Encoding model of temporal processing in human visual cortex

Anthony Stigliani[a], Brianna Jeska[a], and Kalanit Grill-Spector[a,b,1]

[a]Department of Psychology, Stanford University, Stanford, CA 94305; and [b]Stanford Neurosciences Institute, Stanford University, Stanford, CA 94305

Edited by Thomas A. Carlson, University of Sydney, and accepted by Editorial Board Member Marlene Behrmann November 1, 2017 (received for review March 24, 2017)

How is temporal information processed in human visual cortex? Visual input is relayed to V1 through segregated transient and sustained channels in the retina and lateral geniculate nucleus (LGN). However, there is intense debate as to how sustained and transient temporal channels contribute to visual processing beyond V1. The prevailing view associates transient processing predominately with motion-sensitive regions and sustained processing with ventral stream regions, while the opposing view suggests that both temporal channels contribute to neural processing beyond V1. Using fMRI, we measured cortical responses to time-varying stimuli and then implemented a two temporal channel-encoding model to evaluate the contributions of each channel. Different from the general linear model of fMRI that predicts responses directly from the stimulus, the encoding approach first models neural responses to the stimulus from which fMRI responses are derived. This encoding approach not only predicts cortical responses to time-varying stimuli from milliseconds to seconds but also, reveals differential contributions of temporal channels across visual cortex. Consistent with the prevailing view, motion-sensitive regions and adjacent lateral occipitotemporal regions are dominated by transient responses. However, ventral occipitotemporal regions are driven by both sustained and transient channels, with transient responses exceeding the sustained. These findings propose a rethinking of temporal processing in the ventral stream and suggest that transient processing may contribute to rapid extraction of the content of the visual input. Importantly, our encoding approach has vast implications, because it can be applied with fMRI to decipher neural computations in millisecond resolution in any part of the brain.

fMRI | V1 | V4 | hMT | transient

How does the visual system process the temporal aspects of the visual input? In the retina (1) and lateral geniculate nucleus (LGN) (2–4), temporal processing is thought to be mediated predominately by a magnocellular (M) pathway distinguished by its large transient responses (3, 4) and a parvocellular (P) pathway, which has larger sustained responses than the M pathway (3, 4) [in addition to a smaller koniocellular (5) pathway]. While M and P pathways remain segregated up to striate cortex (V1), there is intense debate as to how these pathways contribute to visual processing in extrastriate cortex. The prevailing view suggests that the dorsal stream, particularly motion-sensitive middle temporal (MT), is M-dominated (6–8), and that the ventral stream, particularly V4, is P-dominated (9). However, based on evidence for M and P contributions to both V4 (5, 9) and MT (10, 11), an opposing view suggests that these pathways are not segregated in extrastriate cortex (5, 8).

Since M and P pathways are associated with transient and sustained responses, respectively, these theories make predictions regarding temporal processing in human visual cortex. The prevailing view predicts that human MT complex (hMT+) will have large transient but small sustained responses and conversely, that human V4 (hV4) will have large sustained but small transient responses. However, the opposing view predicts substantial transient and sustained responses in both hMT+ and hV4. These predictions are derived from studies of the macaque brain; whether the same predictions can be made for the human brain is uncertain, because the organization of human visual cortex differs from the macaque in three notable ways. (*i*) V4 and MT neighbor in the macaque brain, but hV4 and hMT+ are separated by ∼3 cm in the human brain; (*ii*) whether macaque V4 and hV4 are homologous is subject to debate (12–15); and (*iii*) the human brain contains several additional visual regions neighboring hV4 and hMT+ that are not found in the macaque [VO-1/VO-2 (13) and LO-1/LO-2 (16), respectively]. Thus, generating a complete model of temporal processing in human visual cortex necessitates measurements in humans.

Understanding temporal processing in human visual cortex has seen little progress for two main reasons. First, the temporal resolution of noninvasive fMRI measurements is in the order of seconds (17), an order of magnitude longer than the timescale of neural processing, which is in the tens to hundreds of milliseconds range. Second, while fMRI responses are largely linear for long stimulus presentations (18, 19), they exhibit marked nonlinearities for short and transient stimuli (18–23). Since the general linear model (GLM) for fMRI (18, 24) is thought to be inadequate for modeling responses to such stimuli and fMRI is slow, the temporal processing characteristics of human visual cortex has remained elusive.

If the observed nonlinearities are of neural [rather than blood oxygen level dependent (BOLD)] origin, a new encoding approach applied to fMRI (25–27), which uses computational models to predict neural responses (even if they are nonlinear),

Significance

How is temporal information processed in human visual cortex? To address this question, we used fMRI and a two temporal channel-encoding model. This approach not only explains cortical responses for time-varying stimuli ranging from milliseconds to seconds but finds differential temporal processing across human visual cortex. While motion-sensitive regions are dominated by transient responses, ventral regions that process the content of the visual input surprisingly show both sustained and transient responses, with the latter exceeding the former. This transient processing may foster rapid extraction of the gist of the scene. Importantly, our encoding approach marks a transformative advancement in the temporal resolution of fMRI, as it enables linking fMRI responses to the timescale of neural computations in cortex.

Author contributions: A.S. and K.G.-S. designed research; A.S. and B.J. performed research; A.S. and K.G.-S. analyzed data; and A.S. and K.G.-S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. T.A.C. is a guest editor invited by the Editorial Board.

Data deposition: The data and code reported in this paper are available via the Open Science Framework (OSF) at https://osf.io/mw5pk.

[1]To whom correspondence should be addressed. Email: kalanit@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1704877114/-/DCSupplemental.

could surmount these issues. Different from the GLM method, which predicts fMRI signals directly from the stimulus, the encoding approach first models neural responses to the stimulus and then, from the predicted neural responses, calculates fMRI responses. The encoding approach (25–29) has been influential for three reasons: (*i*) it provided an important insight that accurately modeling neural responses at a subvoxel resolution better predicts fMRI responses at the voxel resolution, (*ii*) it has advanced understanding of neural mechanisms by building explicit quantitative models of neural computations, and (*iii*) it can be applied to predict fMRI responses to dynamic stimuli, such as movies, by using millisecond variations in the stimulus to precisely model neural responses (30, 31).

Here, we sought to leverage the encoding approach to characterize temporal processing in human visual cortex. Thus, we built a temporal encoding model of neural responses to time-varying visual stimuli in millisecond resolution and used this model to predict fMRI responses in second resolution. The model is based on estimation of the transient and sustained neural channels' impulse response functions (IRFs) from measurements in macaque V1 (32–34) and psychophysics in humans (35–38). To determine temporal processing in human visual cortex, we implemented three experiments aimed to measure fMRI responses to time-varying visual stimuli that were sustained (one continuous image per trial, durations ranging from 2 to 30 s) (Fig. 1, experiment 1), transient [30 flashed, 33-ms long images per trial, interstimulus intervals (ISIs) ranging from 33 to 967 ms] (Fig. 1, experiment 2), or contained both transient and sustained components (30 continuous images per trial, durations ranging from 67 to 1,000 ms per image) (Fig. 1, experiment 3). We first determined if millisecond temporal variations in visual stimuli generate substantial modulations of fMRI responses in visual cortex. Then, we used experiments 1 and 2 data to estimate the parameters of the two temporal channel-encoding model. Finally, we evaluated how well the model predicts fMRI responses to stimuli that vary in their temporal properties from milliseconds to seconds in data from experiment 3. Additionally, to verify the utility of the two-temporal channel model, we compare its performance with that of several other models of fMRI responses (17, 18, 39, 40) featuring different channel architectures and nonlinearities. After we established the model's validity, we derived the contributions of sustained and transient channels to neural responses across striate and extrastriate visual cortex to test the competing theoretical hypotheses.

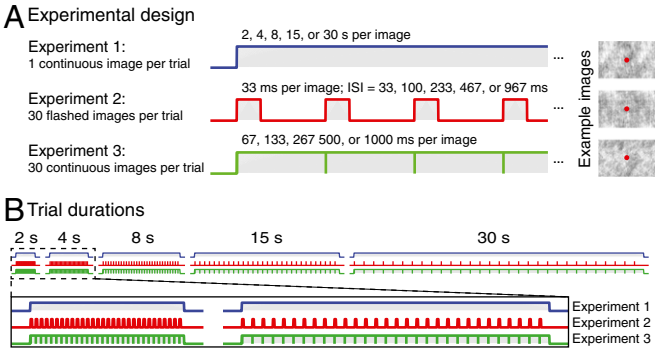

Fig. 1. Measuring brain responses to combinations of sustained and transient visual stimuli. (*A*) Participants fixated centrally and viewed phase-scrambled gray-level images that were presented in trials of different durations interleaved with 12-s periods of a blank screen. The same fixation task (detecting change in fixation color) was used in all three experiments. Experiment 1: a single phase-scrambled image was shown for the duration of a trial. Experiment 2: 30 briefly presented images (33 ms each), each followed by a blank screen, were presented in each trial. As the trial duration lengthens, the gap between images increases, causing the fraction of the trial containing visual stimulation to decrease. Experiment 3: 30 continuous images (with no gaps between consecutive stimuli) were presented in each trial. As the block duration lengthens, the duration of each image progressively increases. (*B*) The same trial durations (2, 4, 8, 15, or 30 s) were utilized across all three experiments, while the rate and duration of visual presentation varied between experiments. Corresponding trials in experiments 1 and 3 have the same overall duration of stimulation but different numbers of stimuli, whereas trials in experiments 2 and 3 have the same number of stimuli but different durations of stimulation. (*Upper*) Stimulation durations for example trials in each experiment; (*Lower*) zoomed-in view of the 2- and 4-s trials.

