

RESEARCH ARTICLE

Dynamic divisive normalization circuits explain and predict change detection in monkey area MT

Udo A. Ernst^{1*}, Xiao Chen¹, Lisa Bohnenkamp¹, Fingal Orlando Galashan², Detlef Wegener²

1 Computational Neurophysics Lab, Institute for Theoretical Physics, University of Bremen, Bremen, Germany, **2** Brain Research Institute, University of Bremen, Bremen, Germany

* udo@neuro.uni-bremen.de



OPEN ACCESS

Citation: Ernst UA, Chen X, Bohnenkamp L, Galashan FO, Wegener D (2021) Dynamic divisive normalization circuits explain and predict change detection in monkey area MT. *PLoS Comput Biol* 17(11): e1009595. <https://doi.org/10.1371/journal.pcbi.1009595>

Editor: Martin Giese, University of Wales, UNITED KINGDOM

Received: September 2, 2020

Accepted: October 27, 2021

Published: November 12, 2021

Copyright: © 2021 Ernst et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting Information](#) files. Running the code requires Matlab R2020a (The MathWorks, Inc.) and Python (Version 3.6). To perform all data analyses and fits, execute script 'PLOS_run_all.m' from the main directory.

Funding: The work was supported by BMBF (Bundesministerium für Bildung und Forschung) grant 01GQ1106 to UAE, and DFG (Deutsche Forschungsgemeinschaft) grants WE 5469/2-1 and

Abstract

Sudden changes in visual scenes often indicate important events for behavior. For their quick and reliable detection, the brain must be capable to process these changes as independently as possible from its current activation state. In motion-selective area MT, neurons respond to instantaneous speed changes with pronounced transients, often far exceeding the expected response as derived from their speed tuning profile. We here show that this complex, non-linear behavior emerges from the combined temporal dynamics of excitation and divisive inhibition, and provide a comprehensive mathematical analysis. A central prediction derived from this investigation is that attention increases the steepness of the transient response irrespective of the activation state prior to a stimulus change, and irrespective of the sign of the change (i.e. irrespective of whether the stimulus is accelerating or decelerating). Extracellular recordings of attention-dependent representation of both speed increments and decrements confirmed this prediction and suggest that improved change detection derives from basic computations in a canonical cortical circuitry.

Author summary

The world is a dynamic place: visual scenes are changing rapidly, and decisions have to be taken quickly and reliably to ensure successful behavior and, finally, survival. This creates a challenge for the brain, since it has to fulfill two requirements at the same time: It has to detect changes regardless of their magnitude and sign of change, and it has to primarily focus on behaviorally relevant changes. We here studied transient stimulus-speed change responses of neurons in motion-sensitive area MT and identify a mechanism supporting both of these requirements. This mechanism can be realized by an elementary neural model circuit which closely fits physiological data. The model is based on dynamic divisive inhibition generating fast transient rate modulations in response to rapid input changes. We analyzed this circuit mathematically and arrived at the formal prediction that attention will consistently increase the steepness of the transient, irrespective of the magnitude of the pre-change activation and the sign of the input change, thus allowing for

5469/3-1 to DW. 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.

faster, reliable reaction to any attended event. By performing single-cell recordings with both stimulus accelerations and decelerations, which could either be attended or not attended, the experimental data fully confirmed the predictions from modelling.

Introduction

For change detection in behaviorally relevant situations, stimulus-induced, strong transient firing rate modulations provide a powerful signal to downstream visuomotor areas on short timescales [1,2]. Such transients are particularly information-rich [3,4], and the temporal sensitivity of neurons to stimulus changes is suggested to constitute a major function of neuronal tuning properties [5]. Human behavioral performance to detect speed changes [6] correlates with the size of area MT change transients [7] and strongly improves with both spatial and feature-directed attention [8]. Accordingly, both forms of attention were found to increase sustained and transient MT firing rates before and after a stimulus change [9,10], and the transient response after the change also correlates with reaction times [9].

This close relation between neuronal transients and behavioral change detection performance on the one hand, and the general effect of attention to increase neuronal response rates [11,12] on the other, raises the question how attention exerts its beneficial effects under conditions where firing rate increases seem to impede the formation of pronounced transients and counteract behavioral performance. For example, because attention-dependent enhancement of firing rates brings a neuron closer to its maximum activity, a stimulus-induced transient firing rate increase under attention might be smaller than without attention. Furthermore, a negative transient would start from a larger pre-change activation level and is presumably having a smaller absolute negative peak than under conditions of no or remote attention. Absolute firing rates of the transient, therefore, may only allow for a poor prediction of attention-dependent improvements in behavioral change detection. As a consequence, the neuronal circuit processing the change depends on some form of normalization to compensate for differences in absolute activity. Moreover, to facilitate change detection, attentional modulation of the circuit should induce a consistent effect on neuronal transients that is independent from the stimulus-induced activation level before the change. It is unclear presently, which feature of a change-transient can be most consistently modulated by attention independent of the specific stimulus condition, and which neuronal mechanism might underlie its corresponding dynamics.

To investigate this issue, we set up and mathematically analyze a biologically plausible, canonical circuit providing divisive inhibition to an excitatory unit, and apply the model to a wide range of different stimulus conditions under, at first, passive viewing conditions [7]. By then introducing an input response gain to simulate top-down modulation of change detection by visual attention, the model circuit predicts a main effect on response rise times not only for positive transients but also for the case of negative transients, i.e. rapid rate decreases in a population of neurons. To test this prediction experimentally, we recorded neurons from motion-sensitive area MT while monkeys were engaged in a change detection task for both speed increments and decrements, eliciting large positive and negative population transients, respectively. We show that MT activity exactly follows the prediction of the model, having steeper response slopes irrespective of the sign of the transient. Thus, the model circuit consistently explains change detection under different stimulus and attention conditions by a rather simple computational unit, realized in the canonical circuitry of the cortex.

Results

Transient neural responses in area MT

Neurons in area MT are well-activated by moving stimuli such as localized drifting Gabor patches presented inside their receptive fields (RFs). Neural responses are tuned to the direction of movement, exhibiting maximal activity when the stimulus is moving in the cell's preferred direction [13]. The additional preference of MT neurons for a particular stimulus speed can be approached by a Gaussian function in log-velocity space [14,15]. Rapid changes in the input to these neurons result in pronounced, transient changes of their activation. An example from previous work [7] is showing the population response of an ensemble of MT neurons ($N = 94$) to a range of sudden stimulus accelerations and decelerations (Fig 1A). Transients show up as fast increases (or decreases) in firing rate, followed by a slower decrease (or increase) in firing rate back to a sustained level of activation. Such transients are visible in response to any instantaneous change in stimulation (black arrows in Fig 1A): stimulus onset, motion onset, speed change, motion offset, and stimulus offset. Due to their causal relation to visual perception and change detection [16], we aimed to study their properties and non-linear dynamics in a theoretical framework providing access to a comprehensive mathematical analysis, and to experimentally test predictions derived from it.

Model for transient activation in area MT

For modeling transient responses, we consider a circuit where an input I activates an excitatory and an inhibitory unit, with output A_i of the inhibitory unit providing divisive inhibition to

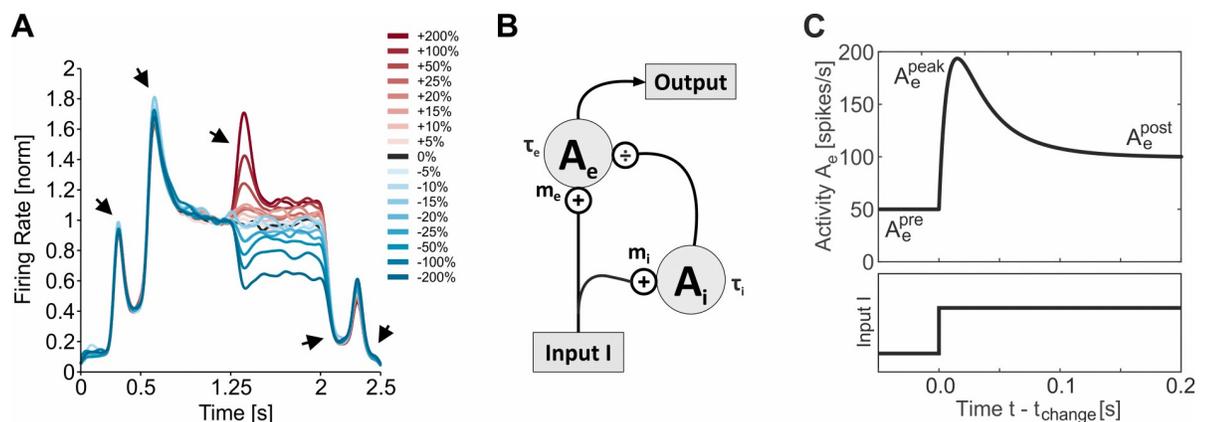


Fig 1. Transient responses in area MT and model circuit. (A) Population response of 94 MT neurons exhibiting large transients (black arrows) in response to stimulus onset (at $t = 0.25$ s), movement onset (at $t = 0.5$ s), speed change (at $t = 1.25$ s), movement offset (at $t = 2.0$ s) and stimulus offset (at $t = 2.25$ s). Color scale indicates sign and magnitude of the speed change (blue for decelerations, red for accelerations). Speed change magnitudes are given in Weber fractions, i.e. with the lower speed as the reference. Modified from [7]. (B) Model circuit providing divisive inhibition A_i to an MT neuron with activation A_e . τ_e and τ_i denote time constants and m_e and m_i denote gains of the corresponding units. (C) Example response of the model circuit (upper graph) to a positive input change at time t_{change} (lower graph). Starting from activation level $A_e^{pre} = 50$ spikes/s, the circuit exhibits a fast transient increase in activation reaching a peak response A_e^{peak} before decreasing more slowly to a post-change, sustained activation level $A_e^{post} = 100$ spikes/s (with parameters: $\tau_e = 10$ ms, $\tau_i = 40$ ms, $A_e^{max} = 120$ spikes/s).

<https://doi.org/10.1371/journal.pcbi.1009595.g001>

the excitatory unit (Fig 1B). The dynamics of the circuit is given by the following differential equations:

$$\tau_e \frac{dA_e(t)}{dt} = -A_e(t) + g_e \left(\frac{I(t)}{A_i(t) + \sigma} \right) \tag{1}$$

$$\tau_i \frac{dA_i(t)}{dt} = -A_i(t) + g_i(I(t)), \tag{2}$$

where output $A_e(t)$ of the excitatory unit represents the instantaneous firing rate A_e of a neuron in MT at time t , with I being the external input provided by the visual stimulus. Terms τ_e, τ_i denote time constants, σ is a constant positive offset, and g_e, g_i represent gain functions realized by threshold-linear rectification, with m_e, m_i indicating gain factors, and θ_e, θ_i indicating thresholds:

$$g_e(I) = m_e(I - \theta_e) \text{ for } I > \theta_e, \text{ and } 0 \text{ otherwise} \tag{3}$$

$$g_i(I) = m_i(I - \theta_i) \text{ for } I > \theta_i, \text{ and } 0 \text{ otherwise,} \tag{4}$$

With constant suprathreshold input I_0 , the steady-state solution

$$A_e^{steady} = A_e(t \rightarrow \infty) = m_e \left(\frac{I_0}{m_i(I_0 - \theta_i) + \sigma} - \theta_e \right) \tag{5}$$

for Eq (1) is equivalent to a standard divisive normalization model [17,18], where A_i represents feedback from co-activated neighboring MT columns. As such, our model can be interpreted as a dynamical reformulation of static divisive normalization, with relaxation dynamics similar to those proposed for recurrent divisive normalization or low-pass filtering the output of a divisive normalization circuit [19,20]. The advantage of having two time constants for excitation and inhibition is the explicit representation of transients, allowing activation A_e to quickly follow changes in I on a fast time scale τ_e , while divisive inhibition is acting on a longer time scale τ_i , bringing the output towards a sustained, steady-state level of activation (Fig 1C).

