Behavioral/Systems/Cognitive

Temporal Tuning Properties along the Human Ventral Visual Stream

Baptiste Gauthier,[1,2,3] **Evelyn Eger,**[1,2,3] **Guido Hesselmann,**[1,2,3,4] **Anne-Lise Giraud,**[5] **and Andreas Kleinschmidt**[1,2,3,6]

[1]Institut National de la Santé et de la Recherche Médicale Unité 992, Cognitive Neuroimaging Unit, and [2]NeuroSpin, Institut d'Imagerie Biomédicale/Direction des Sciences du Vivant/Commissariat à L'Énergie Atomique, Saclay, F-91191 Gif-sur-Yvette, France, [3]Université Paris-Sud, F-91405 Orsay, France, [4]Visual Perception Laboratory, Department of Psychiatry, Charité Campus Mitte, D-10117 Berlin, Germany, [5]Institut National de la Santé et de la Recherche Médicale Unité 960, Ecole Normale Supérieure, F-75005 Paris, France, and [6]Department of Clinical Neurosciences, University Hospital (HUG) and University of Geneva, CH-1211 Geneva, Switzerland

Both our environment and our behavior contain many spatiotemporal regularities. Preferential and differential tuning of neural populations to these regularities can be demonstrated by assessing rate dependence of neural responses evoked during continuous periodic stimulation. Here, we used functional magnetic resonance imaging to measure regional variations of temporal sensitivity along the human ventral visual stream. By alternating one face and one house stimulus, we combined sufficient low-level signal modulation with changes in semantic meaning and could therefore drive all tiers of visual cortex strongly enough to assess rate dependence. We found several dissociations between early visual cortex and middle- and higher-tier regions. First, there was a progressive slowing down of stimulation rates yielding peak responses along the ventral visual stream. This finding shows the width of temporal integration windows to increase at higher hierarchical levels. Next, for fixed rates, early but not higher visual cortex responses additionally depended on the length of stimulus exposure, which may indicate increased persistence of responses to short stimuli at higher hierarchical levels. Finally, attention, which was recruited by an incidental task, interacted with stimulation rate and shifted tuning peaks toward lower frequencies. Together, these findings quantify neural response properties that are likely to be operational during natural vision and that provide putative neurofunctional substrates of mechanisms that are relevant in several psychophysical phenomena as masking and the attentional blink. Moreover, they illustrate temporal constraints for translating the deployment of attention into enhanced neural responses and thereby account for lower limits of attentional dwell time.

Introduction

Some of the first positron emission tomography and functional magnetic resonance imaging (fMRI) experiments studied stimulus rate-dependent modulations of hemodynamic responses (Fox and Raichle, 1985; Kwong et al., 1992). These and most subsequent studies used simple stimuli and in a given range found near-linear response increases in low-level cortical areas with stimulation frequency (Singh et al., 2000; Ozus et al., 2001; Liu and Wandell, 2005; Mullen et al., 2010; D'Souza et al., 2011). These findings are congruent with predominantly phasic rather than tonic neural responses and established that hemodynamic signals can measure temporal tuning of neural populations.

To drive not just early but also high-level visual cortex, two studies varied rate in serial presentations of different face or house images

(Mukamel et al., 2004; McKeeff et al., 2007). We built on this earlier work but chose periodic stimulation by alternating a single face and a single house image. From the low-level perspective, this paradigm has the advantage of a clearly defined sensory modulation at a single frequency, comparable with a flicker or grating as used in studies of lower-tier visual areas. We measured regional temporal tuning functions over a very wide frequency range and tested how much they change along the ventral visual stream. Figure 1 schematically illustrates the predictions as to how this type of periodic stimulation should translate into fMRI responses as a function of stimulation rate and regional temporal sensitivity. We expected that just as perception and behavior express upper limits of tractable frequencies, cortical population responses should increase with stimulation rate but saturate when approaching fusion frequencies with activity reductions at rates beyond peak.

Two mechanisms could underlie temporal tuning. It could be that, to achieve full amplitude, neural responses require a minimal time of sustained sensory evidence (i.e., exposure to a given sensory stimulus). Or neural processing or integration time could be the limiting factor such that full response strength depends not so much on exposure time but is diminished if other input arrives too early after stimulus onset and interferes. To discriminate effects of stimulus duration and rate, we therefore studied response modulation by duration but for fixed rates.

In addition to "bottom–up" factors like stimulus duration and rate, the "top–down" effect of attention on hemodynamic brain signals has also been investigated since early on (Corbetta et al., 1990) and progressively been tracked down to the earliest accessible stages of the sensory processing chain (O'Connor et al., 2002). Many studies have established that attention increases responses to stimulation, and some studies have attempted to disentangle contributions from background activity and response gain (Chawla et al., 1999; Kastner and Ungerleider, 2000). However, we are not aware of studies that have shown whether attention changes neural temporal sensitivity. In a second experiment, we tested whether the effects of attention interact with those of stimulation rate. As attention relies at least in part on hierarchical feedback loops, its local cortical correlates manifest with greater latency than stimulus-driven feedforward responses (Martínez et al., 1999; Noesselt et al., 2002). We therefore speculated that attention might introduce additional time constraints for achieving maximal responses and hence shift tunings peak to slower rates.

Materials and Methods

Subjects. Twenty-four healthy volunteers (age, 22 ± 5 years; 10 females; 5 left-handers) participated in the study, 11 in Experiment 1 and 13 in Experiment 2. Subjects had normal or corrected-to-normal visual acuity. All subjects were in good health with no history of psychiatric or neurological disorders and gave written informed consent. The principal investigator (A.K.) had local ethics committee approval for this study.

Experimental paradigm. Visual stimuli were presented via a backprojection display (1024 × 768 resolution, 60 Hz refresh rate) with a uniform background (black in Experiment 1, gray in Experiment 2). E-prime software (Psychology Software Tools) was used for presenting stimuli, recording behavioral responses, and synchronizing experimental timing with scanner pulse timing. Stimuli covered 6.5 × 9° of visual angle. Participants were instructed to maintain fixation on a central mark (a small square in Experiment 1; a short bar in Experiment 2). Unfortunately, no eye-tracking equipment was available for this study, which would have permitted monitoring subject compliance with the instruction.

Both experiments used a face and a house stimulus adapted from the study by Kriegeskorte et al. (2007). In the first part of Experiment 1, stimulation followed a block design: visual stimulation blocks (20 s) alternated with 10 s baseline blocks with a gray screen (Fig. 2*A*). During a given stimulation block, the two pictures, of a face and a house, were alternated at a fixed frequency without any gap. We tested nine alternation rates across different blocks: frame lengths of 33, 50, 100, 200, 400, 800, 1600, 3200, and 4800 ms per picture.

