PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

# Differential sustained and transient temporal processing across visual streams

**Anthony Stigliani**[1], **Brianna Jeska**[1], **Kalanit Grill-Spector**[1,2]*

**1** Psychology Department, Stanford University, Stanford, California, United States of America, **2** Stanford Neurosciences Institute, Stanford University, Stanford, California, United States of America

* kalanit@stanford.edu



OPEN ACCESS

**Citation:** Stigliani A, Jeska B, Grill-Spector K (2019) Differential sustained and transient temporal processing across visual streams. PLoS Comput Biol 15(5): e1007011. https://doi.org/10.1371/journal.pcbi.1007011

**Editor:** Saad Jbabdi, Oxford University, UNITED KINGDOM

**Received:** October 26, 2018

**Accepted:** April 7, 2019

**Published:** May 30, 2019



**Data Availability Statement:** Data will be held in a public repository at https://osf.io/mw5pk/ and code is available at https://github.com/VPNL/TemporalChannels.

**Funding:** This research was supported by National Eye Institute Grant 1R01-EY02391501A1 awarded to KGS (https://nei.nih.gov/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

How do high-level visual regions process the temporal aspects of our visual experience? While the temporal sensitivity of early visual cortex has been studied with fMRI in humans, temporal processing in high-level visual cortex is largely unknown. By modeling neural responses with millisecond precision in separate sustained and transient channels, and introducing a flexible encoding framework that captures differences in neural temporal integration time windows and response nonlinearities, we predict fMRI responses across visual cortex for stimuli ranging from 33 ms to 20 s. Using this innovative approach, we discovered that lateral category-selective regions respond to visual transients associated with stimulus onsets and offsets but not sustained visual information. Thus, lateral category-selective regions compute moment-to-moment visual transitions, but not stable features of the visual input. In contrast, ventral category-selective regions process both sustained and transient components of the visual input. Our model revealed that sustained channel responses to prolonged stimuli exhibit adaptation, whereas transient channel responses to stimulus offsets are surprisingly larger than for stimulus onsets. This large offset transient response may reflect a memory trace of the stimulus when it is no longer visible, whereas the onset transient response may reflect rapid processing of new items. Together, these findings reveal previously unconsidered, fundamental temporal mechanisms that distinguish visual streams in the human brain. Importantly, our results underscore the promise of modeling brain responses with millisecond precision to understand the underlying neural computations.

## Author summary

How does the brain encode the timing of our visual experience? Using functional magnetic resonance imaging (fMRI) and a generative temporal model with millisecond resolution, we discovered that visual regions in the lateral and ventral processing streams fundamentally differ in their temporal processing of the visual input. Regions in lateral temporal cortex process visual transients associated with the beginning and ending of the stimulus, but not its stable aspects. That is, lateral regions appear to compute moment-to-moment changes in the visual input. In contrast, regions in ventral temporal cortex

process both stable and transient components of the visual input, even as the response to the former exhibits adaptation. Surprisingly, the model predicts that in ventral regions responses to stimulus endings are larger than beginnings. We suggest that ending responses may reflect a memory trace of the stimulus, when it is no longer visible, and the beginning responses may reflect processing of new inputs. Together, these findings (i) reveal a fundamental temporal mechanism that distinguishes visual streams and (ii) highlight both the importance and utility of modeling brain responses with millisecond precision to understand the temporal dynamics of neural computations in the human brain.

## Introduction

How do high-level visual areas encode the temporal characteristics of our visual experience? The temporal sensitivity of early visual areas has been studied with electrophysiology in non-human primates [1–4] and recently using fMRI in humans [5, 6]. However, the nature of temporal processing in high-level visual regions remains a mystery for two main reasons. First, the temporal resolution of noninvasive fMRI measurements is in seconds [7], an order of magnitude longer than the timescale of neural processing, which is in the order of tens of milliseconds. Second, while fMRI responses are roughly linear for stimuli lasting 3–10 s [8], responses in visual cortex exhibit nonlinearities both for briefer stimuli, which generate stronger than expected responses [5, 6, 8–13], as well as for longer stimuli, which get suppressed due to adaptation [14]. Since the standard approach using a general linear model (GLM) to predict fMRI signals from the stimulus [8] is inadequate for modeling responses to such stimuli, the temporal processing characteristics of human high-level visual cortex have remained largely elusive (but see [12, 14–17]).

We hypothesized that if nonlinearities are of neural (rather than BOLD) origin, a new approach that predicts fMRI responses by modeling neural nonlinearities can be applied to characterize temporal processing in high-level visual cortex. Different than the GLM, which predicts fMRI signals directly from the stimulus, the encoding approach first models neural responses to the stimulus and from them predicts fMRI responses. Recent studies show that accurately modeling neural responses to brief visual stimuli at millisecond resolution better predicts fMRI responses than the GLM [5, 6, 18]. The encoding approach also enables testing a variety of temporal models and quantifying which model best predicts brain responses. Further, generative computational models of neural processing offer a framework that can provide key insights into multiple facets of temporal processing including integration time windows [19–21], temporal channel contributions [5, 18, 22–25], and response nonlinearities [5, 6, 9–12, 18].

One fundamental way in which visual regions differ is in how they process sustained and transient visual stimuli. In the retina [26], LGN [27–29], and V1 [18], there are separate temporal channels for processing transient and sustained components of the visual input that are associated with magnocellular (M) and parvocellular (P) pathways, respectively [22–24]. While these channels are likely combined outside early visual cortex (V1-V3), it is thought that regions that process dynamic stimuli such as hMT+ largely receive input from the transient channel in V1 [23], and regions in the ventral stream such as hV4 receive inputs from both transient and sustained channels [24]. Indeed, using an encoding model with two temporal channels—one sustained and one transient—we were able to successfully predict fMRI responses in early and intermediate visual areas (V4, hMT+) to phase scrambled stimuli varying in duration from 33 ms to 30 s [5].

Motivated by the success of this approach in early and intermediate visual areas, we considered three hypotheses regarding temporal processing in high-level visual cortex. One possibility is that temporal processing characteristics are similar across high-level visual regions but differ from those of earlier stages of the visual hierarchy. This hypothesis is based on results from animal electrophysiology showing longer latencies of responses in higher-level visual regions compared to primary visual cortex, V1 [1], as well as research in humans showing longer temporal receptive windows [19, 20] and integration times [21] in ventral temporal cortex (VTC) and lateral temporal cortex (LTC) compared to early visual areas. A second possibility is that temporal processing is uniform across high-level regions that process a shared category (e.g. face-selective regions in VTC and LTC), but differs across regions that process different categories (e.g. face- vs. body-selective regions). This prediction is based on data showing differential responses to long-duration (21 s) images in face- vs. place-selective regions in VTC [14], as well as differential response characteristics to fast (8 Hz) visual stimulation in body-selective regions vs. other category-selective regions [15]. A third possibility is that temporal processing differs across ventral and lateral visual streams rather than across categories. A large body of literature has documented that regions in LTC along the superior temporal sulcus (STS) show heightened responses to biological motion compared to stationary stimuli and other types of motion [30–37], unlike regions in VTC that are thought to represent the static aspect of the stimulus [34, 35, 38]. This predicts that lateral regions may show larger transient responses than ventral regions, which instead may show larger sustained responses.

To test these predictions, we measured fMRI responses in high-level visual areas to images of faces, bodies, and words that were either sustained (one continuous image per trial, durations ranging from 3–20 s) (Fig 1*A* and 1*B*, *experiment 1*), transient [30 flashed, 33 ms long images per trial with interstimulus intervals (ISIs) ranging from 67–633 ms] (Fig 1*A* and 1*B*, *experiment 2*), or contained both transient and sustained components (30 semi-continuous images per trial, durations ranging from 67–633 ms per image with 33 ms ISIs) (Fig 1*A* and 1*B*, *experiment 3*). We also collected a separate functional localizer experiment to independently define regions selective to faces and bodies in VTC and LTC (Fig 1*C*; Materials and methods). We used face- and body-selective regions as a model system as there are multiple clusters of selectivity to these categories across the temporal lobe, and face and body regions neighbor on the cortical sheet [39]. This organization enabled us to (i) measure how the temporal dynamics of stimuli affect responses in each region and (ii) test if temporal processing characteristics vary across regions selective to different categories (e.g., faces or bodies) or across regions in different anatomical locations (e.g., ventral vs. lateral temporal cortex).

## Results

### Responses in high-level visual cortex exhibit temporal nonlinearities

To assess the feasibility of our approach, we first used a standard widely-used GLM [8] to predict fMRI responses in the three main experiments. Then, we compared these predictions to measured fMRI responses from two sample functional regions of interest: a ventral body-selective region and a lateral body-selective region.

In general, the GLM predicts longer responses for longer trials and similar responses in experiments 1 and 3 (Fig 2*A*, *blue* and *green*). Responses in experiment 1 are predicted to be slightly higher than in experiment 3 because the 33 ms gaps between images in the latter experiment make up 1 s of baseline within each trial. Due to the nature of the hemodynamic response function (HRF), the GLM also predicts that peak response amplitudes in experiments 1 and 3 will increase gradually from 3 s to 10 s trials and subsequently plateau for longer trial durations. In contrast, this model predicts substantially lower responses in experiment 2

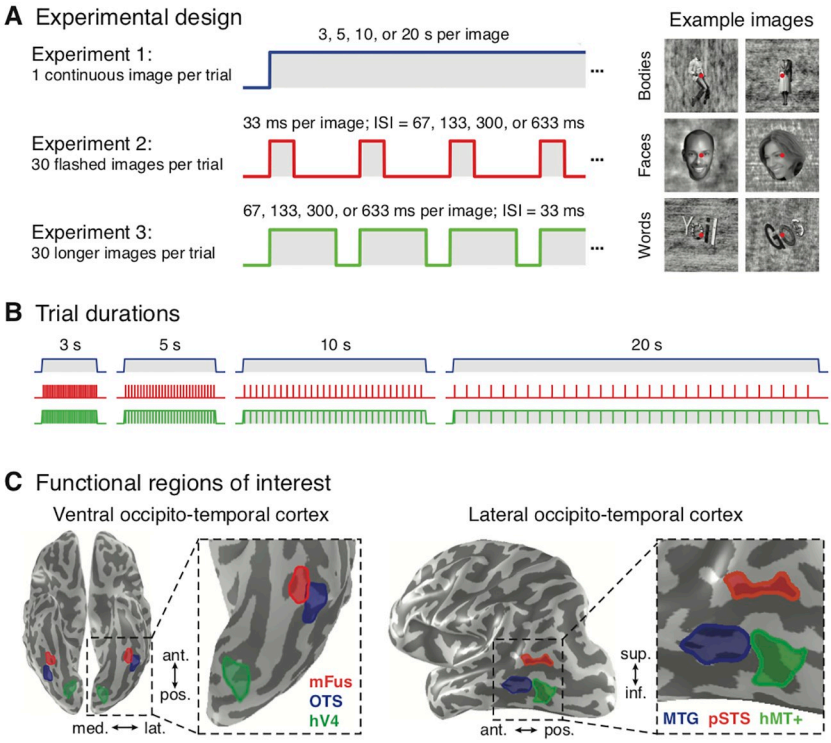


**Fig 1. Measuring brain responses to combinations of sustained and transient visual stimuli in high-level visual cortex.** (*A*) Participants fixated centrally and viewed images of bodies, faces, and pseudowords (*right*) that were presented in trials of different durations interleaved with 12-s periods of a blank screen (*left*). *Experiment 1*: a single image was shown for the duration of a trial. *Experiment 2*: 30 briefly presented images from the same category (33 ms each), each followed by a blank screen, were presented in each trial. As the trial duration lengthens, the gap between images increases, causing the fraction of the trial containing visual stimulation to decline. *Experiment 3*: 30 semi-continuous images from the same category were presented in each trial with a constant 33-ms blank screen between consecutive images. As the block duration lengthens, the duration of each image progressively increases but the gap does not. (*B*) The same trial durations (3, 5, 10, or 20 s) were utilized across all three experiments, while the rate and duration of visual presentation varied between experiments. Corresponding trials in experiments 1 and 3 have almost the same overall duration of stimulation but different numbers of stimuli, whereas trials in experiments 2 and 3 have the same number of stimuli but different durations of stimulation. The same fixation task was used in the three main experiments. (*C*) Functional regions of interest in ventral temporal cortex (*left*) and lateral temporal cortex (*right*) selective to bodies (OTS and MTG, blue) and faces (pSTS and mFus, red), as well as human V4 (hV4) and human motion-sensitive area (hMT+). Regions in each anatomical section are shown in an example subject.

https://doi.org/10.1371/journal.pcbi.1007011.g001

compared to the other experiments because the transient 33 ms stimuli in this experiment comprise only a small fraction of each trial duration (Fig 2*A*, *red*). Therefore, the GLM predicts a progressive decrease in response amplitude from 3 s to 20 s trials in experiment 2, as the fraction of the trial in which stimuli are presented decreases (from 1/3 to 1/20 of the trial).

