Title:
Temporal dynamics of neural responses in human visual cortex

Abbreviated title:
Temporal dynamics in human visual cortex

Authors:
Iris I.A. Groen[1,2], Giovanni Piantoni[3], Stephanie Montenegro[4], Adeen Flinker[4], Sasha Devore[4], Orrin Devinsky[4-6], Werner Doyle[5], Nick Ramsey[3], Natalia Petridou[7], Jonathan Winawer[1]

Affiliations:
[1] Department of Psychology, New York University, New York, NY 10003, USA
[2] Video & Image Sense Lab, Institute for Informatics, University of Amsterdam, 1098 XH Amsterdam, The Netherlands
[3] Department of Neurology and Neurosurgery, Utrecht Brain Center, University Medical Center Utrecht, 3584 CX, Utrecht, The Netherlands
[4] Department of Neurology, New York University Grossman School of Medicine, New York 10016, NY, USA
[5] Department of Neurosurgery, New York University Grossman School of Medicine, New York 10016, NY, USA
[6] Department of Psychiatry, New York University Grossman School of Medicine, New York 10016, NY, USA
[7] Department of Radiology, Center for Image Sciences, University Medical Center Utrecht, 3584 CX, Utrecht, The Netherlands

Corresponding author:
IIA Groen, i.i.a.groen@uva.nl

Conflicts of interest statement:
The authors declare no competing financial interests.

Acknowledgements:
This work was funded by BRAIN Initiative, the National Institute Of Mental Health of the National Institutes of Health under Award Number R01MH111417. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Abstract

Neural responses to visual stimuli exhibit complex temporal dynamics, including sub-additive temporal summation, response reduction with repeated or sustained stimuli (adaptation), and slower dynamics at low contrast. These phenomena are often studied independently. Here, we demonstrate these phenomena within the same experiment and model the underlying neural computations with a single computational model. We extracted time-varying responses from electrocorticographic (ECoG) recordings from patients presented with stimuli that varied in contrast, duration, and inter-stimulus interval (ISI). Aggregating data across patients yielded 88 electrodes with robust visual responses, covering both earlier (V1-V3) and higher-order (V3a/b, LO, TO, IPS) retinotopic maps. In all regions, the temporal dynamics of neural responses exhibit several non-linear features: peak response amplitude saturates with high contrast and longer stimulus durations; the response to a second stimulus is suppressed for short ISIs and recovers for longer ISIs; response latency decreases with increasing contrast. These features are accurately captured by a computational model comprised of a small set of canonical neuronal operations: linear filtering, rectification, exponentiation, and a delayed divisive normalization. We find that an increased normalization term captures both contrast- and adaptation-related response reductions, suggesting potentially shared underlying mechanisms. We additionally demonstrate both changes and invariance in temporal response dynamics between earlier and higher-order visual areas. Together, our results reveal the presence of a wide range of temporal and contrast-dependent neuronal dynamics in the human visual cortex, and demonstrate that a simple model captures these dynamics at millisecond resolution.

Significance Statement

Sensory inputs and neural responses change continuously over time. It is especially challenging to understand a system that has both dynamic inputs and outputs. Here we use a computational modeling approach that specifies computations to convert a time-varying input stimulus to a neural response time course, and use this to predict neural activity measured in the human visual cortex. We show that this computational model predicts a wide variety of complex neural response shapes that we induced experimentally by manipulating the duration, repetition and contrast of visual stimuli. By comparing data and model predictions, we uncover systematic properties of temporal dynamics of neural signals, allowing us to better understand how the brain processes dynamic sensory information.

Introduction

The manner in which neural responses change over short time-scales, from milliseconds to seconds, is important for understanding cognitive, perceptual, and motor functions. Neural dynamics are critical for decision-making (Gold and Shadlen, 2007; Wang, 2012), motor planning (Churchland et al., 2012), and perception (Heeger, 2017), and are important for achieving improved performance in artificial neural network models (Kubilius et al., 2019; Spoerer et al., 2020). Even for simple static stimuli, neural responses in sensory cortex exhibit interesting and complex temporal dynamics. For example, neural responses in visual cortex start to decrease when a static visual stimulus is prolonged in time (subadditive temporal summation), reduce to stimuli that are repeated (adaptation), and rise less rapidly for low contrast stimuli (phase delay).

Studies of temporal dynamics in visual cortex typically measure neural responses with a tailored set of stimuli designed to investigate one particular kind of temporal dynamics. For example, one study might show visual stimuli that vary in duration to investigate how neural activity sums over time (e.g., Tolhurst et al., 1980), while another study might show repeated stimuli to investigate neural response reductions due to adaptation (e.g., Motter, 2006), and yet another study might vary the level of stimulus contrast to study how input strength affects the rise and fall of neural responses over time (e.g., Albrecht et al., 2002). These phenomena are also often studied with different measurement techniques and different computational models.

Even when the same computational framework is used across different stimulus manipulations, model parameters may be fit separately to each stimulus type (e.g., Zhou et al., 2019), leaving open the question of whether a single model with a single set of parameters can account for the multiple phenomena observed across experiments. Here we investigated temporal dynamics in neural responses to visual stimuli that varied systematically in three different ways: duration, repetition, and contrast. We measured electrocorticographic (ECoG) recordings of human visual cortex, which track neural responses at the millisecond scale with high spatial and temporal precision. Using different stimulus manipulations in the same electrodes allows us to investigate links between phenomena. For example, neural dynamics vary with contrast and with stimulus

repetition. Our approach allows us to directly compare these effects and ask how similar they are and whether they can be explained by the same computational mechanism.

By mapping the electrodes to a probabilistic retinotopic atlas within each participant and then aggregating measurements over multiple participants, we collected a large, comprehensive sample of neural responses from multiple visual areas in human visual cortex, covering earlier (V1-V3a/b) and higher-order (LO, TO, IPS) retinotopic maps. From each electrode and for each stimulus, we extracted the time-varying broadband power. Testing the same stimuli in the same participants across multiple visual areas clarifies the qualitative similarities across areas as well as the quantitative differences. For example, our approach allows us to address discrepant claims about temporal window length in the visual hierarchy, as there is some evidence both in support of (Hasson et al., 2008; Weiner et al., 2010; Honey et al., 2012), and against (Fritsche et al., 2020) the claim that temporal window length increases along the visual hierarchy.

The paper is structured as follows. We first examine the different temporal dynamics resulting from each stimulus manipulation (changes in duration, repetition and contrast) and their corresponding non-linear effects in area V1. We show that these modulations are well captured by a dynamic normalization model fit simultaneously to all stimulus types, and perform a systematic comparison with reduced versions of the model, demonstrating the contribution of each individual canonical computation to its ability to fit the data. Finally, we investigate to what extent these temporal dynamics vary both across and within visual areas.

Materials and Methods

Subjects

Data were measured from 11 participants (6 females) who were undergoing subdural electrodes implantation for clinical purposes. 8 participants (5 females) were included in the dataset after data preprocessing (see below for details). Data from 9 participants were collected at New York University Grossman School of Medicine, New York, USA, and 2 participants were tested at the University Medical Center Utrecht (UMCU), Utrecht, The Netherlands. Written informed consent to participate in this study was given by all the patients. The study was approved by the NYU Grossman School of Medicine Institutional Review Board and the ethical committee of the

UMCU, in accordance with the Declaration of Helsinki (2013). All participants were implanted with standard clinical subdural grid and/or strip electrodes. Several NYU participants were additionally implanted with standard clinical depth electrodes. For most participants, electrode implantation and location were guided solely by clinical requirements. Two NYU participants consented to the additional placement of a small, high-density grid (PMT Corporation), which provided denser sampling of underlying cortex. Detailed information about each participant and their implantation is provided in **Table 1** and in **Extended Data Figure 1-1 and 1-2**.

ECoG recordings

NYU

Stimuli were shown on a 15 in. MacBook Pro laptop. The laptop was placed 50 cm from the participant's eyes at chest level. Screen resolution was 1280 × 800 pixels (33 × 21 cm). Prior to the start of the experiment, the screen luminance was linearized using a lookup table based on spectrophotometer measurements (Cambridge Research Systems). Recordings were made using one of two amplifier types: NicoletOne amplifier (Natus Neurologics, Middleton, WI), bandpass filtered from 0.16-250 Hz and digitized at 512 Hz, and Neuroworks Quantum Amplifier (Natus Biomedical, Appleton, WI) recorded at 2048 Hz, bandpass filtered at 0.01- 682.67 Hz and then downsampled to 512 Hz. Stimulus onsets were recorded along with the ECoG data using an audio cable connecting the laptop and the ECoG amplifier. Behavioral responses were recorded using a Macintosh wired external numeric keypad that was placed in a comfortable position for the participant (usually their lap) and connected to the laptop through a USB port. Participants self-initiated the start of the next run by pushing a designated response button on the number pad.

UMCU

Stimuli were shown on a NEC MultiSync® E221N LCD monitor positioned 75 cm from the participant's eyes. Screen resolution was 1920 x 1080 pixels (48 x 27cm). Stimulus onsets were recorded along with the ECoG data using a serial port that connected the laptop to the ECoG amplifier. As no spectrophotometer was available at the UMCU, screen luminance was linearized by reverting the built-in gamma table of the display device. Data was recorded using a MicroMed amplifier (MicroMed s.p.a., Treviso, Italy) at 2048 Hz with a high-pass filter of 0.15Hz and low-pass filter of 500Hz. Responses were recorded with a custom-made response pad.

Stimuli

All code used for generating the stimuli and for experimental stimulus presentation can be found at https://github.com/BAIRR01/BAIR_stimuli and https://github.com/BAIRR01/vistadisp.

Visual stimuli for the purpose of estimating changes in neural temporal dynamics were generated in Matlab 2018b. Stimuli were shown within a circular aperture with a radius of 8.3 degrees of visual angle using Psychtoolbox-3 (https://psychtoolbox.org/) and were presented at a frame rate of 60 Hz. Custom code was developed in order to equalize visual stimulation across the two recording sites as much as possible; for example, all stimuli were constructed at high resolution (2000x2000 pixels) and subsequently downsampled in a site-specific manner such that the stimulus was displayed at the same visual angle at both recording sites (see *stimMakeSpatiotemporalExperiment.m* and *bairExperimentSpecs.m*). Stimuli consisted of grayscale bandpass noise patterns that were created following procedures outlined in Kay et al., (2013). Briefly, the pattern stimuli were created by low-pass filtering white noise, thresholding the result, performing edge detection, inverting the image polarity such that the edges are black, and applying a band-pass filter centered at 3 cycles per degree (see *createPatternStimulus.m*). All stimuli were presented within a circular aperture; the remainder of the display was filled with neutral gray (**Fig. 1*A***). The spatial pattern we used has been shown to effectively elicit responses in most retinotopic areas (Kay et al., 2013b; Zhou et al., 2018).

After generating independent pattern stimuli for each trial, stimuli were assigned to one of three different stimulus conditions (**Fig. 1*B***): duration, inter-stimulus-interval (ISI) or contrast (*stimMakeSpatiotemporalExperiment.m*). Duration and ISI trials were generated using the same parameters as in Zhou et al., (2018). For duration trials, a single pattern stimulus was shown at full contrast for one of 6 durations that ranged between 17-533 ms. For ISI trials, a single full contrast pattern stimulus was shown for a fixed duration (134 ms) twice in a row, with one of 6 inter-stimulus intervals whose duration varied between 17-533 ms. Duration and ISIs were powers of 2 times the monitor dwell time (1/60s). For contrast trials, the contrast varied from 6.25% to 100% in 5 steps of powers of 2. Contrast trials were shown at a fixed duration of 500 ms. Together, these trial-types amounted to 17 stimulus conditions (6 duration, 6 ISIs, and 5 contrasts).

Experimental design and statistical analysis

Experimental procedure

Duration, repetition, and contrast conditions were divided across two runtypes: one containing the duration- and ISI-varying trials, and one containing the contrast-varying trials as well as an additional set of grating and density-varying stimuli that were not analyzed for the purpose of the current paper. Trials were presented with an inter-trial-interval (ITI) that was randomly picked from a uniform distribution varying between 1.25 and 1.75 seconds. For each runtype, two unique stimulus sequences were created containing randomly ordered trials and ITIs. Each run contained 36 stimuli with 3 instances per stimulus condition and lasted about 60 seconds.

All but one participant (see **Table 1**) completed an additional task in which they viewed visual stimuli for the purpose of estimating the spatial selectivity for individual electrodes (population receptive field mapping; Dumoulin and Wandell, 2008; Kay et al., 2013a). In these pRF mapping runs, grayscale bandpass noise patterns were viewed through a bar aperture that swept across the visual field eight times in twenty-eight 0.85 s steps, which also included 8 blanks each. Each bar stimulus was displayed for 0.5 s followed by a 0.35 s blank period showing a grey mean luminance image. Each pRF run lasted 196 seconds (about 3 minutes).

For all tasks, participants were instructed to fixate on a cross located in the center of the screen and press a button every time the cross changed color (from green to red or red to green). Fixation cross color changes were created independently from the stimulus sequence and occurred at randomly chosen intervals ranging between 1 and 5 s. For each task, participants completed each unique run at least once. Runs were then repeated several times within the same experimental session or in an additional session recorded on a different day. The amount of data collected for each patient is provided in **Table 1**. The experimenter stayed in the room throughout the experiment. Patients were encouraged to take short breaks in between runs.

ECoG data analysis

Data preprocessing

Data were read into Matlab 2020b using the Fieldtrip Toolbox (Oostenveld et al., 2011) and preprocessed with custom scripts available at https://github.com/WinawerLab/ECoG_utils. Raw data traces obtained in each recording session were visually inspected for spiking, drift or other artifacts. Electrodes that showed large artifacts or showed epileptic activity were marked as bad and excluded from analysis. Data were then separated into individual runs and formatted to conform to the ieeg-BIDS format(Holdgraf et al., 2019). Data for each run were re-referenced to the common average across electrodes for that run, whereby a separate common average was calculated per electrode group (e.g., one for grid and one for strip electrodes, see *bidsEcogRereference.m*). Next, a time-varying broadband estimate was computed for each run in the following way (see *bidsEcogBroadband.m*). First, the voltage traces were band-pass filtered using a Butterworth filter (passband ripples < 3 dB, stopband attenuation 60 dB) for 10 Hz-wide bands that ranged between 50-200 Hz. Bands that included frequencies that were expected to carry line noise and their harmonics were excluded (for NYU: bands 60, 120 and 180 Hz; for UMCU, 50, 100 and 150 Hz). The power envelope of each band-pass filtered voltage time course was calculated as the square of the time course's Hilbert transform. The resulting envelopes were then averaged across bands by taking the geometric mean (see *ecog_extractBroadband.m*). Unlike the arithmetic mean, taking the geometric mean ensures that the resulting average is not biased towards the lower frequencies. The re-referenced voltage and broadband traces for each run were written to BIDS derivatives directories.

Electrode localization

Intracranial electrode arrays from NYU participants were localized based on pre- and post-implantation structural MRI images (Yang et al., 2012). Electrodes from UMCU patients were localized from the postoperative CT scan and co-registered to the preoperative MRI (Hermes et al., 2010). Electrode coordinates were computed in native T1 space and visualized onto pial surface reconstructions of the T1 scans generated using Freesurfer (http://freesurfer.net). Visual maps of striate and extrastriate cortex were generated for each individual participant based on the preoperative anatomical MRI scan by aligning the surface topology with two atlases of retinotopic organization: an anatomically defined atlas (Benson et

al., 2014; Benson and Winawer, 2018) and a probabilistically defined atlas derived from a retinotopic fMRI mapping dataset (Wang et al., 2015); **Fig. 1*C***. Using the alignment of the participant's cortical surface to freesurfer's fsaverage subject, atlas labels defined on the fsaverage were interpolated onto their cortical surface via nearest-neighbour interpolation.

Electrodes were then matched to both the anatomical and the probabilistic atlases using the following procedure (see *bidsEcogMatchElectrodesToAtlas.m*): For each electrode, the distance to all the nodes in the freesurfer pial surface mesh were calculated, and the node with the smallest distance was determined to be the matching node. The atlas value for the matching node was then used to assign the electrode to one of the following visual areas in the anatomical atlas (henceforth referred to as the Benson atlas): V1, V2, V3, hV4, VO1, VO2, LO1, LO2, TO1, TO2, V3a, V3b, or none; and to assign it a probability of belonging to each of the following visual areas in the probabilistic atlas (henceforth referred to as the Wang atlas)): V1v, V1d, V2v, V2d, V3v, V3d, hV4, VO1, VO2, PHC1, PHC2, TO2, TO1, LO2, LO1, V3b, V3a, IPS0, IPS1, IPS2, IPS3, IPS4, IPS5, SPL1, FEF, or none.

Data selection

The preprocessed data were analyzed for the purposes of this paper using custom Matlab code available at https://github.com/irisgroen/temporalECoG. First, a dataset was created by reading in the voltage and broadband time courses of each run from each participant from the corresponding BIDS derivatives folders (*tde_getData.m*). Only electrodes that had a match with one of the visual areas from the anatomical atlas (Benson and Winawer, 2018) or one the visual areas in the maximum probabilistic atlas from (Wang et al., 2015), were selected for preprocessing. In addition, only (high-density) grid or strip electrodes were included (i.e., depth electrodes were excluded).