Results

Do Millisecond Temporal Variations in the Visual Stimulus Modulate V1 Responses? To test the feasibility of this approach, we first examined V1 responses during the three experiments. Predicted fMRI responses from a GLM depend only on the type and duration of stimuli. Thus, the GLM approach predicts longer responses for longer trials and identical responses in experiments 1 and 3 (Fig. 2*A*, blue and green, respectively), which use the same visual stimuli and trial durations and just vary by the number of images per trial (1 vs. 30, respectively). Furthermore, the model predicts that the amplitude of responses in experiments 1 and 3 will increase from 2- to 8-s trials and will remain largely the same for longer trials. While this model predicts the same response durations in experiment 2, it predicts substantially lower response amplitudes in experiment 2 than experiments 1 and 3, because the transient visual stimuli are presented for only a fraction of each trial. Furthermore, the model predicts a progressive decrease in response amplitude during experiment 2 from 2- to 30-s trials as the fraction of the trial in which visual stimuli are presented decreases (from 1/2 to 1/30 of the trial) (Fig. 2*A*, red).

While V1 responses to sustained visual stimulation in experiment 1 largely followed the predictions of a GLM (Fig. 2*A*, blue), responses in experiments 2 and 3 deviated from GLM predictions. First, responses in experiment 3 (Fig. 2*B*, green) were higher than responses in experiment 1 for all trial durations. Second, responses to transient stimuli in experiment 2 (Fig. 2*B*, red) were substantially higher than predicted by a GLM. In fact, V1 responses during 2- to 8-s trials of experiment 2 were equal or higher than those of experiment 1, although the cumulative duration of stimulation across images in experiment 2 was a fraction of the duration of stimulation in experiment 1. Third, different from the predictions of the GLM, response amplitudes in experiment 2 did not systematically decline with trial duration but instead, peaked for the 4-s trials.

These data show that (*i*) varying the temporal characteristics of visual presentation in the millisecond range has profound effects on V1 fMRI responses and (*ii*) the GLM approach is inadequate in predicting measured fMRI responses for these stimuli, in agreement with prior data. Furthermore, the higher responses in experiment 3 (which has both sustained and transient visual stimulation) compared with experiments 1 and 2 (which have either sustained or transient stimuli, respectively) suggest that both transient and sustained components of the visual input contribute to the fMRI signals, consistent with our hypothesis.

An Encoding Model for Temporal Processing in Visual Cortex. To accurately predict fMRI responses in all three experiments, we built a temporal encoding model of neural responses in millisecond resolution and used this model to predict fMRI responses in second resolution (Fig. 3). Our model consists of two neural temporal channels, each of which can be characterized by a linear systems approach using a temporal IRF (32–34, 36–38). The sustained channel is characterized by a monophasic IRF (Fig. 3*B*, blue channel IRF), peaking at around 40 ms and lasting

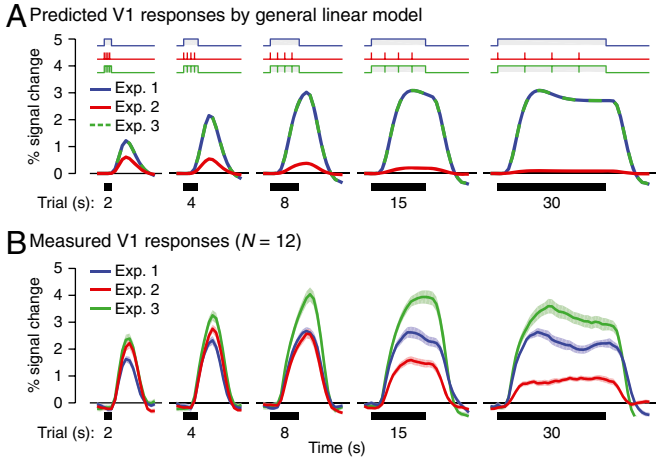

Fig. 2. V1 responses to transient stimuli differ from the predictions of a GLM. (*A*) The GLM approach predicts the same response in trials of the same duration across experiment 1 (blue) and experiment 3 (green), since both present stimuli continuously for the same total duration in each trial. However, responses in experiment 2 (red) are predicted to be much lower, because stimuli are spaced apart and are only presented for a fraction of each trial duration. Illustrations of the stimulus are schematic representations. (*B*) The mean V1 response in experiments 1–3 averaged across 12 participants. Each curve is data from a different experiment. Shaded regions around the curves indicate ±1 SEM across 12 participants. In both *A* and *B*, the onsets and lengths of the trials are illustrated as thick black bars below each graph, and curves extend 2 s before the onset and 12 s after the offset of each trial.

100–150 ms; convolving this channel with a visual stimulus will produce a sustained neural response for the duration of the stimulus. The transient channel is characterized by a biphasic IRF, akin to a derivative function, with the positive part peaking at around 35 ms and the negative part peaking at around 70 ms (Fig. 3*B*, red channel IRF). A quadratic nonlinearity (squaring) is added, as both stimulus onset and offset lead to increased neural firing and consequently, increased metabolic demands (38, 41). This nonlinearity produces a positive neural response when there is an onset or offset of the visual stimulus but zero response in between when the stimulus is presented for durations longer than the duration of the IRF. The predicted fMRI response is generated by convolving the output of each neural channel with the hemodynamic response function (HRF) and summing the responses of the two temporal channels (Fig. 3*C*).

Our procedure for evaluating the two temporal channel-encoding model had three stages. First, we estimated the contributions of the two temporal channels to fMRI signals using concatenated data from experiments 1 and 2 that were designed to largely drive sustained or transient channels, respectively. Second, we cross-validated the model by testing how well it predicted data from experiment 3 that had both transient and sustained visual stimulation. Third, we compared the performance with two other families of models: a family of three hemodynamic models and a family of three nonlinear single-channel neural models implementing compressive temporal summation (CTS). Hemodynamic models are (*i*) the standard GLM approach (18, 24), (*ii*) the hemodynamic temporal derivative (HTD) model that has been used to model temporal variability of hemodynamic responses across cortex (39) and contains two linear channels—the canonical HRF and its derivative—and (*iii*) the balloon model (17), which models changes in blood flow, volume, and oxygenation to account for nonlinearities in fMRI signals. CTS single-channel neural models that were derived to explain subadditive temporal summation and adaptation effects in fMRI responses are (*i*) compressive temporal summation with a static power law (CTS-p), (*ii*) compressive temporal summation with divisive normalization (CTS-n) (40), and (*iii*) dynamic compressive temporal summation (dCTS), which has a prominent onset response and a subsequent declining continued response for the duration of the stimulus (40).

Does a Two-Temporal Channel Model Explain V1 fMRI Responses to Time-Varying Stimuli? Comparing the predictions of the two-temporal channel model with V1 responses reveals three findings. First, the two-temporal channel model containing one sustained predictor (weighted by β_S) and one transient predictor (weighted by β_T) generated fMRI signals that tracked both the duration and amplitude of V1 responses in experiments 1 and 2 [Fig. 4*A*, compare model prediction (black) with measured V1 data (gray)]. Consistent with our predictions, the sustained channel accounted for the majority of responses in experiment 1 (blue in Fig. 4*A*, *Upper*), while the transient channel contributed the bulk of the response in experiment 2 (red in Fig. 4*A*, *Lower*). Second, the two-temporal channel model with β_S and β_T fit from experiments 1 and 2 (Fig. 4*C*) accurately predicted independent data from experiment 3 (Fig. 4*B*). The sustained contribution (Fig. 4 *A* and *B*, blue) in experiment 3 was comparable with experiment 1 (these experiments have the same total duration of stimulation per trial), and the transient contribution (Fig. 4 *A* and *B*, red) in experiment 3 was similar to experiment 2 (these experiments have the same number of transients per trial). Since both temporal channels provided a significant contribution to V1 and the contributions of the two channels are additive, experiment 3 responses were higher than both experiments 1 and 2 across all trial durations. Analysis of cross-validated R^2 showed that the two-temporal channel model explained an average of 72 ± 2% of variance in experiment 3 (Fig. 4*F*), although the channel weights were estimated from responses to independent data with different temporal characteristics. Third, the two-temporal channel model outperformed the other models that we tested. Notably, while all models predicted V1 responses to the long and sustained stimulus presentations in the first experiment, hemodynamic models failed to capture responses to

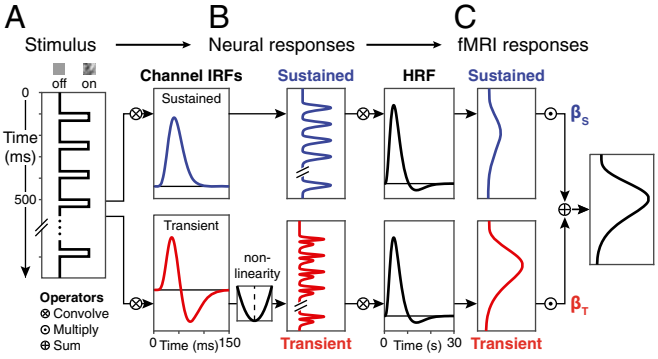

Fig. 3. The two-temporal channel model. (*A*) Transitions between stimulus and baseline screens are coded as a step function representing when a stimulus was on vs. off with millisecond temporal resolution. In the example illustrated here, each stimulus is presented for 33 ms and followed by a 100-ms blank screen. (*B*) Separate neural responses for the sustained (blue) and transient (red) channels are modeled by convolving the stimulus vector with IRFs for the sustained and transient channels, respectively, estimated from human psychophysics. A squaring nonlinearity is applied in the transient channel to rectify offset deflections (*Materials and Methods*). (*C*) Sustained and transient fMRI response predictors are generated by convolving each channel's neural responses with the HRF and down-sampling to match the temporal acquisition rate of fMRI data. The total fMRI response is the sum of the weighted sustained and transient fMRI predictors for each channel. To estimate the contributions (β weights) of the sustained (β_S) and transient (β_T) channels in V1, we fit the two-temporal channel model to data concatenated across experiments 1 and 2.

