

Derivation of visual timing-tuned neural responses from early visual stimulus representations

Evi Hendrikx

Utrecht University

Jacob Paul

University of Melbourne

Martijn van Ackooij

Utrecht University

Nathan van der Stoep

Utrecht University <https://orcid.org/0000-0002-0412-2078>

Ben Harvey (✉ b.m.harvey@uu.nl)

Utrecht University <https://orcid.org/0000-0003-2694-618X>

Article

Keywords: visual stimulus, visual timing-tuned neural responses, visual processing

Posted Date: June 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-579436/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on July 8th, 2022.
See the published version at <https://doi.org/10.1038/s41467-022-31675-9>.

Title: Derivation of visual timing-tuned neural responses from early visual stimulus representations

Authors: Evi Hendrikx¹⁺, Jacob M. Paul¹⁺, M. van Ackooij¹, Nathan van der Stoep¹, Ben M. Harvey^{1*}

⁺ Equal contribution

¹ Experimental Psychology, Helmholtz Institute, Utrecht University, Heidelberglaan 1, Utrecht 3584 CS, Netherlands

*Corresponding author:

Ben M. Harvey

Experimental Psychology, Helmholtz Institute, Utrecht University
Heidelberglaan 1
Utrecht 3584 CS
Netherlands
b.m.harvey@uu.nl

Abstract:

Quantifying the timing (duration and frequency) of brief visual events is vital to human perception, multisensory integration and action planning. Tuned neural responses to visual event timing have been found in areas of the association cortices implicated in these processes. Here we ask whether and where the human brain derives these timing-tuned responses from the responses of early visual cortex, which monotonically increase with event duration and frequency. Using 7T fMRI and neural model-based analyses, we find a gradual transition from monotonically increasing to timing-tuned neural responses beginning in area MT/V5. Therefore, successive stages of visual processing gradually derive timing-tuned response components from the inherent modulation of sensory responses by event timing. This additional timing-tuned response component was independent of retinotopic location. We propose that this hierarchical derivation of timing-tuned responses from sensory processing areas quantifies sensory event timing while abstracting temporal representations from the spatial properties of their inputs.

Introduction

Quantifying spatial and temporal information of events is vital to perceiving and interacting with our environment. For example, accurately determining the duration and frequency of sub-second events is crucial for multisensory integration and motor action planning^{1,2}.

Converging behavioural, computational and neuroimaging results suggest that the human brain's representation of visual event timing is derived from visual responses rather than a centralised (internal) clock³⁻¹¹. Visual responses are strongly modulated by event timing^{12,13}, but it remains unclear how we quantify the timing of sensory events from the neural responses in sensory cortices.

Early visual cortical responses increase monotonically but sub-linearly with event frequency and duration, which can be described in terms of the summed amplitude of transient and sustained neural responses respectively for a stimulus of a fixed strength^{12,13}. Therefore, established early visual neural response dynamics provide a signal from which event timing can be quantified. These monotonic responses are likely to be restricted to the retinotopic location of the stimulus as these responses are encoded in neural responsivity to the stimulus in retinotopically organized early visual areas.

Beyond early visual areas, an extensive network throughout the human association cortices shows visual timing-tuned responses, with maximum response amplitudes at specific preferred durations (the time from event onset to offset) and periods (the time between repeating event onsets; i.e., $1/\text{frequency}$)¹⁴. These responses are topographically mapped, such that the preferred durations and periods of neural populations gradually progresses across the cortical surface^{14,15}. Such timing-tuned responses and maps are found in areas that also show tuning and mapping for other quantities, including visual numerosity, visual object size and haptic numerosity¹⁶⁻¹⁸. These areas have a more abstracted representation of multiple quantities that likely allows their neural responses to interact regardless of the spatial,

temporal or sensory origins^{17,18}, potentially underlying perceptual interactions^{19,20}. As these quantity-tuned responses are topographically mapped by the state of stimulus quantity, rather than its position on the retina or skin, we expect them not to be restricted to the stimulus' position on the sensory organ: they are encoded in an abstracted, quantity-based frame of reference.

Regarding the computation of timing-tuned responses, it seems likely that they are derived from the dynamics of early sensory neural responses. However, it is unclear how and where tuned representations to visual event timing are computed and transformed from monotonic responses in early visual cortices. Computational models for other quantities suggest that excitatory and inhibitory non-linear monotonic responses can be compared to give numerosity-tuned responses^{21,22}. Indeed, monotonic responses of early visual areas that closely follow numerosity are transformed into tuned responses to numerosity in lateral occipital areas²³. A range of motor event timing-dependent neural response profiles, from timing-tuned to monotonic, has been described in macaque premotor cortex^{24,25}. These tuned and monotonic neurons are intermixed and together transform from abstract, timing-tuned response to action-triggering thresholds to determine motor action timing.

Here we ask how monotonic and tuned neural responses to visual event timing are related throughout the brain's hierarchy of both timing maps and visual field maps: whether and where monotonic responses are transformed into tuned responses. We further ask how the responses to timing and visual field position are related. We do this by comparing the fits of monotonically increasing and tuned neural response models on ultra-high-field (7T) fMRI data that was acquired during the presentation of repetitive visual events which gradually varied in event duration and/or period.

Results

Timing responses transform from monotonic to tuned along the visual field map hierarchy

For each voxel, we used both a monotonic and a tuned response model (Fig. 1 and Supplementary Fig. 1) to predict the time course of response amplitudes to the changes in event timing. Each of these response models uses a set of candidate response model parameters to predict the neural response amplitudes to every combination of event duration and frequency (Fig. 1a, b). At every event's offset, we evaluate this prediction (Fig. 1c, d) to generate a prediction of the neural response time course that would be seen for this set of candidate response model parameters (Fig. 1e). Convolution with a hemodynamic response function (Fig. 1f) predicts an fMRI response time course (Fig. 1g), which we correlate to the measured responses (Fig. 1h). We repeat this for a large set of candidate response model parameters to find those parameters that most closely predict each voxel's response time course. We then compared the goodness of fit of the resulting monotonic and tuned response models on cross-validated data to distinguish between these models' performance despite differences in model complexity.

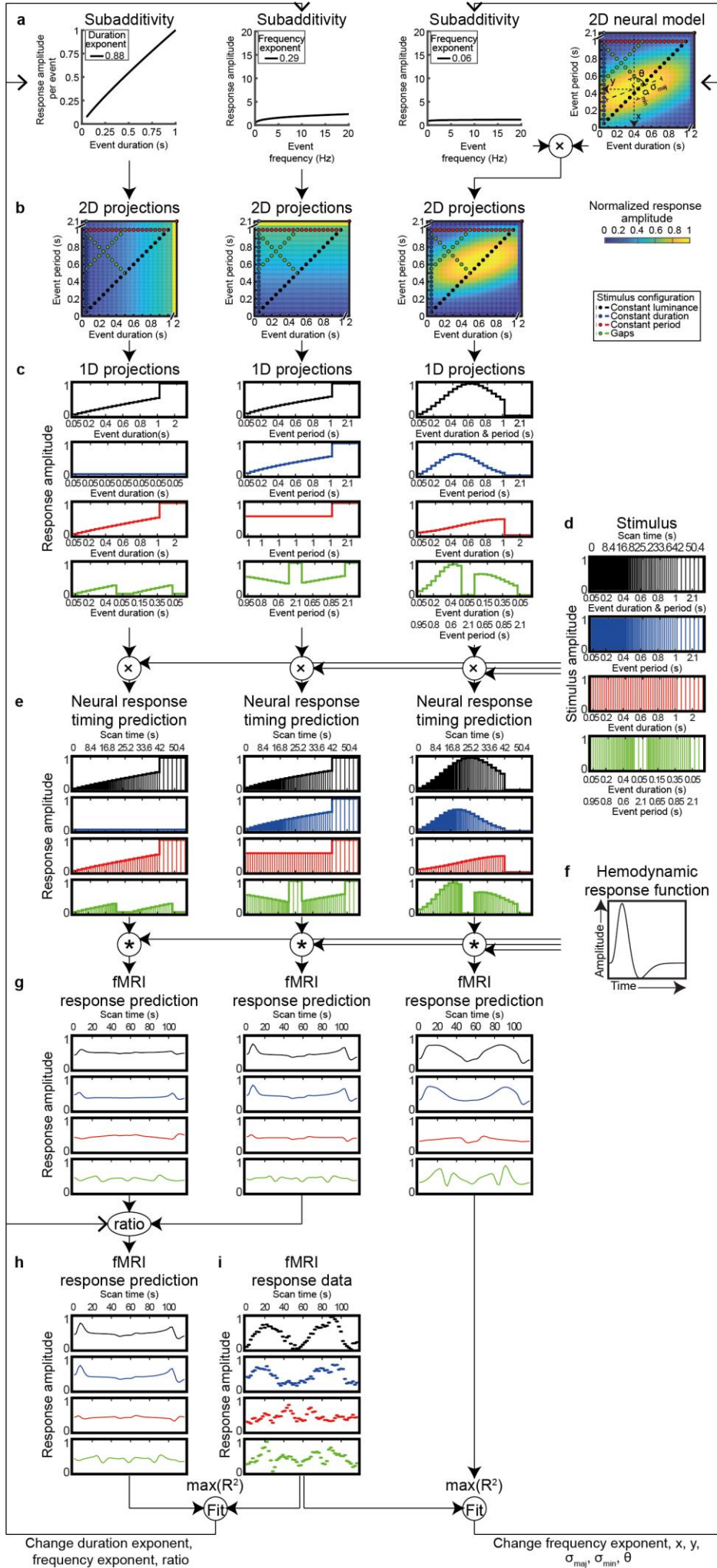


Fig. 1 Monotonic and tuned response model fitting procedures. **a** Monotonic response model for duration (left) and frequency (middle) response components, and tuned response model (right). **b** In each case, a compressive exponent parameter captures a non-linear scaling with increases in event duration (left) or frequency (middle and right). Together, these make a prediction of the per-event response amplitude for every event timing shown in the stimulus (dots). **c** This gives a prediction of the response amplitudes that would be seen for each stimulus condition for a specific candidate set of response model parameters. **d** The times of event offsets in each stimulus condition, which vary in frequency. **e** Combining the per-event response amplitudes predicted by the response models (in c) and the times of the event offsets (in d) gives neural response amplitude predictions for each condition, equal to the amount of color under the curve. **f** The hemodynamic response function. **g** The neural response amplitude predictions (in e) are convolved with a hemodynamic response function (in f) to get the predicted fMRI response time courses. Note here that these predictions are for both ascending and descending sweeps of duration and/or period, while neural response predictions (in c and e) are for ascending sweeps only. **h** The monotonic response predictions for event duration and frequency changes are combined as a weighted sum. **i** The recorded fMRI response time course for an example voxel. The monotonic and tuned predictions were compared to the recording from each voxel. The free parameters of both the monotonic and tuned response models were found that maximize the correlation (R^2) between predicted and measured fMRI response time courses. For the monotonic response model these parameters were the compressive exponents on duration and frequency components, and the weighting of these two components (ratio). For the tuned response model, the free parameters were the compressive exponent on event frequency, preferred duration (x), preferred period (y), response function extent along its major (σ_{maj}) and minor (σ_{min}) axes, and major axis orientation (θ).