To combine the UMCU and NYU participants in a single dataset, the UMCU data was resampled at 512 Hz prior to epoching. Visual inspection of the data indicated a shift in response onset between the UMCU and NYU participants, due to a delay caused by the KVM switch at the UMCU clinic (assessed with a photodiode, Black Box ToolKit v2 www.blackboxtoolkit.com). To correct the delay, UMCU data were aligned to the NYU data based on a cross-correlation on the average event-related potential (ERP) across all stimulus conditions from the V1 and V2 electrodes from three participants (1 UMCU, 2 NYU). The delay in stimulus presentation was

estimated to be 72ms, and stimulus onsets of the UMCU participants were shifted accordingly (see *s_determineOnsetShiftUMCUvsNYU.m*)

Broadband and voltage data were epoched between [-0.1 and 1.2] seconds relative to stimulus onset. Voltage epochs were baseline-corrected by subtracting the average pre-stimulus amplitude within each trial. Broadband epochs were converted to percent signal change by pointwise dividing and subtracting the average prestimulus baseline across all epochs within each run. We then performed two consecutive data trimming steps: 1) epoch selection and 2) electrode selection (see *tde_selectData.m)*.

Epoch selection was performed using both the voltage and broadband epochs. Epochs were automatically rejected according to two criteria (see *ecog_selectEpochs.m*): 1) a difference in voltage between consecutive samples larger than 200 mV anywhere in the epoch; 2) the maximum broadband amplitude in the epoch *outside* the time window [0.05 0.55] relative to stimulus presentation exceeds the average of the maximum response across trials *inside* the stimulus presentation window by 3 standard deviations. Across participants, on average 1.65% of epochs were rejected (sd = 0.86%, min = 0.24%, max =3.1%).

Electrode selection was performed based on the broadband epochs and the electrode locations. To be included in data analysis, electrodes had to meet one criterion (see *ecog_selectElectrodes.m*): an above-threshold correlation between two independent halves of the data (split-half reliability). To compute this measure for each electrode, all epochs for a given trial type were randomly assigned to one of two splits. Within each half, epochs of the same trial-type were averaged and then concatenated, and the coefficient of determination (R2) was calculated between the two concatenated time series. Electrodes with a low split-half reliability (R2 < 0.2) were excluded.

For three participants, electrode selection resulted in all their electrodes getting rejected from analysis. For the 8 remaining participants, 71% of electrodes on average were rejected (sd = 14.5%, min = 50%, max = 91%; see **Table 1** for subject-specific data).

Data summary

The data preprocessing, electrode localization and data selection procedures outlined above resulted in a total of 88 electrodes with robust visual responses covering visual areas V1, V2, V3, V3a, V3b, hV4, LO1, LO2, TO1, TO2, IPS0-4 (see **Table 2** for an overview of electrodes per area). The broadband epochs of selected electrodes were averaged across all trials within a stimulus condition, resulting in 17 response time courses per electrode that constituted the data for computational model fitting (see below).

Probabilistic electrode assignment

When assigning individual electrodes to visual areas, we used a bootstrapping procedure that took into account the probability of each electrode overlapping with a visual retinotopic region. In this procedure, we repeatedly (for n bootstraps) assigned electrodes to visual areas in accordance with each electrode's matched node's probability distribution across areas determined by the Wang probability atlas (Wang et al., 2015), resulting in a distribution of bootstrapped electrodes for each visual area. From these distributions we then computed summary statistics (median or mean) and confidence intervals across neural responses or model predictions for each visual area. Before conducting the repeated assignments, we rescaled the probability values in the atlas to exclude the 'none' probabilities, which ensured that each electrode would always get assigned to an area (e.g. if an electrode had 60% probability of belonging to V3b, 20% probability of belonging to V3a, and 20% probability of belonging to 'none', it would get assigned to V3b 75% of the time and to V3a 25% of the time). To guarantee robust estimates, we only report results from areas that had at least 10 electrodes with a non-zero probability of being positioned on that area. This resulted in the exclusion of area hV4; and since only two electrodes overlapped with non-zero probability with area TO2, we grouped together electrodes for TO1 and TO2 into a single area TO.

Computational modeling

Model fitting: temporal dynamics

We fit models of temporal dynamics to the 17 broadband time courses for each individual electrode (see *tde_fitModel.m*). Each model takes stimulus time courses as input and predicts neural response time courses as output. For analyses focusing on describing different temporal

phenomena in V1 (Figures 2-4), comparisons of model performance (Figure 7) and visual field locations (Figure 11), models were fit separately to individual electrodes, and parameters or metrics derived from these fits were then averaged within visual areas using the probabilistic assignment bootstrapping procedure described above. For analyses that characterized changes in temporal dynamics along the visual hierarchy (Figures 8-10), we instead conducted repeated (n=1000) fits to the average time courses across electrodes within an area, whereby for each fit the average time course per area was computed after performing the probabilistic assignment, yielding a slightly different average and thus different model fit each time. While results and conclusions were not qualitatively different, we found that in higher visual areas, this averaging-then-fitting procedure yielded slightly more robust estimates compared to fitting individual electrodes and subsequently averaging them. For analyses illustrating similarities in contrast and adaptation time courses (Figures 5-6), model predictions were generated after electrodes were first averaged probabilistically within V1.

Models were fit using Bayesian adaptive direct search (BADS; Acerbi and Ma, 2017). BADS alternates between a series of fast, local Bayesian optimization steps and a systematic, slower exploration of a mesh grid, and we found that it resulted in slightly more robust fits to the data compared to two built-in Matlab alternatives *fmincon* and *lsqnonlin*; however, the overall results and conclusions did not change qualitatively when using these optimizers, and all three options are implemented in the fitting code (see *tde_fitModel.m*). Model performance was evaluated by computing the cross-validated coefficient of determination (R2) through 17-fold leave-one-out cross-validation whereby each of the stimulus conditions was once left out of the fitting procedure and then predicted based on the parameters estimated from the remaining 16 stimulus conditions. Model R2 was computed by averaging across the R2 values for the 17 left out stimuli. Model parameter values and summary parameters (see below) were estimated based on fits to the full dataset.

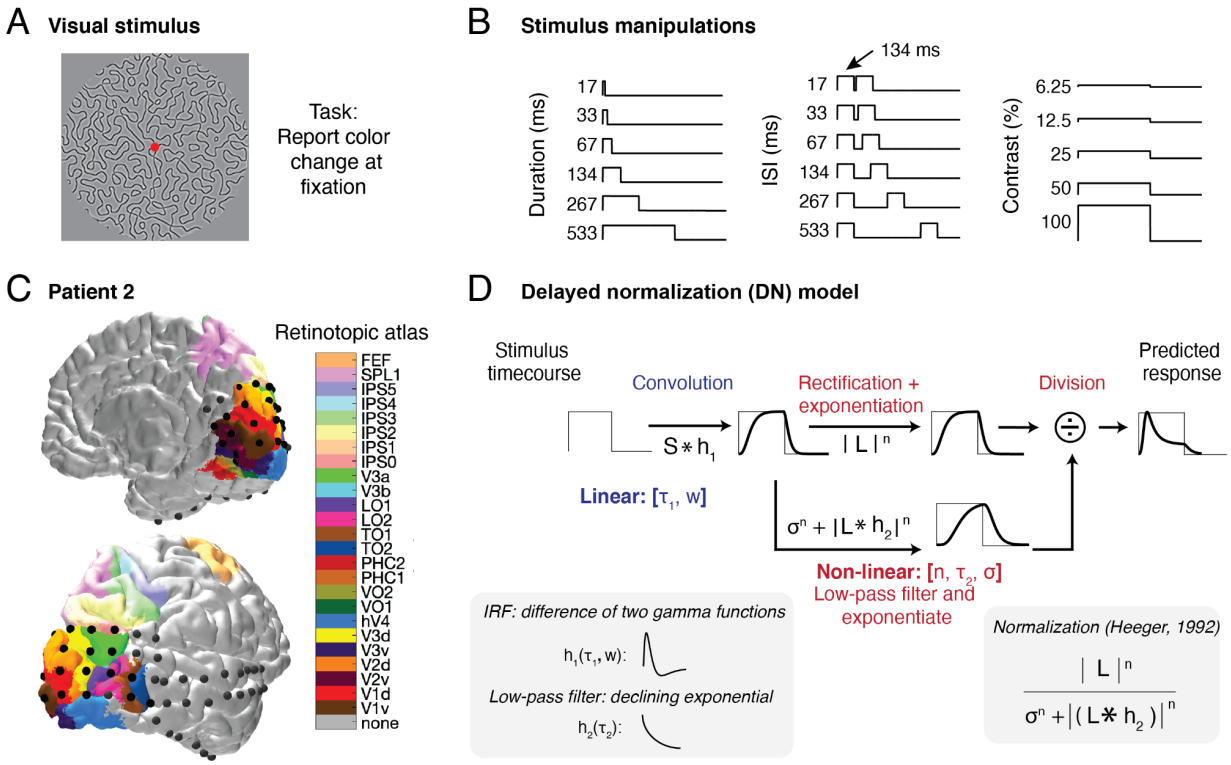


Figure 1: Experimental design, ECoG data and delayed normalization model.
A) Participants were shown large (16°) grayscale pattern stimuli consisting of connected curved line segments. Unique individual stimuli were used for each stimulus condition. **B**) Stimulus time courses were varied in three different ways. For duration trials, a single full contrast stimulus was presented for different durations ranging between 17-533 ms. For repetition trials, a full contrast stimulus was presented twice in a row for 134 ms, with an inter-stimulus interval (ISI) ranging between 17-533 ms. For contrast trials, a single stimulus was presented at a fixed duration (500 ms) with levels of contrast ranging between 6.25% to 100%. **C)** Electrode positions from an example participant overlaid on a pial surface reconstruction with predicted visual area locations**.** In this visualization, a surface node in the pial mesh was assigned a color if it had a non-zero probability of being on a visual region according to the full probability map from (Wang et al., 2015). If the electrode had a non-zero probability of being in multiple regions, the region with the highest probability was assigned. High-resolution images of each patient's electrode positions and atlas projections are provided in **Extended Data Figure 1-1** and **Extended Data Figure 1-2**; interactive 3D figures can be generated using *tde_mkFigure1_1.m* in https://github.com/irisgroen/temporalECoG**. D)** Visual depiction of the delayed normalization (DN) model, first presented in (Zhou et al., 2019). This linear-non-linear gain model takes a stimulus time course as input and produces a predicted response time course as output, by applying both linear and non-linear computations. The initial, linear part consists of convolution with an impulse response function (IRF; see lower left inset), parameterized as a difference of two gamma functions, with free parameters τ_1 (time constant) and *w* (weight of negative to positive gamma). This is followed by a non-linear part consisting of rectification, exponentiation with exponent *n*, and division by a semi-saturation contrast σ, plus a delayed copy of the input that are also rectified and exponentiated. The delay is implemented as a convolution of the linear response with an exponentially decaying function of time constant τ_2 (see lower left inset). This model form constitutes a temporal implementation of divisive normalization, as first proposed by Heeger (1992) (see lower right inset).

Computational models of temporal dynamics

The DN model and its reduced versions (see below) are implemented as standalone functions in https://github.com/irisgroen/temporalECoG, folder 'temporal_models'. Each model is paired with a JSON metadata file containing parameter descriptions, starting points and upper and lower bounds that were used when fitting the model. As detailed below, models differ in the number and types of parameters fitted. In addition to the model-specific parameters, two nuisance parameters were fitted for all models: *shift* (delay in response onset relative to stimulus onset) and *scale* (gain of the response), to take into account differences between electrodes or visual areas in response onset latency and overall response magnitude (i.e. responses in V1 are much higher than in later visual areas). We tested the following models:

Delayed normalization model

Our main model of interest was the delayed normalization (DN; **Figure 1D**) model previously described in Zhou et al., (2019), who demonstrated that this model was able to predict compressive temporal summation (in ECoG data to stimuli of a single duration), adaptation (in fMRI responses to stimuli of varying durations and interstimulus-intervals) and contrast dynamics (in non-human primate MUA recordings from V1). Here, we tested whether the same model could predict all three of these phenomena *simultaneously* in ECoG data. The DN model is defined by an LNG structure (Linear, Nonlinear, Gain control) and is described by the following set of equations:

First, a linear response (R_L) is computed by convolving an input stimulus time course *S* of length *t* (dimensions 1x*t*, with values ranging between 0 and 1) with an impulse response function (IRF) h_1, consisting of a weighted difference of two gamma functions:

$$R_L = S * h_1(\tau_1, w) \quad ,$$
$$h_1(\tau_1, w) = te^{-t/\tau_1} - w \cdot te^{-t/1.5\tau_1} \qquad (1)$$

in which t is time, w the weight on the second (negative) gamma function, and τ_1 is a time constant determining the shape of the IRF. Symbols in parentheses are parameters. R_L, S, and h_1 are functions of time; we omit t for simplicity. Note that each individual gamma function here is a simplification of the gamma function provided in Boynton et al., (1996); their equation 3, whereby the phase delay of the IRF is fixed at value 2 (as in Zhou et al., 2019). Moreover, as in Zhou et al., (2019), we assumed the peak timing of the second (negative) gamma function to be 1.5 times the first (positive) one, leaving only τ_1 and w as the free parameters of the linear step. This version of the DN model (*DN.m*) was used for all analyses with the exception of the model reduction analysis (Figure 7), where we compare this model to models with a more flexible IRF, which has two additional free parameters: one for the phase delay and one for the time constant of the negative gamma function (see next section).

The linear response (R_L) is then converted into a non-linear response R_{LN} by applying a full-wave rectification and an exponentiation with exponent n, which is fitted:

$$R_{LN} = |R_L|^n \tag{2}$$

The final step is divisive normalization of the non-linear response R_{LN} with a low-pass filtered version of R_L that is also rectified and exponentiated to the same n. In addition, an exponentiated semi-saturation constant σ is added to the denominator:

$$R_{DN} = \frac{R_{LN}}{\sigma^n + |R_L * h_2(\tau_2)|^n} \quad , \tag{3}$$
$$h_2(\tau_2) = e^{-t/\tau_2}$$

The low-pass filtering here is achieved by means of convolution with an exponential decay function h_2 with time constant τ_2.

As noted in **Fig. 1*D*,** equation 3 has the form of canonical divisive normalization (Carandini and Heeger, 2012), in which a numerator (reflecting the rectified and exponentiated linear input) is divided by a denominator consisting of a normalization pool and a semi-saturation constant, which each are also raised to the same exponent n as the numerator. Here, the normalization pool consists of a delayed version of the numerator, which means that the input drive is

essentially normalized by a delayed version of itself. This yields an output time course that is characterized by a transient response rise followed by a decay to a sustained response level.

To summarize, for the main version of the DN model used throughout, five parameters were fitted in total: τ_1 (time constant of the IRF), *w* (weight of the negative and positive IRFs), *n* (exponent), σ (semi-saturation constant), and τ_2 (time constant of the exponential decay). Model parameters and bounds were initialized using values used in Zhou et al., (2019).

Delayed normalization model, fully parameterized and deconstructed

In addition to the original model proposed by Zhou et al., (2019), we tested reduced versions of that same model, in which non-linear computations were consecutively added on top of the purely linear first step of the model. The motivation for including this model construction analysis was to investigate how much each step contributed to the overall predictive performance of the model. As explained above, we additionally adapted the model such that IRF used for linear convolutions was less constrained, allowing the time constant of the negative gamma function and the phase delay of both gamma functions to be fitted in optimization, such that it could capture as much variance in the data as possible with a maximally flexible IRF (total number of fitted parameters = 4: $\tau_{1\text{pos}}$, $\tau_{1\text{neg}}$, phase delay *r*, and *w*; see *LINEAR.m*). We then added full-wave rectification (*LINEAR+RECTF.m*; same parameters as *LINEAR.m*), exponentiation (*LINEAR+RECTF+EXP.m;* adding parameter *n*), normalization without delay (*LINEAR+RECTF+EXP+NORM.m;* adding parameter σ) and finally, normalization with delay (*LINEAR+RECTF+EXP+NORM+DELAY.m*; adding parameter τ_2), which is equivalent to the DN model (*DN.m*), except that the IRF is less constrained. Model parameters and bounds were initialized using the same values as for the DN model and can be found in the JSON meta-data.

Summary metrics

To characterize the temporal dynamics of the broadband time courses, we computed several summary metrics on both data and model predictions (see *tde_computeDerivedParams.m*):

Time-to-peak: The time interval between stimulus onset and the maximum (peak) of the response time course (in seconds). We computed this based on responses or model predictions

for the longest duration, full contrast stimulus (533 ms), which is the condition of maximal stimulation and thus the conditions where the response is overall expected to be greatest..

Full-width half max (Fwhm): The difference between the time-point (relative to stimulus onset) at which the response has risen to half of the maximum, and the time-point at which the response has decayed again to the same half-way point (in seconds). This metric was computed based on neural response or model prediction time courses to short duration stimuli (e.g. 17 ms), which have the least amount of sustained level activation and therefore the most symmetric response shapes.

Ratio sustained/transient: Response magnitude at the sustained level divided by the transient response magnitude (maximum/peak). For estimates of the sustained level, we took the response level at stimulus offset of the longest duration in the dataset (533 ms). To achieve robust estimates of the offset, data and model predictions were smoothed using a moving average filter with a span of 150 time samples (see *tde_mkFigure4.m*).