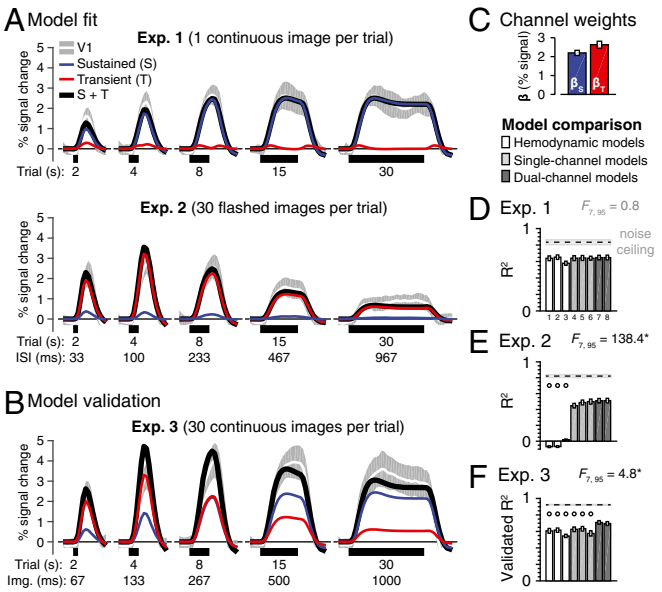

Fig. 4. Sustained and transient contributions to V1 fMRI responses. (*A*) Measured V1 responses in experiments 1 and 2 are plotted as the mean (white) ±1 SD (gray) across 12 participants for each trial duration. Superimposed are the predictions of the two-temporal channel model (quadratic nonlinearity) (as in Fig. 3) fit across data from both experiments. Blue, sustained predictor weighted by β_S; red, transient predictor weighted by β_T; black, prediction of the two-temporal channel model, which is the addition of the two channels. (*B*) Measured V1 responses and cross-validated model prediction for experiment 3. The sustained and transient predictors are weighted with β_S and β_T, respectively, fitted from experiments 1 and 2. Trial durations in *A* and *B* are illustrated below the *x* axis, and curves extend 2 s before the onset and 12 s after the offset each trial. (*C*) The model solution (β_S and β_T) for V1 fit with the two-temporal channel model using data concatenated across experiments 1 and 2. (*D–F*) Comparison of the variance explained (R^2) across three families of models for each experiment with results of corresponding ANOVA indicating the main effect of model. White, hemodynamic models (1, GLM; 2, HTD model; 3, balloon model); light gray, single-channel models (4, CTS-p; 5, CTS-n; 6, dCTS); dark gray, dual channel models (7, two-temporal channel model with rectification; 8, two temporal channel with quadratic power law). Dashed lines indicate the noise ceiling. Open circles indicate models with significantly worse performance vs. dual channel models ($P < 0.05$). Error bars in *C–F* indicate ±1 SEM across participants. *F* value of an ANOVA comparing model family performance (hemodynamic, single channel with a compressive nonlinearity, two temporal channel with a nonlinearity) for each experiment is indicated. *Significant difference across models ($P < 0.001$).

transient stimuli in the second experiment (Fig. 2*A* and Fig. S1*A*). That is, the two-temporal channel model fit to both experiments explained an average of 65 ± 3% (mean ± 1 SEM across participants) of V1 response variance in experiment 1 (Fig. 4*D*) and 52 ± 3% of the variance in experiment 2 (Fig. 4*E*). In contrast, the hemodynamic models fit to both experiments explained an average of 62 ± 3% of the variance in experiment 1 (Fig. 4*D*) but less than 1 ± 1% of the variance in experiment 2 (Fig. 4*E*). Furthermore, although CTS models explained an average of 64 ± 3% of the variance in experiment 1 (Fig. 4*D*) and 48 ± 3% of the variance in experiment 2 (Fig. 4*E*), all CTS models underestimated responses in experiment 3 (Fig. S1*B*). Indeed, in experiment 3, the average cross-validated R^2 of the two-temporal channel models was significantly higher than each of the hemodynamic or CTS models ($ts > 2.06$, $ps < 0.01$, paired *t* tests) (Fig. 4*F*). In general, the dual channel models outperformed the other families of models in all visual areas tested (Figs. S2 and S3), with no significant differences in performance between the quadratic and rectification nonlinearities on the transient channel (Fig. 4 *D–F* and Fig. S1). Thus, a two-temporal channel model with a linear sustained channel and a linear–nonlinear transient channel predicts fMRI responses to visual stimuli across a threefold range of presentation durations ranging from tens of milliseconds to tens of seconds with greater accuracy than several alternative models.

Do Temporal Processing Characteristics Differ Across Intermediate Visual Areas? We next examined hV4 and hMT+ responses to the time-varying visual stimuli in experiments 1–3, as the competing theories make different predictions regarding the contributions of sustained and transient channels to these regions. hV4 and hMT+ illustrated distinct patterns of responses. Like V1, hV4 showed higher responses in experiment 3 (30 continuous images per trial) than either experiment 1 (1 continuous image per trial) or experiment 2 (30 flashed images per trial). Different from V1, hV4 exhibited equal or stronger responses to the brief transient visual stimuli in experiment 2 than the sustained single images in experiment 1 (Fig. 5*A*). Different from both V1 and hV4, hMT+ exhibited close to zero evoked responses for the sustained stimuli in experiment 1 (except for

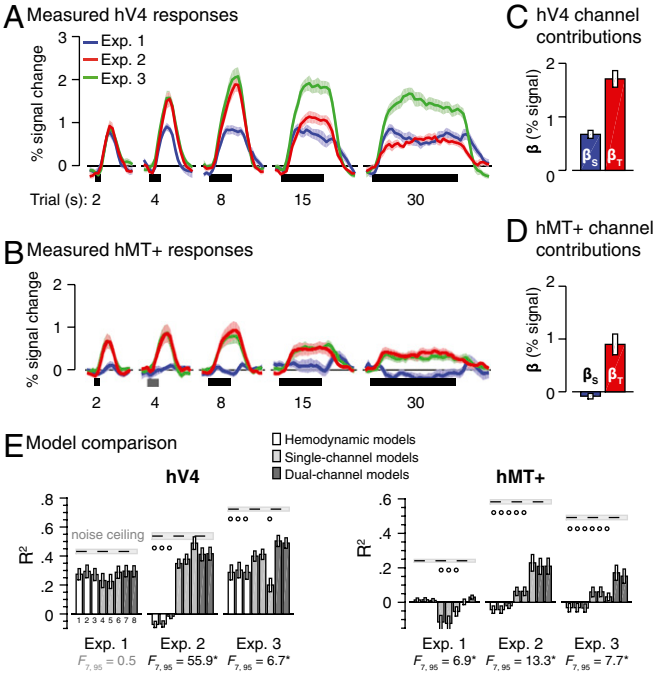

Fig. 5. Differential sustained and transient contributions across hV4 and hMT+. (*A*) hV4 and (*B*) hMT+ responses for experiment 1 (blue), experiment 2 (red), and experiment 3 (green). Curves show the mean (solid line) ±1 SEM across 12 participants (shaded area). Trial durations are indicated by the thick black bars below the *x* axis, and curves extend 2 s before the onset and 12 s after the offset each trial. The two-temporal channel model solution (β_S and β_T; quadratic nonlinearity) for (*C*) hV4 and (*D*) hMT+ is plotted with error bars representing ±1 SEM across participants. (*E*) Comparison of the variance explained (R^2) by each of the models tested. All models were fit using data from experiments 1 and 2. ANOVAs for each experiment for hV4 (*Left*) and hMT+ (*Right*) indicate significant differences in performance across models. White, hemodynamic models (1, GLM; 2, HTD model; 3, balloon model); light gray, single-channel models (4, CTS-p; 5 CTS-n; 6 dCTS); dark gray, dual channel models (7, two-temporal channel model with rectification; 8, two temporal channel with quadratic power law). Dashed lines indicate the noise ceiling. Open circles indicate models with significantly worse performance vs. dual channel models ($P < 0.05$). Error bars in *C–F* indicate ±1 SEM across participants. *F* value of an ANOVA comparing model family performance for each experiment is indicated. *Significant difference across model family ($P < 0.001$).

onset and offset responses that are visible in trials of 8 s and longer) (Fig. 5*B*). However, hMT+ showed substantial responses for transient stimuli in experiment 2 that were comparable with experiment 3, which had both transient and sustained stimulation. Together, these data suggest differences in temporal processing across hV4 and hMT+.

