the stimulus set (Fig. 1A, example stimuli in second row) to simultaneously control for multiple low-level differences between stimuli of different categories including their (1) contrast, (2) luminance, (3) similarity to other stimuli of the same category, (4) visual field coverage, and (5) spatial frequency power distributions. In doing so, we also considered the impact of image manipulations on high-level factors and made efforts to equate stimuli of different categories in terms of (6) familiarity and (7) memorability. Below, we detail the results from the quantification of these seven effects from Experiment 2.

(1) **Michelson Contrast:** Calculated as the ratio between the difference and sum of the maximum and minimum pixel intensities, was matched across all images by scaling the grayscale values of each image such that 1% of pixels is saturated at the lowest and highest intensities, respectively. Further image processing by the SHINE toolbox (Willenbockel et al., 2010) and spatial filtering introduced a small amount of variability in contrast across images, but there is no significant difference in contrast level among images of different categories (no main effect of category, $F_{(4,1435)} = 0.73$, $p > 0.05$; Fig. 1C, Contrast).

(2) **Luminance:** We used the SHINE toolbox (Willenbockel et al., 2010) to match histograms of grayscale values across all stimuli such that the mean luminance of images from different categories does not differ (no main effect of category, $F_{(4,1435)} = 0.97$, $p > 0.05$; Fig. 1C, Luminance).

(3) **Similarity:** To estimate visual similarity among images of a category, we measured the mean Euclidean distance between normalized grayscale values (scaled to be between 0 and 1) of each stimulus and every other stimulus of the same category (Grill-Spector et al., 1999). The distance between a pair of images was calculated using the following formula in which p represents the p th pixel in an image, n represents the total number of pixels in an image, and I_p and I'_p represent the p th pixel in two different images:

$$D = \frac{\sqrt{\sum_{p=1}^n (I_p - I'_p)^2}}{n}$$

Similarity (S) was defined as follows: $S = 1 - D$. Thus, pairwise similarity values range from 0 (images with inverted intensity difference at each pixel) to 1 (identical images). Although there are numerically significant differences in mean similarity across categories (main effect of category, $F_{(4,101529)} = 788.52$, $p < 0.001$), the range of similarity values is similar across categories (Fig. 1C, Similarity), and differences in the mean values

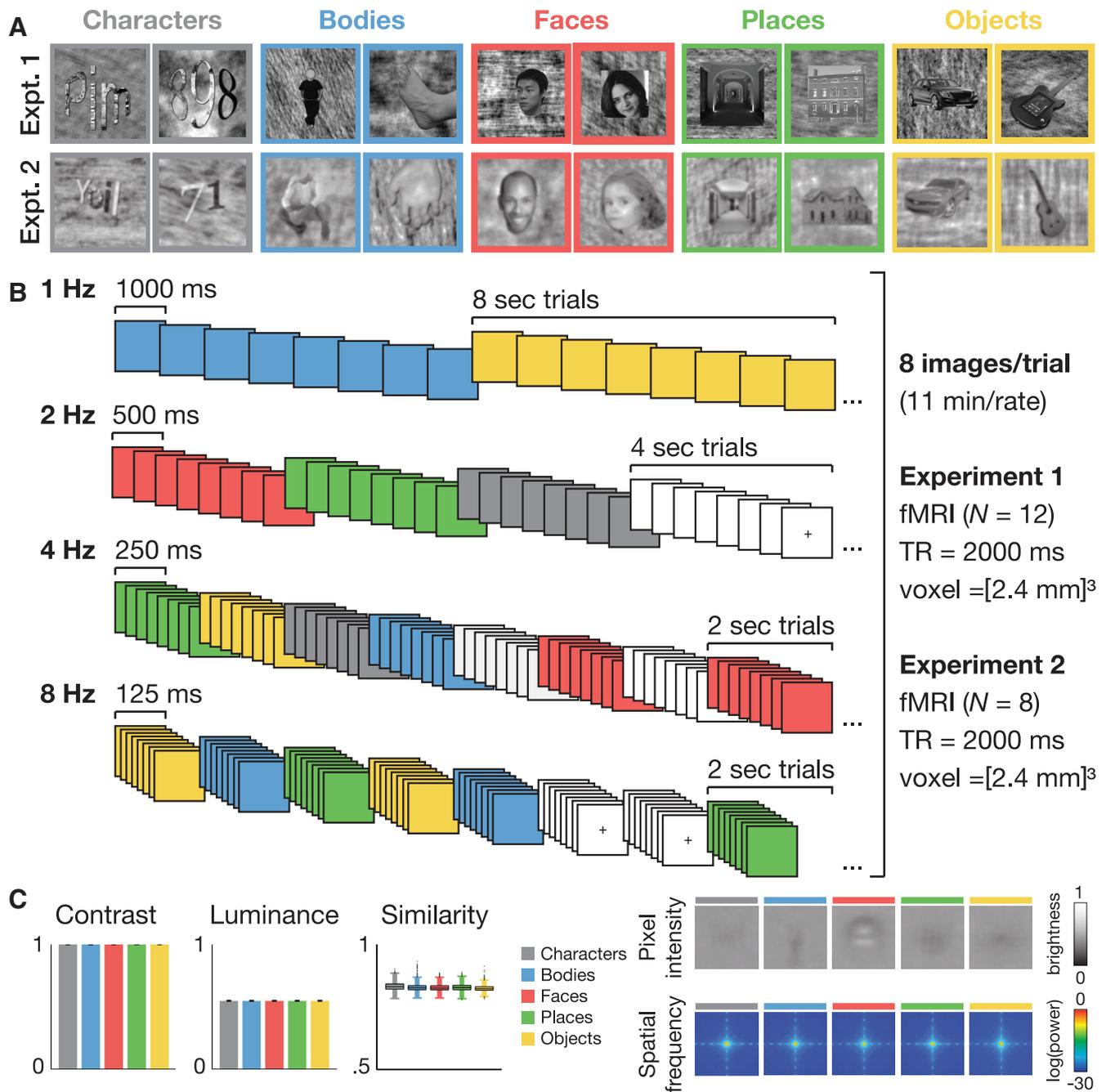


Figure 1. Experimental design. **A**, Example images used in experiments color coded by category for illustration. **B**, Partial stimulus sequences from fMRI studies. Images in a given run were displayed in trials of eight stimuli each, all from the same category. In different runs, images in a trial were shown at a single rate of 1 Hz, 2 Hz, 4 Hz, or 8 Hz. Stimuli presented at 8 Hz were shown for 1 s followed by a 1 s blank screen to complete the 2 s TR. **C**, Image analysis of stimulus set used in Experiment 2. Contrast: Ratio between the difference and sum of the maximum and minimum pixel intensities of an image. Luminance: Average luminance across an image. Error bars indicate standard deviation (SD) across images. Similarity: 1 – (mean distance between all pairs of images in a category). Boxes depict the upper and lower interquartile range of values for a given category with the medians represented by black dashes; whiskers extend beyond this range ± 2 times the size of the upper and lower interquartile ranges, respectively; black dots represent outlier data points falling outside of the range of the whiskers. Pixel Intensity: Mean luminance at each pixel location averaged across all images in a category. Spatial Frequency: Spectrograms of the mean proportion of total power in each spatial frequency for images in each category.

are not perceptually relevant because the size of the effect (0.0029, standard deviation (SD) of category means in normalized grayscale values) is below the resolution of the display (0.0039, minimum pixelwise distance resolved by the display in normalized values, corresponding to the change in one gray-level value of a possible 255 increments). Boxplots in Figure 1C, Similarity, show distributions of similarities for all possible pairwise comparisons in each category.

(4) Visual field coverage: While we varied the view, size, and position of stimuli used in Experiment 1 to maximize variation within a category and reduce variation across categories, we further refined the cropping and place-

ment of stimuli in the visual field in Experiment 2. This refinement further reduced biases related to intrinsic regularities in the shape and aspect ratio of different categories. Figure 1C, Pixel Intensity, shows the mean luminance of each pixel location, averaged across all images in a category.

(5) Spatial frequency: Was controlled by filtering the images with a low-pass Gaussian filter with a visual angle cutoff frequency of 8 cycles/degree. After filtering, the mean spatial frequency distribution across images of each category was matched. Figure 1C, Spatial Frequency, shows mean proportion of total power in each spatial frequency, averaged across all images in a category.