Model fit to transients from experimental data

To investigate how well the model explains experimental data, $A_e(t)$ was fitted to MT stimulus onset transients. Because MT neurons exhibit spontaneous activity with low firing rates even in absence of a visual stimulus, we assume $\theta_e = \theta_i = 0$ and $I(t) > 0$, which allows to replace Eqs (3) and (4) by linear gain functions to simplify the analysis. By modeling stimulus onset as an instantaneous change in external input at $t = t_{change}$ from I^{pre} to I^{post} , Eq (2) can be explicitly solved, assuming that for $t < t_{change}$ the system is in its steady state for a constant input I^{pre} . The result can be rewritten in terms of the sustained activation A_e^{pre} before stimulus onset, and sustained activation A_e^{post} after decay of the transient response, to account for the fact that experimental access is given to the output of the neuron rather than to its input:

$$\tau_e \frac{dA_e}{dt} = -A_e + A_e^{max} \left[\frac{\frac{1}{A_e^{post}} - \frac{1}{A_e^{pre}}}{\frac{1}{A_e^{pre}} - \frac{1}{A_e^{max}}} \exp\left(-\frac{t - t_{change}}{\tau_i}\right) + \frac{A_e^{max}}{A_e^{post}} \right]^{-1} \tag{6}$$

The ratio of the two gain factors $A_e^{max} = m_e/m_i$ designates the theoretical maximum sustained activation of the circuit. We used a grid search to find the remaining free parameters τ_e, τ_i , and A_e^{max} to minimize the average quadratic error between model activation and recorded

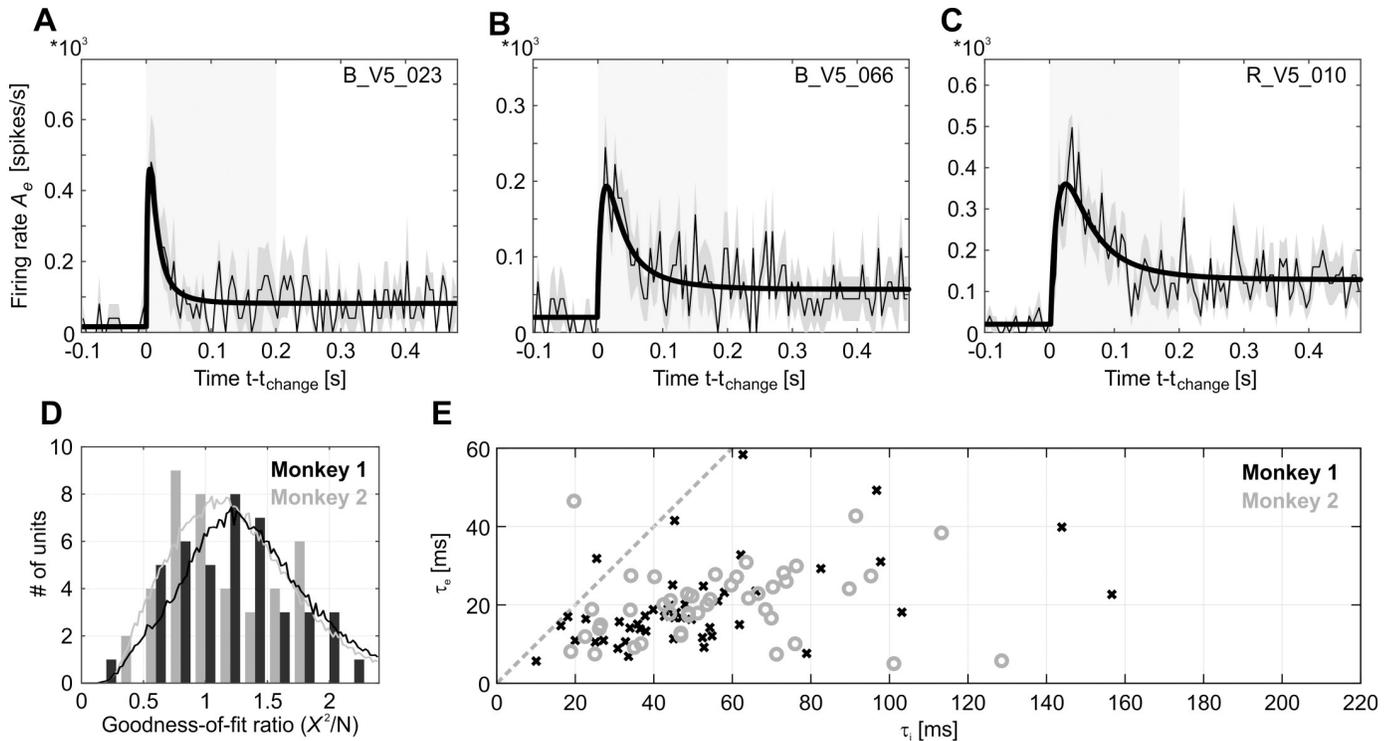


Fig 2. Model fits to stimulus onset transients. (A–C) Activity of three example multi-units (thin lines) and corresponding model fit (solid lines). Gray shading indicates standard error over 5–10 trial repetitions, light gray box indicates fitting interval. Time axis is shown relative to neural response onset. Note that actual responses were not smoothed for visual purposes. (D) Model Chi-square distributions of the recorded units for two monkeys (M1, black bars; M2, gray bars). Solid lines show surrogate data distributions for artificially generated spike data from models with known parameters (M1, black; M2, gray). (E) Scatter plot for fitted time scales τ_e versus τ_i (M1, crosses; M2, open circles).

<https://doi.org/10.1371/journal.pcbi.1009595.g002>

MT firing rate during the transient response, excluding units yielding total spike counts that were too low to allow for successful fits (cf. Methods).

This structurally very simple model allowed for a close approximation of transient and sustained MT responses to motion onsets, despite the significant differences in their shape, caused by, among other things, the different tuning of individual neurons to actual stimulus speeds. Three examples of multi-unit responses and corresponding model fits are given in Fig 2A–2C, and the Chi-squared distributions for the total of 48 (monkey 1 (M1)) and 44 (M2) remaining units are shown in Fig 2D. The χ^2/N_t ratio was sufficiently close to 1 for most units (M1: 1.49 ± 0.74 SD, M2: 1.37 ± 0.5 SD), indicating that fits were estimating the mean response nearly as good as the experimental data. This conclusion was confirmed by a comparison to surrogate distributions with the same statistical power and known "ground truth" (Fig 2D, see Methods for more details).

Excitatory time scales were much faster than inhibitory ones, as to be expected for the dynamics of transients (Fig 2E). For M1, mean τ_e was 17 ± 16 ms SD, mean τ_i was 45 ± 30 ms, average ratio τ_e/τ_i was 0.4 with SD 0.98, and average A_e^{max} was 87 ± 54 spikes/s. For M2, mean τ_e was 20 ± 9 ms, mean τ_i was 52 ± 25 ms, average ratio τ_e/τ_i was 0.38 with SD 0.34, and average A_e^{max} was 127 ± 99 spikes/s.

We also tested whether less than three free parameters per individual MT neuron would be sufficient to explain the different shapes of transient responses across the population of neurons. We therefore fixed one or more parameters (τ_e , τ_i , A_e^{max}) for all units in a population, and repeated the fitting procedure. With one fixed global parameter, the time-averaged chi-square

measure between model fit and experimental data increased by 3.3% (M1) and 2.8% (M2) on average. Which of the three parameters was globally defined was largely irrelevant, but fixing the time constant ratio τ_e/τ_i had the smallest impact, increasing the fitting error by only 1.6% (for both monkeys). With two or three global parameters fixed, however, the error increased strongly (14.7% (M1) and 13.8% (M2) for two fixed parameters, 42.3% (M1) and 36.6% (M2) for three fixed parameters).

Transient response characteristics to stimulus changes

After the model had been calibrated to motion onset transients, we were next interested to investigate whether it is capable to predict and explain the complex non-linear scaling of transients in response to changes in speed, which depend on the sign and magnitude of the physical speed change and the individual neurons' tuning characteristics [7]. Because time scales τ_e were shorter than τ_i , an idealized version of the model can be considered by assuming that the excitatory unit reacts infinitely fast to input changes. With this approximation, it is possible to solve Eq (6) explicitly for $\tau_i > 0$ and obtain the peak of the transient shortly after $t = t_{change}$ analytically (cf. Eqs (12) & (13) in Methods).

Peak amplitudes obtained from the model in this manner reproduced an important and so far, unexplained characteristic of MT change transients. In MT, peak amplitudes in response to speed changes exceed those expected from the neuron's speed tuning profile significantly if the speed before the change is away from the neuron's preferred speed. Neurons well-tuned to the pre-change speed are very poor change detectors, while neurons for which the pre-change speed is on the flank of their tuning curve have a strong impact on the population response [7]. This tuning-dependent, non-linear relationship between pre-change stimulus speed and individual neuronal tuning profiles is well captured by the model. Simulated speed changes, realized by step functions applied to the circuit's input at time t_{change} (Fig 1C), predict speed change transients very closely matching the experimental data with regard to sign and amplitude of the peak for both changes occurring on the ascending and the descending flank of the tuning curve (Fig 3). Thus, the model reproduces the full dynamics of physiological change transients, including the over- and undershooting of peak amplitudes.

Model predictions for attention-dependent modulation of change transients

Because transients signal stimulus changes in a rapid and pronounced manner, they were previously suggested as a possible target for attentional modulation and a neuronal mechanism to speed up reaction times [9,21,22]. In area MT, using speed changes of 100% magnitude, attention was found to modulate both the peak amplitude and latency of a change transient [9,10]. Modulations due to other magnitudes of change, including negative changes, have not been investigated yet. Therefore, in our model, we next studied the dynamical effects of attention on transients for speed changes of arbitrary magnitude (without separation of time scales as in the previous subsection, allowing for arbitrary combinations of excitatory and inhibitory time constants). We included attention by simply assuming a multiplicative scaling [11] of the input I by a factor $\alpha > 1$, $I \rightarrow \alpha I$ (Fig 4A, top left).