Next, we quantified hV4 and hMT+ responses with the two temporal channel-encoding model. The model fits revealed (*i*) that, in hV4, both channels contributed to responses, with the contribution of the transient channel about double that of the sustained channel (Fig. 5*C*) and (*ii*) that, in hMT+, the transient channel substantially contributed to responses but that the sustained channel had close to zero contribution (Fig. 5*D*). Across both hV4 and hMT+, the two-temporal channel models fit data from experiment 2 better than the hemodynamic models [$ts > 4.88$, $ps < 0.001$, paired t tests on R^2 values for each region of interest (ROI)] and predicted data from experiment 3 equally or better than all hemodynamic and CTS models (Fig. 5*E*). These results show that not only does the two-temporal channel model perform significantly better than a variety of alternate models at intermediate stages of the visual hierarchy but that the contributions of transient and sustained channels differ across hV4 and hMT+.

What Is the Topology of Sustained and Transient Channels Across Visual Cortex? To complement the ROI approach, we next visualized the spatial topology of the sustained and transient channels across visual cortex, which enables mapping channel contributions at the voxel level.

Examining the contribution of sustained and transient channels across ventral and lateral occipitotemporal cortex revealed two main findings. First, lateral occipitotemporal cortex was devoid of contributions from the sustained channel but had substantial contributions from the transient channel (Fig. 6 *A* and *B*). This effect was widespread and not only included voxels in hMT+, as predicted by the prior analysis, but also, extended (*i*) posteriorly into portions of lateral occipital areas LO-1 and LO-2 and (*ii*) ventrally into the inferior occipital gyrus and lateral fusiform gyrus. Dorsal regions along the intraparietal sulcus also showed negligible sustained responses (Fig. 6*B*). Second, in ventral occipitotemporal cortex, regions along the posterior collateral sulcus and medial fusiform gyrus (where hV4, VO-1, and VO-2 are located) showed both transient and sustained responses, with larger contributions from the transient than sustained channel (Fig. 6*A*). Third, a similar pattern of results was observed when participants viewed complex stimuli, such as faces, bodies, and pseudowords (Fig. S4). This indicates that the differential contributions of the sustained and transient channels to early and intermediate regions generalize across stimuli.

We quantified the mean contributions of transient and sustained channels across visual areas spanning occipitotemporal cortex. Our results showed differences in the contributions of sustained and transient channels across early visual cortex (V1–V3), ventral occipitotemporal cortex (hV4, VO-1, and VO-2), lateral occipitotemporal cortex (LO-1, LO-2, and hMT+), and dorsal occipitotemporal cortex [V3A and V3B; significant temporal channel by cluster interaction, $F_{3,\ 33} = 10.29$, $P < 0.001$, two-way ANOVA on β weights with factors of temporal channel (sustained/transient) and visual cluster (early/ventral/lateral/dorsal)] (Fig. 6*C*). From V1 to higher-order areas, there was a larger drop in the contribution of the sustained channel than the transient channel (Fig. 6*C*). Nevertheless, there were significant differences among clusters: in ventral areas, both sustained and transient channels contributed to responses, but in lateral areas, responses were largely dominated by the transient channel (Fig. 6*C*). As the overall amplitude of responses varied across regions, we quantified the relative contribution (ratio) of sustained to transient channel across regions (Fig. 6*D*). Dorsal and lateral regions have a larger transient preference than V1, as predicted by prior research. However, it is surprising that ventral regions hV4, VO-1, and VO-2 also have a greater transient preference than V1 (Fig. 6*D* and Fig. S4).

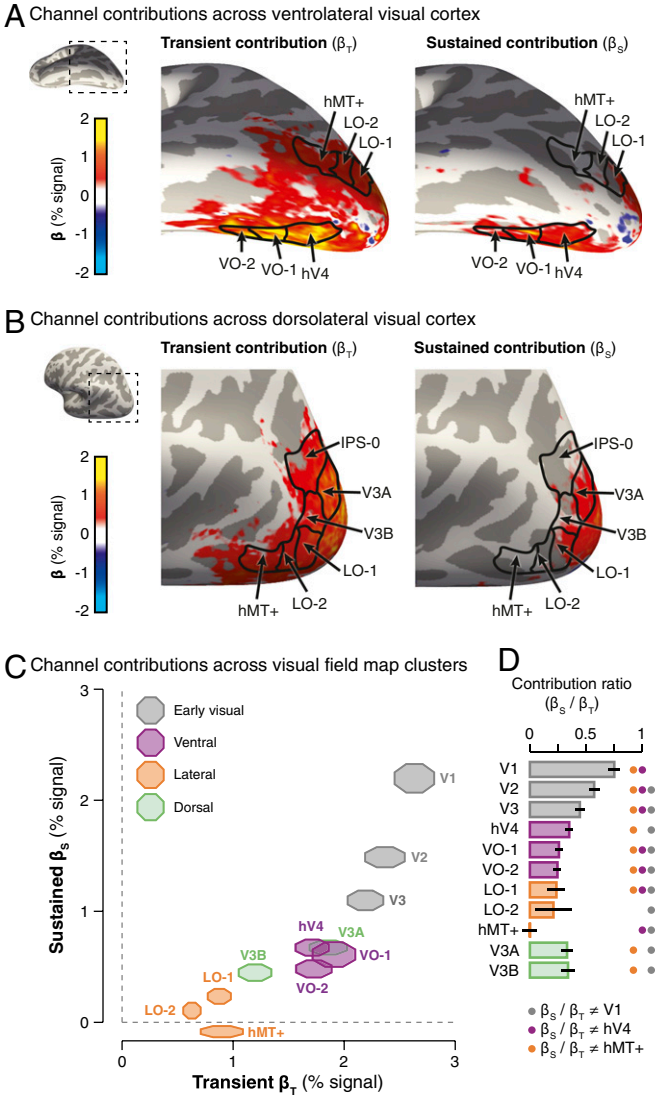

Fig. 6. Differential transient and sustained contributions across visual regions. (*A*) Ventrolateral view of occipitotemporal cortex (*Inset*) depicting group-averaged ($n = 12$) maps of the contributions of transient (*Left*) and sustained (*Right*) channels. We estimated β weights for each channel in each voxel in each participant's native brain space; β-weight maps were then transformed to the FreeSurfer average brain using cortex-based alignment and averaged across participants in this common cortical space. The resulting group maps were thresholded to exclude voxels with weak contributions ($-0.1 > \beta > 0.1$). Boundaries of ventral and lateral regions (black) are drawn from the Wang atlas (69), with hMT+ as the union of TO-1 and TO-2. (*B*) Same as *A* for the dorsolateral view (*Inset*). (*C*) Contributions (β weights) of transient (x axis) and sustained (y axis) channels to each visual area as estimated by the two-temporal channel model. Marker size spans ±1 SEM across 12 participants in each axis, and β weights were solved by fitting the two temporal channel using data concatenated across experiments 1 and 2. (*D*) The ratio of sustained to transient channel contributions was computed for each subject and visual area. Data show the average across subjects per area. A ratio of one indicates no preference, a ratio of less than one indicates a transient preference, and a ratio equal to zero indicates no sustained responses. Colored dots indicate a significantly ($P < 0.05$) different ratio vs. V1 (gray), hV4 (purple), and hMT+ (orange).

The spatial topology of sustained and transient channels also revealed differences in temporal processing within regions. Specifically, in early visual cortex (V1–V3), the sustained channel was robust in eccentricities <20° but declined in more peripheral eccentricities (Fig. 7*A*, *Right*). In contrast, the transient channel contributed to responses across a larger range of eccentricities that extend farther into the periphery (>20°) (Fig. 7*A*, *Left*). We quantified these effects by measuring the contributions of the two channels across eccentricities using uniformly sized disk ROIs defined along the horizontal meridian representations in V1 and V2/V3 (Fig. 7*B*). This quantification showed that, in early visual areas, the magnitude of the sustained channel declined more rapidly with eccentricity than the transient channel to the extent that, at eccentricities of 40°, there still was a 0.90 ± 0.17% transient response but less than 0.26 ± 0.07% of a sustained response (Fig. 7*B*). Furthermore, the decline of the sustained channel with eccentricity occurred more rapidly in V2/V3 than V1. Together, we find a differential contribution of transient and sustained channels across eccentricities and areas [significant three-way interaction of temporal channel (sustained or transient), visual area (V1 or V2/V3), and eccentricity (5°, 10°, 20°, or 40°), $F_{3,\ 33} = 3.18$, $P < 0.05$, three-way ANOVA on β weights].

Discussion

Our results show that a two-temporal channel model of neural responses, containing a sustained linear channel and transient channel with a nonlinearity, is a parsimonious encoding model that predicts fMRI responses in human visual cortex to visual stimuli across a broad range of durations from tens of milliseconds to tens of seconds. Critically, our data address the ongoing debate regarding the contribution of sustained and transient channels in extrastriate cortex. Consistent with the prevailing view (2, 6, 7), we find that the transient channel dominates hMT+ responses and peripheral eccentricity representations of V1 (38). Importantly, we show that this temporal processing characteristic extends to human lateral occipitotemporal cortex as well as peripheral eccentricity representations of V2 and V3.

