

Temporal Limitations in Object Processing Across the Human Ventral Visual Pathway

Thomas J. McKeeff,¹ David A. Remus,¹ and Frank Tong²

¹Department of Psychology, Princeton University, Princeton, New Jersey; and ²Psychology Department, Vanderbilt University, Nashville, Tennessee

Submitted 29 May 2006; accepted in final form 8 May 2007

McKeeff TJ, Remus DA, Tong F. Temporal limitations in object processing across the human ventral visual pathway. *J Neurophysiol* 98: 382–393, 2007. First published May 9, 2007; doi:10.1152/jn.00568.2006. Behavioral studies have shown that object recognition becomes severely impaired at fast presentation rates, indicating a limitation in temporal processing capacity. Here, we studied whether this behavioral limit in object recognition reflects limitations in the temporal processing capacity of early visual areas tuned to basic features or high-level areas tuned to complex objects. We used functional MRI (fMRI) to measure the temporal processing capacity of multiple areas along the ventral visual pathway progressing from the primary visual cortex (V1) to high-level object-selective regions, specifically the fusiform face area (FFA) and parahippocampal place area (PPA). Subjects viewed successive images of faces or houses at presentation rates varying from 2.3 to 37.5 items/s while performing an object discrimination task. Measures of the temporal frequency response profile of each visual area revealed a systematic decline in peak tuning across the visual hierarchy. Areas V1–V3 showed peak activity at rapid presentation rates of 18–25 items/s, area V4v peaked at intermediate rates (9 items/s), and the FFA and PPA peaked at the slowest temporal rates (4–5 items/s). Our results reveal a progressive loss in the temporal processing capacity of the human visual system as information is transferred from early visual areas to higher areas. These data suggest that temporal limitations in object recognition likely result from the limited processing capacity of high-level object-selective areas rather than that of earlier stages of visual processing.

INTRODUCTION

Although observers can recognize a briefly flashed object quite quickly and accurately (Grill-Spector and Kanwisher 2005; Potter and Faulconer 1975; Thorpe et al. 1996), behavioral studies have revealed the capacity-limited nature of visual object recognition (Nothdurft 1993; Tong and Nakayama 1999; Treisman 1988). Rapid serial visual presentation (RSVP) has been used to estimate the rate at which the visual system can process a series of objects. Observers can reliably identify objects at moderate presentation rates of ~8–10 items/s (McMains and Somers 2004; Potter 1975), whereas basic visual changes involving flicker or motion can be detected at rates as high as 30–50 Hz (Kelly 1961, 1979).

Many cognitive theories have been proposed to account for the temporal limitations of visual object recognition. These include proposals that a requisite amount of time is needed to attend to an object (Duncan et al. 1994; Raymond et al. 1992), to classify an

object's appearance as a distinct event (Kanwisher 1991), or to transfer object information into working memory (Chun and Potter 1995; Marois and Ivanoff 2005). Although attention and memory can influence recognition performance at moderate presentation rates, these theories do not explain the global loss in recognition performance at high temporal rates.

Instead, this basic temporal limit in object recognition may reflect a fundamental limit in the processing capacity of the visual system. Capacity-limited processing might occur in early visual areas that encode the local features of objects or high-level areas that encode the global shapes of objects. Neurons in the primary visual cortex (V1) are sensitive to a much lower range of temporal frequencies than those in the lateral geniculate nucleus (LGN), indicating a loss of temporal processing capacity in V1 (Hawken et al. 1996). Early visual areas (V1–V4) seem to show similar temporal frequency response profiles to drifting gratings, with average peak tuning ranging from 3 to 10 Hz depending on the study (Foster et al. 1985; Gegenfurtner et al. 1997; Hawken et al. 1996; Levitt et al. 1994; Singh et al. 2000). Recordings in inferotemporal cortex have revealed that very brief presentations (<20 ms) of an object can evoke reliable neuronal responses, although longer stimulus durations lead to stronger responses (Keysers et al. 2001; Kovacs et al. 1995; Rolls and Tovee 1994). Because of the different stimulus paradigms used across studies, it has proven difficult to compare the temporal sensitivity of early and high-level visual areas directly.

We used functional MRI (fMRI) to characterize the temporal response properties of the ventral visual pathway, with the goal of understanding the relationship between cortical processing and capacity limits in object recognition. Subjects were presented with natural images of faces and houses at varying temporal rates to evoke responses in retinotopic visual areas (V1–V4) as well as object-selective areas, specifically the fusiform face area (FFA) and parahippocampal place area (PPA) (Epstein and Kanwisher 1998; Kanwisher et al. 1997; McCarthy et al. 1997; Tong et al. 1998). Temporal rate response profiles were quantified to characterize the temporal sensitivity of multiple visual areas throughout the ventral pathway.

METHODS

Participants

Seven right-handed, healthy adults (1 female; age, 21–32 yr) with normal or corrected-to-normal visual acuity participated in the main

Address for reprint requests and other correspondence: Frank Tong, Psychology Dept., 301 Wilson Hall, 111 21st Avenue South, Nashville, TN 37203 (E-mail: frank.tong@vanderbilt.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

experiment. Five volunteers (1 female), four from the main study, participated in an additional control experiment that involved passive viewing of the same visual stimuli as in the main study. The study was approved by the Institutional Review Panel at Princeton University. All subjects provided written informed consent.

Experimental design and stimuli

In each experimental fMRI run, subjects viewed stimulus sequences of either faces or houses presented at varying temporal rates of 2.3, 4.7, 9.4, 18.8, or 37.5 items/s (monitor frame rate, 75 Hz). No blank period or visual mask occurred between successive images; therefore stimulus presentation rate was inversely proportional to the presentation duration of each image.

Each run started with a 16-s fixation-baseline period, followed by alternating periods of stimulus presentation (8 s) and fixation rest (16 s). The duration of an entire run was 256 s. Presentation rates either increased across successive stimulus blocks within a run (i.e., 10 blocks consisting of 2.3, 4.7, 9.4, 18.8, 37.5, 2.3, 4.7, 9.4, 18.8, and 37.5 items/s) or decreased within a run. The order of presentation rates was counterbalanced across runs. Each subject performed a total of eight experimental runs, four face runs, and four house runs, for a total of eight stimulus blocks for each combination of stimulus type and temporal rate. In addition, subjects performed four runs of a control experiment in which they viewed image sequences that alternated between face and house repeatedly, shown at the same presentation rates.

Visual stimuli were rear-projected onto a screen in the scanner bore using a luminance-calibrated Epson Powerlite 7250 LCD projector driven by a Macintosh computer. A limitation of LCD projectors, which are easier to adapt to MR environments, is that they provide less precise control over the timing of visual presentation than traditional CRT displays. Nonetheless, recent studies have shown that LCD projectors still provide a high correspondence between desired and actual target durations, with trial-by-trial variability on the order of ~ 5 ms (Wiens et al. 2004). Using a photodiode connected to an oscilloscope, we confirmed that our LCD projector was able to present changes in luminance up to the highest presentation rates used in the study (37.5 Hz). Although some trial-by-trial variability in the duration of each image is to be expected with LCD presentation, the average presentation rate on each trial corresponded to the desired presentation rate.

Subjects were instructed to maintain fixation on a central fixation point (0.56° diam), which remained present throughout each experimental run, while stimuli were presented within a $10.6 \times 10.6^\circ$ -sized window at the center of the screen on a uniform white background. Stimuli consisted of digitized grayscale images of 30 different faces and 30 different houses. Because front-view faces are visually homogenous, we used face images of varying depth-plane rotations to reduce the amount of contour overlap across successive images and thereby lessen the extent of visual masking. Possible face orientations consisted of 0° , $\pm 22.5^\circ$, and $\pm 45^\circ$ rotated views, relative to a front-view perspective.

The subject's task was to report which of two possible target images appeared in each stimulus sequence by pressing one of two buttons on a response box at the end of each stimulus block. Subjects were allowed to view the two target images freely before starting each experimental run. Each stimulus block consisted of a randomly generated sequence of distractor images, selected from the same object category as the targets (i.e., faces or houses). Sequences were generated by randomly drawing from the full set of distractor images without replacement and repeating this procedure until the full stimulus sequence was created. The identity and temporal position of the target was randomly determined for each sequence, with the constraint that the target could not appear as the first or last image in the stimulus sequence. Before the actual fMRI study, each subject performed a

practice run involving the face discrimination task and a practice run involving the house discrimination task.

Five subjects participated in a separate control experiment involving passive viewing of the same visual stimuli. In the passive viewing version of this experiment, target images were not presented to the subject, and subjects were instructed simply to stay alert and to maintain steady fixation while viewing all stimuli.

