Modeling systematic differences in temporal summation and adaptation in human visual cortex: evidence from fMRI and intracranial EEG

Jingyang Zhou[1*], Noah C. Benson[1], Kendrick Kay[2], Jonathan Winawer[1,3,4]

1. Department of Psychology, New York University
2. Department of Radiology, University of Minnesota, Twin Cities
3. Center for Neural Science New York University
4. Stanford Human Intracranial Cognitive Electrophysiology Program (SHICEP)

* Corresponding Author, jingyang.zhou@nyu.edu

Significance Statement

Combining sensory inputs over time is fundamental to seeing. Due to temporal integration, we do not perceive the flicker in fluorescent lights nor the discrete sampling of movie frames; instead we see steady illumination and continuous motion. As a result of adaptation, elements of a scene that suddenly change in appearance are more salient than elements that do not. Here we investigated how the human nervous system combines visual information over time, measuring both functional MRI and intracortical EEG. We built predictive models using canonical neural computations, and account for temporal integration and adaptation. The models capture systematic differences in how information is combined in different visual areas, and generalize across instruments, subjects, and stimuli.

Abstract

The visual system analyzes image properties across multiple spatial and temporal scales. Population receptive field ("pRF") models have successfully characterized spatial representations across the human visual pathways. Here, we studied temporal representations, measuring fMRI and electrocorticographic ("ECoG") responses in posterior, lateral, ventral, and dorsal visual areas to briefly viewed contrast patterns. We built a temporal pRF model employing linear summation and time-varying divisive normalization. Our model accurately predicts the fMRI amplitude and ECoG broadband time-course, accounting for two phenomena – accumulation of stimulus information over time (summation), and response reduction with prolonged or repeated exposure (adaptation). We find systematic differences in these properties: summation periods are increasingly long and adaptation more pronounced in higher compared to earlier visual areas. We propose that several features of temporal responses – adaptation, summation, and the timescale of temporal dynamics – can be understood as resulting from a small number of canonical neuronal computations.

Keywords: Electrocorticography, Functional MRI, Population Receptive Fields, Temporal Summation, Visual Cortex, Normalization, Adaptation, Temporal dynamics, Repetition suppression

1. Introduction

A successful visual system extracts meaning from stimuli that vary across space and time. This requires integrating and segregating features at multiple scales. The classic visual perception example that requires flexible spatial pooling is object recognition: recognizing an object requires grouping features across space (1), but grouping over too large a region results in a jumbled, or 'crowded', percept that interferes with recognition (2).

Information is also integrated at multiple time scales to achieve a coherent interpretation of visual stimuli. At a fine scale, interpreting scenes requires combining and segregating features across image changes that arise from eye movements and blinks. At a longer scale, features must be appropriately combined across occlusion events and extended actions. Some visual features, such as texture boundaries, are integrated in short temporal windows (< ~20 ms) (3), whereas other stimuli, such as words, are integrated over much longer periods (> ~100 ms) (4). Abnormalities in temporal processing can cause perceptual deficits in patients with optic neuritis (5) and amblyopia (6), and may be a contributing factor in dyslexia (7). A model of how the different areas in the human visual system combine stimulus information over time is necessary for understanding the recognition process, and for establishing norms against which to compare disorders.

In the spatial studies, two large-scale trends emerge, and may serve as correlates of achieving increasingly invariant representation of objects and scenes (8). First, along the visual hierarchy, from striate to extrastriate areas, receptive field size increases, measured using both electrophysiology (9) and fMRI (10-12). Second, when two or more stimuli are presented together, spatial summation becomes increasingly more subadditive in later visual areas (11, 13, 14).

Here, we investigated the scale and properties of how neuronal populations pool information over time. We characterized responses to brief stimuli at the time scale of neuronal dynamics (ten to hundreds of ms) in many visual areas, measured with fMRI and electrocorticography (ECoG). fMRI measurements have the advantage of being non-invasive and recording from many visual areas in parallel. But the fMRI measurements also have limits for interpreting the neuronal response. First, subadditivities in the fMRI response can arise from the stimulus-to-neuronal transform or neuronal-to-BOLD transform. Second, the slow hemodynamics does not enable us to characterize the detailed time course of the neuronal response. The ECoG measurements complement fMRI by providing much greater temporal resolution and by not compounding nonlinearities in the neuronal response with nonlinearities in the hemodynamics.

To quantify and understand how temporal information is encoded across visual cortex, we built temporal population receptive field ("pRF") models which predict the fMRI and ECoG responses to arbitrary stimulus time courses, and we examined the model parameters in visual areas spanning V1 to IPS. Together, the temporal pRF model reveals a systematic hierarchy of increasingly large temporal windows and increasingly large deviations from linear summation, paralleling the hierarchy of spatial receptive fields.

2. Results

We present two variants of a temporal pRF model. The first was fit to fMRI data, and captures subadditivities of the BOLD amplitude for stimuli with different temporal profiles (sections 2.1-2.4). Fitting the ECoG responses required expanding the model to account for temporal dynamics at the ms scale (section 2.5). Finally, we test how accurately the expanded, dynamic model predicts the fMRI responses (section 2.6).

2.1 Measuring temporal summation in visual cortex

In each trial of the fMRI experiment, participants viewed either one or two pulses of a static spatial contrast pattern. Each pattern was an independently generated band-pass noise image (24° diameter), used in prior studies of spatial encoding (11, 15), except that for the two-pulse stimuli, the two spatial patterns were identical. Each trial used one of thirteen distinct time courses (Figure 1A). The durations of the one-pulse stimuli and the ISIs of the two-pulse stimuli were the same: 0, 17, 33, 67, 134, 267, 533ms, and each pulse in the 2-pulse stimuli was 134ms. The 0-ms one-pulse stimulus was a blank (mean luminance), and the two-pulse stimulus with 0 ISI was identical to the one-pulse stimulus of twice the length (267ms). Four participants were scanned, and data were binned into nine bilateral, eccentricity-restricted (2-10°) visual areas defined from a separate retinotopy scan.

The fMRI data were analyzed in two stages. First, we extracted the amplitude (ß-weight) for each stimulus condition using a variation of the general linear model, “GLM denoise” (16), a technique that improves the signal-to-noise ratio by including noise regressors in the GLM. Second, we fitted the temporal pRF model to the GLM ß-weights, averaged across voxels within ROIs.

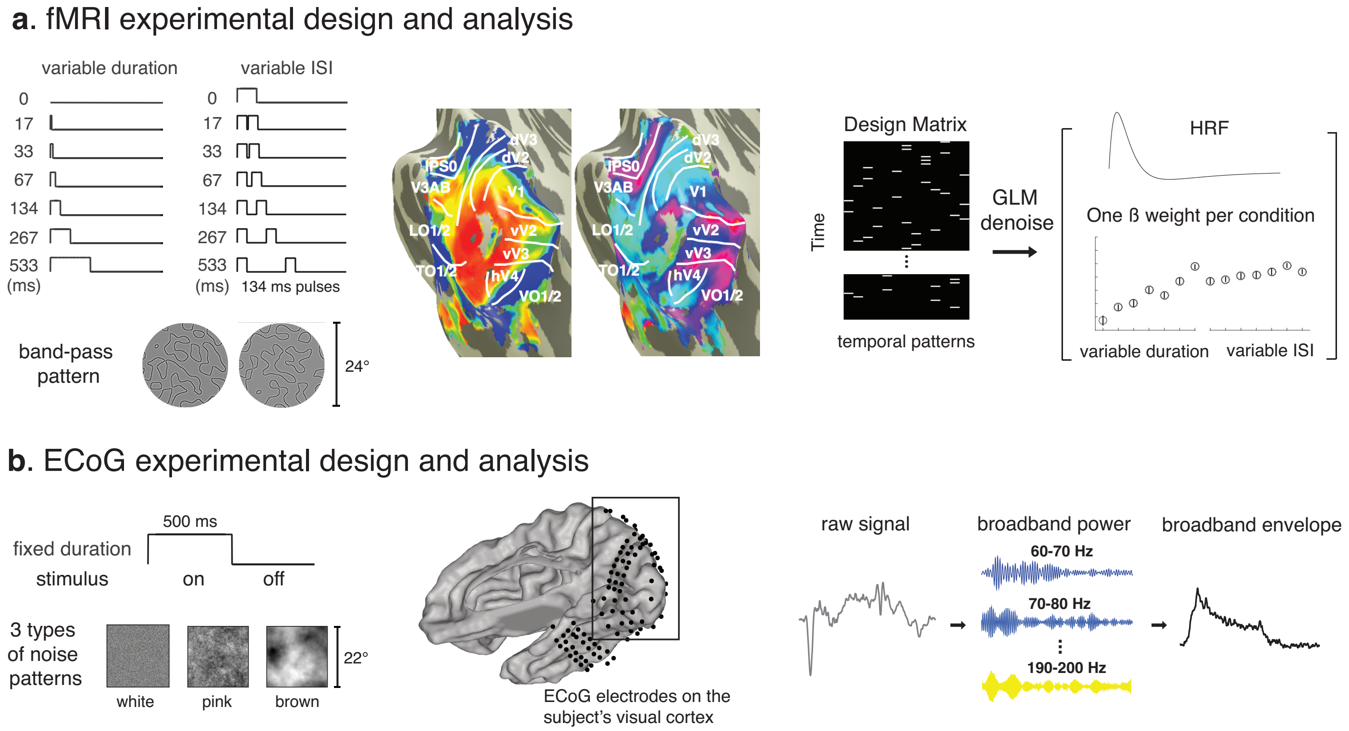


Figure 1. Experimental design and analysis. *(A) fMRI.* Participants were presented with one or two pulses of large field (24°) spatial contrast patterns. One-pulse stimuli were of varying durations and two-pulse stimuli were of varying ISI (with each pulse lasting 134ms). Nine visual field maps or visual field maps pairs were bilaterally identified for each participant (V1; V2; V3; hV4; VO-1/2; V3A/B; IPS-0/1; LO-1/2; TO-1/2). The temporal conditions were presented in random order, indicated by the white bars in the 13-column design matrix (one column per temporal condition). To analyze the data, we extracted a ß-weight for each temporal condition per area using a variant of the general linear model, GLM denoise. *(B) ECoG.* In the ECoG experimental, one 500ms pulse of a large field (22°) noise pattern (either white, pink or brown noise) was presented at the beginning of each 1s trial. We summarized the ECoG signal as the envelope of the whitened broadband response (60-200 Hz), averaged across stimulus class, trials, and electrodes within the same retinotopically defined visual areas.

2.2 Temporal summation in visual cortex is subadditive

We tested the linearity of the fMRI BOLD signal in each visual area. To do so, we assume a time-invariant linear system such that the BOLD amplitude (GLM ß-weight) is proportional to the total stimulus duration within the trial[1]. For example, the linear prediction is that a stimulus of duration *2t* produces twice the amplitude as a stimulus of duration *t,* and the same amplitude as two-pulse stimuli, with total duration 2*t* (Figure 2A). This prediction is not borne out by the data. The response to a stimulus of length *2t* is about 75% of the linear prediction in V1 and 50% in TO (Figure 2B, left panel). This failure of linearity is found in all visual areas measured, with temporal summation ratios below 0.8 for all ROIs, and a tendency toward lower ratios in later areas (Figure 2C).

[1] Because the stimulus events are short (≤800 ms), and the hemodynamic response function (hRF) is low-pass (on the order of seconds), the convolution of the stimulus time course with a neural impulse response function, followed by the convolution of this output with an hRF, is approximately the same as summing the stimulus time course (to create a scaled impulse), followed by convolution of the impulse with the hRF.

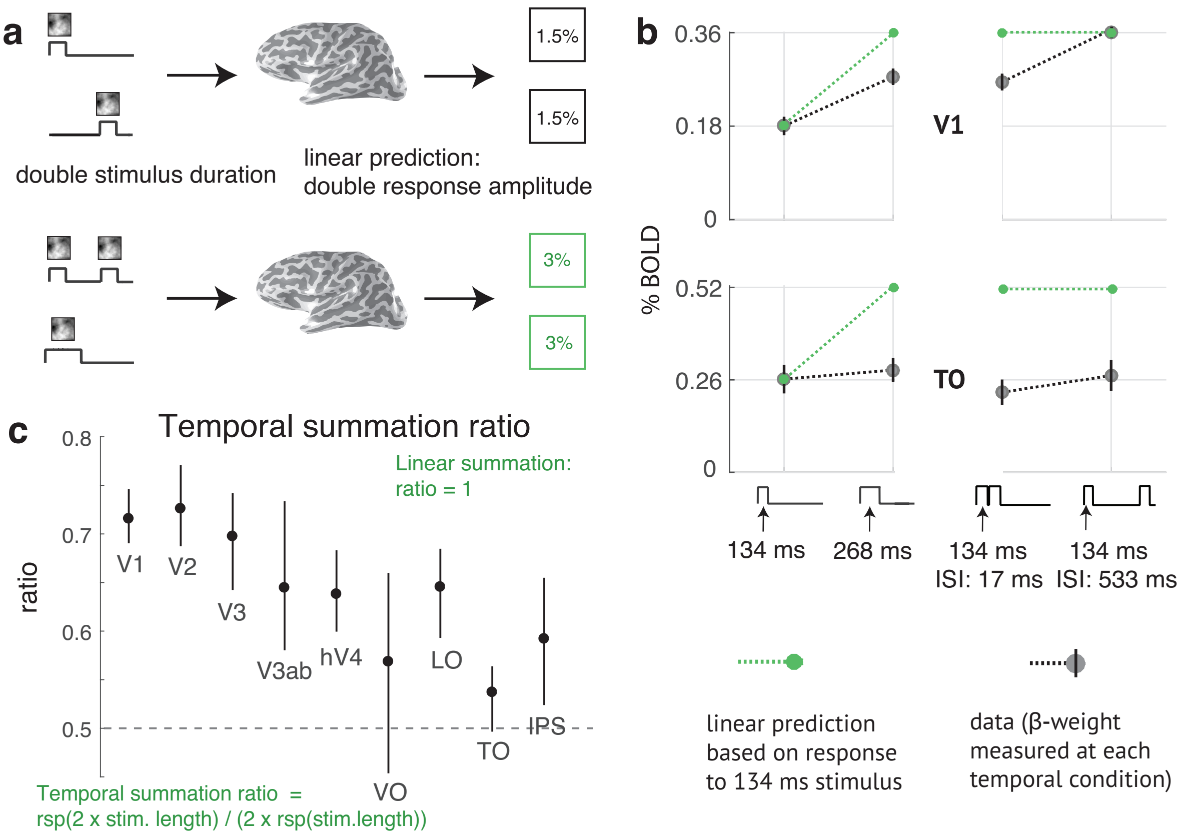


Figure 2. Sub-linear temporal summation in visual cortex. *(A) Linear temporal summation prediction.* The sum of the response to two separate events (top) is equal to the response to the two events in the same trial, with or without a brief gap between them (bottom). *(B) Sub-linear temporal summation.* Gray dots are the measured responses to a 134-ms pulse, a 268-ms pulse, and two 134-ms pulses, with either a 17-ms or 134-ms gap between them. Plots show the mean across subjects and 50% CI (bootstrapped across repeated scans within each subject). The green circles and dotted lines are the linear prediction based on the response to the single 134-ms pulse. For V1, the measured responses are less than the linear prediction except when there is a long gap. For TO, all responses are less than the linear prediction. *(C) Temporal summation ratio.* Temporal summation ratio is the response to a stimulus of length 2x divided by twice the response to a single pulse stimulus of length x, averaged across 5 stimulus pairs (e.g., 17 and 34ms, 34 and 68ms, etc.). Linear summation occurs when the temporal summation ratio is 1. Error bars represent the 50% CI (bootstrapped across scans). The temporal summation ratio is higher in early visual areas (~0.7 in V1-V3), and lower in later areas (between 0.5 and 0.65). The ROIs on the X-axis are arranged in order of increasing spatial pRF size at 5 deg eccentricity, as a proxy for order in the visual hierarchy.