Strikingly, responses to body images in a ventral body-selective region (OTS-bodies; Fig 2*B*) and a lateral body-selective region (MTG-bodies; Fig 2*C*) both deviate from the predictions of the GLM, but in different ways. Although these regions prefer the same category, we observe differences in their maximal response to the different timing conditions in our experiments [significant three-way interaction, $F_{6,\ 54} = 2.28$, $P < .05$; three-way ANOVA on peak trial response amplitude for each participant with factors of trial duration (3/5/10/20 s), experiment (1/2/3), and ROI (OTS-bodies/MTG-bodies); Fig 2*B* and 2*C*].

In contrast to the predictions of the GLM, responses in OTS-bodies to trials of 30 flashed images in experiment 2 (Fig 2*B*, *red*) are substantially higher than in corresponding trial durations in experiment 1, when one stimulus is shown per trial (Fig 2*B*, *blue*). This occurs despite

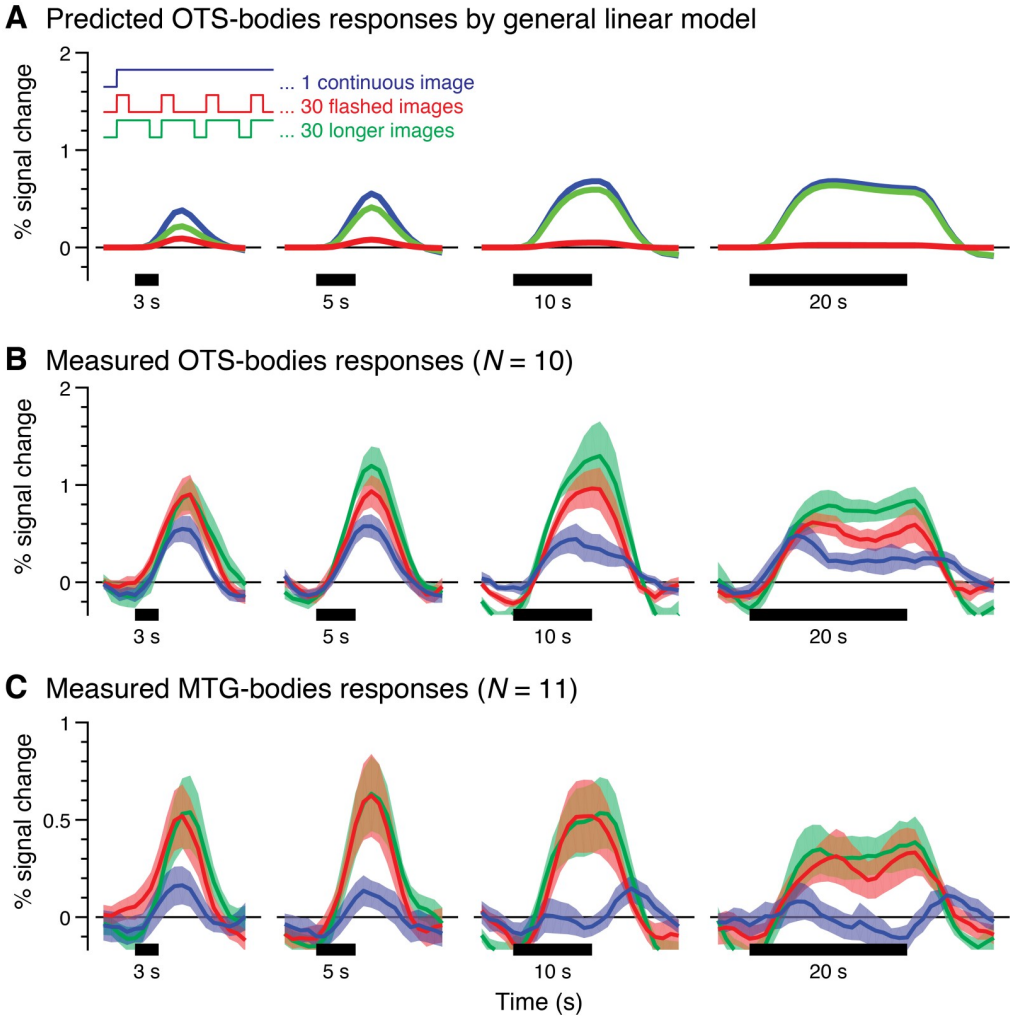


**Fig 2. Responses of body-selective regions in ventral and lateral temporal cortex exhibit nonlinearities that are not predicted by a linear model.** (*A*) Predicted responses by a GLM for trials containing one continuous image (*blue*), thirty flashed (33 ms) images (*red*), and thirty longer images that span then entire trial duration except for a 33 ms interstimulus interval (ISI) following each image (*green*). Predictors are fit to OTS-bodies responses using data concatenated across all three experiments shown in (B). (*B*) Measured responses during Exp1-Exp 3 from an independently defined ventral region on the occipitotemporal sulcus (OTS) selective to bodies (OTS-bodies). (*C*) Measured responses during Exp1-Exp 3 from an independently defined lateral region on the middle temporal gyrus (MTG) selective for bodies (MTG-bodies). In (B-C), *lines*: mean response time series across participants for trials with body images; *shaded areas*: standard error of the mean (SEM) across participants; *Horizontal black bars*: stimulus duration.

https://doi.org/10.1371/journal.pcbi.1007011.g002

the fact that stimuli are presented for only a small fraction of each trial duration in experiment 2 compared to experiment 1. Furthermore, peak response amplitudes do not increase with trial duration in experiment 1 as predicted by the GLM. Instead, we observe a systematic decrease in response after the first few seconds of stimulation in the 10 s and 20 s trials, which is consistent with prior reports of fMRI adaptation for prolonged stimuli in nearby face- and place-selective regions [14]. Lastly, responses in experiment 3 (Fig 2*B*, *green*) exceed responses in both experiment 1 (which has only one image per trial but similar overall durations of stimulation) and experiment 2 (which has the same number of images per trial but shorter stimulus durations). This observation suggests that both the number of stimuli in a trial and their duration impact response amplitudes, as in earlier visual areas such as V1 and hV4 (S1 Fig).

Unlike OTS-bodies, MTG-bodies illustrates a largely transient response characteristic with substantially lower responses to the prolonged single images in experiment 1. Notably, for the 10 s and 20 s trials, we observe a transient response following both the onset and the offset of the image but no elevation of response in the middle of the trial (Fig 2*C*, *blue*). In contrast to the lack of robust responses in experiment 1, MTG-bodies shows surprisingly large responses to briefly flashed stimuli in experiment 2 (Fig 2*C*, *red*). Additionally, responses in MTG-bodies during experiment 2 (Fig 2*C*, *red*) and experiment 3 (Fig 2*C*, *green*), which have 30 stimuli per trial but of different stimulus durations, are similar and both exceed responses in experiment 1, which has a single stimulus per trial. This suggests that, unlike ventral regions, stimulus duration has little impact on MTG-bodies responses, which resemble responses in neighboring motion-sensitive hMT+ (S1 Fig).

These data demonstrate that (i) varying the temporal properties of visual presentations in the millisecond range has a profound effect on fMRI responses in high-level visual cortex, (ii) the standard GLM is inadequate for predicting measured fMRI responses to these types of stimuli in high-level regions, in agreement with prior data in earlier visual areas [5, 6, 8–13], and (iii) even though OTS-bodies and MTG-bodies prefer the same stimulus category, their temporal response characteristics vastly differ.

## An encoding model of temporal processing in high-level visual cortex

Motivated by the recent success of encoding models that predict fMRI responses in earlier visual areas by modeling neural temporal nonlinearities [5, 6, 18], we applied a similar approach to predict responses in high-level visual areas. Different than the GLM, the temporal encoding approach first models the neural response in millisecond resolution and then convolves the estimated neural response with a HRF to predict fMRI responses (Fig 3).

Our encoding model consists of two temporal channels [5, 18]–a sustained channel and a transient channel—each of which can be modeled using a neural temporal impulse response function (IRF) followed by a nonlinearity [2, 3, 5, 18, 40]. The sustained channel is modeled with a monophasic IRF (Fig 3*B*, *blue channel IRF*), which predicts a sustained neural response for the duration of the stimulus. To capture the gradual decay (adaptation, A) of the response observed in ventral regions for sustained images (Fig 2*B*, *blue*), we apply a nonlinearity to the sustained channel in the form of an exponential decay function (Materials and methods). The transient channel is characterized by a biphasic IRF (Fig 3*B*, *red channel IRF*) that identifies changes to the visual input. That is, it acts like a derivative function, predicting no further increase in the neural response once a stimulus has been presented for longer than the duration of the IRF [5, 18]. This channel too has a nonlinearity, as we hypothesize an increase in neural response at both the appearance (onset) and disappearance (offset) of a stimulus. To account for the pronounced transient responses in high-level visual regions (S1 Fig), we apply a flexible compressive nonlinearity on the transient channel using a pair of sigmoid (S) functions, one for the onset and another for the offset (Materials and methods). Thus, we refer to this two-temporal channel encoding model as the A+S model. The predicted fMRI response is generated by convolving the neural response predictors for each channel with the HRF and summing the responses of the two temporal channels (Fig 3C). Since the HRF effectively acts as a low-pass temporal filter, predicted fMRI responses can be downsampled with minimal distortion to match the slower sampling rate of fMRI measurements.

We estimated optimized A+S model parameters separately for each participant and region using nonlinear programming and a cross-validation approach. In our procedure, we use half the data from all three experiments to estimate model parameters. Using optimization, we estimate a time constant for the neural IRFs ($\tau$), a time constant controlling adaptation of

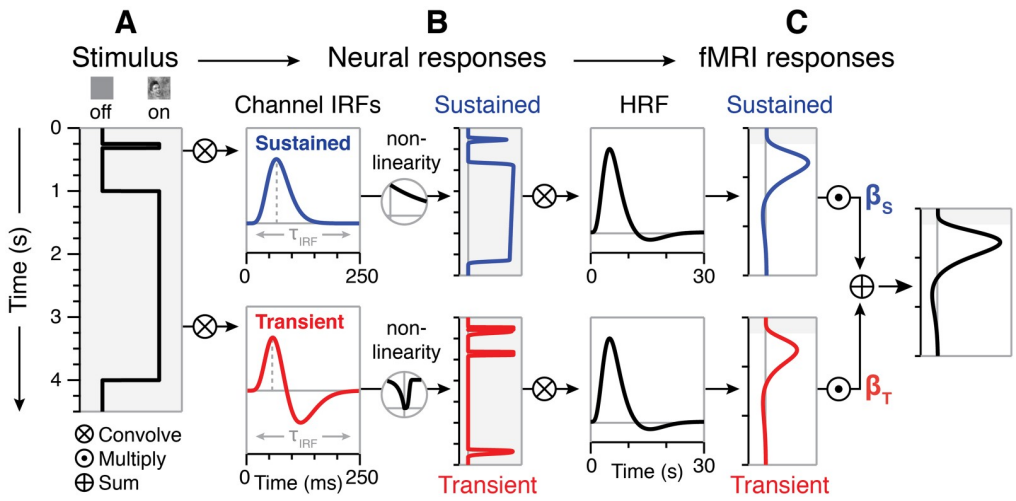


**Fig 3. Optimized two-temporal channel A+S model with adaptation and sigmoid nonlinearities.** (*A*) Transitions between stimulus and baseline screens are coded as a step function representing when a stimulus was on vs. off with millisecond temporal resolution. (*B*) Separate neural responses for the sustained (*blue*) and transient (*red*) channels are modeled by convolving the stimulus vector with an IRF for each channel. An exponential decay function is applied to the sustained channel to model response decrements related to neural adaptation, and a compressive sigmoid nonlinearity is applied to the transient channel to vary the temporal characteristics of "on" and "off" responses (Materials and methods). (*C*) Predictors of sustained and transient fMRI responses are generated by convolving each channel's neural response predictors with the HRF and down-sampling to match the sampling rate of measured fMRI data. The total fMRI response is the sum of the weighted sustained and transient fMRI predictors for each channel. To optimize model parameters and estimate the contributions ($\beta$ weights) of the sustained ($\beta_S$) and transient ($\beta_T$) channels, we fit the model to different splits of the data including runs from all three experiments.

https://doi.org/10.1371/journal.pcbi.1007011.g003

sustained responses (α), and three parameters controlling compression of transient responses ($k_{on}$, $k_{off}$, and λ). After optimizing these parameters, we use a GLM to estimate the magnitude of response ($\beta$ weight) for each channel and stimulus category in our experiments, resulting in three $\beta$ weights for the sustained channel (one $\beta_S$ for each category) and three $\beta$ weights for the transient channel (one $\beta_T$ for each category). These parameters and weights are then used to predict responses in left-out data and evaluate the model's goodness-of-fit (cross-validated variance explained, x-$R^2$).