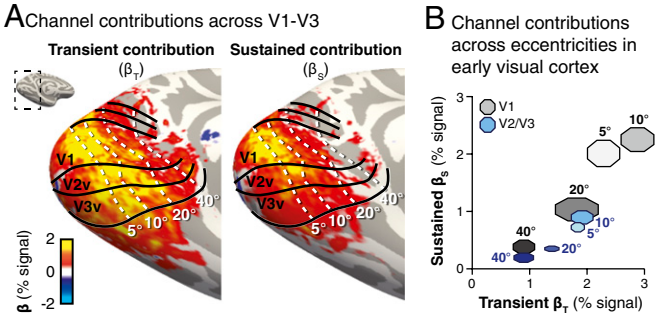

Fig. 7. Differential transient and sustained contributions across central and peripheral eccentricities. (*A*) Medial cortical surface zoomed on the occipital lobe (*Inset*) depicting group-averaged ($n = 12$) maps of the contributions of transient (*Left*) and sustained (*Right*) channels. We first estimated β weights of each channel in each voxel in each participant's native brain space; β-weight maps were transformed to the FreeSurfer average brain using cortex-based alignment and averaged across participants in this common cortical space. The resulting group maps were thresholded to exclude voxels with weak contributions ($-0.1 > \beta > 0.1$). Regional boundaries (black) and eccentricity bands (white) of early visual areas are derived from the Benson atlas (70). (*B*) Contributions of transient (*x* axis) and sustained (*y* axis) channels across eccentricities along the horizontal representation in V1 (grays) and V2/V3 (blues) as estimated by the two-temporal channel model. Eccentricities range from 5° (lightest markers) to 40° (darkest markers). Marker size spans ±1 SEM across 12 participants in each dimension, and β weights were solved by fitting the two temporal channel using data concatenated across experiments 1 and 2.

In contrast to the prevailing view, we find that both sustained and transient channels drive responses in not only hV4 but also, ventral occipitotemporal regions VO-1 and VO-2, with a surprisingly larger transient preference in ventral regions compared with V1. This finding argues against the view that the ventral stream primarily codes static visual information and suggests a rethinking of the role of transient processing in the visual system.

One prior experiment (38) had used a similar two-temporal channel approach to model luminance transients in visual cortex during an experiment in which subjects were placed in the scanner with semitransparent, hemisphere-shaped diffusers over their eyes and presented with alternating 24-s blocks of low and high luminance with no spatial contrast. Because the study by Horiguchi et al. (38) contained a single experimental condition, their data do not enable making inferences about the validity of the model to other conditions or stimuli. In contrast, our study used an innovative experimental design that revealed four important findings. (*i*) A key finding of our study is that the two-temporal channel model explains the observed nonlinearities in fMRI responses for transient and rapid visual stimuli, resolving an outstanding gap in knowledge in understanding BOLD signals. We made the data and the code implementing this model freely available: https://osf.io/mw5pk. (*ii*) We report differences in the contributions of the sustained and transient channels across multiple visual areas in the human brain. (*iii*) We tested a range of conditions within and across experiments and show that the two-temporal channel model not only predicts the responses to data (experiment 3) but that it generalizes across a variety of low- and high-level visual stimuli. (*iv*) We quantitatively evaluated the performance of this model compared with other models (17, 18, 39, 40) and show that it outperforms them, especially for conditions containing brief and rapid visual stimulation.

Differential Transient and Sustained Responses Across Visual Cortex. Our research fills a large gap in knowledge regarding temporal processing in human visual cortex by showing (*i*) that the two-temporal channel model is applicable to at least 10 additional visual areas beyond V1 (32–34, 38, 42) and (*ii*) that temporal processing is a key functional attribute that differentiates visual areas.

Our observation that hMT+ responses to sustained stimuli are close to zero is consistent with the prevailing view (*i*) that hMT+ is involved in processing visual dynamics rather than static information and (*ii*) that inputs to hMT+ are M-dominated (7). Notably, we found that neighboring regions, LO-1 and LO-2, also have close to zero sustained responses. This finding is interesting, because LO-2, which is thought to be involved in visual processing of objects (16, 43) and body parts (44), shows more robust responses to rapidly presented stimuli compared with nearby category-selective regions (45). These data suggest that this characteristic may be an outcome of a dominant transient channel—a hypothesis that can be tested in future research using more optimal stimuli to drive this region.

Inconsistent with the prevailing view, we found that hV4 showed not only sustained responses as expected (8, 23, 46, 47) but also, large transient responses. While this observation is consistent with reports that macaque V4 receives both P and M inputs (5, 9), it is unexpected that the transient preference in hV4 is larger than V1. Furthermore, unlike the macaque brain, where portions of V4 are adjacent to MT, hV4 is ∼3 cm away from hMT+, indicating that this finding cannot be explained by the proximity of these two regions. While the sustained channel is often associated with coding static visual input and the transient channel is often associated with coding visual dynamics (46, 48, 49), visual transients can also indicate changes to the content of the visual input. Indeed, in our experiments, transients occurred when stimuli changed (i.e., when a new image was shown or an image was replaced by a uniform gray screen). Since the

function of the ventral stream is to derive the content of the visual input, fast, transient visual processing in ventral stream regions may enable rapid processing of visual changes (50), which in turn, may foster detection of novel stimuli and rapid extraction of the gist of the visual scene.

It is interesting that temporal processing in dorsolateral occipitotemporal regions, like that of far peripheral eccentricities (>20°) in early visual cortex, was dominated by the transient channel, and temporal processing in ventral occipitotemporal regions, like central eccentricities in early visual cortex, showed a dual channel contribution. These functional characteristics may be anatomically supported by white matter connections from peripheral representations of early visual areas to MT and nearby regions (51) that are separate from white matter connections from central representations of early visual cortex to ventral occipitotemporal regions (52). Furthermore, our observations of diminished sustained responses in the periphery of early visual cortex are consistent with prior findings showing reduced P inputs (53) and diminished sustained luminance responses (38) in peripheral compared with central V1 as well as faster perception in the periphery (54).

Implications for Modeling fMRI Signals: Millisecond Timing Matters. Our data have important implications regarding modeling fMRI signals and understanding temporal processing in the human brain, because they show (*i*) that varying the temporal characteristics of the visual stimulus in the milliseconds range has observable effects on fMRI responses in the seconds range and (*ii*) that, by considering the contribution of a transient neural channel, an encoding model can account for nonlinearities in fMRI responses for rapid and short visual stimuli (18–20, 22, 55). It would be interesting to extend our design using rapid event-related paradigms and accelerated MRI acquisitions to measure and model responses to single brief images in future studies.

Our data extend the original GLM of fMRI signals (18, 24) by showing the importance of modeling the temporal properties of neural responses at millisecond resolution to accurately predict fMRI signals. In their original study, Boynton et al. (18) noted that, for short durations (3–6 s), fMRI responses deviated from the predictions of the linear model, which underestimated fMRI signals. These nonlinearities are exacerbated in experiments using even shorter stimuli [1/4–2 s (20, 22, 55)]. Likewise, in experiment 2 of our study, linear hemodynamic models fail to account for the large fMRI responses to brief, transient stimuli (Fig. 4 and Figs. S1–S3). Boynton et al. (18) suggested that nonlinearities in BOLD responses, neural adaptation, or transient neural responses may explain deviations from linearity in fMRI responses. We favor the interpretation that transient responses account for nonlinearities for four reasons. (*i*) Taking into account the neural transient channel resolves this nonlinearity and can predict not only our fMRI measurements but also, prior data showing nonlinearities (22) (Fig. S5). (*ii*) Adaptation would have resulted in declining responses during long trials of continuous presentation of a single stimulus (50). However, we observed negligible adaptation in striate (56) and extrastriate (e.g., hV4) areas, even during the 30-s single continuous image trials. (*iii*) While the family of single-channel CTS models examined here was developed to model subadditive accumulation of fMRI responses caused by adaptation (40), their generalization across a range of temporal conditions is more limited than the two-temporal channel model (Fig. 4 and Figs. S1–S3). (*iv*) The balloon model, explicitly derived to model neurovascular nonlinearities in the BOLD signal, did not account for experimental conditions containing transient stimuli.

Of the other models tested, dCTS (40) came closest in performance to the two-temporal channel model. Like the two-temporal channel model, dCTS has a prominent onset response. Different from the two-temporal channel model, it predicts a declining continuous response across the duration of the stimulus and predicts no offset response. Notably, the two-temporal channel model has several advantages over dCTS. (*i*) It is a more parsimonious model: across ROIs and experiments, its performance is equal or better than the dCTS. (*ii*) It is a more interpretable model, as it is clear which component of the response is caused by the transient and which is caused by the sustained part, which is difficult to untangle with the dCTS model. (*iii*) It is faster and easier to solve, as the solution is analytical rather than requiring complex nonlinear optimization.

While the two-temporal channel model provides a significant improvement in modeling fMRI signals, we acknowledge that this model does not explain the entire variance of fMRI signals (Fig. S3) and does not account for all nonlinearities. For example, in experiment 3, the model overestimates responses in extrastriate regions during short trials (durations <8 s) (Fig. S2). Furthermore, temporal processing dynamics may differ across brain areas (23, 45, 57–59), which may require developing area-specific temporal encoding models with area-specific neural IRFs. Additionally, while we did not find differences between two-temporal channel models with rectification vs. quadratic nonlinearities, our experiments were not designed to discriminate between specific nonlinearities. Thus, future experiments optimized for isolating "on" and "off" neural responses could distinguish these alternatives. Another direction for future research would be to implement models that consider both nonlinear neural responses and nonlinearities between neural activity and BOLD responses (60), as presently, we examined each type of nonlinearity separately. Finally, an important goal for future research would be to combine the two-temporal channel model with a spatial receptive field model (25–27) to generate a complete spatiotemporal understanding of visual responses.