A separate eye-tracking version of this experiment was performed by four subjects outside of the scanner to determine whether observers can maintain stable fixation while discriminating targets embedded in these rapid serial visual displays. Eye position was recorded using an Applied Systems Laboratory EYE-TRAC 6000 120-Hz video-based eye-tracking system (Bedford, MA). Results confirmed that observers could maintain very stable fixation throughout each run. For presentation rates of 2.3, 4.7, 9.4, 18.8, and 37.5 items/s, mean eye position values were -0.08 ± 0.40 , -0.18 ± 0.49 , -0.22 ± 0.37 , -0.21 ± 0.31 , and $-0.16 \pm 0.28^\circ$ (SD), respectively, for horizontal positions relative to fixation and -0.26 ± 0.67 , -0.22 ± 0.55 , -0.19 ± 0.67 , -0.06 ± 0.46 , and $-0.17 \pm 0.46^\circ$ for vertical positions relative to fixation. No reliable differences in fixation stability were found between the five temporal rates.

MRI acquisition

Subjects were scanned at the Princeton Center for the Study of Brain, Mind and Behavior on a 3.0-Tesla Siemens MAGNETOM Allegra scanner using a standard head coil. A high-resolution anatomical scan was collected using a T1-weighted 3D SPGR sequence with 1-mm isotropic voxels. Standard T2*-weighted gradient-echo echoplanar imaging was used to measure BOLD contrast for whole-brain functional imaging (TR 2,000 ms, TE 30 ms, flip angle 90° , in-plane resolution 3×3 mm, 28 slices, slice thickness 5 mm, gap between slices 1 mm). Head movement was minimized using either a custom bite-bar system or a forehead strap.

Data analysis

Functional data were motion-corrected using Automated Image Registration (AIR) (Woods et al. 1998). Subsequent analyses were conducted using Brain Voyager (Brain Innovation, Maastricht, The Netherlands). Preprocessing of the functional data included the removal of linear trends, mean intensity adjustment, and slice scan-time correction. fMRI data from individual subjects were aligned to their retinotopic visual maps, collected in a separate session, through the coregistration of high-resolution three-dimensional (3D) anatomical scans. Automated alignment procedures were followed by careful visual inspection and fine-tuned manual adjustments, after which all data were transformed into Talairach coordinate space. Statistical maps were created using the general linear model with specified predictors for each stimulus condition. Predictors were determined by convolving the stimulus time course with a standard gamma function to account for the BOLD hemodynamic response.

fMRI response amplitudes for each experimental condition, visual area, and subject were calculated by averaging the response amplitudes of individual stimulus blocks. The amplitude of the fMRI time course for each stimulus block was measured relative to the preceding fixation-baseline period by calculating the amplitude of the best-fitting sinusoid function. Because the hemodynamic response is better approximated by a sinusoid function than a simple boxcar function, this approach provides a more stable and robust estimate of fMRI response amplitudes. The frequency, phase, amplitude, and vertical displacement of the best-fitting sinusoid were determined using standard procedures in Matlab to minimize the mean squared error between the actual data and estimated fits. To ensure reasonable fitted values and the exclusion of noisy fMRI trials, limits were set on the allowable range of phase and frequency values for the fitted sinusoidal curve. Data from individual trials were excluded if the fitted sinusoid had a

minimum value that fell outside of time-points -4 to $+4$ s relative to stimulus onset (equivalent to ± 2 TRs), a maximum value that occurred after 16 s poststimulus onset or a period that was <14 or >32 s. This resulted in the removal of $<7\%$ of the data. Average response amplitudes were calculated for each presentation rate, stimulus type, visual area, and subject, and the resulting data were analyzed using within-subjects ANOVA and planned comparisons. The same pattern of results was obtained when response amplitudes were estimated using a simple boxcar function that averaged over time-points 4–12 s poststimulus onset. However, the reliability of our amplitude estimates, both within and across subjects, was improved by using the sine-fitting procedure.

We calculated mean response amplitudes as a function of temporal rate for each subject, visual area, and stimulus condition. Peaks in these temporal rate response functions were determined by fitting a third-order polynomial to the data and identifying the peak of the fitted function. Polynomial fitting provided an efficient and effective method to fit the data given the variety of possible shapes of the temporal rate response profiles across visual areas and the small number of data points defining each curve. While most temporal rate response functions followed an inverted-U profile, some subjects showed linearly increasing activity as a function of temporal rate in areas such as V1 or linearly decreasing activity in the FFA or PPA. Polynomial fitting effectively captured the variance of both linear and nonlinear components and provided estimates of peak temporal sensitivity that agreed well with our own evaluations based on visual inspection. It should be noted that polynomial fitting is quite robust because it takes into account the value of all points along a curve; thus points that lie somewhat distant from the peak can influence the estimated location of the peak to some extent. In cases where activity levels are greater at the lowest temporal frequencies and weaker at the highest temporal frequencies (cf. Fig. 4), polynomial fitting may lead to a slight leftward shift in the estimated location of peak activity to provide a better fit of the entire curve. Additional analyses verified that similar results were obtained with other methods to determine peak temporal tuning (Supplementary Fig. 1).¹

Regions of interest

The primary regions of interest in this study consisted of areas along the ventral pathway: retinotopic areas V1v–V4v and category-selective regions of the ventral temporal cortex, specifically the FFA and PPA (Fig. 1). We performed additional analyses of dorsal retinotopic areas V1d–V3a to provide a more detailed characterization of the temporal response properties of areas throughout the visual system.

LOCALIZATION OF THE FFA AND PPA. The FFA and PPA were localized using well-documented procedures (Epstein and Kanwisher 1998; Kanwisher et al. 1997; Tong et al. 1998, 2000). Subjects completed two localizer runs in which they passively viewed alternating stimulus blocks of faces and houses. The FFA was identified in individual subjects as the region in the fusiform gyrus that responded significantly more to faces than houses, using a minimum statistical threshold of $t > 5.2$, $P < 0.025$ corrected. The PPA was identified as the region in the parahippocampal gyrus that responded significantly more to houses than faces using a similar statistical threshold.

RETINOTOPIC MAPPING OF VISUAL AREAS. Retinotopic visual areas of each subject were delineated in a separate experimental session using well-established methods (DeYoe et al. 1996; Engel et al. 1997; Sereno et al. 1995). Details of our specific procedures for retinotopic mapping and cortical flattening can be found in previous reports from our laboratory (Awater et al. 2005). In brief, subjects maintained fixation while viewing “traveling wave” stimuli consisting of rotating wedges and expanding rings, which were used to construct phase-

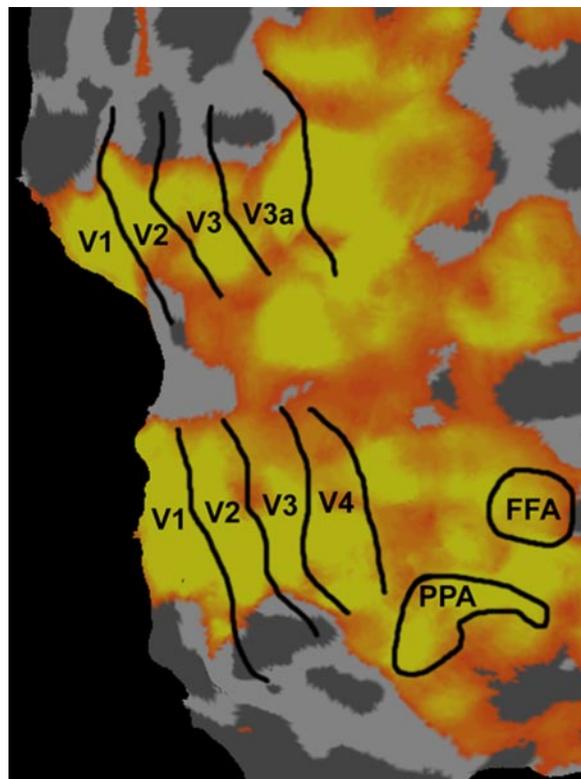


FIG. 1. Visual areas of interest shown on the flattened cortical representation of a representative subject. Primary regions of interest consisted of ventral areas V1v–V4v (lower bank areas), the fusiform face area (FFA), and parahippocampal place area (PPA). Additional analyses were performed on dorsal retinotopic areas V1d–V3a (upper bank areas). Pseudocolor statistical map shows activations to centrally presented faces and houses while subjects maintained fixation on a static fixation point (t -value range: 4.0–16.0).

encoded retinotopic maps of polar angle and eccentricity, respectively. Subjects typically completed eight polar-angle runs and four eccentricity mapping runs (10 stimulus cycles/run, 32 s/cycle). Boundaries between visual areas were delineated on flattened cortical representations using field-sign mapping, which identifies reversals in polar-angle preference relative to topographic changes in eccentricity preference (Sereno et al. 1995).