Recovery from adaptation: The relative magnitude of response of the second stimulus relative to the first stimulus in the repeated stimuli conditions. To obtain a robust estimate of the response to the first stimulus, we averaged together the response time courses for the 134ms duration stimulus (same duration as the ISI stimuli) and each of the ISI stimuli from trial onset *up to* the the onset of the second stimulus (see *tde_computeISIrecovery.m)*. We then subtracted this average time course from the response in each ISI stimulus condition, yielding an estimate of the response of the second stimulus 'corrected for' the response to the first stimulus. We then computed relative recovery by taking the maximum (peak) of the corrected second stimulus time courses and comparing it with the peak response of the average first stimulus estimate. Taking the sum rather than the peak of the time course yielded qualitatively similar results.

C50: Percentage contrast for which response magnitude reaches 50% of the response to a 100% contrast stimulus. For both the neural data and the model predictions, this point was estimated by fitting a NakaRushton function to the maxima of the response time courses for each of the 5 contrast levels we measured (see *fitNakaRushton.m*). Taking the sum rather than the peak of the time course yielded qualitatively similar results.

pRF analysis

Data analysis and model fitting

ECoG recordings in pRF mapping runs were preprocessed the same way as the runs with duration, ISI and contrast-varying stimuli described above, yielding time-varying broadband signal change epochs of length [-0.2 0.6] seconds relative to the onset of each bar stimulus (see *tde_prfrun.m*). These broadband time courses were then averaged between [0.05-0.55] seconds, resulting in a single power estimate for each bar position in the stimulus sequence (see *tde_computePRFtimecourses.m*). The resulting pRF timeseries were used to compute a pRF model(Dumoulin and Wandell, 2008), using the Compressive Spatial Summation (CSS) variant (Kay et al., 2013a).

Briefly, this involved (1) converting the stimulus into a sequence of binary contrast apertures, (2) projecting the contrast apertures onto the best-fitting 2D isotropic Gaussian population receptive field (pRF), and (3) passing the output through a static nonlinearity (power function) to predict the response. The pRF models were fit to each electrode's ECoG responses separately by minimizing the difference between the predicted response and the observed response, using nonlinear optimization as implemented in the analyzePRF toolbox (Kay et al., 2013a). For each electrode, estimates of pRF angle, eccentricity and size were estimated using analyzePRF. Eccentricity is defined as the distance of the center point of isotropic Gaussian pRF from the fixation point at the center of the visual stimulus, while pRF size is defined as $\sigma/\sqrt{n}$, whereby σ is the standard deviation of the Gaussian pRF and n is the exponent of the power-law function. Before fitting the model, data were averaged across the runs. **Extended Data 10-1** shows summary plots of the fitted parameters for the electrodes analyzed in the current study.

We then used the estimated eccentricity for each individual electrode to separate electrodes into foveal (<2 degrees) and peripheral (>2 degrees, <16.6 degrees) bins. To ensure that eccentricity estimates were reliable, we only included electrodes for which the pRF model explained at least 30% of the variance in the pRF time series. This resulted in a total of 64 electrodes being included in the foveal-peripheral comparison (42 foveal, 22 peripheral). These electrodes were then assigned to either lower (V1-V3) or higher (V3b and beyond) visual areas and averaged using the probabilistic assignment procedure described above. Since electrodes can have

overlap with both lower and higher regions (for example, an electrode could have 20% probability of being in V3 and 80% probability of being in V3b), this means that some electrodes contributed to average estimates in both lower and higher visual areas. In contrast, the assignment to foveal or periphery was binary, i.e. each electrode was either foveal or peripheral.

Code and data availability statement

All code used for the purpose of this paper can be found at the GitHub repositories mentioned above. The ieeg-BIDS-formatted data and derivatives will be made available on the OSF.

Results

We first examined how each of our three stimulus manipulations (duration, repetition and contrast) affected the temporal dynamics of ECoG broadband time courses in area V1.

Neural responses in human V1 exhibit subadditive temporal summation

First, systematic variations of stimulus duration demonstrated the presence of *subadditive temporal summation* in neural responses in V1 (**Fig. 2**). When a stimulus doubles in duration, its resulting neural response is reduced relative to the linear prediction, i.e., a sum of the responses to the shorter stimulus and a time-shifted copy of that response (**Fig. 2*A***). This demonstrates that the additional visual exposure resulting from a longer presentation duration does not accumulate linearly. Comparison of summed neural responses across all stimulus durations indeed shows evidence of subadditive summation; throughout the 30-fold range measured in our experiment (17 ms to 533 ms), the obtained responses deviate systematically from the linear prediction, with responses to brief stimuli much larger than the linear prediction fit to the longest stimulus (**Fig. 2*B*)**.

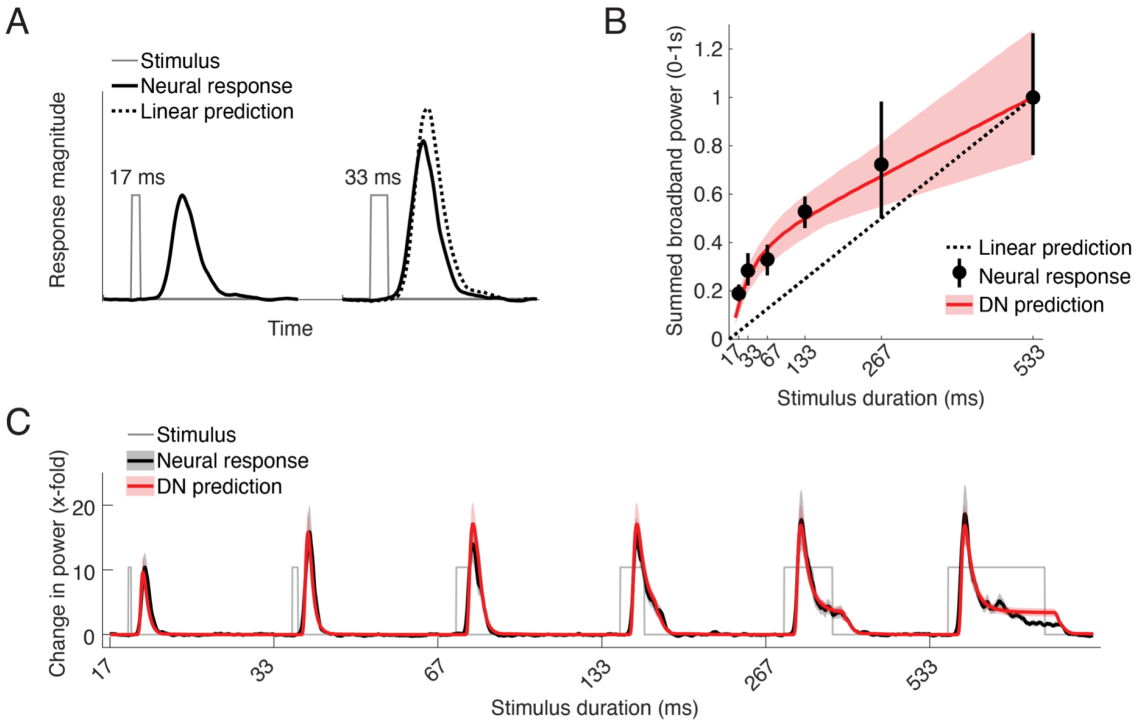


Figure 2: Sub-additive temporal summation in neural responses in human V1.
A) Compressive temporal summation in neural responses illustrated by two stimulus time courses (17 and 33 ms durations, gray thin lines) and corresponding ECoG broadband time courses (black thick lines). Doubling stimulation duration does not result in a doubling of the response, as predicted by a linear prediction, computed as the sum of the shortest duration response and a shifted and scaled copy of that response (dotted line). **B**) Sum of neural responses for each stimulus duration (black) and the predicted response by the DN model (red). Compressive temporal summation is evident as deviation of both data and model from the linear prediction (dotted line). Data points reflect summed broadband response time courses between 0-1 s after stimulus onset, and are normalized to the response for the longest duration (533 ms). Data points indicate median across repeated samples of V1 electrodes, taking into account each electrode's probability of overlap with a retinotopic atlas (see Methods). Error bars and shaded regions reflect 68% confidence intervals across 10,000 bootstraps of electrode assignments. **C**) Average ECoG broadband time courses in V1 (black) for all stimulus durations measured (17-533 ms), along with the average prediction by the DN model (red). Shaded regions reflect 68% confidence intervals across 1000 bootstraps of electrode assignments. This figure is produced using *tde_mkFigure2.m* in https://github.com/irisgroen/temporalECoG.

The reason for this can be understood by looking at the average response time courses themselves (**Fig. 2*C***). Going from a very short stimulus duration to a slightly longer duration results in an increase in the magnitude of the initial response transient (i.e. peak). However, as stimulus duration increases further, the transient part of the response saturates, and a sustained response emerges on its downslope, which remains elevated at a steady-state level until

response offset. Summed responses in **Fig. 2*B*** reflect the combination of both transient and sustained level activity; since the sustained response magnitude is lower compared to the transient, and keeps decreasing over time, responses to longer durations accumulate relatively less over time.

We find that this temporal subadditivity in ECoG broadband responses is well captured by a delayed normalization (DN) model (**Fig. 1*D***). The model accurately predicts the rapid increase in response at short durations and the slower increase at longer durations (**Fig. 2*B***). Moreover the model time courses accurately fit both the transient and the sustained levels of neural response across different stimulus durations (**Fig. 2*C***).

Neural responses in human in V1 exhibit response suppression due to adaptation

Second, repetitions of stimuli with different inter-stimulus intervals (ISIs) demonstrated the presence of *adaptation* in neural responses in V1 (**Fig. 3**). Adaptation is evident as a reduction in stimulus-evoked response due to the presence of a preceding stimulus, and is most pronounced when the two stimuli are presented close together in time (**Fig. 3*A***). Adaptation effects are non-linear: while a linear model predicts no change in response magnitude due to preceding stimuli, the neural data reveal strong suppression for a stimulus that is presented in close temporal proximity to another stimulus, which gradually recovers as the ISI increases (**Fig. 3*B***). These adaptation effects are clearly visible in the response time courses to the separate stimuli (**Fig. 3*C***), where it can be seen that the response to the second stimulus achieves near-full recovery by the longest ISI tested (533 ms). Notably, these adaptation effects are well captured by the delayed normalization model, as indicated by the red lines in **Fig. 3*B-C***, although there are slight underpredictions for the response to the second stimulus at short ISIs.

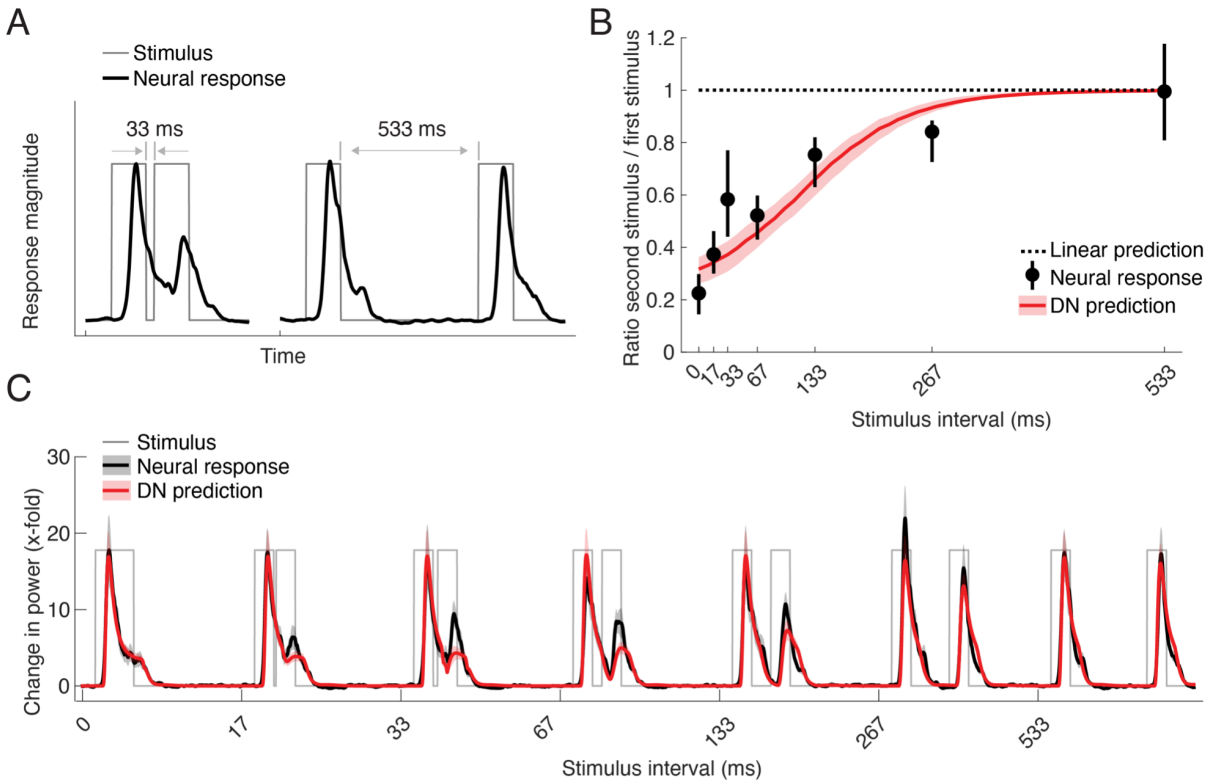


Figure 3: Adaptation of neuronal responses to stimulus repetition in human V1.
A) Example of adaptation to a repeated stimulus, showing time courses for two stimuli of equal duration with a short (33 ms) and a long (533 ms) inter-stimulus interval (gray thin lines) and their corresponding ECoG broadband time courses (black thick lines). Short intervals lead to a strong suppression of the response to the second stimulus, and the response recovers with longer intervals. **B**) Recovery from adaptation for each condition, expressed as the difference in maximum response for the second stimulus compared to the maximum response of the first stimulus (black dots) along with the predicted recovery by the DN model (red line). Adaptation is evident from the deviation of both data and model from the linear prediction (dotted line). Data points indicate median across probabilistically assigned V1 electrodes. Error bars and shaded regions indicate 68% confidence intervals across 10,000 bootstraps of electrode assignments. **C**) Average broadband time courses (black) in V1 for inter-stimulus intervals between 0-533 ms, along with predictions by the DN model (red), computed as in **Fig. 2*B***. This figure is produced using *tde_mkFigure3.m*.

Neural responses in human in V1 exhibit slower dynamics with reduced contrast

Third, we observed *contrast-dependent temporal dynamics* of neural responses in V1, in spite of the fact that the temporal structure of the stimulus time course itself did not differ across the contrast conditions (each stimulus was presented at a single duration of 500 ms). Increasing stimulus contrast resulted not only in an increase in peak response magnitude (**Fig. 4*A***, left), but the response shape also changed in other ways independent of the peak magnitude, showing increased time-to-peak with lower contrast (phase delay) and a relatively less

pronounced transient relative to the sustained level response (**Fig. 4*A***, right). This result differs from the descriptive model in Albrecht et al., (2002) in which the effect of contrast is modeled as a shift and scale in the response shape.