Comparing the predictions of our optimized A+S model with measured fMRI responses in high-level visual cortex reveals two notable findings. First, our model generates signals that closely track the amplitude of fMRI responses in all three experiments in the left-out data. Second, analysis of x-$R^2$ shows that our optimized A+S model consistently outperforms other optimized temporal encoding models.

We illustrate these results for one region, OTS-bodies (Fig 4, S2 Fig); Results for other regions are in S3–S5 Figs. Notably, the A+S model closely tracks response amplitudes in all three experiments [Fig 4*A*–4*C*, compare overall model prediction (*black*) with measured data from OTS-bodies (*gray*)]. Consistent with our predictions, the sustained channel accounts for the bulk of the response in experiment 1 (Fig 4*A*, *blue*); The transient channel contributes most of the response in experiment 2 (Fig 4*B*, *red*), and both channels contribute to responses in experiment 3 (Fig 4*C*).

We compared the performance of our A+S model to other models of fMRI responses: the standard GLM [8], the balloon model [7], four single-channel models (L, CTS [6], A, S; S3*A* Fig), and three alternative two-channel models (L+Q [5, 18], C+Q, A+Q) across all three experiments (Materials and methods; S3–S5 Figs). For simplicity, Fig 4*D*–4*F* compares

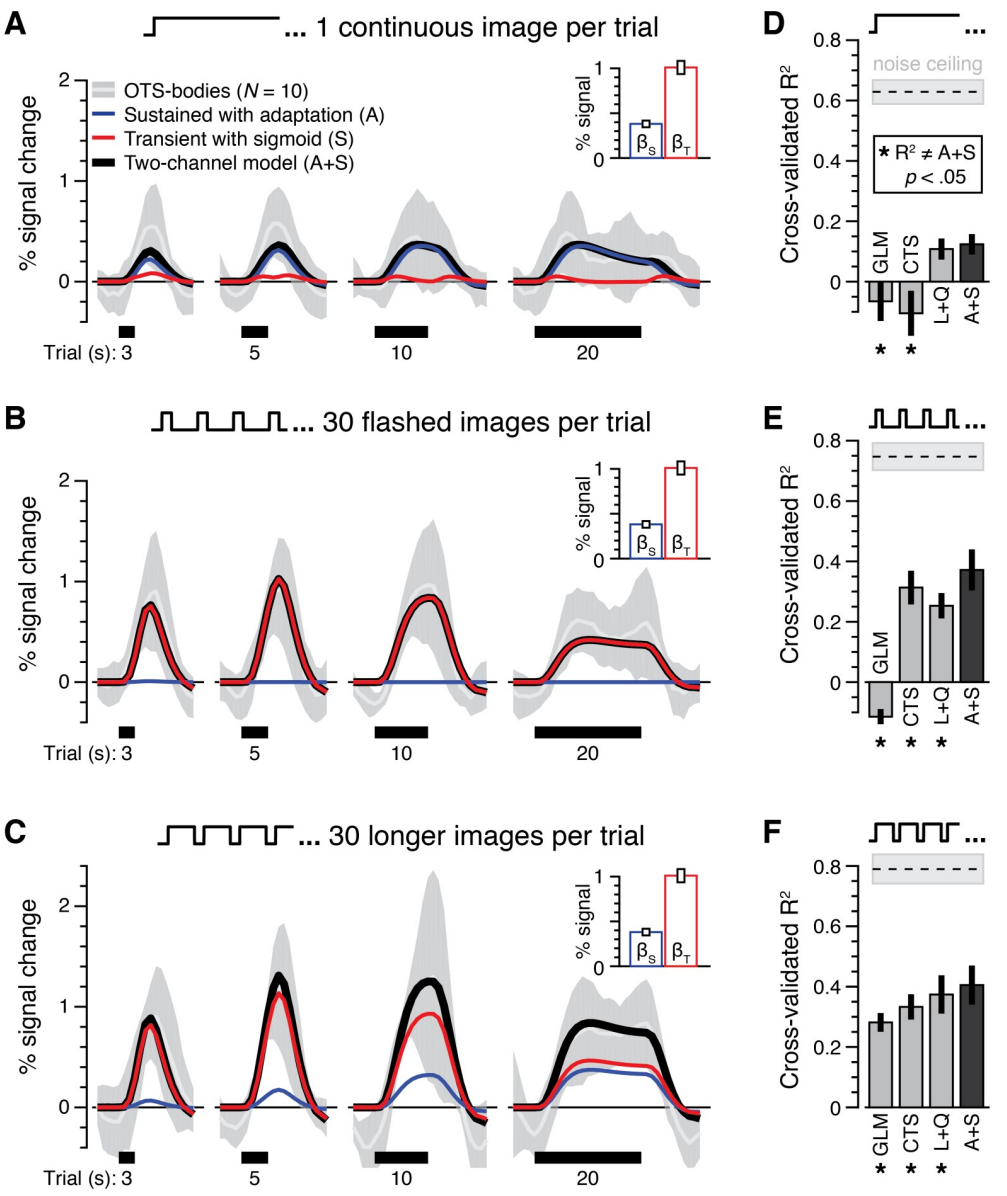


**Fig 4. Two-temporal channel model with nonlinearities on both sustained and transient channels predicts responses in ventral temporal cortex.** (A-C) Responses and model predictions for body images in OTS-bodies. *White curve*: mean response across 10 participants. *Shaded gray*: standard deviation across participants. *Blue*: predicted response from the sustained channel. *Red*: predicted response from the transient channel: *Black*: sum of responses from both channels. *Inset*: *mean* contribution ($\beta$ weight) for each channel ±1 SEM across participants. (*A*) Experiment 1 data, 1 continuous image per trial. (*B*) Experiment 2 data, 30 flashed images per trial. (*C*) Experiment 3 data, 30 longer images per trial. (D-F) Model comparison. Bars show the performance of various models for each experiment presented in (A-C). Models are fit using runs from all three experiments, and cross-validation performance (x-$R^2$) is calculated in left-out data from each experiment separately. (*D*) Experiment 1. (*E*) Experiment 2. (*F*) Experiment 3. Single-channel models: G*LM*, general linear model [8]; *CTS*, a sustained channel with compressive temporal summation [6]. Dual-channel models: *L+Q*, a linear sustained channel and a transient channel with quadratic nonlinearity [5]; *A+S*: a sustained channel with adaptation and a transient channel with sigmoid nonlinearities. Asterisks denote models with significantly different performance compared to A+S (paired *t*-tests comparing x-$R^2$ of each model vs. A+S in each experiment).

https://doi.org/10.1371/journal.pcbi.1007011.g004

performance in OTS-bodies for our model vs. three others: the standard GLM [8], a single-channel model with compressive temporal summation (CTS) [6], and a two-channel model composed of a linear sustained channel and a transient channel with a quadratic nonlinearity (L+Q) [5, 18]. The latter two models have been recently used to model temporal dynamics of early and intermediate visual areas. Comparing the performance of these models in OTS-bodies (Fig 4*D*–4*F*), we observe significant differences across experiments [significant main effect of model type, $F_{6,\ 54} = 13.79$, $P < .001$, two-way ANOVA with factors of model type (GLM/CTS/L+Q/A+S) and experiment (experiment 1/2/3)]. Notably, the A+S model predicts OTS-bodies responses in left out data significantly better than the GLM [8], which overestimates responses in experiment 1 and underestimates responses in experiment 2 (GLM vs. A+S: all $t$s $> 2.64$, $P$s $< .05$, paired $t$-tests on x-$R^2$ separately for each experiments) (S2*A* Fig). The A+S model also outperforms the recently proposed CTS model [6] that enhances early and late portions of the neural response to a stimulus (CTS vs. A+S: all $t$s $> 2.60$, $P$s $< .05$). While the CTS model performs considerably better than the GLM in experiment 2, it overestimates responses in experiment 1 with a single continuous image per trial and underestimates responses in experiment 3 with 30 longer images per trial (S2*B* Fig). In experiments 2 and 3, we also observe a significant advantage of the A+S model compared to the two-temporal channel L+Q model [5, 18], which underestimates the large responses to transient stimuli in experiment 2 (S2*C* Fig) (L+Q vs. A+S: $t$s $> 4.06$, $P$s $< .05$; the difference fell short of significance for experiment 1, $t_9 = 1.98$, $P = .08$).

Thus, an optimized two-temporal channel model with an adaptation nonlinearity in the sustained channel and compressive sigmoid nonlinearities in the transient channel predicts fMRI responses to visual stimuli ranging from milliseconds to seconds in high-level visual cortex with greater accuracy than alternative models.

## How do channel contributions differ across ventral and lateral category-selective regions?

Examination of response time series (S1 Fig) and channel weights (Fig 5) in body- and face-selective regions in VTC and LTC reveals prominent differences across ventral and lateral temporal regions.

First, comparing the response time courses of different category-selective regions shows that ventral temporal regions (e.g., OTS-bodies and mFus-faces) respond strongly to both the sustained stimuli in experiment 1 and the transient stimuli in experiment 2, whereas lateral temporal regions (MTG-bodies and pSTS-faces) respond strongly to the transient stimuli but minimally to the sustained stimuli (S1 Fig). The ratio of sustained and transient channel amplitudes, $|\frac{\beta_S}{\beta_T}|$, also differs across regions in ventral and lateral aspects of temporal cortex [significant main effect of processing stream, $F_{1,\ 107} = 14.27$, $P < .01$, three-way ANOVA with factors of processing stream (ventral/lateral), stimulus category (faces/bodies/words), and preferred category (bodies/faces); all other effects failed to reach significance; $F$s $< 1.39$, $P$s $> .05$]. That is, while both sustained and transient channels contribute to responses in ventral temporal regions (Fig 5*B*), the transient channel dominates responses in lateral temporal regions (Fig 5*C*). In fact, zeroing the contribution of the sustained channel slightly improves model performance in lateral regions (i.e. x-$R^2$ of the S model is marginally better than the A+S model in MTG-bodies and pSTS-faces; S3*B* Fig). In contrast, zeroing the sustained channel detrimentally affects the prediction of responses in ventral regions for prolonged visual stimulation as in experiment 1 (S6*A* Fig). Finally, ventral temporal regions show a response characteristic similar to both hV4 (S4*A* Fig) and occipital category-selective

