nature neuroscience

A normalization model of multisensory integration

Tomokazu Ohshiro[1], Dora E Angelaki[2] & Gregory C DeAngelis[1]

Responses of neurons that integrate multiple sensory inputs are traditionally characterized in terms of a set of empirical principles. However, a simple computational framework that accounts for these empirical features of multisensory integration has not been established. We propose that divisive normalization, acting at the stage of multisensory integration, can account for many of the empirical principles of multisensory integration shown by single neurons, such as the principle of inverse effectiveness and the spatial principle. This model, which uses a simple functional operation (normalization) for which there is considerable experimental support, also accounts for the recent observation that the mathematical rule by which multisensory neurons combine their inputs changes with cue reliability. The normalization model, which makes a strong testable prediction regarding cross-modal suppression, may therefore provide a simple unifying computational account of the important features of multisensory integration by neurons.

In an uncertain environment, organisms often need to react quickly to subtle changes in their surroundings. Integrating inputs from multiple sensory systems (for example, vision, audition and somatosensation) can increase perceptual sensitivity, enabling better detection or discrimination of events in the environment[1–3]. A basic question in multisensory integration is: how do single neurons combine their unisensory inputs? Although neurophysiological studies have revealed a set of empirical principles by which two sensory inputs interact to modify neural responses[4], the computations performed by neural circuits that integrate multisensory inputs are not well understood.

A prominent feature of multisensory integration is the principle of inverse effectiveness, which states that multisensory enhancement is large for weak multimodal stimuli and decreases with stimulus intensity[4–7]. A second prominent feature is the spatial/temporal principle of multisensory enhancement, which states that stimuli should be spatially congruent and temporally synchronous for robust multisensory enhancement to occur, with large spatial or temporal offsets leading instead to response suppression[8–10]. Although these empirical principles are well established, the nature of the mechanisms required to explain them remains unclear.

We recently measured the mathematical rules by which multisensory neurons combine their inputs[11]. These studies were performed in the dorsal medial superior temporal area (MSTd), where visual and vestibular cues to self-motion are integrated[12–14]. We found that bimodal responses to combinations of visual and vestibular inputs were well described by a weighted linear sum of the unimodal responses, consistent with recent theory[15]. Notably, however, the linear weights appeared to change with reliability of the visual cue[11], suggesting that the neural 'combination rule' changes with cue reliability. It is unclear whether this result implies dynamic changes in synaptic weights with cue reliability or whether it can be explained by network properties.

We propose a divisive normalization model of multisensory integration that accounts for the apparent change in neural weights with cue reliability, as well as several other important empirical principles of multisensory integration. Divisive normalization[16] has been successful in describing how neurons in primary visual cortex (V1) respond to combinations of stimuli having multiple contrasts and orientations[17,18]. Divisive normalization has also been implicated in motion integration in area MT[19], as well as in attentional modulation of neural responses[20]. Our model extends the normalization framework to multiple sensory modalities, demonstrates that a simple set of neural operations can account for the main empirical features of multisensory integration and makes predictions for experiments that could identify neural signatures of normalization in multisensory areas.

RESULTS

Brief description of the model

The model consists of two layers of primary neurons, each sensitive to inputs from a different sensory modality (for example, visual or auditory), and one layer of multisensory neurons that integrate the primary sensory inputs (**Fig. 1a**). In our basic version of the model, we assume that a pair of primary neurons with spatially overlapping receptive fields provides input to the same multisensory neuron. Therefore, each multisensory neuron has spatially congruent receptive fields, similar to neurons in the superior colliculus[9].

The unisensory inputs to each multisensory neuron increase monotonically, but sublinearly, with stimulus intensity (**Fig. 1b**). This input nonlinearity models response saturation in the sensory inputs[21], which could be mediated by means of synaptic depression[22] or normalization within the unisensory pathways. This assumption has little effect on the multisensory properties of model neurons, but it plays an important role in the response to multiple unisensory inputs.

Following the input nonlinearity, each multisensory neuron performs a weighted linear sum (E) of its unisensory inputs with weights, d_1 and d_2, that we term modality dominance weights

$$E = d_1 \cdot I_1(x_0, y_0) + d_2 \cdot I_2(x_0, y_0) \quad (1)$$

Here, $I_1(x_0, y_0)$ and $I_2(x_0, y_0)$ represent, in simplified form, the two unisensory inputs to the multisensory neuron, indexed by the spatial location of the receptive fields (see Online Methods for a detailed

[1]Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York, USA. [2]Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, Missouri, USA. Correspondence should be addressed to G.C.D. (gdeangelis@cvs.rochester.edu).

Received 28 January; accepted 21 March; published online 8 May 2011; doi:10.1038/nn.2815