In the second part of Experiment 1 (Fig. 2*A*), we used only two different alternation rates (every 400 or 800 ms) but varied for each block with a given rate the duration for which the pictures were presented, with values of 33, 50, 100, 200, and 400 ms for the rate every 400 ms and of 33, 50, 100, 400, and 800 ms for the rate every 800 ms. Except when frame length of pictures was equivalent to the rate,
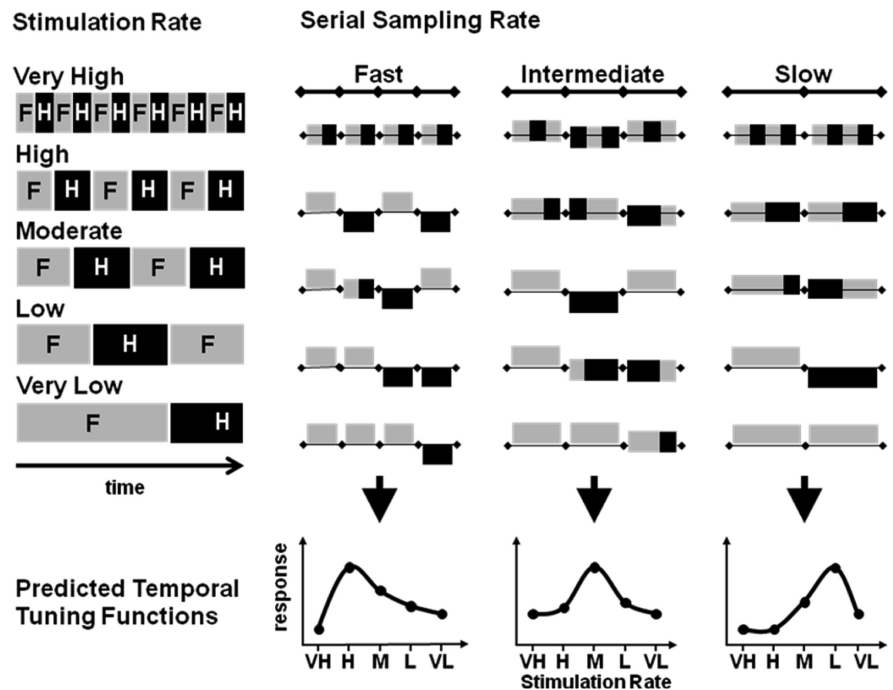
Figure 1. Schematic illustrating predicted determinants of cortical temporal tuning. These predictions are grounded in a framework assuming that cortical regions differ intrinsically in the width of the temporal window over which they process information, that the processing of a continuous flow of sensory information involves serial sampling at a rate corresponding to window width, and that responses are phasic (i.e., determined by changes of resampled sensory signals). The left-hand column indicates a periodic input, the alternation of a single face and a single house image (indicated by F and H) at rates ranging from very high (VH) to very low (VL). The second to fourth columns illustrate the sensory content of successive windows of integration for areas differing in their intrinsic sampling rate. The sensory content can correspond entirely to one image if the neural sampling window is at least as short as stimulus duration or contain mixtures of the two frames. Sensory content is graphically indicated by the vertical position of each integration window relative to the horizontal line. We assume that fMRI signal is sensitive to the degree of variation of sensory content occurring over successive windows of integration. The ensuing tuning functions for different cortical areas are indicated at the bottom of the figure.

the screen returned to baseline gray during the gap before appearance of the next picture.

In Experiment 2 (Fig. 2*B*), we compared two sets of conditions. In the first one, stimulation blocks (20 s) alternated with baseline blocks (10 s) of uniform low-luminance gray screen. During stimulation blocks, face and house pictures alternated at a fixed frequency without any gap in the same way as in Experiment 1. We tested nine conditions of alternation rates across different blocks: frame lengths per picture of 50, 75, 100, 125, 150, 175, 200, 250, and 400 ms. In the second set, we repeated the same stimulation but additionally introduced randomly twice per block targets events during which the fixation bar presented a 50 ms tilt by 90°. Subjects were asked to detect and report these events by pressing a button (Fig. 2*B*). These conditions of active viewing were grouped into sessions and sessions of active and passive viewing were randomly interspersed with the according instruction presented at the onset of each session. We deliberately chose such an incidental task so as not to disrupt the stream of face/house alternations, not to bias categorical attention toward either of the two stimuli, and to avoid gross rate condition by task difficulty interactions. In the context of our questions, the only purpose of the task manipulation was to ensure that participants maintained a high degree of attention directed toward central parts of the visual stimulus stream they were exposed to.

In both experiments, functional localizer sessions were organized as a block design with each stimulation block comprising 12 alternations of a picture (500 ms) and a white screen (500 ms), followed by 6 s baseline blocks with a white screen. Pictures in a block belonged to one of four categories: faces, places, objects, and scrambled pictures (made by randomly reassembled parts of images picked from one of the three other categories). Each block with a given category and the following baseline block ensemble (12 s + 6 s = 18 s) were repeated eight times per subject.
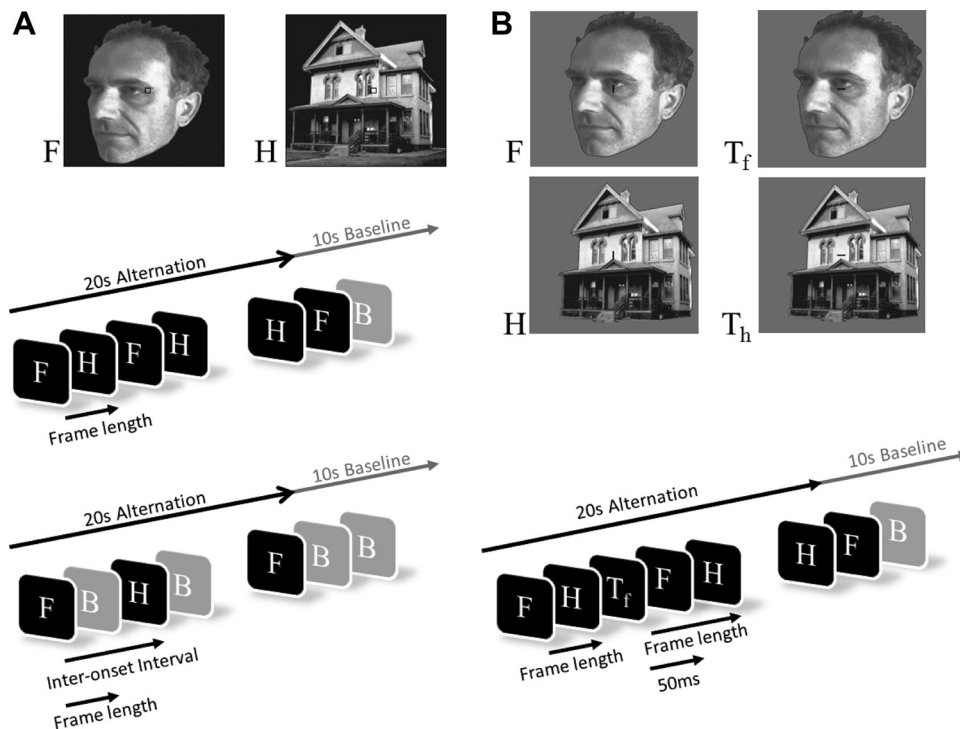
Figure 2. Experimental paradigm. *A*, Face and house pictures used in Experiment 1 (top) along with schematic time line of one block in the first (middle) and second part (bottom) of Experiment 1: F, H, and B stand for face, house, and baseline, respectively. *B*, Face and house pictures used in Experiment 2 (top) along with schematic timeline of one block in the experiment that was repeated once with and once without a task for all stimulation rates (bottom). T_f and T_h indicate 50 ms frames that were inserted into the overall frame length with a fixed stimulation rate and that contained a tilted fixation bar that participants were instructed to report.