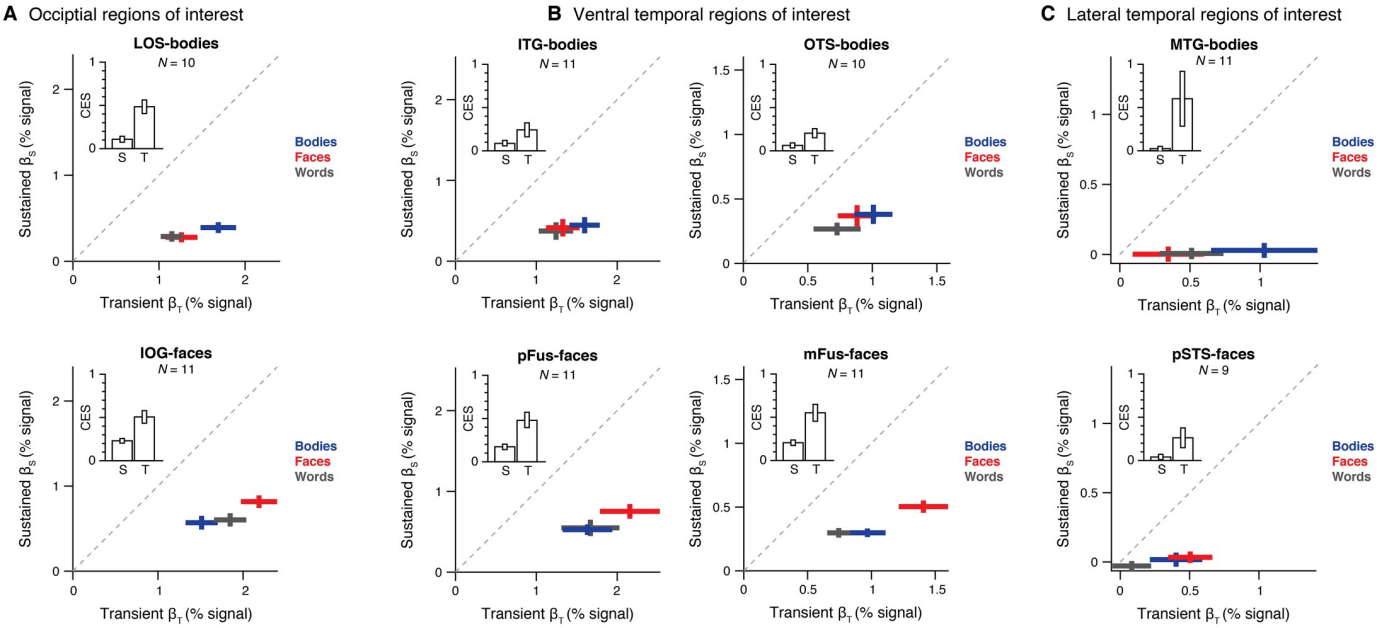


**Fig 5. Differential contributions of transient and sustained temporal channels across ventral and lateral regions selective to face and body stimuli.** Contributions ($\beta$ weights) of transient (*x* axis) and sustained (*y* axis) channels for each stimulus category estimated by the two-temporal channel A+S model in (*A*) occipital body-selective region on the lateral occipital sulcus (LOS) and a face-selective region on the inferior occipital gyrus (IOG), (*B*) ventral-temporal body-selective regions on the inferior temproal gyrus (ITG) and occipito-temporal sulcus (OTS) and face-selective regions on the posterior and mid fusiform gyrus, pFus- and mFus-faces, respectivevely, and (*C*) a lateral temporal body-selective region on the mid temporal gyrus (MTG) and a face-selective region on the posterior aspect of the superior temporal sulcus (pSTS-faces). Crosses span ±1 SEM across participants in each axis, and $\beta$ weights were solved by fitting the model using split halves of the data including runs from all three experiments. Data show average model weights across both splits of the data for each participant. *Red*: response to faces. *Blue*: response to bodies. *Gray*: response to words. *Dashed line*: identity line ($\beta_S = \beta_T$). *Inset*: bars indicate mean contrast effect size (CES) of $\beta$ weights for the preferred vs. nonpreferred categories in each channel ±1 SEM across participants.

https://doi.org/10.1371/journal.pcbi.1007011.g005

regions (Fig 5*A*), whereas lateral temporal regions show a characteristic similar to motion-sensitive hMT+ (S4*A* Fig).

Second, in VTC (Fig 5*B*), category selectivity—or higher responses to a preferred category vs. other categories—is evident in both sustained and transient channels [all $ts > 2.26$, $Ps < .05$, one-tailed *t*-tests comparing the contrast effect size (CES) of $\beta$ weights for the preferred vs. nonpreferred categories separately for each channel; Fig 5*B*, *insets*]. For example, in both channels, the weighting of the predicted response to faces in mFus is significantly higher than the average weighting of responses to words and bodies ($ts > 5.25$, $Ps < .001$, paired *t*-test for each channel; Fig 5*A*, *right*). Likewise, in both channels, the predicted weighting of responses to bodies in OTS is higher than the average weighting of responses to other categories ($ts > 2.59$, $Ps < .05$, paired *t*-test for each channel; Fig 5*A*, *left*). Selectivity across both temporal channels was also observed in nearby word-selective regions (IOS-words and pOTS-words; S7*A* Fig). Interestingly, category selectivity in the sustained channel was higher in ventral face-selective regions as compared to body-selective regions. In contrast, in LTC (MTG-bodies and pSTS-faces; Fig 5*C*), there is a significant difference in the CES across sustained and transient channels [significant main effect of channel, $F_{1, 8} = 14.88$, $P < .01$, two-way ANOVA with factors of channel (sustained/transient) and preferred category (bodies/faces)]. That is, higher responses to the preferred category are observed only in the transient channel ($ts > 1.99$, $Ps < .05$; the effect was not significant in the sustained channel, $ts < 1.37$, $Ps > .10$; Fig 5*B*, *insets*). Thus,

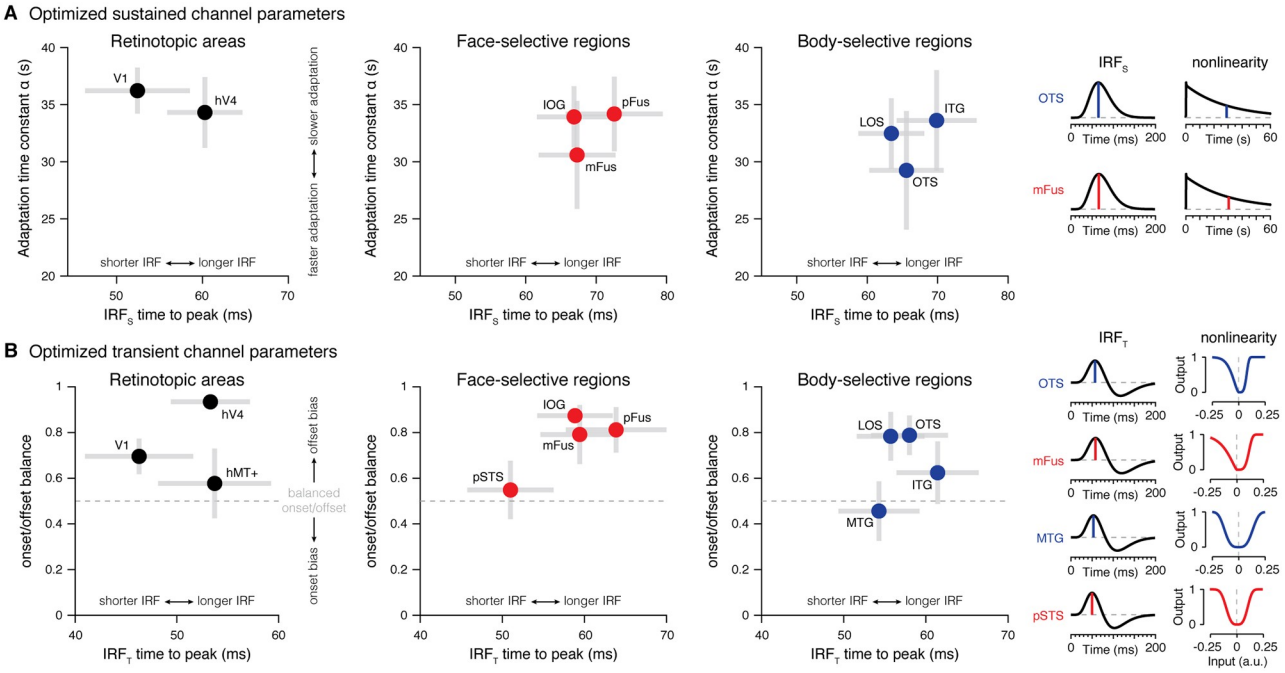


**Fig 6. Optimized two-temporal channel model parameters differ across visual cortex.** (*A*) Optimized sustained channel parameters. Time to peak of sustained $IRF_S$ (*x* axis) and exponential time constant of the adaptation function (*y* axis) for each set of regions estimated by the two-temporal channel A+S model. Crosses span ±1 SEM across participants in each axis, and parameters were optimized using split halves of the data containing runs from all experiments. Data show model parameters averaged across both splits of the data for each participant. (*B*) Optimized transient channel parameters. Time to peak of transient $IRF_T$ (*x* axis) and onset/offset balance (*y* axis) for each set of regions estimated by the two-temporal channel A+S model (with a zeroed sustained channel in lateral regions). The onset/offset balance metric captures differences in the shapes of the sigmoid nonlinearities used to compress transient "on" and "off" responses, where values larger than 0.5 refect elongation of offset responses compared to onset responses. Crosses span ±1 SEM across participants in each axis, and parameters were optimized using split halves of the data from all experiments. Plots show average model parameters across all splits of the data for each participant. Sample IRFs and nonlinearities shown to the right of (A-B) are generated by averaging optimized model parameters across participants.

https://doi.org/10.1371/journal.pcbi.1007011.g006

these results reveal differential contributions of transient and sustained channels across ventral and lateral category-selective regions.

## How do timing parameters vary across ventral and lateral face and body regions?

We next examined the optimized timing and compression parameters for each channel to test if there are functional differences across regions. The parameters in our A+S model were optimized separately for each region within each participant. Thus, for a given ROI, we optimized one time constant for the channel IRFs, one time constant for the adaptation decay function, as well as three sigmoid parameters (Materials and methods).

For the sustained channel, we assessed how the time to peak of the neural IRF ($IRF_S$) and the adaptation decay constant (α) vary across occipital and ventral temporal regions, omitting lateral temporal regions that did not have significant sustained responses. We discovered a hierarchical progression of longer time to peak and stronger adaptation in the sustained channel ascending from early to later stages of the ventral hierarchy (Fig 6*A*). That is, the time to peak of the sustained IRF tended to be shorter in V1 than hV4 and shorter in hV4 than in ventral regions OTS-bodies and mFus-faces (Fig 6*A*, *x axis*). At the same time, the adaptation decay constant decreased from V1 to ventral temporal regions, indicating more adaptation in

mFus-faces and OTS-bodies than in V1 (Fig 6A, *y axis*). We also observed a decreasing adaptation constant from occipital regions IOG-faces/LOS-bodies to the ventral regions mFus-faces/OTS-bodies. Thus, analyzing timing parameters in the sustained channel revealed differences in processing across the ventral stream.

In the transient channel, we examined how the time to peak of the $IRF_T$ varies across regions and if there are asymmetries in the compression of "on" compared to "off" neural responses controlled by the sigmoid shape parameters $k_{on}$ and $k_{off}$, respectively. Since lower $k$ values generally elongate transient responses, the relative contribution of the offset component can be indexed by a balance metric, $\frac{k_{on}}{k_{on}+k_{off}}$, where a ratio of 0.5 indicates equal contributions from the onset and the offest of a stimulus to BOLD signals ($k_{on} = k_{off}$). A ratio $< 0.5$ indicates a larger contribution of onset than offset responses, and a ratio $> 0.5$ indicates a larger contribution of offset than onset responses.

First, like the sustained channel, the transient channel also shows an increase in the time to peak of $IRF_T$ going from V1 to face- and body-selective regions in VTC and LTC (Fig 6B, *x axis*). Second, VTC face-and body-selective regions tended to show longer time to peak of their transient $IRF_T$ as compared to LTC face-and body-selective regions. Third, interestingly, transients in lateral regions, pSTS-faces and MTG-bodies, show balanced contributions of onset and offset responses (balance metric = 0.50±0.09; Fig 6B, *y axis* and *insets*). In contrast, transients in ventral regions, pFus/mFus-faces and ITG/OTS-bodies, as well as occipital face-selective IOG and body-selective LOS are dominated by offset responses (balance metric = 0.77 ±0.09; Fig 6B and *insets*). The surprisingly large offset contribution predicted by our model indicates that the bulk of VTC responses for the brief stimuli in experiment 2 can be attributed to neural responses that occur after the stimuli are no longer visible, rather than during the initial response to these stimuli.

Thus, comparison of optimized A+S model parameters reveals functional differences between early and later stages of the visual hierarchy, as well as distinct nonlinearities across ventral and lateral regions with the same category preference.

## Discussion

Using a temporal encoding approach to explain responses in high-level visual regions, we discovered that an optimized two-temporal channel model consisting of a sustained channel with an adaptation nonlinearity and a transient channel with compressive sigmoid nonlinearities successfully predicts fMRI responses in human high-level visual cortex for stimuli presented for durations ranging from tens of milliseconds to tens of seconds. Critically, the innovative temporal encoding framework we introduce combines in a single computational model several components of temporal processing including time windows of temporal integration [12, 16, 19–21], channel contributions [5, 18, 22–25], and nonlinearities in temporal summation [5, 6, 9–12, 18]. Using this approach, we (i) uncover the temporal sensitivity of neural responses in human high-level visual cortex, (ii) find differential temporal characteristics across lateral and ventral category-selective regions, and (iii) propose a new mechanism—temporal processing—that functionally distinguishes visual processing streams in the human brain.