Within each region of interest, we selected all voxels that were reliably activated in the main experiment by centrally presented faces and houses, relative to fixation baseline, using a minimum statistical threshold $t > 4.0$, $P < 0.0001$ uncorrected. For all subsequent analyses, we used the average MR response of all visually active voxels within each region of interest.

RESULTS

Behavioral results

In each experimental fMRI run, subjects viewed visual sequences of either faces or houses presented at varying temporal rates. Subjects were required to discriminate which of two target images appeared within each stimulus sequence among a set of randomly ordered distractor images from the same visual category. As expected, discrimination performance declined as a function of presentation rate for both faces and houses (Fig. 2). Behavioral performance was near ceiling at slow presentation rates of 2.3 and 4.7 items/s, dropped to $\sim 75\%$ accuracy at intermediate rates of 9.4 items/s, and fell to chance levels when stimuli were rapidly presented at rates of

¹ The online version of this article contains supplemental data.

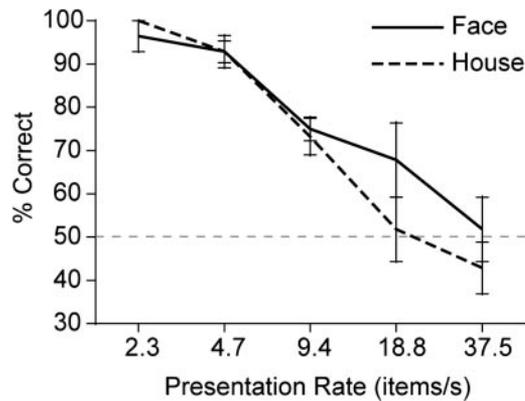


FIG. 2. Behavioral performance. Target recognition performance for both faces (solid lines) and houses (dashed lines) revealed a monotonic decline in accuracy as a function of stimulus presentation rate ($F = 30.8$, $P < 10^{-8}$). Significant decreases in performance were observed between presentation rates of 4.7 and 9.4 items/s ($F = 11.7$, $P < 0.005$) and 9.4 and 18.8 items/s ($F = 6.8$, $P < 0.05$). At rapid presentation rates of 18.8 items/s and higher, performance no longer reliably differed from chance level of 50%. Error bars indicate \pm SE.

18.8 or 37.5 items/s. These behavioral results are consistent with previous studies showing that visual recognition begins to decline at presentation rates of ~ 8 – 10 items/s and falls sharply at faster presentation rates (McMains and Somers 2004; Potter 1975).

fMRI results

Temporal rate response profiles for each visual area were constructed by plotting fMRI response amplitudes as a function of presentation rate for viewed faces and houses separately. Response amplitudes on individual stimulus blocks were mea-

sured relative to fixation baseline, and mean amplitudes for each experimental condition were calculated by averaging first within and then across subjects (see METHODS). Although the primary areas of interest in this study lay within the ventral visual pathway (areas V1v–V4v, FFA, and PPA), response amplitudes for dorsal retinotopic areas V1d–V3a were also analyzed to provide a more comprehensive description of the temporal response properties of the visual system.

Figure 3 provides a comparison of temporal rate response profiles for visual areas within the ventral and dorsal pathways, separated by stimulus type. In general, all visual areas showed above-baseline levels of activity across the full range of temporal rates tested. These responses seemed to be quite broadly tuned, as one might expect for fMRI measures of population activity given that individual neurons in a cortical area can greatly differ in their temporal tuning preferences (Foster et al. 1985; Gegenfurtner et al. 1997; Hawken et al. 1996). Responses across all presentation rates were also expected because our randomized image sequences led to stimulus energy over a broad range of temporal frequencies, with fall-offs in power occurring at frequencies exceeding one half of the presentation rate (for details, see Fig. 6B). Nonetheless, fMRI response amplitudes of each visual area were strongly modulated as a function of temporal rate.

Focusing on ventral visual areas, a common pattern of results for faces and houses can be seen (Fig. 3, A and B, respectively). Early visual areas, V1 and V2, showed increasing activity as a function of temporal rate, with peak amplitudes occurring at high rates of roughly 18.8 items/s. In comparison, higher visual areas showed evidence of a leftward shift in peak activity toward lower temporal rates. This shift in peak tuning toward lower temporal rates was especially evident

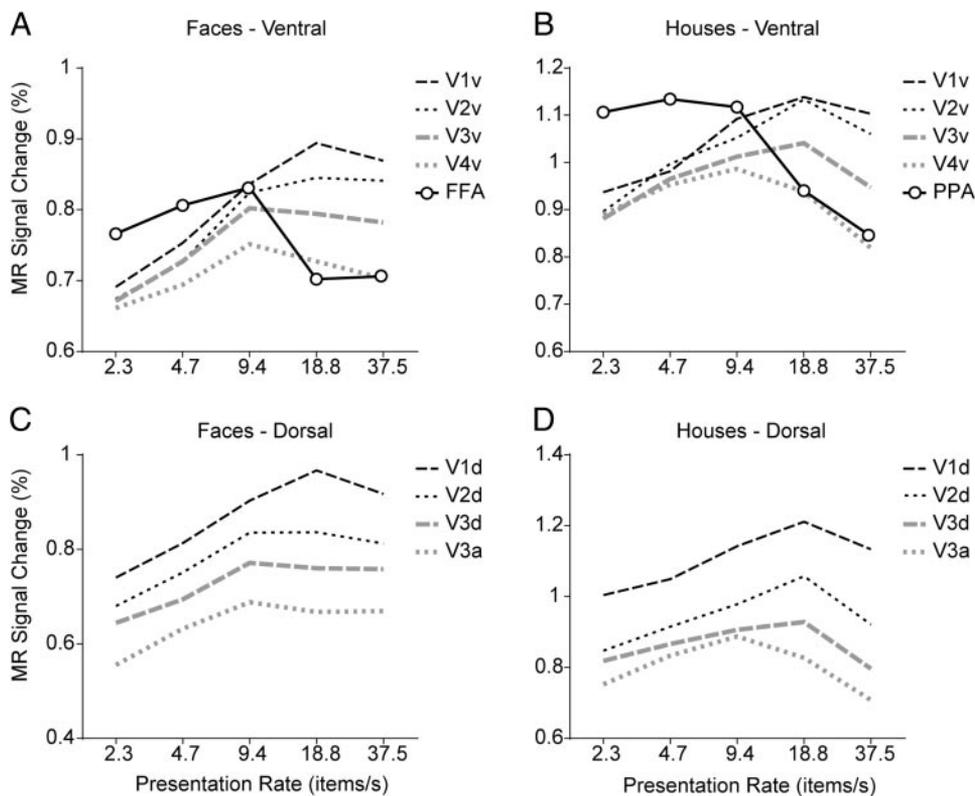


FIG. 3. Average functional MRI (fMRI) response amplitudes of all subjects, plotted as a function of temporal rate for individual visual areas. Subplots show the temporal rate response profiles of ventral visual areas for faces (A) and houses (B) and of dorsal retinotopic areas for faces (C) and houses (D). Mean response amplitudes indicate the percent change in MR signal relative to fixation baseline for each stimulus presentation rate.

in the FFA response to face stimuli and the PPA response to house stimuli. These results are suggestive of poorer processing of object information at high temporal rates, as was indicated by the subject's behavioral performance (Fig. 2).

In theory, one might predict that activity in object-selective areas should increase linearly as a function of the number of preferred stimuli that are presented within a fixed time window or at least increase in a monotonic fashion (Mukamel et al. 2004). That is, if the total amount of neural activity in a region were to increase steadily as a function of object presentation rate, BOLD activity would be expected to increase in a corresponding fashion, or to saturate at a some level because of physiological limitations in the maximum possible blood flow. Instead, however, we found that presenting a greater number of stimuli, at the expense of presentation duration, actually led to a decrease in the response of these areas. The weaker responses found at high temporal rates cannot be explained in terms of saturation of the BOLD response, and instead, indicate that underlying neural responses are attenuated at high temporal rates.

ANOVAs were performed on data from each of the subplots in Fig. 3 to assess whether response amplitudes reliably varied across visual areas or temporal rates. The response of ventral areas to face stimuli (Fig. 3A) revealed no reliable difference in overall response amplitudes across visual areas ($F = 0.52$, $P = 0.72$), but a significant main effect of temporal rate ($F = 5.78$, $P < 0.005$) and a highly significant interaction between visual area and temporal rate ($F = 8.51$, $P < 10^{-11}$). The response of ventral areas to house stimuli (Fig. 3B) revealed similar results: no main effect of visual area ($F = 0.58$, $P = 0.67$), a main effect of temporal rate ($F = 10.41$, $P < 10^{-6}$), and a highly significant interaction between visual area and temporal rate ($F = 21.49$, $P < 10^{-16}$). The robust interaction effects indicate that temporal rate response profiles reliably differ across ventral visual areas; subsequent analyses will focus on the nature of these differences. Additional analyses confirmed that every single visual area of interest showed reliable modulations in response amplitude as a function of temporal rate, with the sole exception of area V4v, which showed a reliable effect for houses ($F = 10.86$, $P < 0.0001$) but nonsignificant trend for faces ($F = 2.3$, $P = 0.109$).