To ensure active processing, subjects were engaged in a one-back working memory task in which they pressed a button whenever a picture was immediately repeated.

Data acquisition. Imaging was performed on a 3 T MRI scanner (TIM Trio; Siemens). Each participant underwent 7 min anatomical imaging using a T1-weighted MPRAGE sequence (160 slices; repetition time, 2300 ms; echo time, 2.98 ms; FOV, 256; voxel size, $1 \times 1 \times 1$ mm^3) and five sessions of functional imaging with blood oxygen level-dependent contrast using a T2*-weighted gradient-echo echo-planar imaging sequence (25 slices; repetition time, 1500 ms; echo time, 30 ms; voxel size, $3.5 \times 3.5 \times 3.5$ mm^3). For Experiment 1, each of the four experimental fMRI sessions with 403 volumes lasted 10 min and 5 s. For Experiment 2, each of the four experimental fMRI sessions with 410 volumes lasted 10 min and 9 s. Localizer sessions with 390 volumes took 9 min and 45 s.

Functional data analysis. We used statistical parametric mapping (SPM5; Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/) for image preprocessing with slice timing correction, realignment, coregistration with the structural image, normalization to Montreal Neurological Institute stereotactic space, and spatial smoothing with a 5 mm full-width half-maximum isotropic Gaussian kernel.

General linear models were estimated subjectwise on the basis of a design matrix that covered all five fMRI sessions and included regressors for every experimental condition of the paradigm (convolved with a canonical hemodynamic response function) as well as nuisance covariates from the motion parameters and their first derivatives and session blocks. A high-pass filter (128 s cutoff) was applied to remove slow drifts unrelated to the paradigm. Regressors of interest for the first part of Experiment 1 were each frequency condition (RATE) and a baseline condition comprising all baseline blocks from the first part, for the second part each duration condition (DUR) per frequency and a condition comprising all baselines blocks. In Experiment 2, we modeled regressors for the baseline condition as well as for each block with a given frequency and as a function of the task, active or passive viewing, as well as for the occurrence of target events, separately for misses and hits. For the local-

izer session, we modeled five regressors of interest: for each picture category and for the baseline blocks. Analysis of condition-dependent fMRI signal differences were based on the estimated β weights for peak voxels defined in contrasts of interest in both left and right brain hemispheres since each region of interest (ROI) was activated bilaterally. Peak voxels in occipital calcarine sulcus were identified subjectwise from the contrast "all RATEs minus baseline" in the functional runs of the main experiment to represent early visual cortex ("V1/V2"). The localizer scans permitted to define peak coordinates for category-specific regions. The peak voxel of the fusiform face area (FFA) was defined as the most active voxel in a subjectwise *t* contrast of "face minus scrambled pictures" masked with "face minus baseline." Equivalent procedures with "place" and "object" conditions were used to define peak voxels from parahippocampal place area (PPA) and lateral occipital complex (LOC). For the latter, given its composition of subregions, we defined in each subject both a ventral temporal and a lateral occipital peak voxel ($_v$LOC and $_l$LOC).

To build tuning profiles, we first contrasted the β weight of each of the conditions with that of the baseline and then z-transformed these values for each session and subject. As no significant differences were found between hemispheres or within tiers of the visual hierarchy, further analyses were conducted after collapsing individual β profiles between the two hemispheres and within each of three levels of the ventral visual hierarchy, low- (V1/V2), mid- (LOC), and high-level (FFA, PPA). We then derived peak frequency values from subject by subject polynomial fits to the tuning profiles for the three tiers [similar to the studies by Hagenbeek et al. (2002), McKeeff et al. (2007), and D'Souza et al. (2011)]. A leave-one-out cross-validation approach confirmed that third-order was the most suitable polynomial degree to minimize fit error while obtaining acceptable goodness of fit values. To account for our nonlinear spacing of frequencies, these values were log-transformed before fitting the curves, providing a better goodness of fit.

ANOVAs were performed both on the tuning curves (all frame lengths) and fitted peaks (extrapolated peak frame lengths) to explore the effects of the factors relevant for each experiment. Specifically, in the factorial statistical analyses, we explored the effects of two factors

in part 1 of Experiment 1, RATE with nine levels and ROI with three levels: "low level," "medium level," and "high level." In part 2 of Experiment 1, we used the same ROI factor and complemented the two-level factor RATE (400 or 800 ms) by embedding a duration factor (DUR) with five levels for each of the two RATEs. In Experiment 2, we complemented the ROI and RATE factors by an additional third factor, the task-induced effect (TASK) with two levels corresponding to passive viewing (no task) and active viewing (during the detection task). To address adaptation effects, we conducted further analyses for both experiments in which β weights were determined for the first and second repetition of each condition, which permitted introducing a REPETITION factor into the related ANOVAs. Significance levels in pairwise *post hoc t* tests were corrected for multiple comparisons with the Holm–Bonferroni method.

Results

Localizer and main activations: ventral stream subdivision

Figure 3, *A* and *B*, provides an overview of the topography of stimulus-driven activations across all conditions versus baseline in Experiment 1 and Experiment 2, respectively. As detailed above, the subsequent analysis of interest was based on condition-dependent signal changes from functionally identified peak response voxels defined subject by subject and contrast by contrast. Table 1 lists average stereotactic coordinates of these voxels for Experiment 1. Their spatial distribution across subjects and per region is presented in Figure 4. The mean coordinates are similar to those reported in the literature for early visual cortex and category-specific regions (Downing et al., 2006) and, while slightly variable, nonetheless largely consistent across subjects.

Temporal tuning along the ventral visual stream (Experiment 1)

As a test of our first question, we analyzed mean regional fMRI responses as a function of stimulation frequency (Fig. 4*B*). Each region along the ventral stream showed the same overall tuning profile in which, with increasing frame durations, responses first became stronger up to a maximum and then decreased to lower values with even longer frames. In overall accord with a previous report in the literature (McKeeff et al., 2007), the rates at which responses peaked, changed with regions and showed a progressive slowing down of the best stimulation rate along the ventral visual stream.