### Differences in temporal processing across visual streams

Our results suggest two key differences between temporal processing in the ventral and lateral visual processing streams which project to ventral and lateral temporal cortex, respectively [41]. First, there are differences in channel contributions. Lateral temporal cortex is dominated by responses to visual transients, while ventral temporal cortex responds to both sustained

and transient visual information. Transient processing in LTC is consistent with the view that face and body-selective regions in the STS and MTG, respectively, are involved in processing dynamic visual information [30–37]. However, different than prior theories that have implicated these lateral regions in specialized processing of biological motion [31–33, 42], our data suggest that there is a more fundamental difference between high-level visual regions in lateral and ventral temporal cortex that is driven by differential temporal channel contributions. Second, there are also differences in the dynamics of transient processing across visual streams. LTC regions show equal increases in neural responses due to the onset and offset of a visual stimulus, suggesting they carry information about moment-to-moment changes in the visual input. However, VTC regions exhibit surprisingly asymmetric contributions from the onset and offset of the stimulus. That is, the accumulation of fMRI responses due to the termination of a stimulus is more pronounced than responses associated with its onset. This difference suggests the intriguing possibility that transient responses in LTC code progressive changes to the visual input, while transient offset responses in VTC may reflect memory traces that are maintained in high-level regions after a stimulus is no longer visible. This prediction is consistent with results from ECoG studies showing that high frequency broadband responses (>60 Hz) in VTC continue for 100–200 ms after the stimulus is off [43–46] and carry stimulus-specific information that may be modulated by attention [45, 46].

Observing a strong transient response in LTC regions, MTG-bodies and pSTS-faces, is interesting in the context of classic theories that propose differential contributions of magnocellular (M) and parvocellular (P) inputs to parallel visual streams in the primate visual system [22–24]. In macaques, the M pathway is thought to code transient visual information and projects from V1 to MT, while the P pathway is thought to code sustained information and projects from V1 to V4 and IT. Our results reveal that the transient channel, associated with the M pathway, dominates responses not just in hMT+ [5] but also in LTC category-selective regions. In turn, this suggest the intriguing possibility that there may be substantial M projections not only to hMT+ as predicted by classic theories [22–24], but also to surrounding face- and body-selective regions.

Different from the predictions of classic theories of a predominant P input to the primate ventral stream [22–24], we find significant contributions from both transient and sustained channels in VTC as well as evidence for category selectivity in VTC in both channels. This finding is consistent with later studies in macaques that reported that both M and P inputs propagate to ventral visual areas such as V4 [5, 25]. Surprisingly, our data in Fig 5 suggests that the contribution of the transient channel in VTC appears to be larger than the sustained channel. We note that while interpreting the relative amplitude of responses within a channel is straightforward (e.g. comparing $\beta$ weights for the different categories within the transient channel), interpreting the relative weight of sustained vs. temporal channels is complex, as it depends on the specific implementation of the model and the experimental design. Nonetheless, we are confident that there are both sustained and transient responses in VTC for three reasons. First, examination of raw BOLD responses during our experiments (S1 Fig), which are model free, shows that VTC regions respond strongly both to sustained single images (experiment 1) as well as transient, briefly flashed images (experiment 2). Second, responses in experiment 3, which had a combination of sustained and transient stimulation, exceed those of either experiment 1 or 2, suggesting additive contributions of the two channels. Finally, a two-channel model with sustained and transient channels performs better in VTC than single-channel models with only a sustained channel or only a transient channel (S6 Fig).

Critically, finding substantial transient responses in VTC suggests a rethinking of the role of transient processing in the ventral visual stream. That is, this finding provides evidence

against the prevailing theoretical view that the role of the ventral stream is just to process static visual information. We hypothesize that transient responses in the ventral steam may serve two purposes. First, onset transients may reflect the processing of novel stimuli, which underlie rapid extraction of the gist of the visual input. Second, offset transients in VTC may reflect the ignition of a memory trace of the stimulus after it is no longer visible.

## Differences in temporal processing across early and high-level stages of visual processing

Notably, the estimated timing parameters from our experiments are largely consistent with parameters of neural IRFs derived from compressive temporal models applied to fMRI [6], as well ECoG and electrophysiology data [47], which have millisecond temporal resolution. Another aspect of our results shows that temporal parameters of neural responses vary across early and high-level areas in the visual processing hierarchy [1, 19–21]. Evidence for hierarchical differences in temporal processing is reflected in two ways. First, our model estimates that the time of peak responses of neural IRFs is later in both intermediate visual areas and high-level VTC regions relative to V1 (Fig 6). Second, our data suggests faster adaptation in the sustained channel in VTC regions compared to V1 (Fig 6A).

Nonetheless, not all of our data follow the predictions of hierarchical differences in temporal processing. For example, the time to peak of neural IRFs in mFus-faces (OTS-bodies) is earlier than pFus-faces (ITG-bodies), even though the former two are thought to be higher in the processing hierarchy than the latter. This deviation from the hierarchical view may be due to the impact of additional factors on neural response latencies, which may also vary across areas. For example, the contrast of images may affect the time to peak in V1 more than in higher-level visual regions [47, 48].

## What are the implications for modeling fMRI responses beyond visual cortex?

Our data has critical implications for computational models of the brain. We developed a parsimonious yet powerful encoding model that can be applied to estimate nonlinear neural responses and temporal integration windows across cortex with millisecond resolution. While our two-temporal channel model provides a significant improvement in predicting fMRI signals compared to other models, we acknowledge that it does not explain the entire variance of the data. Future research may build upon the present results and improve model predictions by incorporating additional nonlinearities and channels. In terms of nonlinearities, future research could examine if there are also adaptation effects in the transient channel by modeling transient responses to repeated vs. non repeated stimuli [49]. In terms of processing channels, combining the temporal encoding approach with models of spatial processing such as population receptive field models [50] and featural processing models [51, 52] may be important for accounting for the remaining unexplained variance of fMRI responses. Further, encoding models with temporal, spatial, and featural components may be necessary to accurately predict brain responses to dynamic real-world visual inputs in higher-level regions [19, 20, 52].

Given the pervasive use of the standard GLM in fMRI research, our results have broad implications for fMRI studies of any part of the brain. We find that varying the timing of stimuli in the millisecond range has a substantial impact on the magnitude of fMRI responses. However, by estimating neural responses in millisecond resolution, we can accurately predict fMRI responses in second resolution for both brief and long visual stimuli. Thus, the temporal encoding approach we pioneered marks a transformative advancement in using fMRI to elucidate temporal processing in the brain as it links fMRI responses to the timescale of neural

computations. In other words, our approach could be applied to study other brain regions, For example, neurons in auditory cortex are sensitive not only to the frequency of tones, but also to their timing and duration [53]. By varying the timing parameters of auditory stimuli and fitting a temporal (or spectral-temporal [54]) encoding model to brain responses, the framework we developed here could be used to uncover the shape and timing of neural impulse response functions that characterize auditory cortex, as well as temporal processing of complex stimuli such as speech and music [55, 56].

Additionally, as parallel streams occur not just in the visual system but throughout the brain, our data raise the intriguing hypothesis that temporal processing may also segregate other brain systems such as auditory or somatosensory cortex. For example, temporal computations in the auditory ventral stream are thought to differ from those in the auditory dorsal stream [57], whereby the latter may be dominated by processing of auditory transients. Others have also suggested hemispheric differences in auditory cortex; in particular, that the temporal resolution of neural processing is higher in left than right auditory cortex [58]. These hypotheses can be explicitly tested by developing temporal encoding models for auditory cortex like the ones we have developed here for visual cortex.

Overall, our innovative approach offers a quantitative framework to identify functional and computational differences across cortex [59, 60] in many domains such as audition [61] and working memory [62]. Importantly, the encoding approach can also be applied to study impairments in high-level abilities like reading [63] and mathematical processing [64] that require integrating visual information over space and time.

In sum, our results provide the first comprehensive computational model of temporal processing in high-level visual cortex. Our findings propose a fundamental new mechanism —temporal processing—that distinguishes visual processing streams. We propose that lateral category-selective regions process moment-to-moment visual transitions, but ventral category-selective regions respond to both sustained and transient components. Visual transients in ventral category-selective regions may reflect rapid detection of changes to the visual content at stimulus onset and a memory trace of a recent stimulus at stimulus offset, which together suggest a new role of transient processing in the visual system beyond processing of dynamic stimuli. Finally, the encoding approach we introduce underscores the importance of modeling brain responses with millisecond precision to better understand the underlying neural computations.

## Materials and methods

### Ethics statement

The Stanford University Institutional Review Board approved of the study (Protocol #29458—Functional Neuroanatomy of High-Level Visual Cortex: A quantitative multi-model approach). We obtained written informed consent by each subject.

### Participants

Twelve participants (6 males, 6 females) with normal or corrected-to-normal vision participated in the main experiments (experiments 1–3). Each individual provided written informed consent and participated in two fMRI sessions: one session for experiments 1 and 2 and another session for experiment 3 and a functional localizer experiment [15]. Seven participants from the main experiments (3 males, 4 females) also underwent population receptive field (pRF) mapping [50] to define retinotopic cortical regions and another experiment to define human motion-sensitive area (hMT+) [65]. The Stanford Internal Review Board on Human Subjects Research approved all protocols.

## Temporal channels experiments

**Visual stimuli.** Stimuli consisted of well-controlled grayscale images of faces, bodies, and pseudowords (Fig 1*A*, *right*) used in our previous publications [15]. Stimuli were presented using an Eiki LC-WUL100L projector (resolution: 1920 x 1200; refresh rate: 60 Hz) that was controlled by an Apple MacBook Pro using MATLAB (http://www.mathworks.com/) and functions from Psychophysics Toolbox [66] (http://psychtoolbox.org). Participants viewed images through an auxiliary mirror mounted on the RF coil with stimuli spanning ~20˚ of visual angle in each dimension.

**Experimental design.** To develop a temporal encoding model for high-level visual cortex, we adapted a fMRI paradigm previously used to model contributions of sustained and transient temporal channels in early visual cortex [5]. The three main experiments in this study all used the same stimuli, trial durations, and task but varied in the temporal presentation of the images. Critically, a 12-s baseline period (blank gray screen) always came before and after each trial. In all three experiments, participants were instructed to fixate on a small, central dot and respond by button press when it changed color (occurring randomly once every 2–14 s, 8 s on average).

Experiment 1 – one continuous image per trial: Stimuli were shown in trials of varying durations (3, 5, 10, or 20 s per trial) in which a single image was shown for the entire trial. Across trial durations the number of stimuli and transients (at the onset and offset of each stimulus) are matched but the duration of stimulation varies (Fig 1*A* and 1*B*, *blue*). This experiment was designed to enable measurement of sustained responses as well as fMRI-adaptation for prolonged images [14].

Experiment 2 – 30 flashed images per trial: used the same trial durations as experiment 1, but in each trial we presented 30 different images from the same category. Each image was shown for 33 ms and followed by a blank interstimulus interval (ISI). Across trial durations the number of stimuli, number of transients, and total duration of visual stimulation are matched, but the ISI between consecutive images varied. Each ISI was 67 ms in the 3-s trials, 133 ms in the 5-s trials, 300 ms in the 10-s trials, and 633 ms in the 20-s trials (Fig 1*A* and 1*B*, *red*).

Experiment 3 – 30 longer images per trial: used the same design as experiment 2, except that in each trial we presented 30 images from the same category for longer durations with a constant ISI of 33 ms between images. Image durations varied across trials and were each shown for 67 ms in the 3-s trials, 133 ms in the 5-s trials, 300 ms in the 10-s trials, and 633 ms in the 20-s trials (Fig 1*A* and 1*B*, *green*).

**Data acquisition.** Functional data were acquired using a simultaneous multi-slice EPI sequence with a multiplexing factor of 3 to obtain near whole-brain coverage with a TR of 1 s. Participants viewed four 270-s runs of each experiment. Each run of each experiment contained one instance of every permutation of stimulus category (face/body/word) and trial duration (3, 5, 10, or 20 s) presented in random order.

**Category localizer experiment.** To functionally define cortical regions that respond preferentially to specific stimulus categories, we collected three 300-s runs of a standard fMRI category localizer experiment used in our previous publications [15]. Participants were instructed to fixate on a central dot and respond by button press when an image repeated randomly within a block. Code for the experiment is available at https://github.com/VPNL/fLoc.

**pRF mapping and hMT+ localizer.** To delineate retinotopic boundaries, we acquired four 200-s runs of pRF mapping [50] in a subset of participants from the main experiments. In this experiment, a bar with flickering black and white checkerboards swept across a circular aperture (40˚ × 40˚ of visual angle) in eight directions as participants performed a fixation

task. To functionally define hMT+ in the same subset of participants, we collected one 300-s run of a fMRI motion localizer experiment as detailed in our prior publications [5, 65].