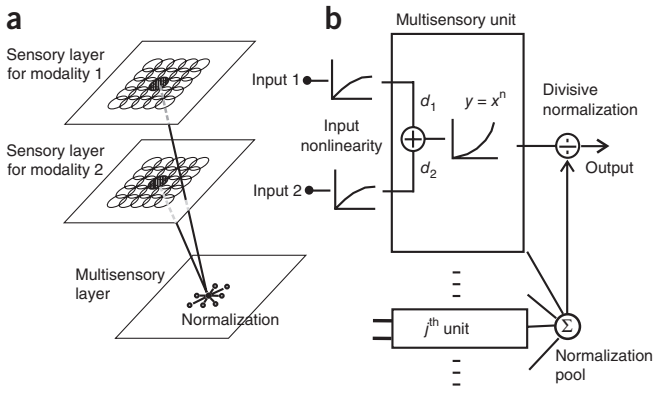


Figure 1 Schematic illustration of the normalization model of multisensory integration. (**a**) Overview of network architecture. The model consists of two layers of primary neurons that respond exclusively to sensory modalities 1 and 2. These primary sensory units feed into a layer of multisensory neurons that integrate responses from unisensory inputs with matched receptive fields. (**b**) Signal processing at the multisensory stage. Each unisensory input first passes through a nonlinearity that could represent synaptic depression or normalization in the unisensory pathways. The multisensory neuron then performs a weighted linear sum of its inputs with modality dominance weights, d_1 and d_2. Following an expansive power-law nonlinearity that could represent the transformation from membrane potential to firing rate, the response is normalized by the net activity of all other multisensory neurons.

formulation). The modality dominance weights are fixed for each multisensory neuron in the model (they do not vary with stimulus parameters), but different neurons have different combinations of d_1 and d_2 to simulate various degrees of dominance of one sensory modality. Following an expansive power-law output nonlinearity, which simulates the transformation from membrane potential to firing rate[17,23], the activity of each neuron is divided by the net activity of all multisensory neurons to produce the final response (divisive normalization[16])

$$R = \frac{E^n}{\alpha^n + \left(\frac{1}{N}\right)\sum_{j=1}^{N} E_j^n} \qquad (2)$$

The parameters that govern the response of each multisensory neuron are the modality dominance weights (d_1 and d_2), the exponent (n) of the output nonlinearity, the semi-saturation constant (α) and the locations of the receptive fields (x_0, y_0). The semi-saturation constant, α, determines the neuron's overall sensitivity to stimulus intensity, with larger α shifting the intensity-response curve rightward on a logarithmic axis. We found that this simple model accounts for key empirical principles of multisensory integration that have been described previously.

Inverse effectiveness

The principle of inverse effectiveness states that combinations of weak inputs produce greater multisensory enhancement than combinations of strong inputs[4–7]. In addition, the combined response to weak stimuli is often greater than the sum of the unisensory responses (super-additivity), whereas the combined response to strong stimuli tends toward additive or sub-additive interactions. Note, however, that inverse effectiveness can hold independent of whether weak inputs produce super-additivity or not[7,24]. The normalization model accounts naturally for these observations, including the dissociation between inverse effectiveness and super-additivity.

We examined responses of a representative model neuron ($d_1 = d_2$) as a function of the intensities of the two unisensory inputs (**Fig. 2a**). For inputs of equal strength (**Fig. 2b**), bimodal responses exceed the corresponding unimodal responses at all intensities, consistent with physiological results obtained with balanced unisensory stimuli that are centered on the receptive fields[24]. For low stimulus intensities, the bimodal response exceeds the sum of the unimodal responses, indicating super-additivity (**Fig. 2b**). However, as stimulus intensity increases, the bimodal response becomes sub-additive, demonstrating inverse effectiveness.

To quantify this effect, we computed an additivity index, which is the ratio of the bimodal response to the sum of the two unimodal responses. The normalization model predicts that super-additivity (additivity index > 1) only occurs when both sensory inputs have low intensities (**Fig. 2c**), which may explain why super-additivity was not seen in previous studies in which one input was fixed at a high intensity[11]. Furthermore, the degree of super-additivity is determined by the exponent parameter (n) of the output nonlinearity (**Fig. 2d**). For an exponent of 2.0, which we used as a default[25], the model predicts an additivity index ≈ 2 for low intensities (**Fig. 2d**). Larger exponents produce even greater super-additivity, whereas the model predicts purely additive responses to low intensities when the exponent is 1.0 (**Fig. 2d**). Thus, the degree of super-additivity is determined by the curvature of the power-law nonlinearity and greater super-additivity can be achieved by adding a response threshold to the model (data not shown).

For large stimulus intensities, responses become sub-additive (additivity index < 1) regardless of the exponent (**Fig. 2d**) and this effect is driven by divisive normalization. Thus, all model neurons exhibit inverse effectiveness, but super-additivity is seen only when responses are weak, such that the expansive output nonlinearity has a substantial effect. These predictions are qualitatively consistent with physiological data from the superior colliculus, where neurons show inverse effectiveness regardless of whether or not they show super-additivity, and only neurons with weak multisensory responses exhibit super-additivity[24].

To evaluate performance of the model quantitatively, we compared model predictions to population data from the superior colliculus[26]. Response additivity was quantified by computing a z score[7] that quantifies the difference between the bimodal response and the sum of the two unimodal responses (a z score of zero corresponds to perfect additivity, analogous to additivity index = 1). For combined visual-auditory stimuli, significant super-additivity (z score > 1.96) was observed for weak stimuli and additivity was seen for stronger stimuli (**Fig. 2e**), thus demonstrating inverse effectiveness. After adding Poisson noise and adjusting parameters to roughly match the range of firing rates, the normalization model produces very similar results (**Fig. 2f**). Thus, the model accounts quantitatively for the transition from super-additivity at low intensities to additivity (or sub-additivity) at high intensities, with a single set of parameters. Although these simulations assumed specific model parameters, inverse effectiveness is a robust property of the model even when stimuli are not centered on the receptive fields, or modality dominance weights are unequal (**Supplementary Figs. 1** and **2**).