A further failure of linearity occurs for trials with two pulses and variable ISI: the response is larger when the ISI is longer, especially in V1, whereas the linear prediction is that the amplitudes are the same, and double the response to the one-pulse (Figure 2B, right). When the ISI is long, the response in V1 is close to the linear prediction made from the one-pulse stimulus. In TO, even with a long ISI the response is well below the linear prediction. This pattern, whereby the response to a second stimulus is reduced for short ISIs, and larger for longer ISIs, is often called adaptation and recovery (17, 18). For TO, the recovery time is longer than V1.

2.3 The temporal subadditivity is captured by a compressive temporal summation model (CTS)

We modeled the temporal subadditivity with a compressive temporal summation model ("CTS"), analogous to the compressive spatial summation model (CSS) used to predict fMRI responses to spatial patterns (11, 19). The model predicts the neuronal response by convolving the stimulus time course with a temporal impulse response function, and then passing the output through a power-law static non-linearity (Figure 3). The model is linear if the exponent equals 1 and subadditive if less than

1. Finally, we summed the time-varying neuronal prediction to derive a single value, which, when scaled, represents the predicted BOLD amplitude.

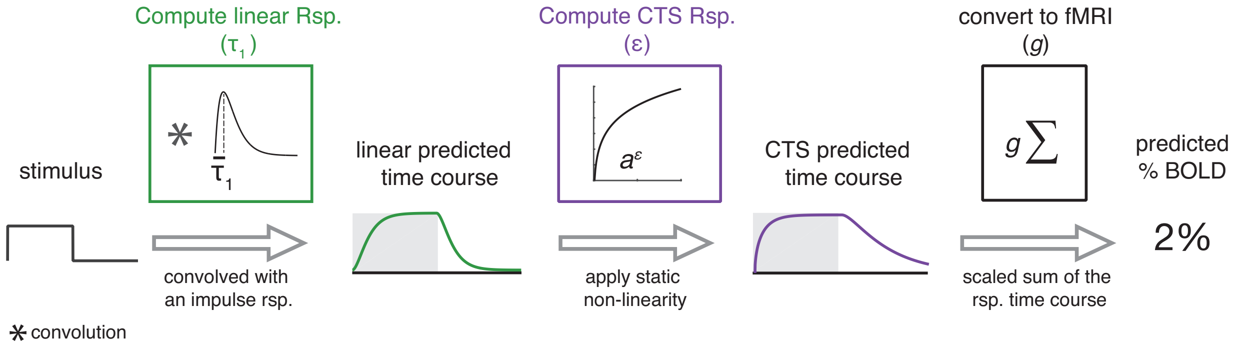

Compute linear Rsp.
(τ1)
Compute CTS Rsp.
(ε)
convert to fMRI
(g)
stimulus
*
τ1
linear predicted time course
a^ε
CTS predicted time course
g∑
predicted % BOLD
convolved with an impulse rsp.
apply static non-linearity
scaled sum of the rsp. time course
2%
* convolution

Figure 3. Compressive temporal summation (CTS) model. The CTS model takes the stimulus time course for a trial as input (1 when the contrast pattern is present, 0 when it is absent). The input is convolved with an impulse response function parametrized by τ_1, to produce a linear prediction. The linear prediction is then point-wise exponentiated, (parameterized by ε) to make the CTS prediction. Finally, the time-varying CTS prediction is summed and scaled (g) to predict the percent BOLD response. If ε is 1, the CTS prediction is identical to the linear prediction. In this special case, the value of τ_1 has no effect on the predicted BOLD, since the output will always be proportional to the total stimulus duration. The CTS model was fit for each ROI by finding the values of τ_1, ε, and g that minimized the squared error between the model predictions and the GLM ß-weights.

We compared the CTS model (fitted exponent) to a linear model (exponent fixed at 1) by measuring cross-validated accuracy. The CTS model is more accurate than the linear model for all areas (Figure 4A). The linear model substantially underpredicts responses to short durations and overpredicts responses to long durations, whereas the CTS model does not. Further, the predictions of the linear model do not depend on ISI, whereas the CTS model correctly predicts that the response amplitude increases with longer ISI. The cross-validated predictions of the CTS model capture more than 90% of the variance of the left-out data for all 9 ROIs. This represents an improvement of 8-17% compared to the linear model. The improvement is more pronounced in later than early areas (LO/TO/IPS vs. V1-V3).

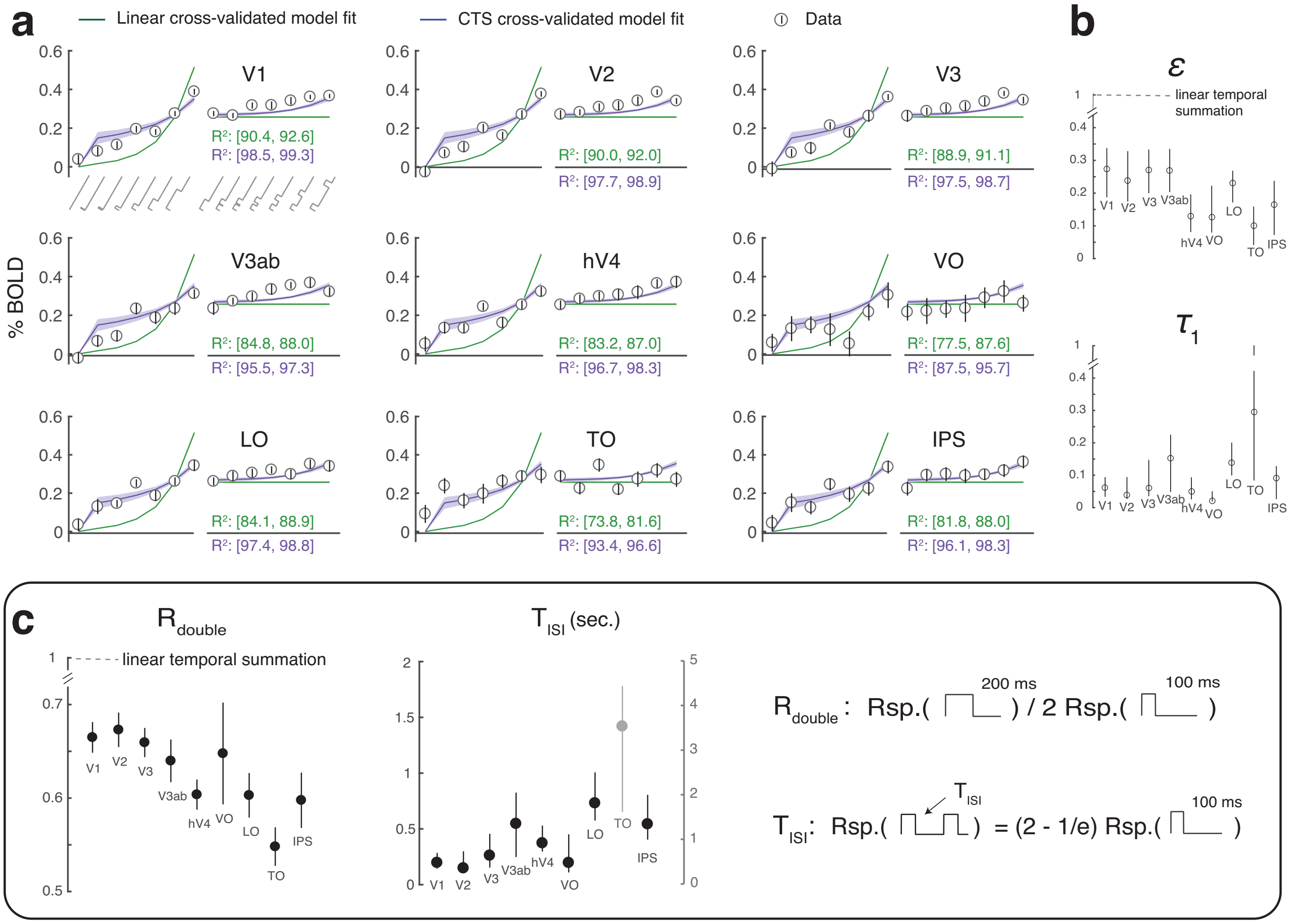


Figure 4. CTS model fits to BOLD data across visual areas. *(A) Data and predictions.* BOLD responses to each temporal condition averaged across subjects are plotted as circles. The temporal conditions on the x-axis show increasing durations of one-pulse stimuli (0 to 533ms; left) and increasing ISI of two-pulse stimuli (0 to 533ms, right). Error bars show the 50% CI bootstrapped across repeated scans. Predictions for the linear (green) and CTS (purple) model fits are computed by leave-out-one-condition cross-validation. Shaded regions represent the 50% CI in predictions across bootstraps (not visible for the linear fit because the CI is narrow). The cross-validated accuracy (R^2) is higher for the CTS model in each area. *(B) CTS model parameter estimates.* The estimated exponent ε is below 1 in each area and lower (more sub-linear) in later areas (~0.15, hV4-IPS versus ~0.25, V1-V3ab). The time constant τ_1 is short in V1-V3. *(C) Summary metrics*. Two summary metrics of the CTS model reveal a pattern across ROIs. R_{double} is the ratio of the predicted response to a 200-ms pulse divided by twice the response to a 100-ms pulse. R_{double} is below 1 for all ROIs, indicating sub-additivity, and decreases along the visual hierarchy (V1-V3, ~0.67, LO-IPS, < 0.6). T_{ISI} is the length of ISI required for the response to two 100-ms pulses to approach the linear prediction. T_{ISI} is short in the earlier areas (V1-V3, ~250 ms) compared to most of the later areas. See figure S1 for fits to individual subjects. In the T_{ISI} panel, the data for TO is outside the range of other areas and is plotted on the right y-axis.

2.4 The CTS model fits capture systematic differences between areas

The CTS model is parameterized by τ_1, ε, and a gain factor, *g*. τ_1 is the latency to peak in the temporal impulse response function, and therefore is related to temporal summation window length; ε is the exponent, and represents how compressive the temporal summation is. The exponent ε is less than 1 for all ROIs, and is smaller in later (hV4-IPS) than in earlier areas (V1-V3), consistent with the pattern found for spatial summation (11) (Figure 4b; see Figure S1 for individual subject fits). The same pattern was also found in a second experiment using identical temporal conditions but different spatial patterns, including noise stimuli and face images (Figure S2). A consequence of more compressive temporal summation is that the response amplitude varies less with minor changes in stimulus duration, just as greater compression of spatial summation predicts more tolerance to changes in size and position (11).

From the current fMRI data set, we did not observe systematic variation in τ_1. Our interpretation is that we do not have enough power to accurately fit τ_1 due in part to the coarse temporal resolution of fMRI. (See Figure S8A for CTS parameter recovery.) Because fitting a parameter that is not well-constrained by the data can affect the fit to other parameters, we re-fit the CTS model with τ_1 fixed at 0.05, 0.1, or 0.2 s; in each case, ε is below 1 for all ROIs, and lower in later areas than early areas, just as observed in the full model fit.

To further examine the differences in temporal processing between ROIs, we summarized the CTS model in terms of two metrics that have more directly interpretable units: R_{double} and T_{ISI} (Figure 4b). R_{double} is the ratio between the CTS-predicted BOLD response to a 100-ms stimulus and a 200-ms stimulus. Lower R_{double} means more compressive temporal summation. Later visual areas have lower R_{double} than earlier ones. T_{ISI} is the minimal duration separating two 100-ms pulses such that the response to the paired stimuli is close to the linear prediction from the single stimulus. Similar to previous measurements at longer time scales (20, 21), the recovery time is longer for later than earlier visual areas.

In a separate analysis, we asked whether model parameters differed as a function of eccentricity, as suggested by differential temporal sensitivity in V1 between fovea and periphery (22). We did not find reliable differences for parafovea (2-5 deg) versus periphery (5-10 deg) (Figure S3). This may be due to the limited range of eccentricities; as Horiguchi et al (22) found the biggest difference in temporal sensitivity between fovea and the far periphery (20-60 deg), whereas we only tested out to 10 deg.

2.5 Temporal dynamics of normalization

There are at least two potential sources of subadditivity contributing to the BOLD response: subadditivity of the neuronal response with respect to the stimulus time course, and subadditivity of the fMRI amplitude with respect to the neuronal response. To evaluate additivity of the neuronal response in isolation, and to characterize the neuronal response at a finer temporal scale, we re-analyzed data from a published ECoG experiment (22) (Figure 1B). We analyzed data from 45 electrodes in visual cortex (Figure S4, ECoG subject 1; Figure S7, ECoG subject 2). In each trial, a static texture (22°-diameter) was presented for 500ms followed by a 500-ms blank. We analyzed trials with noise patterns of $1/f^n$ amplitude spectra, with n=0, 1, or 2 (white, pink, or brown noise). We summarized the ECoG signal as the time-varying envelope of the broadband response (60-200 Hz), averaged across stimulus class, trials, and electrodes within visual areas, as the broadband response is a correlate of the multiunit spiking activity (23). Because there were fewer electrodes in anterior ROIs than in V1-V3, we grouped the anterior electrodes into lateral, ventral, and dorsal regions.

Across all visual areas, the time course of the ECoG broadband signal consisted of a large transient power increase, followed by a lower sustained response (e.g., Figure 5A, left). This transient/sustained pattern is similar to that observed for electrophysiological spiking data ((e.g., 24, 25, 26)). The CTS model predictions fail to capture the sharp onset transient (Figure 5A, middle panels). To account for the temporal pattern of the ECoG response, we implemented a dynamic variation of the CTS model, “dCTS”.