We excluded voxels where the cross-validated variance explained (R^2) was 0.2 or less for both models. We found a clear modulation of responses by visual event timing throughout the visual field map hierarchy, with the monotonic response model best capturing responses of occipital areas and the tuned response model best capturing most of the responses in the parietal and frontal areas (Fig. 2a).

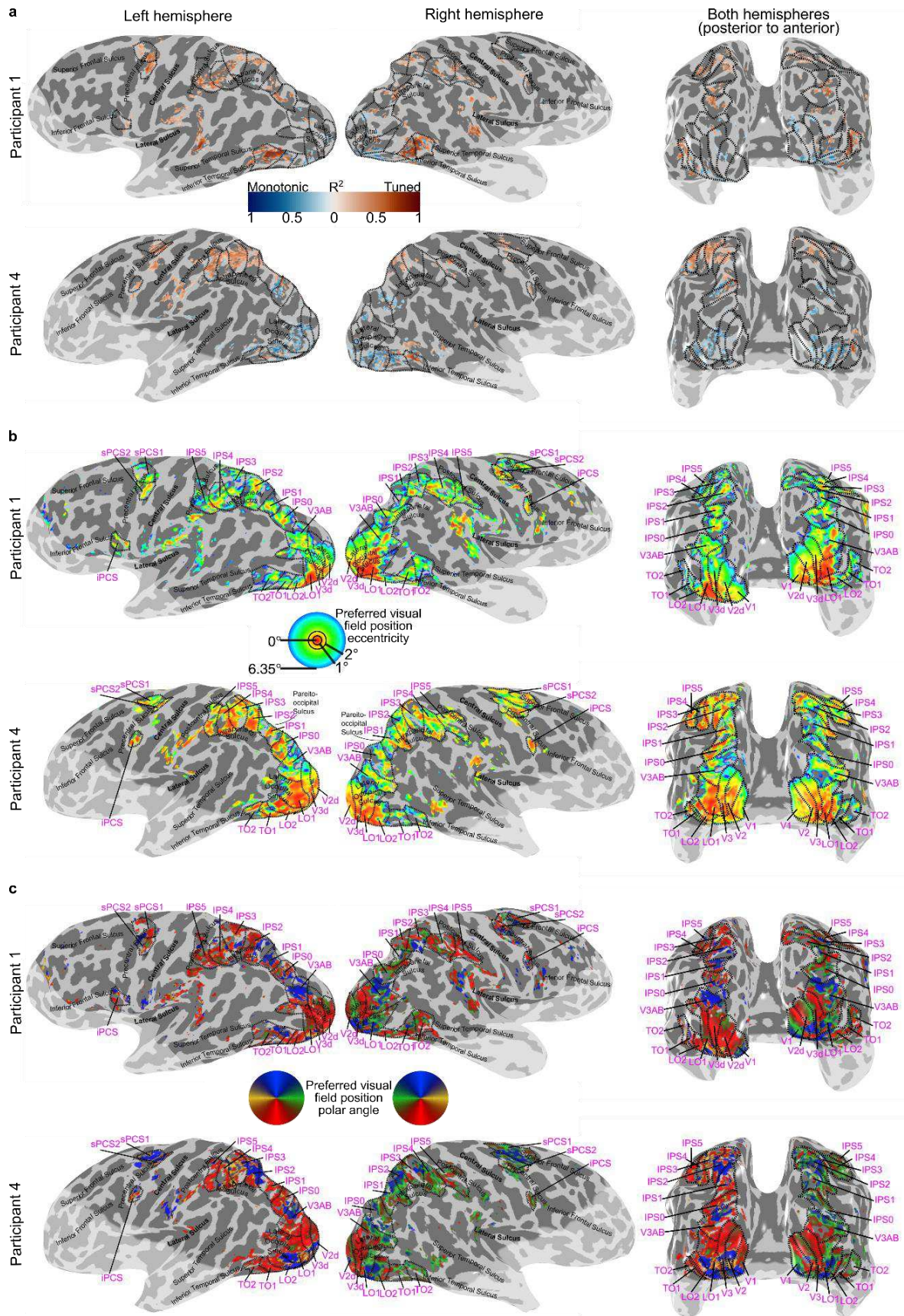


Fig. 2 Progressions of best-performing model on timing-selective responses and visual field position preferences. a Cross-validated variance explained of the best-performing

model within each voxel is displayed (averaged across both cross-validation splits, for voxels with a variance explained above 0.2 for either model). The variance explained of the monotonic response model is given in blue and of the tuned response model is given in red. The intensity of the color relates to the magnitude of the variance explained. **b** Eccentricity preferences for voxels with over 0.1 variance explained by the response model. **c** Polar angle preferences. Visual field map borders are shown as black dashed lines, and named in magenta text. The light shaded region is outside the fMRI recording volume. See also Supplementary Figs. 2-4.

We used standard visual field mapping and population receptive field (pRF) modelling procedures^{26,27} to determine the preferred visual field positions of every voxel. We grouped these voxels into visual field maps (Fig. 2 and Supplementary Figs. 2 and 3), then averaged the fits in each visual field map and used a three-factor ANOVA to assess how model fits differed between visual field maps, models and participants. Model fits (R^2) differed between visual field maps ($F_{(16, 1021)} = 27.95, p < 10^{-10}, \eta_p^2 = 0.31$) and models ($F_{(1, 1036)} = 4.77, p = 0.029, \eta_p^2 = 0.005$), and there was an interaction between visual field maps and model ($F_{(16, 1021)} = 59.26, p < 10^{-10}, \eta_p^2 = 0.49$). Post-hoc multiple comparisons demonstrated that early visual and lateral visual field maps had significantly better fits for the monotonic response model (V1, V2, V3, LO1, LO2, TO1, and V3AB; Fig. 3). In contrast, the responses of several parietal and frontal visual field maps were better captured by the tuned response model than the monotonic response model (IPS2, IPS3, IPS4, IPS5, sPCS1, sPCS2, and iPCS; Fig. 3). Regions that lay in between the early visual (monotonic) and parietal/frontal (tuned) maps (TO2, IPS0 and IPS1), showed no significant difference between the fit of the monotonic and tuned response model (V1: mean difference (μ) = -0.17, standard error ($\sigma_{\bar{x}}$) = 0.01, $t_{(23)} = -14.28, p < 10^{-10}, d' = -2.91$, 95% confidence interval (CI) = [-0.20, -0.15]; V2: $\mu = -0.17, \sigma_{\bar{x}} = 0.01, t_{(28)} = -11.83, p < 10^{-10}, d' = -2.20$, CI = [-0.20, -0.15]; V3: $\mu = -0.12, \sigma_{\bar{x}} = 0.01, t_{(31)} = -9.94, p = 10^{-10}, d' = -1.76$, CI = [-0.15, -0.10]; LO1: $\mu = -0.17, \sigma_{\bar{x}} = 0.01, t_{(31)} = -12.24, p < 10^{-10}, d' = -2.16$, CI = [-0.20, -0.14]; LO2: $\mu = -0.15, \sigma_{\bar{x}} = 0.01, t_{(29)} = -10.99, p < 10^{-10}, d' = -2.01$, CI = [-0.18, -0.13]; TO1: $\mu = -0.09, \sigma_{\bar{x}} = 0.02, t_{(31)} = -$

$5.18, p = 2 \times 10^{-5}, d' = -0.91, CI = [-0.12, -0.05]$; TO2: $\mu = -0.01, \sigma_{\bar{x}} = 0.02, t_{(31)} = -0.48, p = 0.635, d' = -0.08, CI = [-0.04, 0.03]$; V3AB: $\mu = -0.06, \sigma_{\bar{x}} = 0.02, t_{(31)} = -3.25, p = 0.004, d' = -0.57, CI = [-0.10, -0.02]$; IPS0: $\mu = -0.02, \sigma_{\bar{x}} = 0.02, t_{(31)} = -0.94, p = 0.376, d' = -0.17, CI = [-0.05, 0.02]$; IPS1: $\mu = 0.06, \sigma_{\bar{x}} = 0.03, t_{(31)} = 2.09, p = 0.051, d' = 0.37, CI = [0.00, 0.11]$; IPS2: $\mu = 0.17, \sigma_{\bar{x}} = 0.02, t_{(31)} = 8.23, p = 5 \times 10^{-9}, d' = 1.46, CI = [0.13, 0.21]$; IPS3: $\mu = 0.19, \sigma_{\bar{x}} = 0.02, t_{(31)} = 9.71, p = 10^{-10}, d' = 1.72, CI = [0.15, 0.23]$; IPS4: $\mu = 0.11, \sigma_{\bar{x}} = 0.02, t_{(31)} = 5.38, p = 10^{-5}, d' = 0.95, CI = [0.07, 0.15]$; IPS5: $\mu = 0.14, \sigma_{\bar{x}} = 0.02, t_{(28)} = 6.50, p = 8 \times 10^{-7}, d' = 1.21, CI = [0.10, 0.18]$; sPCS1: $\mu = 0.22, \sigma_{\bar{x}} = 0.02, t_{(30)} = 12.40, p < 10^{-10}, d' = 2.23, CI = [0.19, 0.26]$; sPCS2: $\mu = 0.19, \sigma_{\bar{x}} = 0.02, t_{(27)} = 10.41, p = 10^{-10}, d' = 1.97, CI = [0.15, 0.22]$; iPCS: $\mu = 0.06, \sigma_{\bar{x}} = 0.02, t_{(27)} = 3.01, p = 0.007, d' = 0.57, CI = [0.02, 0.11]$.