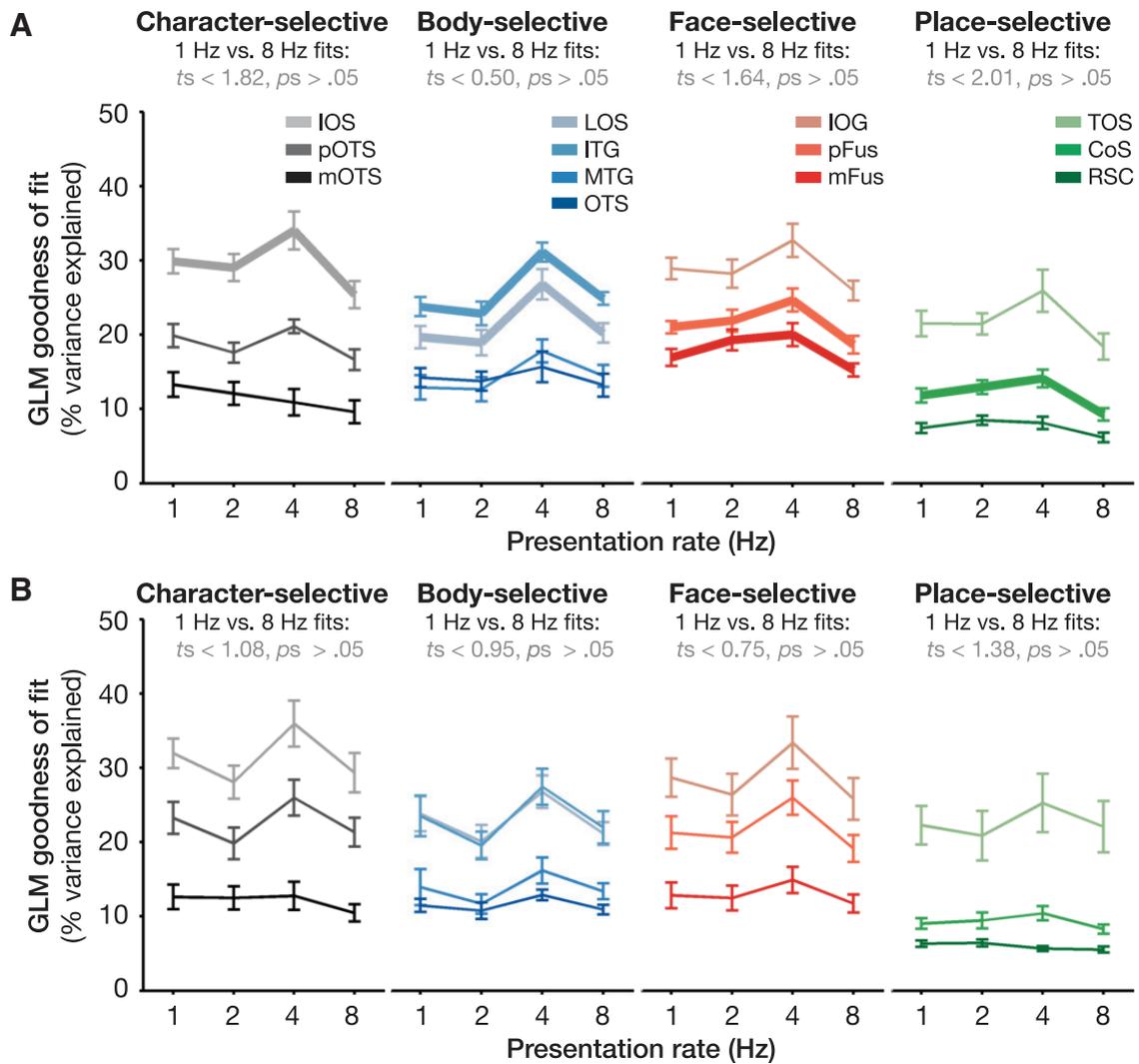


Figure 2. Consistent GLM model fits between fast (8 Hz) and slow (1 Hz) presentation rates. The mean proportion of variance explained by the GLM fits in ROIs selective to characters, bodies, faces, and places in Experiment 1 (A) and Experiment 2 (B). Error bars indicate SEM across subjects. Regions in which there was a significant main effect of rate ($p < 0.05$) are plotted with thick lines. Results of paired t tests comparing the variance explained by 1 and 8 Hz GLMs in individual ROIs are summarized for sets of regions above each plot. These analyses rule out the possibility that any differences that we find across regions reflect differences in model fits between fast (8 Hz) and slow (1 Hz) presentation rates or between short (2 s) and long (8 s) trial durations.

(6) Familiarity: To diminish familiarity effects across categories, all stimuli were selected to be equally unfamiliar to participants. None of the face or place stimuli were personally familiar to our participants, and pronounceable pseudowords and uncommon numbers were used as character stimuli instead of more familiar English words and frequently encountered numbers.

(7) Memorability: We tested whether the memorability of stimuli belonging to different categories was matched by comparing recognition memory for stimuli of various categories presented at 1 Hz (see below, Behavioral testing). Performance did not significantly differ across categories at 1 Hz (no main effect of category, $F_{(3,31)} = 0.79, p > 0.05$).

Functional localizer. All subjects participated in four runs of an independent functional localizer experiment (6 min per run) collected on a different day using 16 s blocks, 1 Hz presentation, and an oddball task (Weiner and Grill-Spector, 2010, 2011).

Data acquisition. Subjects were scanned on a GE 3 tesla Signa scanner at the Center for Cognitive and Neurobiological Imaging at Stanford University using a custom-built, phase-array 32-channel head coil. We acquired 34 slices covering occipitotemporal cortex (resolution: $2.4 \times 2.4 \times 2.4$ mm; one-shot T2*-sensitive gradient echo acquisition sequence: FOV = 192 mm, TE = 30 ms, TR = 2000 ms, and flip angle = 77°). A whole-brain, anatomical volume was acquired (T1-weighted

BRAVO pulse sequence; resolution: $1 \times 1 \times 1$ mm, TI = 450 ms, flip angle = 12°, 1 NEX, FOV = 240 mm).

Data analysis. Data were analyzed with MATLAB using the mrVista toolbox (<http://github.com/vistalab>) as in our prior publications (Weiner and Grill-Spector, 2010, 2011).

Functional ROIs were defined in individual subjects from the functional localizer using anatomical and functional criteria (Weiner and Grill-Spector, 2010, 2012; Grill-Spector and Weiner, 2014). A common threshold ($t > 4$, voxel level) was used to define all regions across all subjects, and subsequent analyses were conducted in ROIs identifiable in at least 9 of the 12 subjects (ROIs from a representative subject are shown in Fig. 3A).

Character-selective ROIs (characters > others) included a bilateral region in the inferior occipital sulcus (IOS, $N = 12$), a bilateral region in the posterior occipital temporal sulcus (pOTS; $N = 12$), corresponding to the visual word form area (Cohen et al., 2000; Ben-Shachar et al., 2011), and a more anterior patch often lateralized to the left hemisphere localized to mid OTS (mOTS; $N = 9$). In our prior studies, we use anatomical nomenclature with the corresponding selectivity to refer to face-, place-, and body-selective ROIs (e.g., mFus-faces). Here, we extend this nomenclature to character-selective regions. Thus, we use the label

pOTS-characters for visual word form area 1 (VWFA-1) and mOTS-characters for VWFA-2.

Body-selective ROIs (bodies > others) were found bilaterally in the lateral occipital sulcus (LOS; $N = 12$), inferior temporal gyrus (ITG; $N = 12$), and middle temporal gyrus (MTG; $N = 12$) as in Weiner and Grill-Spector (2011), collectively corresponding to the extrastriate body area (Downing et al., 2001). An additional body-selective region was defined in the OTS ($N = 9$), referred to elsewhere as the fusiform body area (Peelen and Downing, 2005; Schwarzlose et al., 2005).

Face-selective ROIs (faces > others) were observed on the inferior occipital gyrus (IOG; $N = 11$), corresponding to the occipital face area (Gauthier et al., 2000), and on the posterior (pFus; $N = 12$) and mid (mFus; $N = 12$) lateral fusiform gyrus, collectively corresponding to the fusiform face area (Kanwisher et al., 1997).

Place-selective ROIs (places > others) were observed near the transverse occipital sulcus (TOS; $N = 12$) as described in previous studies (Hasson et al., 2003), collateral sulcus (CoS; $N = 12$), corresponding to the parahippocampal place area (Epstein and Kanwisher, 1998), and medial parietal/retrosplenial cortex (RSC; $N = 10$), corresponding to the retrosplenial complex (O'Craven and Kanwisher, 2000).

ROIs were considered in three anatomical sections arranged from posterior to anterior corresponding to putative stages of the visual hierarchy: LOTC (IOS-characters, LOS-bodies, IOG-faces, and TOS-places), posterior VTC (pOTS-characters, ITG-bodies, and pFus-faces), and mid VTC (mOTS-characters, OTS-bodies, mFus-faces, and CoS-places). Two additional regions were located outside these anatomical expanses (MTG-bodies and RSC-places).

Estimating responses to experimental conditions. We ran a GLM using the experimental conditions to generate a design matrix that was then convolved with the HRF implemented in SPM (<http://www.fil.ion.ucl.ac.uk/spm>). This procedure was done on the time course of each voxel (units of percentage signal change) and separately for each rate. Response amplitudes (betas) and residual variance of each voxel were estimated from the GLM. The contrast effect size (CES) was measured as the difference in response amplitudes to a preferred category versus all other categories (in units of percentage signal change). Selectivity (t -value) was measured as the ratio of the CES versus its standard error. The standard error was estimated from the residual variance of the GLM without whitening (Worsley, 2001).

To examine the goodness of our model fits across rates, we compared the variance explained by the GLM fits in each voxel for each rate. Across both experiments, this analysis revealed no significant differences in the proportion of variance explained by models fit separately for 1 and 8 Hz data in any ROI (no significant differences, $t_s < 2.01$, $p_s > 0.05$, paired t test for each ROI comparing proportion of variance explained at 1 and 8 Hz; Fig. 2). Goodness of fit of the GLM tended to be highest at 4 Hz in most category-selective ROIs even though these regions were defined using a localizer with stimuli presented at 1 Hz. This fit at 4 Hz was significantly better than in other rates in some ROIs in Experiment 1 (main effect of rate, $F_s > 2.99$, $p_s < 0.05$, one-way ANOVA on proportion of variance explained by GLM for each ROI; Fig. 2A, thick lines) but not in any ROI in Experiment 2 (no main effect of rate, $F_s < 2.49$, $p_s > 0.05$; Fig. 2A, thin lines). These findings rule out the possibility that our results reflect differences in model fits between fast (8 Hz) and slow (1 Hz) presentation rates or between short (2 s) and long (8 s) trial durations.