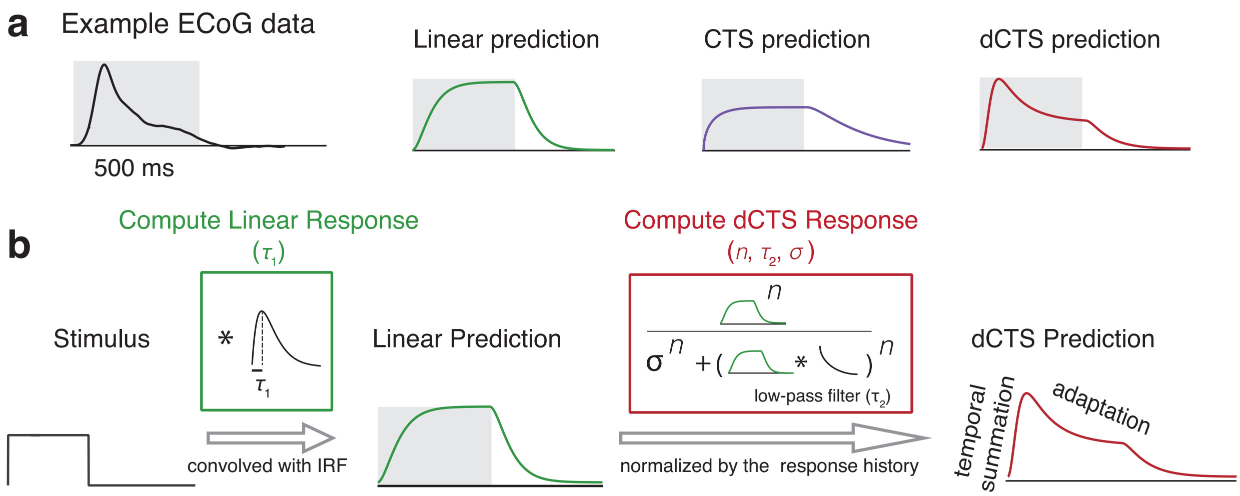


Figure 5. Dynamic CTS (dCTS) model. *(A)* The broadband envelope to ECoG data (left) contains an early transient and then a lower level sustained response, lasting until after stimulus offset (example response from a V2 electrode for a 500-ms stimulus). The linear and CTS predictions (middle panels; exponent 1 and 0.2, respectively) do not capture the transient-sustained pattern observed in the data. A CTS model with a dynamic rather than static nonlinearity (dCTS; right panel) qualitatively matches the data. *(B) DCTS model.* The first step of the dCTS model computation is the same as the CTS model – linear convolution of the stimulus time course with an impulse response function (parametrized by τ_1). The dCTS model uses divisive normalization rather than a static power law to achieve temporal compression. The numerator is the linear response time course raised point-wise to a power *n*, assumed to be greater than 1. We use the symbol *n* for the dCTS exponent rather than ε to indicate that the exponent here is greater than 1 (expansive), whereas in the CTS model ε is less than 1 (compressive). This predicted response is then divisively normalized, with the normalization being the sum of a semi-saturation constant (σ), and a low-pass filtered linear response (parametrized by time constant τ_2), each raised to the same power *n*. The low-pass exponential causes the normalization to be delayed, so that the early response is large (un-normalized), reflecting temporal summation, and the later response is reduced, reflecting adaptation or normalization. This pattern matches the transient-sustained pattern in the time series data.

The dCTS model, like the CTS model, is linear-nonlinear. But in contrast to CTS, in which the non-linearity is applied uniformly in time as a power law, the dCTS non-linearity was implemented as a divisive normalization, with the normalization signal low pass-filtered (Figure 5B). The low-pass filtering causes the response reduction to lag the linear response, producing an onset transient. This feedforward model with delayed normalization approximates a feedback normalization proposed by Heeger (27). The numerator contains the linear (un-normalized) response parameterized by τ_1. The denominator contains the sum of a semi-saturation constant (σ) and the low-passed linear response (parameterized by τ_2). All three terms are raised to the power *n*. Following stimulus onset, the response increases rapidly due to the exponent *n*, and then reduces due to normalization (controlled by σ and *n*). The time constant τ_2 controls the time scale of normalization. Because we are modeling the population response summed via the ECoG electrode, we treat the normalization pool (denominator) and the response pool (numerator) as the same, as previously assumed in spatial models of the fMRI signal (11).

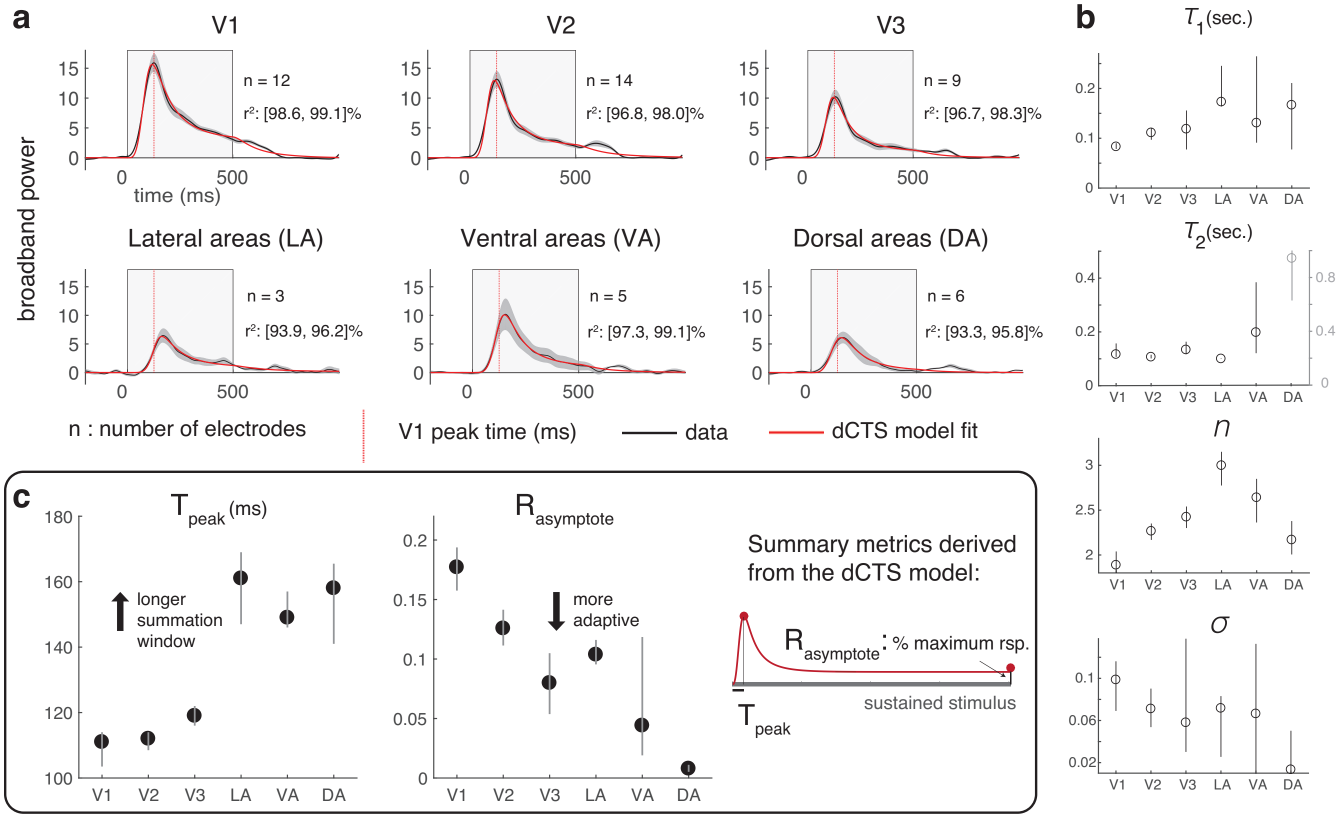


Figure 6. Dynamic CTS model fits to ECoG data across visual areas. *(A) Model fits.* The dCTS model fits (red) accurately describe the ECoG broadband time course (black) in all visual areas. Due to lower numbers of electrodes in ROIs beyond V3, anterior ROIs are grouped into lateral (LO-1/2), ventral (hV4, VO-1/2), and dorsal (V3A/B, IPS-0-4). Data are averaged across trials and electrodes within ROIs, and models are fit to the average data. Each trial had a 500-ms stimulus (gray box) followed by a 500-ms blank. Plots show the mean and 50% CI for data (bootstrapped 100 times across electrodes within an ROI), and the model fit averaged across the 100 bootstraps. The number of electrodes per ROI and the 50% CI of model accuracy (r^2 per bootstrap) are indicated in each subplot. *(B) DCTS model parameters.* τ_1, τ_2, n, and σ. *(C) Re-parameterized dCTS model.* The model fits were summarized by two derived constants, T_{peak}, $R_{asymptote}$. T_{peak} is the duration from the onset of a sustained stimulus to the peak response. T_{peak} is longer for later ROIs, ranging from ~115ms (V1/V2) to ~160ms (lateral and dorsal ROIs). $R_{asymptote}$ is the level at which the response asymptotes for a sustained stimulus, as a fraction of the peak response. A smaller $R_{asymptote}$ indicates a greater extent of normalization. $R_{asymptote}$ is largest in V1 (~0.18) and declines in extrastriate areas.

The dCTS model, fitted to the ECoG broadband time series, captures the main features of the temporal dynamics in all ROIs - an initial transient followed by a sustained response (Figure 6A) – explaining 93% to 99% of the variance in the time courses. In some electrodes, especially those with peripheral receptive fields (Figure S4), there is a small positive deflection 100-200ms after stimulus offset. This is consistent with the finding that peripheral V1 has a relatively greater sensitivity to visual transients (28). This feature of the data is not captured by our model. A variant of the model, in which the linear impulse response function is biphasic, predicts the offset transient (Figure S5). Because the offset response is not evident for most electrodes, we use the monophasic response function for primary analyses.

Although the time-courses in all ROIs follow a transient-sustained pattern, they differ in detail. These differences are reflected in model parameters (Figure 6B). This is clearest for the time-scale of the impulse response function, τ_1, which generally increases along the visual hierarchy, from ~90ms (V1) to ~150ms in later areas. The parameters n, σ, and τ_2, do not follow as clear a pattern. However, the relationship between a single model parameter and the predicted response depends on the other parameters. For example, the level of the sustained response increases with n and decreases with σ.

To clarify the effect of the fitted parameters on the resultant time series, we derived two summary metrics for each model fit (Figure 6C): For a sustained stimulus, the model predictions were summarized by the time to peak (T_{peak}) and the asymptotic response amplitude ($R_{asymptote}$). A longer T_{peak} indicates a longer temporal summation window, and increases slightly from V1 to V3, and substantially in more anterior areas. A smaller $R_{asymptote}$ corresponds to a lower sustained response, indicative of more normalization. $R_{asymptote}$ is highest in V1, and decreases substantially in extrastriate areas.

In a separate analysis, we assessed the effect of our signal processing pipeline on the parameter estimates. Because the broadband envelope is derived from a modulating signal, its temporal resolution is limited by the period of the oscillations. Simulations show that this has a small but measurable effect on parameter estimates of the dCTS model, with no change in the general pattern of results (Figure S6).

2.6 Integration of fMRI and ECoG

The fMRI and ECoG data sets were fit with different variants of the CTS model. The two variants were chosen for practical reasons – the slow time scale of the fMRI response limits our ability to resolve the dynamics of the nonlinearity, and the static non-linearity used to fit the fMRI data is a poor fit to the ECoG time course. Here we asked how accurately the dCTS model, fit to ECoG data, predicts the fMRI responses. In each ROI, the dCTS parameters derived from ECoG data were used to generate time-course predictions for the 13 distinct temporal stimuli used in the fMRI experiment. We converted these time courses to predicted BOLD amplitudes assuming one of two fMRI transforms: either linear, which was shown to be a reasonable approximation for relatively long ISIs (a few to many seconds) (29-31), or a square root transform, as recently proposed (32) (Figure 7). Because the dCTS model parameters were derived from the ECoG data alone, there were no free parameters other than a gain factor. Although the models were solved with different participants, different stimuli, and a different instrument, they nonetheless accurately fit the BOLD data, with r^2 ranging from 67% to 94% for the linear fMRI transform, and 80% to 96% for the square root transform. For every ROI, the square root transform was slightly more accurate than the linear transform. The most accurate fits for both transforms are for V1-V3.

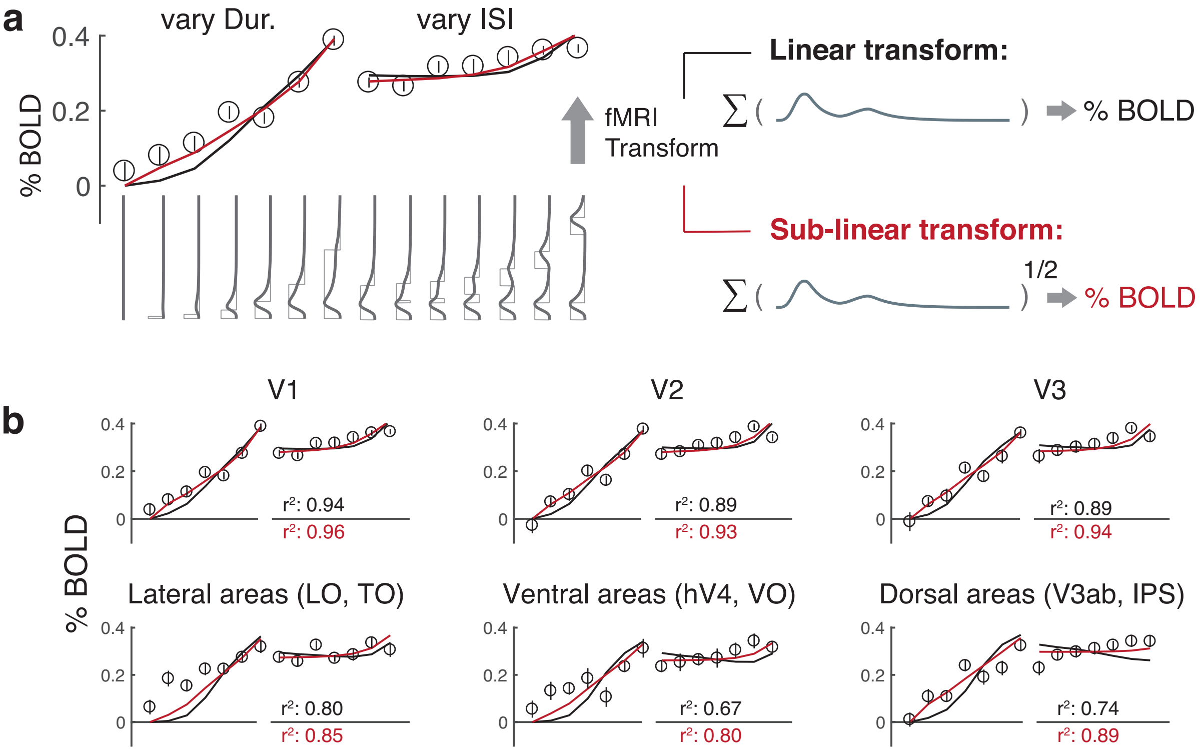


Figure 7. DCTS model fit to ECoG data predicts fMRI responses. *(A) Predicting BOLD amplitude from dCTS models fitted to ECoG data.* We predicted the fMRI responses using dCTS model parameters fitted to ECoG data. We used two types of transforms to relate the predicted ECoG time series to a percent BOLD response: 1. Linear transform. We summed the predicted ECoG time series for each temporal condition, and fit a gain factor to convert the sums to percent BOLD responses. 2. Sublinear transform. Same as linear transform, except that the predicted ECoG response was point-wise square-rooted prior to summing. Data are from V1. *(B) Predictions across ROIs.* Using parameters fitted to the ECoG experiment only, the dCTS predictions (lines) are well matched to the fMRI ß-weights (circles). The red and black curves represent the linear and sublinear transform predictions. In each ROI, the sublinear transform fits the data slightly better than the linear transform. Data from V1 are replotted from panel A.