Strikingly, the amount of variance explained (i.e. the goodness of model fit) by both models increased along the visual field map hierarchy from V1 to TO1 (hMT+) (Fig. 3a), after which the tuned response model fit better. Tuned response model fits then remained relatively stable across frontal and parietal visual field maps while monotonic response model fits declined for more anterior visual field maps. As a result of these two changes, the difference in model fits (i.e. the additional variance explained by a timing-tuned response component) increased from occipital to frontal visual field maps (Fig. 3b).

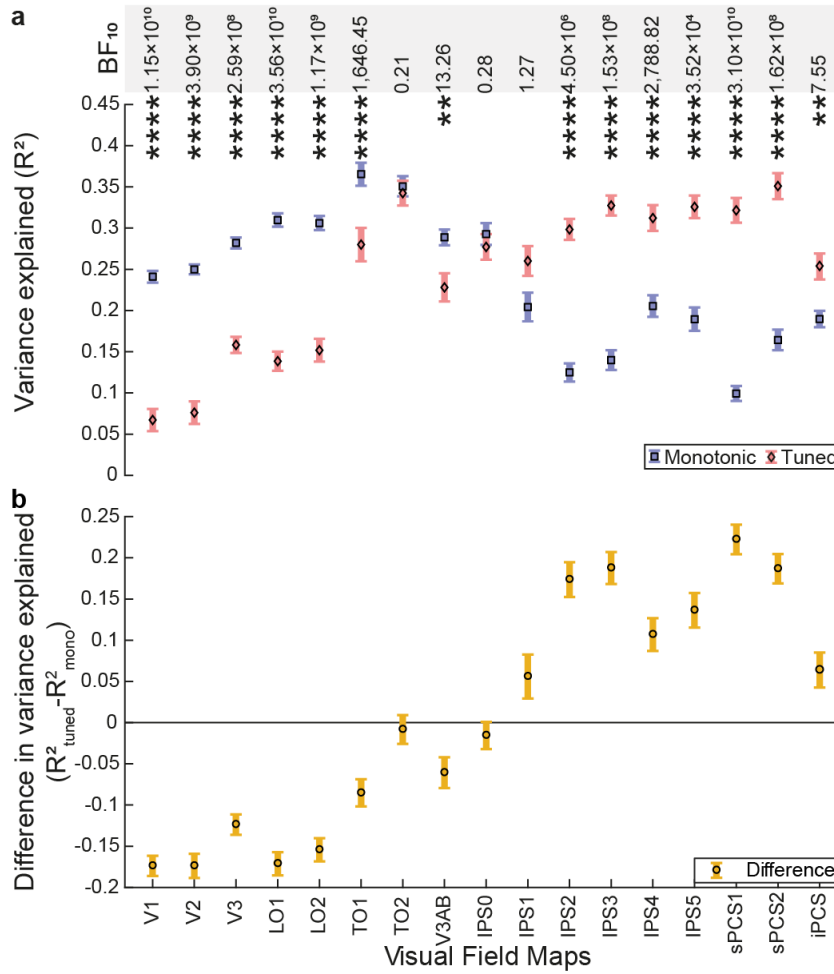


Fig. 3 Comparison of fits of monotonic and tuned response models in each visual field map. a Cross-validated variance explained by monotonic and tuned response models. **b** Difference in cross-validated variance explained between models (tuned – monotonic) progressively changes from monotonic to tuned up the visual hierarchy. Markers show the mean variance explained of each visual field map, error bars show the standard error. Only the voxels that had a variance explained above 0.2 for either model are included. Significant differences between the models (in a paired t-test, FDR corrected for multiple comparisons): * $p < .05$ ** $p < .01$, *** $p < .001$ and **** $p < .0001$. Bayes factors for each paired comparison are presented in the top row.

Relationship between model fits and retinotopic location

Since the monotonic response model performed well in early visual fields maps, it is likely that monotonic responses are computed from low-level stimulus properties. This suggests they would be limited to the retinotopic location of the stimulus (the central visual field representation), rather than elsewhere (the peripheral visual field representation). We therefore investigated whether the fits of the monotonic and tuned response models differed

between central and peripheral representations within each visual field map. We excluded voxels with a cross-validated variance explained (R^2) of 0 in both models.

First, we visualized the progressions of variance explained of each response model throughout the visual field by plotting the variance explained against eccentricity (Fig. 4 and Supplementary Fig. 5). It is clear that each model's fit decreased as the voxel's preferred visual position moved away from the stimulus area. This decrease was steep and sudden in early visual and lateral occipital visual field maps, which have smaller spatial receptive fields and where responses to event timing are best captured by monotonic response models. A decrease in model fits with eccentricity was also apparent in higher visual field maps, where the tuned response model begins to fit better, but this decrease was more gradual.

The gradual progression from monotonic to tuned response model fits along the visual field map hierarchy suggests tuned responses are derived from monotonic inputs. Furthermore, it is important to remember that both response models describe the responses per event, so even in a monotonic response model (where total response amplitude increases with event frequency) the response per event actually decreases with frequency. As a result of this strong relationship between frequency and response amplitude in both models, much of the fMRI response of a timing-tuned neural population can be captured by a monotonic response model. Furthermore, as timing-tuned responses are found overlapping with visual field maps it may be that a spatial representation responding monotonically to temporal contrast is intermixed with a temporal representation with timing-tuned responses. We therefore also assess how much extra response variance the tuned response model can capture beyond the prediction of a monotonic response model: the difference between the tuned and monotonic response model fits.

This difference between the monotonic and tuned response models' fits decreased with eccentricity in early visual and lateral occipital visual field maps where responses to

event timing are best captured the monotonic response model. After TO1 the difference in goodness-of-fit does not show a consistent relationship to the voxel's preferred visual position eccentricity.

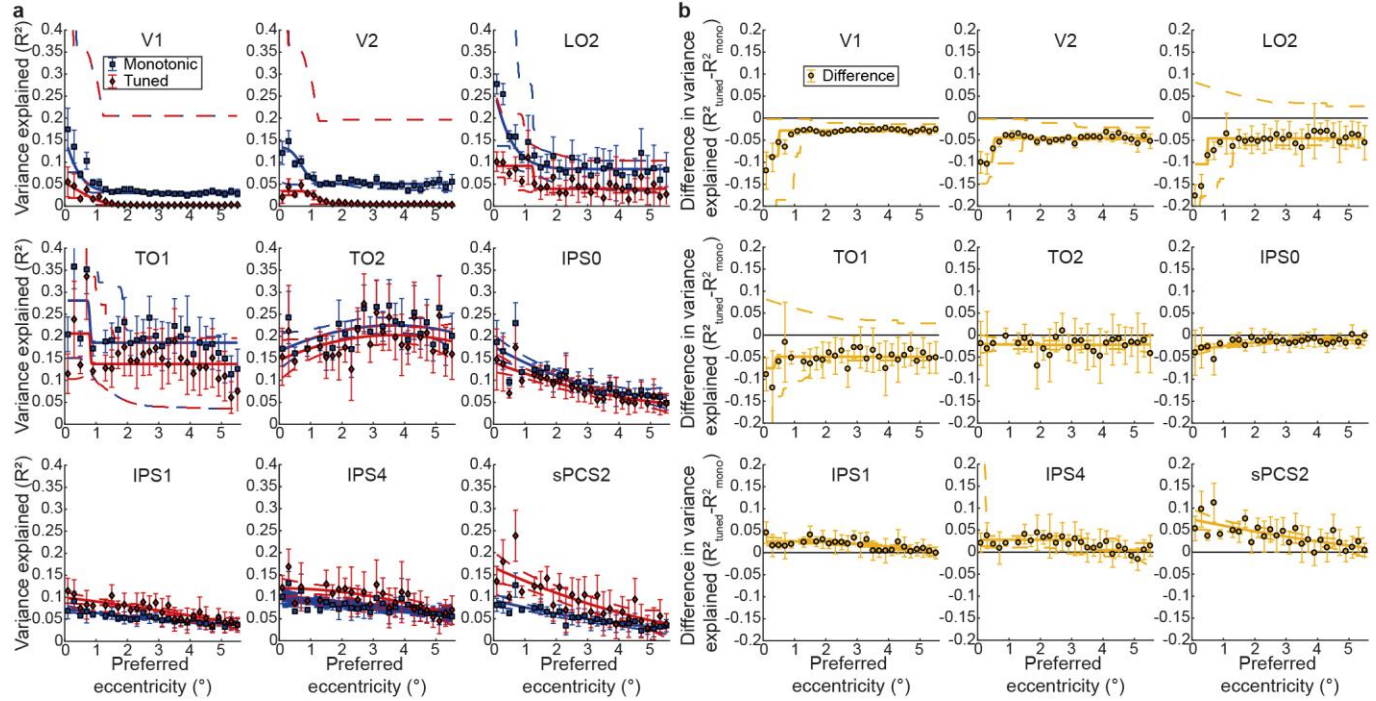


Fig. 4 Progression of timing response model fits with preferred visual field eccentricity in a representative selection of visual field maps. **a** Early and lateral occipital visual field maps show a sharp decrease of model fits moving away from the retinotopic representation of the stimulus position. This decrease then becomes more gradual where tuned response model fits begin to improve. **b** The difference between the response model fits (tuned - monotonic) also decreases with eccentricity in the early and lateral occipital visual field maps, but shows no consistent relationship with eccentricity after TO1. Markers show mean variance explained per eccentricity bin, error bars show the standard error of the mean. Solid lines show the best fit to changes with eccentricity, dashed lines are their 95% confidence intervals. Note that the data for these plots are not thresholded to a variance explained above 0.2, but above 0. See also Supplementary Fig. 5.

Second, to assess the significance of the effect of eccentricity on timing response model fits within each map, we compared the variance explained in voxels near ($<1^\circ$) and far ($>2^\circ$) from the center of the visual field using 3-way ANOVAs for monotonic response model fits, tuned response model fits, and their difference (factors: visual field map, eccentricity range and participant; interaction: visual field map and eccentricity range).