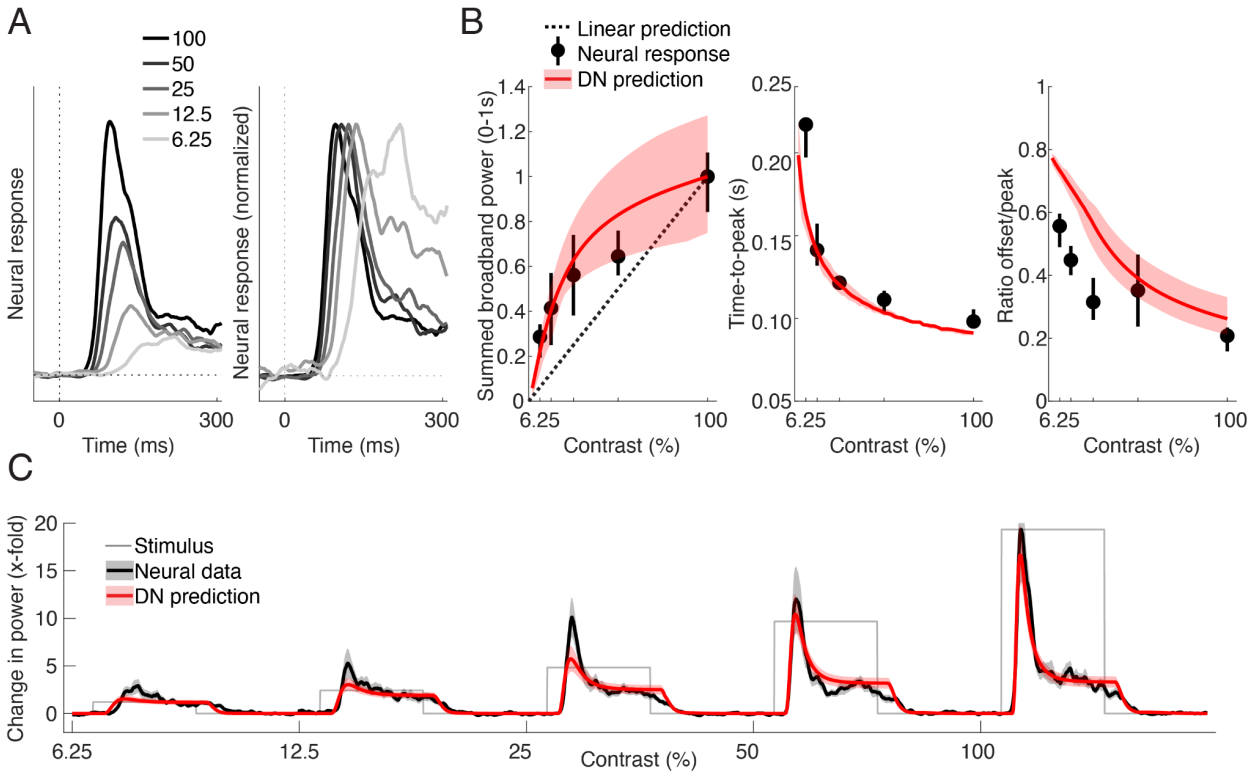


Figure 4: Contrast-dependent temporal dynamics of neuronal responses in human V1.
A) Average ECoG broadband responses in V1 for different levels of contrast (6.25%-100%), superimposed (left) and superimposed and peak-normalized (right). Decreasing stimulus contrast results not only in a decrease in peak amplitude (left), but also a shift in peak latency and a decrease in the ratio of transient to sustained level of response (right). **B**) Summary statistics of contrast-dependent responses derived from individual electrodes (black dots) and DN model predictions (red line). Left: contrast response functions, measured as summed broadband power between 0.05-1s after stimulus onset. Middle: Time to peak. Right: Ratio of sustained response level (response magnitude at stimulus offset) to transient response level (maximum response level). Data points indicate median across probabilistically assigned V1 electrodes. Error bars and shaded regions reflect 68% confidence intervals across 10,000 bootstraps of electrode assignments. **C**) Average ECoG broadband time courses in V1 (black) along with predictions by the DN model (red), computed as in **Fig. 2*B***. This figure is produced using *tde_mkFigure4.m*

We quantified these dynamics by computing three different summary statistics from each electrode's response time course at each contrast level, which each again showed deviations from linearity (**Fig. 4*B***). Summed responses over time give rise to a contrast response function

that progressively increases for higher contrasts (**Fig. 4*B***, left). Estimates of time-to-peak at each contrast level show that peak latency sharply decreases with increasing contrast (**Fig. 4*B***, middle). A comparison of the sustained level response (at stimulus offset) compared to the transient level (peak response) shows that the difference between these two measures decreases as contrast decreases, due to a less pronounced transient (**Fig. 4*B***, right). All three effects are again also qualitatively captured by the DN model, resulting in highly accurate model predictions across all contrast levels (**Fig. 4*C***). The model predictions are not perfect, however; they slightly underestimate the height of the transient part of the response for lower contrasts, resulting in an underprediction of the sustained/transient response ratio (**Fig. 4*B***). This slight model failure parallels the underprediction of the response to a second stimulus at short ISIs.

In sum, the V1 results show that there are several robust non-linear temporal dynamics in ECoG responses in human V1 that are all consistent with a time-dependent normalization, and thus well described by the DN model. We note that the model captures well-known nonlinear phenomena (e.g. contrast response function), as well as phenomena that are perhaps less well-characterized in population responses in human visual cortex, such as temporal summation and short-latency visual adaptation.

Contrast and repetition effects lead to similar response reductions in V1

An advantage of testing many stimuli in the same experiments is that we can directly compare the effects of different stimulus manipulations. We showed nonlinearities with respect to contrast manipulations (**Fig. 4**) and with respect to stimulus repetition (**Fig. 3**). Here we ask how the two effects compare, and how the DN model simultaneously accounts for them both.

First, zooming into the broadband time courses contrast responses (same data as in **Fig. 4*A***) at the first 100 ms after stimulus onset clearly shows that lower stimulus contrast leads to more slowly rising responses, and to less distinction between transient and sustained responses (**Fig. 5**, top left). Interestingly, responses to repeated stimuli, after subtracting out the influence of the first stimulus (see Methods) have remarkably similar shapes (**Fig. 5**, top right). The similarity of the effects of both contrast and repetition on the response dynamics is predicted by the DN model (**Fig. 5**, bottom row). Therefore, both the data and model suggest the possibility that V1 responds to a recently viewed stimulus similarly to a novel but low-contrast stimulus. One difference between model prediction and data is that in the predictions, the responses for all

stimuli start rising at the same time (but at different rates), whereas in the data, the onset latency appears to be delayed for low contrast and short repetitions.

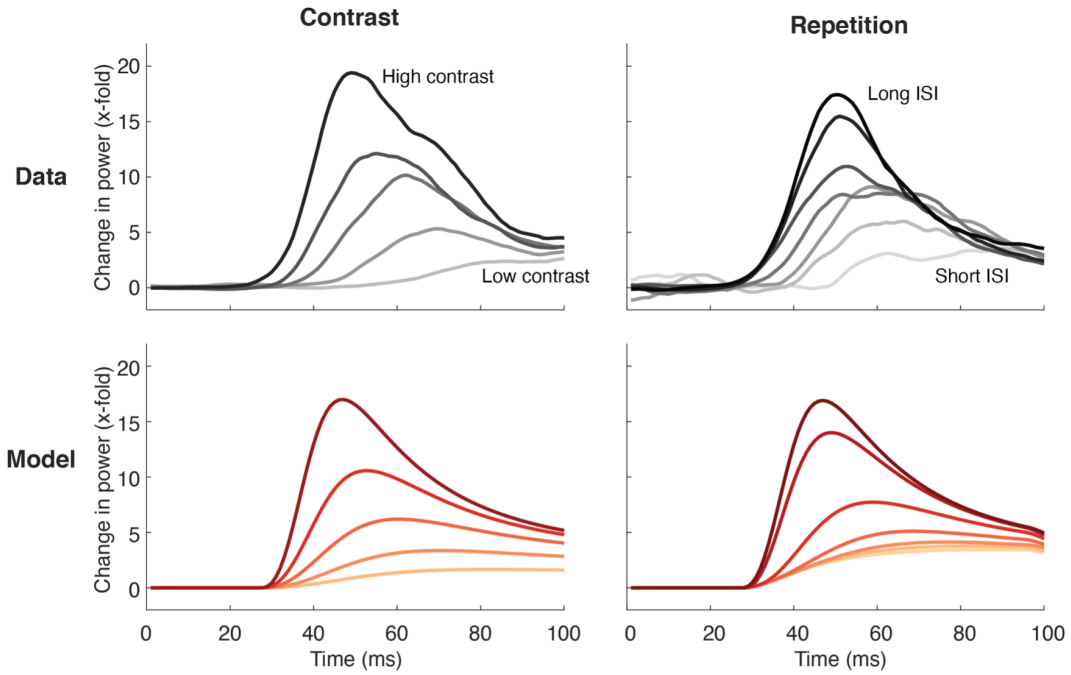


Figure 5: Comparison of contrast and repetition effects in ECoG data and DN model predictions. Top left: Average broadband time courses in V1 for the contrast-varying stimuli during the first 100 ms after stimulus onset. Darker lines indicate increasing contrast. Responses are stronger and rise faster with increasing contrast. Top right: Average broadband time course for the second stimulus in each repetition condition, after subtracting out the response to the first stimulus (see Methods). Darker lines indicate increasing ISI. Responses are stronger and rise faster with longer ISIs. Bottom row: DN model predictions derived using parameters from fitting to the average time courses in top row, for contrast-varying stimuli (left) and repeated stimuli (right). This figure is produced using *tde_mkFigure5.m*.

Inspecting the inner workings of the delayed normalization model helps understand why these responses look so similar and how the model accounts for the two effects simultaneously. In the model, the input drive is represented by the numerator, while the normalization pool is represented by the denominator (see **Fig. 1*D***). In **Figure 6**, these two components of the model are plotted separately for two different contrasts and two different ISIs. This shows that contrast and adaptation effects are both driven by differences in the normalization pool (denominator), but that for contrast, the effect can be attributed to the semi-saturation constant (i.e., left hand of denominator in **Fig. 1*D***), while for repetition, the effect can be attributed to the normalized input drive (i.e., the right hand of the denominator).

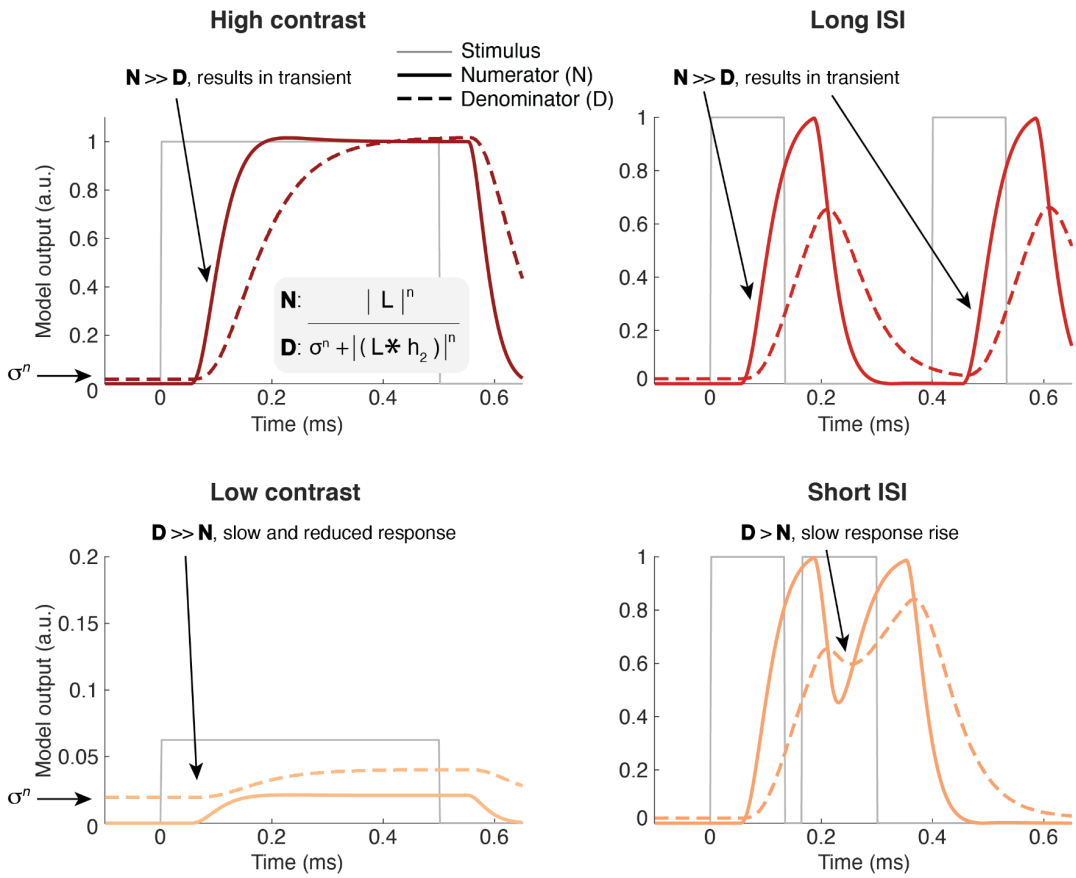


Figure 6. Illustration of how the DN model predicts similar effects of low-contrast and repetition. DN model numerator (solid lines) and denominator (dashed lines) for high contrast (top left), low contrast (bottom left), long ISI (top right) and short ISI (bottom right), for the model parameters fit to the average V1 response. (But note that no gain factor is applied, so the output units are relative to stimulus contrast.) With high contrast, the numerator quickly rises and thereby diminishes the influence of the denominator, which is slightly elevated at onset due to the semi-saturation constant σ^n. With low contrast, the denominator term is bigger relative to the input drive, resulting in a slower reduced response. For long ISIs, the denominator has sufficient time to decrease in between stimuli, but with shorter ISI, the denominator is still increased due to delayed normalization from the first stimulus , leading to a reduced response to the second stimulus. In both cases, the response rises more slowly and response magnitude is repressed due to a larger denominator term. This figure is produced using *tde_mkFigure6.m*.

At high stimulus contrast, the semi-saturation constant is negligible, and the numerator rises faster than the denominator, leading to a transient (**Fig. 6**, top left). At low contrast, the numerator never rises above the minimum possible value of the denominator, which is set by σ^n (**Fig. 6**, bottom left; see Methods and Materials). The main difference between the two contrast conditions thus is the driven response relative to σ^n. For repeated stimuli, the driven response is effectively the same irrespective of ISI, and the denominator rises and decays similarly for both

the first and second stimulus. However, when ISI is long (**Fig. 6**, top right), the denominator has decayed more than when it is short (**Fig. 6**, bottom right). Hence for short ISIs, the ratio of numerator to denominator for the second stimulus is lower. The main difference between the two repetition conditions is thus the level of pre-existing normalization.

In summary, neural responses rise more slowly and are suppressed both when contrast is reduced and when stimuli are repeated with short intervals. According to the DN model, for contrast, this is due to an enhanced left term of the denominator relative to the input drive (e.g. due to background neural activity reflected in σ^n), while for repetition, the reduction is due to an enhanced right term reflecting pre-existing delayed normalization. Thus, the DN model is able to simultaneously predict the effects of both of these stimulus manipulations because it can achieve response reductions through either one of these terms.

Delayed normalization predicts responses throughout visual hierarchy

Next, we examined to what extent each of the canonical computations in the DN model contribute to its ability to predict responses for variations in duration, repetition and contrast. To this end, we computed cross-validated explained variance for a number of reduced versions of the model, starting with a linear model with fully parameterized IRFs (see Methods and Materials), after which each of the non-linear operations in the DN model (rectification, exponentiation, normalization, normalization with delay; see **Fig. 1*D***) are added in turn (**Fig. 7*A***). We find that delayed normalization is a key component in raising the prediction accuracy. Adding non-linear operations to a purely linear model gradually increases the ability to explain responses in V1, and a clear gain in explanatory power is observed when adding the final step of delayed rather than instantaneous normalization.

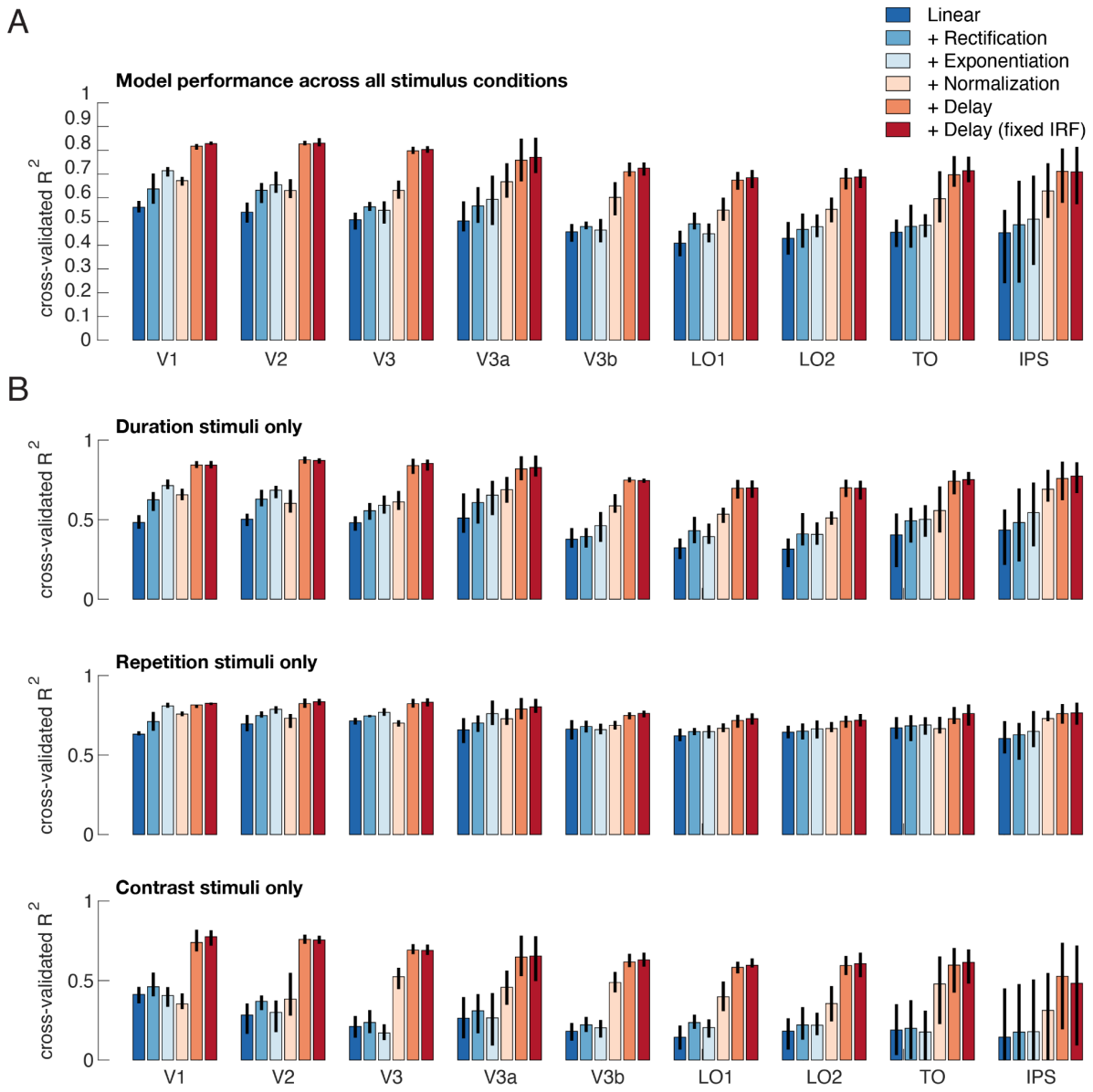


Figure 7: Delayed normalization explains temporal dynamics in multiple visual areas.
A) Cross-validated explained variance (coefficient of determination) across all stimulus conditions in V1-IPS for the DN model along with deconstructed versions of the model from which each of the canonical non-linear computations depicted in **Fig. 1*D*** was removed in turn. Relative to a purely linear model (dark blue), adding non-linear canonical operations of rectification, exponentiation, and delayed normalization each increase the model's ability to capture the variance in the dataset. Moreover, a full delayed normalized model with a constrained IRF (dark red) performs equally well as the more unconstrained IRF (2 additional free parameters, dark orange). **B)** Explained variance of DN model components, separated by stimulus manipulation. Note that the models were fit to all conditions simultaneously, but that model fits here are evaluated separately for the duration, repetition and contrast conditions. Error bars indicate 68% confidence intervals across 10000 bootstraps of electrode assignments. This figure is produced by *tde_mkFigure7.m*.