Spatial principle of multisensory integration

The spatial principle of multisensory enhancement states that a less effective stimulus from one sensory modality (for example, a stimulus placed off the receptive field center) can suppress the response to a highly effective stimulus from the other modality[9,10]. Divisive normalization accounts naturally for this effect (**Fig. 3**).

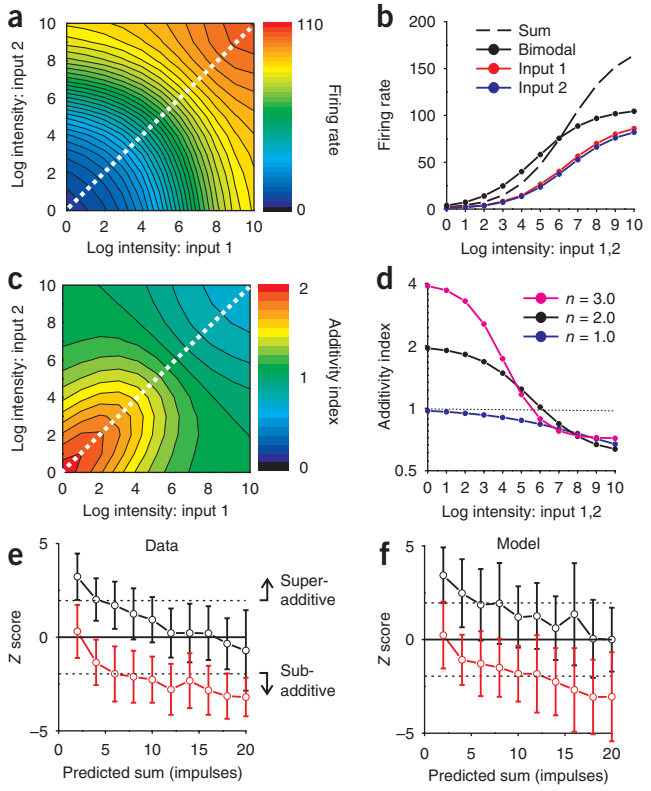


Figure 2 Normalization accounts for the principle of the inverse effectiveness. (**a**) The bimodal response of a model unit is plotted as a function of the intensities of input 1 and input 2. Both inputs were located in the center of the receptive field. Diagonal line, inputs with equal intensities. Exponent, $n = 2.0$. (**b**) The bimodal response (solid black curve) and the unimodal responses (red and blue curves) are plotted as a function of stimulus intensity (from the diagonal of **a**). The sum of the two unimodal responses is shown as the dashed black curve. The red and blue curves have slightly different amplitudes to improve clarity. (**c**) Additivity index is plotted as a function of both input intensities. Additivity index > 1 indicates super-additivity and additivity index < 1 indicates sub-additivity. (**d**) Additivity index values (from the diagonal of **c**) are plotted as a function of intensity for three exponent values: $n = 1.0$, 2.0 and 3.0. (**e**) Data from cat superior colliculus, demonstrating inverse effectiveness (replotted from ref. 26). The *z*-scored bimodal response (± s.d.) is plotted against the predicted sum of the two unimodal responses, both for cross-modal (visual-auditory) inputs (black curve) or pairs of visual inputs (red). *z* score values > 1.96 represent significant super-additivity and values < –1.96 denote significant sub-additivity. (**f**) Model predictions match the data from cat superior colliculus. For this simulation, model neurons had all nine combinations of dominance weights from the set (d_1, d_2 = 0.50, 0.75 or 1.00), and the exponent, n, was 1.5.

In this simulation, one of the unimodal inputs (input 1) is presented in the center of the receptive field, while the other input (input 2) is spatially offset from the receptive field center by different amounts (**Fig. 3a**). When both inputs are centered on the receptive field (**Fig. 3b**), the combined response exceeds the unimodal responses for all stimulus intensities (as in **Fig. 2b**). As input 2 is offset from the receptive field center, the bimodal response decreases relative to that of the more effective input 1. Notably, when the stimulus offset substantially exceeds 1 s.d. of the Gaussian receptive field profile (**Fig. 3b**), the combined response becomes suppressed below the unimodal response to input 1. Hence, the model neuron exhibits the spatial principle. The intuition for this result is simple: the less effective (that is, offset) input contributes little to the underlying linear response of the neuron, but contributes strongly to the normalization signal because the normalization pool includes neurons with receptive fields that span a larger region of space. Note that the model neuron exhibits inverse effectiveness for all of these stimulus conditions (**Fig. 3c**), although super-additivity declines as the spatial offset increases.

We examined data from two cat superior colliculus neurons that illustrate the spatial principle[9,27] (**Fig. 3d**). Both neurons show cross-modal enhancement when the spatial offset between visual and auditory stimuli is small and a transition toward cross-modal suppression for large offsets. The normalization model captures the basic form of these data nicely (**Fig. 3d**). We are not aware of any published data that quantify the spatial principle for a population of neurons.