Evaluating the effect of rate on category selectivity. We measured the effect of rate on category selectivity using two different repeated-measures ANOVAs. Category selectivity was defined as the mean voxel t -value of preference to one category versus others, averaged across voxels in a given ROI. The first ANOVA was conducted to determine whether selectivity varies differentially as a function of rate across regions located in the same putative hierarchical tier. This ANOVA included factors of rate and preferred domain using data from subjects with all regions defined in a particular anatomical section for repeated-measures analyses (anatomical sections—LOTC: IOS-characters, LOS-bodies, IOG-faces, and TOS-places; posterior VTC: pOTS-characters, ITG-bodies, and pFus-faces; mid VTC: mOTS-characters, OTS-bodies, mFus-faces, and CoS-places; Fig. 3B). To determine the reproducibility

of domain-specific effects of rate on selectivity after controlling low-level stimulus properties, we directly compared selectivity data from Experiments 1 and 2 for sets of regions in the same anatomical section using a three-way ANOVA with factors of rate, preferred domain, and experiment. The second ANOVA was conducted to determine whether selectivity varies differentially as a function of rate across regions with the same category preference, but in different hierarchical tiers (preferred categories—characters: IOS, pOTS, and mOTS; bodies: LOS, ITG, MTG, and OTS; faces: IOG, pFus, and mFus; places: TOS, CoS, and RSC; Fig. 4).

To gain a finer-grained estimate of the presentation rate producing maximal selectivity in a given region within the range of rates measured in the present study, we next fitted third-order polynomial functions to the temporal frequency tuning curve of each ROI in individual subjects as in previous studies (McKeffe et al., 2007; Gauthier et al., 2012). We then identified the rate corresponding to the local peak of the function within the 1–8 Hz range in which we presented our stimuli (Table 1, Selectivity). We compared the distributions of these local maxima across regions in the same anatomical section and across regions with the same category preference using a series of one-way ANOVAs and data from all subjects.

Evaluating the effect of rate on response amplitude. Selectivity is affected by (1) differences in response amplitudes across conditions and (2) the noise level. Consequently, we measured the effect of rate on response amplitudes to preferred and non-preferred categories in each ROI and the effect of rate on the CES or the difference in amplitude between preferred and non-preferred stimuli. We first determined whether the effect of presentation rate on response amplitude differs for preferred and non-preferred stimuli in sets of regions preferring the same category using a series of three-way repeated-measures ANOVAs with factors of rate, preference, and ROI separately for Experiment 1 (Fig. 5) and Experiment 2 (Fig. 7). We then tested whether this interaction significantly differed between experiments using a series of four-way ANOVAs with factors of rate, preference, ROI, and experiment. Likewise, we determined the consistency of rate modulations for preferred stimuli across experiments using a three-way ANOVA with factors of rate, ROI, and experiment. For Experiment 2, we also examined if the effect of rate on response amplitudes in a given ROI varied across each individual stimulus domain using a two-way repeated-measures ANOVA with factors of rate and domain for each ROI (Fig. 6).

The same curve-fitting procedures and statistical analyses used for selectivity data were also applied to amplitude data to estimate the rate eliciting peak responses to the preferred category (Table 1, Preferred) as well as peak CES (Table 1, CES) for each ROI. To further quantify these effects, we compared these additional metrics of temporal capacity across regions processing different domains in the same hierarchical tier and across regions processing the same domain in different hierarchical tiers using a series of one-way ANOVAs. Results using amplitudes to the preferred category or CES (Table 1, CES) were similar (no significant differences, $t_{(12)} = 0.50$, $p > 0.05$, paired t test comparing estimated peak rates derived from amplitude to the preferred category and CES computed at the subject level and then averaged across subjects for each ROI in Experiment 1). Therefore, we focus our analysis on response amplitudes to the preferred category in the main text, but present all data in Table 1.

Finally, to assess the level of consistency between different neural measures of temporal capacity within and across experiments, we compared the capacity estimates derived from various metrics (peak amplitude to the preferred category, peak CES, and peak selectivity) in Experiments 1 and 2 among sets of regions located in the same anatomical section using a series of three-way ANOVAs with factors of preferred domain, experiment, and metric. Complementary analyses were then performed for sets of regions with the same category preference using three-way ANOVAs with factors of ROI, experiment, and metric.

Behavioral testing

Eight of the 12 subjects participated in a behavioral experiment conducted on a different day outside the scanner to estimate subjects' behavioral encoding capacities (Fig. 8). Encoding capacities were measured as a function of rate for a subset of the categories used in the fMRI experiments. These categories were as follows: characters, bodies, faces, and

places. Subjects viewed trials comprised of three stages: (1) an encoding phase consisting of an RSVP sequence of eight stimuli from a single category presented at a single rate of 1 Hz, 2 Hz, 4 Hz, or 8 Hz; (2) a 3 s delay containing a blank screen; and (3) a probe image (Fig. 8A). Subjects were instructed to indicate if the probe image occurred in the preceding sequence or not (0.5 probability). Subjects participated in four runs, each consisting of 64 trials at a single rate with 16 trials per category. The order of runs at different rates was counterbalanced across subjects. In Figure 8B, we plot accuracy for each domain as a function of the presentation rate during encoding.

We chose this type of behavioral test because it requires the subject to process all eight stimuli during the encoding phase. In contrast, having a probe at the beginning of the trial would not require the subjects to process all images during encoding and pilot behavioral testing indicated that subjects' performance was at ceiling for all rates, consistent with prior findings (Potter and Levy, 1969). Ideally, we would have had subjects respond and report their percept for each and every stimulus. However, this is impractical for rates of 4 and 8 Hz, which are faster than the response times of typical subjects.

Results

Are temporal processing capacities in high-level visual cortex domain general or domain specific?

We examined the effect of presentation rate on category selectivity and tested if (1) presentation rate affects the degree of category selectivity, (2) regions in the same anatomical location and processing stage of the ventral stream have the same processing capacity (domain-general hierarchy hypothesis), or (3) regions processing the same category in different anatomical locations have the same processing capacity (domain-specific temporal hypothesis). To address these questions, we independently defined functional ROIs selective to characters, bodies, faces, and places in each subject and measured the degree of selectivity in each ROI as a function of presentation rate. We divided the ventral stream into three anatomical sections: LOTC, posterior VTC, and mid VTC (Fig. 3A). Each section corresponds to a putative tier of the visual hierarchy and contains a cluster of character, body, face, and place ROIs (with the exception of posterior VTC that did not contain a place-selective ROI).

We find that presentation rate significantly modulates the degree of category selectivity in all ventral stream ROIs regardless of their location in the hierarchy or category preference (main effect of rate, $F_s > 25.28$, $p_s < 0.001$, repeated-measures ANOVA for regions in each anatomical section with factors of rate and preferred domain). Thus, a property of category-selective regions is that they are more selective to their preferred category when presented at a certain rate.

Contrary to predictions of the hierarchy hypothesis, data from Experiment 1 also reveal that each anatomical section contains regions with different temporal capacities (rate by domain interaction, $F_s > 6.16$, $p_s < 0.001$, repeated-measures ANOVA for regions in each anatomical section with factors of rate and preferred domain; Fig. 3B). In LOTC, for example, regions selective to bodies, faces, and places show highest selectivity at 4 Hz, but the character-selective IOS exhibits maximal selectivity at 2 Hz (Fig. 3B; LOTC). Likewise in mid VTC, regions selective to characters, bodies, faces, and places are anatomically adjacent (Fig. 3A, mid VTC) yet show different peaks and troughs in their temporal processing characteristics (Fig. 3B, mid VTC). Moving from medial to lateral: CoS-places and mFus-faces have comparable selectivity at 2 and 4 Hz with a decline at 8 Hz, OTS-bodies reveals higher selectivity at 4 and 8 Hz than 1 and 2 Hz, and mOTS-characters shows peak selectivity at 2 Hz with a decline at 4 Hz.

Is temporal processing capacity organized hierarchically among sets of regions processing the same domain?

To further discount the hierarchy hypothesis, we must rule out the possibility that domain-specific temporal processing capacity systematically declines across the ventral stream hierarchy when considering sets of regions with the same category preference. Therefore, for each ROI and subject we estimated the presentation rate that produced maximal selectivity as a neural marker of temporal capacity. We then examined if this peak varied across regions in the same anatomical section selective to different categories and across regions at different stages of the hierarchy processing the same domain.

Our results indicate that the rates estimated to elicit maximal selectivity are similar across regions processing the same domain across anatomical locations (no main effect of ROI, $F_s < 0.89$, $p_s > 0.05$, ANOVA on peak rates estimated from selectivity for sets of regions with the same category preference; Fig. 4, dashed horizontal bars), but vary considerably across regions processing different domains at each stage of the hierarchy (main effect of ROI, $F_s > 15.03$, $p_s < 0.001$, ANOVA on peak rates estimated from selectivity for sets of regions in each anatomical section). As shown in Table 1, regions processing characters show maximal selectivity at a similar rate in Experiment 1 (IOS: 1.94 ± 0.15 Hz, average \pm SEM; pOTS: 1.97 ± 0.14 Hz; mOTS: 2.07 ± 0.30 Hz). The same is true for regions processing bodies (LOS: 5.12 ± 0.14 Hz; ITG: 4.84 ± 0.41 Hz; MTG: 4.79 ± 0.37 Hz; OTS: 5.10 ± 0.55 Hz), faces (IOG: 2.98 ± 0.39 Hz; pFus: 3.20 ± 0.36 Hz; mFus: 3.09 ± 0.30 Hz), or places (TOS: 2.28 ± 0.35 Hz; CoS: 3.07 ± 0.28 Hz; RSC: 2.65 ± 0.35 Hz). Consistent with these findings, selectivity does not differ as a function of rate across regions processing the same domain (no rate by ROI interaction, $F_s < 2.22$, $p_s > 0.05$, repeated-measures ANOVA for sets of regions with the same category preference with factors of rate and ROI; Fig. 4), providing further evidence against the hierarchy hypothesis and in support of domain-specific temporal processing capacities in high-level visual cortex.