For the monotonic response model there was a main effect of eccentricity range ($F_{(1, 1086)} = 97.97, p < 10^{-10}, \eta_p^2 = 0.09$) and map ($F_{(16, 1071)} = 59.07, p < 10^{-10}, \eta_p^2 = 0.47$), and an interaction between eccentricity range and visual field map ($F_{(16, 1071)} = 5.81, p < 10^{-10}, \eta_p^2 = 0.08$). This interaction demonstrated that the difference between eccentricity ranges differed between maps. Not all maps showed a difference between eccentricity ranges, but post-hoc multiple comparisons demonstrated a higher variance explained near the central visual field representation than in the periphery in several maps (V1, V2, V3, LO1, LO2, V3AB, IPS0, IPS2, IPS3, sPCS2, iPCS; Fig. 5a) (V1: median difference between near and far (M) = 0.01, $Z = 2.69, p = 0.012, r = 0.08$, 95% confidence interval of the median difference computed from 1000 bootstrap iterations (CI) = [0.00, 0.03]; V2: $M = 0.02, Z = 3.44, p = 0.001, r = 0.11$, CI = [0.01, 0.05]; V3: $M = 0.06, Z = 4.51, p = 6 \times 10^{-5}, r = 0.14$, CI = [0.04, 0.07]; LO1: $M = 0.10, Z = 4.94, p = 10^{-5}, r = 0.15$, CI = [0.08, 0.13]; LO2: $M = 0.05, Z = 4.24, p = 0.0001, r = 0.13$, CI = [0.03, 0.06]; TO1: $M = 0.03, Z = 1.98, p = 0.062, r = 0.06$, CI = [-0.02, 0.07]; TO2: $M = -0.03, Z = -2.09, p = 0.051, r = -0.07$, CI = [-0.07, 0.02]; V3AB: $M = 0.05, Z = 3.63, p = 0.0007, r = 0.11$, CI = [0.01, 0.06]; IPS0: $M = 0.03, Z = 2.94, p = 0.006, r = 0.09$, CI = [0.01, 0.04]; IPS1: $M = 0.01, Z = 1.83, p = 0.071, r = 0.06$, CI = [-0.00, 0.02]; IPS2: $M = 0.01, Z = 2.58, p = 0.015, r = 0.08$, CI = [-0.00, 0.02]; IPS3: $M = 0.02, Z = 3.66, p = 0.0007, r = 0.11$, CI = [0.00, 0.04]; IPS4: $M = 0.02, Z = 1.83, p = 0.071, r = 0.06$, CI = [0.00, 0.03]; IPS5: $M = 0.00, Z = 1.87, p = 0.071, r = 0.06$, CI = [0.00, 0.01]; sPCS1: $M = 0.01, Z = 1.25, p = 0.210, r = 0.04$, CI = [-0.00, 0.01]; sPCS2: $M = 0.02, Z = 4.13, p = 0.0002, r = 0.13$, CI = [0.01, 0.03]; iPCS: $M = 0.03, Z = 3.95, p = 0.0003, r = 0.12$, CI = [0.02, 0.04]; all $n = 32$ pairs).

For the tuned response model, there were main effects of eccentricity range ($F_{(1, 1086)} = 31.33, p = 3 \times 10^{-8}, \eta_p^2 = 0.03$) and map ($F_{(16, 1071)} = 34.31, p < 10^{-10}, \eta_p^2 = 0.34$) but the interaction between eccentricity range and map did not reach significance ($F_{(16, 1071)} = 1.62, p$

= 0.058, $\eta_p^2 = 0.02$). Again, post-hoc multiple comparisons demonstrated a higher variance explained near the central visual field representation than in the periphery in several maps (V1, V2, V3, LO1, LO2, V3AB, IPS3, IPS5, sPCS2, iPCS; Fig. 5b) (V1: $M = 0.00$, $Z = 4.34$, $p = 6 \times 10^{-5}$, $r = 0.14$, $CI = [0.00, 0.01]$; V2: $M = 0.01$, $Z = 4.73$, $p = 10^{-5}$, $r = 0.15$, $CI = [0.01, 0.03]$; V3: $M = 0.03$, $Z = 4.92$, $p = 7 \times 10^{-6}$, $r = 0.15$, $CI = [0.02, 0.06]$; LO1: $M = 0.04$, $Z = 4.94$, $p = 7 \times 10^{-6}$, $r = 0.15$, $CI = [0.03, 0.05]$; LO2: $M = 0.01$, $Z = 2.99$, $p = 0.007$, $r = 0.09$, $CI = [0.00, 0.02]$; TO1: $M = 0.03$, $Z = 0.88$, $p = 0.379$, $r = 0.03$, $CI = [-0.02, 0.07]$; TO2: $M = -0.05$, $Z = -1.93$, $p = 0.084$, $r = -0.06$, $CI = [-0.08, -0.00]$; V3AB: $M = 0.03$, $Z = 3.33$, $p = 0.003$, $r = 0.10$, $CI = [0.01, 0.04]$; IPS0: $M = 0.01$, $Z = 1.78$, $p = 0.107$, $r = 0.06$, $CI = [-0.00, 0.02]$; IPS1: $M = 0.00$, $Z = 1.66$, $p = 0.126$, $r = 0.05$, $CI = [-0.00, 0.03]$; IPS2: $M = 0.00$, $Z = 1.35$, $p = 0.196$, $r = 0.04$, $CI = [-0.02, 0.04]$; IPS3: $M = 0.03$, $Z = 3.03$, $p = 0.007$, $r = 0.09$, $CI = [0.02, 0.06]$; IPS4: $M = 0.02$, $Z = 1.55$, $p = 0.147$, $r = 0.05$, $CI = [0.01, 0.04]$; IPS5: $M = 0.02$, $Z = 2.84$, $p = 0.008$, $r = 0.09$, $CI = [0.01, 0.03]$; sPCS1: $M = 0.01$, $Z = 1.33$, $p = 0.196$, $r = 0.04$, $CI = [-0.01, 0.02]$; sPCS2: $M = 0.02$, $Z = 2.90$, $p = 0.008$, $r = 0.09$, $CI = [0.00, 0.04]$; iPCS: $M = 0.02$, $Z = 2.73$, $p = 0.011$, $r = 0.09$, $CI = [0.00, 0.05]$; all $n = 32$ pairs).

Finally, the difference in variance explained between the models revealed a main effect of eccentricity range ($F_{(1,1086)} = 10.57$, $p = 0.001$, $\eta_p^2 = 0.01$) and map ($F_{(16, 1071)} = 54.85$, $p < 10^{-10}$, $\eta_p^2 = 0.46$), and an interaction between the eccentricity range and map ($F_{(16, 1071)} = 2.90$, $p = 0.0001$, $\eta_p^2 = 0.04$). Here, post-hoc multiple comparisons demonstrated that the difference in variance explained between the models was larger near the center of the visual field than in the periphery only in V3, LO1, LO2 and IPS0 (Fig. 5c). Notably, these are areas in which the monotonic response model outperformed the tuned response model. Areas in which the tuned response model outperformed the monotonic response model showed no significant difference between eccentricity ranges. The Bayes factors of the paired comparisons in the latter areas are below one, indicating that such a difference between

eccentricity ranges is indeed absent. This suggests that the additional response component explained by tuning was location-independent. (V1: $M = -0.00$, $Z = -1.72$, $p = 0.181$, $r = -0.05$, $CI = [-0.02, 0.00]$; V2: $M = -0.01$, $Z = -0.93$, $p = 0.541$, $r = -0.03$, $CI = [-0.02, 0.00]$; V3: $M = -0.02$, $Z = -2.77$, $p = 0.024$, $r = -0.09$, $CI = [-0.04, -0.01]$; LO1: $M = -0.05$, $Z = -4.88$, $p = 2 \times 10^{-5}$, $r = -0.15$, $CI = [-0.08, -0.03]$; LO2: $M = -0.04$, $Z = -3.85$, $p = 0.0010$, $r = -0.12$, $CI = [-0.05, -0.02]$; TO1: $M = -0.01$, $Z = -1.93$, $p = 0.153$, $r = -0.06$, $CI = [-0.04, 0.00]$; TO2: $M = -0.01$, $Z = -0.37$, $p = 0.753$, $r = -0.01$, $CI = [-0.02, 0.01]$; V3AB: $M = -0.01$, $Z = -2.04$, $p = 0.141$, $r = -0.06$, $CI = [-0.02, -0.00]$; IPS0: $M = -0.01$, $Z = -2.90$, $p = 0.021$, $r = -0.09$, $CI = [-0.02, -0.00]$; IPS1: $M = -0.00$, $Z = -0.37$, $p = 0.753$, $r = -0.01$, $CI = [-0.01, 0.01]$; IPS2: $M = -0.01$, $Z = -0.75$, $p = 0.594$, $r = -0.02$, $CI = [-0.02, 0.00]$; IPS3: $M = 0.02$, $Z = 1.72$, $p = 0.181$, $r = 0.05$, $CI = [0.01, 0.03]$; IPS4: $M = 0.01$, $Z = 0.97$, $p = 0.541$, $r = 0.03$, $CI = [-0.01, 0.02]$; IPS5: $M = 0.01$, $Z = 1.53$, $p = 0.236$, $r = 0.05$, $CI = [-0.00, 0.02]$; sPCS1: $M = 0.00$, $Z = -0.13$, $p = 0.896$, $r = -0.00$, $CI = [-0.01, 0.01]$; sPCS2: $M = 0.00$, $Z = 0.80$, $p = 0.594$, $r = 0.03$, $CI = [-0.01, 0.02]$; iPCS: $M = -0.00$, $Z = -0.45$, $p = 0.753$, $r = -0.01$, $CI = [-0.02, 0.01]$; all $n = 32$ pairs).