3. DISCUSSION

3.1 Summation and adaptation in visual cortex

We report subadditive temporal summation throughout human visual cortex. Across 9 areas, responses to long stimuli were less than the linear prediction from briefer stimuli, with more pronounced sub-additivities in areas anterior to V1-V3. We captured this effect in a new temporal receptive field model, with a static non-linearity to explain the fMRI amplitude and a dynamic non-linearity to explain the ECoG time course. The dynamic implementation is more general, as it accurately predicts responses in both modalities. Nonetheless, the simpler instantiation of the model (CTS) is adequate to make highly accurate predictions for the fMRI data (cross-validated R^2 ~ 90%); an adequate model can be useful and is commonly employed in science and engineering, even when the model is known to fail for certain conditions (which all models do) (12, 33).

The two variants of the model, CTS and dCTS, account for two phenomena: first, areas accumulate information over time (summation, modeled as temporal convolution), and second, response levels reduce from prolonged or repeated exposures (adaptation, modeled with an exponent or divisive normalization). Both phenomena, and the corresponding model parameters, vary systematically across the visual hierarchy: the summation window lengthens and the effect of adaptation grows more pronounced in later compared to earlier visual areas.

3.2 Subadditivities in fMRI

We observed temporal subadditivities for fMRI and ECoG and therefore these effects cannot be solely due to hemodynamic nonlinearities. For the fMRI model fits, we assumed a linear transformation from the neural to BOLD response, as proposed previously (29, 31). A recent alternative proposal is a square root transformation (32). We compared fMRI predictions from ECoG models using linear and square root transforms, and found both fit well, with slightly better fits for the square root transform. There are numerous differences between the ECoG and fMRI experiments so we do not consider this a compelling reason to reject the linear assumption. If we do assume the square root transform as the last stage of the CTS model (conversion to fMRI), the CTS model parameters would differ, with exponents between 0.2 and 0.5, rather than 0.1 and 0.25, still consistent with significant temporal subadditivities across visual cortex. Thus, the fMRI results, as well as the ECoG results, provide strong evidence for temporal nonlinearities in the neural response.

3.3 Subadditivities in Temporal Summation

Prior literature has characterized temporal subadditivities in several ways. For example, the fMRI response to a long presentation of a reversing contrast pattern is less than the prediction from a short presentation (29); the fMRI response to contrast patterns is larger for short ISIs than long ISIs (34); the response of V1 neurons to a steady flash is not predicted by its temporal frequency tuning and decreases over time (24); the response of a neuron to a repeated stimulus is less than the response to the first stimulus (17, 25). Our model accounts for effects such as these with a small number of components – temporal summation (convolution) and a normalization that depends on response history. By formulating a quantitative, forward model, we can then ask whether a phenomenon is unexpected, requiring additional explanation, or is already predicted by the model. For example, repetition suppression and fMRI adaptation at a long time-scale (several seconds (35, 36)) might not be predicted by our model, and hence may be distinct from the short-term adaptation effects we observe.

A phenomenon as ubiquitous as subadditive temporal summation (adaptation) is likely to be a critical part of neural information processing (37). For example, adaption may serve to prioritize new information or act as a gain control (38). An interesting consequence of subadditive temporal summation is that responses to stimuli of different durations are more similar to one another than they would be if summation were linear. This may be thought of as a form of duration tolerance or timing tolerance, analogous to size and position tolerance in spatial encoding, which are increasingly prominent in higher visual areas (11).

3.4 Multiple Scales of Temporal Dynamics

Our finding that temporal windows lengthen across the visual hierarchy is consistent with prior work measuring temporal dynamics at a larger scale. For example, temporal receptive window length was studied by measuring response reliability to scrambled movie segments (39, 40): In visual cortex, responses depended on information accumulated over ~1s, whereas in anterior temporal, parietal and frontal areas the time scale ranged from ~12-36s. Similarly, in event related fMRI, the influence of prior trials was modeled with an exponential decay, with longer time constants in later areas: Boynton et al (29) reported a time constant of ~1s in V1 for contrast reversing checkerboards, and Mattar et al (21), using static face images, reported short time constants in V1 (~0.6s) and much longer constants in face areas (~5s). In macaque, the timescale of fluctuations in spike counts was longer for areas higher in the hierarchy compared to sensory areas (41).

Analyzing visual information at multiple temporal scales has benefits. First, accumulating information in the past is necessary for predicting the future, and a hierarchy of temporal windows may thus be useful for predictions over different time-scales (42). Second, signal-to-noise ratios are optimized when the temporal scale of the analysis is matched to the temporal scale of the event of interest (i.e., a “matched filter”); different visual areas extract information about different image properties, which in turn are likely to have different temporal (or spatiotemporal) distributions in natural viewing. Conversely, the time-scale of cortical areas may set the time-scale of integration for behavior. For example, words, faces, and global motion patterns are integrated over periods 5-10 times longer than textures and local motion patterns (43, 44); modeling the time-scale of cortical areas critical for these tasks may help explain these large behavioral effects.

3.5 Models of Temporal Dynamics

Several models have been proposed to account for temporal dynamics (Figure S9). For example, psychophysical temporal sensitivity (45-47) and fMRI responses in V1 (28) and extrastriate cortex (48) can be accounted for by a model with two temporal frequency channels, sustained and transient. This model also captures some features of the ECoG broadband response, but does not match the time series in detail for our 500-ms stimuli (Figure S9). For example, it does not predict a gradual decline in signal amplitude following the peak response. The dCTS model has a different form, which was motivated to capture important phenomena governing temporal dynamics, the timescale of summation and the degree of subadditivity. The model components accounting for these phenomena are grounded in canonical computations used to model visual cortex: linear filtering, exponentiation, and normalization (49-51). The two temporal channels model contains filtering and exponentiation but not normalization. A potential way to assess a specific role for normalization would be an experiment with two stimuli superimposed spatially but with different temporal frequencies: The two-channel model would predict summation, but normalization would predict subadditivity of fMRI responses or frequency tagged MEG or EEG responses. On the other hand, a model with two temporal channels may be useful for capturing differential time courses to stimuli that preferentially drive magno vs parvo pathways, or for differences in foveal vs peripheral sensitivity (28, 48); hence the two types of models are complementary.

The dCTS model we propose is input-referred (12), i.e. a computational description of the output specified in terms of the visual stimulus, rather than a model of how the dynamics arise. Hypotheses about circuit mechanisms giving rise to temporal dynamics in cortex have been proposed (52, 53); these dynamical systems models predict differences in time scales across cortical hierarchies, in agreement with empirical results, though they don’t account for the specific shape of neural temporal responses (e.g., compare Figure 3A in [(53)] to Figure S9). Another way to account for the different time scales across visual areas would be a cascade model, in which the dCTS is a canonical computation, with the output of one stage used as the input to the next stage, with the same model parameters used in each stage. Such a cascade model can account for some of the properties in later visual areas, such as more subadditive temporal summation.

3.6 Generalization and future directions

The dCTS model we fit accurately predicts responses across multiple visual field maps using two different types of measures and many stimulus temporal profiles. An important test of a model is whether it can make informative predictions for conditions it was not designed to account for. The fact that the dCTS model, fit only to ECoG data from 500-ms stimuli, predicts the fMRI responses for many different temporal patterns is an example of successful quantitative generalization. As a test of qualitative generalization to conditions that differ even further from those the model was designed for, the dCTS model predicts different time course shapes as a function of stimulus contrast, similar to multi-unit activity (MUA) observed in human visual cortex (44) (Figure 8A). One reason that our model, developed to account only for temporal patterns, generalizes to contrast is that the model is comprised of elements fundamental in sensory processing (filtering and normalization). Finally, the dCTS model predictions for temporally white noise stimuli have autocorrelation functions that decline with temporal lag, with slower declines for later visual areas, consistent with network models of macaque cortex (53).

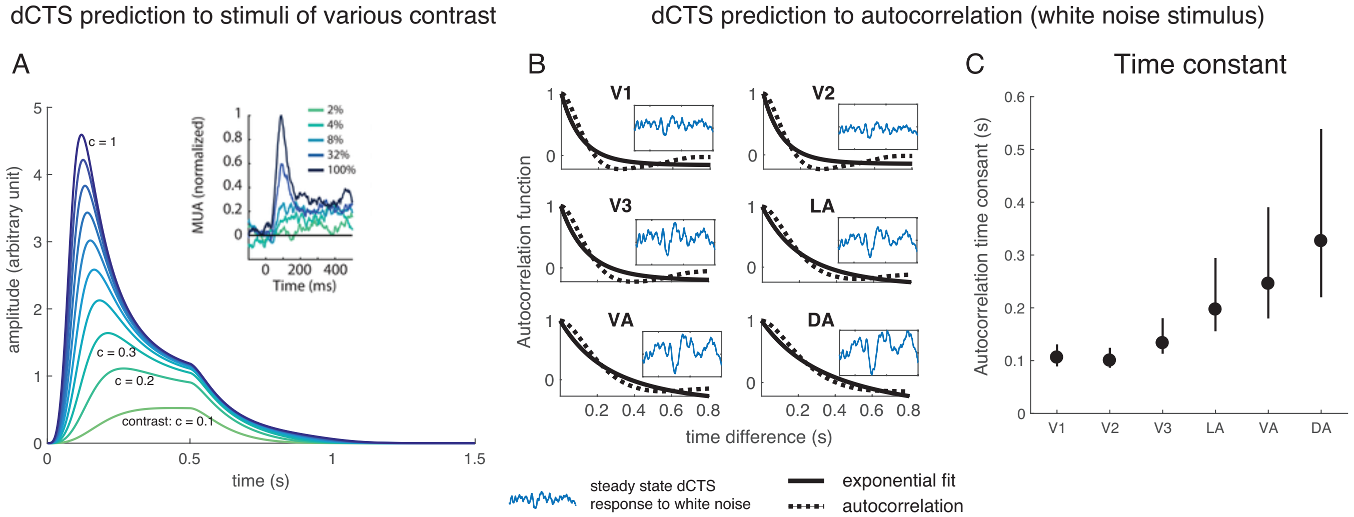


Figure 8. Applications of dCTS models. *(A) dCTS model predictions as a function of contrast.* We plot the dCTS time course for stimuli whose contrast ranged from 0.1 to 1.0. The input time course was a 500-ms boxcar with the height equal to the contrast. For high contrast stimuli, there is an initial sharp transient, followed by a lower sustained response. For lower contrast stimuli, there is little to no transient response. This is because a lower contrast is effectively the same as a higher σ and therefore less normalization. This is qualitatively similar to MUA results from human V1 (inset, from (44)). The dCTS model parameters were similar to our V1 fits (0.1, 0.1, 2, 0.1 for τ_1, τ_2, n, σ, respectively). *(B) Temporal scale of dCTS autocorrelation.* Autocorrelation functions were computed for simulated dCTS response time courses (insets) to temporal white noise inputs (uniform randomization on [0, 1]). The autocorrelation functions (dashed lines) were fit by a declining exponential (solid lines). *(C) Time constants of fitted exponentials from B.* The time constants are short in early visual areas, and longer in later areas, similar to the resting state data and network models from (41, 53). ROI labels as in Figure 6.

However, just as with spatial pRF models, it is likely that our model will fail for certain tasks or stimuli (12). For example, sustained attention to the stimulus (44), presence of a surround (43), non-separable spatiotemporal patterns (motion), and stimulus history of many seconds or more (20), can all affect the time course of the response, phenomena not captured by our current model. However, a model with these limits is still quite useful: By formulating a forward model of responses to large-field contrast stimuli during passive viewing, we provide a quantitative benchmark that can be used to assess how other factors influence response dynamics, and a platform upon which to extend the model to new stimulus or task features. An important goal for future work is to develop a space-time model that simultaneously accounts for nonlinearities in spatial (11) and temporal summation.

METHODS

4.1 fMRI procedure

Participants

Data from four experienced fMRI participants (2 males, age range 21- 48, mean age 31) were collected at the Center for Brain Imaging (CBI) at NYU. All participants had normal or corrected-to-normal visual acuity. The experimental protocol was approved by the University Committee on Activities Involving Human Subjects, and informed written consents were obtained from all participants prior to the study. Each subject participated in one 1.5-hour session for the main experiment, and an additional 1 hour session for visual field map identification and high-resolution anatomical volumes.

Visual Stimuli

Stimuli. In each trial, we used an independently generated large field (24° diameter) band-pass noise pattern (centered at 3 cycles per degree). The pattern was chosen because it was previously shown to be effective in eliciting responses in most visual areas (15). (See ref [(15)] for details on stimulus construction). In each trial of the supplementary fMRI experiment, participants viewed either an independently generated pink noise (1/f amplitude spectrum, random phase) large field image (24° diameter, 768 x 768 pixels), or a face image embedded in the pink noise. Stimulus generation, presentation and response recording were coded using Psychophysics Toolbox (54, 55) and vistadisp (https://github.com/vistalab/vistadisp). We used a MacBook Air computer to control stimulus presentation and record responses from the participants (button presses) during the experiment.

Display. Stimuli were displayed via an LCD projector (Eiki LC_XG250; resolution: 1024 x 768 pixels; refresh rate: 60 Hz) onto a back-projection screen in the bore of the magnet. Participants, at a viewing distance of ~58 cm, viewed the screen (field of view, horizontal: ~32°, vertical: ~24°) through an angled mirror. The images were confined to a circular region with a radius of 12°. The display was calibrated and gamma corrected using a linearized lookup table.

Fixation task. To stabilize attention level across scans and across subjects during the main experiment, all participants were instructed to do a one-back digit task at the center of fixation throughout the experiment. The digit (0.24° x 0.24°) was presented at the center of a neutral gray disk (0.47° diameter). Within a scan, each digit (randomly selected from 0 to 9) was on for 0.5 second, off for 0.167 second before the next digit appeared at the same location. Participants were asked to press a button when a digit repeated. Digit repetition occurred around 2-3%, with no more than two identical digits being presented successively. To reduce visual adaptation, all digits alternated between black and white, and on average participants pressed a button every 30 seconds. During the retinotopy task, the fixation alternated pseudo-randomly between red and green (switches on average every 3s), and the subject pressed a button to indicate color changes.

Experimental Design

We used a randomized event-related experimental design to prevent subjects from anticipating the stimulus conditions. An event is a stimulus presented according to one of thirteen distinct time courses (< 800 ms in total), either a single pulse with variable duration or a double pulse with fixed duration and variable inter-stimulus interval (ISI). Durations and ISIs were multiples of the monitor dwell time (1/60 s). Each pulse in the double-pulse stimuli lasted 134ms. The 0-ms stimulus was a blank (zero-contrast, mean luminance, and hence identical to the preceding and subsequent blank screen between stimulus events). Each participant completed seven scans, and within a scan, each temporal event repeated 4 times. A temporal event started with the onset of a pattern image, and the inter-trial interval (stimulus plus subsequent blank) was always 4.5 seconds. For experiments with two pulses, the two noise patterns were identical. The design was identical for the supplementary fMRI experiment, except that each time course repeated three times per scan, and each participant completed 12 scans.