How does presentation rate affect response amplitude?

In principle, selectivity for a preferred category (e.g., a t -value of face selectivity in face-selective regions) and response amplitudes to a preferred category (e.g., responses to faces in face-selective regions) are closely related. In practice, the presentation rate eliciting peak selectivity may differ from the presentation rate eliciting maximal amplitude if the effect of rate differs across responses to preferred and non-preferred stimuli in that region.

In fact, we find differential effects of rate on responses to preferred and non-preferred stimuli in all category-selective regions in Experiment 1 (Fig. 5). Comparing the effect of rate on response amplitudes to preferred and non-preferred stimuli among sets of regions selective to the same category, we find that rate-dependent modulation of amplitude differs for the preferred category and non-preferred categories [rate by preference interaction, $F_s > 4.55$, $p_s < 0.05$, repeated-measures ANOVA for sets of regions with the same category preference with factors of rate, preference (preferred/non-preferred stimuli), and ROI, but no significant three-way interaction, $F_s < 0.93$, $p_s > 0.05$]. That is, presentation rate generally has a more prominent effect on response amplitudes to the preferred than non-preferred stimuli.

Closely replicating the selectivity data, the rate producing the peak response amplitude for the preferred category does not differ across regions processing the same domain (no main effect of

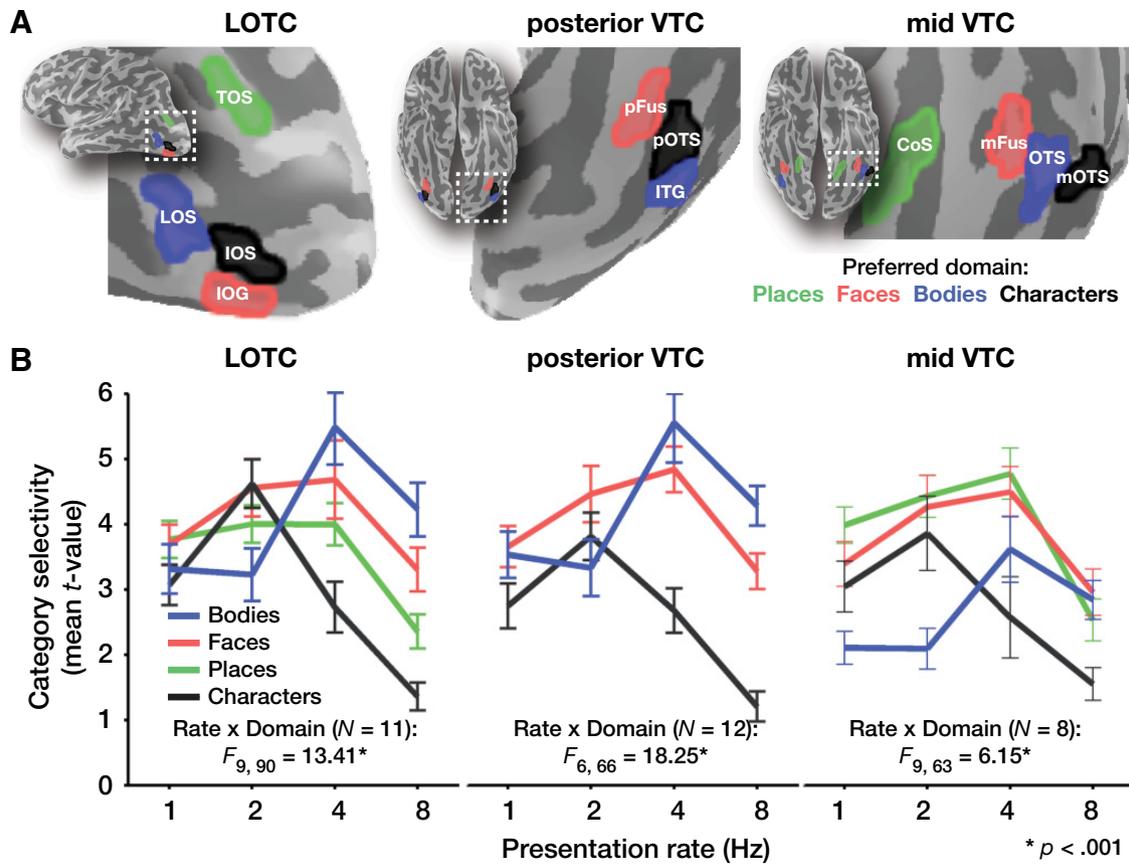


Figure 3. Temporal processing capacity is domain specific. **A**, Functional regions of interest in LOTC, posterior VTC, and mid VTC selective to characters, bodies, faces, and places defined in individual subjects with regions in each anatomical section shown in an example subject. **B**, Mean *t*-value of voxels in each region for the contrast of its preferred category > others as a function of presentation rate in Experiment 1. Error bars indicate SEM across subjects. See data analysis section in the Methods for definition of acronyms.

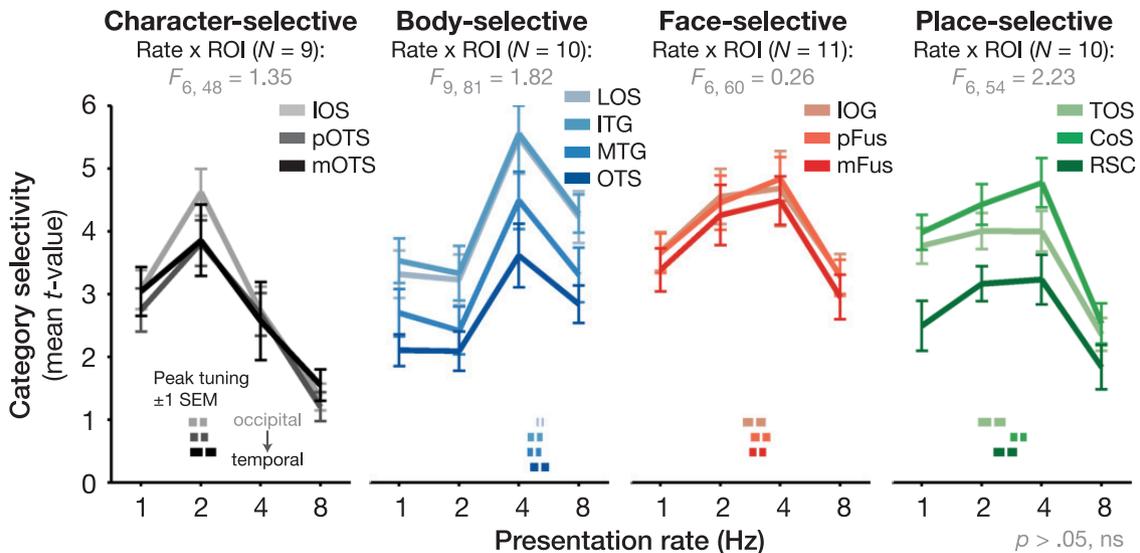


Figure 4. Temporal processing capacity is consistent across regions processing the same domain. Mean selectivity (*t*-value) of voxels in each ROI (preferred category > others). Error bars indicate SEM across subjects. Horizontal bars below curves represent estimates of the average rate producing peak selectivity for each ROI (white dashes) ± 1 SEM across subjects (colored bars).

ROI, $F_s < 0.72$, $p_s > 0.05$, ANOVA on peak rates estimated from amplitude to the preferred category for sets of regions with the same category preference; Fig. 5, dashed horizontal bars). In contrast, the rate producing maximal response to the region's preferred category varies significantly across regions selective to different domains at each stage of the hierarchy (main effect of

ROI, $F_s > 6.03$, $p_s < 0.01$, ANOVA on peak rates estimated from amplitude for regions in each anatomical section). Likewise, the rank ordering of processing capacities associated with different domains observed in selectivity data (Fig. 4) is replicated in the amplitude of response for the preferred category and for the CES (difference in response amplitudes between preferred and non-