Ideally, to improve computation and, ultimately, perception, the effect of attention should be consistent across the entire dynamical range of the circuit. In particular, for attention to be effective, any change in the external input I causing a positive (or negative) transient (as e.g. the traces for the $\pm 100\%$ changes in Fig 1A) should be associated with attention-dependent modulations preserving the sign of the transient independent from the neuron's pre-change activation, as e.g. a consistent increase (or decrease) in the transient's peak amplitude. If,

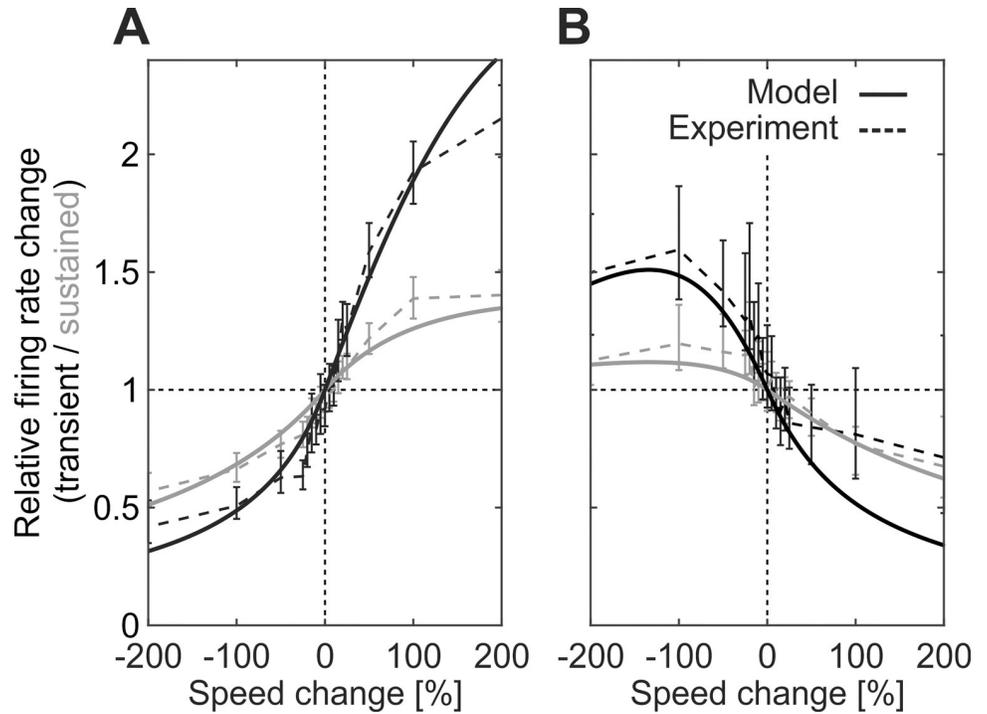


Fig 3. Transient/sustained rate changes following stimulus speed changes. Experimentally observed transient (black) and sustained (gray) rate changes in response to positive and negative speed changes (dashed lines) of different magnitude, and corresponding predictions from model analysis (solid lines). Base speed before the change was located either on (A) the ascending flank, or (B) the descending flank of the cells' speed tuning curves.

<https://doi.org/10.1371/journal.pcbi.1009595.g003>

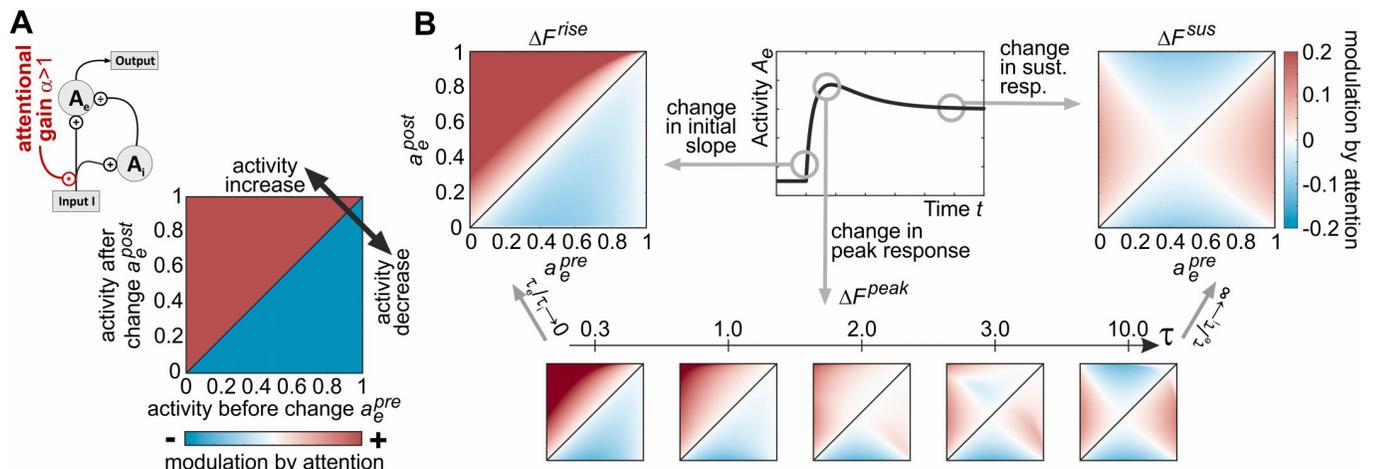


Fig 4. Model predictions for attention-induced changes in slope, peak, and sustained responses. (A) Attention is included into the model by a simple multiplicative gain change of the input (top left schematic). Ideally, attention should elicit a consistent modulation of the neuronal change response, as e.g. enhancing a neural response feature for positive input changes (above diagonal, red shading), and suppressing it for negative input changes (below diagonal, blue shading) irrespective of the particular pre-change activation level. (B) Predicted attention-dependent changes in slope ΔF^{rise} (top left, indicating consistent modulation by attention), sustained activation ΔF^{sus} (top right, indicating inconsistent, pre-change activation-dependent modulation by attention), and relative peak height ΔF^{peak} (bottom plots). Relative peak height critically depends on the time scale ratio τ_e/τ_i and becomes similar to ΔF^{rise} for $\tau_e/\tau_i \rightarrow 0$ and similar to ΔF^{sus} for $\tau_e/\tau_i \rightarrow \infty$. Axes and color scales are identical for all plots in (B), indicated at the top left and right chart.

<https://doi.org/10.1371/journal.pcbi.1009595.g004>

however, peak amplitude is only increased for low pre-change activation levels, but decreased for high pre-change activation levels for otherwise identical stimulus conditions, the effect of attention would be inconsistent. Fig 4A (bottom right) exemplifies a consistent pattern of attentional modulation for stimulus-change responses associated with positive and negative rate changes, normalized to a dynamic range between 0 (lowest level) and 1 (highest level). The colored surface illustrates a consistent (i.e. pre-change activity independent) positive modulation by attention for stimulus-induced rate increases (above diagonal), and a corresponding negative modulation by attention for stimulus-induced rate decreases (below diagonal).

The model was used to investigate three different change-response features regarding their dependence from attention and pre-change activity: the initial slope F^{rise} of the transient (i.e. rise/decay time), the maximal amplitude F^{peak} of the transient, and the sustained activation F^{sus} following the transient. For analysis, pre- and post-change activities A_e^{pre} and A_e^{post} were normalized from absolute values to activations $a_e^{pre} = A_e^{pre}/A_e^{max} \in [0, 1]$ and $a_e^{post} = A_e^{post}/A_e^{max} \in [0, 1]$ relative to the hypothetical maximum response A_e^{max} . Peak and sustained responses were quantified relative to pre-change activation, assuming that any plausible change detection circuit needs to base its computation on the number of spikes exceeding or falling below this level. The initial slope becomes

$$F^{rise}(a_e^{pre}, a_e^{post}, \alpha) = \left. \frac{dA_e}{dt} \right|_{t=t_{change}} = \frac{1}{\tau_e} \frac{a_e^{post} - a_e^{pre}}{1 - a_e^{pre}} \frac{\alpha}{a_e^{pre}(\alpha - 1) + 1} \tag{7}$$

and its attention-induced change ΔF^{rise} is visualized in Fig 4B (left). Similarly, the sustained activation level F^{sus} is given by

$$F^{sus}(a_e^{pre}, a_e^{post}, \alpha) = \frac{\alpha}{\alpha - 1 + (a_e^{post})^{-1}} - \frac{\alpha}{\alpha - 1 + (a_e^{pre})^{-1}} \tag{8}$$

and its attention-induced change ΔF^{sus} relative to the pre-change activity is plotted in Fig 4B (right). Finally, the relative change in maximal amplitude ΔF^{peak} depends on the specific ratio of the excitatory and inhibitory time constants τ_e/τ_i . Numerical evaluations for different values of τ reveal that for $\tau_e/\tau_i \rightarrow 0$, ΔF^{peak} approaches ΔF^{rise} , while for $\tau_e/\tau_i \rightarrow \infty$, it approaches ΔF^{sus} (Fig 4B, bottom panels). These computations reveal two main insights: First, the slope and the sustained response indicate the two extremes of the analysis. The slope is subject to a strong and consistent pattern of attentional modulation, which is independent of both the overall activity and the sign of the rate change, indicating generally faster transients with attention for both speed increments and decrements. In contrast, the sustained response exhibits an inconsistent pattern of attentional modulation, with the sign of the modulation depending on the overall activation of the neuron before the change. Second, as a function of the specific ratio of excitatory and inhibitory time constants, the modulation of the peak amplitude shifts between these extremes. Attentional modulation becomes stronger and more consistent the smaller the ratio τ_e/τ_i , but is attenuated (and, theoretically, even inconsistent) for larger values of τ_e/τ_i . Thus, the model makes the explicit prediction that, like positive transients, negative transients (i.e. rapid decreases in firing rate) will have a steeper rise time, i.e. shorter latencies, under the influence of attention as well as higher relative peaks, assuming τ_e/τ_i being in the previously estimated range (Fig 2E). The inconsistent pattern of attentional modulation as predicted for the sustained response can be explained by having a closer look at the steady state in Eq (5) with θ_e and θ_i set to zero, which can be rearranged into the form of a Naka-Rushton equation. For an attention-scaled input of αI with $\alpha > 1$ for small inputs I , the function rises more steeply (Fig 4B right, red region above diagonal) than with $\alpha = 1$ (no attention). From a critical input

strength $I > \sigma/(\sqrt{\alpha}m_i)$ on, however, this behavior reverses and the function rises more slowly with attention (Fig 4B right, blue region above diagonal).

Interestingly, the qualitative results depicted in Fig 4 remain identical also when considering a more general class of models in which attentional modulation of inputs to the excitatory and inhibitory units can be different (Eqs (16) and (17)). As long as attentional modulation α_e of excitatory input is equal or larger than the corresponding modulation of inhibitory input α_i , the expression in Eq (16) yields a positive result for $a_e^{post} > a_e^{pre}$ (steeper rise), and a negative result (steeper decay) otherwise. Conversely, the expression in Eq (17) can be positive or negative for $a_e^{post} > a_e^{pre}$, depending on the model parameters and pre- and post-change activation. Assuming a different input I_i to the inhibitory population is mathematically equivalent to rescaling the gain factor m_i and leads also to the same qualitative results as long as the corresponding input to the excitatory population I_e is equal or larger.

Experimental investigation of model predictions

Because attention is generally found to facilitate the response of visual neurons to the initial stimulus, any decrease in the population firing induced by a corresponding change of the behaviorally relevant stimulus would be antagonized by the opposite effect of attention, in terms of absolute firing. For this problem, the model's prediction of generally faster rise times and potentially more pronounced relative peak firing rate changes as a result of attention offers a particularly attractive solution for effective detection of stimulus changes. Changes in rise time (and, under appropriate conditions, relative peak firing rate changes), provide a mechanism independent of absolute firing to transmit attentionally selected information to downstream areas of the visuomotor pathway.

To test the model predictions, we recorded neuronal responses from motion-sensitive area MT of two macaques (M3: $N = 45$ units, M4: $N = 25$ units). Monkeys performed a speed-change detection task requiring them to either attend towards or away from the recorded unit's RF and to detect increments or decrements of the target's speed (Fig 5A). Peri-stimulus time histograms (PSTHs) aligned to the speed change of the stimulus displayed higher pre-change firing rates when the stimulus inside the RF was attended, and strong, transient increases and decreases of the firing rate in response to increments (accelerations) and decrements (decelerations) of motion speed, respectively (Fig 5B). Interestingly, neither pre-change firing rates (Wilcoxon signed rank tests, [attend-in (M3 M4), attend-out (M3 M4)]: $P = [1\ 0.77\ 0.53\ 0.36]$, $N = [43\ 21\ 43\ 21]$), nor attentional modulation before the change (Wilcoxon signed rank tests on attentional modulation index $AI = (A_e^{pre,A} - A_e^{pre,N}) / (A_e^{pre,A} + A_e^{pre,N})$, M3, M4: $P = [0.1\ 0.7]$, $N = [43\ 21]$) differed between blocks of speed increments and decrements. This indicates that both activity and attentional modulation during the pre-change period were independent from the sign of the speed change to be detected, despite the block-wise design of the experiment. Spike counts for all attentional conditions and all 25 ms time intervals between 400ms before and 200ms after a speed change exhibited a variance close to their mean (Fig 5B, insets), compatible with the statistical properties of a Poisson process (used below for assessing significance of spike count differences).