Given the pervasive use of the GLM approach in fMRI research, our results have broad implications for fMRI studies in any part of the brain. We find that timing of stimuli in the millisecond range has a large impact on the magnitude of fMRI responses, which has important implications for interpreting results of studies that vary the temporal characteristics of stimuli across conditions (61, 62). Critically, we show that, rather than ignoring fast cortical processing because of nonlinearities that are not accounted for in the GLM, it is possible to generate neural predictions at subsecond resolution and use them to accurately predict fMRI responses (40). These encoding approaches open exciting opportunities for investigating fast cortical mechanisms using fMRI in many domains, including somatosensory, auditory, and high-level cognitive processing.

In sum, our experiments elucidate the characteristics of temporal processing across human visual cortex. These findings are important, because they (*i*) explicate the contribution of transient and sustained visual responses across human visual cortex beyond V1 and (*ii*) show that accounting for neural responses in the millisecond range has important consequences for understanding fMRI signals in the second range in any part of the brain.

Materials and Methods

Participants. Twelve participants (six males, six females) with normal or corrected to normal vision participated in this study. All participants provided written informed consent, and the experimental protocol was approved by the Stanford University Institutional Review Board. Each individual participated in three fMRI sessions: two used to fit and validate the two-temporal channel model and one session in which we conducted population receptive field (pRF) mapping (25) to define retinotopic cortical regions and another experiment to define human motion-sensitive area (hMT+) (63–65). Detailed materials and methods can be found in *SI Materials and Methods*.

Temporal Channels Experiments. To obtain data that can be used to estimate and test the two temporal channel-encoding model, we introduce an fMRI paradigm that estimates independent sustained and transient contributions to fMRI responses across visual cortex using three experiments. All three experiments used the same stimuli, trial durations, and task and only varied

in their temporal presentation of the stimuli as detailed below and illustrated in Fig. 1.

Experiment 1—largely sustained stimulation. Phase-scrambled images were shown in trials of varying durations (2, 4, 8, 15, or 30 s per trial), in which a single phase-scrambled image was shown for the entire duration of the trial (Fig. 1, blue). Before and after each trial, there was a 12-s baseline period (blank gray screen matched to the mean luminance of the stimuli). Across trials, the numbers of stimuli (one per trial) and transients (at the onset and offset of each stimulus) are matched; just the duration of sustained stimulation varies. This experiment was designed to primarily activate the sustained channel, especially in the long trials.

Experiment 2—largely transient stimulation. Experiment 2 used the same trial durations and general experimental design as experiment 1, except that, in each trial, 30 different phase-scrambled images were shown briefly, each for 33 ms. Thus, the number of stimuli, number of transients, and total duration of visual stimulation are matched across trial durations in experiment 2. The only factor that varied across trials was the ISI between consecutively presented images. The ISI consisted of a blank mean luminance screen that was 33-ms long in the 2-s trials, 100 ms in the 4-s trials, 233 ms in the 8-s trials, 467 ms in the 15-s trials, and 967 ms in the 30-s trials (Fig. 1, red). This experiment was designed to maximally drive the transient channel and minimally the sustained channel, since each image was shown for only 33 ms.

Experiment 3—combined sustained and transient stimulation. Experiment 3 used the same design as experiment 2, except that, in each trial, we presented 30 different phase-scrambled images in a continuous fashion without an ISI between sequential images. The durations of images (67, 133, 267, 500, or 1,000 ms per image) varied across trials that were matched in length to experiment 1, whereby the 67-ms presentations occurred in the 2-s trials and the 1,000-ms presentations occurred in the 30-s trials (Fig. 1, green). This experiment was designed to drive both the sustained and transient channels, because (*i*) during the entire trial duration, there was always a stimulus on the screen and (*ii*) there were always 30 different images per trial.

Task. In all three experiments, participants were instructed to fixate on a small central dot and respond by button press when it changed color (occurring randomly once every 2–14 s, 8 s on average).

Data Acquisition. MRI data were collected using a 3-T GE Signa MR750 scanner at the Center for Cognitive and Neurobiological Imaging at Stanford University.

fMRI. We used a Nova 16-channel visual array coil (novamedical.com) to give participants a large unobstructed visual field of view. In each participant, we acquired two partially overlapping oblique slice prescriptions in separate scan sessions that together fully cover occipitotemporal cortex [resolution: 2.4 × 2.4 × 2.4 mm; one-shot T2*-sensitive gradient echo acquisition sequence: field of view (FOV) = 192 mm, echo time (TE) = 30 ms, repetition time (TR) = 1,000 ms, and flip angle = 73°]. We also collected T1-weighted in-plane images with the same prescription as the functional data to align each participant's data to their high-resolution whole-brain anatomy.

In a separate session, we obtained pRF mapping and hMT+ localizer data with the same receptive field coil setup and spatial resolution using 28 oblique slices covering the same brain volume but with a longer TR (resolution: 2.4 × 2.4 × 2.4 mm; one-shot T2*-sensitive gradient echo acquisition sequence: FOV = 192 mm, TE = 30 ms, TR = 2,000 ms, and flip angle = 77°). We again collected T1-weighted in-plane images in the same prescription to finely align in-plane data to the whole-brain anatomy of each participant.

Anatomical MRI. We acquired a whole-brain anatomical volume in each participant using a Nova 32-channel head coil [resolution: 1 × 1 × 1 mm; T1-weighted BRAVO pulse sequence: inversion time (TI) = 450 ms, flip angle = 12°, number of excitations (NEX) = 1, FOV = 240 mm].

Data Analysis. Data were analyzed with MATLAB using code from vistasoft (https://github.com/vistalab) and FreeSurfer (freesurfer.net).

Data preprocessing. Functional data were aligned to each participant's native anatomical space using T1-weighted in-plane images, and volumes acquired within the first 8 s of each run were discarded to allow time for magnetization to stabilize. We then performed slice time correction and motion compensation (within and between scans) and transformed voxel time series to units of percentage signal change. To normalize the baseline level of response across experiments, we subtracted from time points in each run the mean signal across the 4-s periods preceding the trial onsets in each run. This baseline removal procedure centers the mean response for the blank screen around zero to improve cross-validation performance (66) and to enable comparison of trial responses relative to the blank baseline.

Two-temporal channel model. In typical analysis of fMRI responses (18, 24), the stimulus vector is convolved with the HRF to obtain a prediction of the fMRI response. However, this model does not account for distinct temporal channels of neural responses (32–34, 67). To generate predicted fMRI responses accounting for the temporal channels, we implemented an encoding approach similar to that of Horiguchi et al. (38). Code for implementing the two-temporal channel model is freely available online: https://github.com/VPNL/TemporalChannels.

The model illustrated in Fig. 3 shows the procedure. First, we estimate the neural response of each channel by convolving the stimulus (Fig. 3*A*) separately with the neural IRF for the sustained channel (Fig. 3*B*, blue channel IRF) and the transient channel (Fig. 3*B*, red channel IRF). This generates the predicted neural response to the visual stimulus for each channel. Second, the estimated neural responses for each channel are convolved with the HRF (Fig. 3*C*) and summed to generate a prediction of the fMRI response. We use a GLM to solve for the contributions of the sustained and transient channels (β weights) given the measured fMRI responses. Thus, the BOLD response can be expressed as

$$\text{BOLD Response} = \beta_S([\text{stimulus} \otimes IRF_S] \otimes HRF) + \beta_T\left([\text{stimulus} \otimes IRF_T]^2 \otimes HRF\right),$$

where β_S and β_T are fitted response amplitude scalars for the sustained and transient channels, respectively; IRF_S and IRF_T are the impulse responses functions for the sustained and transient channels, respectively; and *HRF* is the canonical HRF.

The sustained neural channel is characterized by a monophasic IRF_S that generates a response for the entire duration of a stimulus. The transient neural channel is characterized by a biphasic IRF_T that generates a brief response at the onset and offset of an image (32–34, 36, 37). The transient channel also contains a nonlinearity (squaring operation) that generates positive responses both from the onset and from the offset of the stimulus, as firing rates associated with transient on or off responses are positive (68) and metabolically demanding (38, 41). We also implemented an otherwise identical model, except for a rectification to the transient channel (instead of squaring) to generate a positive on response but no off response. In both cases, the nonlinearities in this model are at the neural level, and a linear relationship is assumed between the neural and BOLD responses. The predictions and performance of these two dual channel models are indistinguishable in our data, and therefore, results from the original quadratic implementation are presented unless otherwise noted.

Modeling the neural impulse response. Our model used IRFs estimated from human psychophysics (37) (Fig. 3*B*) to approximate the temporal sensitivity of the human visual system. These IRFs are expressed as the difference between excitatory and inhibitory linear filters. The excitatory filter is expressed as

$$h_1(t) = u(t) \cdot [\tau(n_1 - 1)!]^{-1} \cdot \left(\frac{t}{\tau}\right)^{n_1 - 1} \cdot e^{-\frac{t}{\tau}},$$

where $u(t)$ is the unit step function at time t, τ is a fitted time constant, and n_1 is the number of stages in the excitatory filter. The inhibitory filter incorporates the same time constant and is expressed as

$$h_2(t) = u(t) \cdot [\kappa\tau(n_2 - 1)!]^{-1} \cdot \left(\frac{t}{\kappa\tau}\right)^{n_2 - 1} \cdot e^{-\frac{t}{\kappa\tau}},$$

where κ is the ratio of time constants for the two filters and n_2 is the number of stages in the inhibitory filter. Both the sustained and transient channel IRFs are derived with the formula

$$h_c(t) = \xi[h_1(t) - \zeta h_2(t)],$$

where the normalization parameter ξ is used to match the height of the functions and is equal to 1 for IRF_S and 1.44 for IRF_T. The transience parameter ζ is equal to zero for IRF_S and one for IRF_T.