To further test for regional tuning differences, we collapsed data from left and right hemispheres as well as within tiers, with V1/V2 as low-level, the two portions of LOC as mid-level, and FFA and PPA as high-level visual cortex. A repeated-measures ANOVA with RATE and ROI as factors showed a significant effect of RATE on the response profile ($F_{(8,80)} = 19.6$; $p < 0.001$) and revealed a RATE by ROI interaction ($F_{(16,160)} = 10.59$; $p < 0.001$).

As our experiment involved only two images and hence extensive stimulus repetition, we also assessed adaptation effects on tuning functions. Similar to the above analysis, we used an ANOVA with the additional factor of session but obtained no
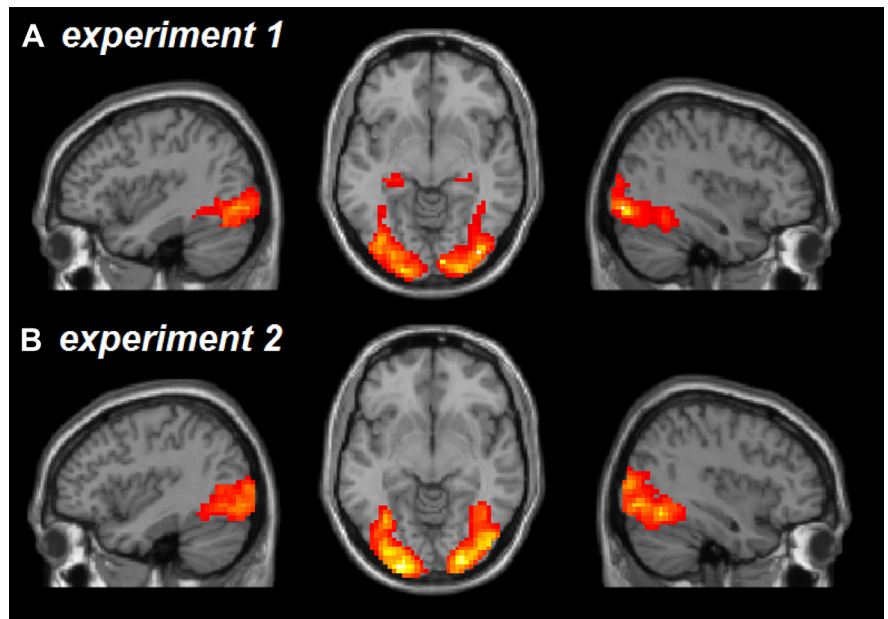

Figure 3. Topographical overview of the extent of stimulus-driven brain activation in both Experiment 1 (*A*) and Experiment 2 (*B*). Color-coded overlay of results from a second-level analysis across all participants and conditions (random-effects model, contrast of all conditions of interest vs baseline, thresholded at $p < 0.001$, uncorrected). Activations are superimposed onto structural brain images and displayed in three representative sections (left sagittal at $x = -38.5$ mm, transverse at $z = -7$ mm, and right sagittal at $x = 38.5$ mm).

Table 1. Stereotactic coordinates of voxels from the ventral visual stream regions that were studied in Experiment 1 and Experiment 2

ROI	x (SD)	y (SD)	z (SD)
Left			
V1/V2	−15.3 (6.2)	−97.6 (4.0)	−1.3 (6.4)
lLOC	−46.5 (3.0)	−76.0 (5.5)	−3.4 (4.8)
vLOC	−39.4 (5.5)	−57.0 (7.2)	−16.3 (5.0)
PPA	−25.7 (3.0)	−48.7 (5.2)	−9.0 (4.2)
FFA	−37.3 (3.0)	−49.9 (5.5)	−19.1 (3.1)
Right			
V1/V2	20.3 (5.7)	−96.3 (4.0)	0.6 (6.7)
lLOC	48.1 (3.8)	−72.8 (5.9)	−4.1 (4.6)
vLOC	39.5 (4.3)	−56.6 (7.6)	−15.3 (4.2)
PPA	26.3 (4.0)	−47.5 (7.8)	−8.9 (4.1)
FFA	38.2 (3.1)	−46.8 (7.5)	−18.2 (4.6)

Values represent mean ± SD and are expressed in millimeters.

significant main effect or interaction involving session. Since repetition effects in our setting might well be most pronounced during the first trials (Grill-Spector and Malach, 2001), we estimated separately β weights for the first and second repetition of each condition in the first session. With this approach, an ANOVA with the factors REPETITION, RATE, and ROI again showed no main effect of REPETITION nor an interaction with RATE but a significant interaction with ROI ($F_{(2,20)} = 6.86$; $p < 0.01$). This interaction was driven by the fact that adaptation effects occurred only in mid- and higher-tier cortex.

Finally, we estimated and compared the peaks of regional tuning functions (i.e., the frame durations for which maximal fMRI responses were observed). Subject by subject, we fitted third-order polynomial functions to the β weights averaged for each tier in the ventral visual hierarchy, and extracted tuning peak rates from them (Fig. 5*A,B*). We further confirmed that tuning profile peaks differ significantly across ROIs by an ANOVA ($F_{(2,10)} = 13.4$; $p < 0.001$) and ensuing pairwise comparisons with paired *t* tests performed on
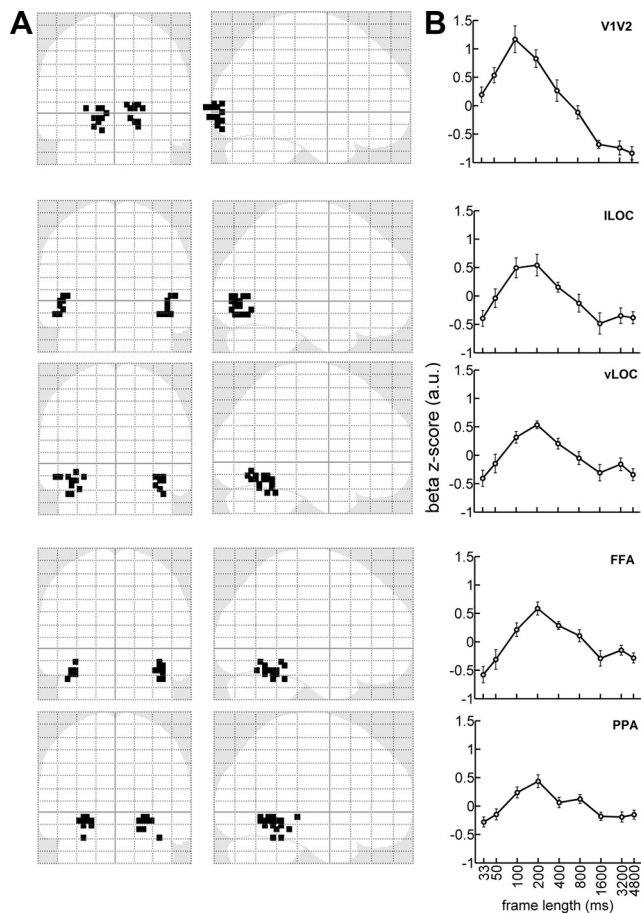
Figure 4. Overview of the tuning curves from the first part of Experiment 1 across selected brain regions along the ventral stream. *A*, Coronal and sagittal glass-brain views of the most activated voxels selected for each brain region (1 voxel per region and participant). *B*, Normalized mean β z-scores ($\pm$SE) from Experiment 1, for each region, averaged across all participants.