For the example neurons (**Fig. 3d**), responses to the offset stimulus were not presented[9,27], so it is not clear whether cross-modal suppression occurs while the non-optimal stimulus is excitatory on its own. However, the normalization model makes a critical testable prediction: in a specific stimulus domain, the less effective input 2 evokes

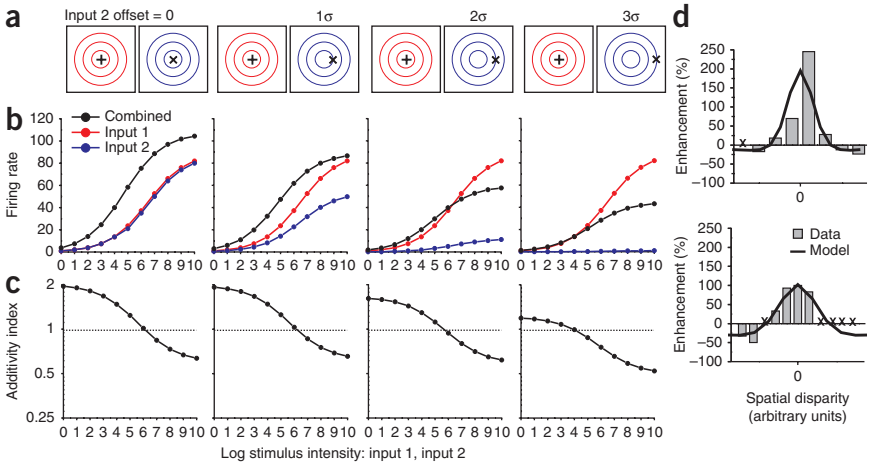


Figure 3 Normalization and the spatial principle of multisensory enhancement. (**a**) Schematic illustration of stimulus conditions used to simulate the spatial principle. Input 1 (+) was located at the center of the receptive field for modality 1 (red contours). Input 2 (×) was offset by various amounts relative to the receptive field for modality 2 (blue contours). Contours defining each receptive field are separated by 1 s.d. (σ) of the Gaussian. The modality dominance weights were equal ($d_1 = d_2 = 1$). (**b**) Responses of the model neuron to the stimuli illustrated in **a** (format as in **Fig. 2b**). Response is plotted as a function of intensity for input 1 (red), input 2 (blue) and the bimodal stimulus (black). Input 2 can be excitatory on its own (blue), but suppress the response to input 1 (red) when the two are combined (black, third column). (**c**) Additivity index as a function of stimulus intensity. (**d**) Two examples of the spatial principle for neurons from cat superior colliculus, re-plotted from refs 9,27. The response enhancement index (%) is plotted as a function of the spatial offset between visual and auditory stimuli (gray bars). Locations marked with an 'x' denote missing data in the original dataset. Predictions of the normalization model are shown as black curves. Model parameters (fit by hand) were $d_1 = d_2 = 1.0$, $\alpha = 1.0$ and $n = 2.0$. Stimulus intensity was set at 16 for the top neuron and 64 for the bottom neuron.

a clearly excitatory response on its own (**Fig. 3b**), but suppresses the response to the more effective input 1 when the two inputs are presented together. Analogous interactions between visual stimuli were demonstrated in V1 neurons and attributed to normalization[17]. Notably, the cross-modal suppression (**Fig. 3b**) appears to be a signature of a multisensory normalization mechanism, as alternative model architectures that do not incorporate normalization[28,29] fail to exhibit this behavior (**Supplementary Figs. 3** and **4**).

An analogous empirical phenomenon is the temporal principle of multisensory integration, which states that multisensory enhancement is strongest when inputs from different modalities are synchronous and declines when the inputs are separated in time[8]. The normalization model also accounts for the temporal principle as long as there is variation in response dynamics (for example, latency and duration) among neurons in the population, such that the temporal response of the normalization pool is broader than the temporal response of individual neurons (**Supplementary Fig. 5**). More generally, in any stimulus domain, adding a non-optimal (but excitatory) stimulus can produce cross-modal suppression if it increases the normalization signal enough to overcome the additional excitatory input to the neuron.

Multisensory suppression in unisensory neurons

Multisensory neurons are often much more responsive to one sensory modality than the other. Responses of such neurons to the more effective input can be suppressed by simultaneous presentation of the seemingly non-effective input[5,30]. In the normalization model, each neuron receives inputs from primary sensory neurons with modality dominance weights that are fixed (**Fig. 1**), but the specific combination of weights (d_1, d_2) varies from cell to cell. We found that normalization accounts for response suppression by the non-effective input.

We simulated responses for four model neurons (**Fig. 4**) with different combinations of modality dominance weights, ranging from balanced inputs ($d_1 = 1.0$, $d_2 = 1.0$) to strictly unisensory input ($d_1 = 1.0$, $d_2 = 0$). When modality dominance weights are equal ($d_1 = 1.0$, $d_2 = 1.0$), the model shows multisensory enhancement (**Fig. 4b**). As the weight on input 2 is reduced, the bimodal response declines along with the unimodal response to input 2. Notably, when d_2 is approximately 0.5 or less, the bimodal response becomes suppressed below the best unimodal response (**Fig. 4b**). For the unisensory neuron with $d_2 = 0$, input 2 evokes no excitation, but suppresses the combined response. This effect is reminiscent of cross-orientation suppression in primary visual cortex[18,31,32].

Normalization accounts for cross-modal suppression in unisensory neurons by similar logic used to explain the spatial principle: although input 2 makes no contribution to the linear response of the neuron when $d_2 = 0$, it still contributes to the normalization signal via other responsive neurons with nonzero d_2. This effect is robust as long as the normalization pool contains neurons with a range of modality dominance weights. Response additivity (**Fig. 4c**) again shows inverse effectiveness in all conditions, with super-additivity for weak, balanced inputs.