Importantly, this pattern of results holds not only in V1, but in all retinotopic areas in which we had sufficient electrode coverage, including V2, V3, and higher-order lateral-occipital and parietal-occipital regions. Note that because the accuracy is computed on left out data (cross-validation), the result is not guaranteed simply due to adding more free parameters for each model. Indeed, there are some instances where adding parameters makes the fits less accurate (V1, +Exponentiation vs. +Normalization). In general, improvements in model performance are most pronounced for the duration and contrast conditions, with reduced models performing relatively well when predicting responses to stimuli varying in repetition interval (see **Fig. 7*B***).

In addition, **Fig. 7** shows that the version of the delayed normalization model with fully parameterized IRFs is on par with the more constrained DN model (in which the phase delay and time constants of the negative gamma function are fixed) as used in our analyses so far (see Materials and Methods), indicating that the data do not sufficiently constrain the detailed shape of the IRF, and that the chosen IRF constraints are adequate for the current dataset.

Having established that the DN model predicts ECoG responses with high accuracy not only in V1, but in multiple visual regions, we next used it to investigate how temporal dynamics change along the visual hierarchy. Note that we use the term “visual hierarchy” as an approximation. By nearly any metric, V2 is a later area than V1, and V3 later than V2, and all the other areas later than V3. However the hierarchical relationship among the areas beyond V3, if any, is uncertain.

Temporal summation windows increase in higher visual areas

The responses to stimuli that vary in duration revealed a systematic change in temporal summation between visual areas (**Fig. 8**). Compared to V1, responses in, for example, V3b rise more slowly and stay elevated for a longer period of time, resulting in wider responses and higher sustained levels of response relative to the transient (**Fig. 8*A***). We quantified these differences in response shapes across visual areas by computing three different summary metrics, which each capture a different aspect of the response dynamics, on both the data and DN model predictions (see Methods and Materials).

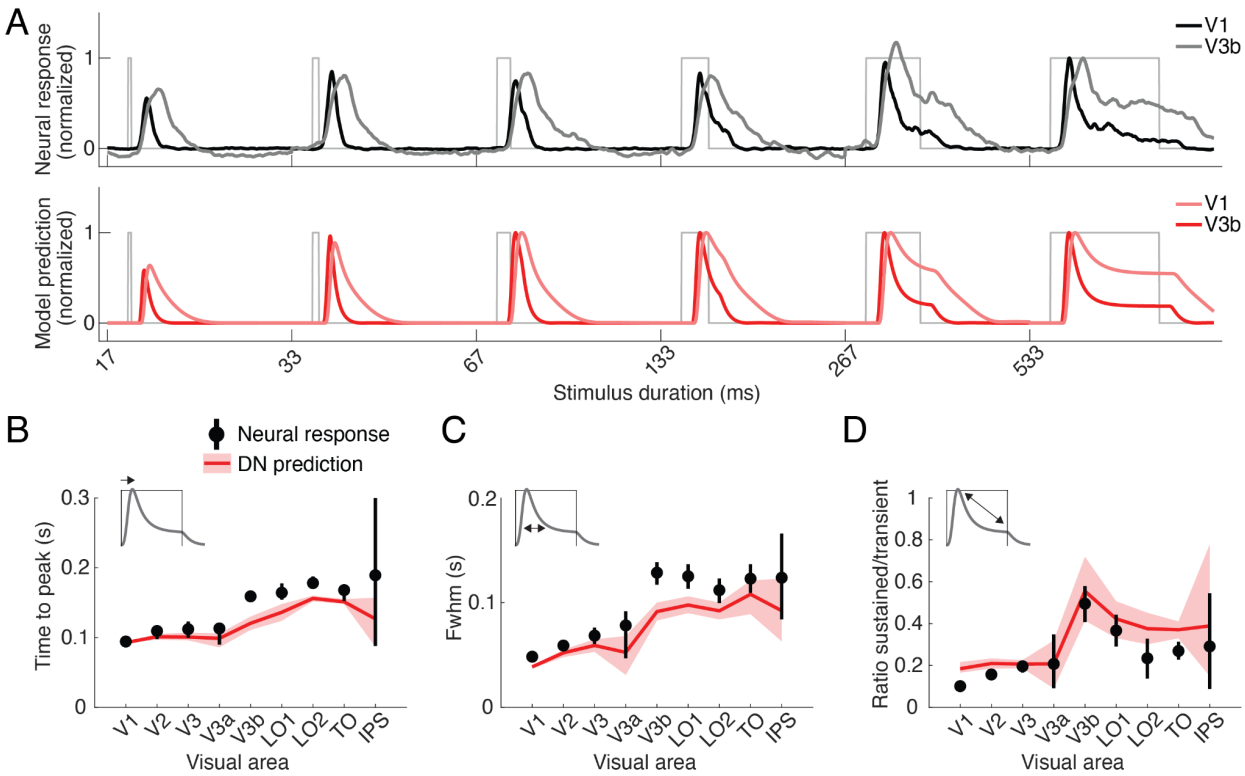


Figure 8: Temporal summation window sizes increase in higher visual areas.
A) Top: Average broadband ECoG responses for stimuli of increasing duration for an early (V1) and higher visual area (V3b). Bottom: DN model predictions for these same data. Compared to V1, responses in V3b rise more slowly and stay up longer, resulting in wider responses and higher sustained levels of response relative to the transient level. **B-D)** Three summary metrics derived from both the neural responses (black) and DN model predictions (red), separately for all visual areas in the dataset. **B)** Time to peak, computed from the neural response and model prediction to the longest duration stimulus (533 ms). **C)** Full-width half max, computed from the neural response and model prediction to the shortest duration stimulus (17 ms). **D)** Ratio of sustained (response level at stimulus offset) relative to the maximum response (peak), computed from the neural response and model prediction to the longest duration stimulus. Data points indicate means and error bars indicate 68% confidence intervals across 1000 response averages and corresponding DN model fits using probabilistically assigned electrodes. Summary statistics computed separately for individual patients are provided in **Extended Data Figure 8-1**. This figure can be reproduced using *tde_mkFigure8.m*.

Relative to V1, all three measures increased in higher visual areas: the slower rise was reflected in increasing time-to-peak (**Fig. 8*B***); the broadening of the transient was reflected in increasing full width at half max (**Fig. 8*C***); and relatively higher sustained activity was reflected in a sustained/transient ratio that tended to become larger (**Fig. 8*D***), although this latter summary metric was more variable across electrodes compared to the first two metrics. Overall, these summary metrics reflecting differences in temporal summation differ most between a group of

earlier visual areas (V1-V3, V3a) and the later areas. These patterns were largely consistent when computed separately within individual patients (see **Extended Data Figure 8-1**).

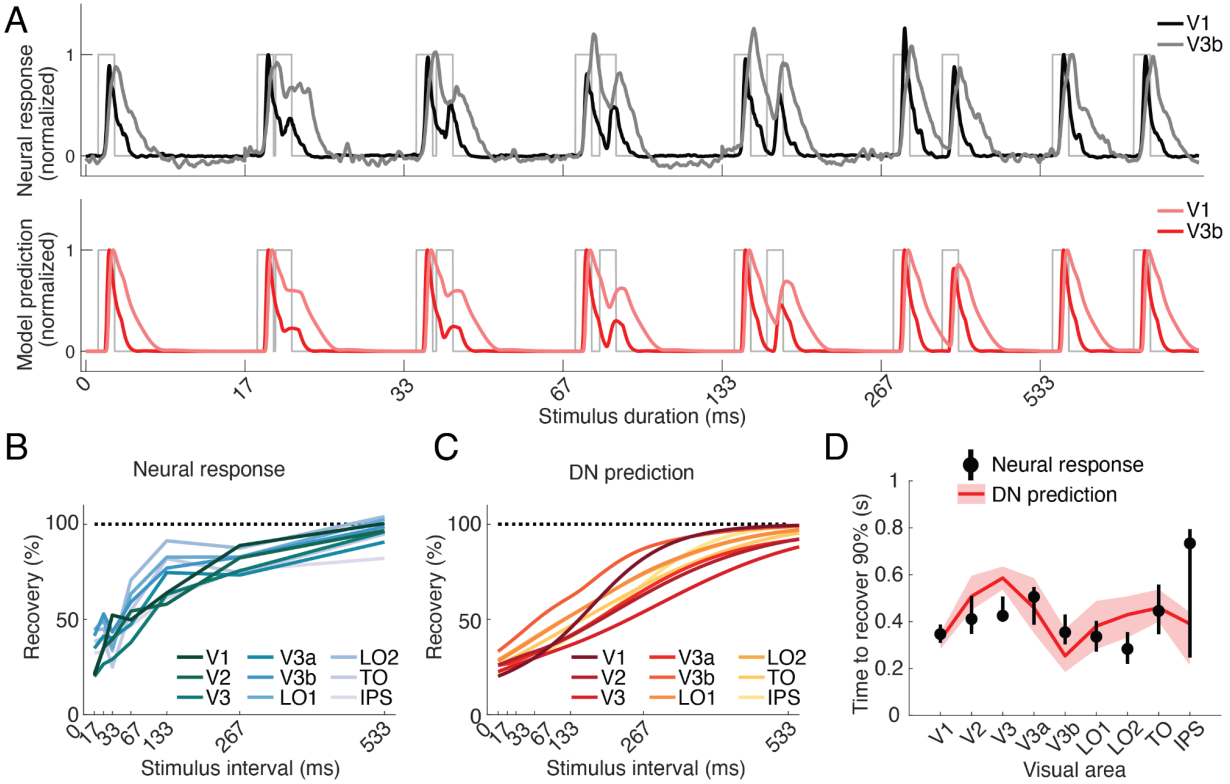


Figure 9: Temporal adaptation windows do not differ systematically between visual areas. **A)** Top: Average broadband ECoG responses to repeated stimuli with increasing ISI for an early (V1) and a higher visual area (Vb). Bottom: DN model predictions corresponding to these same data. Compared to V1, responses in V3b recover at a similar rate, with full recovery being achieved at the longest ISI measured (533ms). **B-D)** Recovery from adaptation across visual areas. While most higher visual areas show a somewhat slower recovery compared to V1 (e.g. V3a) some areas show a similar or faster recovery (e.g. V3b, LO1) and error bars are highly overlapping, suggesting no systematic change in recovery across the visual hierarchy. **B)** Measured recovery from adaptation across ISIs for V1-V3b (calculated the same way as in **Fig. 3*B***). **C)** Predicted recovery from adaptation based on the DN model when simulating responses across a range of ISIs. **D)** Estimation of the ISI at which recovery is 90% of the response to the first stimulus, calculated for both neural data and DN model. Data points indicate means and error bars indicate 68% confidence intervals across 1000 response averages and corresponding DN model fits using probabilistically assigned electrodes. This figure is produced using *tde_mkFigure9.m*.

Temporal adaptation windows do not differ systematically between visual areas

In contrast to the pattern with temporal summation, adaptation to repeated stimuli showed less evidence of varying systematically across areas (**Fig. 9**). While the response to the second

stimulus in V3b is on average higher than in V1 (**Fig. 9*A***), this appears to result from the continued response to the first stimulus even after the second stimulus onset in V3b, rather than to less adaptation (as observed in **Fig. 8**) or a systematically different rate of recovery. In both areas, full recovery was achieved by the longest ISI measured (533 ms). Consistent with these observations, we did not observe a clear relation between the rate of recovery from adaptation across visual areas. To quantify this, we expressed the magnitude of the response to the second stimulus relative to the first stimulus separately for each ISI and visual area, yielding a measure of recovery from adaptation over time per area. While areas V2, V3 and V3a recovered somewhat more slowly than V1 in both the data (**Fig. 9*B***) and DN model predictions (**Fig. 9*C***), area V3b did not. Indeed, when comparing the estimated ISI at which the response is nearly fully recovered (90%) for all areas shows that, for both the data and the model, no systematic change in this measure is observed, with several higher-order visual regions having a similar recovery level as V1 (**Fig. 9*D***).

Opposite effects of contrast on response amplitude and latency across visual areas

In a third comparison of visual areas, we examined responses as a function of contrast (**Fig. 10**). Interestingly, we observe opposite effects of contrast on response amplitude and timing across visual areas. Amplitude differences between low and high contrast stimuli become smaller in higher-visual areas (greater invariance to contrast), whereas time-to-peak differences tend to get larger (greater sensitivity to contrast).

The amplitude effects can be seen by comparing responses across different contrast conditions within a given area: compared to V1, responses in V3b are relatively enhanced for low contrast stimuli (**Fig. 10*A***); similarly, compared to V2, responses in TO are relatively enhanced for low contrast stimuli (**Fig. 10*B***). We quantified this pattern by fitting a Naka-Rushton equation to the peak responses for both the data and the DN model predictions (see Methods and Materials). The C50 parameter is the contrast at which the response reaches half its maximum, with lower values indicating steeper contrast response functions. This measure showed a downwards trend along the visual hierarchy in both the data and the model predictions (**Fig. 10*C***), indicating that less contrast is needed to elicit a robust response in higher compared to lower visual areas (i.e. responses become more contrast invariant). While the data and model predictions agree in showing decreasing C50 in higher areas, there is an overall offset between data and model, with the model predicting slightly higher C50 parameters overall. This matches our previous

observation of small but systematic underpredictions of the transient response at low contrast noted in **Fig. 4*C***.

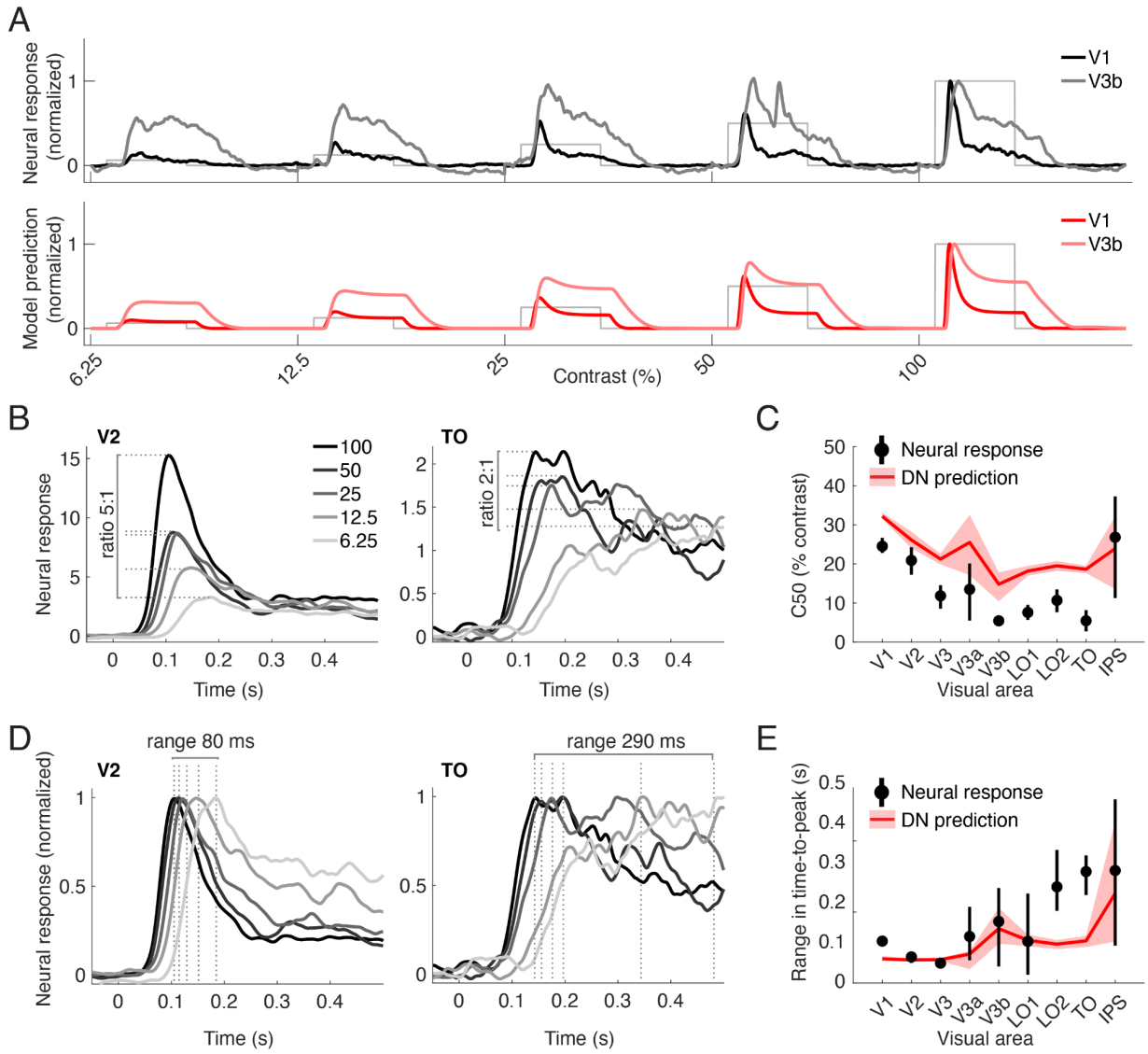


Figure 10: Opposite effects of contrast on response amplitude and timing across visual areas. **A)** Top: Average broadband ECoG responses to stimuli with increasing contrast for an early (V1) and a higher visual area (V3b). Bottom: DN model predictions corresponding to these same data. **B)** Average broadband ECoG responses to stimuli with different contrasts in early visual area V2 (left) and higher visual area TO (right), superimposed. Lines indicate maximum amplitude in each condition. Time courses were slightly smoothed for illustrative purposes only. **C)** C50 for each visual area, computed both from the data and model predictions. **D)** Same data as in B), with each time course normalized to its peak, which illustrates differences in response rise. Lines indicate time-to-peak in each condition. **E)** Range in time-to-peak across contrast for each visual area for both data and DN model predictions. Data points in C) and E) indicate means and error bars indicate 68% confidence intervals for 1000 response averages of probabilistically assigned electrodes. This figure is produced using *tde_mkFigure10.m*.