MRI Data Acquisition

All fMRI data were acquired at NYU Center for Brain Imaging (CBI) using a Siemens Allegra 3T head-only scanner with a Nova Medical phased array, 8-channel receive surface coil (NMSC072). For each participant, we collected functional images (1500 ms TR, 30 ms TE, and 72-degree flip angle). Voxels were 2.5mm^3 isotopic, with 24 slices. The slice prescription covered most of the occipital lobe, and the posterior part of both the temporal and parietal lobes. Images were corrected for B0 field inhomogeneity using CBI algorithms during offline image reconstruction.

In a separate session, we acquired two to three T1-weighted whole brain anatomical scans (MPRAGE sequence; 1mm^3). Additionally, a T1-weighted "inplane" image was collected with the same slice prescription as the functional scans to aid alignment of the functional images to the high-resolution T1-weighted anatomical images. This scan had an inplane resolution of 1.25 x 1.25 mm and a slice thickness of 2.5 mm.

Data Preprocessing and Analysis

Data preprocessing. We co-registered and segmented the T1-weighted whole brain anatomical images into gray and white matter voxels using FreeSurfer's auto-segmentation algorithm (surfer.nmr.mgh.havard.edu). Using custom software, vistasoft (https://github.com/vistalab/vistasoft), the functional data were slice-time corrected by resampling the time series in each slice to the center of each 1.5s volume. Data were then motion-corrected by co-registering all volumes of all 7 scans to the first volume of the first scan. The first 8 frames (12 seconds) of each scan were discarded for analysis to allow longitudinal magnetization and stabilized hemodynamic response.

GLM analysis. We used a variant of the GLM procedure—GLM denoise (16), a technique that improves signal-to-noise ratios by entering noise regressors into the GLM analysis. Noise regressors were selected by performing principle component analysis on voxels whose activities were unrelated to the task. The optimal number of noise regressors was selected based on cross-validation R^2 improvement. The input to GLM denoise was the pre-processed EPI data and a design matrix for each scan (13 distinct temporal profiles x number of time points per scan), and the output was ß-weights for each temporal profile for each voxel, bootstrapped 100 times across scans. For analysis, we normalized all

13 ß-weights per voxel by the vector length and selected a subset of voxels (see *Voxel selection*). We then averaged the ß-weights for a given temporal condition from the first bootstrap across voxels within each ROI and across all subjects to get a mean; this gives one estimate of the mean response per ROI for a given condition. This was repeated for each condition, and then repeated for each of the 100 bootstraps, yielding a matrix of 100 x 13 for each ROI (bootstraps by temporal condition).

ROI identification. We fitted a linear pRF model (10) to each subject's retinotopy data (average of two scans). We made an initial guess of ROI locations by first projecting the maximum likelihood probabilistic atlas from Wang et al (56) onto the cortical surface. Then we visualized eccentricity and polar angle maps derived from the pRF model fits and modified ROI boundaries based on visual inspection. For each participant, we defined nine bilateral ROIs (V1, V2, V3, hV4, VO-1/2, LO-1/2, TO-1/2, IPS-0/1).

Voxel selection. All analyses were restricted to voxels that satisfy the following three criteria. First voxels be must located within 2-10° (eccentricity) based on the pRF model. Second, voxels must have positive bootstrapped ß-weights (averaged across bootstraps) for all non-blank temporal conditions. Third, voxels must have > 3% GLM R^2. Voxels that satisfy all criteria were pooled across subjects, and the group average (bootstrapped) ß-weights were analyzed and plotted.

4.2 ECoG Procedure

We re-analyzed previously published ECoG data (22).

Preprocessing. The data were pre-processed as in the original paper. In brief, electrodes that had large artifacts or epileptic activity, as identified by the neurologist, were excluded from analysis. From the remaining electrodes, we re-referenced the time series to the common average, and then down sampled the data from the recorded frequency 3052/1528 Hz (Subject 1/Subject 2) to 1,000 Hz.

Trial structure. At the beginning of each 1-second trial, a large field (22°) noise image was randomly selected from one of 8 image classes. Several of these image classes were chosen for studying gamma oscillations in the original paper, which was not the purpose of this study. For this study, we analyzed data from 3 of the 8 image classes, those that were most similar to the noise stimuli in the fMRI experiment: white, pink, and brown noise (amplitude spectra proportional to $1/f^0$, $1/f^1$, $1/f^2$). Each image was presented for 500ms followed by a 500ms blank. We analyzed data in 1200 ms epochs, beginning 200 ms prior to stimulus onset and ending 500 ms after stimulus offset.

Broadband envelope. We computed the time varying broadband envelope in several steps, as follows. First, we band-pass filtered the time series in 12 adjacent 10-Hz bins from 80 Hz to 200 Hz (80-90 Hz, 90-100 Hz, etc) using a Butterworth filter (passband ripples < 3 dB, stopband attenuation 60 dB). For each filtered time series, we computed the envelope as the magnitude of the analytic function (Hilbert transform). We then normalized the envelope of each bin by dividing by the variance, so that each envelope had a variance of 1. We normalized the variance to compensate for the fact that the power in field potentials declines with frequency. We then summed the 12 envelopes to derive a single, time-varying broadband envelope. Finally, we defined the baseline as the average value of the envelope in the 200 ms prior to stimulus onset and subtracted this baseline value from the time series at all points.

Broadband units. Because of the normalization of the 12 bands, the broadband power is the sum of 12 z-scores. So, for example, a stimulus-driven power increase of 12 means an average increase in power of 1-zcore per each of the 10-Hz frequency bands.

Electrode selection. We selected all electrodes located in identifiable visual areas based on separate retinotopy scans, and whose stimulus-triggered broadband response, averaged across trials, reached at least a power of 3 (see *broadband units*, above).

4.3 Temporal pRF Models

We used three variants of a temporal pRF model, one linear and two non-linear, to predict neuronal summation measured using fMRI and ECoG. All model forms take the time course of a spatially uniform contrast pattern as input (T_{input}), and produce a predicted neuronal response time course as output. To predict the fMRI data (BOLD), we summed the predicted time course within a trial (< 1 s) to yield one number per temporal condition. These numbers were compared to the fMRI ß-weights for model fitting (see below). For ECoG data, the predicted time course was compared directly to the broadband time series for model fitting.

Models

Linear model. The linear model prediction is computed by convolving a neuronal impulse response function (IRF) with the stimulus time course (T_{input}), and scaling by a gain factor (*g*)

$$R_{linear} = g\left[IRF \ * T_{input}\right]$$

The time course is then summed for the fMRI predictions (plus an error term, *e*):

$$BOLD_{linear} = \sum g\left[IRF \ * T_{input}\right] + e$$

For the IRF, we assumed a gamma function, parameterized by τ_1, of the form,

$$IRF = t * \exp(-t/\tau_1)$$

Because the IRF was assumed to have unit area, the specific shape of the IRF has no effect on the predictions, and the prediction reduces to:

$$BOLD_{linear} = g\left[\sum T_{input}\right] + e$$

and the only value solved for is the gain factor. We did not fit the linear model to ECoG data because the linearly predicted time courses clearly differ from broadband traces.

Compressive summation model (CTS). To compute the CTS predicted neuronal response, we first computed the linear response by convolving an IRF (gamma function with variable time to peak τ_1) with an input stimulus time course. Then an exponent ε is applied point-wise to the predicted linear output.

$$R_{CTS} = g\left[IRF(\tau_1) \ * T_{input}\right]^{\varepsilon}$$

To fit the CTS model to the fMRI data, we again summed the predicted response time series:

$$BOLD_{CTS} = \sum g\left[IRF(\tau_1) \ * T_{input}\right]^{\varepsilon} + e$$

and solved for τ_1, ε, and *g*. We did not fit the CTS model to ECoG broadband traces because CTS-predicted neuronal response differs from the measurements qualitatively.

Dynamic compressive temporal summation (dCTS). This variant of the CTS model implemented the compressive nonlinearity with a divisive normalization rather than a compressive power law. The numerator contains the linear neuronal response (same computation as the linear part in CTS). The denominator is the sum of two terms, a semi-saturation constant (σ) and an exponentially filtered (low-pass) linear response. The rate of the exponential decay is determined by a parameter τ_2. All three terms (one in the numerator, two in the denominator), are raised to the power *n*, assumed to be greater than 1.

$$R_{dCTS} = \frac{[R_{linear}(\tau_1)]^n}{\sigma^n + [R_{linear}(\tau_1) * \exp(\tau_2)]^n}$$

We fit the 4 parameters as well as a gain factor, *g*, to the ECoG broadband time series. To predict the fMRI response from the dCTS model (Figure 7), we used the parameters fitted from ECoG data for each ROI, generated a neuronal time course for each of the 13 distinct temporal profiles from the fMRI experiment. Then we either summed each predicted time course (linear assumption) or point-wise square-rooted the time course and then summed, and finally scaled the sum by a gain factor.

Parameter estimation

CTS model for fMRI. Models were fit in two steps, one to obtain seed parameters, and one to fit parameters.

In the first step, we obtain seed values for τ_1 and ε for each ROI. To do so, we generated 1000 seeds by randomly selecting τ_1 from [0.01 1] and ε from [0, 1]. These were then used to make 1000 sets of model predictions for the 13 temporal stimuli. For each ROI, the 1000 sets of model predictions were compared to the 13 ß-weights. Using linear regression, we then derived the gain factor, *g*, and the variance explained for each of the 1000 sets of predictions. The model parameters τ_1, ε, *g* were averaged from all models with variance explained greater than 95%. This gave us seeds for the three parameters for each ROI.

We then did a search fit using Matlab's *fminsearch*, 100 times per ROI, using the 100 sets of bootstrapped ß-weights, and the seeds as derived above. The search finds the parameters which minimize the squared error between predicted and measured ß-weights. This gave us 100 estimates of each model parameter for each ROI, which we summarized by the median and 50% confidence interval.

Linear model for fMRI. The linear model does not require a search or seeds. Instead, we fit the 100 bootstrapped data sets per ROI by linear regression, giving us 100 estimates of the gain factor, *g*, per ROI.

dCTS model for ECoG. We again used a two-stage approach to fitting the dCTS model, first to obtain seeds and then to estimate parameters. For each ROI, we averaged the broadband envelope across electrodes and trials, yielding one time course per ROI. We then generated 1000 model predictions by randomly selecting each parameter: τ_1 from [0.01, 1], τ_2 from [0.01, 1], *n* from [0.5, 5], and σ from [0.01, 0.5]. Using linear regression on the ECoG data, we derived the gain factor, *g*, and the variance explained for each of the 1000 predicted time series. For each ROI, the sets of parameters that generated reasonably accurate model predictions (> 80% variance explained) were averaged and served as the seed for the search fit.

For the search fit, we did 100 bootstraps per ROI over the electrodes in that ROI. For each of the 100 bootstrapped time courses per ROI, we used *fminsearch* to find the parameters that minimized the squared error between the predicted and observed time series. In addition to the four parameters above, we included a nuisance shift parameter, which delays the onset of the response. In principle, this delay is important, since the time at which the signal from the stimulus reaches cortex is delayed, and the delay varies across visual field maps, and could be as high as 50-150ms. However, the impulse response function includes a slow ramp, and the broadband envelope extraction contains a small amount of blur. Hence in practice, the shifts were quite small (< 10 ms), and not informative about the latency of neuronal response.

Model accuracy

fMRI experiment. For the fMRI experiment, we compared model accuracy of the CTS and the linear model. Because the models have different numbers of free parameters, it is important to obtain an unbiased estimated of model accuracy, which we did by leave-one-out cross validation. For each ROI, and for each of the 100 bootstrapped sets of ß-weights, we fit 13 linear models and 13 CTS models by leaving out each of the 13 temporal stimuli. For each bootstrap, we thus obtain 13 left-out predictions, which were compared to the 13 ß-weights by coefficient of determination, R^2:

$$R^2 = 100 \times \left[1 - \frac{\sum (MODEL - DATA)^2}{\sum DATA^2} \right]$$

This yielded 100 R^2's per ROI, and we summarized model accuracy as the median and 50% confidence interval derived from these values.

For the dCTS model fit to the ECoG data, there was only one temporal condition, and no model comparison, so we did not cross-validate the model fits. Instead, we summarized model accuracy as the variance explained, r^2, the square of the Pearson-correlation coefficient *r*.

Note that the coefficient of determination, R^2, is bounded by [$-\infty$, 1], as the residuals between model and data can be larger than the data. In contrast, r^2 is bounded by [0, 1].

Public Data Sets and Software Code

To ensure that our computational methods are reproducible, all data and all software will be made publicly available via an open science framework site, https://osf.io/v843t/. The software repository will include scripts of the form *trf_MakeFigure2* to reproduce figure 2, etc., as in prior publications (57).

Acknowledgements

We thank Dora Hermes for helpful discussion and for helping us analyze ECoG data from prior work. We also thank Josef Parvizi and the Stanford Human Intracranial Cognitive Electrophysiology Program for helping us with ECoG data acquisition for a prior paper, which was re-analyzed for this paper. We thank David Heeger, Brian Wandell, and Mike Landy for comments on an earlier draft of this manuscript. We also thank Bosco Tjan, David Heeger, XJ Wang, Denis Pelli, Rachel Denison, and Geoff Aguirre for helpful discussions and feedback as we developed our models and analyses. The research was supported by NIH grants R00-EY022116 and R01-MH111417 (J.W.)

References

1. Rock I (1984) *Perception* (Scientific American Library : Distributed by W.H. Freeman, New York) pp x, 243 p.
2. Pelli DG & Tillman KA (2008) The uncrowded window of object recognition. *Nat Neurosci* 11(10):1129-1135.
3. Motoyoshi I & Nishida S (2001) Temporal resolution of orientation-based texture segregation. *Vision research* 41(16):2089-2105.
4. Holcombe AO & Judson J (2007) Visual binding of English and Chinese word parts is limited to low temporal frequencies. *Perception* 36(1):49-74.
5. Raz N, Dotan S, Chokron S, Ben-Hur T, & Levin N (2012) Demyelination affects temporal aspects of perception: an optic neuritis study. *Ann Neurol* 71(4):531-538.
6. Bonneh YS, Sagi D, & Polat U (2007) Spatial and temporal crowding in amblyopia. *Vision research* 47(14):1950-1962.
7. Farmer ME & Klein RM (1995) The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychon Bull Rev* 2(4):460-493.
8. Riesenhuber M & Poggio T (1999) Hierarchical models of object recognition in cortex. *Nature Neuroscience* 2(11):1019-1025.
9. Maunsell JH & Newsome WT (1987) Visual processing in monkey extrastriate cortex. *Annual Reviews of Neuroscience* 10:363-401.
10. Dumoulin SO & Wandell BA (2008) Population receptive field estimates in human visual cortex. *Neuroimage* 39(2):647-660.
11. Kay KN, Winawer J, Mezer A, & Wandell BA (2013) Compressive spatial summation in human visual cortex. *Journal of neurophysiology* 110(2):481-494.
12. Wandell BA & Winawer J (2015) Computational neuroimaging and population receptive fields. *Trends Cogn Sci* 19(6):349-357.
13. Britten KH & Heuer HW (1999) Spatial summation in the receptive fields of MT neurons. *J Neurosci* 19(12):5074-5084.
14. Rolls ET & Tovee MJ (1995) The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. *Exp Brain Res* 103(3):409-420.
15. Kay KN, Winawer J, Rokem A, Mezer A, & Wandell BA (2013) A two-stage cascade model of BOLD responses in human visual cortex. *PLoS Comput Biol* 9(5):e1003079.
16. Kay KN, Rokem A, Winawer J, Dougherty RF, & Wandell BA (2013) GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Frontiers in neuroscience* 7:247.
17. Priebe NJ, Churchland MM, & Lisberger SG (2002) Constraints on the source of short-term motion adaptation in macaque area MT. I. the role of input and intrinsic mechanisms. *Journal of neurophysiology* 88(1):354-369.
18. Kohn A (2007) Visual adaptation: physiology, mechanisms, and functional benefits. *Journal of neurophysiology* 97(5):3155-3164.