To analyze these transients according to model predictions, we assessed first, response slopes and second, relative firing rate changes during the transient time period of 50 to 200 ms following the speed change. First, for analyzing slopes, we calculated excess cumulative spike counts representing the number of spikes over- or undershooting the mean firing rate before the speed change as a proxy for the initial slope. Because rates increase or decrease almost linearly for the first 40 ms following the population transient onset, a larger (smaller) cumulative count is equivalent to a steeper positive (negative) slope. In contrast to estimating slopes

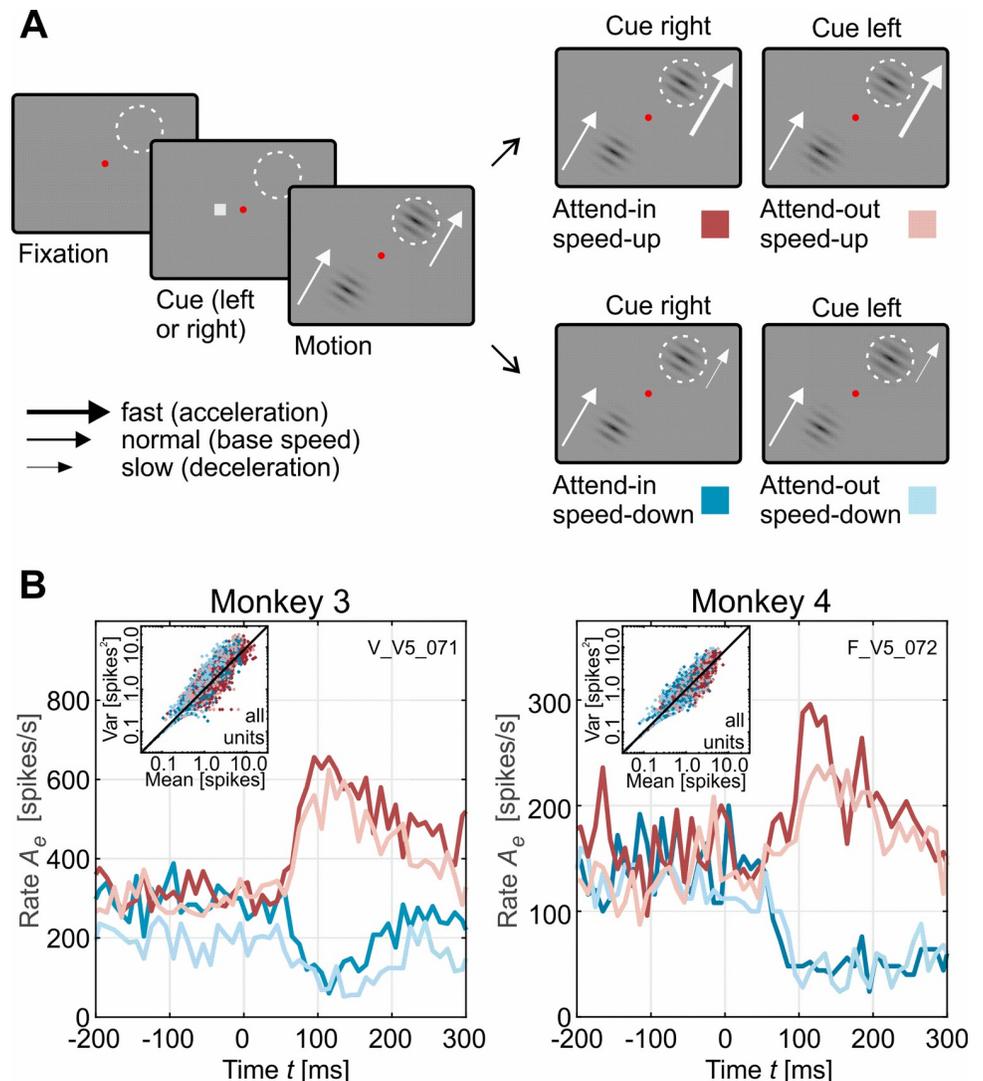


Fig 5. Experimental speed change detection paradigm and example PSTHs. (A) Monkeys were detecting positive and negative speed changes (presented block-wise) at pre-cued locations. Each trial started with a cue indicating the hemifield of the rewarded stimulus (small gray box, left or right of the central, red fixation spot). After monkeys properly fixated and pressed a lever, two moving gratings appeared, one of which was placed inside the RF of the recorded neuron (dashed white circle), while the other was placed in the opposite hemifield, mirrored across the fixation spot. Following a pseudo-randomized delay of 0.66 to 5.5 s, the RF-stimulus rapidly increased (top row) or decreased speed (bottom row). Any speed change of the stimulus in the uncued hemifield had to be ignored. Keeping fixation and releasing the lever within 750 ms after the speed change was rewarded with some drops of water or diluted grape juice. Depending on the cued stimulus location, the RF stimulus was either attended or non-attended, giving rise to four experimental conditions (right plots, color-coded). (B) PSTHs for two example multi-unit sites (one from each monkey), illustrating rapid firing rate adjustments in response to speed changes. Note that because negative speed changes result in a decrease of the firing rate, the stimulus-induced modulation of the firing rate is opposite to the attention-induced modulation before the change. Insets show spike counts and variances assessed in time intervals of 25 ms from -400 ms to +200 ms relative to stimulus change for all units. Color coding of the different attentional conditions is indicated in panel (A).

<https://doi.org/10.1371/journal.pcbi.1009595.g005>

directly from the PSTHs, the cumulative measure has two advantages. First, because accumulation is based on integration, noise does not become amplified but attenuated. Second, accumulation does not need smoothing of data to more reliably calculate specific response parameters, which is problematic if transients are faster than the width of the PSTH filter kernel used.

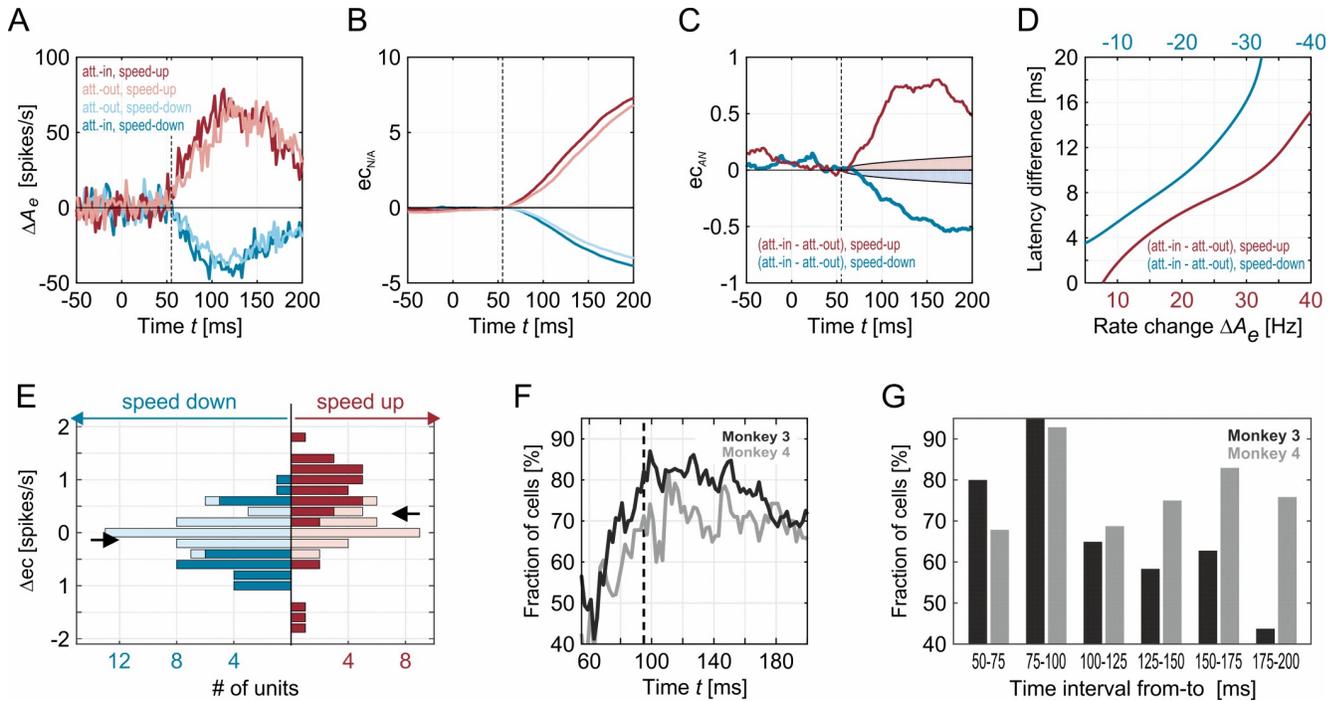


Fig 6. Quantification of attention-induced rate changes. (A) Population average of speed change-induced rate modulations for the four attentional conditions. Data taken from M3. (B) Corresponding excess cumulative spike counts relative to the pre-change firing rate. (C) Attention-induced differences in excess cumulative spike count, separately for accelerations (red) and decelerations (blue). Differences are significantly larger than 0 ($p < 0.01$) outside shaded regions. (D) Time advantage (i.e. latency difference) induced by attention for different read-out thresholds (i.e. absolute rate changes ΔA_e), as measured in the experimental data shown in (A–C). (E) Histogram of attention-induced differences in excess cumulative spike count over individual sites, separately for acceleration and deceleration conditions. Dark colors indicate samples with values being significantly different from zero. Samples were taken at 95 ms after stimulus change. (F) Percentage of significantly modulated individual units for which cumulative counts were significantly larger with attention in the speed-up condition or significantly smaller with attention in the speed-down condition. Black line, M3; gray line, M4. Dashed line indicates the time for which the samples for the histogram in panel (E) were taken. (G) Percentage of significantly modulated units for which spike count differences between activities before and after the stimulus change were significantly larger in the speed-up condition or significantly lower in the speed-down condition, separately for successive time intervals of 25 ms length.

<https://doi.org/10.1371/journal.pcbi.1009595.g006>

Consistent with the prediction of the model, the excess cumulative spike count was found to rise more steeply for positive transients, and to decay more rapidly for negative transients when attention was directed to the stimulus (Fig 6A and 6B). Onset of the population transient was around 55 ms after the stimulus change, which is well in the range of typical MT response latencies [23,24]. Attention-dependent differences in excess cumulative spike counts become significantly different from 0 already shortly after response onset (Fig 6C). The steepening of the PSTH’s slope by attention causes the neurons to reach a certain rate increase or decrease earlier than without attention. To quantify the induced time advantage, we computed the average smoothed PSTH (Gaussian kernel, RMS width 10ms) for each condition and monkey, and extracted the time difference in reaching a certain rate increase or decrease with and without attention. Attention provided a time advantage of up to 10 ms and more for both accelerations and decelerations, which corresponds to the typical attention-dependent latency reduction in area MT and goes along with a reaction time decrease of 50 ms and more (cf. Table 1 in Ref. [9]). Significantly different cumulative spike counts following speed increments were found for 60% of all recorded units during the onset of the transient (Fig 6E), and of those, 85% confirmed the predictions of the model (Fig 6F). Likewise, for speed decrements, 46% of all units displayed significantly different spike counts (Fig 6E), and 76% of those were confirming the model’s prediction of steeper slopes (Fig 6F). This result was evident for both monkeys,

with 84% (M3) and 72% (M4) of all significantly modulated units being in accordance with the model prediction. Interestingly, units with strong rate increases in the acceleration condition also showed strong decreases in the deceleration condition. Performing a linear regression analysis on the excess spike count for deceleration vs. acceleration revealed a consistent anticorrelation with slope -0.20 ± 0.02 STD and offset -0.11 ± 0.02 SD for monkey M3 and slope -0.13 ± 0.04 SD and offset 0.00 ± 0.02 SD for monkey M4 (variability assessed by bootstrap using one-leave-out method).