Differences in the time to peak between low and high contrast stimuli are *more* pronounced in the highest visual areas. This is illustrated in **Fig. 10*D***, which shows the same data as in **Fig. 10*B*** but with each condition normalized to its peak (see **Fig. 4*A***, right panel for same data in V1). Both V2 and TO show a greater and shallower slope at lower contrast (**Fig. 10*D***). However, in area TO the range in latencies is larger compared to V2. To quantify these effects, we calculated the time-to-peak for each contrast (i.e., the dashed vertical lines in **Fig. 10*D***) and measured the range (difference between minimum and maximum value) across contrast levels in each area, for both the data and the model (**Fig. 10*E***). While the range is relatively constant in early areas V1-V3, both data and model show an upwards trend in the time-to-peak range across contrast conditions in the highest visual areas. This suggests that, unlike response amplitude, response latency becomes more sensitive to contrast in higher visual areas.

Changes in temporal dynamics across visual areas

To summarize the comparisons across areas, we find that the temporal dynamics of neural responses differ across the visual areas we measured in several different ways. Most changes seem to reflect increasing invariance in higher visual areas, demonstrating, for example, less dramatic effects of changing stimulus duration on neural response shapes, or less effect of changing contrast intensity on response amplitude, compared to lower visual areas. However, not all of our analyses were indicative of increased invariance along the visual hierarchy. In particular, recovery from adaptation remained fairly stable, suggesting that the rate of adaptation does not change systematically across areas, or that our measurements were not sensitive enough to pick up differences in this response property. (We return to this in the Discussion). In addition, response latency becomes more sensitive to stimulus contrast in higher visual areas. Importantly, changes in response properties between low and higher visual areas were to a large extent recapitulated by the DN model, suggesting that the model captures temporal response dynamics not just in early visual regions but across visual cortex more broadly.

Faster temporal dynamics for peripheral versus foveal visual field locations

In a final set of analyses, we explored differences in temporal dynamics of neural responses dependent on retinotopic tuning. This is motivated by neural and psychophysical studies reporting differences in temporal dynamics dependent on eccentricity (McKee and Taylor, 1984;

Horiguchi et al., 2009; Sinha et al., 2017; Stigliani et al., 2017; Himmelberg and Wade, 2019). A population receptive field (pRF) was estimated using independent data collected in the same session (see Methods and Materials). We verified that the resulting pRF estimates were consistent with known retinotopic organization of visual cortex, such as being positioned in the visual field contralateral to the recorded hemisphere, and showing increasing pRF size along the visual hierarchy (see **Extended Data Fig. 11-1**). Using the estimated eccentricity of the centers of each of these pRFs, we separated electrodes into two different groups: foveal (<2 degrees of visual angle) and peripheral (>2 degrees and <16.6 degrees, which was twice the stimulus extent in the main experiment). Comparison of neural response time courses across all conditions (**Fig. 11*A***) reveals subtle differences between foveal and peripheral electrode groups for V1-V3, and more pronounced differences between foveal and peripheral electrodes for higher areas. In higher visual regions (bottom row), foveal electrodes exhibit higher responses than peripheral electrodes in the duration and repetition conditions, and for these conditions, responses in peripheral electrodes decay faster (are more transient). In early visual cortex (top row), the responses for low-contrast stimuli are higher in foveal than peripheral electrodes.

To quantify these effects, we compute summary statistics reflecting different aspects of the temporal dynamics (similar to those used in **Fig. 8*B-D***). In particular, we used the DN model predictions to compute time-to-peak, fwhm and the sustained/transient ratio separately for foveal and peripheral electrodes. As shown in **Fig. 11*B-D***, time-to-peak and fwhm were consistently higher for foveal compared to peripheral electrodes, with more pronounced differences in higher areas than V1-V3. Statistical comparison of bootstrapped median parameter values between electrode groups indeed indicated significant differences for time-to-peak ($p=0.03$) and a trend for full width at half max ($p=0.07$) in higher visual regions, but not early regions (all $p>0.21$). The sustained/transient ratio did not differ between the more foveal and more peripheral electrodes in either group (all $p>0.36$). Because the placement of electrodes was not under experimental control, we do not equally sample all visual areas foveally and peripherally, adding potential sources of sampling noise to the data. Despite this considerable limitation, there are trends that are broadly consistent with prior findings distinguishing between foveal and peripheral temporal dynamics.

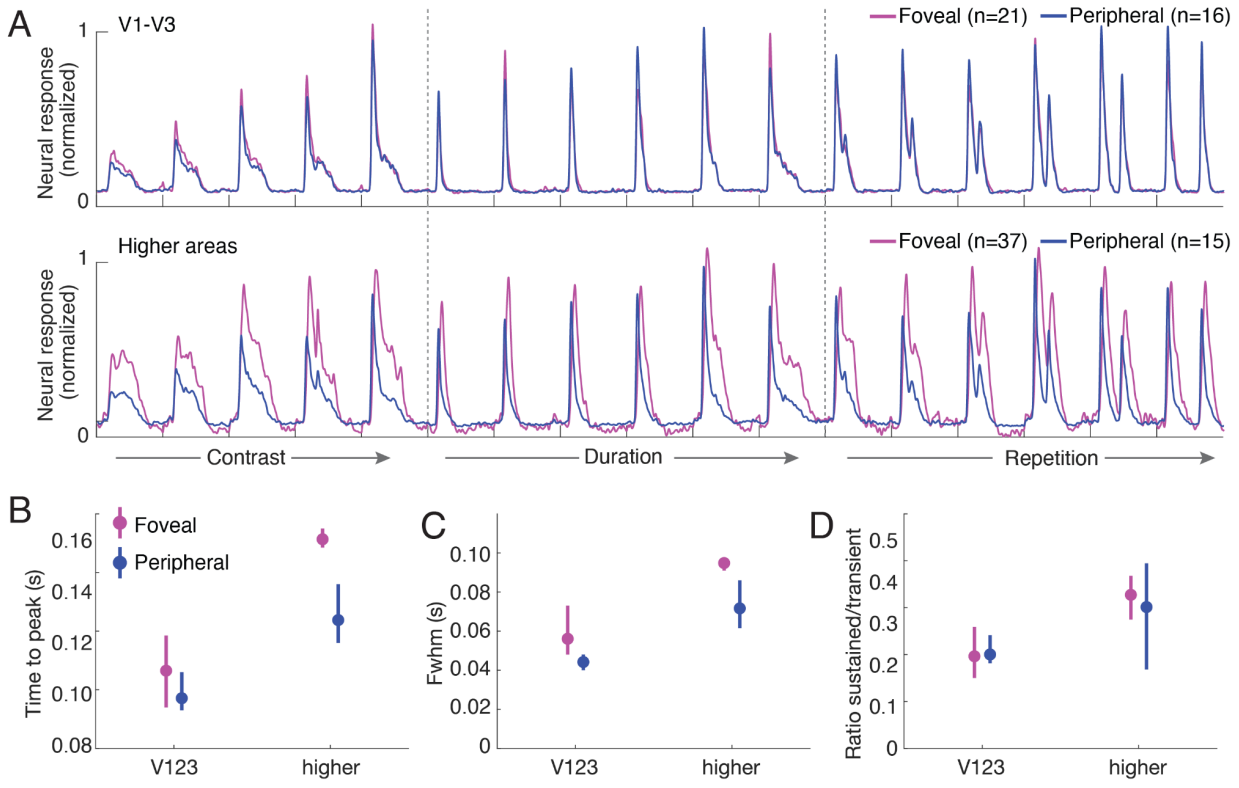


Figure 11: Temporal dynamics of neural responses as a function of pRF location.
A) Average broadband ECoG responses across all conditions separately for foveal (magenta) and peripheral (blue) electrodes in areas V1-V3 (top row) and higher visual areas (bottom row). Responses are ordered by the contrast-varying conditions (left; contrast increasing from left to right), duration-varying conditions (middle; duration increases from left to right) and repetition conditions (right; ISI increases from left to right). Foveal electrodes tend to exhibit stronger responses and slower temporal dynamics, especially in higher visual areas. Legends indicate how many electrodes contributed to the average. Note that due to probabilistic assignment, electrodes can contribute to both the V1-3 average and the higher visual area average (see Methods and Materials). **B-D)** Three summary metrics derived from the DN model predictions separately for foveal and peripheral electrodes. **B)**: Time to peak, computed from the simulated response to a long duration stimulus (2s). **C)**: Full-width half max, computed from model prediction to a short duration stimulus (5 ms). **D)**: Ratio of sustained (response level at stimulus offset) relative to the maximum response (transient peak), computed from the DN model prediction to a long duration stimulus (2s). Data points indicate median across probabilistically assigned electrodes to V1-V2-V3 or higher visual areas. Error bars indicate 68% confidence intervals across 1000 bootstraps of electrode assignments. This figure is produced using *tde_mkFigure11.m*.

Discussion

We used three different experimental manipulations to investigate the temporal dynamics of neural responses in human visual cortex. Having multiple manipulations within the same participants and electrodes allowed us to systematically compare how different stimulus properties affect neural responses, and how response properties differ across visual areas.

Fitting a single model to many stimuli

The multiple, controlled stimulus manipulations enabled us to demonstrate several non-linearities in the neural dynamics of visual cortex: responses increase more gradually as contrast and stimulus duration increase (saturation), the response to a repeated stimulus is suppressed for short intervals and recovers for longer intervals (adaptation), and responses slow at lower contrast (phase delay). While each of these phenomena has been shown with other kinds of experiments (e.g., Albrecht and Hamilton, 1982; Albrecht et al., 2002; Zhou et al., 2018); an important goal in systems neuroscience is to understand how apparently disparate phenomena might be linked (Wandell et al., 2015). Here we showed that a delayed normalization model consisting of a small set of canonical neural computations predicts these responses with high accuracy in early visual areas and parts of higher-level visual cortex.

This model was previously developed to account for similar temporal phenomena in separate, independent studies (compiled and analyzed by Zhou et al., 2019). While that work showed that a single model *form* could account for a wide variety of neural data, a single model (i.e., including the same parameters) was not fit to the various data sets because they came from different experiments, measurement types, and species, including single unit macaque firiting rates, human fMRI, and human ECoG. For example, the time to peak of the single unit data was tens of ms, whereas for the ECoG broadband it was over 100 ms even in V1, the cortical visual area with the shortest response latency. This leaves open the possibility that differences in parameter fits are found for different phenomena because the phenomena have different causes, and cannot be explained by a single, simple model. Thus, a key advance over this prior work is that by fitting the same model *simultaneously* to all three conditions, we show that, for a given site on cortex, a single instance of our delayed normalization model captures temporal

dynamics arising from different types of stimulus manipulations, thereby simultaneously accounting for temporal summation, adaptation, and slower dynamics at low contrast.

Parallels between responses to repeated stimuli and low contrast stimuli

A surprising observation is that the response time course to a repeated stimulus at high contrast is remarkably similar to the time course of a non-repeated stimulus at low contrast. These two very different stimulus manipulations delay the responses, reduce the overall amplitude, and change the response shape (the early transient is reduced more than the later sustained level). Having a single model offers a possible explanation for the similarities. In particular, the delayed normalization model accounts for these patterns in terms of the time-varying relationship between stimulus drive and normalization: At low contrast, the stimulus drive never gets large relative to the normalization, and for repeated stimuli, there is lingering normalization from the first stimulus.

The similarity between the two types of responses does not imply that the two stimulus dimensions are perfectly conflated in the brain. A recent study in inferotemporal cortex in macaque monkey (Mehrpour et al., 2020) shows that despite the fact that contrast and repetition both reduce neural population responses, monkeys are still able to behaviorally distinguish repeated stimuli independent of their contrast level. These results indicate that there is some memory information independent of contrast in the population responses. Nonetheless, the parallels we observe in the ECoG responses raise the prospect that behavioral responses, for example detection thresholds or response time, will show similar patterns for the two stimulus manipulations, and might be explained by a single model. Such a model could potentially help understand how stimulus properties such as contrast interact with cognitive phenomena like the attentional blink (Raymond et al., 1992). A dynamic normalization model has recently been used to explain the behavioral effects of temporal attention (Denison et al., 2021). Our findings suggest that a dynamic normalization model might also explain how effects of temporal attention interact with stimulus contrast.

Links between ECoG neural measures, fMRI and single-unit electrophysiology

We quantified all the ECoG responses as the time-varying broadband envelope. Prior work has shown that the response patterns can differ dramatically depending on how the ECoG response is quantified. For example, the visual steady state evoked potential shows approximately linear

spatial summation whereas the broadband response is sub-additive (Winawer et al., 2013); Narrowband gamma oscillations and broadband power elevations show quite different stimulus selectivity (Ray and Maunsell, 2011; Hermes et al., 2015, 2019). We focused on the broadband response in this study because it appears to be most strongly connected to other measures of interest, including fMRI (Hermes et al., 2017) and single- and multi-unit spiking (Miller et al., 2009; Ray and Maunsell, 2011), and because of its link to function across cortical areas and behavior (Miller et al., 2014). The relationships of the ECoG broadband response to both the fMRI signal and to spiking data facilitates comparisons between our data and the literatures from these other domains.

In fact, the temporal phenomena we report here have been demonstrated before with fMRI and single-unit physiology. However, these methods leave open questions that are better addressed with our ECoG data. For example, sub-additive temporal summation of responses to visual stimuli increasing in duration has been observed in human visual cortex with fMRI (Boynton et al., 1996; Huettel and McCarthy, 2000; Birn et al., 2001; Mukamel, 2004; Stigliani et al., 2017; Zhou et al., 2018). These effects could result from a mixture of non-linear summation in the BOLD signal and the underlying neural responses. Here, we clearly show that temporal subadditivity is present at the level of neural population responses in human ECoG, paralleling single-cell recordings from early visual primate cortex (Tolhurst et al., 1980). Similarly, adaptation of visual responses resulting from long- and short-lagged repetitions has been widely studied using fMRI (Grill-Spector et al., 2006). However, due to the sluggish nature of the BOLD response, responses to stimuli separated by short intervals cannot be estimated independently. In contrast, our ECoG data allowed for separate estimations of the time course of the first and second stimulus. Consistent with single-cell recordings in animals (Motter, 2006), we find that adaptation is not only associated with a reduction of the amplitude of the response, but also a delay of the rise (slower response dynamics). The ECoG measures also provide important data not easily obtained in animal studies. We studied responses to the same stimulus manipulations across many visual areas, which is typically not feasible in microelectrode recordings, enabling us to systematically compare temporal dynamics across the visual pathways.

Changes in sensitivity and invariance across the visual pathways

In comparing responses and model fits across visual areas, we focused on two questions: Do temporal windows become longer in higher visual areas? Do responses become more invariant

to stimulus contrast? The data and the model predictions provide insight into these questions, showing that for each property–temporal windows and contrast responses–the answers are more subtle and interesting than a simple yes or no.