To assess model performance quantitatively, we compared predictions to an extensive dataset of multisensory responses of macaque ventral intraparietal (VIP) neurons to visual and tactile stimuli[30]. In this dataset, a measure of response additivity is plotted against a measure of multisensory enhancement (**Fig. 4d**). The pattern of data across the population of VIP neurons is largely reproduced by a subset of model neurons that vary along just two dimensions: the semi-saturation constant (α) and the ratio of dominance weights (d_2/d_1). Increasing the value of α shifts the intensity-response curve to the right and yields greater super-additivity for a fixed stimulus

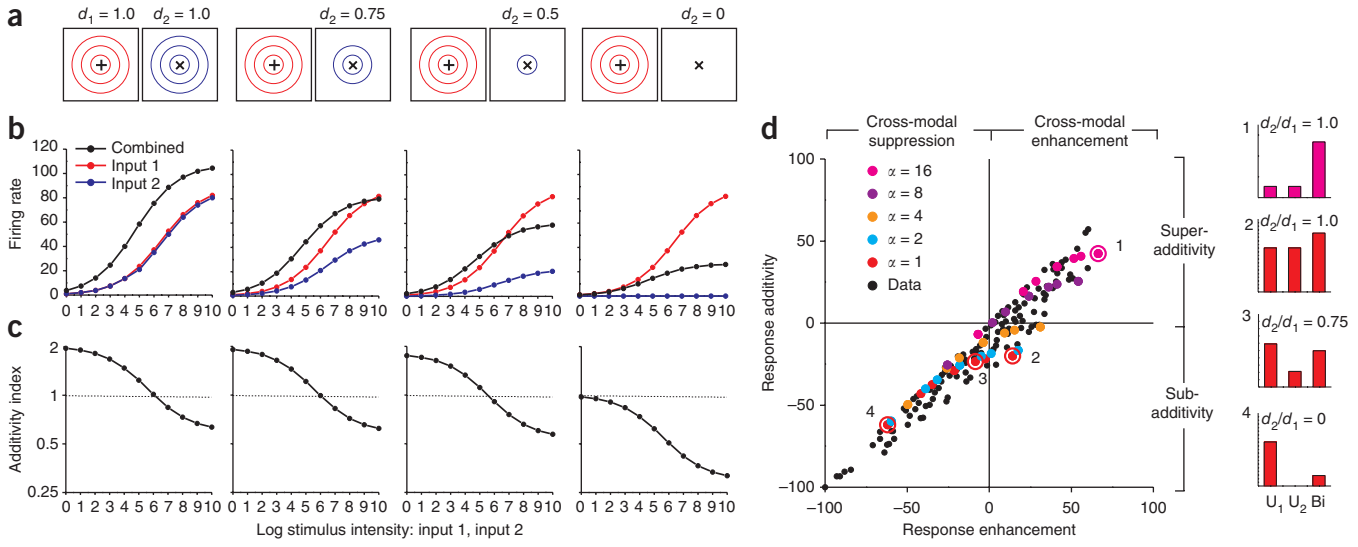


Figure 4 Multisensory suppression in unisensory neurons. (**a**) Responses were simulated for four model neurons. The dominance weight for modality 1, d_1, was fixed at unity while the dominance weight for modality 2, d_2, decreased from left to right (denoted by the number of receptive field contours). Input 1 (+) and input 2 (×) were presented in the center of the receptive fields. (**b**) Responses as a function of intensity are shown for input 1 (red), input 2 (blue) and both inputs together (black). Data are presented as in **Figure 3b**. (**c**) Additivity index is plotted as a function of intensity for the four model neurons. Data are presented as in **Figure 3c**. (**d**) Summary of multisensory integration properties for a population of neurons from area VIP (black symbols), re-plotted from ref. 30. The ordinate shows a measure of response additivity: $(Bi - (U_1 + U_2)) / (Bi + (U_1 + U_2)) \times 100$, for which positive and negative values indicate super-additive and sub-additive interactions, respectively. Bi: bimodal response; U_1, U_2: unimodal responses. The abscissa represents a measure of response enhancement: $(Bi - \max(U_1, U_2)) / (Bi + \max(U_1, U_2)) \times 100$, for which positive and negative values denote cross-modal enhancement and cross-modal suppression, respectively. Colored symbols represent predictions of the normalization model for units that vary in the ratio of dominance weights (d_2/d_1, ranging from 0 to 1) and the semi-saturation constant, α, ranging from 1 to 16. The exponent, n, was 2.5. Numbered symbols correspond to model neurons for which responses are shown as bar graphs (right).

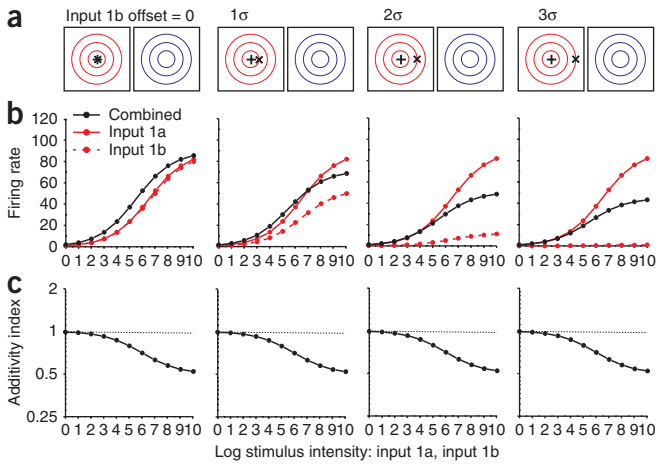


Figure 5 Interactions among within-modality inputs. (**a**) The stimulus configuration was similar to that of **Figure 3a** except that two stimuli of the same sensory modality, input 1a (+) and input 1b (×) were presented, and one was systematically offset relative to the receptive field of modality 1 (red contours). No stimulus was presented to the receptive field of modality 2 (blue contours). (**b**) Responses of a model neuron are shown for input 1a alone (solid red curve), Input 1b alone (dashed red curve) and both inputs together (black curve). (**c**) Additivity index as a function of stimulus intensity shows that model responses to pairs of within-modality inputs are additive or sub-additive with no super-additivity.