the individually fitted peak response rates from the different tiers. Low-level visual cortex tuning peaks were found at significantly higher rates (average peak, 109 ms frame length) than medium-level (average peak, 153 ms; $t_{(10)} = -2.89; p < 0.05$) and high-level cortex (average peak, 213 ms; $t_{(10)} = -4.33; p < 0.005$). High-level tuning peaks were at significantly lower rates than those from mid-level ROIs ($t_{(10)} = -2.94; p < 0.05$).

Sensory exposure versus neural integration time
We used data from the second part of Experiment 1 to address the question of whether the tuning profiles result from the duration of sensory exposure or from the time available for processing and integration before novel incoming sensory information. If, for instance, higher-level visual cortex simply needed longer exposure times (DUR) to develop peak responses than lower levels, then the response within a given rate (RATE) should increase linearly with the length of stimulus exposure. We found the opposite to be true. Both in mid- and high-level ventral visual cortex, response strength for a given rate was near maximal even with the shortest durations of stimulus exposure that we tested and did not increase significantly with longer stimulus exposure times (Fig. 6). Conversely, in low-level visual cortex, there was a near-linear benefit from longer sensory exposure and this effect did not pass ceiling within the RATE values tested (400 and 800 ms). In other words, in this setting with a fixed rate, responses were max-

imal at stimulus durations well above those that had shown a peak in the tuning profiles studied in the first part of this experiment in which stimulus duration served to implement rate by way of periodic alternation.

An ANOVA performed on these data with the factors of RATE, DUR, and ROI revealed a three-way interaction RATE by DUR by ROI ($F_{(6,60)} = 5.918; p < 0.001$) and hence showed a clear dissociation between low- and mid- and high-level regions for the effect of duration. When exploring this interaction separately for the two RATE levels, two-way ANOVAs revealed DUR by ROI interactions for both 400 and 800 ms rates ($F_{(8,80)} = 28.9$, $p < 0.001$; $F_{(8,80)} = 42.5$, $p < 0.001$, respectively). Mid- and high-level cortex showed no significant effect of duration for either of the two rates but low-level visual cortex did for both ($F_{(4,40)} = 49.8, p < 0.001$; $F_{(4,40)} = 55.7, p < 0.001$, respectively, for 400 and 800 ms). Moreover, we found that mean responses of low-level cortex in the 400 ms conditions were significantly higher than in the 800 ms conditions ($t_{(54)} = 7.27; p < 0.001$), as predicted by the tuning curve obtained in the first part of the experiment. This observation simply reflects the greater number of stimulus transitions occurring in a given block length of the 400 ms compared with the 800 ms conditions. In low-level visual cortex, fMRI responses were readily linearly fit to duration (adjusted $R^2 = 0.41$, $F_{(4,50)} = 10.34$, $p < 0.001$, and adjusted $R^2 = 0.67$, $F_{(4,50)} = 28$, $p < 0.001$, for RATE 400 and 800 ms, respectively).

Attentional modulation of temporal tuning functions (Experiment 2)
In Experiment 2, we addressed whether and how such temporal tuning profiles are modulated by attention. We therefore replicated a similar design as in the first experiment but with a reduced and more finely graded range of stimulation rates. In addition, we engaged participants not only in a "passive viewing" condition as in Experiment 1 but also obtained data with the same rates while they were engaged in active viewing due to a task that required maintaining attention allocated to the central part of the stimulus. Accuracy on this very demanding task was good but not at ceiling (78% hits) and not affected by stimulation rate (ANOVA on RATE effect on hits, $F_{(8,96)} = 7.93, p = 0.25$). Reaction times were affected by rate (ANOVA, $F_{(8,96)} = 3.08, p < 0.005$), but pairwise t testing showed that only RTs corresponding to the longest frame length (400 ms; $p < 0.05$) were significantly slower than the others. We did not observe any significant change in accuracy or reaction time between the first and second session involving active viewing.

From previous findings in the literature, it seemed conceivable that attentional modulation of neural responses to stimulation could manifest in an upward shift [baseline offset (Kastner et al., 1999)] or in a multiplicative effect [gain control (Hillyard et al., 1998)] or a mixture of both. Neither of these scenarios should necessarily result in a change of temporal sensitivity as indexed by the peak of the tuning function. Our first analysis was therefore based on the estimated peaks of the regional tuning functions. These tuning peaks were again, as in the first experiment, derived from fitting subject-by-subject and tier-by-tier third-order polynomials for each of the conditions, passive and active viewing (Fig. 7). An ANOVA was performed on the peak response values derived from these fits and was tested for effects of the factors TASK and ROI. We found no interaction but significant main effects of both ROI ($F_{(2,24)} = 5.5; p < 0.005$) and TASK ($F_{(1,12)} = 8.3; p < 0.05$). This finding confirms the result of Experiment 1 that tuning peaks move to slower rates along the ventral visual

hierarchy. Additionally, this finding establishes that attention slows down tuning peak values throughout all levels of the hierarchy. This effect was qualitatively consistent across regions but offset in accordance with the regional differences in absolute tuning peak positions during passive viewing. In other words, attention preserved the overall slowing down of tuning peaks along the visual hierarchy.

To obtain a more fine-grained analysis of what accounted for these attention-related effects, we analyzed fMRI responses across all different stimulation rates. The tuning functions confirmed that attention shifted maximal responses to longer frame durations (lower rates) and hence a more complex scenario than one would have predicted for a baseline shift or increased response gain (Fig. 8). An ANOVA on the basis of ratewise fMRI responses and with the additional factors of TASK and ROI revealed a significant three-way interaction involving all factors ($F_{(16,192)} = 2.46$; $p < 0.005$). In other words, the RATE by ROI interaction that we had already established in the first experiment was significantly modulated by TASK. We further explored this effect by testing at each level of the visual hierarchy for RATE by TASK interactions. We found a significant RATE by TASK interaction for higher-tier cortex ($F_{(8,96)} = 2.49$; $p < 0.05$) but not for middle and lower tiers in which only the main effect of RATE was significant ($F_{(8,96)} = 6.3$, $p < 0.001$; $F_{(8,96)} = 11.56$, $p < 0.001$, respectively).