A number of reports using different methodologies have found that the time-scale of temporal processing becomes increasingly long ascending the visual hierarchy (Hasson et al., 2008; Weiner et al., 2010; Honey et al., 2012; Mattar et al., 2016; Zhou et al., 2018), and cortical hierarchies more generally (Murray et al., 2014), paralleling the finding that spatial receptive fields become larger in higher areas (Maunsell and Newsome, 1987). In contrast, a recent fMRI study reported no differences in the recovery time of repetition suppression across visual areas (Fritsche et al., 2020). Here, we replicate both patterns, showing evidence for longer temporal windows by some metrics but not others, consistent with the interpretation that there is not a single processing time-scale per area. When temporal dynamics are characterized by the period over which responses sum (akin to a temporal receptive field), there are indeed systematic increases in time scale from earlier to later visual areas: compared to V1-V3, responses in higher visual areas rose more slowly, the transient part of the response decayed less rapidly, and the sustained part of the response was less reduced relative to the peak magnitude (Figure 8). However, when characterized as the recovery time in a repetition suppression paradigm, we find the time scales to be relatively constant across many visual areas, consistent with Fritsche et al. (2020). These patterns are not contradictory, as both empirical observations are also found in the model fits, indicating that even a relatively simple model (a few canonical neuronal computations–filtering, exponentiation, and normalization) can result in complex behavior.

Another encoding property that tends to vary across visual areas is contrast sensitivity, with a tendency for higher visual areas to show similar responses for low and high contrast stimuli (greater invariance) (Tootell et al., 1998; Avidan et al., 2002; Lu and Roe, 2007). Interestingly, as with temporal windows, we find that the degree of invariance depends on the quantification. Specifically, we find that response amplitudes become more invariant in higher visual areas, but response latencies do not; if anything, higher areas show a greater latency difference between low and high contrast than do early areas. The fact that sensitivity to contrast remains in the timing of the responses but not the amplitudes means that measurements which average across trials (like fMRI or mean spike rates) will tend to miss this feature of the responses. It also means that information about stimulus contrast remains in higher visual areas, and potentially

the temporal dynamics of the response in higher visual areas could be used to infer the stimulus contrast.

The fact that the same model form accurately fits responses in several cortical areas does not imply that each area is computing the same thing. Rather, it implies that some aspects of the responses are captured by a common model form (though differing in parameters). There is in fact evidence that cortical areas with very different stimulus selectivity, V1 and MT, manifest a similar computational architecture, differing primarily in their inputs rather than computations (Heeger et al., 1996).

Space and time

Some of the observations about changes in dynamics across visual areas parallel findings in the spatial domain, for example spatial receptive fields and temporal summation windows increase in higher visual areas. However, it is unlikely that spatial and temporal properties are perfectly separable. To make the experiments and analyses tractable, we kept the spatial properties of the stimuli constant (other than the contrast level). The choice of stimuli–textures with first and second order contrast and with many orientations–was made because these stimuli were shown to be effective in driving responses in multiple visual areas, especially V1-hV4 (Kay et al., 2013b; Zhou et al., 2018). We limited our analyses here to visual areas defined by retinotopic mapping. Higher visual areas, however, even those with retinotopic maps, are likely more sensitive to more complex and meaningful stimuli (Sayres and Grill-Spector, 2008; Arcaro et al., 2009; Silson et al., 2016). It is possible that some of the observed patterns, for example recovery time for repeated images, would differ using stimuli that are optimal for particular brain areas. Indeed, in a recent study using images of natural scenes, we find that the ECoG broadband responses in ventral temporal and lateral-occipital areas are several times larger than for simple textures, and that the amount of adaptation depends on the area's preferred stimulus category (faces, objects, scenes; Brands et al., 2021). This implies that a complete characterization of temporal dynamics is not possible without also accounting for spatial properties of the stimuli (and *vice versa*). Therefore, an important outstanding goal for future work is to develop a space-time model that simultaneously accounts for spatial summation, stimulus selectivity and temporal dynamics throughout the visual cortex.

Model failures and limitations

While the DN model was able to capture many response variations, we also noted some model failures. In particular, the DN model underpredicts the transient level response for low contrasts; similarly, it underpredicts the peak response magnitude of a second stimulus presented at short ISIs. These failures are small but systematic, suggesting that additional model components are needed to fully explain the data. The DN model accounts for both transient and sustained responses by the temporal dynamics of the driven responses (numerator) and normalization (denominator). Other models account for transient and sustained responses by including multiple channels, one more sustained and one more transient, in the driven responses (Horiguchi et al., 2009; Stigliani et al., 2017). While these models fail to account for the slower dynamics at low contrast, incorporating multiple channels into the driven responses might help resolve discrepancies between model predictions and data at low contrast and short ISIs.

Our results are subject to several limitations. First, we did not have experimental control over electrode placements over eccentricity and our stimuli extended over a limited part of the visual field (8.3º eccentricity). Prior work has shown that the largest differences in temporal dynamics as a function of eccentricity are found when comparing data across a much larger eccentricity range (out to 60º; Horiguchi et al., 2009). Hence the relatively small variation in dynamics we observe across eccentricity may be a consequence of limited range. Second, visual area was not fully crossed with subject, such that not all subjects contributed data to all visual areas. When relatively small differences in response properties or model parameters are expected between areas (say, V1 vs. V2), the variability across subjects may be larger than the differences between areas. For this reason, we avoided drawing strong conclusions about comparisons between each individual visual area, but instead focused primarily on differences between earlier areas (V1-V3) and later areas. Third, there is some uncertainty in the assignment of electrodes to visual areas as we did not measure retinotopic maps in individual subjects. For this reason, we developed a probabilistic resampling method to incorporate this uncertainty into our measures of variability. Fourth, our dataset did not include electrodes with sufficiently reliable visual responses in ventral stream regions beyond V3v; it is possible that visual responses in those regions exhibit temporal dynamics that are different from higher-order lateral-occipital and dorsal-parietal maps (Stigliani et al., 2019), that may or may not be well captured by the DN model. Thus, an important direction for future work is to precisely measure

temporal dynamics in neural population responses with more extensive sampling of higher-order visual regions, especially in ventral visual cortex.

Finally, it is important to compare the performance of the DN model to other models of temporal dynamics, either those that have been previously proposed in the literature (e.g. two-channel models; Horiguchi et al., 2009; Stigliani et al., 2017, 2019) or potential future models. To facilitate such comparisons, we make our dataset publicly available, and along with our data, we provide an open code repository that allows to implement models in a modular fashion, allowing for direct comparison of multiple models fit to the same data and analyzed with the same methods in a standardized manner, including automated cross-validation and visual area comparison.

References

Acerbi L, Ma WJ (2017) Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search. Adv Neural Inf Process Syst 30:1834–1844.

Albrecht DG, Geisler WS, Frazor RA, Crane AM (2002) Visual Cortex Neurons of Monkeys and Cats: Temporal Dynamics of the Contrast Response Function. J Neurophysiol 88:888–913.

Albrecht DG, Hamilton DB (1982) Striate cortex of monkey and cat: contrast response function. J Neurophysiol 48:217–237.

Arcaro MJ, McMains SA, Singer BD, Kastner S (2009) Retinotopic Organization of Human Ventral Visual Cortex. J Neurosci 29:10638–10652.

Avidan G, Harel M, Hendler T, Ben-Bashat D, Zohary E, Malach R (2002) Contrast Sensitivity in Human Visual Areas and Its Relationship to Object Recognition. J Neurophysiol 87:3102–3116.

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 Einhäuser W, ed. PLoS Comput Biol 10:e1003538.

Benson NC, Winawer J (2018) Bayesian analysis of retinotopic maps. eLife 7:e40224.

Birn RM, Saad ZS, Bandettini PA (2001) Spatial Heterogeneity of the Nonlinear Dynamics in the FMRI BOLD Response. NeuroImage 14:817–826.

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.

Brands AM, Devore S, Devinsky O, Doyle WK, Flinker A, Winawer J, Groen IIA (2021) Adaptation of neural responses to naturalistic visual categories in low- and high visual cortex. In: European Conference on Visual Perception.

Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51–62.

Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV (2012) Neural population dynamics during reaching. Nature 487:51–56.

Denison RN, Carrasco M, Heeger DJ (2021) A dynamic normalization model of temporal attention. Nat Hum Behav Available at: http://www.nature.com/articles/s41562-021-01129-1
Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. NeuroImage 39:647–660.
Fritsche M, Lawrence SJD, de Lange FP (2020) Temporal tuning of repetition suppression across the visual cortex. J Neurophysiol 123:224–233.
Gold JI, Shadlen MN (2007) The Neural Basis of Decision Making. Annu Rev Neurosci 30:535–574.
Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10:14–23.
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.
Heeger DJ (1992) Normalization of cell responses in cat striate cortex. Vis Neurosci 9:181–197.
Heeger DJ (2017) Theory of cortical function. Proc Natl Acad Sci 114:1773–1782.
Heeger DJ, Simoncelli EP, Movshon JA (1996) Computational models of cortical visual processing. Proc Natl Acad Sci 93:623–627.
Hermes D, Miller KJ, Noordmans HJ, Vansteensel MJ, Ramsey NF (2010) Automated electrocorticographic electrode localization on individually rendered brain surfaces. J Neurosci Methods 185:293–298.
Hermes D, Miller KJ, Wandell BA, Winawer J (2015) Stimulus Dependence of Gamma Oscillations in Human Visual Cortex. Cereb Cortex 25:2951–2959.
Hermes D, Nguyen M, Winawer J (2017) Neuronal synchrony and the relation between the blood-oxygen-level dependent response and the local field potential Kohn A, ed. PLOS Biol 15:e2001461.
Hermes D, Petridou N, Kay KN, Winawer J (2019) An image-computable model for the stimulus selectivity of gamma oscillations. eLife 8:e47035.
Himmelberg MM, Wade AR (2019) Eccentricity-dependent temporal contrast tuning in human visual cortex measured with fMRI. NeuroImage 184:462–474.
Holdgraf C et al. (2019) iEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology. Sci Data 6:102.
Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, Doyle WK, Rubin N, Heeger DJ, Hasson U (2012) Slow Cortical Dynamics and the Accumulation of Information over Long Timescales. Neuron 76:423–434.
Horiguchi H, Nakadomari S, Misaki M, Wandell BA (2009) Two temporal channels in human V1 identified using fMRI. NeuroImage 47:273–280.
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.
Kay KN, Winawer J, Mezer A, Wandell BA (2013a) Compressive spatial summation in human visual cortex. J Neurophysiol 110:481–494.
Kay KN, Winawer J, Rokem A, Mezer A, Wandell BA (2013b) A Two-Stage Cascade Model of BOLD Responses in Human Visual Cortex Diedrichsen J, ed. PLoS Comput Biol 9:e1003079.
Kubilius J, Schrimpf M, Kar K, Hong H, Majaj NJ, Rajalingham R, Issa EB, Bashivan P, Prescott-Roy J, Schmidt K, Nayebi A, Bear D, Yamins DLK, DiCarlo JJ (2019) Brain-Like Object Recognition with High-Performing Shallow Recurrent ANNs. Adv Neural Inf Process Syst 32 Available at: http://arxiv.org/abs/1909.06161 [Accessed July 16, 2021].
Lu HD, Roe AW (2007) Optical Imaging of Contrast Response in Macaque Monkey V1 and V2.

Cereb Cortex 17:2675–2695.
Mattar MG, Kahn DA, Thompson-Schill SL, Aguirre GK (2016) Varying Timescales of Stimulus Integration Unite Neural Adaptation and Prototype Formation. Curr Biol 26:1669–1676.
Maunsell JHR, Newsome WT (1987) Visual Processing in Monkey Extrastriate Cortex. Annu Rev Neurosci 10:363–401.
McKee SP, Taylor DG (1984) Discrimination of time: comparison of foveal and peripheral sensitivity. J Opt Soc Am A 1:620.
Mehrpour V, Meyer T, Simoncelli EP, Rust NC (2020) Pinpointing the neural signatures of single-exposure visual recognition memory. Neuroscience. Available at: http://biorxiv.org/lookup/doi/10.1101/2020.07.01.182881
Miller KJ, Honey CJ, Hermes D, Rao RP, denNijs M, Ojemann JG (2014) Broadband changes in the cortical surface potential track activation of functionally diverse neuronal populations. NeuroImage 85:711–720.
Miller KJ, Sorensen LB, Ojemann JG, den Nijs M (2009) Power-Law Scaling in the Brain Surface Electric Potential Sporns O, ed. PLoS Comput Biol 5:e1000609.
Motter BC (2006) Modulation of Transient and Sustained Response Components of V4 Neurons by Temporal Crowding in Flashed Stimulus Sequences. J Neurosci 26:9683–9694.
Mukamel R (2004) Enhanced Temporal Non-linearities in Human Object-related Occipito-temporal Cortex. Cereb Cortex 14:575–585.
Murray JD, Bernacchia A, Freedman DJ, Romo R, Wallis JD, Cai X, Padoa-Schioppa C, Pasternak T, Seo H, Lee D, Wang X-J (2014) A hierarchy of intrinsic timescales across primate cortex. Nat Neurosci 17:1661–1663.
Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput Intell Neurosci 2011:1–9.
Ray S, Maunsell JHR (2011) Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex Ungerleider L, ed. PLoS Biol 9:e1000610.
Raymond J, Shapiro K, Arnell K (1992) Temporary suppression of visual processing in an RSVP task: An attention blink? J Exp Psychol Hum Percept Perform 18:849–860.
Sayres R, Grill-Spector K (2008) Relating Retinotopic and Object-Selective Responses in Human Lateral Occipital Cortex. J Neurophysiol 100:249–267.
Silson EH, Groen IIA, Kravitz DJ, Baker CI (2016) Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. J Vis 16:14.
Sinha R, Hoon M, Baudin J, Okawa H, Wong ROL, Rieke F (2017) Cellular and Circuit Mechanisms Shaping the Perceptual Properties of the Primate Fovea. Cell 168:413-426.e12.
Spoerer CJ, Kietzmann TC, Mehrer J, Charest I, Kriegeskorte N (2020) Recurrent neural networks can explain flexible trading of speed and accuracy in biological vision Isik L, ed. PLOS Comput Biol 16:e1008215.
Stigliani A, Jeska B, Grill-Spector K (2017) Encoding model of temporal processing in human visual cortex. Proc Natl Acad Sci 114:E11047–E11056.
Stigliani A, Jeska B, Grill-Spector K (2019) Differential sustained and transient temporal processing across visual streams Jbabdi S, ed. PLOS Comput Biol 15:e1007011.
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
Tootell RBH, Hadjikhani NK, Vanduffel W, Liu AK, Mendola JD, Sereno MI, Dale AM (1998) Functional analysis of primary visual cortex (V1) in humans. Proc Natl Acad Sci

95:811–817.
Wandell BA, Winawer J, Kay KN (2015) Computational Modeling of Responses in Human Visual Cortex. In: Brain Mapping, pp 651–659. Elsevier.
Wang L, Mruczek REB, Arcaro MJ, Kastner S (2015) Probabilistic Maps of Visual Topography in Human Cortex. Cereb Cortex 25:3911–3931.
Wang X-J (2012) Neural dynamics and circuit mechanisms of decision-making. Curr Opin Neurobiol 22:1039–1046.
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. J Neurophysiol 103:3349–3365.
Winawer J, Kay KN, Foster BL, Rauschecker AM, Parvizi J, Wandell BA (2013) Asynchronous Broadband Signals Are the Principal Source of the BOLD Response in Human Visual Cortex. Curr Biol 23:1145–1153.
Yang AI, Wang X, Doyle WK, Halgren E, Carlson C, Belcher TL, Cash SS, Devinsky O, Thesen T (2012) Localization of dense intracranial electrode arrays using magnetic resonance imaging. NeuroImage 63:157–165.
Zhou J, Benson NC, Kay K, Winawer J (2019) Predicting neuronal dynamics with a delayed gain control model Seidemann E, ed. PLOS Comput Biol 15:e1007484.
Zhou J, Benson NC, Kay KN, Winawer J (2018) Compressive Temporal Summation in Human Visual Cortex. J Neurosci 38:691–709.