As expected, the FFA responded strongly to faces and the PPA responded strongly to houses, and both regions showed weak, unreliable responses to their nonpreferred stimulus category, consistent with previous reports (Epstein and Kanwisher 1998; Kanwisher et al. 1997; Tong et al. 1998). Response amplitudes for nonpreferred stimuli were too weak to estimate reliably using function fitting methods and therefore are not plotted in Fig. 3. However, it was possible to calculate these fMRI amplitudes based on the average response using a simple boxcar function that accounted for hemodynamic lag (4–12 s poststimulus onset, measured relative to a baseline period –4 to 0 s). Across the five presentation rates (2.3, 4.7, 9.4, 18.8, and 37.5 items/s), the percent signal change of mean fMRI responses were 0.44, 0.49, 0.46, 0.37, and 0.33 for FFA responses to houses and 0.33, 0.32, 0.31, 0.32, and 0.27 for PPA responses to faces, respectively. Statistical analyses revealed no reliable differences between temporal rates in the FFA or PPA ($F = 2.0$, $P = 0.12$). If such modulations were present, the low-amplitude of the fMRI responses for nonpreferred stimuli would likely have impaired the ability to detect reliable

differences. In all subsequent analyses of the FFA and PPA, we therefore focused on fMRI responses to the preferred stimulus category.

Dorsal retinotopic areas also showed reliable effects of temporal rate and some evidence of differences in temporal rate response profiles across visual areas (Fig. 3, C and D). For face stimuli, we observed reliable main effects of visual area ($F = 3.39$, $P < 0.05$) and temporal rate ($F = 20.99$, $P < 10^{-10}$), although the interaction between visual area and temporal rate was not statistically significant ($F = 1.40$, $P = 0.19$). Response amplitudes for house stimuli revealed main effects of visual area ($F = 4.20$, $P < 0.05$) and temporal rate ($F = 20.54$, $P < 10^{-10}$), and in this case, the statistical interaction proved to be highly significant ($F = 6.55$, $P < 10^{-7}$). In general, the pattern of results was quite similar to that seen in the ventral pathway, with higher visual areas showing peak activity at lower temporal rates than early visual areas.

To better visualize whether visual areas differed in their temporal response properties, fMRI response amplitudes for each visual area, stimulus type, and subject were normalized to that of the slowest temporal rate. Figure 4, A and B, shows the normalized temporal rate response profile of each ventral visual area for faces and houses, respectively. A clear pattern can be observed across the visual hierarchy. The ventral portion of V1 showed the strongest responses at high temporal rates, followed by areas V2, V3, and V4, with areas FFA and PPA showing the weakest responses at high rates. Inspection of the peak of the tuning curve for each visual area also suggested a gradual leftward shift toward lower temporal rates at successive levels of the visual hierarchy. Normalized responses for dorsal areas V1d–V3a displayed a similar pattern; higher areas showed evidence of a leftward shift in peak responses in favor of lower temporal rates (Fig. 4, C and D). Therefore as visual information is passed from early areas to higher areas, there seems to be a progressive loss in sensitivity to high temporal rates and a gradual shift in peak sensitivity toward lower temporal rates. When presentation rates exceed the peak sensitivity of a given visual area, the response level does not remain steady at asymptote but instead begins to decline, even though more information is being presented that could potentially drive the visual system.

Next, we analyzed the shape of the temporal rate response function to determine the peak temporal tuning of individual visual areas. A third-order polynomial function was used to fit the response profile obtained for each visual area, stimulus type, and subject. The resulting peak of the fitted curve provided an objective estimate of the peak temporal sensitivity (see METHODS). Figure 5A (solid line) reveals a systematic decline in peak temporal sensitivity across the ventral visual pathway for both faces and houses ($F = 33.812$, $P < 10^{-8}$, peak temporal rates analyzed using log values). Peak sensitivity occurred at rapid presentation rates for V1v, V2v, and V3v (average peak rate of 24.9, 19.8, and 18.3 items/s, respectively), intermediate rates for V4v (9.1 items/s), and much lower rates for the FFA and PPA (5.1 and 4.3 items/s, respectively). Planned comparisons indicated that peak temporal sensitivity did not reliably differ across areas V1v–V3v but was significantly lower in V4v than in earlier areas ($F = 37.41$, $P < 10^{-5}$) and lower still for the FFA/PPA ($F = 15.28$, $P < 0.001$). It is interesting to note that at presentation rates greater than 4.7 items/s or the approximate preferred rate of the FFA

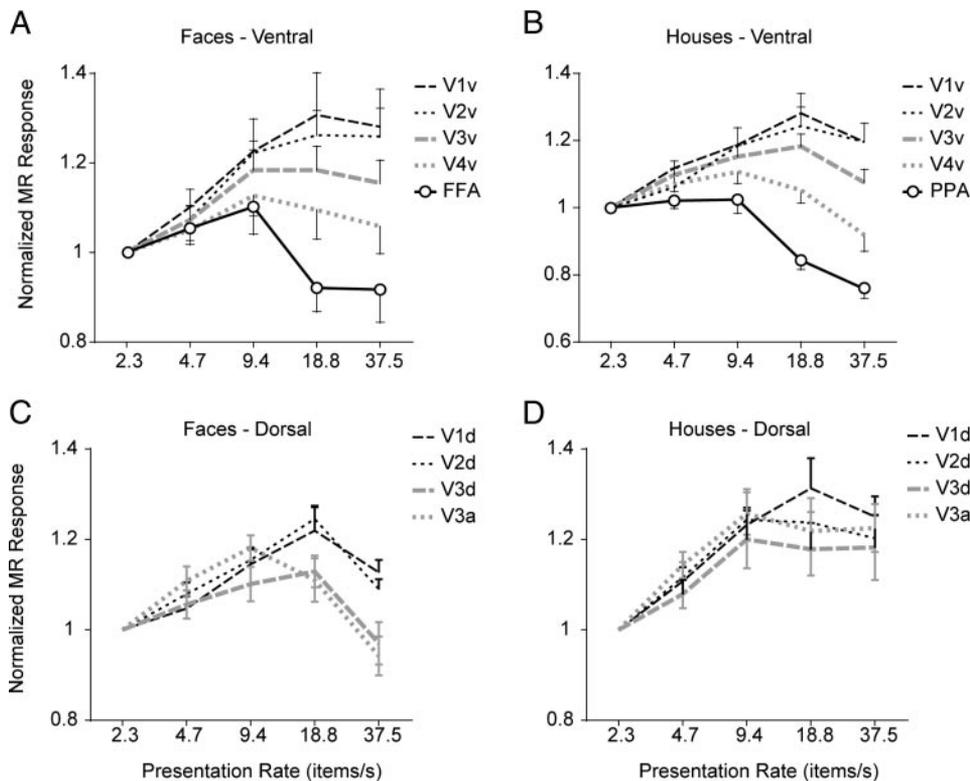


FIG. 4. Normalized fMRI response amplitudes plotted as a function of temporal rate. Response of ventral visual areas to faces (A) and houses (B); response of dorsal visual areas to faces (C) and houses (D). Response amplitudes were normalized relative to the amplitude for the lowest temporal rate. Error bars indicate \pm SE. A systematic decline in responses to high temporal rates, as well as a leftward shift in peak response, can be seen in progressively higher visual areas.

and PPA, behavioral performance on the target discrimination task also began to decline, suggesting a possible link between activity in these object-selective areas and object recognition performance.

A comparison between temporal rate response profiles (Fig. 4, A and B) and estimated peak sensitivities (Fig. 5A) indicated generally good agreement. However, it can be noted that the average fMRI response of the FFA and PPA across all subjects reached its highest value at presentation rates of 9.4 item/s, which was greater than our estimates of average peak sensitivity based on fits of individual temporal response profiles. This difference was largely because of the fact that some subjects showed the strongest FFA or PPA response at the presentation rate of 9.4 items/s (3/7 subjects), whereas others showed maximal responses at slower rates of 4.7 (3/7 subjects) or 2.3 items/s (1/7 subjects). Also, polynomial functions provide a fit of the entire curve; therefore if fMRI response amplitudes are much greater at low than high temporal rates, this could lead to a small leftward shift in the peak of the fitted function. Control analyses confirmed that our estimates of peak temporal sensitivity were robust to the specific method used for function fitting. We compared peaks identified using third-order polynomials, second-order polynomials, and those identified by simply choosing the discrete temporal rate that showed the highest fMRI response level (Supplementary Fig. 1). Although the exact values of estimated peaks varied to a small degree depending on the method used, in all cases we observed the same pattern of decreasing temporal sensitivity across the visual hierarchy.