**Magnetic resonance imaging (MRI).** MRI data were collected using a 3T GE Signa MR750 scanner at the Center for Cognitive and Neurobiological Imaging (CNI) at Stanford University.

fMRI: We used a Nova phase-array 32-channel head coil for the main experiments and functional localizer to obtain near whole-brain coverage (48 slices; resolution: 2.4 × 2.4 × 2.4 mm; one-shot T2*-sensitive gradient echo acquisition sequence: FOV = 192 mm, TE = 30 ms, TR = 1000 ms, and flip angle = 76˚, multiplexing factor of 3). We also collected T1-weighted inplane images to align each participant's functional data to their high-resolution whole brain anatomy.

For pRF mapping and the hMT+ localizer, we used a 16-channel visual array coil (28 slices; resolution: 2.4 × 2.4 × 2.4 mm; one-shot T2*-sensitive gradient echo acquisition sequence: FOV = 192 mm, TE = 30 ms, TR = 2000 ms, and flip angle = 77˚) and collected T1-weighted inplane images in the same prescription.

Anatomical MRI: We acquired a whole-brain, anatomical volume in each participant using a Nova 32-channel head coil (resolution: 1 × 1 × 1 mm; T1-weighted BRAVO pulse sequence: FOV = 240 mm, TI = 450 ms, and flip angle = 12˚).

## Data analysis

Data were analyzed with MATLAB using code from vistasoft (http://github.com/vistalab) and FreeSurfer (http://freesurfer.net). Code used for predicting fMRI responses using a temporal channels approach is available at https://github.com/VPNL/TemporalChannels.

**Regions of interest (ROI) definition.** Category-selective regions were defined in each participant's native anatomical space at a common threshold ($t > 3$, voxel level, uncorrected) using functional and anatomical criteria detailed in prior publications [15] (Fig 1C). Face-selective ROIs (faces > others) were defined bilaterally in the inferior occipital gyrus (IOG-faces, $N = 10$), posterior STS (pSTS-faces, $N = 9$), posterior fusiform gyrus (pFus-faces, $N = 11$), and mid fusiform gyrus (mFus-faces; $N = 11$). Body-selective ROIs (bodies > others) were found bilaterally in the lateral occipital sulcus (LOS-bodies, $N = 10$), inferior temporal gyrus (ITG-bodies, $N = 10$), middle temporal gyrus (MTG-bodies, $N = 11$), and occipitotemporal sulcus (OTS, $N = 10$).

Visual areas V1 and hV4 were defined in each hemisphere in a subset of participant ($N = 7$) using data from the pRF mapping experiment. To match the visual field coverage of the stimuli in the main experiments, we restricted ROIs to only included voxels with pRF centers within the central 10˚. We also defined bilateral hMT+ in the same subset of participants using data from the motion localizer experiment as in previous publications [5, 65].

**Optimized two-temporal channel A+S model.** To predict responses across all three experiments with a single model, we adapted an encoding approach introduced by prior studies [5, 18], which models fMRI responses as the weighted sum of activity across separate sustained and transient temporal channels.

In the procedure illustrated in Fig 3, we first predict neural activity in each channel by convolving the stimulus time course in millisecond resolution (Fig 3A) separately with the neural IRF for the sustained channel (Fig 3B, *blue channel IRF*) and the transient channel (Fig 3B, *red channel IRF*). The sustained channel is characterized by a monophasic $IRF_S$ that generates a response for the entire duration of a stimulus followed by an adaptation nonlinearity. This is implemented by multiplying the predicted neural responses in the sustained channel by an exponential decay function beginning at the onset of each stimulus and extending until the onset of the following stimulus. In contrast, the transient channel is characterized by a biphasic $IRF_T$ that generates a brief response at the onset and offset of an image [2, 3, 40]. Here,

convolved responses are passed through sigmoid nonlinearities that allow different levels of compression to be applied to the "on" and "off" responses. Then, the estimated neural responses for each channel are convolved with a hemodynamic response function (HRF) to generate a prediction of the fMRI response in each channel (Fig 3*C*). As such, there are neural nonlinearities in each channel of this model, but a linear relationship is assumed between the neural activity and BOLD responses. Finally, we use a GLM to solve for the contributions ($\beta$ weights) of the sustained and transient channels, which reflect how much the predicted response from each channel is scaled before the responses of both channels are summed. Thus, the BOLD response to a stimulus can be expressed as

$$\beta_S\{[(stimulus \otimes IRF_S) \cdot e^{-t/\alpha}] \otimes HRF\} + \beta_T\{[\sigma(stimulus \otimes IRF_T)] \otimes HRF\},$$

$$\sigma(x) = \begin{cases} 1 - e^{-\left(x/\lambda\right)^{k_{on}}}, x \geq 0 \\ 1 - e^{-\left(-x/\lambda\right)^{k_{off}}}, x < 0 \end{cases}$$

where $\beta_S$ and $\beta_T$ are the fitted response amplitudes for the sustained and transient channels, respectively; $IRF_S$ and $IRF_T$ are the impulse response functions for the sustained and transient channels, respectively; α determines the exponential decay at time $t$ after stimulus onset; σ is a pointwise sigmoid nonlinearity, and *HRF* is the canonical hemodynamic response function.

**Modeling nonlinearities in the neural response.** We model the IRFs for each channel (Fig 3*B*) using formulas detailed in our prior publications [5]. Here, we optimize the IRF time constant τ for each region, and the other parameters (taken from Watson [40]) are held constant: κ = 1.33, $n_1$ = 9, and $n_2$ = 10.

Adaptation: To capture fMRI-adaptation [14] effects in the sustained channel, we use an exponential decay function, $e^{-t/\alpha}$, where $t$ represents time after stimulus onset, and α indicates when the function declines to a proportion of 1/$e$ (~37%) of the initial response.

Sigmoid nonlinearities: To allow different levels of compression to be applied to "on" and "off" responses in the transient channel, we optimize separate sigmoid nonlinearities for the onset and offset responses using cumulative Weibull distribution functions,

$$\sigma(x) = \begin{cases} 1 - e^{-\left(x/\lambda\right)^{k_{on}}}, x \geq 0 \\ 1 - e^{-\left(-x/\lambda\right)^{k_{off}}}, x < 0 \end{cases}$$

where λ is a sigmoid scale parameter used in both onset and offset nonlinearities; $k_{on}$ is a sigmoid shape parameter that controls the curvature of the onset compression function, and $k_{off}$ is a shape parameter controlling curvature of the offset compression function. Smaller $k$ values produce more compressive nonlinearities that elongate transient "on" and "off" responses compared to larger $k$ values.

**Fitting and optimizing the two-temporal channel model.** Since the HRF acts like a temporal low-pass filter, this allows resampling fMRI response predictors to the lower temporal resolution of the measured fMRI data (TR = 1 s) with minimal distortion. These resampled predictors are then compared with measured fMRI responses to estimate the contributions ($\beta$ weights) of each channel for each category. To normalize the amplitude of predicted fMRI responses for the sustained and transient channels, we match the maximal height of predictors in the design matrix across the two channels. Finally, we used a GLM to estimate $\beta$ weights of the sustained and transient channels for each stimulus category by comparing the predicted responses with the mean response time series of each ROI in each participant. To optimize the A+S model time constants (τ and α) and sigmoid parameters (λ, $k_{on}$, $k_{off}$) for each region, we

used the constrained nonlinear optimization algorithm *fmincon* in MATLAB (*Optimization procedures*).

**Validating the optimized two-temporal channel model.** We assessed the predictive power of the optimized two-temporal A+S channel model by testing how well it predicts responses from separate runs of data from all three experiments. We first generated predicted neural response time courses by coding the visual stimulation in the left-out runs and convolving it separately with the IRFs of the sustained and transient channels (optimized using a separate split of the data). We then applied the adaptation and sigmoid nonlinearities, which were also optimized with independent data (Fig 6). These transformed neural predictors were next convolved with the HRF and down-sampled to 1 s temporal resolution to match our fMRI acquisition. Finally, we multiplied each channel's fMRI predictors with their respective $\beta$ weights (estimated for each category in an independent split of the data) before summing the channel responses to predict fMRI responses. We then quantified how well the predicted responses matched the measured response across all data in the validation split.

Model performance was operationalized as cross-validated $R^2$ (x-$R^2$), which indexes the proportion of variance explained by $\beta$ weights and model parameters that were estimated from independent data. While similar to a typical $R^2$ statistic, x-$R^2$ can be negative when the residual variance of a poor model prediction exceeds the measured variance in the response. Quantification of x-$R^2$ within each experiment is presented in Fig 4*D*–4*F* for OTS-bodies. Performance averaged across all three experiments is shown in S3–S5 Figs for all regions.

**Testing alternative model architectures.** To compare with the performance of our optimized two-temporal channel A+S model with alternatives, we tested five single-channel and three dual-channel models (Figs 2 and 4, S3–S5 Figs).

General linear model (GLM): To first benchmark our model against a common GLM approach [8], we tested a linear model that predicts fMRI responses with a single convolution of the stimulus with the canonical HRF.

Balloon model (B model): To examine if responses in high-level visual cortex can be explained by a nonlinear hemodynamic model, we implemented the balloon model proposed by Buxton an colleagues [7] using standard parameters detailed in prior publications [5].

Linear sustained channel (L model): Similar to the GLM approach [8] but with two stages of convolution, we tested a single-channel model with a linear sustained channel,

$$\beta\,[(\text{stimulus} \otimes IRF_S) \otimes HRF],$$

where $\beta$ is a fitted response amplitude; $IRF_S$ is the impulse response function for the sustained channel, and *HRF* is the canonical HRF.

Sustained channel with compressive temporal summation (CTS, Fig 4, S2–S5 Figs): We also implemented a model proposed by Zhou et al. [6] composed of sustained channel with a compressive static power law,

$$\beta\,[(\text{stimulus} \otimes IRF_S)^{\epsilon} \otimes HRF],$$

where ε is an optimized compression parameter ranging from 0–1.

Sustained channel with adaptation (A model): Identical to the sustained channel shown in Fig 3 (blue), we tested a model composed of a single sustained channel with adaptation,

$$\beta\,[(\text{stimulus} \otimes IRF_S) \cdot e^{-t/\alpha} \otimes HRF],$$

where α determines the exponential decay at *t* seconds after the onset of a stimulus.

Transient channel with sigmoid nonlinearity (S model; S6 Fig): Identical to the transient channel shown in Fig 3 (red), we also tested a single-channel model composed of a transient

channel with the same sigmoid nonlinearities described above,

$$\beta\{[\sigma(stimulus \otimes IRF_T)] \otimes HRF\}$$

where $IRF_T$ is the impulse response function for the transient channel and σ is a pointwise nonlinearity composed of separate sigmoid functions for onset and offset responses.

Alternative dual-channel models (L+Q, C+Q, and A+Q models): To compare the optimized two-temporal channel model shown in Fig 3 (A+S model) to alternative dual-channel models, we tested three variants of our model that all use a transient channel with a quadratic (Q) nonlinearity (squaring) but apply different nonlinearities in the sustained channel (S3*A* Fig). Combining different combinations of the sustained and transient channels described above, we compared two-channel models composed of a transient channel and either a linear sustained channel (L+Q model), a sustained channel with CTS (C+Q model), or a sustained channel with adaptation (A+Q model).

**Optimization procedures.** For all models with a neural IRF, we optimized a single time constant, τ, using formulas described in our prior publications [5]. For models with adaption in the sustained channel (A and A+S), we also optimized an exponential time constant (α). For models with compressive temporal summation (CTS and C+Q models), we instead optimized an exponential compression parameter (ε). For models with a sigmoid nonlinearity in the transient channel (S and A+S models), we optimized three sigmoid parameters (λ, $k_{on}$, $k_{off}$). To optimize model parameters, we used the nonlinear optimization algorithm *fmincon* in MATLAB with the following constraints: τ = 4–20 ms, α = 10–40 s, ε = 0.01–1, λ = 0.01–0.5, $k_{on}$ = 0.1–6, and $k_{off}$ = 0.1–6. The initial values passed to the optimizer for each parameter were τ = 4.93 ms, α = 20 s, ε = 0.1, λ = 0.1, $k_{on}$ = 3, and $k_{off}$ = 3. The cross-validation performance of each model averaged across all three experiments is shown in S3*B* Fig for category-selective regions in VTC and LTC and in S4 and S5 Figs for other regions.

**Statistical analyses.** *Model-free ROI comparison*. To examine differences in the patterns of response between ventral and lateral body-selective regions in Fig 2*B* and 2*C* using a model-free approach, we measured in each participant and region the peak response amplitude to body stimuli for each temporal condition. Then we compared these peaks across regions and conditions using a three-way repeated measure analysis of variance (ANOVA) with factors of trial duration (1, 3, 5, or 10 s), experiment (experiment 1, 2, or 3), and ROI (OTS-bodies vs. MTG-bodies).