Second, for analyzing the transient's peak in response to the stimulus change, the firing rate was calculated in consecutive bins of 25 ms width. For both speed increments and decrements, up to more than 90% of units were found to have significantly higher and lower rates, respectively, in the attend-in condition during the time intervals of 50 to 100 ms post-change (Fig 6G). Yet, because this ratio decreased rapidly for later bins of the transient, consistent attentional modulation was limited to a brief period immediately following the stimulus change.

Taken together, the experimental data confirmed the model predictions for both the effect of attention onto the slope of positive and negative transients and, for the initial part of the transient, onto the modulation of relative peak responses for time constant ratios of $\tau_e/\tau_i \ll 1$, as found in the response onset fits (Fig 2E). As a novel physiological result, they provide evidence that attention modulates the same features of a negative transient than it does for positive transients, suggesting that processing of visual information, and its perception, can rely on information contained in reductions of firing rates.

Discussion

The ability to detect rapid changes in complex, ever-changing environments is fundamental for both animal and human behavior. Neuronal responses to fast stimulus transitions usually come as brief episodes of increased or decreased neuronal activity, followed by a steady-state level of lower absolute amplitude. Pronounced transient changes in neuronal activation were observed in the brain of many different species, spanning the range from invertebrates to primates [25–28], suggesting that they represent a basic principle in neuronal network dynamics. We here show that such a canonical computation can be realized by a very simple circuitry, essentially built of only one excitatory and one inhibitory unit, in which the excitatory unit's output time course is normalized through divisive inhibition. The circuitry can be expressed by a set of equations with only three free parameters, obtained by fitting the model to each neuron's onset response (to cover the individual unit's kinetics). Note that, although we assumed a log-Gaussian speed tuning during the steady state of the neurons to calculate sustained firing rates, the prediction of transients did not depend on any assumption about the underlying speed tuning profile (which may be influenced by higher-order parameters as e.g. stimulus history and rapid response adaptation [29,30]). The simplicity of the model allowed for a comprehensive mathematical analysis of neuronal response dynamics and the reproduction and prediction of physiological transients in response to a wide range of stimulus transitions, including the interesting case of attentional modulation of rate-decreasing events.

Non-stationary normalization by divisive inhibition

A key element of the model is normalization of the circuitry's output by divisive inhibition. Normalization by divisive (shunting) inhibition was initially suggested as a means to explain nonlinearities in the response of neurons in visual cortex [17]. It consists of dividing the response of a given neuron, or group of neurons, by the average response of a pool of normalizing units, either within the same cortical area or between areas [31–35]. Since its introduction, the concept was successfully used to explain neural response characteristics in a range of

different contexts and neural systems, both with and without attention (for overview: [18,36]). Most work, however, has focused on static divisive normalization to describe modulations of sustained neuronal responses, while the circuitry we here introduce explicitly addresses the temporal response dynamics. Static normalization in our model emerges as the fixed point of the activation dynamics for a constant input. Static models may be capable to reproduce MT responses to dynamic stimuli to some extent [37], but they have limits to capture fast neural responses and they are inappropriate to explain non-monotonic transient responses to sudden input changes.

Temporal low-pass filtering combined with delayed inhibition was previously suggested to circumvent these limitations and allowed modelling the time course of MT responses during pursuit eye movements [38]. While low-pass filtering converts an instant input change to an exponentially saturating input current, divisive inhibition kicks in later than excitation and thus allowed us to successfully predict change-transients as observed experimentally. With inhibition following the dynamics of excitation instantaneously [39], it is not possible to reproduce the time course and the peak (or trough) of a change transient. The relaxation differential equations in our model thus have the same effect as in [38]—low-pass filtering the input drive, and delaying inhibition by assuming a larger time constant for the rise of the divisive term. However, the earlier model has a large number of free parameters and multiple functional modules for allowing detailed fits of neural responses to pursuit eye movements of different velocity, while the model we here introduce has its focus on structural simplicity. This property enables thorough mathematical analysis for a large variety of stimulus conditions, with and without the effects induced by attention, and allowed to predict neuronal response dynamics under so far untested experimental conditions.

Dynamic divisive normalization has recently been applied also by other models. A recent study, for example, implemented dynamic divisive normalization in a model to investigate the response to competing visual stimuli under different forms of attention [40], but focused only on stationary states. Other dynamical models incorporate divisive interactions via recurrent inputs. For example, recurrent divisive inhibition was used to study effects of feature-based attention [41], and for creating an abstract model of a cortical column that accounts for signal filtering and nonlinear gain control in combined feedforward and feedback pathways [42]. In decision-related neuronal circuits recurrent divisive inhibition was shown to generate both positive and negative transients [20]. In general, recurrent inhibition allows for very 'rich' model dynamics and, through reverberating activity, is also capable to generate oscillations [19,20,42], whose origin can be understood through linear stability analysis.

Yet, mathematical analysis of the dynamical properties of recurrent models becomes quickly intractable. In contrast, the feedforward model we here introduce is allowing a thorough mathematical analysis, while it still explains the complex firing rate dynamics of our experimental data. The anatomical origin, however, of divisive normalization may be rooted in a recurrent circuit. A possible test of this, and a continuation of our work, is to implement divisive normalization in a recurrent, 'ring-model'-like network, which consists of identical elementary circuits to represent neuronal columns with different preferred motion directions as in [31,36,43], yet including temporal dynamics for inhibitory feedback [19]. In conjunction with appropriate experimental data, this would allow 1) to test whether the effects we here describe are compatible with a recurrent origin of divisive inhibition, 2) how attentional modulation in this network would affect the response of a heterogeneously tuned population, and 3) further illuminate the physiological mechanism underlying dynamic divisive inhibition. Note, however, with regard to the latter, that it was recently put into question whether divisive inhibition constitutes a physiological mechanism on its own, or whether it may simply emerge as a network effect [44], possibly in conjunction with noise [45].

Relation between transients and information processing, perception, and behavior

Biologically relevant signals usually occur on very short time scales, and the brains of both vertebrates and invertebrates generate rapid visual percepts, decisions, and motor behaviors. Flies, for instance, track and chase other flies with response times as short as 30 ms [46], carnivorous vertebrates possess extremely fast sensorimotor programs for visually guided pursuit predation [47], and primates, both human and non-human, categorize objects and perform appropriate motor responses within tens of milliseconds [48–50]. These findings imply that, while different behaviors in different species may involve different neuronal substrates and mechanisms, the brain strongly relies on fast neuronal codes to account for the strong temporal variability of sensory input during eye movements, self- and object-motion. Transient firing rate changes of small groups of neurons in response to sudden changes in sensory input are likely part of this code. Accordingly, such transients not only participate in detection of objects and events but carry detailed information about stimulus properties. In monkey temporal cortex, transients were shown to exhibit specificity for different head views within 25 ms following onset of the population response, and to contain more information than later epochs of the response [51,52]. A corresponding pattern was found in primary visual cortex V1, reaching a peak for detectability and discriminability of oriented gratings within 150 ms of the onset response, and in area MT, where most information on motion direction is available within the first 100 to 200 ms after stimulus onset [3,4,53]. This higher information content of onset transients is likely due to a larger gain and smaller variance as compared to steady-state activation levels during continuous stimulation [53,54]. Accordingly, because brief episodes of coherent motion or rapid speed changes were found to induce firing rate changes significantly correlating with behavioral choices, transients were linked to perceptual judgments [9,55–59]. The results of the current study show that the simple circuitry we used to implement rapid firing changes in response to input changes is fully reproducing the experimentally observed MT responses, including the over- and undershooting during change transients in comparison to firing rate changes expected from steady-state tuning properties [7]. Furthermore, although the excess spike counts in Fig 6 might seem small, they will have a substantial effect in a population of neurons: assuming only 10 to 100 presynaptic neurons projecting to a target unit, about 100ms after a speed change this unit would have received 10 to 100 spikes more than without attention—a number which can readily make the difference for the target unit to either remain silent or fire an action potential. Assuming a fixed threshold for excess spikes, such as in standard race models, this would allow to detect changes 10 to 20ms earlier, given the size of effects in our data (graphs not shown). Because we recently showed that thresholding of transients permitted the read-out of information in full accordance with human behavioral performance both for rate increments and decrements [6], our results provide a mechanistic explanation for computations within sensory cortex underlying the perceptual process of change detection and discrimination, basically realizable by non-stationary divisive inhibition within a simple cortical circuitry.

Modulation of change transients by selective attention

Due to the close relation between transient firing rate patterns and perceptual judgments on the one hand, and the influence of selective attention on neuronal processing and behavioral performance on the other, transients are likely targets for attentional modulation. Recent monkey neurophysiological studies reported attention-dependent modulation of amplitudes, latencies, and gamma coherence during stimulus onset or stimulus change responses in various visual areas [9,21,60–63], and a close relation between reaction times to attended speed increments and the latency of the change transient [9]. All of these results, however, were obtained

by investigating stimulus events eliciting an increase in firing rate. If specific parameters of transient firing rate changes indeed underlie perceptual performance, the question arises how attention influences a population of neurons for stimuli inducing a decrease in mean activation. The mathematical analysis of our model dynamics predicted—not without surprise—that negative transients would basically be modulated by attention in the same way as positive ones, with the slope and the relative peak height being the response features to allow for consistent attentional modulation independent of the pre-change activation level of the neuron, and independent of the sign of the transient. The experimental confirmation of this prediction provides new physiological insights for understanding change detection and its modulation by attention. As an important result, they indicate a relevant constraint of task-dependent modulation. Because attention was implemented as a positive gain to the input of the circuitry in the model, firing rates during the pre-change epoch were always larger with attention than without, regardless of the type of change occurring later. These new physiological results show that, albeit speed increments and decrements were presented block-wise and allowed the animals to make correct predictions on the sign of the upcoming change, pre-change firing rates were not different between acceleration and deceleration trials and attention consistently increased neuronal responses during the pre-change epoch by about the same factor. These results strongly suggest that attention-dependent modulation in early visual cortex is generally associated with a positive gain of neuronal responses, as opposed to a mechanism modulating responses in the same direction as the sensory event. This conclusion is consistent with recent results on both spatial [9] and feature-directed [10] attention, which were found to exert very similar effects during both the pre-change epoch and transient firing. Particularly, it was recently shown that feature-directed attention exerts a positive, tuning-independent gain even in the absence of a visual stimulus [10]. This tuning-independent effect likely constitutes the main source of top-down modulation during feature-directed attention, while tuning dependent modulation, as proposed by the feature-similarity hypothesis [64,65], is likely evolving on top of this and seems to be limited to early visual areas [66]. The new results on consistent pre-change attentional modulation in acceleration and deceleration trials are fully in line with this.