Dorsal retinotopic areas showed a similar decline in peak temporal tuning across the visual hierarchy (Fig. 5B; $F = 4.483$, $P = 0.05$). Averaged across face and house conditions, peak tuning values for areas V1d, V2d, V3d, and V3a were

24.5, 18.5, 15.8, and 13.6 items/s, respectively. Therefore a consistent decline in peak temporal sensitivity was found at higher levels of visual system, in both ventral and dorsal visual areas, with the lowest temporal sensitivities found in anterior object-selective areas of the ventral temporal cortex.

One might ask whether other object-sensitive areas showed similar peaks in temporal tuning as the FFA and PPA. Although our functional localizer runs were designed to isolate face- and house-selective regions of the ventral visual system, it was possible to identify visually active voxels corresponding to the known anatomical location of the lateral occipital complex (LOC). Area LOC responds more strongly to intact than scrambled objects (Malach et al. 1995) and has been strongly implicated in object perception and successful recognition performance (Grill-Spector et al. 2000). In this anatomically defined region, we observed significant modulations in fMRI amplitudes as a function of temporal rate ($F = 5.11$, $P < 0.005$), with estimated peaks in temporal sensitivity of 10.5 and 5.0 items/s for faces and houses, respectively. Temporal rate response profiles for area LOC did not reliably differ for faces and houses ($F = 1.46$, $P = 0.25$), and peak tuning values were quite comparable with those found in the FFA and PPA.

Group-based activation maps

In general, ROI analyses are better suited for isolating the activity of individual visual areas, because these regions can vary considerably in their Talairach position from subject to subject. Such analyses also provide greater statistical power than post hoc group analyses, which require correcting for multiple voxel-wise comparisons. Nevertheless, it was possible to plot the results of a random-effects group analysis on the cortical flatmap of a representative subject to gain a general

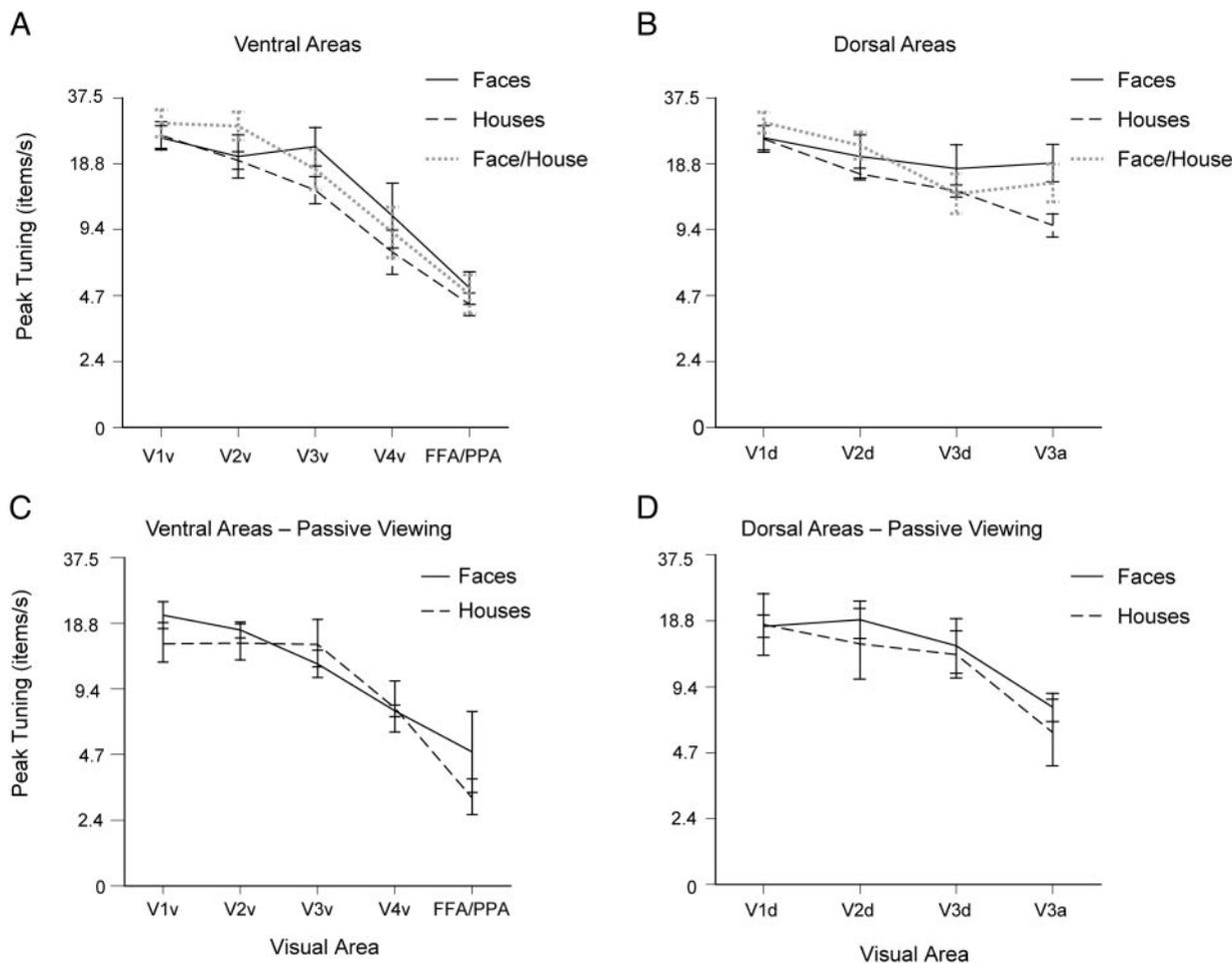


FIG. 5. Peak temporal sensitivity of ventral and dorsal visual areas for the main experiment (*A* and *B*) and a control experiment involving passive viewing (*C* and *D*). Plots show average temporal rate at which peak responses occurred within each visual area for faces (solid lines), houses (dashed lines), and alternating face-house sequences (dotted gray lines in *A* and *B*). Ordinate axis shows the peak temporal rate plotted in log units of items per second. Error bars indicate \pm SE. ANOVA revealed a reliable decline in peak temporal sensitivity across the visual hierarchy for both the main experiment ($F = 33.8, P < 10^{-8}$) and the passive viewing experiment ($F = 23.9, P < 10^{-6}$). Comparisons revealed no reliable differences in fMRI responses between the 2 experiments.

impression of which brain regions tended to prefer low or high temporal rates. From this analysis, it appeared that regions in the vicinity of the FFA and PPA, just anterior to ventral V4, showed stronger responses at low presentation rates (Supplementary Fig. 2A). Regions anterior to V3a and the adjoining foveal representation, in the vicinity of posterior LOC, also preferred lower presentation rates. In contrast, most of V1 and portions of V2 showed greater responses to high than low temporal rates, in agreement with the ROI analysis. Regions preferring higher temporal rates were well centered within the larger retinotopic region that responded positively to stimuli at all presentation rates (Supplementary Fig. 2B), with no evidence of a shift toward the periphery. These findings are consistent with human psychophysical evidence indicating that temporal sensitivity is remarkably uniform across the visual field (Virsu et al. 1982; Wright and Johnston 1983); it is no longer believed that temporal sensitivity is superior in the periphery (Sharpe 1974).

Control experiment: fMRI results for intermixed faces and houses

We ran a control experiment, in which subjects viewed image sequences that alternated between face and house im-

ages repeatedly, to address two key issues. Of particular interest was whether the tuning profiles of the FFA and PPA depended on the presentation rate of preferred stimuli or on the absolute rate of stimulus presentation, independent of the identity of each item. In the case of the FFA, does the temporal tuning of this region specifically reflect the number of faces that must be processed, such that nonpreferred house stimuli can be effectively filtered or ignored, or does it reflect a more fundamental limit in the rate at which items of any type can be processed? We suspected that high-level object-sensitive areas might suffer from a basic limit in temporal processing, just as intermediate visual areas such as V4v showed a general decrease in temporal sensitivity compared with early visual areas.

This control experiment also ensured that large visual changes occurred with each alternation between face and house, because faces and houses do not share a common visual structure. A potential concern with the face-only and house-only sequences is that images from a common category may tend to share overlapping contours or features. As a consequence, when images from a homogenous category are presented at a specified rate, this might lead to slower rates of visual change over local portions of the image. We reasoned that if visual areas show the same tuning profiles for alternating

face-house sequences as for face-only and house-only sequences, this would indicate that visual homogeneity is not a determining factor.