intensity. For a fixed value of α, varying the ratio of dominance weights shifts the data from upper right toward lower left. Model neuron 4 (**Fig. 4d**) is an example of multisensory suppression in a unisensory neuron. Overall, a substantial proportion of variance in the VIP data can be accounted for by a normalization model in which neurons vary in two biologically plausible ways: overall sensitivity to stimulus intensity (α) and relative strength of the two sensory inputs (d_2/d_1).

Response to within-modal stimulus combinations

Previous studies have reported that two stimuli of the same sensory modality (for example, two visual inputs) interact sub-additively, whereas two stimuli of different modalities can produce super-additive interactions[26]. This distinction arises naturally from the normalization model if each unisensory pathway incorporates a sub-linear nonlinearity (**Fig. 1b**) that could reflect synaptic depression or normalization operating at a previous stage.

We examined responses to two inputs from the same modality (input 1a and input 1b) for a model neuron (**Fig. 5**). Input 1a is presented at the center of the receptive field, while input 1b is systematically offset from the receptive field center (**Fig. 5a**). When both inputs are centered in the receptive field (**Fig. 5b**), the combined response is modestly enhanced. The corresponding additivity index curve (**Fig. 5c**) indicates that the interaction is additive for weak inputs and sub-additive for stronger inputs. This result contrasts sharply with the super-additive interaction seen for spatially aligned cross-modal stimuli (**Figs. 2b** and **3b**). As input 1b is offset from the center of the receptive field (**Fig. 5b**), the combined response becomes suppressed relative to the stronger unisensory response. The additivity index curves demonstrate that the interaction is either additive or sub-additive for all spatial offsets (**Fig. 5c**).

Presenting two overlapping inputs from the same modality is operationally equivalent to presenting one input with twice the stimulus intensity. As a result of the sublinear nonlinearity in each unisensory pathway, doubling the stimulus intensity does not double the postsynaptic excitation. As a result, the combined response does not exhibit super-additivity for low intensities, even with an expansive output nonlinearity in the multisensory neuron ($n = 2.0$). For high-stimulus intensities, the combined response becomes sub-additive as a result of normalization. If normalization were removed from the model, combined responses would remain approximately additive across all stimulus intensities (data not shown). For large spatial offsets and strong intensities, the combined response is roughly the average of the two single-input responses (**Fig. 5b**). Similar averaging behavior has been observed for superior colliculus neurons[26], as well as neurons in primary[18] and extrastriate[33,34] visual cortex.

In the superior colliculus[26], super-additivity was substantially reduced for pairs of inputs from the same modality (**Fig. 2e**) relative to cross-modal inputs. This difference is reproduced by the normalization model (**Fig. 2f**) with a single set of model parameters. Hence, the inclusion of an input nonlinearity appears to account quantitatively for the difference in additivity of responses between cross-modal and within-modal stimulation.

Multisensory integration and cue reliability

We found that normalization accounts for key empirical principles of multisensory integration. We then examined whether the model can account for quantitative features of the combination rule by which neurons integrate their inputs. We recently demonstrated that bimodal responses of multisensory neurons in area MSTd are well approximated by a weighted linear sum of visual and vestibular inputs, but that the weights appear to change with visual cue strength[11]. To explain this puzzling feature of the multisensory combination rule, we performed a virtual replication of the MSTd experiment[11] using model neurons. To capture known physiology of heading-selective neurons[13,35], we modified the model architecture such that each cell had spherical heading tuning, lateral heading preferences were more common than fore-aft preferences, and many neurons had mismatched heading tuning for the two cues (see Online Methods and **Supplementary Figs. 6** and **7**).

Responses of model neurons were computed for eight heading directions in the horizontal plane using visual inputs alone, vestibular inputs alone, and all 64 combinations of visual and vestibular headings, both congruent and conflicting. The bimodal response profile of an example neuron, $R_{\text{bimodal}}(\varphi_{\text{vest}}, \varphi_{\text{vis}})$, is plotted as a color contour map, along with the two unimodal response curves, $R_{\text{vest}}(\varphi_{\text{vest}})$ and $R_{\text{vis}}(\varphi_{\text{vis}})$, along the margins (**Fig. 6a**). The intensity of the vestibular cue was kept constant while the intensity of the visual cue was varied to simulate the manipulation of motion coherence used in MSTd[11]. At 100% coherence, the bimodal response is dominated by the visual input, as is typical of MSTd neurons[11]. As visual intensity (motion coherence) is reduced, the bimodal response profile changes shape and becomes dominated by the vestibular heading tuning (**Fig. 6a–c**).