Together, the results from Experiment 2 show that, across the entire ventral visual hierarchy, recruitment of attentional mechanisms by a task shifts tuning peaks to slower rates and that this effect is most pronounced in high-level visual cortex. As for Experiment 1, we also tested for repetition effects. They were comparable with those in Experiment 1 with no REPETITION by RATE interaction (suggesting stability of temporal tuning over time) but a significant REPETITION by ROI interaction ($F_{(2,24)} = 7.12$; $p < 0.001$) because adaptation was only significant in mid- and higher-tier cortex ($F_{(3,36)} = 13.2$, $p < 0.001$; $F_{(3,36)} = 8.8$, $p < 0.001$, respectively). The interaction of TASK with REPETITION was not significant with only a trend toward stronger adaptation in these regions during active compared with passive viewing, thus in line with previous findings (Eger et al., 2004; Murray and Wojciulik, 2004).

Discussion

One of the motivations for studying rate-dependent neural responses is to define and localize neural substrates of psychophysical performance. Temporal sensitivity for visual stimuli has been intensively investigated behaviorally. Its inverted U-shape usually peaks around 10 Hz and reaches up to 30 to 50 Hz for strong low-level stimuli (Hart, 1987). Performance involving grouping
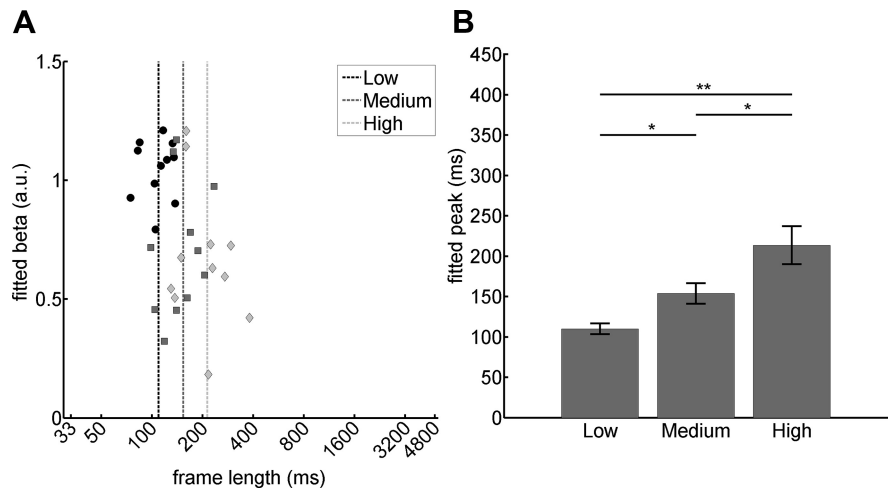

Figure 5. Slowing down of tuning peaks along the ventral visual stream. **A**, Scatterplot with the results from fitted curves from all ventral visual tiers in all participants. Individual fitted peak values (black dots, low-level peaks; dark gray squares, midlevel peaks; light gray diamonds, high-level peaks) are plotted against corresponding β weights. Average tuning peak values are indicated by dotted vertical lines with the same color code for the three tiers. **B**, Average temporal tuning peak values ($\pm$SE) for each of three tiers of the ventral visual stream. *$p < 0.05$; **$p < 0.005$.
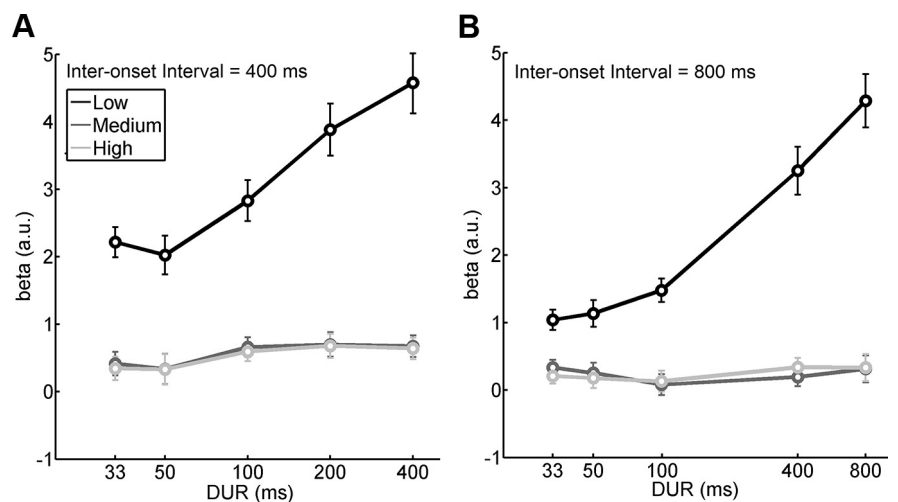

Figure 6. Average response curves from lower-, middle-, and higher-tier regions of the ventral visual stream as a function of varying stimulus duration (DUR) for a fixed stimulation rate either every 400 or 800 ms (second part of Experiment 1). **A**, Mean β weights ($\pm$SE) for RATE of 400 ms. **B**, Same figure for RATE of 800 ms.

or categorization on more complex stimuli in rapid serial visual presentation becomes virtually impossible once rates exceed 10 Hz (Holcombe, 2009). In our paradigm, perception at higher frequencies no longer preserves a sensation of alternating semantic content but instead fuses the face and house image and unties this percept from that of a persistent scene flicker. This effect that we deliberately sought to create motivated the use of a single picture in each category (Fig. 1). It has the additional advantage of avoiding an effect that occurs when serially presenting different pictures rather than periodic alternation of just two and that consists in a smearing of effective visual stimulation into lower frequencies than the upper cutoff from frame length. This difference may explain why in a previous fMRI study using strings of face or house pictures, tuning peaks in primary visual cortex were reported for rates of >20 items/s and in FFA and PPA for rates between 5 and 10 items/s (McKeeff et al., 2007). Importantly, that study showed that it did not matter for tuning in FFA and PPA whether strings were composed only of house or face images or
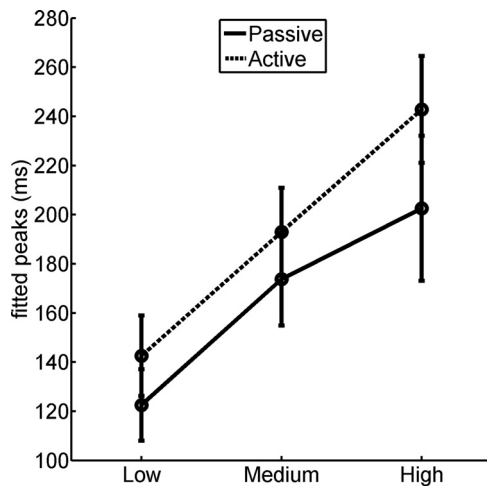
Figure 7. Average fitted peak values (±SE) for each level of visual cortex and for both passive and active conditions (Experiment 2).