Moreover, the model predicted steeper response slopes with attention for both stimulus changes inducing an increase and a decrease in neuronal firing. This prediction was a direct consequence of the model's inherent dynamics, since apart from input gain no other parameter of the model was changed to implement attention. In line with this prediction, the physiological experiments show a significant influence of attention on both the rise and the decay time of the change transient, being steeper with attention than without, i.e. attentional modulation was independent of whether the stimuli induced an increase or a decrease in the firing of neurons. Based on the model's temporal dynamics, a mechanistic explanation of this effect is that the stronger drive of both the excitatory and inhibitory unit allows a faster effect of divisive normalization with attention. Because normalization is acting in the direction of the input change, it is affecting the slope of both positive and negative transients likewise. Both the computational and the physiological data suggests that not only rapid positive changes in firing rate, but also rapid negative changes provide important information to downstream areas that are used for subsequent visual processing.

Methods

Ethics statement

Housing of animals, experimental and surgical procedures were all in accordance with the *Directive 2010/63* issued by the European Commission and with the *Regulation for the Welfare of Experimental Animals* issued by the Federal Government of Germany, and were approved by the local authorities (Der Senator für Gesundheit, Freie Hansestadt Bremen, Az. 522-27-11/02-00).

Subjects

MT data used to develop and test the model were partly recorded in the context of previously published studies [7,10]. Additional data to test model predictions were recorded using non-human primate standard behavioral and neurophysiological procedures. Data were acquired from two male macaque monkeys, six and eight years old, both kept pair-wise with another male. Initial surgical interventions were performed under anesthesia and strictly sterile conditions. Anesthesia was initiated by ketamine/medetomidine and was maintained by propofol, supplemented with isoflurane. Remifentanyl and Carprofen were given for peri- and post-operative analgesia, respectively. Monkeys were given sufficient time to recover before behavioral training in the laboratory. Further details on surgical procedures are given elsewhere [10,67]. Both monkeys were familiar with laboratory standard procedures and general task conditions. They received additional training on detecting negative speed changes during the course of several weeks using water and fruit juice as reinforcer for performing the task. On non-training and nonrecording days, monkeys received fruits and liquid in their home compartments, consisting of large indoor rooms with daily access to equally sized outdoor compartments. All compartments were enriched by a manifold of monkey toys, puzzles, and climbing opportunities. Health and well-being were checked by daily monitoring and regular visits by veterinarians, and body weight was checked several times a week.

Electrophysiological data

Recordings were performed using tungsten microelectrodes (2–5 M Ω , 125 mm shank diameter; Frederic Haer, Bowdoin, ME) and standard electrophysiological equipment. The pre-amplified signal was filtered between 0.7 and 5 KHz and sampled with a frequency of 25 KHz. Spikes were detected online by thresholding the signal. All spike data were then subjected to offline semiautomatic spike sorting using Klustakwik [68], followed by manual adjustment of spike clusters using a custom-made algorithm for spike form and spike parameter illustration [69]. Visual stimulation, control, and documentation of behavioral data was performed using custom-made Matlab scripts and in-house software. Eye movements were controlled by a custom-made video-oculography system with 0.2-degree resolution.

Visual stimulation and behavioral paradigm

Monkeys were tested in a behavioral task requiring detection of a rapid change in the speed of a moving stimulus, either an acceleration (speed change by a factor of ~ 2), or a deceleration (speed change by a factor of ~ 0.5), presented block-wise. The basic task design was the same as in previous studies^{9,10}. Each trial started with appearance of a small red fixation spot (0.14-degree side length) at the center of the screen (22-inch cathode ray tube monitor, resolution 1.280 x 1.024-pixel, 100 Hz refresh rate). Monkeys initiated the trial by gazing at the fixation point, pressing a lever, and keeping it hold. Following a delay of 250 ms, a spatial cue appeared for 700 ms to indicate the location of the upcoming target stimulus, followed by another delay of 200 ms and subsequent appearance of two static Gabor stimuli (sine wave spatial frequency: 2 cycles/degree, Gaussian envelope: $\sigma = 0.75$ degree at half height), one centered above the RF of the recorded neuron and the other one mirrored across the fixation spot. Mean Gabor luminance was identical to background luminance (10 cd/m²). Gabors started to intrinsically move (speed: 2.17 degree/sec) 200 ms after onset, with motion direction adjusted to the preferred direction of the recorded neuron, as described elsewhere [9]. In about 40%–50% of the trials, the uncued stimulus changed speed before the target stimulus, which had to be ignored by the monkeys. Following the speed change of the target, monkeys had to keep fixation for another 300 ms (to avoid contamination of the neuronal post-change response by eye

movements) and to release the lever within a response window of 150–750 ms. Speed changes occurred within 0.66 and 5.5 sec after motion onset. Subsequent trials were separated by an intertrial interval of 3–4 sec. Releasing the lever outside the response window and eye movements of more than one degree from the fixation point caused immediate termination of a trial. Monkeys were rewarded with a few drops of water or diluted fruit juice for each correctly performed trial.

Model steady-state activation and response to step functions

The model consists of two units driven by external input I , with one unit providing divisive inhibition on the other unit (Fig 1B). Their dynamics are described by the differential Eqs (1, 2) and threshold-linear gain functions (3, 4), with the steady-state activation resulting from a constant, suprathreshold input I_0 given by

$$A_e^{steady} = A_e(t \rightarrow \infty) = m_e \left(\frac{I_0}{A_i^{steady} + \sigma} - \theta_e \right), \tag{9}$$

$$A_i^{steady} = A_i(t \rightarrow \infty) = m_i(I_0 - \theta_i) \tag{10}$$

Inserting (10) into (9) provides Eq (5). Assuming zero thresholds $\theta_e = \theta_i = 0$ and a supra-threshold input $I(t) > 0$ allows to replace (3)–(4) by linear gain functions, simplifying further analysis. Instantaneous stimulus changes are considered as step functions providing a change in constant external input from I^{pre} to I^{post} at $t = t_{change}$. Assuming that the model is in its steady state for $t < t_{change}$, Eq (2) can be explicitly solved and Eq (1) for post-change activation ($t \geq t_{change}$) can be written as:

$$\tau_e \frac{dA_e}{dt} = -A_e + m_e \left(\frac{I^{post}}{m_i(I^{post} + (I^{pre} - I^{post})\exp(-(t - t_{change})/\tau_i)) + \sigma} \right). \tag{11}$$

Using Eq (5), I^{pre} and I^{post} , which are unknown in an experiment, can be expressed as functions of their corresponding steady-state (sustained) activations A_e^{pre} and A_e^{post} via Eq (6).

Model fit to physiological data

88 recordings from M1 and 53 recordings from M2 (single- and multi-units) were used to fit the model to stimulus onset transients. Transients were caused by the appearance of a grating inside a neuron’s RF, moving into its preferred direction while monkeys performed a simple fixation paradigm (cf. Ref. [7] for experimental details). While the model responds immediately to a change in its input at time t_{change} , the physiological change response is delayed by the processing time between retina and area MT. To account for this, response onset delays $\Delta\tau$ relative to stimulus onset were estimated by visual inspection of the PSTHs binned at different temporal resolutions, individually for each unit (averages: M1: 31 ms +/- 12 ms SD, M2: 32 +/- 10 ms). Sustained responses A_e^{pre} and A_e^{post} before and after stimulus change were determined by computing the average spike rate over all trials in the intervals [-100 ms, 0 ms] and [200 ms, 500 ms], respectively (time denoted relative to stimulus onset). A_e^{pre} and A_e^{post} allow to numerically solve Eq (6) and compare it directly to the delay-compensated, trial-averaged physiological response by computing the Chi-square function, which relates the quadratic error between observation and prediction to the standard error of the observation via $\chi^2/N_t = \langle \|A_e(t) - A_e^{exp}(t - \Delta\tau)\|_2 / \sigma^{exp}(t - \Delta\tau) \rangle$, averaged over a 200ms interval after response onset. Values of χ^2/N_t around 1 are considered good fits because the average deviation between model and observed data is then comparable to the variability of the data. Experimental A_e^{exp} was sampled within 5 ms time bins and the model’s

response was downsampled to the same temporal resolution for comparison. Neural units were excluded from the fitting procedure if their spike count summed over all available trials was less than 100 (i.e., an average of only 2.5 spikes per time bin, due to a poor response to the applied speed, which was purposefully not adjusted to the neuron’s tuning properties), which provides insufficient data to yield a well-shaped PSTH. Applying this criterium left 48 and 44 units from monkey M1 and M2, respectively, for fitting. Parameters τ_e , τ_i , and A_e^{max} for explaining physiological dynamics were determined by an iterative grid search for minimizing the quadratic error. Because A_e^{max} must be at least the value of A_e^{post} , and values higher than $3 A_e^{post}$ were never reached by the optimization procedure, initial grid search ranges for A_e^{max} were set to $[1.03 A_e^{post}, 3 A_e^{post}]$. Ranges for τ_e and τ_i were set to $[1 \text{ ms}, 100 \text{ ms}]$ and $[1 \text{ ms}, 500 \text{ ms}]$, respectively, with the upper limits being well above the observed rise/decay times of the PSTHs. The resolution of the search grid was 40 bins for A_e^{max} , and 15 bins for the two time constants. Four iterations with subsequently refined grids resulted in a precision of 0.01% for the parameter estimates within the initially chosen range. For determining the goodness-of-fit we compared the empirical distribution of χ^2/N_t with surrogate spike data drawn from model responses generated with parameters sampled from the distributions of the fitted parameters. This surrogate data yielded distributions perfectly shadowing the empirical distributions, and thus all fits of the selected units were considered as appropriate, valid parameter estimates.

Calculation of transient and sustained activation changes

Under the condition $\tau_e \ll \tau_i$ (separation of time scales), the peak response during the transient can be approximated by:

$$A_e^{peak} \approx A_e^{post} \frac{A_e^{max} - A_e^{pre}}{A_e^{max} - A_e^{post}}. \tag{12}$$

Under assumption of a log-Gaussian velocity tuning [14,15], sustained activation can be expressed in terms of the stimulus velocity v :

$$A_e^{steady}(v) = A_0 + A_e^{pref} \exp\left(-\frac{(\log(v_{pref}) - \log(v))^2}{2\sigma_v^2}\right), \tag{13}$$

with spontaneous activity A_0 , tuning width σ_v , and maximum response amplitude A_e^{pref} for preferred speed v_{pref} . Computing A_e^{steady} for stimulus velocities v^{pre} and v^{post} before and after the speed change, respectively, provides $A_e^{pre} = A_e^{steady}(v^{pre})$ and $A_e^{post} = A_e^{steady}(v^{post})$ for Eq (14), to obtain analytical expressions for A_e^{peak}/A_e^{pre} and A_e^{post}/A_e^{peak} for arbitrary acceleration/deceleration ratios v^{pre}/v^{post} (Fig 3). For evaluating A_e^{peak}/A_e^{pre} and A_e^{post}/A_e^{peak} in comparison to experimental data (Fig 3), we chose $A_0 = 0$, a tuning half-width of 2.5 octaves ($\sigma_v = \log(2^{2.5}) \approx 1.73$) and $A_e^{max} = 1.4A_e^{pref}$.