Subject	Age	Sex	Site	Trials	PRF	Hemi	Implantation	Coverage	# Visual	Matching areas	# Included
Patient 1	30	F	UMCU	12	4	L	strip	lat-occ, lat-par, vent-temp	6	V3b, TO1	0
Patient 2	18	F	UMCU	6	2	R	strip	med-occ, lat-occ, lat-temp, vent-temp	23	V1v, V1d, V2v, V2d, V3v, V3a, V3d, LO1, IPS0	9
Patient 3	27	M	NYU	12	0	L	grid, strip, depth	lat-par, ant-temp, post-vent, frontal	17	V2d, V3d, V3a, IPS0, IPS2, IPS3	4
Patient 4	28	M	NYU	10	4	R	grid, strip, depth	lat-occ, ant-temp, frontal	5	V3d, V3b, TO1	2
Patient 5	31	F	NYU	12	2	R	2x grid, strip, depth	lat-occ, lat-par, frontal, ant-temp	22	V2v, V3v, V3d, LO1, TO1, IPS2, IPS3	3
Patient 6	19	F	NYU	12	2	L	grid, strip, depth	med-occ, lat-occ, lat-par, ant-temp, frontal	23	V1d, V2v, V3d, V3b, hV4, TO1, IPS0, IPS3	2
Patient 7	41	F	NYU	24	4	L	grid, strip, depth	lat-occ, lat-par, med-occ, ant-temp, frontal	14	V1v, V1d, V2d, V3d, V3a, IPS0, IPS1, IPS2, IPS3	7
Patient 8	47	M	NYU	12	2	R	grid, strip, depth	lat-par, ant-temp	4	LO2, TO1	0
Patient 9	25	M	NYU	24	4	L,R	bilateral grid, strip, depth	lat-occ, lat-par	5	TO1, LO2	0
Patient 10	23	M	NYU	36	6	R	grid, HD grid, strip, depth	lat-occ, lat-par, post-vent	118	V2d, V3d, V3a, V3b, LO1, LO2, TO1, IPS0, IPS2	42
Patient 11	19	F	NYU	36	6	L	grid, HD grid, strip, depth	lat-occ, lat-par, post-vent	96	V3b, LO1, TO1, IPS2	19

Table 1: Overview of patient data included in this dataset.
Columns refer to the following: *Subject*: Subject code in dataset. *Age*: Age of patient at time of recording in years. *Sex*: Patient sex (M = Male, F = Female). *Site*: hospital in which recording took place (NYU: NYU Langone Hospital; UMCU: University Medical Center Utrecht). *Trials*: Number of

trials collected per condition in main experiment. *PRF*: Number of PRF runs collected. *Hemi*: implanted hemisphere (L = left, R = right). *Implantation*: type of electrodes implanted (grid = standard clinical grid; strip = standard clinical strip; depth = depth electrodes; HD grid = high-density grid). *Coverage*: approximate overview of visual cortex covered (lat = lateral, med = medial, vent = ventral, par = parietal, occ = occipital, temp = temporal, ant = anterior, post = posterior; see **Extended Data Figure 1-1** for visualizations per patient). #*Visual*: number of electrodes that had a match with either one of the retinotopic atlases used for initial electrode selection (see Methods and Materials). *Matching areas*: matched retinotopic maps according to the maximum probability atlas by Wang et al., (2014). #*Included*: number of electrodes included in the final dataset, after epoch and electrode selection (see Methods and Materials). Patients for whom none of the electrodes survived selection are indicated in gray. Note that while these patients do not contribute to the results reported in this paper, we include these patients in the overview because alternative selection data pruning and electrode selection methods on the publicly available data may yield other inclusion results.

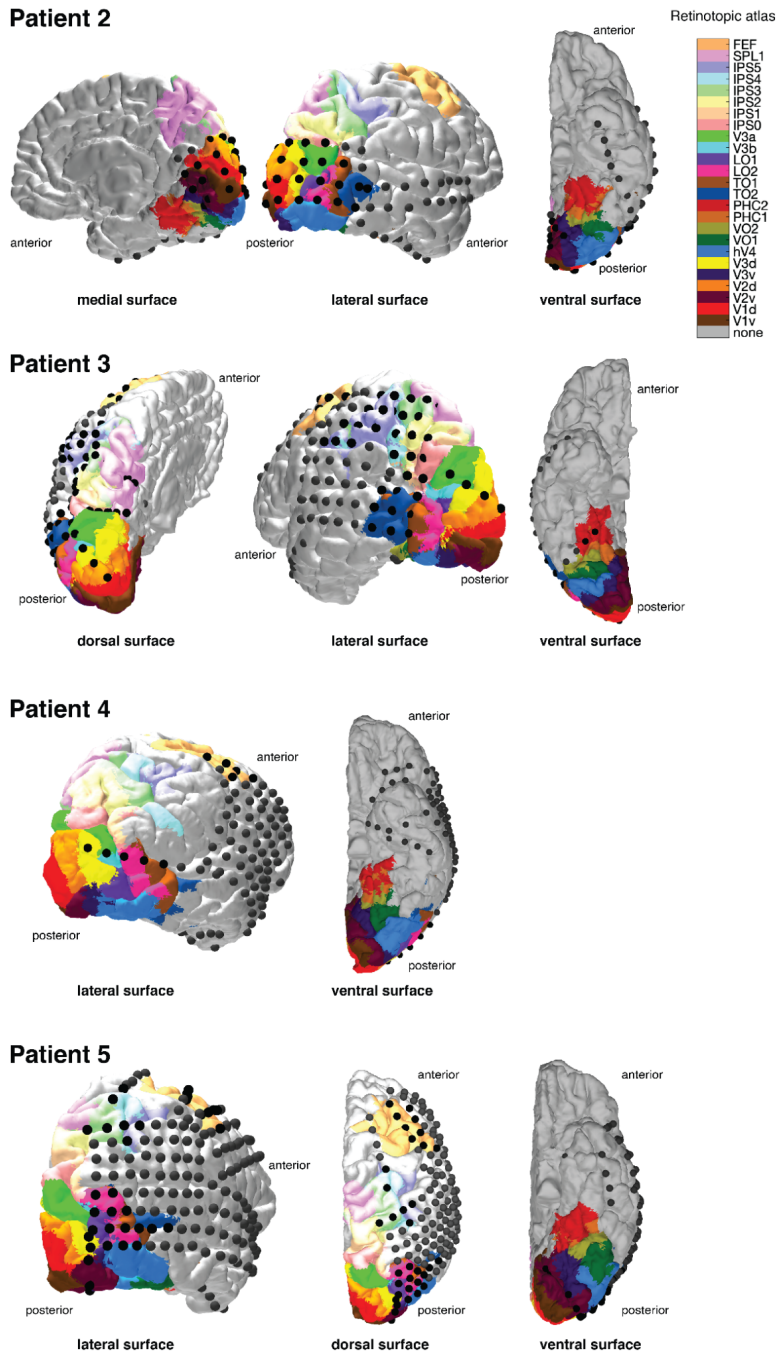
Visual area	**Number of electrodes**	**Contributing patients**
V1	13	patient 2, 3, 7, 10
V2	34	patient 2, 3, 4, 5, 6, 7, 10
V3	50	patient 2, 3, 4, 5, 6, 7, 10, 11
V3a	34	patient 3, 4, 7, 10
V3b	48	patient 3, 4, 7, 10, 11
LO1	64	patient 3, 4, 5, 7, 10, 11
LO2	32	patient 3, 5, 10, 11
TO	12	patient 3, 10, 11
IPS	21	patient 7, 10, 11

Table 2: Overview of electrodes and patients per visual area.
Columns refer to the following: *Visual area:* Name of retinotopic map. Ventral and dorsal maps of early visual areas (V1-3) were combined. *Number of electrodes*: Number of electrodes with non-zero overlap with each retinotopic area, according to the full probability retinotopic atlas by Wang et al., (2014). Note that the sum is larger than the total number of electrodes included in the dataset (n=88), because electrodes can probabilistically map onto several areas. *Contributing patients*: List of patients that contribute electrodes for each visual area. Area hV4 was excluded in preprocessing, and areas TO1 and TO2 were combined into a single area TO (see Materials and Methods).

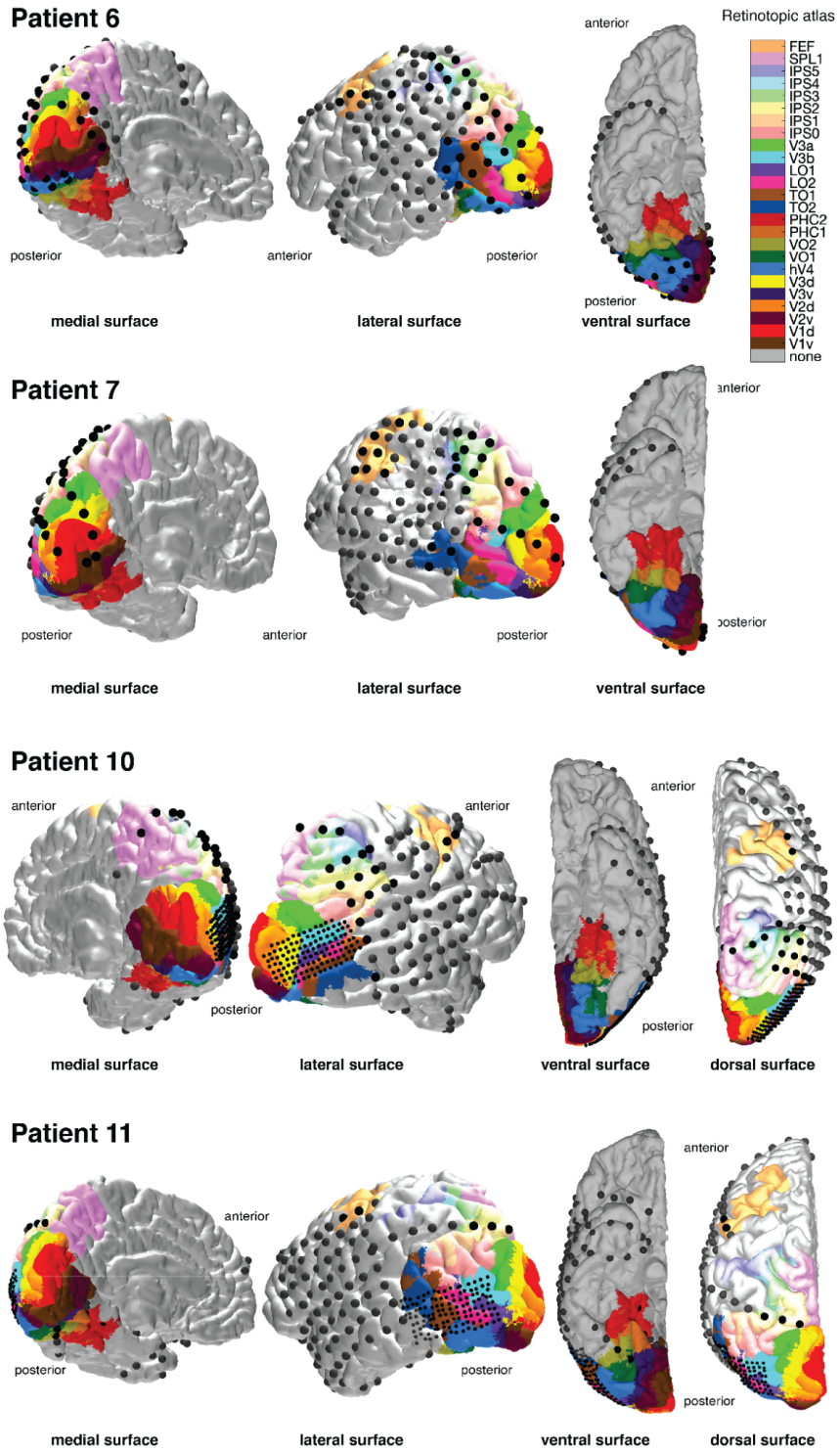
Extended Data

Contains:

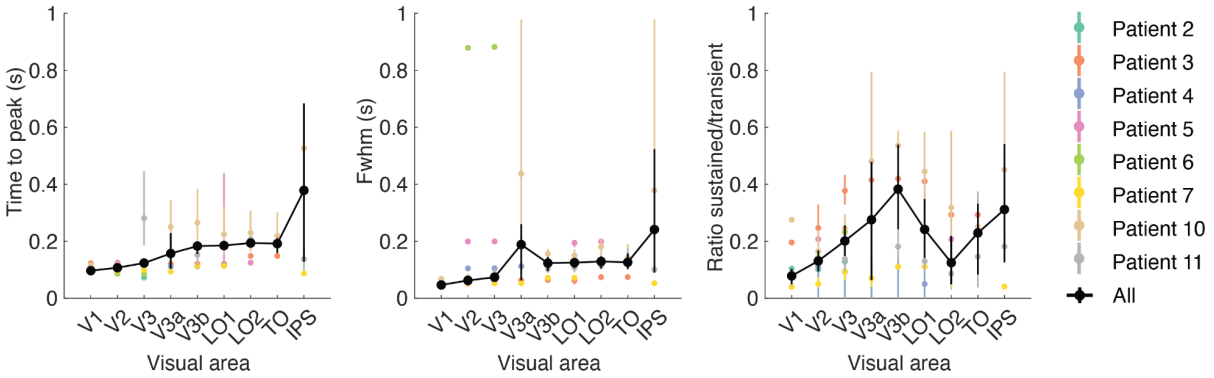
1. Extended Data Figure 1-1 (Electrode reconstructions for patient 2,3,4,5)
2. Extended Data Figure 1-2 (Electrode reconstructions for patient 6,7,10,11)
3. Extended Data Figure 8-1 (Temporal summation parameters for individual patients)
4. Extended Data Figure 11-1 (Fitted pRF parameter summary)



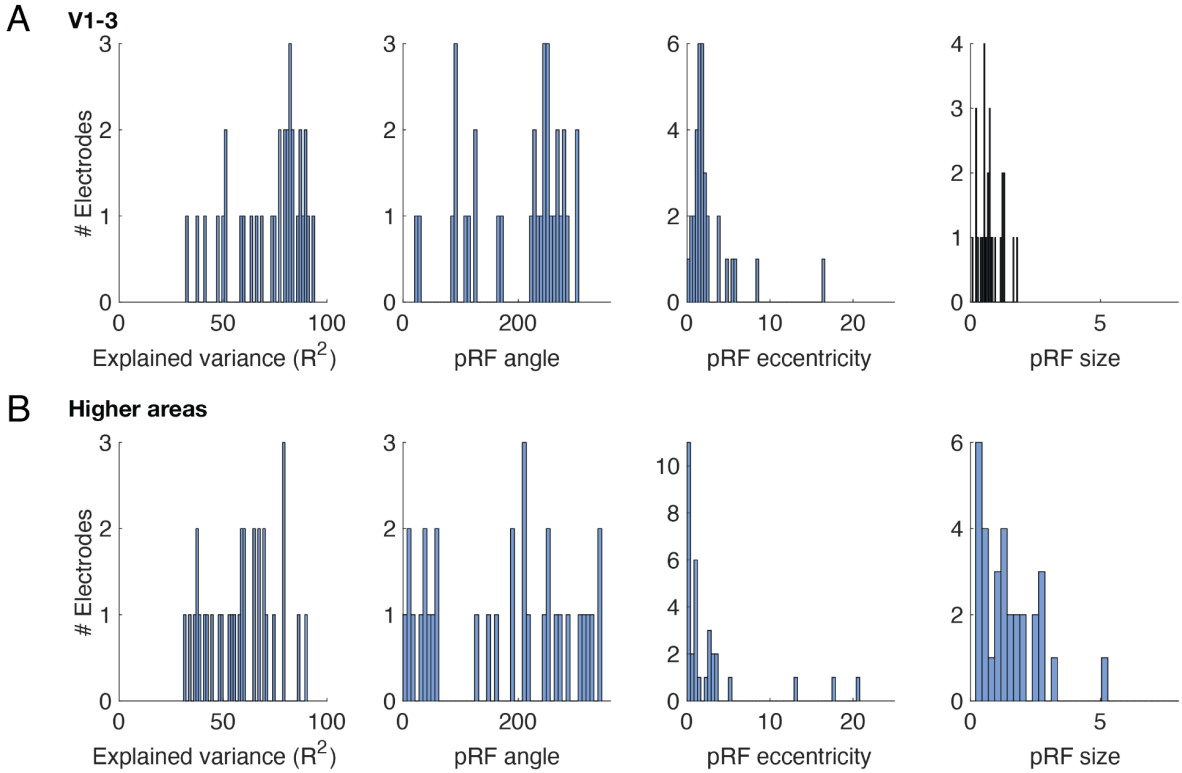
Extended Data Figure 1-1: Electrode reconstruction on pial surfaces of patient 2, 3, 4 and 5.. Different views of the implanted hemisphere(s) are shown and labeled for each patient. The normalized full probability atlas (see Materials and Methods) (from Wang et al., (2014) is projected on each patient's pial surface and depicted using the color legend shown in the upper right. 3D interactive images of the individual meshes can be reproduced using *tde_mkFigure1_1.m*



Extended Data Figure 1-2: Electrode reconstruction on pial surfaces of patient 6, 7, 11 and 11. Same as Extended Data Figure 1-1, but for patients 5-8. 3D interactive images of the individual meshes can be reproduced using *tde_mkFigure1_1.m*



Extended Data Figure 8-1: Temporal summation windows for individual patients.
Time to peak, Fwhm and Sustained/transient ratio computed from the data, but separately for each electrode and then grouped by visual area within each individual patient (colored data points) as well as across all patients (black data points and line). Note that the average here is not identical to the data shown in **Figure 8*B-D*** where summary statistics were computed based on average time courses per area, while here they were calculated per individual electrode and then averaged. Data points indicate median across repeated samples of electrodes, taking into account each electrode's probability of overlap with a retinotopic atlas (see Methods). Error bars and shaded regions reflect 68% confidence intervals across 1,000 bootstraps of electrode assignments. While individual electrode data is noisy, and there is substantial variety amongst individual patients, most patients show trends towards longer temporal summation windows for higher visual areas, as evidenced by increase in time-to-peak, response width and sustained to transient ratios. This figure can be reproduced using *tde_mkFigure8_1.m*.



Extended Data Figure 11-1: PRF properties across electrodes

pRF parameters summarized across all electrodes included in **Figure 11** for **A)** early visual areas V1-V3 and **B)** higher visual areas. Columns show explained variance by the pRF model, estimates of pRF angle, eccentricity and size. This figure can be reproduced using *tde_mkFigure11_1.m*.