Figure 8. Effect of task on temporal tuning curves (Experiment 2) across visual tiers with a peak shift to slower rates. *A–C*, Mean β z-scores (±SE) from Experiment 2, respectively, for low-, medium-, and high-level ventral visual cortex.

alternated between the two categories. Accordingly, we consider that the relevant rate values in our paradigm are provided by frame durations and not by the frequencies of preferred categorical content (one-half the rate of stimulation).

In early visual areas, the tuning functions we obtained seem well in line with most previous neurophysiological investigations with low-level stimuli that have reported neural tuning peaks ranging from 3 to 10 Hz (Foster et al., 1985; Singh et al., 2003). By using a face and a house stimulus in the present study, we additionally could measure tuning functions in response to the same sensory input across all tiers of the ventral visual stream. The profiles we obtained had an inverted U-shape for all visual regions tested and showed a progressive slowing down of tuning peaks when moving up in the ventral visual hierarchy. This decreasing temporal sensitivity along the ventral visual stream matches well the aforementioned material-dependent psychophysical sensitivity variations.

At first glance, these findings might seem to contradict the fact that even ultrashort presentation can suffice for successful recognition of meaningful visual stimuli (Thorpe et al., 1996). One possible explanation for this apparent discrepancy might be that cortical responses differ not only with respect to the length of sensory stimulation that gives maximal responses but also in the width of an integration window over which this information needs to processed. This possibility was addressed in the second part of our first experiment in which we chose stimulation rates that were slower than the tuning peaks of regions along the ventral visual stream and varied the length of sensory exposure and ensuing interstimulus interval while maintaining a constant rate. We found a clear dissociation: Early visual cortex activity increased with sensory exposure length well beyond its intrinsic tuning peak during rhythmic stimulation. Conversely, responses in higher-tier cortex were already near-maximal with much shorter exposure lengths than suggested by the tuning peaks and did not significantly increase further with exposure length. At the level of neural discharges, this distinction could be due to a gradual change from strong stimulus-locking in early visual areas to greater persistence of responses in mid- and higher-tier cortex. The latter effect is well known from direct neural recordings in electrophysiological studies and has been proposed to impact on fMRI signals and to constitute a putative substrate of iconic memory (Coltheart, 1983; Rolls and Tovee, 1994; Mukamel et al., 2004; Keysers et al., 2005).
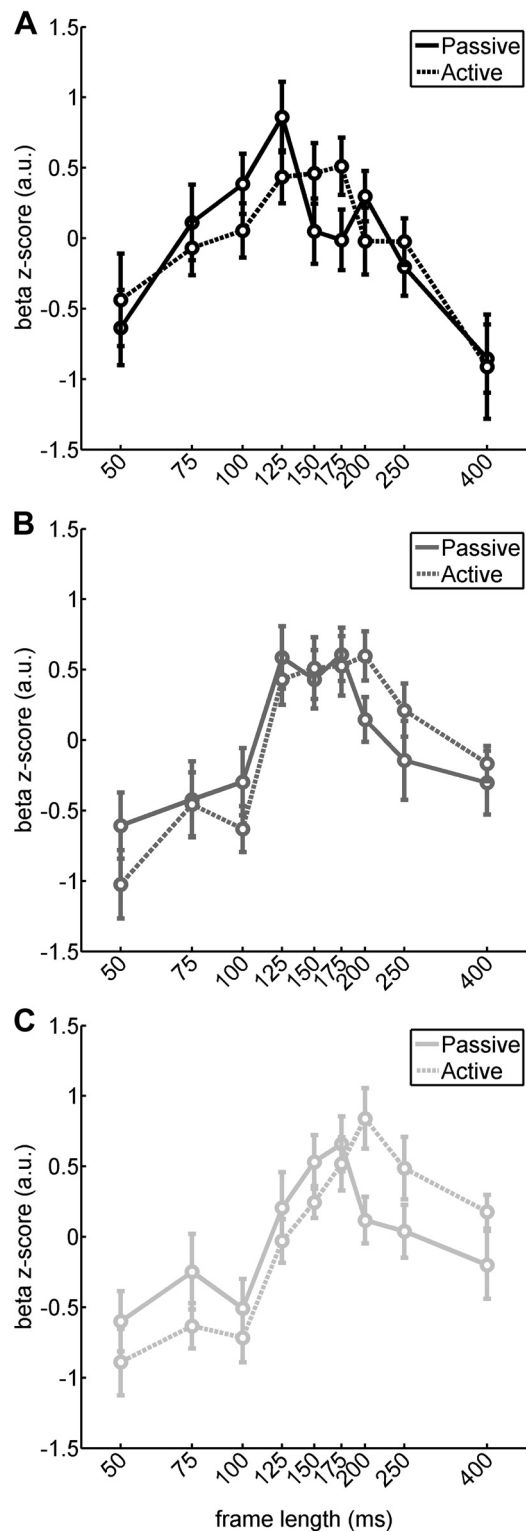
Our observations cover the time range from sufficient duration for recognizing ultrabrief semantic stimuli (Thorpe et al., 1996) to beyond the maximal latency for masking their perception by subsequent sensory input (Enns and Di Lollo, 2000). Persisting neural activity in higher-tier brain areas could constitute a source for this window of perceptual vulnerability that

extends well beyond stimulus exposure. Conversely, within the capacity and time range probed traditionally, subsequent sensory input is not inevitably detrimental nor in itself less perceived, as illustrated by the sparing of the first item after target in a typical attentional blink setting (Potter et al., 1998). This latter phenomenon might suggest that what is being blinked is not the sensory input that immediately follows a target but the content of the next window of integration, an interpretation that suggests the temporal properties of sensory sampling to define constituent chunks for ensuing actions.

The functional significance of temporal sensitivity has maybe been more firmly established in the auditory domain (Giraud et al., 2000), but the present findings from the predominantly "spatial" modality of vision also seem ecologically plausible for natural vision because they suggest that the usual duration of fixations presents an optimal trade-off between complete visual object and scene analysis and fastest possible refresh rate of gaze position for sampling different spatial locations (Dorr et al., 2010). It might appear tenuous to relate our observations of regional cortical temporal sensitivities during passive viewing to temporal parameters that were determined in behavioral settings with demanding perceptual tasks. Our second experiment therefore addressed the relationship of attentional response modulation and temporal sensitivity even though the previously discussed earlier fMRI study (McKeeff et al., 2007) found no significant impact of attention on tuning, apart from a greater number of subjects and hence power in our experiments; it might also be that we succeeded in finding an effect of attention because we sampled in a more fine-grained way the relevant range of frame lengths. Another possible explanation is that attention was oriented toward target items within semantic categories in the McKeeff study (faces or houses) and that there was a huge impact of rate on behavior, whereas we only created a situation of heightened spatial attention due to an incidental task—a brief tilt of the central fixation bar—in which performance was almost independent from background stimulation rate. We hoped this manipulation would result in an "active viewing" throughout all tiers of ventral visual stream with greater permissiveness when processing locations containing central parts of the face and house stimuli.