Modulation by attention

Attention is modeled by multiplicative modulation of input I by a factor $\alpha > 1$, $I \rightarrow \alpha I$ (Fig 4A). Expressing absolute activation A_e relative to A_e^{max} , $a_e := A_e/A_e^{max}$ provides scaled pre- and post-change sustained activities $a_e^{pre} = A_e^{pre}/A_e^{max}$ and $a_e^{post} = A_e^{post}/A_e^{max}$ ranging in the interval $[0, 1]$. The initial slope F_α^{rise} of the transient response as a function of a_e^{pre} and a_e^{post} in dependence on α yields

$$F_\alpha^{rise}(a_e^{pre}, a_e^{post}) := \left. \frac{da_e}{dt} \right|_{t=0} = \frac{1}{\tau_e} \frac{a_e^{post} - a_e^{pre}}{1 - a_e^{post}} \frac{\alpha}{a_e^{pre}(\alpha - 1) + 1}, \tag{14}$$

and the change in sustained activation level F_{α}^{sus} relative to the pre-change level becomes

$$F_{\alpha}^{sus}(a_e^{pre}, a_e^{post}) := \frac{\alpha}{\alpha - 1 + (a_e^{post})^{-1}} - \frac{\alpha}{\alpha - 1 + (a_e^{pre})^{-1}}. \tag{15}$$

For assessing the attention-induced modulation in these quantities (Fig 4B), we computed $\Delta F_{\alpha}^{rise} = F_{\alpha}^{rise} - F_1^{rise}$ and $\Delta F_{\alpha}^{sus} = F_{\alpha}^{sus} - F_1^{sus}$. Peak activation F_{α}^{peak} during transient responses must be computed numerically, thus $\Delta F_{\alpha}^{peak} = F_{\alpha}^{peak} - F_1^{peak}$ was obtained by solving Eq (6) explicitly in dependence on the chosen time constants τ_e and τ_i .

As a control, and for the purpose of generalization, we extended the model to have different attentional modulation factors α_e and α_i as input to the excitatory and inhibitory units, such that the attention-induced modulations become:

$$\Delta F_{\alpha}^{rise}(a_e^{pre}, a_e^{post}) := \frac{(a_e^{post} - a_e^{pre})(a_e^{pre}(\alpha_i - 1) + (1 - \alpha_e))}{\tau_e(a_e^{post} - 1)(a_e^{pre}(\alpha_i - 1) + 1)} \tag{16}$$

$$\Delta F_{\alpha}^{sus}(a_e^{pre}, a_e^{post}) := (a_e^{post} - a_e^{pre}) \left(\frac{\alpha_e}{(a_e^{post}(\alpha_i - 1) + 1)(a_e^{pre}(\alpha_i - 1) + 1)} - 1 \right) \tag{17}$$

Calculation and comparison of physiological response parameters

In total, $N = 45$ sites were recorded in M3 and $N = 25$ sites in M4. For inclusion to data analysis, we required the speed-up condition to be associated with a significant firing rate increase in the attend-out condition (assessed in the time interval 140 to 160 ms after stimulus change, one-tailed test on Poissonian distribution around mean firing rate before stimulus change, $p < 0.05$), which was given for $N_{up} = 36$ sites in M3 and $N_{up} = 19$ sites in M4. Similarly, the speed-down condition was required to be associated with a significant firing rate decrease in the attend-out condition, which was fulfilled for $N_{down} = 42$ sites in M3 and $N_{down} = 21$ sites in M4. Rise times and relative spike counts of the experimentally observed transients (Fig 5) were calculated to compare physiological data against model predictions for ΔF^{rise} and ΔF^{peak} . First, rise times of physiological transients were assessed by determining excess cumulative spike counts following the stimulus change, defined as the number of spikes exceeding the spike count of a neuron continuing to fire with its observed pre-change rate. By visual inspection, we estimated an average response delay of $t_{change} = 55$ ms for both monkeys. For estimating pre-change activity, we computed the summed firing rate F^{pre} over all trials of a given attention condition in the time window $[-400 \text{ ms}, t_{change}]$. Excess cumulative spike count $ec(t)$ was then defined for $t > t_{change}$ by

$$ec(t) = -F^{pre}(t - t_{change}) + \int_{t_{change}}^t \delta(t' - t_k) dt' \tag{18}$$

where t_k denote the times of K spikes $k = 1, \dots, K$, considering all trials of the respective condition. For comparing the initial slopes of two attention conditions N and A , we first computed their cumulative excess responses $ec_N(t)$ and $ec_A(t)$. We then evaluated the difference $\Delta ec_{AN}(t) = ec_A(t) - ec_N(t)$ and tested statistically whether $\Delta ec_{AN}(t)$ was significantly deviating from zero. The transient of condition A was considered to rise significantly faster (or slower) than the transient of condition N , if $\Delta ec_{AN}(t)$ was above (or below) a time-dependent threshold. Thresholds were set to $\pm z \sqrt{(F_A^{pre} + F_N^{pre})(t - t_{change})}$, where $z = 2.32$ was chosen to yield a significance level of $p < 0.01$.

Second, spike counts for individual attention conditions and during different periods of the transient were calculated as the difference Δac between the absolute spike count before and after the speed change, $\Delta ac(itv) = ac^{post}(itv) - ac^{pre}$, where itv denotes intervals of 25 ms length, taken between 50 ms and 200 ms following the speed change for the transient. The steady-state response before the change was estimated during the period from -400 ms to t_{onset} . The difference between two relative spike counts was considered significantly different from zero when it exceeded $\sqrt{\text{var}_A^{pre} + \text{var}_A^{post} + \text{var}_N^{pre} + \text{var}_N^{post}}$, where var designates the spike count variance in the corresponding attentional condition pre- or post-stimulus change.

Supporting information

S1 Data Code Repository. The archive ‘S1_Data_Code_Repository.zip’ contains all relevant data sets and program code for generating the results reported in the manuscript. Running the code requires Matlab R2020a (The MathWorks, Inc.) and Python (Version 3.6). To perform all data analyses and fits, execute script ‘PLOS_run_all.m’ from the main directory. See the comments in the scripts named ‘PLOS_run.m’ in the subdirectories for further information on the result plots and files created by running the code.
(ZIP)

Acknowledgments

The authors thank Franziska Müller, Peter Bujotzek, Katrin Thoss and Ramazani Hakizimani for technical and training support, Andreas Traschütz and Bastian Schledde for recording physiological data used to develop and test the model, and Bonne Habekost for helping to reorganize physiological data.

Author Contributions

Conceptualization: Udo A. Ernst, Lisa Bohnenkamp, Detlef Wegener.

Data curation: Udo A. Ernst, Fingal Orlando Galashan, Detlef Wegener.

Formal analysis: Udo A. Ernst, Xiao Chen, Lisa Bohnenkamp.

Funding acquisition: Udo A. Ernst, Detlef Wegener.

Investigation: Udo A. Ernst, Xiao Chen, Lisa Bohnenkamp, Fingal Orlando Galashan.

Methodology: Udo A. Ernst, Detlef Wegener.

Project administration: Udo A. Ernst, Detlef Wegener.

Resources: Udo A. Ernst, Detlef Wegener.

Software: Udo A. Ernst, Xiao Chen, Lisa Bohnenkamp, Fingal Orlando Galashan, Detlef Wegener.

Supervision: Udo A. Ernst, Detlef Wegener.

Validation: Udo A. Ernst, Detlef Wegener.

Visualization: Udo A. Ernst, Detlef Wegener.

Writing – original draft: Udo A. Ernst, Detlef Wegener.

Writing – review & editing: Udo A. Ernst, Xiao Chen, Lisa Bohnenkamp, Fingal Orlando Galashan, Detlef Wegener.

References

1. Gerstner W. Population dynamics of spiking neurons: Fast transients, asynchronous states, and locking. *Neural Comput.* 2000; 12: 43–89. <https://doi.org/10.1162/089976600300015899> PMID: 10636933
2. Tsodyks M, Sejnowski TJ. Rapid state switching in balanced cortical network models. *Network: Comput Neural Syst.* 1995; 6: 111–124.
3. Osborne LC, Bialek W, Lisberger SG. Time course of information about motion direction in visual area MT of macaque monkeys. *J Neurosci.* 2004; 24: 3210–3222. <https://doi.org/10.1523/JNEUROSCI.5305-03.2004> PMID: 15056700
4. Buracas GT, Zador AM, DeWeese MR, Albright TD. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron.* 1998; 20: 959–969. [https://doi.org/10.1016/s0896-6273\(00\)80477-8](https://doi.org/10.1016/s0896-6273(00)80477-8) PMID: 9620700
5. Motter BC. Modulation of transient and sustained response components of V4 neurons by temporal crowding in flashed stimulus sequences. *J Neurosci.* 2006; 26: 9683–9694. <https://doi.org/10.1523/JNEUROSCI.5495-05.2006> PMID: 16988039
6. Traschütz A, Zinke W, Wegener D. Speed change detection in foveal and peripheral vision. *Vision Res.* 2012; 72: 1–13. <https://doi.org/10.1016/j.visres.2012.08.019> PMID: 22982685
7. Traschütz A, Kreiter AK, Wegener D. Transient activity in monkey area MT represents speed changes and is correlated with human behavioral performance. *J Neurophysiol.* 2015; 113: 890–903. <https://doi.org/10.1152/jn.00335.2014> PMID: 25392161
8. Wegener D, Ehn F, Aurich MK, Galashan FO, Kreiter AK. Feature-based attention and the suppression of non-relevant object features. *Vision Res.* 2008; 48: 2696–2707. <https://doi.org/10.1016/j.visres.2008.08.021> PMID: 18824190
9. Galashan FO, Saßen HC, Kreiter AK, Wegener D. Monkey area MT latencies to speed changes depend on attention and correlate with behavioral reaction times. *Neuron.* 2013; 78: 740–750. <https://doi.org/10.1016/j.neuron.2013.03.014> PMID: 23719167
10. Schledde B, Galashan FO, Przybyla M, Kreiter AK, Wegener D. Task-specific, dimension-based attentional shaping of motion processing in monkey area MT. *J Neurophysiol.* 2017; 118: 1542–1555. <https://doi.org/10.1152/jn.00183.2017> PMID: 28659459
11. Maunsell JHR. Neuronal mechanisms of visual attention. *Ann Rev Vis Sci.* 2015; 1: 373–391.
12. Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Ann Rev Neurosci.* 1995; 18: 193–222. <https://doi.org/10.1146/annurev.ne.18.030195.001205> PMID: 7605061
13. Mikami A, Newsome WT, Wurtz RH. Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *J Neurophysiol.* 1986; 55: 1308–1327. <https://doi.org/10.1152/jn.1986.55.6.1308> PMID: 3016210
14. Nover H, Anderson CH, DeAngelis GC. A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. *J Neurosci.* 2005; 25: 10049–10060. <https://doi.org/10.1523/JNEUROSCI.1661-05.2005> PMID: 16251454
15. Lagae L, Raiguel S, Orban GA. Speed and direction selectivity of macaque middle temporal neurons. *J Neurophysiol.* 1993; 69: 19–39. <https://doi.org/10.1152/jn.1993.69.1.19> PMID: 8433131
16. Kanai R, Verstraten FAJ. Visual transients without feature changes are sufficient for the percept of a change. *Vision Res.* 2004; 44: 2233–2249. <https://doi.org/10.1016/j.visres.2004.04.005> PMID: 15208010
17. Heeger DJ. Normalization of cell responses in cat striate cortex. *Vis Neurosci.* 1992; 9: 181–197. <https://doi.org/10.1017/s0952523800009640> PMID: 1504027
18. Carandini M, Heeger DJ. Normalization as a canonical neural computation. *Nat Rev Neurosci.* 2011; 13: 51–62. <https://doi.org/10.1038/nrn3136> PMID: 22108672
19. Heeger DJ, Zemlianova KO. A recurrent circuit implements normalization, simulating the dynamics of V1 activity. *Proc Natl Acad Sci USA.* 2020; 117: 22494–22505. <https://doi.org/10.1073/pnas.2005417117> PMID: 32843341
20. Louie K, LoFaro T, Webb R, Glimcher PW. Dynamic divisive normalization predicts time-varying value coding in decision-related circuits. *J Neurosci.* 2014; 34: 16046–16057. <https://doi.org/10.1523/JNEUROSCI.2851-14.2014> PMID: 25429145
21. Sundberg KA, Mitchell JF, Gawne TJ, Reynolds JH. Attention influences single unit and local field potential response latencies in visual cortical area V4. *J Neurosci.* 2012; 32: 16040–16050. <https://doi.org/10.1523/JNEUROSCI.0489-12.2012> PMID: 23136440
22. Lee J, Williford T, Maunsell JHR. Spatial attention and the latency of neuronal responses in macaque area V4. *J Neurosci.* 2007; 27: 9632–9637. <https://doi.org/10.1523/JNEUROSCI.2734-07.2007> PMID: 17804623