The other parameters are taken from Watson (37) and are $\tau = 4.94$ ms, $\kappa = 1.33$, $n_1 = 9$, and $n_2 = 10$.

Modeling the visual input. Since the neural impulse response to a stimulus occurs on a millisecond timescale, we code each stimulus sequence in milliseconds. The stimulus is coded as a binary vector of ones and zeros, where one represents the presence of a stimulus and zero indicates when there is no stimulus, just a blank mean luminance screen (Fig. 3*A*). To capture the digital transitions of the display (constrained by the 60-Hz refresh rate of the projector), a 17-ms gap is coded at the offset of each image. Next, the stimulus vector is convolved separately with each channel IRF to generate separate sustained and transient neural response predictors (Fig. 3*B*). To model the corresponding fMRI responses from each channel, each of the two neural response predictors is convolved with an HRF (Fig. 3*C*) that was

sampled at the same high (millisecond) temporal resolution of the neural response predictors. Here, we slightly adapted the parameters of the canonical HRF implemented in SPM8 (www.fil.ion.ucl.ac.uk/spm/software/spm8) to better capture the rise and fall of the BOLD response in our measurements (delay of peak response = 5 s, delay of undershoot = 14 s, kernel length = 28 s).

Fitting the two-temporal channel model. Since the HRF acts as a low-pass temporal filter, this enables us to resample the predicted fMRI response to the lower temporal resolution of the acquired fMRI data (TR = 1 s). This resampled fMRI response predictor is compared with measured fMRI responses to solve for the contributions (β weights) of each channel. We normalized the predicted fMRI responses across the two channels, such that the maximal height is the same across both predictors. Then, we used a GLM to estimate the β weights of the sustained (β_S) and transient (β_T) predictors by comparing the predicted responses with the measured response using data concatenated across all runs of experiments 1 and 2. For ROI analyses, the GLM is applied to the mean response of each visual area in each participant. Quantification of model performance in each of experiments 1 and 2 is presented in Fig. 4 *D* and *E* for V1, Fig. 5*E* for hV4 and hMT+, and Fig. S3 for all ROIs. The predicted fMRI responses generated by the model are shown in Fig. 4*A* for V1 and Fig. S2 for other ROIs.

Validating the two-temporal channel model. We assessed the predictive power of the two-temporal channel model by testing how well it predicts responses in independent data obtained in experiment 3. Thus, we coded the visual stimulation of experiment 3 in the same manner described above and convolved it separately with the IRFs of the sustained and transient neural channels to generate the neural predictors. These neural predictors were then convolved with HRF and down-sampled to 1 s. Then, we multiplied each channel's fMRI response predictor with its respective β weight (β_S or β_T) that was estimated with independent data concatenated across experiments 1 and 2. We then tested how well the predicted responses matched the measured response in experiment 3. Model performance was operationalized as cross-validated R^2, also known as the coefficient of determination (that is, the proportion of response variance explained using β weights that were estimated from independent data). Although conceptually like the classical R^2 statistic, cross-validated R^2 can be negative when the residual variance of an inaccurate prediction exceeds the variance in the measured response. Quantification of cross-validation performance is shown in Fig. 4*F* for V1, Fig. 5*E* for hV4 and hMT+, and Fig. S3*D* for all ROIs. The predicted fMRI responses generated by the model are shown in Fig. 4*B* and Fig. S1 for V1 and Fig. S2 for other ROIs.

Hemodynamic models.

GLM. For model comparison with the linear systems approach used in fMRI, we fit a GLM to the data. This model predicts fMRI responses by convolving a stimulation vector with the HRF (Fig. 3 *A* and *C*) and can be expressed as

$$\text{BOLD Response} = \beta(\text{stimulus} \otimes HRF),$$

where β is a fitted response amplitude scalars and *HRF* is the canonical HRF.

HTD model. To test a hemodynamic model with two temporal channels, we fit an extension of the GLM proposed by Henson et al. (39) that incorporates an additional temporal derivative predictor to account for differences in the latency of BOLD responses across brain regions. This approach predicts fMRI responses as the weighted sum of a stimulation vector convolved with the canonical HRF and another factor convolved with the temporal derivative of the HRF (*HRF′*), which can be expressed as

$$\text{BOLD Response} = \beta_1(\text{stimulus} \otimes HRF) + \beta_2(\text{stimulus} \otimes HRF'),$$

where β_1 and β_2 are fitted response amplitude scalars and *HRF′* is the temporal derivative of the canonical *HRF*.

Balloon model. To test if our results can be explained by a nonlinear hemodynamic model, we implemented a version of the balloon model proposed by Buxton et al. (17). This input–state–output model treats the brain's vasculature as an inflatable balloon and describes the effect of blood flow on two state variables, *v* and *q*, that represent the blood volume and deoxyhemoglobin content, respectively. The state variables *v* and *q* vary over time as described by a system of flow equations and a balloon equation. The balloon component of the model can be expressed as

$$\text{BOLD Response} = V_0\left(k_1[1-q] + k_2\left[1-\frac{q}{v}\right] + k_3[1-v]\right),$$

where V_0 is the resting blood volume fraction (set to a standard value of 0.03) and the other parameters are calculated specifically for modeling fMRI signals measured at 3-T field strength ($k_1 = 6.7$, $k_2 = 2.73$, and $k_3 = 0.57$). Constants used in the flow equations include the resting net oxygen fraction (E_0), mean transit time (τ_{MTT}), a stiffness parameter (α), and a visoelastic time constant for inflation and deflation (τ_v). Here, we used a standard set of values for these parameters across all regions: $E_0 = 0.4$, $\tau_{MTT} = 2.5$ s, $\alpha = 0.4$, and $\tau_v = 25$ s.

As for the two-temporal channel model, we fit each hemodynamic model to data concatenated across all runs of experiments 1 and 2 and then, cross-validated the β weight using experiment 3 data.

Single-channel models with compressive nonlinearities.

CTS-p. The first implementation of the CTS model uses a static power law to explain subadditive temporal summation and adaptation effects observed in neural responses (40). This model can be expressed as

$$\text{BOLD Response} = \beta(\text{stimulus} \otimes IRF_\tau)^\varepsilon \otimes HRF,$$

where β is a fitted response amplitude scalar, τ is a time constant that determines the shape of the neural IRF, and ε is an exponential term used to compress responses.

CTS-n. The next implementation of the CTS model applies compression using divisive normalization instead of a power law (40). This model can be expressed as

$$\text{BOLD Response} = \beta\frac{(\text{stimulus} \otimes IRF_\tau)^2}{\sigma^2 + (\text{stimulus} \otimes IRF_\tau)^2} \otimes HRF,$$

where β is a fitted response amplitude scalar, τ is a time constant that determines the shape of the neural IRF, and σ is a semisaturation constant.

dCTS. The final implementation of a single-channel CTS model uses dynamic divisive normalization to compress responses (40). The dCTS predicts a prominent onset neural response and a subsequent declining continuous response for the remainder of the stimulus duration. This model can be expressed as

$$\text{BOLD Response} = \beta\frac{(\text{stimulus} \otimes IRF_\tau)^2}{\sigma^2 + (\text{stimulus} \otimes IRF_\tau \otimes LPF_{\kappa\tau})^2} \otimes HRF,$$

where β is a fitted response amplitude scalar, τ is a time constant that determines the shape of the neural IRF, σ is a semisaturation constant, and *LPF* is a low-pass filter parameterized by the time constant $\kappa\tau$. The *LPF* generates the attenuation of the later response.

We fit each CTS model to data concatenated across all runs of experiments 1 and 2 and then, cross-validated the β weight using experiment 3 data. To optimize the values of τ, ε, σ, and κ for each session and ROI during the fitting stage, we used a custom two-stage nonlinear optimization procedure consisting of a grid fit routine followed by gradient descent.

ACKNOWLEDGMENTS. We thank Justin Gardner, Gary Glover, Kendrick Kay, Anthony Norcia, Brian Wandell, and Kevin Weiner for fruitful discussions, and Ben Harvey as well as two anonymous reviewers for constructive feedback. This research was supported by National Eye Institute Grant 1R01EY02391501A1.

1. Kaplan E, Benardete E (2001) The dynamics of primate retinal ganglion cells. *Prog Brain Res* 134:17–34.
2. Hubel DH, Wiesel TN (1972) Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. *J Comp Neurol* 146:421–450.
3. Schiller PH, Malpeli JG (1978) Functional specificity of lateral geniculate nucleus laminae of the rhesus monkey. *J Neurophysiol* 41:788–797.
4. Derrington AM, Lennie P (1984) Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *J Physiol* 357:219–240.
5. Nassi JJ, Callaway EM (2009) Parallel processing strategies of the primate visual system. *Nat Rev Neurosci* 10:360–372.
6. Zeki S, Shipp S (1988) The functional logic of cortical connections. *Nature* 335:311–317.
7. Maunsell JH, Nealey TA, DePriest DD (1990) Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *J Neurosci* 10:3323–3334.
8. Merigan WH, Maunsell JH (1993) How parallel are the primate visual pathways? *Annu Rev Neurosci* 16:369–402.
9. Ferrera VP, Nealey TA, Maunsell JH (1994) Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. *J Neurosci* 14:2080–2088.
10. Nassi JJ, Lyon DC, Callaway EM (2006) The parvocellular LGN provides a robust disynaptic input to the visual motion area MT. *Neuron* 50:319–327.
11. Seidemann E, Poirson AB, Wandell BA, Newsome WT (1999) Color signals in area MT of the macaque monkey. *Neuron* 24:911–917.