19. Winawer J*, et al.* (2013) Asynchronous broadband signals are the principal source of the BOLD response in human visual cortex. *Curr Biol* 23(13):1145-1153.
20. Weiner KS, Sayres R, Vinberg J, & Grill-Spector K (2010) fMRI-adaptation and category selectivity in human ventral temporal cortex: regional differences across time scales. *Journal of neurophysiology* 103(6):3349-3365.
21. Mattar MG, Kahn DA, Thompson-Schill SL, & Aguirre GK (2016) Varying timescales of stimulus integration unite neural adaptation and prototype formation. *Current Biology* 26(13):1669-1676.
22. Hermes D, Miller KJ, Wandell BA, & Winawer J (2015) Stimulus dependence of gamma oscillations in human visual cortex. *Cerebral cortex* 25(9):2951-2959.
23. Mukamel R*, et al.* (2005) Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. *Science* 309(5736):951-954.
24. Tolhurst DJ, Walker NS, Thompson ID, & Dean AF (1980) Non-linearities of temporal summation in neurones in area 17 of the cat. *Exp Brain Res* 38(4):431-435.
25. Motter BC (2006) Modulation of transient and sustained response components of V4 neurons by temporal crowding in flashed stimulus sequences. *J Neurosci* 26(38):9683-9694.
26. Burns SP, Xing D, & Shapley RM (2010) Comparisons of the dynamics of local field potential and multiunit activity signals in macaque visual cortex. *J Neurosci* 30(41):13739-13749.
27. Heeger DJ (1993) Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of neurophysiology* 70(5):1885-1898.
28. Horiguchi H, Nakadomari S, Misaki M, & Wandell BA (2009) Two temporal channels in human V1 identified using fMRI. *Neuroimage* 47(1):273-280.
29. Boynton GM, Engel SA, Glover GH, & Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16(13):4207-4221.
30. Dale AM & Buckner RL (1997) Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp* 5(5):329-340.
31. Boynton GM, Engel SA, & Heeger DJ (2012) Linear systems analysis of the fMRI signal. *Neuroimage* 62(2):975-984.
32. Bao P, Purington CJ, & Tjan BS (2015) Using an achiasmic human visual system to quantify the relationship between the fMRI BOLD signal and neural response. *Elife* 4.
33. Box GEP & Draper NR (1987) *Empirical model-building and response surfaces* (Wiley, New York) pp xiv, 669 p.
34. Heckman GM*, et al.* (2007) Nonlinearities in rapid event-related fMRI explained by stimulus scaling. *Neuroimage* 34(2):651-660.
35. Grill-Spector K & Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107(1-3):293-321.
36. Miller EK, Gochin PM, & Gross CG (1991) Habituation-like decrease in the responses of neurons in inferior temporal cortex of the macaque. *Vis Neurosci* 7(4):357-362.
37. Webster MA (2015) Visual Adaptation. *Annu Rev Vis Sci* 1:547-567.
38. Solomon SG & Kohn A (2014) Moving sensory adaptation beyond suppressive effects in single neurons. *Curr Biol* 24(20):R1012-1022.
39. Hasson U, Yang E, Vallines I, Heeger DJ, & Rubin N (2008) A hierarchy of temporal receptive windows in human cortex. *J Neurosci* 28(10):2539-2550.
40. Honey CJ*, et al.* (2012) Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* 76(2):423-434.
41. Murray JD*, et al.* (2014) A hierarchy of intrinsic timescales across primate cortex. *Nat Neurosci* 17(12):1661-1663.
42. Heeger DJ (2017) Theory of cortical function. *Proc Natl Acad Sci U S A*.
43. Bair W, Cavanaugh JR, & Movshon JA (2003) Time course and time-distance relationships for surround suppression in macaque V1 neurons. *J Neurosci* 23(20):7690-7701.
44. Self MW*, et al.* (2016) The Effects of Context and Attention on Spiking Activity in Human Early Visual Cortex. *PLoS Biol* 14(3):e1002420.

45. Watson AB (1986) Temporal Sensitivity. *Handbook of perception and human performance*, eds Boff KR, Kaufman L, & Thomas JP (Wiley, New York), pp 6.1-6.43.
46. Watson AB & Robson JG (1981) Discrimination at threshold: labelled detectors in human vision. *Vision research* 21(7):1115-1122.
47. Hess RF & Plant GT (1985) Temporal frequency discrimination in human vision: evidence for an additional mechanism in the low spatial and high temporal frequency region. *Vision research* 25(10):1493-1500.
48. Stigliani A, Jeska B, & Grill-Spector K (2017) An encoding model of temporal processing in human visual cortex. *bioRxiv*:108985.
49. Albrecht DG & Geisler WS (1991) Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Vis Neurosci* 7(6):531-546.
50. Heeger DJ (1992) Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9(2):181-197.
51. Carandini M & Heeger DJ (2012) Normalization as a canonical neural computation. *Nature reviews. Neuroscience* 13(1):51-62.
52. Kiebel SJ, Daunizeau J, & Friston KJ (2008) A hierarchy of time-scales and the brain. *PLoS Comput Biol* 4(11):e1000209.
53. Chaudhuri R, Knoblauch K, Gariel MA, Kennedy H, & Wang XJ (2015) A Large-Scale Circuit Mechanism for Hierarchical Dynamical Processing in the Primate Cortex. *Neuron* 88(2):419-431.
54. Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10(4):433-436.
55. Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10(4):437-442.
56. Wang L, Mruczek RE, Arcaro MJ, & Kastner S (2015) Probabilistic Maps of Visual Topography in Human Cortex. *Cereb Cortex* 25(10):3911-3931.
57. Winawer J & Parvizi J (2016) Linking Electrical Stimulation of Human Primary Visual Cortex, Size of Affected Cortical Area, Neuronal Responses, and Subjective Experience. *Neuron* 92(6):1213-1219.

Supplementary Material

Figure S1. Individual subject fMRI results.

Figure S2. FMRI data and model fits from a second experiment.

Figure S3. CTS model fits by eccentricity.

Figure S4. Individual electrode responses.

Figure S5. ECoG responses to stimulus offset.

Figure S6. Relationship between broadband envelope and neural time series.

Figure S7. ECoG responses from ECoG subject 2.

Figure S8. CTS and dCTS parameter recovery.

Figure S1. Related to Figure 4.

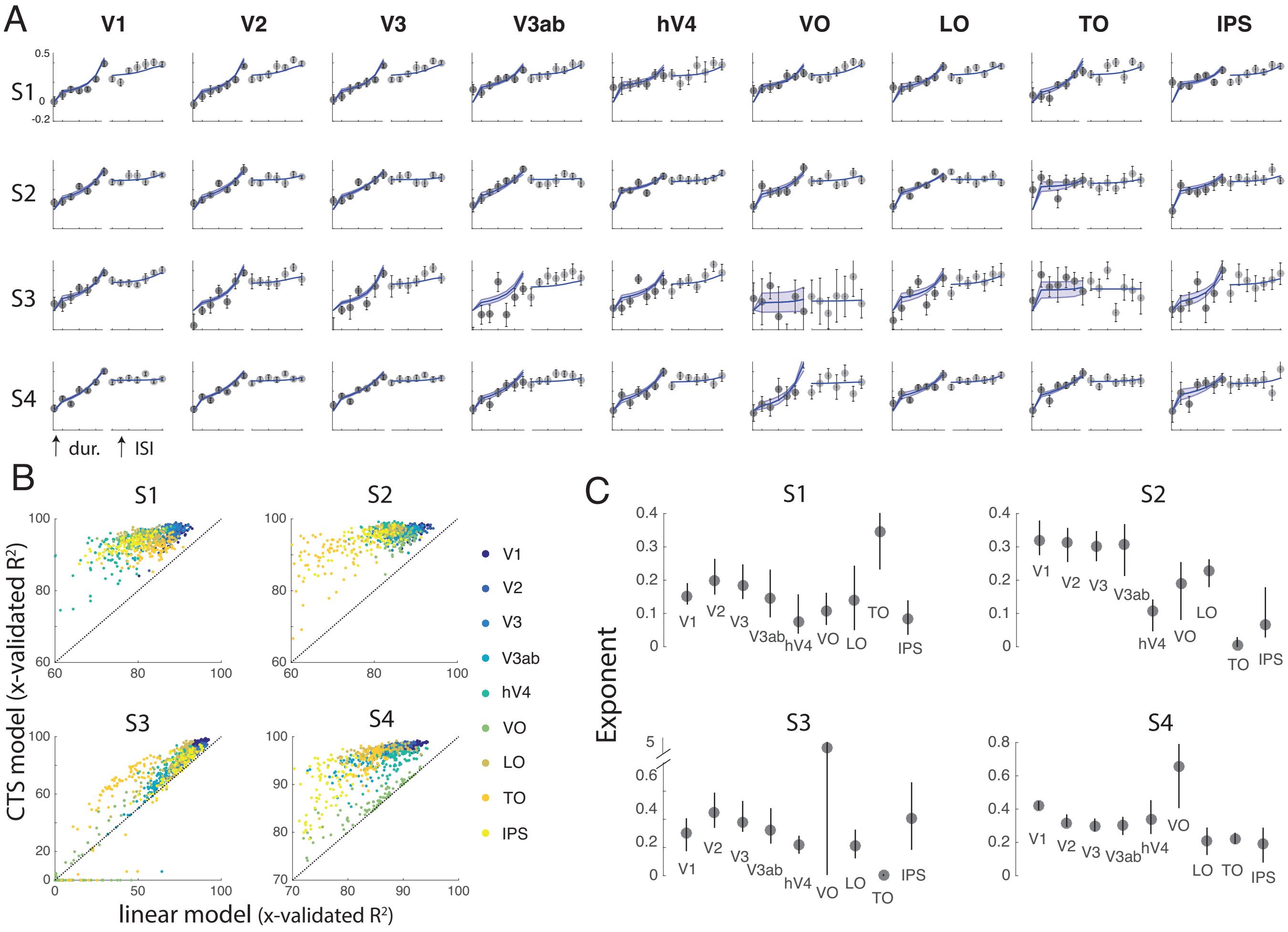


Figure S1. Individual subject fMRI results. *(*A) Four subject's individual ROI data and CTS model fits are presented, one subject per row. Plotting conventions as in Figure 4. (B) The cross-validated accuracy of the CTS models (y-axis) and linear models (x-axis) are plotted for each subject (separate subplots) and each ROI (different colors). Each dot represents the cross-validated R^2 for one bootstrap. Dots above the line indicate higher accuracy for the CTS model than the linear model. (C) The CTS exponent, ε is plotted for each ROI and each subject. Exponents less than 1 indicate subadditive temporal summation.

Figure S2. Related to Figure 4.

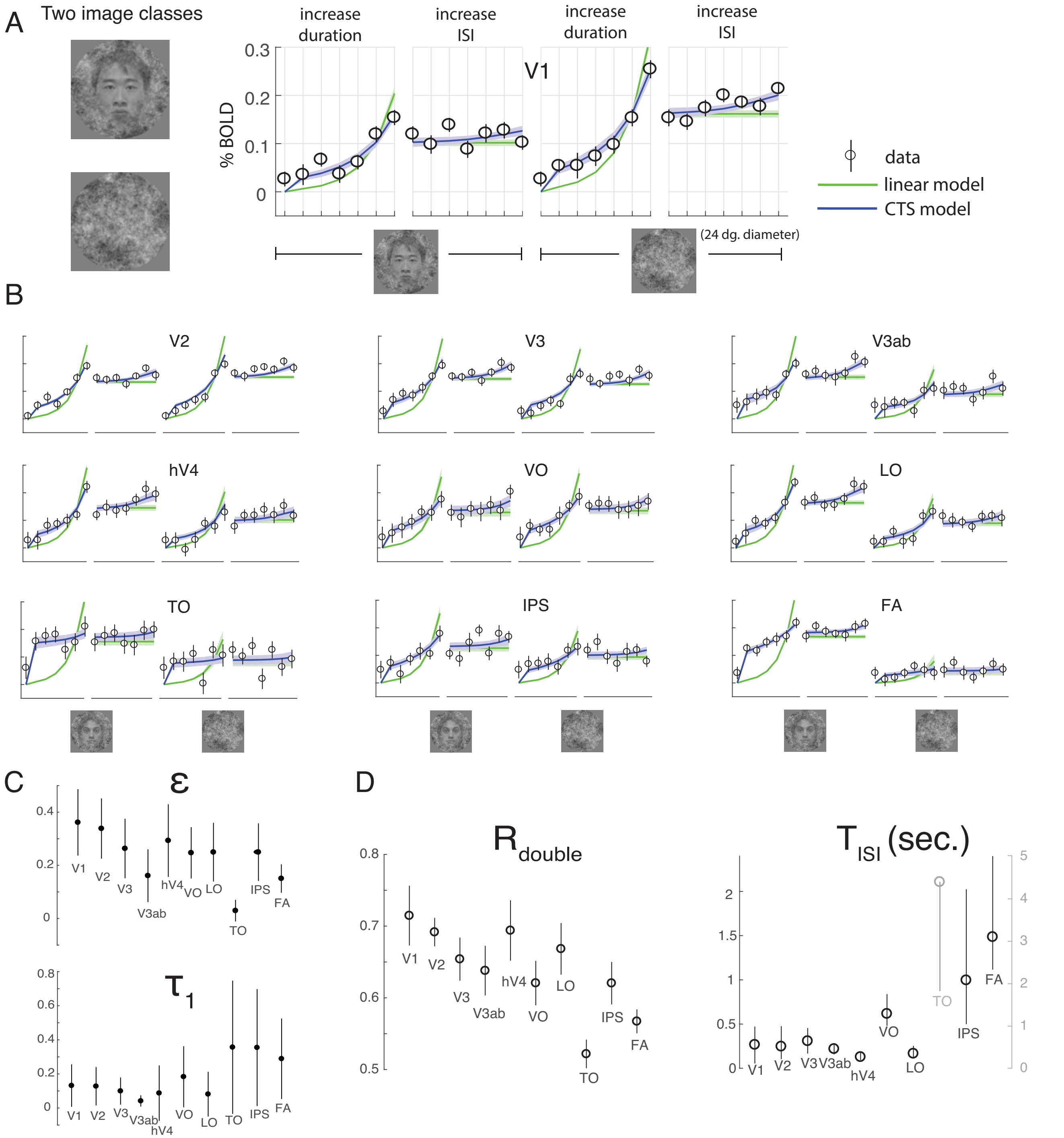


Figure S2. FMRI data and model fits from a second experiment. Four subjects participated in the experiment (two of which participated in the experiment described in the main text). The experiment was the same as the experiment in the main text, except that two different image classes were used: pink noise, and a face embedded in pink noise. Temporal conditions are identical to those in Figure 4. Models were fit simultaneously to both image classes, with a separate gain factor for each class. In general, the response amplitudes are lower than the main experiment due to stimulus selectivity, and the responses are noisier due to fewer trials per condition. In addition to the 9 ROIs in the main text, we also plot data from face areas (union of FFA and OFA). *(A&B)* We fitted both the linear model (green) and the CTS model (purple) to the group averaged data (50% CI from bootstrapping across scans). As in the main fMRI experiment, the CTS model fits the data in each ROI better than the linear model. *(C)* The exponent of the CTS model is below 1, and tends to be lower in extrastriate areas compared to V1. *(D)* The derived metrics, R_{double}, T_{ISI}, show similar patterns as in the main experiment: decreased R_{double} and increased T_{ISI} in higher visual areas.