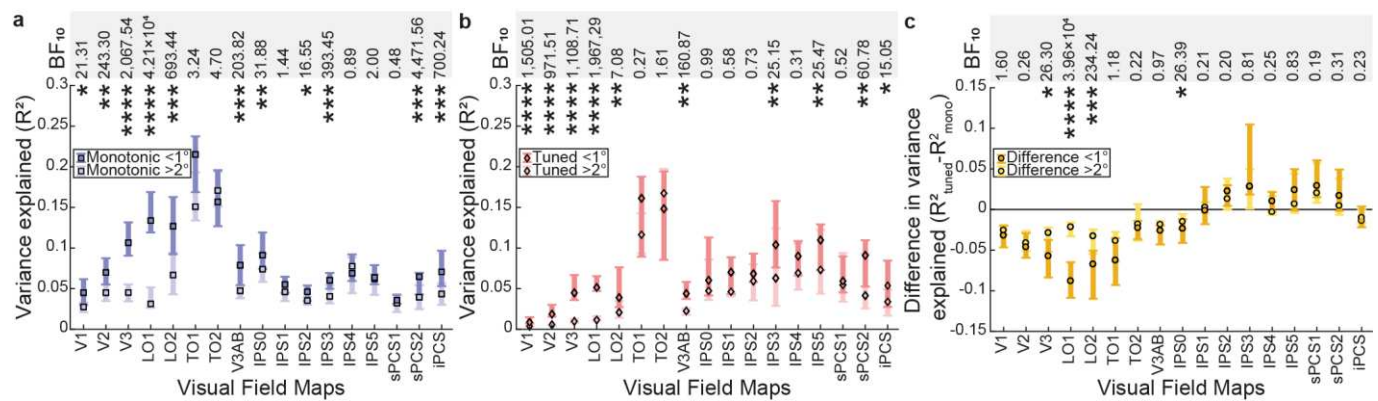


Fig. 5 Model fits decrease with distance from the stimulus area, but the additional response variance explained by the timing-tuned model is only affected in visual field maps where the monotonic response model fits better. a Cross-validated variance explained by the monotonic response model in eccentricity ranges near and far from the stimulus area. **b** Cross-validated variance explained by the tuned response model in eccentricity ranges near and far from the stimulus area. In both cases, responses are clearer

near the stimulus area in many visual field maps. **c** The difference in cross-validated variance explained between the two models was significant only in V3, LO1, LO2 and IPS0, where the monotonic response model fit better than the tuned response model. Markers show the median variance explained in each visual field map, error bars show the 95% confidence interval, computed from 1000 bootstrap iterations. Only voxels that had a variance explained above 0 for either model are included. Significant differences between the near and far eccentricity ranges (Wilcoxon signed-rank test, FDR corrected for multiple comparisons): * $p < .05$ ** $p < .01$, *** $p < .001$ and **** $p < .0001$. Bayes factors for each paired comparison are presented in the top row.

Timing response transforms from monotonic to tuned along the timing map hierarchy

In previous work the locations of timing-tuned responses were defined as timing maps, as the tuned response model performed best over the brain as a whole¹⁴. Such timing maps overlap with visual field maps, but do not have the same borders so do not include the same set of voxels. So far in this study, we have used visual field map borders, which may include both a spatial representation responding monotonically to temporal contrast and a temporal representation with timing-tuned responses. Therefore, to focus on the temporal representation more specifically, we compared the models specifically within the timing map borders. We excluded voxels where the cross-validated variance explained (R^2) was 0.2 or less for both models. We used a three-factor ANOVA to assess how model fits differed between timing maps, models and participants. Model fits differed between timing maps ($F_{(9, 596)} = 24.65, p < 10^{-10}, \eta_p^2 = 0.28$) and models ($F_{(1, 604)} = 145.37, p < 10^{-10}, \eta_p^2 = 0.20$), and there was an interaction between maps and model ($F_{(9, 596)} = 35.63, p < 10^{-10}, \eta_p^2 = 0.36$). Post-hoc multiple comparisons demonstrated that responses of anterior maps (TLS, TPCI, TPCM, TPCS, TFI and TFS) were better fit by the tuned response model (Fig. 6). However, posterior timing maps (TLO and TTOP) were better fit by the monotonic response model, which was not evident in previous analyses that did not compare models in each map separately. Responses in timing maps TTOA and TPO (which largely overlap with TO2, IPS0 and IPS1) showed no significant difference between the fits of the monotonic and tuned response models. (TLO: $\mu = -0.11, \sigma_{\bar{x}} = 0.01, t_{(31)} = -7.29, p = 8 \times 10^{-8}, d' = -1.29, 95\%$

confidence interval (CI) = [-0.13, -0.08]; TTOP: $\mu = -0.06$, $\sigma_{\bar{x}} = 0.02$, $t_{(31)} = -3.25$, $p = 0.004$, $d' = -0.57$, CI = [-0.09, -0.02]; TTOA: $\mu = 0.00$, $\sigma_{\bar{x}} = 0.02$, $t_{(31)} = 0.06$, $p = 0.952$, $d' = 0.01$, CI = [-0.03, 0.03]; TPO: $\mu = 0.02$, $\sigma_{\bar{x}} = 0.02$, $t_{(31)} = 1.05$, $p = 0.337$, $d' = 0.19$, CI = [-0.02, 0.06]; TLS: $\mu = 0.07$, $\sigma_{\bar{x}} = 0.02$, $t_{(29)} = 3.11$, $p = 0.005$, $d' = 0.57$, CI = [0.02, 0.11]; TPCI: $\mu = 0.10$, $\sigma_{\bar{x}} = 0.02$, $t_{(25)} = 4.41$, $p = 0.0003$, $d' = 0.87$, CI = [0.06, 0.15]; TPCM: $\mu = 0.14$, $\sigma_{\bar{x}} = 0.02$, $t_{(29)} = 7.47$, $p = 8 \times 10^{-8}$, $d' = 1.36$, CI = [0.10, 0.17]; TPCS: $\mu = 0.21$, $\sigma_{\bar{x}} = 0.01$, $t_{(31)} = 14.76$, $p < 10^{-10}$, $d' = 2.61$, CI = [0.18, 0.23]; TFI: $\mu = 0.10$, $\sigma_{\bar{x}} = 0.02$, $t_{(25)} = 5.08$, $p = 6 \times 10^{-5}$, $d' = 1.00$, CI = [0.06, 0.14]; TFS: $\mu = 0.20$, $\sigma_{\bar{x}} = 0.02$, $t_{(30)} = 13.05$, $p < 10^{-10}$, $d' = 2.34$, CI = [0.17, 0.23]).

We do not compare model fits between eccentricity ranges within timing maps because the voxel selection within the timing maps is not chosen to cover all visual field position eccentricities.

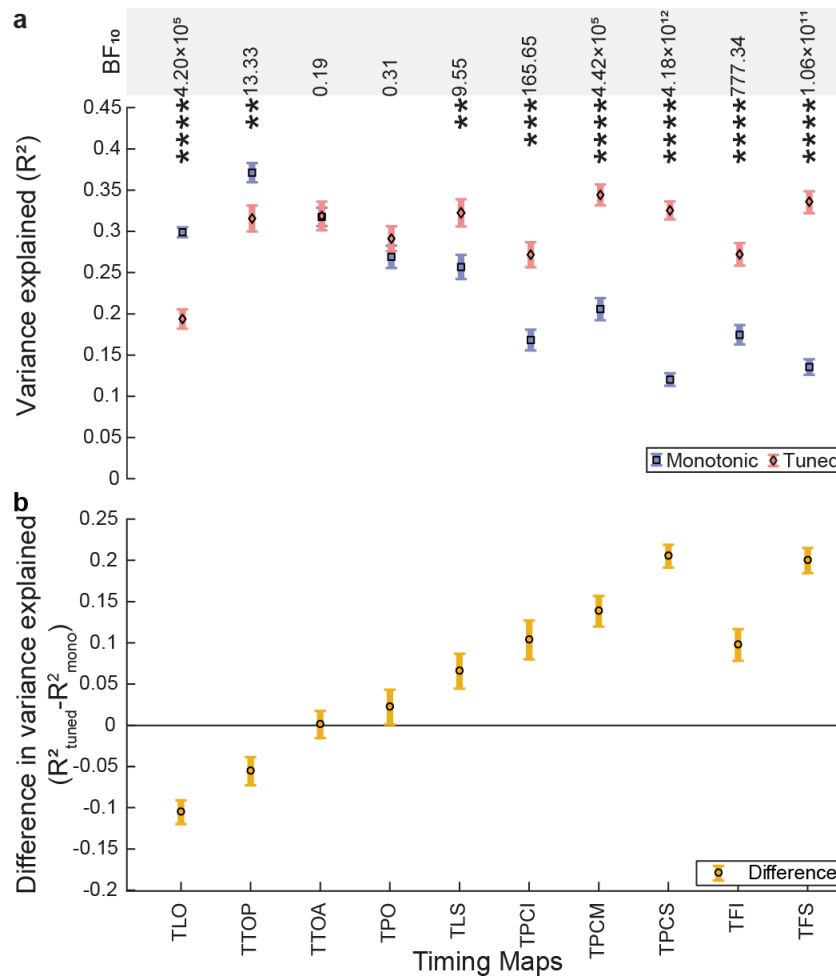


Fig. 6 Comparison of fits of monotonic and tuned response models in each timing map.
a Cross-validated variance explained by the monotonic and tuned response models. **b** Difference in cross-validated variance explained between models (tuned – monotonic) progressively changes from monotonic to tuned up the timing map hierarchy. Markers show the mean variance explained of each timing map, error bars show the standard error. Only the voxels that had a variance explained above 0.2 for either model are included. Significant differences between the models (in a paired t-test, FDR corrected for multiple comparisons): * $p < .05$ ** $p < .01$, *** $p < .001$ and **** $p < .0001$. Bayes factors for each paired comparison are presented in the top row.

Discussion

Tuned responses to visual event timing occur in a hierarchical network of topographically organized maps throughout the human association cortices¹⁴. Responses of early visual field maps are also modulated by event timing, monotonically increasing with event duration and frequency^{12,13}. In the current study we asked how these two sets of responses are related throughout the brain's hierarchy of both timing maps and visual field maps: whether and where monotonic responses are transformed into tuned responses. We also assessed how the responses to timing and visual field position are related. Visual field maps and timing maps largely overlap but have different borders and so include different neural populations which may have slightly different properties. We found increasingly clear monotonic responses to visual event duration and frequency from primary visual cortex to lateral occipital and temporal-occipital visual field maps (LO1, LO2, and TO1). After this, we found a gradual transition from monotonic responses to tuned responses from posterior to anterior brain areas, both when separated into visual field maps and timing maps. Both kinds of responses typically decreased when moving away from the retinotopic location of the stimulus. However, the difference between the model fits in areas where the tuned model outperformed the monotonic response model (i.e. the additional variance explained by the tuned response model) did not consistently depend on retinotopic location of the neural population.