The bimodal response of each model neuron was fit with a weighted linear sum of the two unimodal response curves

$$R_{\text{bimodal}}(\varphi_{\text{vest}},\varphi_{\text{vis}}) = w_{\text{vest}} \cdot R_{\text{vest}}(\varphi_{\text{vest}}) + w_{\text{vis}} \cdot R_{\text{vis}}(\varphi_{\text{vis}}) + C \qquad (3)$$

The mixing weights, w_{vest} and $w_{vis,}$ were obtained for each of the three visual intensities, corresponding to motion coherences of 25, 50 and 100%. This analysis was performed for model neurons with different combinations of modality dominance weights (d_{vest}, d_{vis}; all combinations of values 0.25, 0.5, 0.75 and 1.0). Note that d_{vest}, d_{vis} characterize how each model neuron weights its vestibular and visual inputs and that these modality dominance weights are fixed for each neuron in

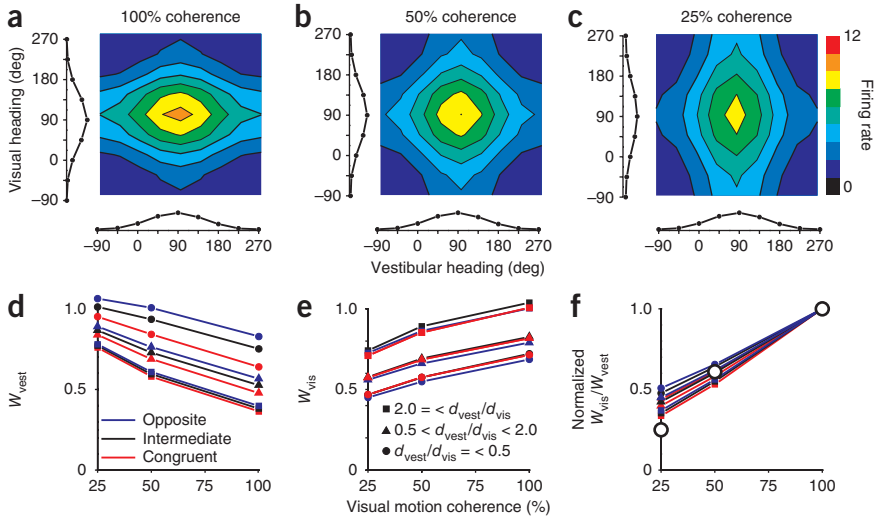


Figure 6 Normalization accounts for apparent changes in the multisensory combination rule with cue reliability. (**a**) Responses of a model MSTd neuron to visual and vestibular heading stimuli (100% visual motion coherence). The bimodal response, $R_{bimodal}(\varphi_{vest}, \varphi_{vis})$, to many combinations of visual (φ_{vis}) and vestibular (φ_{vest}) headings is shown as a color contour plot. Curves along the bottom and left margins represent unimodal responses, $R_{vest}(\varphi_{vest})$ and $R_{vis}(\varphi_{vis})$. This model neuron prefers forward motion (90°). (**b**) Reponses of the same model neuron when visual stimulus intensity is reduced to 50% coherence. (**c**) Responses to 25% coherence. Vestibular stimulus amplitude is constant in **a**–**c** at the equivalent of 50% coherence. (**d**) Bimodal responses were fit with a weighted linear sum of unimodal responses. This panel shows the vestibular mixing weight, w_{vest}, as a function of motion coherence. Red, blue and black points denote model neurons with congruent, opposite and intermediate visual and vestibular heading preferences, respectively. Different symbol shapes denote groups of neurons with different ratios of modality dominance weights: $d_{vest}/d_{vis} \leq 0.5$, $0.5 < d_{vest}/d_{vis} < 2.0$, or $2.0 \leq d_{vest}/d_{vis}$. Note that w_{vest} decreases with coherence. (**e**) The visual mixing weight, w_{vis}, increases with coherence. Data are presented as in **d**. (**f**) The ratio of vestibular and visual mixing weights, w_{vis}/w_{vest}, normalized to unity at 100% coherence, is plotted as a function of motion coherence. Model predictions are qualitatively consistent with data from area MSTd, replotted from ref. 11 as large open symbols.

the model. In contrast, w_{vest} and w_{vis} are weights that characterize the best linear approximation to the model response for each stimulus intensity.

For all visual intensities, the weighted linear fit was a good approximation to responses of model neurons, with average R^2 values of 0.98, 0.96 and 0.96 for simulated coherences of 25, 50 and 100%, respectively. Notably, different values of w_{vest} and w_{vis} were required to fit the data for different coherences. Specifically, w_{vest} decreased with coherence (**Fig. 6d**) and w_{vis} increased with coherence (**Fig. 6e**). The slope of these dependencies was similar for all model neurons, whereas the absolute values of w_{vest} and w_{vis} varied somewhat with the modality dominance weights assigned to each neuron. To summarize this effect, we examined the average weight ratio, w_{vis}/w_{vest}, as a function of coherence, normalized to a value of 1 at 100% coherence (**Fig. 6f**). The results are similar to the data from area MSTd[11], including the fact that weight changes are similar for cells with congruent and opposite heading preferences (**Fig. 6d**,**e**). Because this result depends mainly on response saturation, it could also be predicted by other models that incorporate a saturating nonlinearity.