Figure S3. Related to Figure 4.

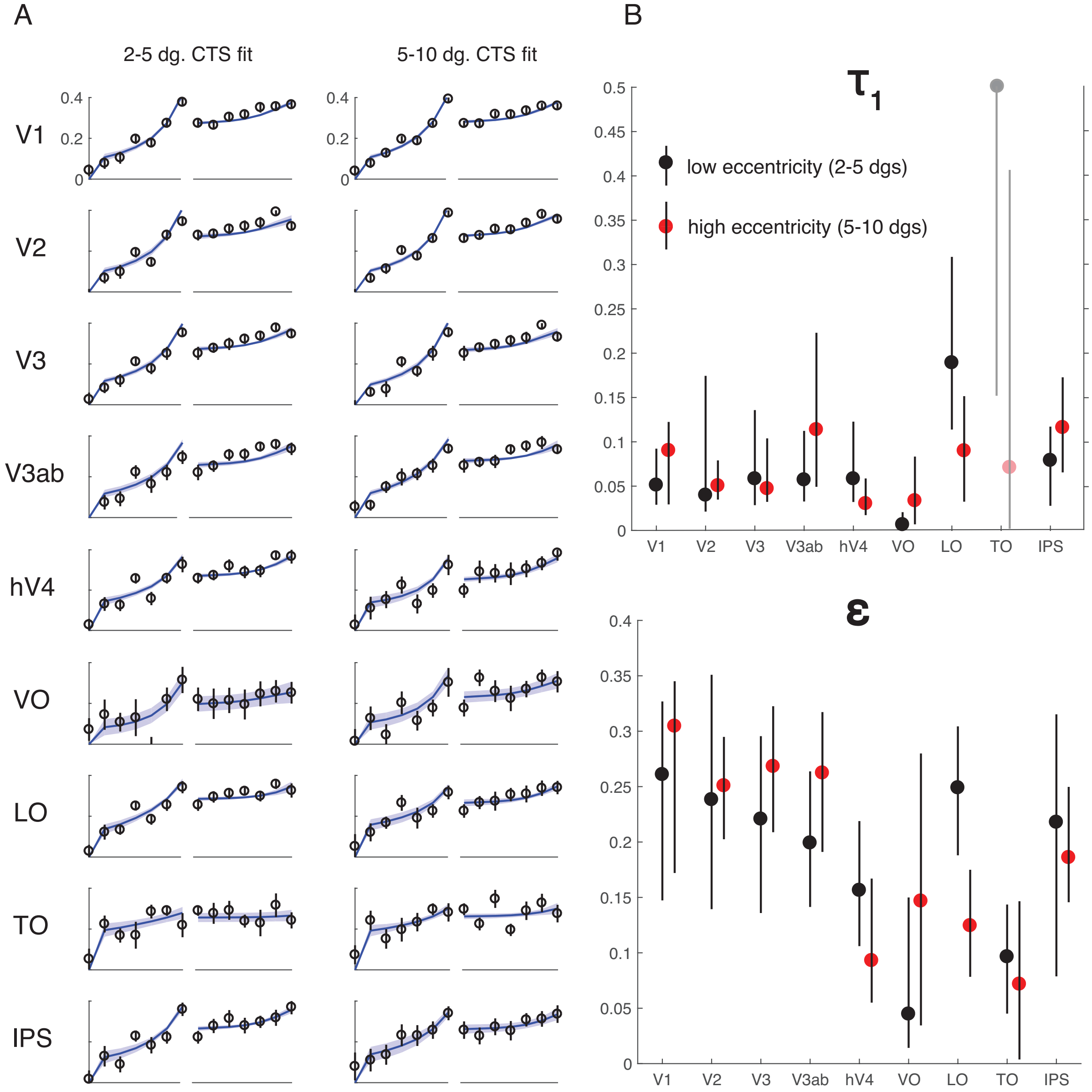


Figure S3. CTS model fits by eccentricity. Data from the main fMRI experiment are replotted separating each ROI into 2 eccentricity bins. *(A)* The left panels are the data and CTS model fits restricted to voxels with population receptive field centers within 2 - 5°. The right panels are data and CTS model fits restricted to voxels with 5 - 10° eccentricity. (*B)* The CTS model parameters do not differ systematically between the two eccentricity ranges.

Figure S4. Related to Figure 6.

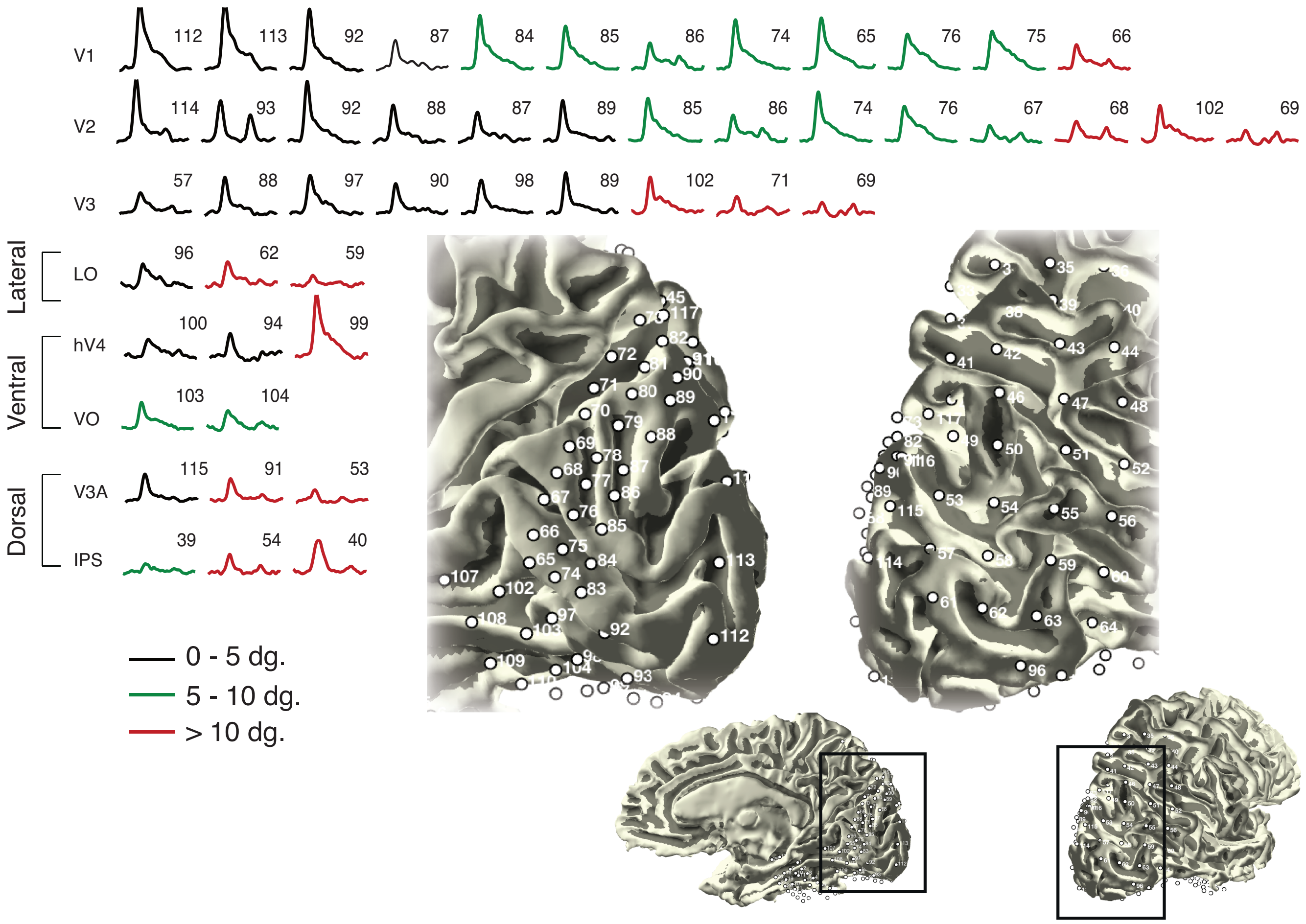
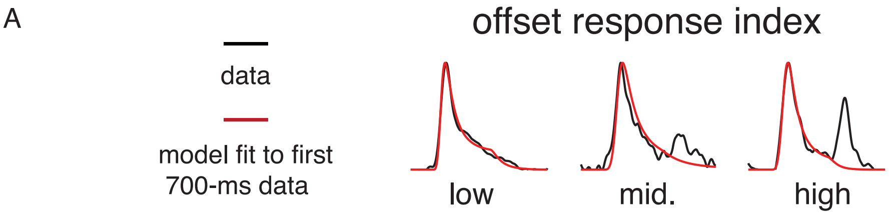


Figure S4. Individual electrode responses. The plots show the ECoG broadband time course in individual electrodes from ECoG subject S1, averaged across 90 trials (30 repeats each of three stimulus types). Each row shows electrodes from one ROI. Some electrodes (e.g., 74) are in two rows, since the electrode was near an ROI boundary. The plots are color coded by eccentricity bin (0-5º, 5-10, >10º). The pRF location was based on a separate ECoG pRF data set published previously (1). The two mesh images show a magnified view of S1's right occipital lobe, exposing the medial surface (left) and lateral surface (right). Insets show the zoomed-out view of the cortical mesh.

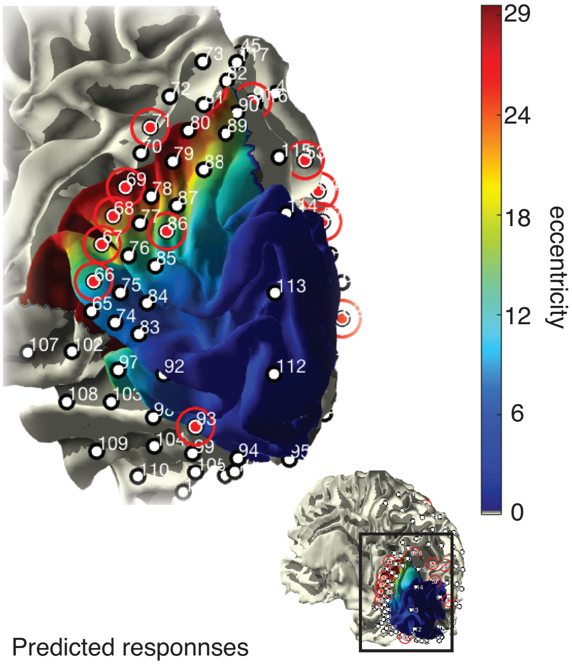
Figure S5. Related to Figure 6.



B

Offset index
(unit : %maximum response)

	0-5 dg.	5-10 dg.	>10 dg.
V1	0.0004	-0.0010	**0.0516**
V2	0.0139	0.0227	**0.0967**
V3	0.0126	--	**0.0874**
V3A	0.0269	--	**0.0536**
hV4	-0.0095	--	**0.0002**
VO	--	-0.0260	--
LO	0.0143	--	**0.0654**
IPS	--	**0.0710**	0.0483



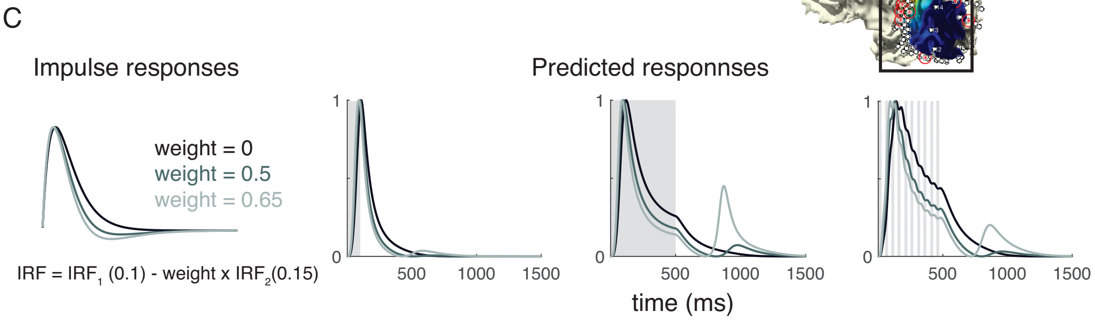


Figure S5. ECoG responses to stimulus offset. Some electrodes show a positive broadband response at stimulus offset. *(A)* We derived an offset index for each electrode as follows. We first normalized the time series by dividing each point by the peak response, so that the time series maximum was 1. We then fit the dCTS model to the response from -200ms to 500ms (200ms pre-stimulus, and 500ms stimulus duration). We then extended the model prediction for the subsequent 500ms post-stimulus. The metric is the mean of the difference between model prediction and data for the 500ms following stimulus offset. The three plots are example electrodes with low, medium, and high offset indices. Red is the model fit and black is the data. *(B)* The table shows the average offset index binned by ROI and by pRF eccentricity. The highest number in each row is shown in red. In most rows, the highest offset index is for the most peripheral electrodes. The mesh image on the right shows S1's right occipital cortex, viewed from behind. The color overlay shows an eccentricity map from fovea (blue) to periphery (red), derived from a V1-V3 atlas template (2, 3). The ECoG electrodes with high offset indices are circled in red. In most cases (with a few exceptions) these electrodes have anterior locations with high eccentricity. *(C)* Although the dCTS model in the main text does not predict an offset response, a slight variant of the model with a biphasic rather than monophasic impulse response function (IRF) does predict the offset response. We computed biphasic IRFs as the difference of 2 gamma functions, with the negative lobe having a time constant 1.5 times longer than the positive lobe, shown on the left. We used three IRFs for simulations with different weights on the negative lobe: weight 0 (monophasic), 0.5 (biphasic) and 0.65 (biphasic). We then show the model outputs for three stimuli: a 100-ms stimulus pulse (left), a 500-ms pulse (middle), and a train of 25-ms pulses (right). When the weight on the negative lobe of the IRF is high (0.65), there are significant offset responses predicted for the 500-ms pulse and the train of 25-ms pulses. For simulations, we assumed the following dCTS parameters: τ_1=0.1s; τ_2=0.1s; *n*=2; σ=0.1.

Figure S6. Related to Figure 6.