We found that monotonic responses to visual stimulus event timing occurred in early visual field maps (as is also reported by Stigliani and colleagues¹²). More specifically, the

amplitude of these monotonic responses accumulated sub-linearly with event duration and frequency (in line with Zhou and colleagues¹³). In contrast to these previous experiments, we characterised the time between events in terms of the frequency of event onsets, rather than interstimulus interval (the time from one single event's offset to another single event's onset). This describes our stimulus design more straightforwardly because we use a repeating periodic stimulus that includes a large variation in event durations (the time from event onset to offset) for any particular frequency.

The fits of the monotonic response model increased from the V1 to the TO visual field maps. This is in line with the relative increase of transient responses to temporal contrast (stronger for magnocellular inputs), compared to sustained responses to temporal contrast (stronger for parvocellular inputs), when moving up the visual field map hierarchy¹². Indeed, in macaques the inputs to neurons in area MT are mainly from the magnocellular stream (the TO visual field maps are functionally homologous to MT)²⁸⁻³⁰, which has benefits for an area specialized in motion processing. A greater proportion of magnocellular inputs increases the modulation of responses by stimulus timing and may likewise be beneficial for subsequently deriving representations of stimulus timing.

A transition from monotonic to tuned responses occurs in the TO2, IPS0 and IPS1 visual field maps or, similarly, the TTOA and TPO timing maps. Strikingly, visual field map TO2 and timing map TTOA overlap with the human MT+^{31,32}. This area appears to be required for visual event timing perception as transcranial magnetic stimulation here disrupts timing judgements⁵.

We had previously reported that the responses of the timing maps were better fit by tuned than monotonic functions¹⁴. In the current study, we find that the responses in the timing maps TLO and TTOP are better captured by monotonic rather than tuned response models. We had previously grouped all timing maps together in this comparison, and used a

monotonic response model that assumed a linear accumulation of response amplitude with event duration¹⁴. Based on the results from the current study, we are no longer convinced that TLO and TTOP are topographically organized by preferred event timing because their monotonic response functions have no timing preference.

Nevertheless, there may be timing-tuned neurons with specific timing preferences in TLO and TTOP. The transition from monotonic to tuned responses is remarkably gradual across the visual field map and timing map hierarchies. This suggests that tuned and monotonic timing-selective responses may not be mutually exclusive. Intermixed monotonic and motor timing-tuned neuronal populations have been found in macaque medial premotor cortex during a rhythm continuation task^{24,25}. fMRI's spatial resolution groups the responses of large neural populations, which prevents us from distinguishing a hierarchical change in the proportions of intermixed monotonic and tuned neurons from a hierarchical transformation of event timing representations of single neurons. Both of these possibilities have independent components reflecting tuning for event timing and the accumulation of neural response amplitude with event frequency. Regardless of whether these components are found in the responses of distinct neurons or contribute to the responses of the same neuron, we propose that the difference between the fits of the monotonic and tuned response models reflects neural tuning for event timing after accounting for the underlying frequency-dependence of sensory responses.

The goodness of fit of both the monotonic and tuned response models are greatest near the retinotopic location of the visual field center, where the stimulus was presented, in most visual field maps. However, in visual field maps where the tuned response model outperforms the monotonic response model, the extra response variance captured by the tuned response model is independent of retinotopic organization. Based on the location-invariance of this additional tuned response component (beyond the monotonic frequency-

dependent response component), the location-dependence of the tuned response model's fit may be rooted in the location-dependence of the monotonic, frequency-dependent response component.

Previous studies of responses to timing have not compared the temporal modulation of responses to visual position selectivity on an individual subject level, which is important given the fine spatial scale of changes in visual position selectivity in visual field maps. A recent meta-analysis found dorsal-ventral gradients of responsivity to spatial and temporal tasks in the intraparietal sulcus and the right frontal operculum and posterior-anterior gradients around the supplementary motor area³³. We expect this discrepancy in locations to be due to the difference in task demands, as the studies in the meta-analysis were only included if they used a task that was clearly linked to spatial or temporal processing. Conversely, the current study did not require the participants to make any timing-related judgements when measuring temporal modulation of neural responses, nor spatial judgements during visual field mapping. Still, we do find a posterior-anterior gradient, but across the entire brain, such that spatial responses to the temporal contrast of stimuli (i.e., monotonic responses) are found posterior, while timing-selective tuned responses are found more anterior.

The emergence of visual numerosity tuned responses shows striking similarities with the emergence of visual timing tuned responses we describe here. Both numerosity and event timing are quantities, but spatial and temporal quantities respectively. Both show tuned responses in higher-level association areas^{14,16}. While these responses largely overlap, there are clear differences in their location. This suggests that, although the different quantities are not processed using the exact same neural populations, there may be similar computational mechanisms for estimating different quantities in the brain. In both cases, early visual responses monotonically follow numerosity and timing, and tuned responses emerge later²³.

However, the transition from monotonic responses to numerosity to numerosity-tuned responses takes place around the lateral occipital areas, rather than the temporal occipital areas that we see for timing. This transition is very sudden in the case of numerosity responses, in contrast to the gradual transition we see in the emergence of timing-tuned responses. As a result, the numerosity-tuned responses are not partly captured by or intertwined with monotonic responses. Therefore, only monotonic numerosity response model fits decrease when moving away from the retinotopic location of the center of the visual field, while the numerosity-tuned response model fits are consistent throughout the visual field.

To begin estimating stimulus timing from visual inputs requires an analysis of neural responses to temporal changes in those inputs. Early retinotopic visual areas will only show such responses at the retinotopic location of the stimulus. However, a true sense of event timing should generalize across all stimulus locations, separating timing information from other stimulus properties. Here we find here that the human brain appears to transform simple monotonic responses to timing into tuned temporal representations, where the tuned component is independent of the event's location in the visual field. As such, similar to numerosity, the brain's responses to stimulus timing reflect a transformation and abstraction from low-level sensory information. Once this abstracted tuned representation of event timing has been derived, it is propagated into brain areas implicated in allocating attention, multisensory integration and planning actions, suggesting it benefits a wide variety of cognitive functions.

Methods

Participants

We collected data from eight participants (1 female, aged 25 to 35). All participants were right-handed and had normal or corrected to normal vision. Two participants were co-authors, familiar with the goals of the study. All participants were briefly trained in duration discrimination tasks before scanning, to encourage attention to stimulus timing and avoid learning or habituation effects at the start of data collection. All subjects gave written informed consent. All experimental procedures were cleared by the ethics committee of University Medical Center Utrecht.

Stimuli

Timing mapping stimuli

All visual stimuli were generated using MATLAB and Psychtoolbox-3³⁴. The visual stimuli were presented on a 27.0 x 9.5 cm screen inside the MRI bore (resolution 1600 x 538 pixels) at 41 cm from the participant's eyes.

Participants were asked to fixate at the center of a red fixation cross that crossed the whole display on a gray background. Visual events comprised the presentation of a filled circle with a diameter of 0.4° that appeared for a variable duration at a variable temporal period¹⁴. The position of the circle changed pseudo-randomly between stimulus events, but was always within 0.75° of this fixation cross and 0.25° or more away (edge to edge) from the previous position. Every 21 s on average, the presented circle was white rather than the usual black. Participants had to respond by pushing a button when they saw this white target stimulus. No timing judgements were ever required.

Each event timing was repeated in a 2100 ms time frame, such that an entire TR (=2100 ms) contained the same event duration and period. The number of events presented within the 2100 ms between timing changes varied with event period and increments in event period sometimes fell slightly before or after 2100 ms. The maximum drift of event onset

timing was only 300 ms, and the increments in event period were only 50 ms, so this deviation was not perceptible. The presented event timing was used for analysis.

To distinguish responses to specific event timings from responses to other stimulus parameters, we used four conditions where event duration and period were related in different ways (Supplementary Movie 1). First, the “constant luminance” condition, in which event duration and period were equal and both changed together (50-1000 ms, 50 ms steps) (Fig. 1b, d, black points). Second, the “constant duration” condition, in which event period changed (50-1000 ms, 50 ms steps), while event duration remained 50 ms (Fig. 1b, d, blue points). Third, the “constant period” condition, in which the event duration changed (50-1000 ms, 50 ms steps), while the period always remained 1000 ms (Fig. 1b, d, red points). In all the aforementioned conditions, increasing progressions of event duration and/or period were followed by a 16.8 s interval of events with a 2100 ms period. The event duration was 50 ms in the “constant duration” condition and 2000 ms in the other two conditions. These extreme timings help to distinguish very small response functions (which would respond briefly and with low amplitude in the 50-1000 ms range) and very large response functions (which respond continuously and with high amplitude) in fMRI responses^{26,35}. Furthermore, these long event durations and periods produce little response from neural populations preferring sub-second timing. Conversely, populations whose response monotonically increases with duration, period, or mean luminance should respond most strongly to these stimuli. After this, the same event timings were presented in a decreasing order, followed by another 16.8 s interval of events with a 50 or 2000 ms duration and a 2100 ms period. We used a single model to capture responses to both increasing and decreasing event duration and/or period progressions. Including responses to both increasing and decreasing timings in the same model counterbalanced adaptation effects with stimuli that give both higher and lower responses preceding presentation of any timing.

Finally, the “gaps” condition was designed to sample further combinations of event periods and durations (Fig. 1b, d, green points). This stimulus configuration consisted of four progressions with timing changing in 50 ms steps. In chronological order: increasing event durations (50-500 ms) and decreasing event periods (950-500 ms); increasing event durations (50-500 ms) and increasing event periods (550-1000 ms); decreasing event durations (500-50 ms) and increasing event periods (500-950 ms); and decreasing event durations (500-50 ms) and decreasing event periods (1000-550 ms). Each of these progressions was separated by 6.3 s of events with 50 ms duration and 2100 ms period, and the last was followed by 14.7s with that timing.

We tested each of 24 possible orders of all four stimulus configurations once per participant, so each participant’s data included 24 scanning runs, each totaling 470.4 seconds and acquired in four sessions.

Visual field mapping stimuli

We acquired visual field mapping responses to examine the relationship between our voxels’ responses to visual event timing and visual field position, and to delineate visual field maps.

The visual field mapping paradigm was identical to that described in previous studies^{16,17}.