Within this setting, we found that attention and temporal tuning interact, which in turn suggests that temporal tuning properties in the ventral visual stream are not completely hardwired intrinsic properties but affected by functional context (Besle et al., 2011). Specifically, we found that "active viewing" due to spatial allocation of attention shifts tuning peaks to slower rates, as evidenced by a main effect of attention on peak values of the tuning functions across all tiers of the ventral visual stream. When analyzing entire tuning functions separately for each visual tier, the effect was significant only in high-level visual cortex. We suggest that this is because the effects of attention are in general more readily detected the higher one moves along the visual hierarchy but also because the tuning shift in earlier areas seems to operate on a shorter timescale.

One possible interpretation of our observations is that, for attentional modulation of a stimulus-driven response to unfold, more processing time is required. This notion is compatible with timing differences observed previously for early "bottom–up" and delayed "top–down" responses in visual areas (Martínez et al., 1999; Noesselt et al., 2002). Transposed into our setting of periodic stimulation, this means that the time constraints for the deployment of attention would shift the optimal width of the integration window into a slower range than for mere feedforward volleys. Assuming a cascade of feedback loops to underpin attentional modulation rather than a single source (Bullier, 2001), this effect should scale with the intrinsic regional differences in temporal sensitivity and hence become more pronounced at higher levels of the hierarchy. Whatever the precise mechanism, however, our findings complement the existing views according to which attention is cortically implemented by baseline shifts or gain control and illustrate that, in continuous stimulation, the benefit from attention may hit temporal boundaries and require a minimum dwell time to be effective (Theeuwes et al., 2004).

References

Besle J, Schevon CA, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE (2011) Tuning of the human neocortex to the temporal dynamics of attended events. J Neurosci 31:3176–3185.

Bullier J (2001) Integrated model of visual processing. Brain Res Brain Res Rev 36:96–107.

Chawla D, Rees G, Friston KJ (1999) The physiological basis of attentional modulation in extrastriate visual areas. Nat Neurosci 2:671–676.

Coltheart M (1983) Iconic memory. Philos Trans R Soc Lond B Biol Sci 302:283–294.

Corbetta M, Miezin FM, Dobmeyer S, Shulman GL, Petersen SE (1990) Attentional modulation of neural processing of shape, color, and velocity in humans. Science 248:1556–1559.

Dorr M, Gegenfurtner KR, Barth E (2010) Variability of eye movements when viewing dynamic natural scenes. J Vis 10:1–17.

Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N (2006) Domain specificity in visual cortex. Cereb Cortex 16:1453–1461.

D'Souza DV, Auer T, Strasburger H, Frahm J, Lee BB (2011) Temporal frequency and chromatic processing in humans: an fMRI study of the cortical visual areas. J Vis 11:1–17.

Eger E, Henson RN, Driver J, Dolan RJ (2004) BOLD repetition decreases in object-responsive ventral visual areas depend on spatial attention. J Neurophysiol 92:1241–1247.

Enns JT, Di Lollo V (2000) What's new in visual masking? Trends Cogn Sci 4:345–352.

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.

Fox PT, Raichle ME (1985) Stimulus rate determines regional brain blood flow in striate cortex. Ann Neurol 17:303–305.

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A (2000) Representation of the temporal envelope of sounds in the human brain. J Neurophysiol 84:1588–1598.

Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol (Amst) 107:293–321.

Hagenbeek RE, Rombouts SA, van Dijk BW, Barkhof F (2002) Determination of individual stimulus–response curves in the visual cortex. Hum Brain Mapp 17:244–250.

Hart WM Jr (1987) The temporal responsiveness of vision. In: Adler's physiology of the eye, clinical application, Ed 8 (Moses RA, Hart WM, eds), pp 429–457. St. Louis, MO: Mosby.

Hillyard SA, Vogel EK, Luck SJ (1998) Sensory gain control (amplification) as a mechanism of selective attention: electro-physiological and neuroimaging evidence. Philos Trans R Soc Lond B Biol Sci 353:1257–1270.

Holcombe AO (2009) Seeing slow and seeing fast: two limits on perception. Trends Cogn Sci 13:216–221.

Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. Annu Rev Neurosci 23:315–341.

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751–761.

Keysers C, Xiao DK, Foldiak P, Perrett DI (2005) Out of sight but not out of mind: the neurophysiology of iconic memory in the superior temporal sulcus. Cogn Neuropsychol 22:316–332.

Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci U S A 104:20600–20605.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R (1992) Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci U S A 89:5675–5679.

Liu J, Wandell BA (2005) Specializations for chromatic and temporal signals in human visual cortex. J Neurosci 25:3459–3468.

Martínez A, Anllo-Vento L, Sereno MI, Frank LR, Buxton RB, Dubowitz DJ, Wong EC, Hinrichs H, Heinze HJ, Hillyard SA (1999) Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat Neurosci 2:364–369.

McKeeff TJ, Remus DA, Tong F (2007) Temporal limitations in object processing across the human ventral visual pathway. J Neurophysiol 98:382–393.

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.

Mullen KT, Thompson B, Hess RF (2010) Responses of the human visual cortex and LGN to achromatic and chromatic temporal modulations: an fMRI study. J Vis 10:1–19.

Murray SO, Wojciulik E (2004) Attention increases neural selectivity in the human lateral occipital complex. Nat Neurosci 7:70–74.

Noesselt T, Hillyard SA, Woldorff MG, Schoenfeld A, Hagner T, Jäncke L, Tempelmann C, Hinrichs H, Heinze HJ (2002) Delayed striate cortical activation during spatial attention. Neuron 35:575–587.

O'Connor DH, Fukui MM, Pinsk MA, Kastner S (2002) Attention modulates responses in the human lateral geniculate nucleus. Nat Neurosci 5:1203–1209.

Ozus B, Liu HL, Chen L, Iyer MB, Fox PT, Gao JH (2001) Rate dependence of human visual cortical response due to brief stimulation: an event-related fMRI study. Magn Reson Imaging 19:21–25.

Potter MC, Chun MM, Banks BS, Muckenhoupt M (1998) Two attentional deficits in serial target search: the visual attentional blink and an amodal task-switch deficit. J Exp Psychol Learn Mem Cogn 24:979–992.

Rolls ET, Tovee MJ (1994) Processing speed in the cerebral cortex and the neurophysiology of visual masking. Proc Biol Sci 257:9–15.

Singh KD, Smith AT, Greenlee MW (2000) Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. Neuroimage 12:550–564.

Singh M, Kim S, Kim TS (2003) Correlation between BOLD-fMRI and EEG signal changes in response to visual stimulus frequency in humans. Magn Reson Med 49:108–114.

Theeuwes J, Godijn R, Pratt J (2004) A new estimation of the duration of attentional dwell time. Psychon Bull Rev 11:60–64.

Thorpe S, Fize D, Marlot C (1996) The speed of visual processing. Nature 381:520–522.