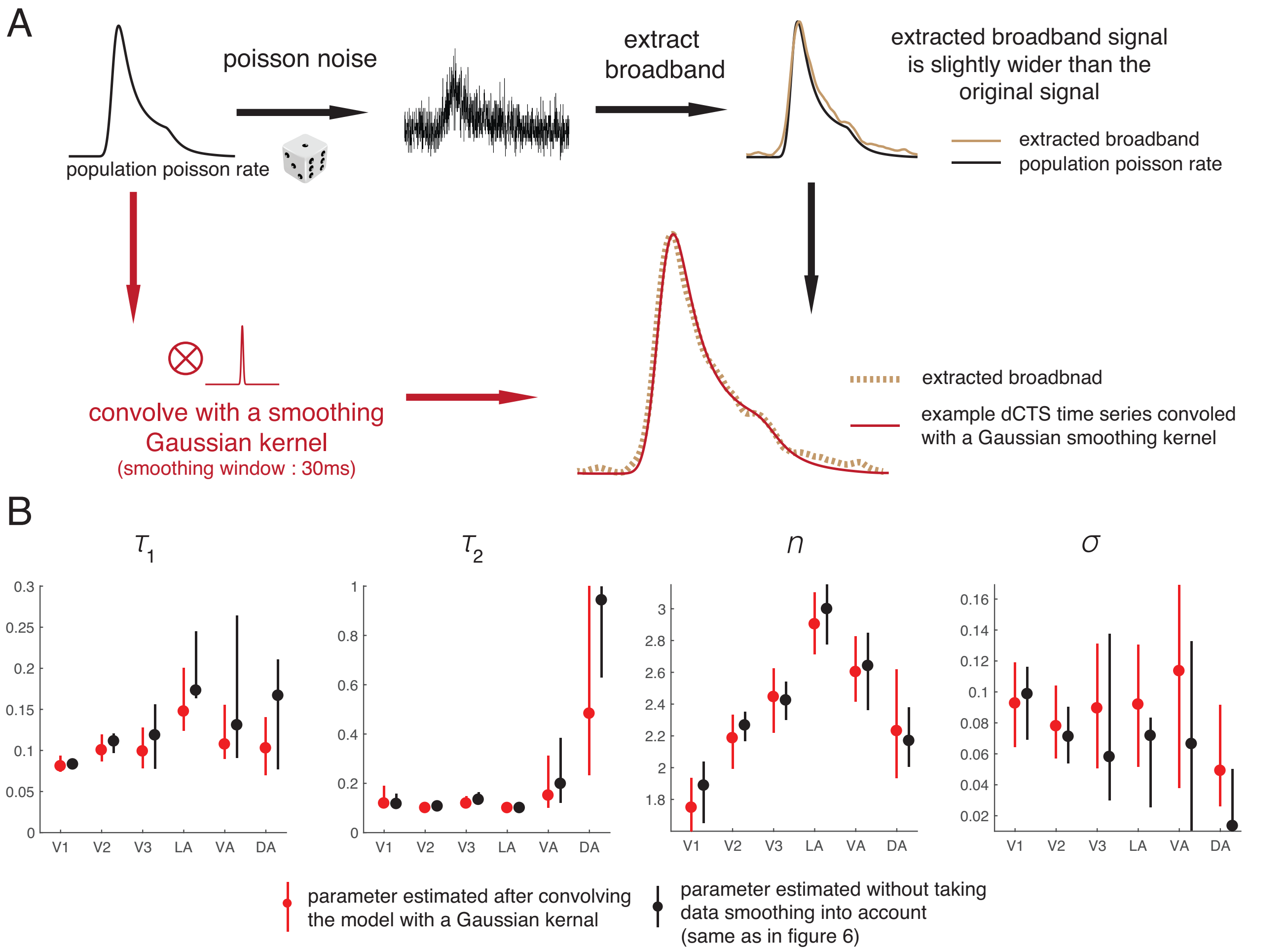


Figure S6. Relationship between broadband envelope and neural time series. *(A)* To understand the effect of extracting the broadband envelope, we simulated a neural response and extracted the envelope. We assumed that a population of neurons has a time-varying average firing rate governed by a Poisson process. The Poisson rate is a latent variable, plotted in the upper left. The result of Poisson sampling, summed over a population of neurons, is shown to the right. This is the presumed neural time series. From this time series, we extract the broadband envelope the way we do from ECoG data, shown on the right in the yellow. The envelope is slightly wider than the original Poisson rate (black). This widening is similar to the effect of smoothing the Poisson time series with a Gaussian blur kernel with a 30ms standard deviation (below, red), which results in a similar time series to the extracted broadband (yellow). *(B)* We replot the dCTS parameters for each ROI in black (same as main text), and compare this to the parameter fits if we insert a Gaussian smoothing step between the model prediction and comparison to the broadband data. The parameters computed this way, shown in red, tend to have slightly shorter τ_1 and τ_2, and smaller *n*, but the general pattern of results is very similar whether this step is inserted or not.

Figure S7. Related to Figure 6.

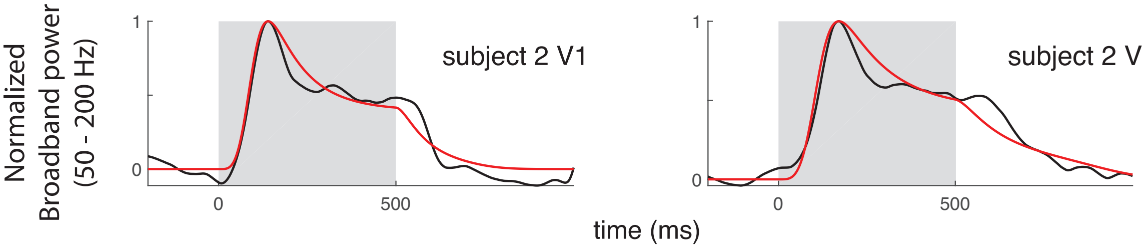


Figure S7. ECoG responses from ECoG subject 2. ECoG broadband responses and dCTS model fits are shown for 2 electrodes in ECoG subject 2. Because there are only two electrodes, one per ROI, we cannot estimate the range of parameters. However, the general form of the time series is comparable to those for S1 in the main text (Figure 6), and the time series is well explained by the model fits.

Figure S8. Related to Figures 4 and 6.

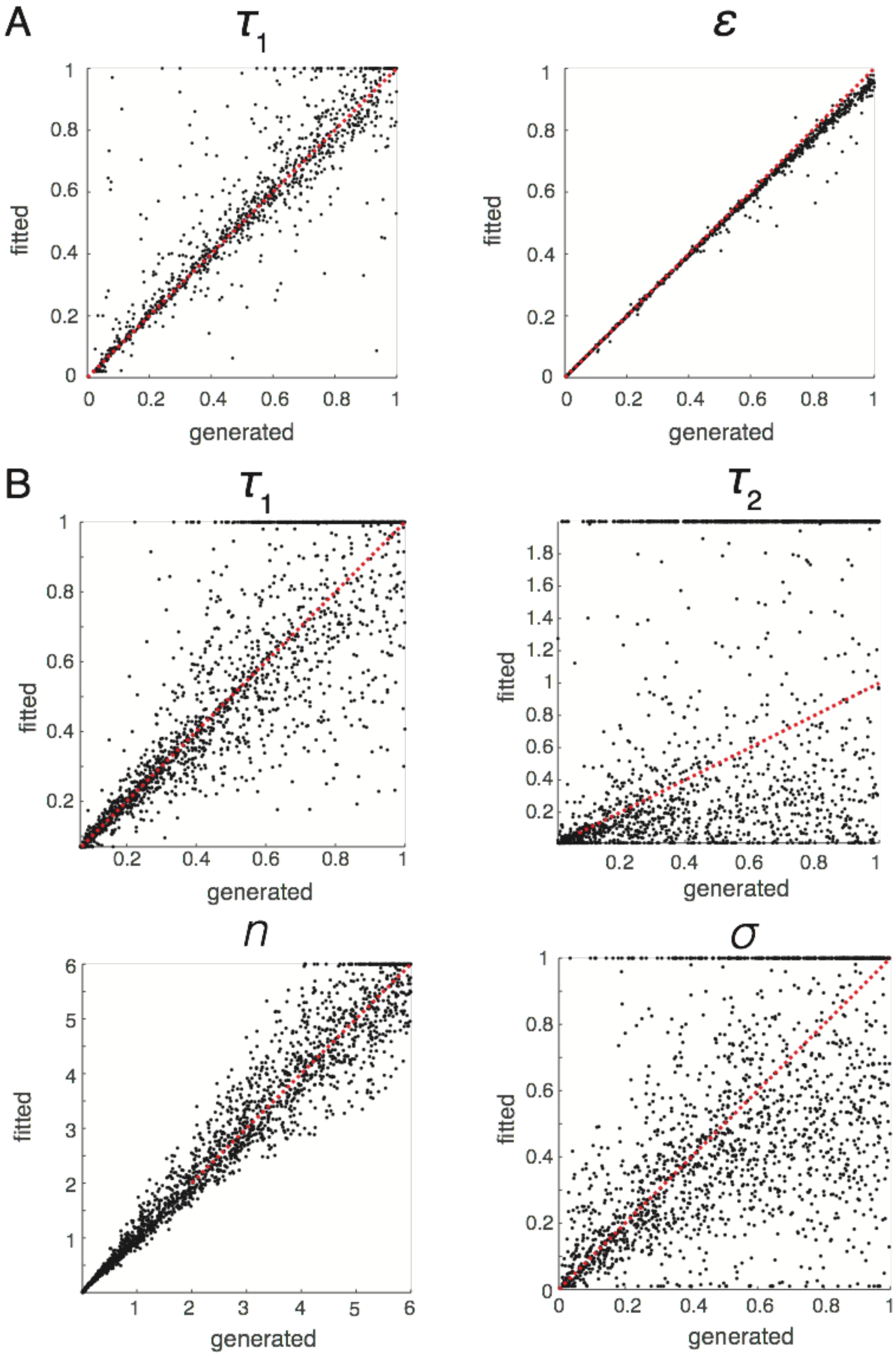


Figure S8. CTS and dCTS parameter recovery. To estimate how accurately we should expect to be able to recover model parameters, we simulated experimental data. (A) We simulated responses to the 13 temporal conditions uses the CTS model with random values for τ_1 from the range [0.01 1], and ε from [0.01 1]. We then added Gaussian white noise to the predictions, and then solved the models on the noisy predictions. The standard deviation of the noise equal to the average residuals between the fMRI data and model fits across all ROIs (as plotted in Figure 4). The plots show the values of the parameter used to generate the predictions (x-axis) and the values recovered from the model fits (y-axis). The ε parameter is recovered more accurately than τ_1. (B) The same simulations were done for ECoG using the dCTS model, selecting parameters from the range [0.01 1] for τ_1, τ_2, and σ, and [0.01 6] for *n.* The added noise was pink (1/f) for each time series, scaled to match the residuals between data and model fits across all ROIs. The parameters *n* and τ_1 are recovered most accurately, σ and τ_2 least accurately.

Figure S9. Related to Figure 5.

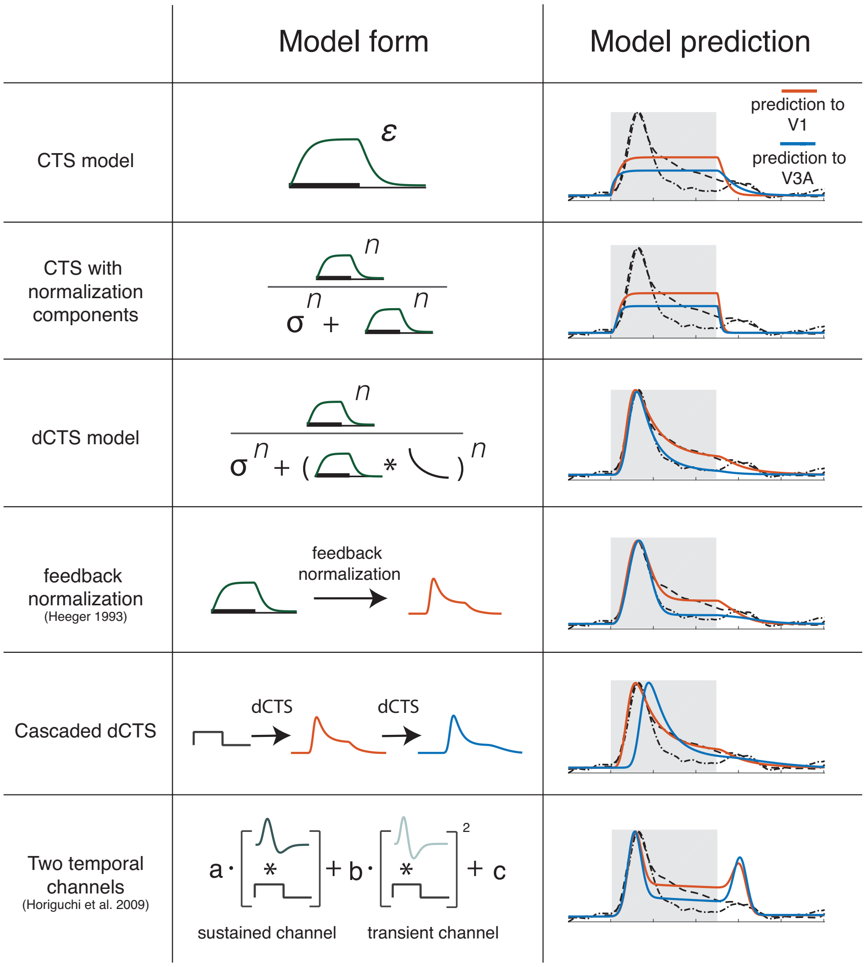


Figure S9. Model comparisons. We illustrate predicted time courses for 6 types of models using a common stimulus - a 500-ms static contrast pattern followed by a 500-ms blank. For each type of model, two sets of parameters were used, generating two time courses: one fit to ECoG data in V1 (red), and one fit to ECoG data in V3A (blue). (i) The CTS model, used to fit fMRI data in this paper. (ii) A close variant of the CTS model, except that the static non-linearity is implemented by divisive normalization rather than a power function. (iii) The dCTS model, used to fit ECoG data in this paper. (iv) Feedback normalization, as proposed by Heeger (1993). This is similar to the dCTS model, except that the normalization is feedback rather than feedforward. (v) A cascade model, which uses the dCTS model to account for V1 data, and computes the downstream response by using the V1 output as input to the identical dCTS model. (vi) A weighted sum of two temporal channels. One channel is sustained and linear; the other is transient with a squaring non-linearity. This model was used by Horiguchi et al (2009) to explain fMRI data, and adapted from related models that account for psychophysical data (Watson, 1986). Overall, the dCTS and feedback normalization models are most similar to the ECoG data. The two temporal channels model captures some features but not others.

Supplementary References

1. Winawer J*, et al.* (2013) Asynchronous broadband signals are the principal source of the BOLD response in human visual cortex. *Curr Biol* 23(13):1145-1153.
2. 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(3):e1003538.
3. Benson NC*, et al.* (2012) The retinotopic organization of striate cortex is well predicted by surface topology. *Curr Biol* 22(21):2081-2085.