The stimulus consisted of drifting bar apertures at various orientations, which exposed a moving checkerboard pattern. The stimulus had a radius of 6.35° , much larger than the timing mapping stimuli (0.75° radius). Two diagonal red lines, intersecting at the center of the display, were again presented throughout the entire scanning run. Participants fixated the center of the cross and pressed a button when these lines changed color, and detected 80-100% of the color changes that were presented within each scanning run.

fMRI data collection and pre-processing

Acquisition procedures were similar to procedures described elsewhere^{17,35}. Briefly, data was acquired with a 7T Philips Achieva scanner with a repetition time (TR) of 2100 ms, echo time (TE) of 25 ms, and a flip angle of 70°. The T1-weighted anatomical scans were automatically segmented with Freesurfer and manually edited to minimize segmentation errors. The T2*-weighted functional scans were acquired using a 32-channel head coil at a resolution of 1.77x1.77x1.75 mm, with 41 interleaved slices of 128x128 voxels. The resulting field of view was 227x227x72 mm. We used a single shot gradient echo sequence with SENSE acceleration factor 3.0 and anterior-posterior encoding, plus a top-up scan with the opposite phase-encoding direction to correct for image distortion in the gradient encoding direction³⁶. We also acquired a T1-weighted anatomical image with the same resolution, position and orientation as the functional data. We used a 3rd-order image-based B0 shim of the functional scan's field of view (in-house IDL software, v6.3, RSI, Boulder, CO, USA). The anterior temporal and frontal lobes were excluded from acquisition due to the fact that 7T fMRI has a low response amplitude and large spatial distortions in these areas. For timing mapping, we used 24 runs, 224 TRs (470.4 s) each, separated into 4 sessions with typically 6 runs each. For visual field mapping, we used 8-10 runs, 182 TRs (382.2 s) each, in a single separate session.

The functional data was co-registered to the anatomical space using AFNI (afni.nimh.nih.gov³⁷). A single transformation matrix was constructed, incorporating all the steps from the raw data to the cortical surface model to reduce the number of interpolation steps to one. A T1 image with the same resolution, position and orientation as the functional data was first used to determine the transformation to a higher resolution (1 mm isotropic) whole-brain T1 image (3dUnifize, 3dAllineate). For the fMRI data, we first applied motion correction to two series of images that were acquired using opposing phase-encoding directions (3dvolreg). Subsequently, we determined the distortion transformation between the

average images of these two series (3dQwarp). We then determined the transformation in brain position between and within functional scans (3dNwarpApply). Then we determined the transformation that co-registers this functional data to the T1 acquired in the same space (3dvolreg). We applied the product of all these transformations to every functional volume to transform our functional data to the whole-brain T1 anatomy. We repeated this for each fMRI session to transform all their data to the same anatomical space.

The resulting data was imported into Vistasoft's mrVista framework (github.com/vistalab/vistasoft). For timing response data, we identified the parts of each scanning run where each stimulus configuration was presented and averaged these fMRI responses together across all runs and sessions¹⁴. We also separately averaged data from odd and even runs to allow cross-validation in subsequent modelling. For visual field mapping data, we averaged all scan runs together.

Visual field mapping analysis and visual field map definitions

We analyzed visual field mapping data following a standard pRF modelling procedure^{26,27}. We identified visual field map borders based on reversals in the cortical progression of the polar angle of voxels' visual field position preferences, manually identifying these on an inflated rendering of each participant's cortical surface (Fig. 2 and Supplementary Figs. 2 and 3). These formed our main regions of interest (ROIs). As well as the early visual field maps (V1, V2, V3), we identified higher visual field maps in the lateral/temporal occipital (LO1-LO2, TO1-TO2), parietal association (V3A/B, IPS0-IPS5), and frontal (sPCS1-sPCS2, iPCS) cortices with reference to landmarks identified in previous studies^{31,38-40}.

Timing response models

The current study compared the fits of established monotonic¹³ and tuned¹⁴ models of neural responses to visual event timing in different brain areas (Fig. 1 and Supplementary Fig. 1). Each of these models describes a particular relationship between event timing and the neural response amplitude to every event. We predict response amplitudes on a per-event basis, but these responses accumulate over a few seconds due to fMRI's measurement of slow changes in blood flow and oxygenation. The models predicted neural response amplitudes at the offset of events, since this is when the information about duration and period of the events was completely available to the participants.

These predictions were then convolved with a standard hemodynamic response function (HRF) to construct a prediction for the fMRI response. Then, any free parameters were fit to maximize the correlation between the predicted response and the actual, recorded data (variance explained), giving a fit neural response model. Since HRF parameters substantially differ between participants, the resulting neural response model was used to determine participant-specific HRF parameters as described elsewhere²⁷. Using these participant-specific HRF parameters, the neural response model's free parameters were refit.

Given that tuned response model has more free parameters than the monotonic response model, both model fits were cross-validated before comparison. This cross-validation was achieved by fitting each response model's free parameters on the even or odd scans and evaluating the resulting model's fit on the complementary half. Because fMRI response amplitudes change arbitrarily between scans and sessions⁴¹, the scaling between the predicted fMRI time course and complementary scan data was refit during cross-validation.

Monotonic response model

The monotonic response model has been demonstrated to capture effects of event timing on fMRI responses in early visual areas. This model has two components that scale

independently with event duration and frequency (1/period)¹². The frequency and duration components were each scaled by free compressive exponent parameters¹³. The neural response amplitude to each event was given by equation (1):

$$Amplitude \propto Duration^{expDur} \times AmplitudeRatio + \frac{Frequency^{expFreq}}{Frequency} \quad (1)$$

Where expDur and expFreq are the compressive exponent on duration and frequency respectively, in the range 0-1. AmplitudeRatio captures the relative amplitudes of the duration and frequency components, and was linearly solved by dividing the duration component's response amplitude by the frequency component's response amplitude from a general linear model. The compressive exponents were fit by testing a large set of candidate values to find the parameter combination that best predicted the measured response of each voxel.

Note that monotonically decreasing responses would imply a response that decreases with increasing duration or frequency, so these should not be found in the transient and sustained responses of early visual areas. Therefore, we restricted the response amplitudes of the duration and frequency components to positive values. If, for a certain voxel, one of the components' scaling factors (e.g. the response amplitude of the duration component) was fit as negative, we set this scaling factor to 0 and fit the response amplitude to the other component alone. If the other scaling factor (e.g. the response amplitude of the frequency component) was positive after any refitting, these scaling factors were used to compute the amplitude ratio in the final model. If the other scaling factor (e.g. the response amplitude of the frequency component) was negative after any refitting, it was also set to zero and the variance explained by the model in this voxel was zero. As a result, the amplitude ratio cannot be below zero. Since no monotonically decreasing responses were allowed, the monotonically increasing model will be referred to simply as the monotonic response model.

Tuned response model

In the tuned response model, the neural response amplitude to each event is described by a 2-dimensional anisotropic Gaussian function, whose response is scaled by a compressive exponent on frequency. We chose the tuned model response function that best predicted neural responses throughout the timing map network¹⁴. The function can be described using the following equations (2-4):

$$X = (Duration - Duration_{pref}) \times \cos(\theta) - (Period - Period_{pref}) \times \sin(\theta) \quad (2)$$

$$Y = (Duration - Duration_{pref}) \times \sin(\theta) - (Period - Period_{pref}) \times \cos(\theta) \quad (3)$$

$$Amplitude \propto e^{-0.5 \times \left(\left(\frac{Y}{\sigma_{maj}} \right)^2 + \left(\frac{X}{\sigma_{min}} \right)^2 \right)} \times \frac{Frequency^{expFreq}}{Frequency} \quad (4)$$

The six free parameters of the model are: the preferred duration ($Duration_{pref}$) and the preferred period ($Period_{pref}$) around which the Gaussian function's mean is centered; the standard deviations along its major and minor axes (σ_{maj} and σ_{min}); the angulation of its major axis (θ); and the compressive exponent on frequency to which the response was scaled. The fitting procedure consisted of testing a large combination of free parameters, followed by a gradient descent between the best-fitting parameter combination and its neighbors.

After fitting this model, it is possible that some voxels are best described by a Gaussian function with a preferred duration and/or period on the boundaries or outside of the presented stimulus range (i.e. a duration or period below 60 ms or above 990 ms). However, it would be impossible to find the exact preferred duration and/or period belonging to this Gaussian²⁶. Also, this would increase the risk that a tuned response model produces monotonic-like responses by simply adhering to large values for these free parameters. Therefore, the variance explained of the tuned response model in voxels where the preferred duration or period was outside the presented stimulus range was set to 0 during cross-validation.

The fit parameters of this tuned response model allowed us to define the borders of timing map regions of interest, as previously described¹⁴. Here we use the same timing maps defined in that study (Supplementary Fig. 4). Briefly, we took the variance explained values for each vertex on the cortical surface and performed surface-based clustering on these values. In some cases, we merged two adjacent clusters into a single ROI, or split a single large cluster into two parts where it contained two contiguous maps (common in TTOP/TTOA and TPCS/TPCM).

Model Comparisons

We removed voxels for which the variance explained of both models was 0.2 or less from model comparisons. Then, the model fits were statistically compared to each other in each visual field map and timing map ROI. For each hemisphere, the model variance explained was averaged across the voxels in each ROI, separately for the two cross-validation splits. If a hemisphere did not have any voxels with a variance explained above 0.2 in a specific ROI and a specific cross-validation split, that ROI in that hemisphere and cross-validation split was excluded from subsequent comparison (n=1038 visual field map measurements; n=606 timing map measurements).

To assess how model fits differed between ROIs, the average variance explained by the two models was then compared using a 3-factor ANOVA (factors: participant, ROI, and model; interaction: ROI and model) using MATLAB's *anovan* function. The interaction factor here demonstrated that different ROIs' responses were best captured by different response models. As the data was normally distributed (Jarque-Bera test with FDR correction), post-hoc two-sided paired t-tests with FDR correction were then used to assess which model best captured the responses in each ROI, using hemispheres and cross-validation splits as independent measures. Furthermore, in order to demonstrate the absence

of a difference between models, we computed Bayes factors for each of these paired comparisons in JASP⁴² using a Cauchy prior distribution with $r = 1/\sqrt{2}$.

Progression of model fits within visual field maps

To study whether the model fits changed throughout the visual field, variance explained was plotted against the distance to the center of the visual field (eccentricity), where the stimulus was shown. To achieve this, we first grouped all voxels within a map across hemispheres and cross-validation splits. For each visual field map, we then binned voxels according to their preferred eccentricity. Each bin had a range of 0.2° , centered on eccentricities from 0.1° to 5.5° . Voxels that did not have a variance explained above 0 for either model were excluded. For each bin, the mean variance explained and its standard error were computed for each model. Bins with less than 50 voxels were excluded.