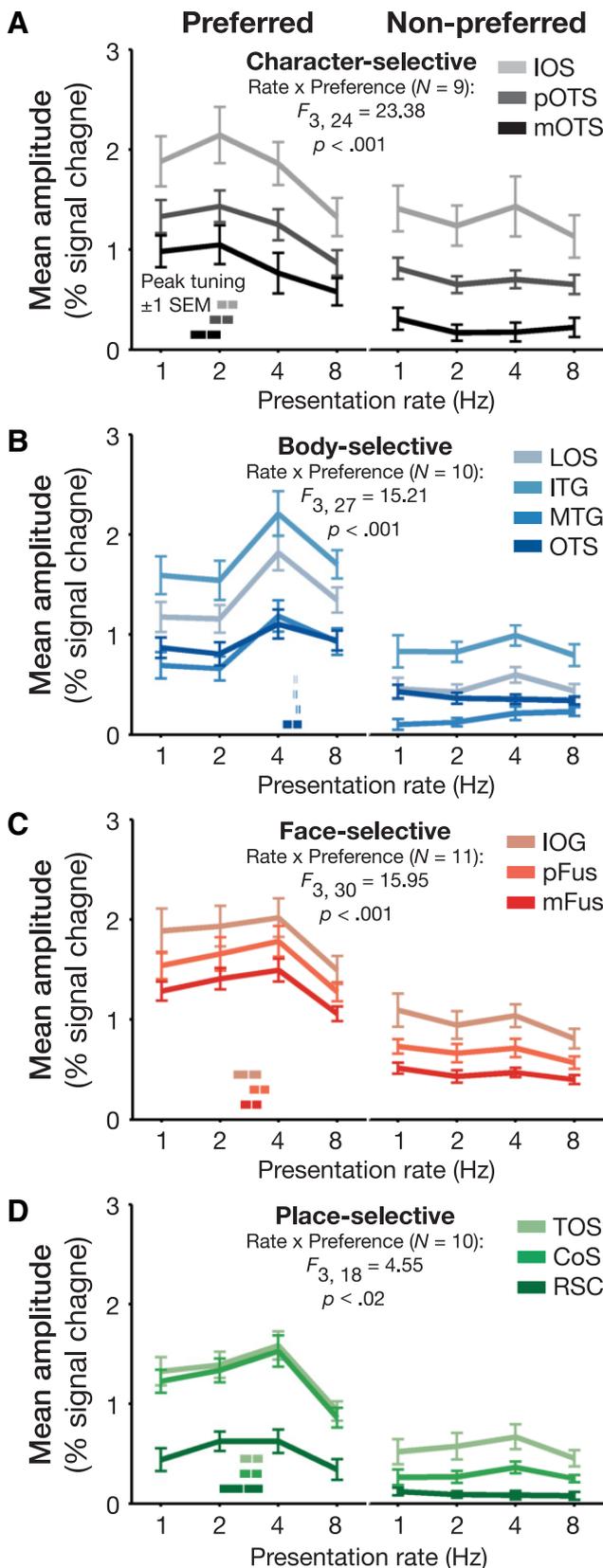


Figure 5. Differential effects of rate on response amplitudes to preferred and non-preferred categories. Mean response amplitudes (percentage signal change) to the preferred category (left) and non-preferred category (right) as a function of presentation rate in Experiment 1. **A**, Character-selective ROIs. **B**, Body-selective ROIs. **C**, Face-selective ROIs. **D**, Place-selective ROIs. Error bars indicate SEM across subjects. Horizontal bars below responses to preferred categories represent estimates of the average rate producing peak amplitude for each ROI (white dashes) ± 1 SEM across subjects (colored bars).

preferred stimuli) as summarized in Table 1. That is, regardless of the metric used to estimate temporal capacity, we find that the presentation rates estimated to elicit peak responses are slowest for regions processing characters, intermediate among regions preferring places and faces, and fastest for regions processing bodies.

Are domain-specific temporal frequencies in high-level visual cortex driven by low-level differences among stimulus categories?

Low-level differences among images of different categories may contribute to the domain-specific temporal processing capacities that we observe in high-level visual cortex. To address this concern, we conducted a second control experiment in eight of the subjects who participated in Experiment 1 using a modified stimulus set designed to minimize categorical differences across multiple low-level properties including contrast, luminance, within-category image similarity, visual field coverage, and spatial frequency (see Materials and Methods; Fig. 1C). Analyses of Experiment 2 data indicate that it is unlikely that low-level image properties drive capacity differences across domains.

First, to validate the low-level controls implemented in Experiment 2, we examined V1 responses. There are no significant differences in the response amplitudes of V1 to different stimulus categories across rates (no main effect of domain, $F_{(3,21)} = 1.46$, $p > 0.05$, repeated-measures ANOVA with factors of rate and domain; Fig. 6, V1), and we do not find variable effects of rate on V1 responses for different categories (no rate by domain interaction, $F_{(9,63)} = 1.79$, $p > 0.05$). In contrast to V1, category-selective regions illustrate higher response amplitudes to particular domains across rates (main effect of domain, $F_s > 16.26$, $ps < 0.001$, repeated-measures ANOVA with factors of rate and domain for each ROI; Fig. 6). Further, category-selective regions, even those in LOTC that are considered lower in the hierarchy compared with VTC, show a significant rate by domain interaction ($F_s > 2.05$, $ps < 0.05$; Fig. 6).

Second, despite changing our stimuli in Experiment 2, we find that presentation rate significantly modulates category selectivity in a domain-specific manner. Each hierarchical tier contains regions with different temporal capacities (rate by domain interaction, $F_s > 2.10$, $ps < 0.03$, ANOVA for regions in each anatomical section with factors of rate, preferred domain, and experiment), and these effects are consistent across experiments (no three-way interaction between rate, domain, and experiment, $F_s < 0.92$, $ps > 0.05$).

Third, examination of response amplitudes among regions selective to the same category shows that presentation rate differentially modulates responses to preferred and non-preferred stimuli (rate by preference interaction, $F_s > 2.89$, $ps < 0.04$, ANOVA for sets of regions with the same category preference with factors of rate, preference, ROI, and experiment). These effects are similar across experiments (no three-way interaction between rate, preference, and experiment, $F_s < 1.44$, $ps > 0.05$; compare Figs. 5 and 7). Similar to Experiment 1, we also find that responses to the preferred category show the same rate modulation among regions preferring the same category (main effect of rate, $F_s > 9.32$, $ps < 0.001$, ANOVA for sets of regions with the same category preference with factors of rate, ROI, and experiment, but there are no two-way or three-way interactions, $F_s < 2.73$, $ps > 0.05$; Fig. 7, Preferred).

These analyses validate that the domain-specific temporal processing capacities observed in Experiment 1 are reproduc-

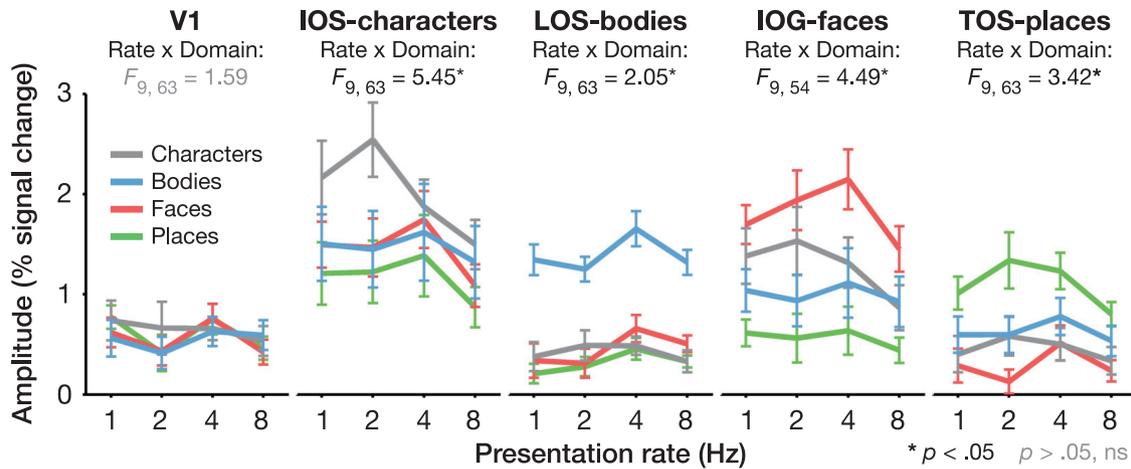


Figure 6. Domain specificity is not driven by low-level image properties or inherited from V1. Mean response amplitude for each domain as a function of presentation rate in V1 and category-selective regions in LOTC. Error bars indicate SEM across eight subjects.

Table 1. Peak estimates derived from different capacity metrics retain the temporal rank ordering across domains

	Experiment 1			Experiment 2			Average ^d
	Preferred	CES	Selectivity	Preferred	CES	Selectivity	
Characters	1.99 ^b	2.08 ^b	1.99 ^b	1.59 ^b	1.93 ^b	2.14 ^b	1.95 ^b
IOS	2.20 (0.26)	1.93 (0.16)	1.94 (0.15)	1.50 (0.15)	1.54 (0.06)	1.50 (0.07)	1.77
pOTS	2.05 (0.28)	2.04 (0.16)	1.97 (0.14)	1.57 (0.15)	1.55 (0.12)	1.55 (0.13)	1.79
mOTS	1.72 (0.30)	2.26 (0.33)	2.07 (0.30)	1.70 (0.23)	2.69 (0.90)	3.38 (0.90)	2.30
Places	2.83 ^b	2.42 ^b	2.67 ^b	2.76 ^b	2.31 ^b	2.44 ^b	2.57 ^b
TOS	2.93 (0.38)	2.54 (0.37)	2.28 (0.35)	2.80 (0.45)	2.57 (0.54)	2.50 (0.59)	2.60
CoS	2.91 (0.37)	2.62 (0.33)	3.07 (0.28)	2.72 (0.54)	2.04 (0.43)	2.37 (0.50)	2.62
RSC	2.66 (0.66)	2.10 (0.29)	2.65 (0.35)				2.47
Faces	2.96 ^b	3.08 ^b	3.09 ^b	3.12 ^b	3.57 ^b	3.99 ^b	3.30 ^b
IOG	2.79 (0.45)	2.99 (0.38)	2.98 (0.39)	3.26 (0.51)	3.33 (0.50)	4.51 (0.76)	3.31
pFus	3.20 (0.37)	3.26 (0.31)	3.20 (0.36)	3.37 (0.47)	3.42 (0.48)	3.50 (0.50)	3.33
mFus	2.90 (0.35)	3.00 (0.32)	3.09 (0.30)	2.74 (0.62)	3.95 (0.44)	3.95 (0.72)	3.27
Bodies	4.90 ^b	4.92 ^b	4.96 ^b	4.15 ^b	4.21 ^b	4.04 ^b	4.53 ^b
LOS	4.91 (0.09)	5.06 (0.12)	5.12 (0.14)	4.10 (0.49)	4.09 (0.81)	4.42 (0.76)	4.62
ITG	4.88 (0.10)	4.84 (0.41)	4.84 (0.41)	3.66 (0.66)	4.06 (0.94)	4.13 (0.95)	4.40
MTG	5.05 (0.12)	4.76 (0.36)	4.79 (0.37)	4.96 (0.55)	4.14 (0.97)	4.28 (0.96)	4.66
OTS	4.74 (0.51)	5.00 (0.50)	5.10 (0.55)	3.88 (0.94)	4.56 (0.83)	3.33 (0.68)	4.44

The mean presentation rate estimated to elicit peak amplitude to the preferred category, contrast effect size (CES), and selectivity in individual subjects from Experiments 1 and 2, with values in parentheses denoting the SEM of this peak estimate across subjects. ^aAverage peak estimates across metrics and experiments for each region (see rightmost column). ^bAverage peak estimates across regions selective to the same domain for each metric. Peaks are not estimated for RSC-places in Experiment 2 because of low signals.

ible and are not driven by low-level differences among stimuli or inherited from V1.