Results of the control experiment can be seen in Fig. 5, A and B; the dotted line shows peak temporal sensitivity of individual areas in the ventral and dorsal pathways, respectively. Peak tuning across the visual hierarchy was essentially the same for image sequences involving faces only, houses only, and alternating faces and houses, indicating that visual homogeneity of the image sequence could not account for these results. Moreover, the FFA and PPA showed essentially the same peak tuning for image sequences involving alternating face-house sequences (4.2 and 5.4 items/s, respectively) as was observed for sequences of preferred stimuli only. If instead, temporal tuning responses were driven by the number of preferred stimuli that activated these areas, one would expect that peak tuning for face-house sequences should occur at twice the rate of single category sequences. (A similar doubling effect would be predicted if one were to argue that neural adaptation to occasional repetitions of individual images might lead to weaker fMRI responses at high temporal rates. Although these occasional repetitions occurred more often at higher presentation rates, caused by random sampling from a fixed number of face and house images, repetitions occurred half as often in the alternating face-house sequences, yet no shift in temporal tuning was observed.)

The results provide compelling evidence that the temporal tuning properties of category-selective areas depend on the *absolute* number of items that must be processed within a short time period rather than the number of preferred items. These peaks in temporal tuning may reflect an upper bound in the efficiency of object processing at high levels of the visual system.

fMRI results for passive viewing experiment

To ensure that the effects found in the FFA and PPA were not caused by the specific attentional demands of the target discrimination task or a possible orienting response to the detected target, we ran five subjects in an additional control experiment that simply required passive viewing of faces and houses presented at varying temporal rates. (The alternating face-house condition was not tested here.) Typically, when subjects are asked to view such displays without any specific task requirements, they informally report that the images appear to blur or blend into one another at the two highest presentation rates, and that it is very difficult to perceive the details of individual items at these rates though the general category of the image sequences (faces or houses) can be readily perceived. In this passive viewing experiment, we again observed a systematic decline in peak temporal sensitivity across both the ventral visual pathway (Fig. 5C; $F = 23.9$, $P < 10^{-6}$) and the dorsal pathway (Fig. 5D, $F = 10.9$, $P < 0.001$). Additional analyses confirmed that the pattern of results did not reliably differ across the two experiments, indicating that these changes in peak temporal sensitivity across the ventral visual pathway likely reflect the inherent temporal properties of these visual areas rather than the specific demands of the experimental task.

Analysis of the temporal spectrum of image sequences across varying spatial scales

There is growing interest in using complex natural images to study the tuning properties of the visual system (Carandini et al. 2005; Simoncelli and Olshausen 2001), even though the spatial-temporal properties of natural stimuli can be more difficult to characterize or to control. In this study, we presented randomized sequences of natural images at varying rates to investigate the temporal response properties of both low-level and high-level areas. A potential concern lies in the possibility that our random image sequences might have led to more rapid changes over time at coarse spatial scales and slower temporal modulations at fine spatial scales. The presence of such systematic differences in the temporal frequency spectrum across spatial scales could lead to an advantage at high presentation rates for early visual areas, which have smaller receptive fields (RFs) that prefer higher spatial frequencies.

To address this potential concern, we measured the extent of temporal change at multiple spatial scales for randomly generated image sequences. Oriented gabor functions were used to simulate the properties of simple and complex cells with RFs of varying size (RF diam of 0.5, 1, 2, or 4°). Larger RFs were tuned to lower spatial frequencies (spat freq fixed to 3 cycles per RF; SD of Gaussian window, ¼ of RF diam). These simulated RFs were used to sample visual signals at each of four possible orientations (0, 45, 90, and 135°), two spatial phases (0 and 90°), and nine locations distributed according to a 3 × 3 grid that evenly divided the image (image size, 10.8 × 10.8°). Next, we generated randomized image sequences of faces, houses, and faces/houses intermixed, and calculated the magnitude of the instantaneous response of each RF to each stimulus by convolving the gabor filter with the image. Negative values were eliminated by applying half-wave rectification to simulate the response of simple cells. Complex cell responses were simulated by summing the squared response of 90° phase-separated oriented pairs. The finite Fourier transform of neuronal responses to each image sequence was calculated and sorted to determine the average power at different temporal frequencies for RFs of varying size.

The results indicate that these natural image sequences led to a broad spectrum of energy at all temporal rates. Figure 6A shows the temporal spectrum for simulated presentation rates of 37.5 items/s. Amplitudes reached maximal levels at moderately low frequencies and maintained this level up to the critical frequency corresponding to one half of the presentation rate. The same pattern of results was obtained for simulated responses of simple cells (Fig. 6A) and complex cells (data not shown). Although coarser spatial scales led to greater overall amplitudes, the temporal spectrum remained essentially flat at all spatial scales of sampling. In other words, these randomized sequences of natural images led to the equivalent of a white-noise spectrum in the temporal domain, across multiple spatial scales. Although there was somewhat more power at low spatial frequencies, just as one would expect for natural images (Field 1987), the distribution of power in the temporal domain was very similar at all spatial scales. In comparison, when the same analysis was applied to images presented at temporal rates slower than 37.5 items/s, we observed a systematic

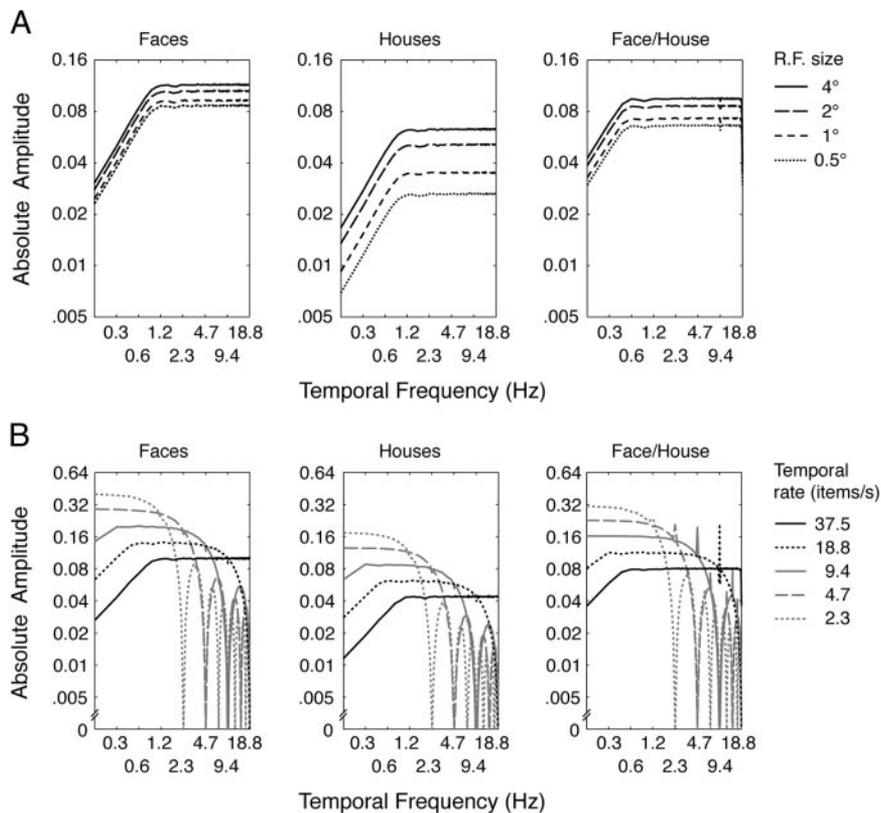


FIG. 6. Analysis of the temporal power spectrum of random image sequences sampled at varying spatial scales (A) and image presentation rates (B). Oriented-gabor receptive fields (RFs) of varying size (0.5, 1.0, 2.0, and 4.0°; spat freq, 3 cycles/RF) were used to calculate the instantaneous response to each image at different spatial scales. Changes in response over time were analyzed in the Fourier domain, for randomized sequences of faces, houses and faces-houses intermixed. (A) Distribution of temporal power was similar across spatial scales of sampling, with amplitudes remaining high up to the critical frequency corresponding to one half of the presentation rate (37.5 item/s for this simulation). (B) slower presentation rates led to a sharp drop-off in power around the critical temporal frequency corresponding to half of the presentation rate. Results for this analysis are pooled across all simulated RF sizes, orientations, and positions.

leftward shift in the peak of the temporal power spectrum, as expected (Fig. 6B).

DISCUSSION

In this study, we used a common set of natural images to compare temporal rate response profiles across multiple sites of the ventral visual pathway, with the goal of understanding the possible origins of the temporal limits in object processing. We observed a systematic decline in temporal frequency tuning across the visual hierarchy. Early visual areas responded best to rapidly presented objects, with peak tuning occurring at rates which were four to five times higher than those found in high-level object-selective areas. Similar results were obtained when subjects viewed faces, houses, or alternating face/house sequences, irrespective of whether they performed a challenging object discrimination task or passively viewed the stimuli, indicating the generality of these findings across variations in visual stimuli and task. These results provide novel evidence indicating that high-level object-selective areas are sensitive to a much lower range of temporal frequencies than early visual areas. It seems that, as visual information is transferred from low- to high-level areas, there is progressive loss in the temporal processing capacity of the human visual system.