12. Tootell RB, Hadjikhani N (2001) Where is 'dorsal V4' in human visual cortex? Retinotopic, topographic and functional evidence. *Cereb Cortex* 11:298–311.
13. Brewer AA, Liu J, Wade AR, Wandell BA (2005) Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nat Neurosci* 8:1102–1109.
14. Hansen KA, Kay KN, Gallant JL (2007) Topographic organization in and near human visual area V4. *J Neurosci* 27:11896–11911.
15. Witthoft N, et al. (2014) Where is human V4? Predicting the location of hV4 and VO1 from cortical folding. *Cereb Cortex* 24:2401–2408.
16. Larsson J, Heeger DJ (2006) Two retinotopic visual areas in human lateral occipital cortex. *J Neurosci* 26:13128–13142.
17. Buxton RB, Wong EC, Frank LR (1998) Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magn Reson Med* 39:855–864.
18. Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4221.
19. Boynton GM, Engel SA, Heeger DJ (2012) Linear systems analysis of the fMRI signal. *Neuroimage* 62:975–984.
20. Huettel SA, McCarthy G (2000) Evidence for a refractory period in the hemodynamic response to visual stimuli as measured by MRI. *Neuroimage* 11:547–553.
21. Ogawa S, et al. (2000) An approach to probe some neural systems interaction by functional MRI at neural time scale down to milliseconds. *Proc Natl Acad Sci USA* 97: 11026–11031.
22. Birn RM, Saad ZS, Bandettini PA (2001) Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. *Neuroimage* 14:817–826.
23. 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.
24. Friston KJ, Frith CD, Turner R, Frackowiak RS (1995) Characterizing evoked hemodynamics with fMRI. *Neuroimage* 2:157–165.
25. Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. *Neuroimage* 39:647–660.
26. Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452:352–355.
27. Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. *Neuron* 63:902–915.
28. Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. *Neuroimage* 56:400–410.
29. Heeger DJ (2017) Theory of cortical function. *Proc Natl Acad Sci USA* 114:1773–1782.
30. Nishimoto S, et al. (2011) Reconstructing visual experiences from brain activity evoked by natural movies. *Curr Biol* 21:1641–1646.
31. Çukur T, Huth AG, Nishimoto S, Gallant JL (2013) Functional subdomains within human FFA. *J Neurosci* 33:16748–16766.
32. De Valois RL, Cottaris NP (1998) Inputs to directionally selective simple cells in macaque striate cortex. *Proc Natl Acad Sci USA* 95:14488–14493.
33. De Valois RL, Cottaris NP, Mahon LE, Elfar SD, Wilson JA (2000) Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision Res* 40:3685–3702.
34. Conway BR, Livingstone MS (2003) Space-time maps and two-bar interactions of different classes of direction-selective cells in macaque V-1. *J Neurophysiol* 89: 2726–2742.
35. Kulikowski JJ, Tolhurst DJ (1973) Psychophysical evidence for sustained and transient detectors in human vision. *J Physiol* 232:149–162.
36. McKee SP, Taylor DG (1984) Discrimination of time: Comparison of foveal and peripheral sensitivity. *J Opt Soc Am A* 1:620–627.
37. Watson AB (1986) Temporal sensitivity. *Handbook of Perception and Human Performance*, eds Boff K, Kaufman L, Thomas J (Wiley, New York), pp 6-1–6-43.
38. Horiguchi H, Nakadomari S, Misaki M, Wandell BA (2009) Two temporal channels in human V1 identified using fMRI. *Neuroimage* 47:273–280.
39. Henson RN, Price CJ, Rugg MD, Turner R, Friston KJ (2002) Detecting latency differences in event-related BOLD responses: Application to words versus nonwords and initial versus repeated face presentations. *Neuroimage* 15:83–97.
40. Zhou J, Benson NC, Kay K, Winawer J (2017) Systematic changes in temporal summation across human visual cortex. bioRxiv:10.1101/108639.
41. Gawne TJ, Martin JM (2002) Responses of primate visual cortical neurons to stimuli presented by flash, saccade, blink, and external darkening. *J Neurophysiol* 88: 2178–2186.
42. Singh KD, Smith AT, Greenlee MW (2000) Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage* 12:550–564.
43. Sayres R, Grill-Spector K (2008) Relating retinotopic and object-selective responses in human lateral occipital cortex. *J Neurophysiol* 100:249–267.
44. Weiner KS, Grill-Spector K (2011) Not one extrastriate body area: Using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *Neuroimage* 56:2183–2199.
45. Stigliani A, Weiner KS, Grill-Spector K (2015) Temporal processing capacity in high-level visual cortex is domain specific. *J Neurosci* 35:12412–12424.
46. Van Essen DC, Gallant JL (1994) Neural mechanisms of form and motion processing in the primate visual system. *Neuron* 13:1–10.
47. Gilaie-Dotan S, Nir Y, Malach R (2008) Regionally-specific adaptation dynamics in human object areas. *Neuroimage* 39:1926–1937.
48. Shipp S, Zeki S (1985) Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature* 315:322–325.
49. Livingstone M, Hubel D (1988) Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science* 240:740–749.
50. Gardner JL, et al. (2005) Contrast adaptation and representation in human early visual cortex. *Neuron* 47:607–620.
51. Ungerleider LG, Desimone R (1986) Projections to the superior temporal sulcus from the central and peripheral field representations of V1 and V2. *J Comp Neurol* 248: 147–163.
52. Weiner KS, et al. (2016) The face-processing network is resilient to focal resection of human visual cortex. *J Neurosci* 36:8425–8440.
53. Azzopardi P, Jones KE, Cowey A (1999) Uneven mapping of magnocellular and parvocellular projections from the lateral geniculate nucleus to the striate cortex in the macaque monkey. *Vision Res* 39:2179–2189.
54. Carrasco M, McElree B, Denisova K, Giordano AM (2003) Speed of visual processing increases with eccentricity. *Nat Neurosci* 6:699–700.
55. Wager TD, Vazquez A, Hernandez L, Noll DC (2005) Accounting for nonlinear BOLD effects in fMRI: Parameter estimates and a model for prediction in rapid event-related studies. *Neuroimage* 25:206–218.
56. Gerber EM, Golan T, Knight RT, Deouell LY (2017) Cortical representation of persistent visual stimuli. *Neuroimage* 161:67–79.
57. McKeeff TJ, Remus DA, Tong F (2007) Temporal limitations in object processing across the human ventral visual pathway. *J Neurophysiol* 98:382–393.
58. Gentile F, Rossion B (2014) Temporal frequency tuning of cortical face-sensitive areas for individual face perception. *Neuroimage* 90:256–265.
59. Grill-Spector K, et al. (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203.
60. Friston KJ, Mechelli A, Turner R, Price CJ (2000) Nonlinear responses in fMRI: The Balloon model, Volterra kernels, and other hemodynamics. *Neuroimage* 12:466–477.
61. Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N (2008) A hierarchy of temporal receptive windows in human cortex. *J Neurosci* 28:2539–2550.
62. Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* 282: 108–111.
63. Tootell RB, et al. (1995) Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci* 15:3215–3230.
64. Huk AC, Dougherty RF, Heeger DJ (2002) Retinotopy and functional subdivision of human areas MT and MST. *J Neurosci* 22:7195–7205.
65. Amano K, Wandell BA, Dumoulin SO (2009) Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *J Neurophysiol* 102: 2704–2718.
66. Brouwer GJ, Heeger DJ (2011) Cross-orientation suppression in human visual cortex. *J Neurophysiol* 106:2108–2119.
67. Saul AB, Carras PL, Humphrey AL (2005) Temporal properties of inputs to direction-selective neurons in monkey V1. *J Neurophysiol* 94:282–294.
68. Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284–299.
69. Weiner KS, et al. (2017) The cytoarchitecture of domain-specific regions in human high-level visual cortex. *Cereb Cortex* 27:146–161.
70. Benson NC, Butt OH, Brainard DH, Aguirre GK (2014) Correction of distortion in flattened representations of the cortical surface allows prediction of V1-V3 functional organization from anatomy. *PLoS Comput Biol* 10:e1003538.
71. Willenbockel V, et al. (2010) Controlling low-level image properties: The SHINE toolbox. *Behav Res Methods* 42:671–684.
72. Brainard DH (1997) The psychophysics toolbox. *Spat Vis* 10:433–436.
73. Weiner KS, Grill-Spector K (2013) Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychol Res* 77:74–97.
74. Kay KN, Winawer J, Mezer A, Wandell BA (2013) Compressive spatial summation in human visual cortex. *J Neurophysiol* 110:481–494.
75. Kay KN, Weiner KS, Grill-Spector K (2015) Attention reduces spatial uncertainty in human ventral temporal cortex. *Curr Biol* 25:595–600.
76. Dumoulin SO, et al. (2000) A new anatomical landmark for reliable identification of human area V5/MT: A quantitative analysis of sulcal patterning. *Cereb Cortex* 10: 454–463.
77. Fischl B, et al. (2008) Cortical folding patterns and predicting cytoarchitecture. *Cereb Cortex* 18:1973–1980.
78. Hinds O, et al. (2009) Locating the functional and anatomical boundaries of human primary visual cortex. *Neuroimage* 46:915–922.
79. Fischl B, Sereno MI, Tootell RB, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8:272–284.
80. Wang L, Mruczek RE, Arcaro MJ, Kastner S (2015) Probabilistic maps of visual topography in human cortex. *Cereb Cortex* 25:3911–3931.