Are domain-specific temporal processing capacity estimates consistent across experiments and metrics?

To establish the overall impact of our stimulus manipulations, we compared capacity estimates across metrics and experiments (Table 1). Consistent with prior analyses, the largest effect is that temporal capacity estimates differ between regions selective to different domains at each stage of the processing hierarchy. This differentiation is consistent across experiments and metrics (main effect of domain, $F_s > 22.60$, $ps < 0.001$, ANOVA on all peak rate estimates from Experiments 1 and 2 for regions in the same anatomical section with factors of preferred domain, experiment, and metric, but no main effect of experiment or metric, $F_s < 3.80$, $ps > 0.05$, and no two-way or three-way interactions, $F_s < 0.61$, $ps > 0.05$).

Nevertheless, controlling for low-level differences among stimuli does produce some numerical variations of peak capacity estimates within a domain. For example, face-selective

ROIs show slightly higher capacities in Experiment 2 compared with Experiment 1, while body-selective ROIs show an opposite trend (main effect of experiment, $F_s > 5.87$, $ps < 0.05$, ANOVA on all peak rate estimates from Experiments 1 and 2 for body- and face-selective regions separately with factors of ROI, experiment, and metric; Table 1). Also, some character- and place-selective regions exhibit shifts in capacity estimates across experiments (ROI by experiment interaction, $F_s > 5.23$, $p < 0.01$, ANOVA on all peak rate estimates from Experiments 1 and 2 for character- and place-selective regions separately with factors of ROI, experiment, and metric, but no main effect of experiment, $F_s < 1.63$, $ps > 0.05$).

Critically, despite numerical shifts within domains, the ranking of processing capacities for regions processing different domains across experiments is consistent: on average and across all metrics, character-selective IOS, pOTS, and mOTS are the slowest three regions; body-selective LOS, ITG, MTG, and OTS are the fastest four regions; and all face-selective regions are intermediate, but faster than place-selective regions (Table 1).

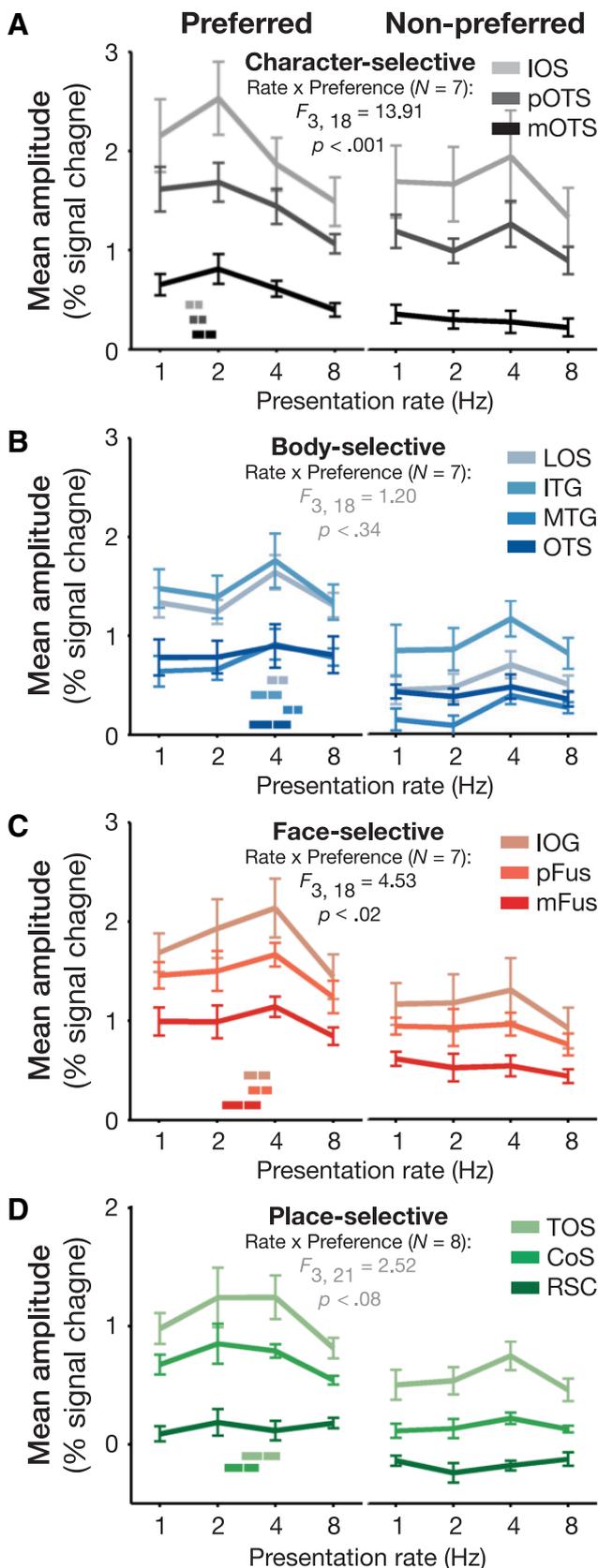


Figure 7. Response amplitudes to preferred and non-preferred categories in Experiment 2. Mean response amplitudes (percentage signal change) to the preferred category (left) and non-preferred category (right) as a function of presentation rate in Experiment 2. **A**, Character-selective ROIs. **B**, Body-selective ROIs. **C**, Face-selective ROIs. **D**, Place-selective ROIs. Considering data from Experiment 2 separately, rate differentially modulates response amplitudes for preferred and non-preferred stimuli in character- and face-selective regions, with a similar

Does domain-specific temporal processing capacity have a behavioral impact?

We hypothesized that if domain-specific temporal processing capacities serve as bottlenecks for perception, there should also be category-specific differences in temporal encoding capacity (i.e., behavioral measures of temporal capacity). To test this hypothesis, we conducted a behavioral experiment during which our subjects viewed encoding stimuli varying in both presentation rate and domain. As in the fMRI experiments, all encoding trials contained eight stimuli, but trials varied in their presentation rate (Fig. 8A; see Materials and Methods). To evaluate temporal encoding capacity, subjects were probed with an image 3 s after the end of the encoding interval and reported if the probe had appeared (or not) in the preceding encoding interval. We reasoned that accuracy should be high for stimuli presented at rates at or below encoding capacity, but would decline at rates exceeding the encoding capacity.

Notably, rate affects encoding performance both when considering accuracy for each category individually (main effect of rate, $F_s > 5.77$, $p_s < 0.05$, ANOVA on accuracy for each domain separately; Fig. 8B) and when analyzing performance across all domains (main effect of rate, $F_{(3,21)} = 57.18$, $p < 0.001$, repeated-measures ANOVA with factors of rate and domain). Interestingly, the effect of rate on encoding capacity differs across domains (rate by domain interaction, $F_{(9,63)} = 4.82$, $p < 0.001$), and several aspects of behavioral encoding capacity mirror the temporal processing capacity of high-level visual cortex. These effects are apparent in three ways. First, behavioral and neural capacity measures for characters peak at presentation rates at ~2 Hz and drop sharply at faster rates. Second, both accuracy and cortical capacity for bodies are highest at 4 Hz. Third, encoding of faces at 8 Hz declines less than encoding of places at 8 Hz, consistent with the observation of higher selectivity to faces than places at this rate in LOTC (Fig. 3).

Despite these striking consistencies, we also observe some deviations between behavioral encoding and neural processing capacities. For example, behavioral encoding capacity for faces and places is highest at ~1 Hz, which is lower than the processing capacity of these stimuli in high-level visual cortex both in selectivity and amplitude measures. Overall, these behavioral data provide striking evidence for domain-specific encoding capacities that are largely consistent with our fMRI findings.

Discussion

According to present hierarchical models, the entirety of high-level visual cortex has a single, domain-general temporal processing capacity that is slower than earlier visual regions. Contrary to this model, our results indicate that domain specificity is a better predictor of temporal processing capacity in high-level visual cortex than anatomical location or processing stage in the visual hierarchy. Specifically, body-selective regions illustrate the fastest capacity, face- and place-selective regions exhibit an intermediate capacity, and regions preferring characters manifest the slowest capacity. In comparison, neural responses within early

trend in place-selective regions but not in body-selective regions. However, the rate by preference interaction is not significantly different between experiments in any of these four domains (no three-way interaction between rate, preference, and experiment, $F_s < 1.44$, $p_s > 0.05$, ANOVA on regions with the same domain preference, with factors of rate, preference, ROI, and experiment). Error bars indicate SEM across subjects. Horizontal bars below responses to preferred categories represent estimates of the average rate producing peak amplitude for each ROI (white dashes) ± 1 SEM across subjects (colored bars).