To visualize the progression of model variance explained across eccentricity, we fitted a cumulative Gaussian sigmoid function (i.e. greater variance explained at low eccentricities, near the stimulus) and a quadratic function (which also allows a maximum variance explained at higher eccentricity) to the progression in each visual field map. The free parameters for the sigmoid function were point of inflection, slope, maximum, and minimum. We computed the 95% confidence interval of this sigmoid from 1000 bootstrap iterations. The free parameters for the quadratic function were intercept, slope, and quadratic term. We computed the 95% confidence interval of the quadratic function linearly using `polyfit` and `polyconf` in MATLAB. For visualizing the progression of model fits with eccentricity, we chose the function that was best correlated with the eccentricity bin means.

Note that the quadratic function can also capture a linear relation, so comparing these progressions is only useful for plotting the data rather than determining whether fits decrease with eccentricity. Therefore, to statistically compare the difference between model fits near

and far from the center of the visual field, we computed the average variance explained in each hemisphere, each cross-validation split and each map for eccentricities lower than 1° and eccentricities higher than 2° for all voxels with a variance explained above 0 for either model ($n=1088$ visual field map measurements). To assess how model fits differed between these eccentricity ranges, these averages were then compared using a 3-factor ANOVA (factors: participant, ROI, and eccentricity range; interaction: ROI and eccentricity range) using MATLAB's *anovan* function. The interaction factor here demonstrated that the different ROIs had better variance explained in different eccentricity ranges. The differences between variance explained at different eccentricity ranges were not normally distributed in all ROIs (Jarque-Bera test with FDR correction). Therefore, we used post-hoc two-sided Wilcoxon signed-rank tests with FDR correction to assess how variance explained differed with eccentricity range in each ROI, using hemispheres and cross-validation splits as independent measures. Here, we calculated the effect size as $r = Z/\sqrt{n}$. Furthermore, in order to demonstrate the absence of a difference between eccentricity ranges, we computed Bayes factors for each of these non-parametric paired comparisons in JASP⁴² with 5 chains of 1000 iterations, using a Cauchy prior distribution with $r = 1/\sqrt{2}$.

We also computed the difference in variance explained between the monotonic and tuned response models for each voxel (tuned minus monotonic). This quantified the component of the voxel's response that reflects timing tuning and cannot be explained by monotonic responses alone. The eccentricity progression of variance explained differences across the visual field map was also assessed using the statistical procedure described above.

Data availability

The data sets generated during the current study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

The code that supports the findings of this study is available from the corresponding author upon reasonable request.

References

1. Spence, C. & Squire, S. Multisensory integration: maintaining the perception of synchrony. *Curr. Biol.* **13**, 519–521 (2003).
2. Vroomen, J. & Keetels, M. Perception of intersensory synchrony: a tutorial review. *Attention, Perception, Psychophys.* **72**, 871–884 (2010).
3. Karmarkar, U. R. & Buonomano, D. V. Timing in the absence of clocks: encoding time in neural network states. *Neuron* **53**, 427–438 (2007).
4. Ivry, R. B. & Schlerf, J. E. Dedicated and intrinsic models of time perception. *Trends Cogn. Sci.* **12**, 273–280 (2008).
5. Buetti, D., Bahrami, B. & Walsh, V. Sensory and association cortex in time perception. *J. Cogn. Neurosci.* **20**, 1054–1062 (2008).
6. Morrone, M. C., Ross, J. & Burr, D. Saccadic eye movements cause compression of time as well as space. *Nat. Neurosci.* **8**, 950–954 (2005).
7. Bruno, A., Ayhan, I. & Johnston, A. Retinotopic adaptation-based visual duration compression. *J. Vis.* **10**, (2010).
8. Bruno, A., Ayhan, I. & Johnston, A. Changes in apparent duration follow shifts in perceptual timing. *J. Vis.* **15**, 1–18 (2015).
9. Heron, J. *et al.* Duration channels mediate human time perception. *Proc. R. Soc. B Biol. Sci.* **279**, 690–698 (2012).
10. Maarseveen, J., Hogendoorn, H., Verstraten, F. A. J. & Paffen, C. L. E. An investigation of the spatial selectivity of the duration after-effect. *Vision Res.* **130**, 67–75 (2017).
11. Roseboom, W. *et al.* Activity in perceptual classification networks as a basis for human subjective time perception. *Nat. Commun.* **10**, (2019).

12. Stigliani, A., Jeska, B. & Grill-Spector, K. Encoding model of temporal processing in human visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11047–E11056 (2017).
13. Zhou, J., Benson, N. C., Kay, K. N. & Winawer, J. Compressive temporal summation in human visual cortex. *J. Neurosci.* **38**, 691–709 (2018).
14. Harvey, B. M., Dumoulin, S. O., Fracasso, A. & Paul, J. M. A network of topographic maps in human association cortex hierarchically transforms visual timing-selective responses. *Curr. Biol.* **30**, 1424-1434.e6 (2020).
15. Protopapa, F. *et al.* Chronotopic maps in human supplementary motor area. *PLoS Biol.* **17**, 1–34 (2019).
16. Harvey, B. M. & Dumoulin, S. O. A network of topographic numerosity maps in human association cortex. *Nat. Hum. Behav.* **1**, (2017).
17. Harvey, B. M., Fracasso, A., Petridou, N. & Dumoulin, S. O. Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13525–13530 (2015).
18. Hofstetter, S., Cai, Y., Harvey, B. M. & Dumoulin, S. O. Topographic maps representing haptic numerosity reveals distinct sensory representations in supramodal networks. *Nat. Commun.* **12**, 1–13 (2021).
19. Arrighi, R., Togoli, I. & Burr, D. C. A generalized sense of number. *Proc. R. Soc. B Biol. Sci.* **281**, 1–7 (2014).
20. Hurewitz, F., Gelman, R. & Schnitzer, B. Sometimes area counts more than number. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19599–19604 (2006).
21. Verguts, T. & Fias, W. Representation of number in animals and humans: a neural model. *J. Cogn. Neurosci.* (2004) doi:10.1162/0898929042568497.

22. Dehaene, S. & Changeux, J. P. Development of elementary numerical abilities: a neuronal model. *J. Cogn. Neurosci.* (1993) doi:10.1162/jocn.1993.5.4.390.
23. Paul, J. M., van Ackooij, M., ten Cate, T. C. & Harvey, B. M. Transformation from local image contrast to location-independent numerosity tuning in human extrastriate cortex. *bioRxiv* 2021.03.28.437364 (2021) doi:10.1101/2021.03.28.437364.
24. Merchant, H., Zarco, W., Pérez, O., Prado, L. & Bartolo, R. Measuring time with different neural chronometers during a synchronization-continuation task. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 19784–19789 (2011).
25. Merchant, H., Pérez, O., Zarco, W. & Gámez, J. Interval tuning in the primate medial premotor cortex as a general timing mechanism. *J. Neurosci.* **33**, 9082–9096 (2013).
26. Dumoulin, S. O. & Wandell, B. A. Population receptive field estimates in human visual cortex. *Neuroimage* **39**, 647–660 (2008).
27. Harvey, B. M. & Dumoulin, S. O. The relationship between cortical magnification factor and population receptive field size in human visual cortex: Constancies in cortical architecture. *J. Neurosci.* **31**, 13604–13612 (2011).
28. Zeki, S. & Shipp, S. The functional logic of cortical connections. *Nature* **335**, 311–317 (1988).
29. Maunsell, J. H. R., Nealey, T. A. & DePriest, D. D. Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *J. Neurosci.* **10**, 3323–3334 (1990).
30. Merigan, W. H. & Maunsell, J. H. R. How parallel are the primate visual pathways? *Annu. Rev. Neurosci.* **16**, 369–402 (1993).
31. Amano, K., Wandell, B. A. & Dumoulin, S. O. Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *J. Neurophysiol.* **102**, 2704–2718 (2009).

32. Dumoulin, S. O. *et al.* A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. *Cereb. Cortex* **10**, 454–463 (2000).
33. Cona, G., Wiener, M. & Scarpazza, C. From ATOM to GradiATOM: cortical gradients support time and space processing as revealed by a meta-analysis of neuroimaging studies. *Neuroimage* **224**, (2021).
34. Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. What's new in Psychtoolbox-3. *Perception* (2007)
doi:<https://doi.org/10.1068/v070821>.
35. Harvey, B. M., Klein, B. P., Petridou, N. & Dumoulin, S. O. Topographic representation of numerosity in the human parietal cortex. *Science* (80-.). **341**, 1123–1126 (2013).
36. Andersson, J. L. R., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* **20**, 870–888 (2003).
37. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
38. Mackey, W. E., Winawer, J. & Curtis, C. E. Visual field map clusters in human frontoparietal cortex. *Elife* **6**, (2017).
39. Silver, M. A., Ress, D. & Heeger, D. J. Neural correlates of sustained spatial attention in human early visual cortex. *J. Neurophysiol.* **97**, 229–237 (2007).
40. Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field maps in human cortex. *Neuron* **56**, 366–383 (2007).
41. McGonigle, D. J. *et al.* Variability in fMRI: an examination of intersession differences. *Neuroimage* **11**, 708–734 (2000).

42. JASP team. JASP (Version 0.14.1). (2020).

Acknowledgments

This work was supported by the Netherlands Organization for Scientific Research (452.17.012 to B.M.H.)

Author Contributions

Conceptualization, J.M.P and B.M.H.; Methodology, E.H., J.M.P, M.A. and B.M.H.; Software, E.H., J.M.P, M.A. and B.M.H.; Validation, E.H., J.M.P and B.M.H.; Formal Analysis, E.H., J.M.P, and B.M.H.; Investigation, E.H., J.M.P and B.M.H.; Data Curation, J.M.P and B.M.H.; Writing – Original Draft, E.H. and B.M.H.; Writing – Review & Editing, E.H., J.M.P, M.A., N.S. and B.M.H.; Visualization, E.H., J.M.P and B.M.H.; Supervision, N.S. and B.M.H.; Project Administration, B.M.H.; Funding Acquisition, B.M.H.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence to B.M.H.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData.xlsx](#)
- [MovieS1.mp4](#)
- [SupplementaryFigures.docx](#)