Our results suggest that temporal limits in object recognition performance may be caused by the limited temporal sensitivity of object-selective areas rather than that of early visual areas, which continue to respond well at fast presentation rates. We found that areas V1–V3 responded best to objects presented at rapid rates of ~ 20 items/s, in general agreement with previous neurophysiological and neuroimaging studies of these areas showing peak tuning at ~ 10 Hz for drifting or flickering

gratings (Hawken et al. 1996; Singh et al. 2000). Taken together, these results suggest that these visual areas show quite comparable temporal response profiles for simple stimuli and complex natural images. Area V4v, which provides considerable input to higher areas in the ventral temporal cortex (Baizer et al. 1991), showed peak tuning at intermediate rates of ~ 9 items/s. In comparison, both the FFA and PPA showed peak tuning at much slower rates of 4–5 items/s. Activity levels in these areas diminished considerably when stimuli were presented at higher rates, indicating degraded processing of these objects.

Interestingly, the temporal tuning of these category-selective areas depended on the absolute rate of stimulus presentation for alternating face-house sequences, rather than the rate at which preferred stimuli were presented. It seems that the FFA and PPA suffer from a fundamental limit in temporal processing, which is not category-specific, such that nonpreferred stimuli can still compete with the processing of preferred stimuli. These competitive interactions may be taking place in the FFA and PPA proper and also at prior sites, such as V4v, which could impair the transmission of category-selective information to these higher areas.

Impairments in behavioral performance at the object discrimination task emerged at intermediate presentation rates. Recognition performance was near perfect at slow presentation rates of 2.3 and 4.7 items/s, moderate at 9.4 items/s, and dropped to chance levels at rates of 18.8 and 37.5 items/s. In comparison, peak temporal tuning in the FFA and PPA occurred at rates of 4–5 items/s, the highest rate at which subjects could still achieve near perfect recognition performance, whereas earlier visual areas peaked at higher temporal rates. Generally speaking, the decline in behavioral performance

seemed to be more consistent with the limited temporal sensitivity found in object-selective areas than with the higher temporal sensitivity found in earlier areas. It should be noted that the correspondence between recognition performance and fMRI responses across presentation rates is unlikely to reveal the exact same pattern of effects, because fMRI response amplitudes depend on both the number of stimuli presented and the presentation rate. Presumably, neural responses should increase as a function of the number of items presented—at very slow presentation rates, fMRI responses summate linearly (Dale and Buckner 1997; Liu and Gao 2000). However, the strength or quality of neuronal responses to individual items may decline at higher presentation rates because of limitations in the temporal processing capacity of the visual system, leading to nonlinear effects in the BOLD response as presentation rate is increased (Liu and Gao 2000; Mukamel et al. 2004). When the cost of increasing the temporal rate outweighs the effects of increasing the number of stimuli, activity levels in a given brain region will begin to decline as presentation rate is further increased. Overall, the temporal response profiles observed in the FFA and PPA seem to provide a better account for the behavioral limitations in object discrimination performance than those of earlier visual areas.

Our neuroimaging results are also consistent with previous behavioral studies suggesting that the human visual system requires more time to process visual information of increasing complexity. People can perceive low-level motion and flicker at rates as high as 30–50 Hz (Kelly 1961, 1979), yet object recognition begins to decline at modest rates of ~8–10 items/s (McMains and Somers 2004; Potter 1975). Recognition of a target face can be disrupted by a subsequent visual mask at much longer delays (upward of 133 ms) if the mask consists of an intact face rather than a scrambled face or visual noise (Loffler et al. 2005). These results are consistent with the notion that competitive interactions between object representations occur over a more extended time period than low-level interactions. Here, we found that early visual areas, which are sensitive to low-level features, can process visual information at much faster rates than high-level areas that are sensitive to complex objects.

This study provides a more detailed picture of the temporal properties of the visual system by revealing how object-selective areas differ from earlier stages of visual processing. Previous neurophysiological studies have used visual masking or rapid serial visual presentation to investigate the temporal sensitivity of object-selective neurons, but it has proven difficult to compare these results to the temporal properties of early visual areas because of differences in the stimuli and experimental paradigms used across studies. One study found that a brief 20-ms presentation of a face followed by a patterned mask could elicit stimulus-specific activity in inferotemporal neurons (Rolls and Tovee 1994). However, these neurons showed much stronger and more selective responses for longer presentation durations, with durations of 100 ms leading to near-asymptotic neural performance. Another study found that inferotemporal neurons can distinguish between different objects in an RSVP sequence at presentation rates as high as 72 items/s (14 ms/item) (Keysers et al. 2001). However, this study also found that stimulus selectivity improved monotonically as a function of the duration of each presented image (i.e., the inverse of presentation rate) and seemed to reach asymptote at

the slowest presentation rate tested of 4.7 items/s. Because slower temporal rates were not tested in these studies, it is difficult to determine the exact rate at which peak activity would occur in inferotemporal regions. Nevertheless, the neurophysiological results seem to be consistent with the low temporal sensitivities of the FFA and PPA that we found using population measures of BOLD activity. A recent fMRI study of visual masking found that activity in both V1 and high-level object-selective areas increased as a function of stimulus duration, reaching near-asymptotic levels at 120 ms (Grill-Spector et al. 2000). Although this presentation duration is somewhat shorter than corresponding duration for which we find peak sensitivity in object selective-areas (5 items/s = 200 ms/item), it is important to note that visual masking and RSVP paradigms provide different measures of temporal processing efficiency. Visual masking provides an estimate of the duration required for effective processing of a single stimulus, whereas RSVP provides an estimate of the rate at which successive stimuli can be processed. These estimates of temporal processing capacity are not necessarily equivalent because the visual system may differ in its capacity to process a single object and multiple objects. This is evidenced by the fact that a target face is more effectively masked by another face than by a visual noise pattern (Loffler et al. 2005). In general, our estimates of temporal processing capacity seem to be consistent with previous neurophysiological and neuroimaging studies of object-selective areas and provide novel evidence that high-level areas are sensitive to much lower temporal rates of visual information than early areas.

This study also adds to a growing body of knowledge regarding the hierarchical functional organization of the visual system (Felleman and Van Essen 1991). It is well documented that, at progressively higher levels of the visual pathway, neurons are likely to be tuned to more complex visual features, conjunctions of features, object parts, or even entire objects (Felleman and Van Essen 1991; Grill-Spector and Malach 2004; Tanaka 1996). Whereas V1 is sensitive to basic features such as orientation, local motion, and binocular disparity (Tong 2003), intermediate visual areas such as V4 are sensitive to more complex features including curvature, basic 2D shape, and orientation in depth (Hinkle and Connor 2002; Pasupathy and Connor 2002), and high-level areas in the ventral temporal cortex show evidence of remarkable selectivity for complex shapes and objects (Grill-Spector and Malach 2004; Tanaka 1996; Tsao et al. 2006). Recent studies also show that contrast sensitivity, position coding, and viewpoint tuning tend to become more flexible or invariant at higher stages of visual processing (Avidan et al. 2002; Grill-Spector et al. 1998, 1999; Gross 1992; Levy et al. 2001; Logothetis et al. 1995). Our results indicate that these increases in tuning complexity and tolerance to image variation found at intermediate and higher levels of the visual pathway are accompanied by a progressive loss in temporal sensitivity.

Arguably, the simplest possible account for this loss of temporal sensitivity is that neurons at all stages of visual processing share a common biophysical limit in their ability to respond at high temporal rates. If every neuron were to resemble a broadly tuned low-pass filter, high-frequency signals that must pass through a series of such filters would undergo progressively greater attenuation. In theory, the biophysical limits of individual neurons might account for the progressive

loss of temporal sensitivity across the visual hierarchy. However, given that many cortical neurons can fire at temporal rates of 75 Hz or greater (Williams et al. 2004), far exceeding the peak temporal sensitivities found here, it is unclear whether limits in individual neuronal firing rates can account for these findings.

Another possible explanation for this progressive loss in temporal processing capacity is that higher visual areas must integrate information from a wide array of neurons projecting from earlier areas to achieve greater tuning complexity and position invariance. Neural integration of information from preceding stages may require more time and thereby lead to an unavoidable loss in temporal sensitivity across successive processing stages. Integrative activity among neurons within each visual area could also contribute to this loss. We believe that a possible organizing principle of the visual system may reflect a fundamental trade-off between the extent of integrative information processing that must be carried out by a cortical region and the time required to process this information.

In conclusion, our results showed a progressive loss in temporal sensitivity across successive stages of processing in the visual pathway, with high-level object-selective areas revealing the lowest range of temporal sensitivities. The limited processing capacity of object-selective areas may account for the temporal limits of object recognition performance. Future studies into the functional organization of the visual system will be important for uncovering the neural computations and processes that underlie the systematic decline in temporal sensitivity across the visual hierarchy.