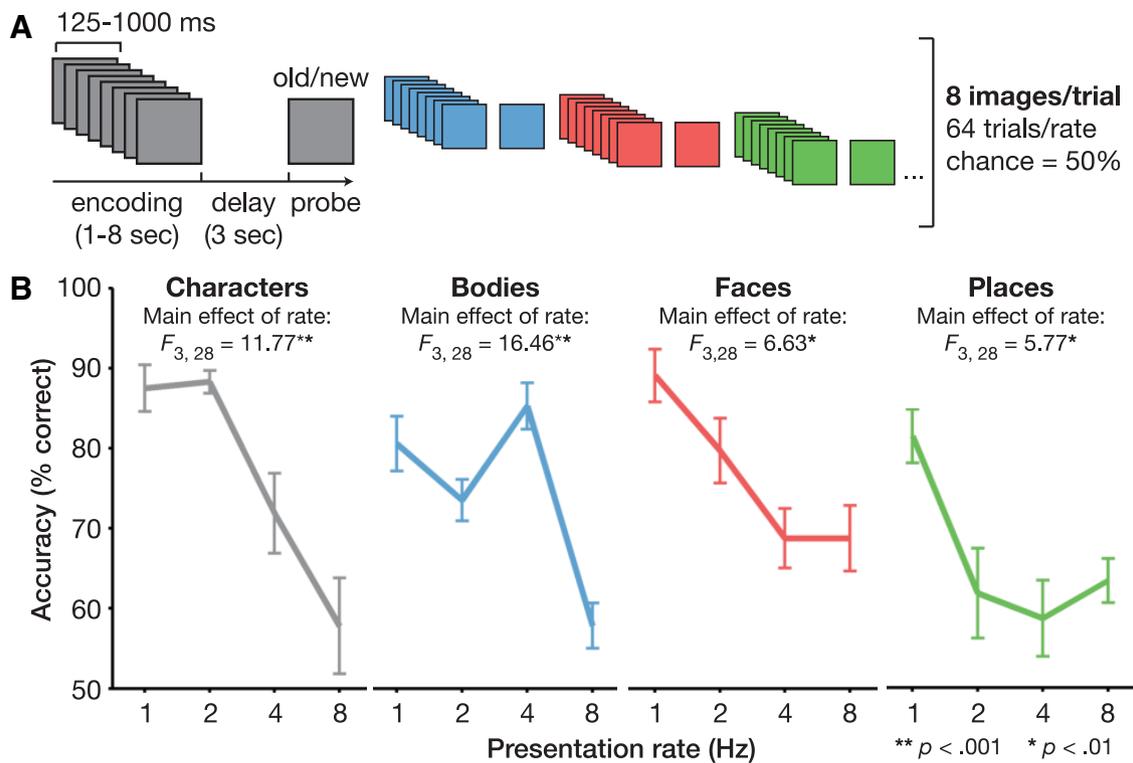


Figure 8. Behavioral encoding capacity is domain specific. **A**, Partial stimulus sequence from behavioral study. Images in a given run were displayed at a single rate of 1 Hz, 2 Hz, 4 Hz, or 8 Hz in trials of eight stimuli from a single category. Participants indicated whether or not a test probe stimulus belonging to the same category appeared in the previous trial following a 3 s blank retention period. **B**, Accuracy of behavioral performance as a function of rate for each domain. Error bars indicate SEM across eight subjects.

visual areas like V1 do not differ across domains as a function of presentation rate. Crucially, our findings are not because of low-level image differences among exemplars of different categories, are behaviorally relevant, and provide evidence for parallel domain-specific encoding capacities.

Estimates of temporal capacity are consistent across amplitude and selectivity metrics

Our data reveal a domain-specific capacity across selectivity, amplitude, and contrast effect size metrics with a consistent rank ordering of capacities of different domains across metrics and experiments (Table 1). Across both experiments, we also observe differential effects of rate on response amplitudes to preferred and non-preferred categories in category-selective ROIs. This difference is a consequence of two factors. First, in some ROIs, especially those in mid VTC, responses to non-preferred categories are low. Thus, we do not find a rate effect modulating the response to non-preferred stimuli (Fig. 5), perhaps because of a floor effect. Second, in other ROIs, especially those in LOTC, responses to non-preferred stimuli are lower than to preferred stimuli, but show a differential rate modulation (Fig. 6). This suggests the intriguing possibility that processing capacity in high-level visual cortex is domain specific even within individual ROIs. Alternatively, as we are measuring fMRI signals with voxels of 2.4 mm on a side, and category-selective regions with different preferences also neighbor one another, it is possible that the differential rate effects across categories in an ROI occur from partial voluming effects. Future fMRI-guided neurophysiology experiments may shed light on these alternatives.

While we provide the first evidence for differences in temporal capacity between body- and character-selective regions, our estimates of similar capacity for face- and place-selective regions

agree with prior studies (McKeeff et al., 2007; Gauthier et al., 2012; Gentile and Rossion, 2014). Our capacity estimates are slightly slower than these previous studies, perhaps because we controlled the number of stimuli across rates (these factors were previously confounded; see Materials and Methods), but the consistency of our findings across amplitude, CES, and selectivity metrics and their stability across experiments illustrates that our measurements are reliable. Nevertheless, we are cognizant that multiple factors may impact capacity. Bottom-up factors such as reduced contrast or increased noise may reduce neural processing capacity. Top-down factors such as familiarity may enhance capacity. For example, high-frequency words might be processed faster than pseudowords. Future experiments can be used to evaluate the effects of additional bottom-up and top-down factors on neural processing capacity.

Visual encoding capacity is coupled with neural capacity in high-level visual cortex

The results of our behavioral experiment reveal that encoding capacities for complex visual stimuli deteriorate at distinct presentation rates for different domains, but we find that performance generally decreases for all categories at 8 Hz. Nevertheless, this should not be interpreted as a fixed capacity limitation on visual processing given the results of previous studies showing that behavioral capacity varies considerably across tasks (Potter, 1976; Potter and Fox, 2009) and is modulated by a range of high-level factors including stimulus complexity and familiarity (Näsänen et al., 2006).

Considering the variety of factors known to modulate temporal capacity, the consistencies we observe between behavioral and neural capacities are striking, but also not unreasonable given the actions we took to eliminate categorical biases in our stimulus set.

Across categories, images were matched on low-level properties such as contrast, luminance, and spatial frequency, and on high-level parameters such as familiarity and memorability. When controlling for these factors, our results suggest that temporal capacities for encoding stimuli with enough detail to allow accurate within-category recognition are closely coupled with neural capacity estimates in regions preferring characters and bodies. Nevertheless, it is possible that neural capacity estimates reflecting the rate at which selectivity or response amplitudes fall to floor (instead of the rate at which they peak) are coupled with capacities for more simple perceptual operations. This hypothesis can be examined in future research comparing behavioral capacities for encoding compared with other tasks such as detection and categorization (Grill-Spector and Kanwisher, 2005; Fei-Fei et al., 2007).

Temporal encoding capacity in high-level visual cortex is domain specific

At the theoretical level, our results suggest a new view of high-level visual cortex in which regions processing related content have common temporal processing capacities that are not hierarchically organized from LOTC to VTC. In general, this finding is consistent with the notion that regions selective to particular categories operate as networks that are synchronized to avoid a processing bottleneck. While the underlying source of capacity differences across high-level visual cortex is uncertain, we hypothesize that two factors may contribute: (1) regularities in the temporal characteristics of exemplars of various domains in the real world and (2) the amount of time required to process domain-specific information.

For instance, lower temporal processing capacities of character-selective regions may be related to the stationary nature of characters. These lower capacities may also be due to the recruitment of additional cortical areas processing linguistic or semantic information when viewing written characters. On the other hand, the high capacity of body-selective regions may result from the necessity to perform fast computations of rapidly evolving, nonrigid body movements. Consistent with this speculation, computational models of biological motion perception posit the existence of “snapshot” neurons selective to brief frames of action sequences (Giese and Poggio, 2003). Interestingly, behavioral encoding performance for pseudowords and bodies peaked at the same rates that produced the highest selectivity in ROIs preferring these domains, respectively. This suggests that the temporal processing capacity of these regions may influence the rate of behavioral encoding specifically for items from these categories. Considering other examined domains, both faces and places contain more malleable features than characters, but their large-scale configurations are more rigid than those of bodies (e.g., eyes and clouds remain above most other features of faces and places, respectively). Thus, the present evidence supports the theory that the natural statistics of temporal information within the visual world may generate domain-specific temporal encoding capacities.

Temporal processing capacity contributes to the organization of high-level visual cortex

We propose that temporal processing capacity contributes to the segregation of category-selective regions in high-level visual cortex. It is interesting to speculate how this segregation may develop across the ventral stream hierarchy. From its origin in the retina and lateral geniculate nucleus, the visual system contains segregated temporal channels for faster processing in the magnocellular (M) pathway and slower processing in the parvocellular (P) pathway (Derrington and Lennie, 1984; Schneider et al., 2004).

This segregation is further propagated into cortex in V1 (Sun et al., 2007) and continues downstream to V2 (Munk et al., 1995) and the middle temporal visual area (MT; Maunsell et al., 1990). Thus, it is possible that the propagation of channels with separable temporal processing capacity continues even further and influences the organization of high-level visual cortex.