**Model comparison.** To test for differences in model cross-validation performance across regions in VTC and LTC, we used a two-way repeated measures ANOVA with factors of model and region (comparing models and regions shown in S3*B* Fig). We then used post-hoc paired two-tailed *t*-tests to compare the x-$R^2$ of our model with others. Fig 4*D*–4*F* contrasts the performance of our model (A+S) in OTS-bodies against three other models (GLM, CTS, L+Q) for each experiment individually. S3–S5 Figs contrast the performance of our model averaged across all three experiments vs. every other model for each region. To assess the level of noise in measurements from different brain regions, we also calculated a noise ceiling for each ROI using the inter-trial variability of responses for each condition as described in our prior publications [5]. The noise ceiling estimate for OTS-bodies in each experiment is plotted in Fig 4*D*–4*F*, and the average noise ceiling across all three experiments is plotted in S3–S5 Figs.

**Parameter comparison.** After establishing the validity of our model, we used paired two-tailed *t*-tests to compare *β* weights estimated by the A+S model for each region's preferred category vs. average contributions for nonpreferred categories, separately for the sustained and transient channels. To test whether selectivity in the two channels differs across regions preferring bodies and faces in either VTC or LTC, we also used two-way ANOVAs with factors of

channel (sustained/transient) and preferred category (bodies/faces) on the difference in channel weights for preferred vs. nonpreferred categories (contrast effect size, CES; Fig 5). To examine whether the proportion of response attributed to sustained vs. transient channels differs across processing streams, stimulus categories, or regions preferring different categories, we then used a three-way ANOVA on channel contribution ratios, $|\frac{\beta_S}{\beta_T}|$, for each category with factors of stream (ventral/lateral), stimulus (faces/bodies/words), and preferred category (bodies/faces).

## Supporting information

**S1 Fig. Responses to time-varying stimuli in occipital, ventral, and lateral regions of interest.** Measured responses in occipital (V1, LOS-bodies, IOG-faces), ventral (hV4, OTS-bodies, mFus-faces), and lateral (hMT+, MTG-bodies, pSTS-faces) regions of interest in experiment 1 (*blue*), experiment 2 (*red*), and experiment 3 (*green*) averaged across all three stimulus categories. *Lines*: mean response time series across participants; *shaded areas*: standard error of the mean (SEM) across participants; *Horizontal black bars*: trial duration.
(TIF)

**S2 Fig. Comparison of temporal encoding models in OTS-bodies.** (A-C) Responses and model predictions for body images in OTS-bodies for each experiment (*left*) with estimated $\beta$ weights for each model (*right*). *White curve*: mean response across 10 participants. *Shaded gray*: standard deviation across participants. *Black curve*: overall model prediction. *Horizontal black bar*: trial duration. (*A*) Predictions of a general linear model (GLM) [8]. (*B*) Predictions of a model with compressive temporal summation (CTS) [6]. (*C*) Predictions of the two-temporal channel L+Q model with linear sustained channel and quadratic transient channel. *Blue curve*: predicted response from the sustained channel. *Red curve*: predicted response from the transient channel: *Black curve*: sum of responses from both channels. In the continuous (left) and flashed images (middle) experiments the model's prediction (black) is obscured by the response of a single channel, as the other channel's contribution is negligible.
(TIF)

**S3 Fig. Comparison of temporal encoding models across high-level visual cortex.** (*A*) Alternative models of sustained (*blue*) and transient (*red*) channels. Schematic depicts neural response predictions generated by different implementations of each channel for both a brief (67 ms) and long (3 s) stimulus. Sustained channel models: *L*, a linear sustained channel; *CTS*, a sustained channel with compressive temporal summation [6]; *A*, a sustained channel with adaptation [8]. Transient channel models: *Q*, a transient channel with a quadratic (squaring) nonlinearity; *S*, a transient channel with a sigmoid nonlinearity. (*B*) Comparison of model performance (cross-validated $R^2$) in each region averaged across all three experiments. Hemodynamic models: *L*, same as in (a); *B*, balloon model [7]. Single-channel neural models: *CTS*, *A*, and *S*, same as in (a). Two-channel neural models: *L*+*Q*, a linear sustained channel and a transient channel with a quadratic nonlinearity [5]; *C*+*Q*, a sustained channel with compressive temporal summation and a transient channel with a quadratic nonlinearity; *A*+*Q*, a sustained channel with adaptation and a transient channel with a quadratic nonlinearity; *A*+*S*, a sustained channel with adaptation and a transient channel with a sigmoid nonlinearity. Cross-validated $R^2$ significantly differs across models in all four regions (significant main effect of model type, $F$s $> 6.80$, $P$s $< .001$, one-way repeated measures ANOVA for each region). Asterisks denote models with significantly different performance compared to the A+S model, $p<0.05$.
(TIF)

**S4 Fig. Contributions of transient and sustained temporal channels across early and intermediate visual areas.** (*A*) Contributions ($\beta$ weights) of transient (*x* axis) and sustained (*y* axis) channels for each stimulus category estimated by the two-temporal channel A+S model in V1, hV4, and hMT+. Crosses span ±1 SEM across participants in each axis, and $\beta$ were solved by fitting the model using data concatenated across all experiments. Data show average model weights across all splits of the data for each participant. *Red*: response to faces. *Blue*: response to bodies. *Gray*: response to words. *Dashed gray*: identity line ($\beta_S = \beta_T$). (*B*) Comparison of model performance (cross-validated $R^2$) in each region averaged across all three experiments. Hemodynamic models: *L* and *B*. Single-channel neural models: *CTS*, *A*, and *S*. Two-channel neural models: *L+Q* [5], *C+Q*, *A+Q*, *and A+S*. Cross-validated $R^2$ significantly differs across models in all three regions (significant main effect of model type, $F$s $> 16.45$, $P$s $< .001$, one-way repeated measures ANOVA for each region). Asterisks denote models with significantly different performance vs. the A+S model.
(TIF)

**S5 Fig. Contributions of transient and sustained temporal channels across other face- and body-selective regions.** (*A*) Contributions ($\beta$ weights) of transient (*x* axis) and sustained (*y* axis) channels for each stimulus category estimated by the two-temporal channel A+S model. Crosses span ±1 SEM across participants in each axis, and $\beta$ were solved by fitting the model using data concatenated across all experiments. Data show average model $\beta$ weights across all splits of the data for each participant. *Red*: response to faces. *Blue*: response to bodies. *Gray*: response to words. *Dashed gray*: identity line ($\beta_S = \beta_T$). (*B*) Comparison of model performance (cross-validated $R^2$) in each region averaged across all three experiments. Hemodynamic models: *L* and *B*. Single-channel neural models: *CTS*, *A*, and *S*. Two-channel neural models: *L+Q* [5], *C+Q*, *A+Q*, *and A+S*. Cross-validated $R^2$ significantly differs across models in all four regions (significant main effect of model type, $F$s $> 36.00$, $P$s $< .001$, one-way repeated measures ANOVA for each region). Asterisks denote models with significantly different performance vs. the A+S model, $p<0.05$.
(TIF)

**S6 Fig. Single-channel model with transient channel and sigmoid nonlinearity applied to OTS-bodies.** (A-C) Responses for body images in OTS-bodies and predictions of a model with a transient channel, but no sustained channel. *White curve*: mean response across 10 participants. *Shaded gray*: standard deviation across participants. *Black*: predicted response from the transient channel: *Inset*: *mean* contribution ($\beta$ weight) for the transient channel ±1 SEM across participants. (*A*) Experiment 1 data, 1 continuous image per trial. (*B*) Experiment 2 data, 30 flashed images per trial. (*C*) Experiment 3 data, 30 longer images per trial. (D-F) Model comparison. Bars show the performance of various models for each experiment presented in (A-C). Models are fit using runs from all three experiments, and cross-validation performance (x-$R^2$) is calculated in left-out data from each experiment separately. (*D*) Experiment 1. (*E*) Experiment 2. (*F*) Experiment 3. Single-channel models: G*LM*, general linear model [8]; *A*, a sustained channel with adaptation; *S*, a transient channel with a sigmoid nonlinearities; *A+S*: a sustained channel with adaptation and a transient channel with sigmoid nonlinearities. Cross-validated $R^2$ significantly differs across models [significant main effect of model type, $F_{6, 54} = 21.88$, $P < .001$, two-way repeated measures ANOVA with factors of model type (GLM/A/S/A+S) and experiment (1/2/3)]. Asterisks denote models with significantly different performance compared to A+S (paired *t*-tests comparing x-$R^2$ of each model vs. A+S in each experiment).
(TIF)

**S7 Fig. Contributions of transient and sustained temporal channels across word-selective regions.** (*A*) Contributions ($\beta$ weights) of transient (*x* axis) and sustained (*y* axis) channels for each stimulus category estimated by the two-temporal channel A+S model in additional word-selective regions. Crosses span ±1 SEM across participants in each axis, and $\beta$ were solved by fitting the model using data concatenated across all experiments. Data show average model weights across all splits of the data for each participant. *Red*: response to faces. *Blue*: response to bodies. *Gray*: response to words. *Dashed gray*: identity line ($\beta_S = \beta_T$). (*B*) Comparison of model performance (cross-validated $R^2$) in each region averaged across all three experiments. Hemodynamic models: *L* and *B*. Single-channel neural models: *CTS*, *A*, and *S*. Two-channel neural models: *L+Q* [5], *C+Q*, *A+Q*, *and A+S*. Cross-validated $R^2$ significantly differs across models in all three regions (significant main effect of model type, $F$s $> 10.17$, $P$s $< .001$, one-way repeated measures ANOVA for each region). Asterisks denote models with significantly different performance vs. the A+S model, $<0.05$.
(TIF)

## Acknowledgments

We thank Jon Winawer and Jing Zhou for fruitful discussions.

## Author Contributions

**Conceptualization:** Anthony Stigliani, Kalanit Grill-Spector.

**Data curation:** Anthony Stigliani, Brianna Jeska, Kalanit Grill-Spector.

**Formal analysis:** Anthony Stigliani, Kalanit Grill-Spector.

**Funding acquisition:** Kalanit Grill-Spector.

**Investigation:** Anthony Stigliani, Kalanit Grill-Spector.

**Methodology:** Anthony Stigliani, Kalanit Grill-Spector.

**Project administration:** Kalanit Grill-Spector.

**Resources:** Kalanit Grill-Spector.

**Software:** Anthony Stigliani, Kalanit Grill-Spector.

**Supervision:** Kalanit Grill-Spector.

**Validation:** Anthony Stigliani, Kalanit Grill-Spector.

**Visualization:** Anthony Stigliani, Kalanit Grill-Spector.

**Writing – original draft:** Anthony Stigliani, Kalanit Grill-Spector.

**Writing – review & editing:** Anthony Stigliani, Kalanit Grill-Spector.

## References

1. Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, et al. Signal timing across the macaque visual system. J Neurophysiol. 1998; 79(6):3272–8. https://doi.org/10.1152/jn.1998.79.6.3272 PMID: 9636126
2. De Valois RL, Cottaris NP. Inputs to directionally selective simple cells in macaque striate cortex. Proc Natl Acad Sci U S A. 1998; 95(24):14488–93. https://doi.org/10.1073/pnas.95.24.14488 PMID: 9826727
3. Conway BR, Livingstone MS. Space-time maps and two-bar interactions of different classes of direction-selective cells in macaque V-1. J Neurophysiol. 2003; 89(5):2726–42. https://doi.org/10.1152/jn.00550.2002 PMID: 12740411