ACKNOWLEDGMENTS

We thank J. R. Kerlin and D. K. Brady for technical assistance and the Princeton Center for Brain, Mind, and Behavior for MRI support.

GRANTS

This research was supported by the following grants from the National Institutes of Health: National Research Service Award MH-065214-2 to T. J. McKeef and P50-MH-62196 and R01-EY-14202 to F. Tong.

REFERENCES

- Avidan G, Harel M, Hendler T, Ben-Bashat D, Zohary E, Malach R. Contrast sensitivity in human visual areas and its relationship to object recognition. *J Neurophysiol* 87: 3102–3116, 2002.
- Awatier H, Kerlin JR, Evans KK, Tong F. Cortical representation of space around the blind spot. *J Neurophysiol* 94: 3314–3324, 2005.
- Baizer JS, Ungerleider LG, Desimone R. Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *J Neurosci* 11: 168–190, 1991.
- Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC. Do we know what the early visual system does? *J Neurosci* 25: 10577–10597, 2005.
- Chun MM, Potter MC. A two-stage model for multiple target detection in rapid serial visual presentation. *J Exp Psychol Hum Percept Perform* 21: 109–127, 1995.
- Dale AM, Buckner RL. Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Map* 5: 1–12, 1997.
- DeYoe EA, Carman GJ, Bandettini P, Glickman S, Wieser J, Cox R, Miller D, Neitz J. Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc Natl Acad Sci USA* 93: 2382–2386, 1996.
- Duncan J, Ward R, Shapiro K. Direct measurement of attentional dwell time in human vision. *Nature* 369: 313–315, 1994.
- Engel SA, Glover GH, Wandell BA. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex* 7: 181–192, 1997.
- Epstein R, Kanwisher N. A cortical representation of the local visual environment. *Nature* 392: 598–601, 1998.
- Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1: 1–47, 1991.
- Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4: 2379–2394, 1987.
- Foster KH, Gaska JP, Nagler M, Pollen DA. Spatial and temporal frequency selectivity of neurons in visual cortical areas V1 and V2 of the macaque monkey. *J Physiol* 365: 331–363, 1985.
- Gegenfurtner KR, Kiper DC, Levitt JB. Functional properties of neurons in macaque area V3. *J Neurophysiol* 77: 1906–1923, 1997.
- Grill-Spector K, Kanwisher N. Visual recognition: as soon as you know it is there, you know what it is. *Psychol Sci* 16: 152–160, 2005.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24: 187–203, 1999.
- Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzhak Y, Malach R. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Map* 6: 316–328, 1998.
- Grill-Spector K, Kushnir T, Hendler T, Malach R. The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci* 3: 837–843, 2000.
- Grill-Spector K, Malach R. The human visual cortex. *Annu Rev Neurosci* 27: 649–677, 2004.
- Gross CG. Representation of visual stimuli in inferior temporal cortex. *Philos Trans R Soc Lond B Biol Sci* 335: 3–10, 1992.
- Hawken MJ, Shapley RM, Gross DH. Temporal-frequency selectivity in monkey visual cortex. *Vis Neurosci* 13: 477–492, 1996.
- Hinkle DA, Connor CE. Three-dimensional orientation tuning in macaque area V4. *Nat Neurosci* 5: 665–670, 2002.
- Kanwisher N. Repetition blindness and illusory conjunctions: errors in binding visual types with visual tokens. *J Exp Psychol Hum Percept Perform* 17: 404–421, 1991.
- Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17: 4302–4311, 1997.
- Kelly DH. Visual response to time-dependent stimuli. I. Amplitude sensitivity measurements. *Rinsho Eiyo* 51: 422–429, 1961.
- Kelly DH. Motion and vision. II. Stabilized spatio-temporal threshold surface. *J Opt Soc Am A* 69: 1340–1349, 1979.
- Keysers C, Xiao DK, Foldiak P, Perrett DI. The speed of sight. *J Cogn Neurosci* 13: 90–101, 2001.
- Kovacs G, Vogels R, Orban GA. Cortical correlate of pattern backward masking. *Proc Natl Acad Sci USA* 92: 5587–5591, 1995.
- Levitt JB, Kiper DC, Movshon JA. Receptive fields and functional architecture of macaque V2. *J Neurophysiol* 71: 2517–2542, 1994.
- Levy I, Hasson U, Avidan G, Hendler T, Malach R. Center-periphery organization of human object areas. *Nat Neurosci* 4: 533–539, 2001.
- Liu H, Gao J. An investigation of the impulse functions for the nonlinear BOLD response in functional MRI. *Magn Reson Imaging* 18: 931–938, 2000.
- Loffler G, Gordon GE, Wilkinson F, Goren D, Wilson HR. Configural masking of faces: evidence for high-level interactions in face perception. *Vision Res* 45: 2287–2297, 2005.
- Logothetis NK, Pauls J, Poggio T. Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5: 552–563, 1995.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci USA* 92: 8135–8139, 1995.
- Marois R, Ivanoff J. Capacity limits of information processing in the brain. *Trends Cogn Sci* 9: 296–305, 2005.
- McCarthy G, Puce A, Gore JC, Allison AT. Face-specific processing in the human fusiform gyrus. *J Cogn Neurosci* 9: 605–610, 1997.
- McMains SA, Somers DC. Multiple spotlights of attentional selection in human visual cortex. *Neuron* 42: 677–686, 2004.
- Mukamel R, Harel M, Hendler T, Malach R. Enhanced temporal nonlinearities in human object-related occipito-temporal cortex. *Cereb Cortex* 14: 575–585, 2004.
- Nothdurft HC. Faces and facial expressions do not pop out. *Perception* 22: 1287–1298, 1993.
- Pasupathy A, Connor CE. Population coding of shape in area V4. *Nat Neurosci* 5: 1332–1338, 2002.
- Potter MC. Meaning in visual search. *Science* 187: 965–966, 1975.
- Potter MC, Faulconer BA. Time to understand pictures and words. *Nature* 253: 437–438, 1975.

- Raymond JE, Shapiro KL, Arnell KM.** Temporary suppression of visual processing in an RSVP task: an attentional blink? *J Exp Psychol Hum Percept Perform* 18: 849–860, 1992.
- Rolls ET, Tovee MJ.** Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc Biol Sci* 257: 9–15, 1994.
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB.** Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268: 889–893, 1995.
- Sharpe CR.** The contrast sensitivity of the peripheral visual field to drifting sinusoidal gratings. *Vision Res* 14: 905–906, 1974.
- Simoncelli EP, Olshausen BA.** Natural image statistics and neural representation. *Annu Rev Neurosci* 24: 1193–1216, 2001.
- Singh KD, Smith AT, Greenlee MW.** Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage* 12: 550–564, 2000.
- Tanaka K.** Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139, 1996.
- Thorpe S, Fize D, Marlot C.** Speed of processing in the human visual system. *Nature* 381: 520–522, 1996.
- Tong F.** Primary visual cortex and visual awareness. *Nat Rev Neurosci* 4: 219–229, 2003.
- Tong F, Nakayama K.** Robust representations for faces: evidence from visual search. *J Exp Psychol Hum Percept Perform* 25: 1016–1035, 1999.
- Tong F, Nakayama K, Vaughan JT, Kanwisher N.** Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron* 21: 753–759, 1998.
- Tong F, Nakayama K, Moscovitch M, Weinrib O, Kanwisher N.** Response properties of the human fusiform face area. *Cogn Neuropsychol* 17: 257–279, 2000.
- Treisman AM.** Features and objects in visual processing. *Sci Am* 255: 114B–125B, 1988.
- Tsao DY, Freiwald WA, Tootell RB, Livingstone MS.** A cortical region consisting entirely of face-selective cells. *Science* 311: 670–674, 2006.
- Virsu V, Rovamo J, Laurinen P, Nasanen R.** Temporal contrast sensitivity and cortical magnification. *Vision Res* 22: 1211–1217, 1982.
- Wiens S, Fransson P, Dietrich T, Lohmann P, Ingvar M, Ohman A.** Keeping it short: a comparison of methods for brief picture presentation. *Psychol Sci* 15: 282–285, 2004.
- Williams PE, Mechler F, Gordon J, Shapley R, Hawken MJ.** Entrainment to video displays in primary visual cortex of macaque and humans. *J Neurosci* 24: 8278–8288, 2004.
- Woods RP, Grafton ST, Holmes CJ, Cherry SR, Mazziotta JC.** Automated image registration: I. General methods and intrasubject, intramodality validation. *J Comput Assist Tomogr* 22: 139–152, 1998.
- Wright MJ, Johnston A.** Spatiotemporal contrast sensitivity and visual field locus. *Vision Res* 23: 983–989, 1983.