Other fundamental properties that prevail across the visual system such as segregation of neurons processing foveal and peripheral stimuli have also been proposed to constrain the topology of high-level visual cortex. For example, viewing biases that are associated with particular domains (e.g., people tend to foveate on faces and characters) may generate the particular eccentricity biases associated with category-selective regions (Levy et al., 2001; Hasson et al., 2002, 2003; Kay et al., 2015). Similar to the eccentricity bias hypothesis (Malach et al., 2002), the topology of category-selective regions may also be determined by “temporal biases” of stimuli in the natural world that are mapped onto regions with differential contribution of M and P inputs. These are not mutually exclusive organizational principles. On the contrary, we propose that the functional organization of high-level visual cortex may be constrained both by spatial and temporal capacity limitations, the combination of which is domain specific. Thus, we suggest that both eccentricity and temporal biases may contribute to the functional organization of high-level visual cortex. For example, regions selective to faces and places have similar temporal processing capacity, but remain segregated as a result of their association with different eccentricity biases and viewing patterns. On the other hand, regions processing faces and characters have similar foveal biases but distinct temporal dynamics, which may guide segregation of associated neural processing in lateral VTC. Our results suggest that temporal processing capacity is one of multiple dimensions (Huth et al., 2012) contributing to segregation of functional regions, and that temporal capacity is a characteristic of domain-specific networks in high-level visual cortex.

Notes

Supplemental material for this article is available at vpnl.stanford.edu/flloc/. This material contains stimuli and code for a functional localizer experiment to define category-selective visual regions. This material has not been peer reviewed.

References

- Ben-Shachar M, Dougherty RF, Deutsch GK, Wandell BA (2011) The development of cortical sensitivity to visual word forms. *J Cogn Neurosci* 23:2387–2399. [CrossRef Medline](#)
- Chance FS, Nelson SB, Abbott LF (1998) Synaptic depression and the temporal response characteristics of V1 cells. *J Neurosci* 18:4785–4799. [Medline](#)
- Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaux MA, Michel F (2000) The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain* 123:291–307. [CrossRef Medline](#)
- Derrington AM, Lennie P (1984) Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *J Physiol* 357:219–240. [CrossRef Medline](#)
- Dilks DD, Julian JB, Paunov AM, Kanwisher N (2013) The occipital place area is causally and selectively involved in scene perception. *J Neurosci* 33:1331–1336a. [CrossRef Medline](#)
- Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. *Science* 293:2470–2473. [CrossRef Medline](#)
- Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. *Neuroimage* 39:647–660. [CrossRef Medline](#)
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601. [CrossRef Medline](#)

- Fei-Fei L, Iyer A, Koch C, Perona P (2007) What do we perceive in a glance of a real-world scene? *J Vis* 7(1):10. [CrossRef Medline](#)
- Foster KH, Gaska JP, Nagler M, Pollen DA (1985) Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J Physiol* 365:331–363. [CrossRef Medline](#)
- Gauthier B, Eger E, Hesselmann G, Giraud AL, Kleinschmidt A (2012) Temporal tuning properties along the human ventral visual stream. *J Neurosci* 32:14433–14441. [CrossRef Medline](#)
- Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000) The fusiform “face area” is part of a network that processes faces at the individual level. *J Cogn Neurosci* 12:495–504. [CrossRef Medline](#)
- Gentile F, Rossion B (2014) Temporal frequency tuning of cortical face-sensitive areas for individual face perception. *Neuroimage* 90:256–265. [CrossRef Medline](#)
- Giese MA, Poggio T (2003) Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* 4:179–192. [CrossRef Medline](#)
- Glazer LS, Jiang X, Riesenhuber M (2009) Evidence for highly selective neuronal tuning to whole words in the “visual word form area.” *Neuron* 62:199–204. [CrossRef Medline](#)
- Grill-Spector K, Kanwisher N (2005) Visual recognition: as soon as you know it is there, you know what it is. *Psychol Sci* 16:152–160. [CrossRef Medline](#)
- Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci* 15:536–548. [CrossRef Medline](#)
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Hasson U, Harel M, Levy I, Malach R (2003) Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37:1027–1041. [CrossRef Medline](#)
- Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A hierarchy of temporal receptive windows in human cortex. *J Neurosci* 28:2539–2550. [CrossRef Medline](#)
- Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. *Trends Cogn Sci* 4:223–233. [CrossRef Medline](#)
- Heller J, Hertz JA, Kjaer TW, Richmond BJ (1995) Information flow and temporal coding in primate pattern vision. *J Comput Neurosci* 2:175–193. [CrossRef Medline](#)
- Huth AG, Nishimoto S, Vu AT, Gallant JL (2012) A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76:1210–1224. [CrossRef Medline](#)
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311. [Medline](#)
- Kay KN, Weiner KS, Grill-Spector K (2015) Attention reduces spatial uncertainty in human ventral temporal cortex. *Curr Biol* 25:595–600. [CrossRef Medline](#)
- Keyser C, Xiao DK, Földiák P, Perrett DI (2001) The speed of sight. *J Cogn Neurosci* 13:90–101. [CrossRef Medline](#)
- Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center-periphery organization of human object areas. *Nat Neurosci* 4:533–539. [Medline](#)
- Liu J, Wandell BA (2005) Specializations for chromatic and temporal signals in human visual cortex. *J Neurosci* 25:3459–3468. [CrossRef Medline](#)
- Malach R, Levy I, Hasson U (2002) The topography of high-order human object areas. *Trends Cogn Sci* 6:176–184. [CrossRef Medline](#)
- Maunsell JH, Nealey TA, DePriest DD (1990) Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *J Neurosci* 10:3323–3334. [Medline](#)
- McKeeff TJ, Remus DA, Tong F (2007) Temporal limitations in object processing across the human ventral visual pathway. *J Neurophysiol* 98:382–393. [CrossRef Medline](#)
- Mukamel R, Harel M, Hendler T, Malach R (2004) Enhanced temporal non-linearities in human object-related occipito-temporal cortex. *Cereb Cortex* 14:575–585. [CrossRef Medline](#)
- Munk MH, Nowak LG, Girard P, Chounlamountri N, Bullier J (1995) Visual latencies in cytochrome oxidase bands of macaque area V2. *Proc Natl Acad Sci U S A* 92:988–992. [CrossRef Medline](#)
- Näsänen R, Ojanpää H, Tanskanen T, Päällysaho J (2006) Estimation of temporal resolution of object identification in human vision. *Exp Brain Res* 172:464–471. [CrossRef Medline](#)
- O’Craven KM, Kanwisher N (2000) Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J Cogn Neurosci* 12:1013–1023. [CrossRef Medline](#)
- Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol* 57:162–178. [Medline](#)
- Peelen MV, Downing PE (2005) Selectivity for the human body in the fusiform gyrus. *J Neurophysiol* 93:603–608. [Medline](#)
- Peelen MV, Downing PE (2007) The neural basis of visual body perception. *Nat Rev Neurosci* 8:636–648. [CrossRef Medline](#)
- Potter MC (1976) Short-term conceptual memory for pictures. *J Exp Psychol Hum Learn* 2:509–522. [CrossRef Medline](#)
- Potter MC, Fox LF (2009) Detecting and remembering simultaneous pictures in a rapid serial visual presentation. *J Exp Psychol Hum Percept Perform* 35:28–38. [CrossRef Medline](#)
- Potter MC, Levy EI (1969) Recognition memory for a rapid sequence of pictures. *J Exp Psychol* 81:10–15. [CrossRef Medline](#)
- Richmond BJ, Optican LM (1990) Temporal encoding of two-dimensional patterns by single units in primate primary visual cortex. II. Information transmission. *J Neurophysiol* 64:370–380. [Medline](#)
- Schmulesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG (1998) Signal timing across the macaque visual system. *J Neurophysiol* 79:3272–3278. [Medline](#)
- Schneider KA, Richter MC, Kastner S (2004) Retinotopic organization and functional subdivisions of the human lateral geniculate nucleus: a high-resolution functional magnetic resonance imaging study. *J Neurosci* 24:8975–8985. [CrossRef Medline](#)
- Schwarzlose RF, Baker CI, Kanwisher N (2005) Separate face and body selectivity on the fusiform gyrus. *J Neurosci* 25:11055–11059. [CrossRef Medline](#)
- Schwarzlose RF, Swisher JD, Dang S, Kanwisher N (2008) The distribution of category and location information across object-selective regions in human visual cortex. *Proc Natl Acad Sci U S A* 105:4447–4452. [CrossRef Medline](#)
- Singh KD, Smith AT, Greenlee MW (2000) Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage* 12:550–564. [CrossRef Medline](#)
- Sun P, Ueno K, Waggoner RA, Gardner JL, Tanaka K, Cheng K (2007) A temporal frequency-dependent functional architecture in human V1 revealed by high-resolution fMRI. *Nat Neurosci* 10:1404–1406. [CrossRef Medline](#)
- Vogels R, Orban GA (1994) Activity of inferior temporal neurons during orientation discrimination with successively presented gratings. *J Neurophysiol* 71:1428–1451. [Medline](#)
- Weiner KS, Grill-Spector K (2010) Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* 52:1559–1573. [CrossRef Medline](#)
- Weiner KS, Grill-Spector K (2011) Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *Neuroimage* 56:2183–2199. [CrossRef Medline](#)
- Weiner KS, Grill-Spector K (2012) The improbable simplicity of the fusiform face area. *Trends Cogn Sci* 16:251–254. [CrossRef Medline](#)
- Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-level image properties: the SHINE toolbox. *Behav Res Methods* 42:671–684. [CrossRef Medline](#)
- Worsley KJ (2001) Statistical analysis of activation images. In: *Functional MRI: an introduction to methods* (Jezzard P, Matthews PM, Smith SM, eds.), pp 251–270. New York: Oxford UP.