4. Nandy AS, Mitchell JF, Jadi MP, Reynolds JH. Neurons in Macaque Area V4 Are Tuned for Complex Spatio-Temporal Patterns. Neuron. 2016; 91(4):920–30. https://doi.org/10.1016/j.neuron.2016.07.026 PMID: 27499085
5. Stigliani A, Jeska B, Grill-Spector K. Encoding model of temporal processing in human visual cortex. Proc Natl Acad Sci U S A. 2017; 114(51):E11047–E56. https://doi.org/10.1073/pnas.1704877114 PMID: 29208714
6. Zhou J, Benson NC, Kay KN, Winawer J. Compressive Temporal Summation in Human Visual Cortex. J Neurosci. 2018; 38(3):691–709. https://doi.org/10.1523/JNEUROSCI.1724-17.2017 PMID: 29192127
7. Buxton RB, Wong EC, Frank LR. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med. 1998; 39(6):855–64. PMID: 9621908
8. Boynton GM, Engel SA, Glover GH, Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci. 1996; 16(13):4207–21. PMID: 8753882
9. Huettel SA, McCarthy G. Evidence for a refractory period in the hemodynamic response to visual stimuli as measured by MRI. Neuroimage. 2000; 11(5 Pt 1):547–53. https://doi.org/10.1006/nimg.2000.0553 PMID: 10806040
10. Ogawa S, Lee TM, Stepnoski R, Chen W, Zhu XH, Ugurbil K. An approach to probe some neural systems interaction by functional MRI at neural time scale down to milliseconds. Proc Natl Acad Sci U S A. 2000; 97(20):11026–31. https://doi.org/10.1073/pnas.97.20.11026 PMID: 11005873
11. Birn RM, Saad ZS, Bandettini PA. Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. Neuroimage. 2001; 14(4):817–26. https://doi.org/10.1006/nimg.2001.0873 PMID: 11554800
12. Mukamel R, Harel M, Hendler T, Malach R. Enhanced temporal non-linearities in human object-related occipito-temporal cortex. Cereb Cortex. 2004; 14(5):575–85. https://doi.org/10.1093/cercor/bhh019 PMID: 15054073
13. Wager TD, Vazquez A, Hernandez L, Noll DC. Accounting for nonlinear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. Neuroimage. 2005; 25(1):206–18. https://doi.org/10.1016/j.neuroimage.2004.11.008 PMID: 15734356
14. Gilaie-Dotan S, Nir Y, Malach R. Regionally-specific adaptation dynamics in human object areas. Neuroimage. 2008; 39(4):1926–37. https://doi.org/10.1016/j.neuroimage.2007.10.010 PMID: 18061482
15. Stigliani A, Weiner KS, Grill-Spector K. Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. J Neurosci. 2015; 35(36):12412–24. https://doi.org/10.1523/JNEUROSCI.4822-14.2015 PMID: 26354910
16. McKeeff TJ, Remus DA, Tong F. Temporal limitations in object processing across the human ventral visual pathway. J Neurophysiol. 2007; 98(1):382–93. https://doi.org/10.1152/jn.00568.2006 PMID: 17493920
17. Gentile F, Rossion B. Temporal frequency tuning of cortical face-sensitive areas for individual face perception. Neuroimage. 2014; 90:256–65. https://doi.org/10.1016/j.neuroimage.2013.11.053 PMID: 24321556
18. Horiguchi H, Nakadomari S, Misaki M, Wandell BA. Two temporal channels in human V1 identified using fMRI. Neuroimage. 2009; 47(1):273–80. https://doi.org/10.1016/j.neuroimage.2009.03.078 PMID: 19361561
19. Hasson U, Yang E, Vallines I, Heeger DJ, Rubin N. A hierarchy of temporal receptive windows in human cortex. J Neurosci. 2008; 28(10):2539–50. https://doi.org/10.1523/JNEUROSCI.5487-07.2008 PMID: 18322098
20. Honey CJ, Thesen T, Donner TH, Silbert LJ, Carlson CE, Devinsky O, et al. Slow cortical dynamics and the accumulation of information over long timescales. Neuron. 2012; 76(2):423–34. https://doi.org/10.1016/j.neuron.2012.08.011 PMID: 23083743
21. Mattar MG, Kahn DA, Thompson-Schill SL, Aguirre GK. Varying Timescales of Stimulus Integration Unite Neural Adaptation and Prototype Formation. Curr Biol. 2016; 26(13):1669–76. https://doi.org/10.1016/j.cub.2016.04.065 PMID: 27321999
22. Merigan WH, Maunsell JH. How parallel are the primate visual pathways? Annu Rev Neurosci. 1993; 16:369–402. https://doi.org/10.1146/annurev.ne.16.030193.002101 PMID: 8460898
23. Maunsell JH, Nealey TA, DePriest DD. Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. J Neurosci. 1990; 10(10):3323–34. PMID: 2213142
24. Van Essen DC, Gallant JL. Neural mechanisms of form and motion processing in the primate visual system. Neuron. 1994; 13(1):1–10. PMID: 8043270
25. Ferrera VP, Nealey TA, Maunsell JH. Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. J Neurosci. 1994; 14(4):2080–8. PMID: 8158258

26. Kaplan E, Benardete E. The dynamics of primate retinal ganglion cells. Prog Brain Res. 2001; 134:17–34. PMID: 11702542

27. Hubel DH, Wiesel TN. Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. J Comp Neurol. 1972; 146(4):421–50. https://doi.org/10.1002/cne.901460402 PMID: 4117368

28. Schiller PH, Malpeli JG. Functional specificity of lateral geniculate nucleus laminae of the rhesus monkey. J Neurophysiol. 1978; 41(3):788–97. https://doi.org/10.1152/jn.1978.41.3.788 PMID: 96227

29. Derrington AM, Lennie P. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. J Physiol. 1984; 357:219–40. https://doi.org/10.1113/jphysiol.1984.sp015498 PMID: 6512690

30. Bonda E, Petrides M, Ostry D, Evans A. Specific involvement of human parietal systems and the amygdala in the perception of biological motion. J Neurosci. 1996; 16(11):3737–44. PMID: 8642416

31. Puce A, Allison T, Asgari M, Gore JC, McCarthy G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J Neurosci. 1996; 16 (16):5205–15. PMID: 8756449

32. Beauchamp MS, Lee KE, Haxby JV, Martin A. Parallel visual motion processing streams for manipulable objects and human movements. Neuron. 2002; 34(1):149–59. PMID: 11931749

33. Grossman ED, Blake R. Brain Areas Active during Visual Perception of Biological Motion. Neuron. 2002; 35(6):1167–75. PMID: 12354405

34. Fox CJ, Moon SY, Iaria G, Barton JJ. The correlates of subjective perception of identity and expression in the face network: an fMRI adaptation study. Neuroimage. 2009; 44(2):569–80. https://doi.org/10.1016/j.neuroimage.2008.09.011 PMID: 18852053

35. Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N. Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage. 2011; 56(4):2356–63. https://doi.org/10.1016/j.neuroimage.2011.03.067 PMID: 21473921

36. Said CP, Moore CD, Engell AD, Todorov A, Haxby JV. Distributed representations of dynamic facial expressions in the superior temporal sulcus. J Vis. 2010; 10(5):11. https://doi.org/10.1167/10.5.11 PMID: 20616141

37. Freiwald W, Duchaine B, Yovel G. Face Processing Systems: From Neurons to Real-World Social Perception. Annu Rev Neurosci. 2016; 39:325–46. https://doi.org/10.1146/annurev-neuro-070815-013934 PMID: 27442071

38. Gilaie-Dotan S, Saygin AP, Lorenzi LJ, Rees G, Behrmann M. Ventral aspect of the visual form pathway is not critical for the perception of biological motion. Proc Natl Acad Sci U S A. 2015; 112(4):E361–70. https://doi.org/10.1073/pnas.1414974112 PMID: 25583504

39. Weiner KS, Grill-Spector K. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. Neuroimage. 2010; 52(4):1559–73. https://doi.org/10.1016/j.neuroimage.2010.04.262 PMID: 20457261

40. Watson AB. Temporal sensitivity. In: Boff K, Kaufman L, Thomas J, editors. Handbook of Perception and Human Performance. New York: Wiley; 1986.

41. Weiner KS, Grill-Spector K. Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. Psychol Res. 2013; 77(1):74–97. https://doi.org/10.1007/s00426-011-0392-x PMID: 22139022

42. Courtney SM, Ungerleider LG, Keil K, Haxby JV. Transient and sustained activity in a distributed neural system for human working memory. Nature. 1997; 386(6625):608–11. https://doi.org/10.1038/386608a0 PMID: 9121584

43. Fisch L, Privman E, Ramot M, Harel M, Nir Y, Kipervasser S, et al. Neural "ignition": enhanced activation linked to perceptual awareness in human ventral stream visual cortex. Neuron. 2009; 64(4):562–74. https://doi.org/10.1016/j.neuron.2009.11.001 PMID: 19945397

44. Jacques C, Witthoft N, Weiner KS, Foster BL, Rangarajan V, Hermes D, et al. Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. Neuropsychologia. 2016; 83:14–28. https://doi.org/10.1016/j.neuropsychologia.2015.07.024 PMID: 26212070

45. Engell AD, McCarthy G. Selective attention modulates face-specific induced gamma oscillations recorded from ventral occipitotemporal cortex. J Neurosci. 2010; 30(26):8780–6. https://doi.org/10.1523/JNEUROSCI.1575-10.2010 PMID: 20592199

46. Davidesco I, Harel M, Ramot M, Kramer U, Kipervasser S, Andelman F, et al. Spatial and object-based attention modulates broadband high-frequency responses across the human visual cortical hierarchy. J Neurosci. 2013; 33(3):1228–40. https://doi.org/10.1523/JNEUROSCI.3181-12.2013 PMID: 23325259

47. Zhou J, Benson NC, Kay K, Winawer J. Unifying Temporal Phenomena in Human Visual Cortex. bioRxiv. 2018; https://doi.org/10.1101/108639.

48. Avidan G, Harel M, Hendler T, Ben-Bashat D, Zohary E, Malach R. Contrast sensitivity in human visual areas and its relationship to object recognition. J Neurophysiol. 2002; 87(6):3102–16. https://doi.org/10.1152/jn.2002.87.6.3102 PMID: 12037211

49. Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object recognition. Vision Res. 2001; 41(10–11):1409–22. PMID: 11322983

50. Dumoulin SO, Wandell BA. Population receptive field estimates in human visual cortex. Neuroimage. 2008; 39(2):647–60. https://doi.org/10.1016/j.neuroimage.2007.09.034 PMID: 17977024

51. Kay KN, Naselaris T, Prenger RJ, Gallant JL. Identifying natural images from human brain activity. Nature. 2008; 452(7185):352–5. https://doi.org/10.1038/nature06713 PMID: 18322462

52. Çukur T, Huth AG, Nishimoto S, Gallant JL. Functional subdomains within human FFA. J Neurosci. 2013; 33(42):16748–66. https://doi.org/10.1523/JNEUROSCI.1259-13.2013 PMID: 24133276

53. Barton B, Venezia JH, Saberi K, Hickok G, Brewer AA. Orthogonal acoustic dimensions define auditory field maps in human cortex. Proc Natl Acad Sci U S A. 2012; 109(50):20738–43. https://doi.org/10.1073/pnas.1213381109 PMID: 23188798

54. Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, et al. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. PLoS Comput Biol. 2014; 10 (1):e1003412. https://doi.org/10.1371/journal.pcbi.1003412 PMID: 24391486

55. Giraud AL, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 2012; 15(4):511–7. https://doi.org/10.1038/nn.3063 PMID: 22426255

56. Norman-Haignere S, Kanwisher NG, McDermott JH. Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. Neuron. 2015; 88(6):1281–96. https://doi.org/10.1016/j.neuron.2015.11.035 PMID: 26687225

57. Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 2009; 12(6):718–24. https://doi.org/10.1038/nn.2331 PMID: 19471271

58. Zatorre RJ, Belin P, Penhune VB. Structure and function of auditory cortex: music and speech. Trends Cogn Sci. 2002; 6(1):37–46. PMID: 11849614

59. Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proc Natl Acad Sci U S A. 2014; 111 (23):8619–24. https://doi.org/10.1073/pnas.1403112111 PMID: 24812127

60. Kriegeskorte N. Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. Annu Rev Vis Sci. 2015; 1:417–46. https://doi.org/10.1146/annurev-vision-082114-035447 PMID: 28532370

61. Werner S, Noppeney U. The contributions of transient and sustained response codes to audiovisual integration. Cereb Cortex. 2011; 21(4):920–31. https://doi.org/10.1093/cercor/bhq161 PMID: 20810622

62. Druzgal TJ, D'Esposito M. Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. J Cogn Neurosci. 2003; 15(6):771–84. https://doi.org/10.1162/089892903322370708 PMID: 14511531

63. Demb JB, Boynton GM, Heeger DJ. Brain activity in visual cortex predicts individual differences in reading performance. Proc Natl Acad Sci U S A. 1997; 94(24):13363–6. https://doi.org/10.1073/pnas.94.24.13363 PMID: 9371851

64. Harvey BM, Klein BP, Petridou N, Dumoulin SO. Topographic representation of numerosity in the human parietal cortex. Science. 2013; 341(6150):1123–6. https://doi.org/10.1126/science.1239052 PMID: 24009396

65. Weiner KS, Grill-Spector K. Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. Neuroimage. 2011; 56(4):2183–99. https://doi.org/10.1016/j.neuroimage.2011.03.041 PMID: 21439386

66. Brainard DH. The Psychophysics Toolbox. Spat Vis. 1997; 10(4):433–6. PMID: 9176952