23. Raiguel SE, Xiao DK, Marcar VL, Orban GA. Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. *J Neurophysiol.* 1999; 82: 1944–1956. <https://doi.org/10.1152/jn.1999.82.4.1944> PMID: 10515984
24. Bair W, Cavanaugh JR, Smith MA, Movshon JA. The timing of response onset and offset in macaque visual neurons. *J Neurosci.* 2002; 22: 3189–3205. <https://doi.org/20026257> PMID: 11943820
25. Kurtz R, Egelhaaf M, Meyer HG, Kern R. Adaptation accentuates responses of fly motion-sensitive visual neurons to sudden stimulus changes. *Proc R Soc B, Biol Sci.* 2009; 276: 3711–3719. <https://doi.org/10.1098/rspb.2009.0596> PMID: 19656791
26. Thiel A, Greschner M, Eurich CW, Ammermüller J, Kretzberg J. Contribution of individual retinal ganglion cell responses to velocity and acceleration encoding. *J Neurophysiol.* 2007; 98: 2285–2296. <https://doi.org/10.1152/jn.01342.2006> PMID: 17596411
27. Petersen SE, Miezin PM, Allman JM. Transient and sustained responses in four extrastriate visual areas of the owl monkey. *Exp Brain Res.* 1988; 70: 55–60. <https://doi.org/10.1007/BF00271847> PMID: 3402568
28. Uchida N, Kepecs A, Mainen ZF. Seeing at a glance, smelling in a whiff: rapid forms of perceptual decision making. *Nat Rev Neurosci.* 2006; 7: 485–491. <https://doi.org/10.1038/nrn1933> PMID: 16715056
29. Schlack A, Krekelberg B, Albright TD. Recent history of stimulus speeds affects the speed tuning of neurons in area MT. *J Neurosci.* 2007; 27: 11009–11018. <https://doi.org/10.1523/JNEUROSCI.3165-07.2007> PMID: 17928442
30. Price NSC, Born RT. Adaptation to speed in macaque middle temporal and medial superior temporal areas. *J Neurosci.* 2013; 33: 4359–4368. <https://doi.org/10.1523/JNEUROSCI.3165-12.2013> PMID: 23467352
31. Simoncelli EP, Heeger DJ. A model of neuronal responses in visual area MT. *Vision Res.* 1998; 38: 743–761. [https://doi.org/10.1016/S0042-6989\(97\)00183-1](https://doi.org/10.1016/S0042-6989(97)00183-1) PMID: 9604103
32. Chance FS, Abbott LF, Reyes AD. Gain modulation from background synaptic input. *Neuron.* 2002; 35: 773–782. [https://doi.org/10.1016/S0896-6273\(02\)00820-6](https://doi.org/10.1016/S0896-6273(02)00820-6) PMID: 12194875
33. Rust NC, Mante V, Simoncelli EP, Movshon JA. How MT cells analyze the motion of visual patterns. *Nat Neurosci.* 2006; 9: 1421–1431. <https://doi.org/10.1038/nn1786> PMID: 17041595
34. Busse L, Wade AR, Carandini M. Representation of concurrent stimuli by population activity in visual cortex. *Neuron.* 2009; 64: 931–942. <https://doi.org/10.1016/j.neuron.2009.11.004> PMID: 20064398
35. Ruff DA, Alberts JJ, Cohen MR. Relating normalization to neuronal populations across cortical areas. *J Neurophysiol.* 2016; 116: 1375–1386. <https://doi.org/10.1152/jn.00017.2016> PMID: 27358313
36. Reynolds JH, Heeger DJ. The normalization model of attention. *Neuron.* 2009; 61: 168–185. <https://doi.org/10.1016/j.neuron.2009.01.002> PMID: 19186161
37. Nishimoto S, Gallant JL. A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *J Neurosci.* 2011; 31: 14551–14564. <https://doi.org/10.1523/JNEUROSCI.6801-10.2011> PMID: 21994372
38. Lisberger SG, Movshon JA. Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. *J Neurosci.* 1999; 19: 2224–2246. <https://doi.org/10.1523/JNEUROSCI.19-06-02224.1999> PMID: 10066275
39. Denison RN, Carrasco M, Heeger DJ. A dynamic normalization model of temporal attention. *bioRxiv.* 2019: 2019.12.21.886051.
40. Beuth F, Hamker FH. A mechanistic cortical microcircuit of attention for amplification, normalization and suppression. *Vision Res.* 2015; 116: 241–257. <https://doi.org/10.1016/j.visres.2015.04.004> PMID: 25883048
41. Bayerl P, Neumann H. A neural model of feature attention in motion perception. *Biosystems.* 2007; 89: 208–215. <https://doi.org/10.1016/j.biosystems.2006.04.018> PMID: 17280774
42. Brosch T, Neumann H. Computing with a canonical neural circuits model with pool normalization and modulating feedback. *Neural Comput.* 2014; 26: 2735–2789. https://doi.org/10.1162/NECO_a_00675 PMID: 25248083
43. Medathati NVK, Rankin J, Meso AI, Kornprobst P, Masson GS. Recurrent network dynamics reconciles visual motion segmentation and integration. *Sci Rep.* 2017; 7: 11270. <https://doi.org/10.1038/s41598-017-11373-z> PMID: 28900120
44. Malo JE-T, J.J.; Bertalmio M. Divisive normalization from Wilson-Cowan dynamics. *arXiv.* 2019: 1906.08246v2 [q-bio.NC].
45. Rubin Daniel B, Van Hooser Stephen D, Miller Kenneth D. The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron.* 2015; 85: 402–417. <https://doi.org/10.1016/j.neuron.2014.12.026> PMID: 25611511

46. Wehrhahn C, Poggio T, Bühlhoff H. Tracking and chasing in houseflies (*Musca*). *Biol Cybern*. 1982; 45: 123–130.
47. Bro-Jørgensen J. Evolution of sprint speed in African savannah herbivores in relation to predation. *Evolution*. 2013; 67: 3371–3376. <https://doi.org/10.1111/evo.12233> PMID: 24152014
48. Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. *Nature*. 1996; 381: 520–522. <https://doi.org/10.1038/381520a0> PMID: 8632824
49. Girard P, Jouffrais C, Kirchner CH. Ultra-rapid categorisation in non-human primates. *Anim Cogn*. 2008; 11: 485–493. <https://doi.org/10.1007/s10071-008-0139-2> PMID: 18259787
50. Kirchner H, Thorpe SJ. Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Res*. 2006; 46: 1762–1776. <https://doi.org/10.1016/j.visres.2005.10.002> PMID: 16289663
51. Oram MW, Perrett DI. Time course of neural responses discriminating different views of the face and head. *J Neurophysiol*. 1992; 68: 70–84. <https://doi.org/10.1152/jn.1992.68.1.70> PMID: 1517829
52. Tovée MJ, Rolls ET, Treves A, Bellis RP. Information encoding and the responses of single neurons in the primate temporal visual cortex. *J Neurophysiol*. 1993; 70: 640–654. <https://doi.org/10.1152/jn.1993.70.2.640> PMID: 8410164
53. Müller JR, Metha AB, Krauskopf J, Lennie P. Information conveyed by onset transients in responses of striate cortical neurons. *J Neurosci*. 2001; 21: 6978–6990. <https://doi.org/10.1523/JNEUROSCI.21-17-06978.2001> PMID: 11517285
54. Churchland MM, Yu BM, Cunningham JP, Sugrue LP, Cohen MR, Corrado GS, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat Neurosci*. 2010; 13: 369–378. <https://doi.org/10.1038/nn.2501> PMID: 20173745
55. Ghose GM, Harrison IT. Temporal precision of neuronal information in a rapid perceptual judgment. *J Neurophysiol*. 2009; 101: 1480–1493. <https://doi.org/10.1152/jn.90980.2008> PMID: 19109454
56. Price NSC, Born RT. Timescales of sensory- and decision-related activity in the middle temporal and medial superior temporal areas. *J Neurosci*. 2010; 30: 14036–14045. <https://doi.org/10.1523/JNEUROSCI.2336-10.2010> PMID: 20962225
57. Herrington TM, Assad JA. Neural activity in the middle temporal area and lateral intraparietal area during endogenously cued shifts of attention. *J Neurosci*. 2009; 29: 14160–14176. <https://doi.org/10.1523/JNEUROSCI.1916-09.2009> PMID: 19906965
58. Harrison IT, Weiner KF, Ghose GM. Inattention blindness to motion in middle temporal area. *J Neurosci*. 2013; 33: 8396–8410. <https://doi.org/10.1523/JNEUROSCI.3409-12.2013> PMID: 23658178
59. Smith JET, Zhan CA, Cook EP. The functional link between area MT neural fluctuations and detection of a brief motion stimulus. *J Neurosci*. 2011; 31: 13458–13468. <https://doi.org/10.1523/JNEUROSCI.1347-11.2011> PMID: 21940439
60. Cook EP, Maunsell JH. Dynamics of neuronal responses in macaque MT and VIP during motion detection. *Nat Neurosci*. 2002; 5: 985–994. <https://doi.org/10.1038/nn924> PMID: 12244324
61. Bisley JW, Krishna BS, Goldberg ME. A rapid and precise on-response in posterior parietal cortex. *J Neurosci*. 2004; 24: 1833–1838. <https://doi.org/10.1523/JNEUROSCI.5007-03.2004> PMID: 14985423
62. Womelsdorf T, Fries P, Mitra PP, Desimone R. Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature*. 2006; 439: 733–736. <https://doi.org/10.1038/nature04258> PMID: 16372022
63. Khayat PS, Martinez-Trujillo JC. Effects of attention and distractor contrast on the responses of middle temporal area neurons to transient motion direction changes. *Eur J Neurosci*. 2015; 41: 1603–1613. <https://doi.org/10.1111/ejn.12920> PMID: 25885809
64. Treue S, Martinez-Trujillo JC. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*. 1999; 399: 575–579. <https://doi.org/10.1038/21176> PMID: 10376597
65. Martinez-Trujillo JC, Treue S. Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol*. 2004; 14: 744–751. <https://doi.org/10.1016/j.cub.2004.04.028> PMID: 15120065
66. Gledhill D, Grimsen C, Fahle M, Wegener D. Human feature-based attention consists of two spatiotemporally distinct processes. *J Vis*. 2015; 15: 8.
67. Wegener D, Freiwald WA, Kreiter AK. The influence of sustained selective attention on stimulus selectivity in macaque visual area MT. *J Neurosci*. 2004; 24: 6106–6114. <https://doi.org/10.1523/JNEUROSCI.1459-04.2004> PMID: 15240802
68. Harris KD, Henze DA, Csicsvari J, Hirase H, Buzsáki G. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J Neurophysiol*. 2000; 84: 401–414. <https://doi.org/10.1152/jn.2000.84.1.401> PMID: 10899214

69. Galashan FO, Rempel HC, Meyer A, Gruber-Dujardin E, Kreiter AK, Wegener D. A new type of recording chamber with an easy-to-exchange microdrive array for chronic recordings in macaque monkeys. *J Neurophysiol.* 2011; 105: 3092–3105. <https://doi.org/10.1152/jn.00508.2010> PMID: 21451061