The effect of coherence on the visual and vestibular mixing weights can be derived from the equations of the normalization model, with a few simplifying assumptions (including $n = 1.0$). The mixing weights, w_{vest} and w_{vis}, can be expressed as (see Online Methods)

$$w_{\text{vest}} = \frac{\alpha + k \cdot c_{\text{vest}}}{\alpha + k \cdot (c_{\text{vest}} + c_{\text{vis}})}, w_{\text{vis}} = \frac{\alpha + k \cdot c_{\text{vis}}}{\alpha + k \cdot (c_{\text{vest}} + c_{\text{vis}})} \quad (4)$$

Clearly, w_{vest} declines as a function of visual intensity (c_{vis}), whereas w_{vis} rises as a function of c_{vis}. In our simulations (**Fig. 6**), the exponent (n) was 2.0. In this case, the mixing weights become functions of the modality dominance weights (d_{vest}, d_{vis}) as well as stimulus intensities, resulting in vertical shifts among the curves (**Fig. 6d**,**e**).

In summary, normalization simply and elegantly accounts for the apparent changes in mixing weights exhibited by MSTd neurons as coherence was varied[11]. For any particular combination of stimulus intensities, the behavior of the normalization model can be approximated as linear summation, but the effective mixing weights appear to change with stimulus intensity as a result of changes in the net activity of the normalization pool.

DISCUSSION

We propose that divisive normalization can explain many fundamental response properties of multisensory neurons, including the empirical principles described in seminal work on the superior colliculus[4,5] and the effect of cue reliability on the neural combination rule in area MSTd[11]. The normalization model is attractive because it relies on relatively simple and biologically plausible operations[36]. Thus, the same basic operations that account for stimulus interactions in visual cortex[17–19] and attentional modulation[20] may also underlie various nonlinear interactions exhibited by multisensory neurons. The normalization model may therefore provide a good computational foundation for understanding multisensory cue integration.

Critical comparison with other models

Despite decades of research, quantitative models of multisensory integration have only recently been proposed[28,29,37–40]. One of the first mechanistic models of multisensory integration[37,40] is a compartmental model of single neurons that accounts for inverse effectiveness and sub-additive interactions between inputs from the same modality. It was also constructed to account for the modulatory effect of top-down cortical input on multisensory integration in the superior colliculus[41,42]. This model shares some elements with ours: it includes a squaring nonlinearity that produces super-additivity for weak inputs and a shunting inhibition mechanism that divides the response by the net input to each compartment. Notably, this model does not incorporate interactions among neurons in the population. Thus, it cannot account for the spatial principle of multisensory integration or cross-modal suppression in unisensory neurons. In addition, this model cannot account for the effects of cue reliability on the neural combination rule, as seen in area MSTd[11].

In contrast with this compartmental model[37,40], a recent neural network architecture[28,29] incorporates lateral interactions among neurons with different receptive field locations. Similar to the

normalization model, this neural network model can account for inverse effectiveness and the spatial principle, but there are important conceptual differences between the two schemes. First, to produce inverse effectiveness, the neural network model[28,29] incorporates a sigmoidal output nonlinearity into each model neuron (see also ref. 39). In contrast, in the normalization model, response saturation at strong intensities arises from the balance of activity in the network, not from a fixed internal property of individual neurons. Second, although the neural network model can produce cross-modal suppression, it appears to do so only when the less effective input is no longer excitatory on its own[29], but rather becomes suppressive as a result of lateral connections that mediate subtractive inhibition. We verified this observation by simulating an alternative model containing the key structural features of the neural network model[28,29]. This alternative model only produces cross-modal suppression when the non-optimal input is no longer excitatory (**Supplementary Figs. 3** and **4**). Thus, the key testable prediction of the normalization model, that an excitatory non-optimal input can yield cross-modal suppression (**Fig. 3**), does not appear to be shared by other models of multisensory integration. The divisive nature of lateral interactions in the normalization model appears to be critical for this prediction. Indeed, the alternative model does not account for the VIP data[30] (**Fig. 4d**) as successfully as the normalization model (**Supplementary Fig. 8**).

A recent elaboration of the neural network model[38] incorporates a number of specific design features to account for the experimental observation[41–43] that inactivation of cortical areas in the cat gates multisensory enhancement by superior colliculus neurons. We have not attempted to account for these results in our normalization model, as it is intended to be a general model of multisensory integration, rather than a specific model of any one system.

A recent computational theory[15] has demonstrated that populations of neurons with Poisson-like spiking statistics can achieve Bayes-optimal cue integration if each multisensory neuron simply sums its inputs. Thus, nonlinear interactions such as divisive normalization are not necessary to achieve optimal cue integration. Because this theory[15] involves simple summation by neurons, independent of stimulus intensity, it cannot account for various empirical principles of multisensory integration discussed here, including the effects of cue reliability on the neural combination rule[11]. It is currently unclear what roles divisive normalization may have in a theory of optimal cue integration and this is an important topic for additional investigation.

Parallels with visual cortical phenomena

Divisive normalization was initially proposed[16] to account for response properties in primary visual cortex. Normalization has often been invoked to account for stimulus interactions in the responses of cortical neurons[17,19,33,44] and has been implicated recently in the modulatory effects of attention on cortical responses[20]. The apparent ubiquity of divisive normalization in neural circuits[36,45] makes normalization operating at the level of multisensory integration an attractive general model to account for cross-modal interactions.

Perhaps the clearest experimental demonstration of normalization comes from a recent study[18] that measured responses of V1 neurons to orthogonal sine-wave gratings of various contrasts. This study found that population responses to any pair of contrasts can be well fit by a weighted linear sum of responses to the individual gratings. However, as the relative contrasts of the gratings varied, a linear model with different weights was required to fit the data[18], as predicted by divisive normalization. This result closely parallels the finding[11] that the multisensory combination rule of MSTd neurons depends on the relative strengths of visual and vestibular inputs. Here, we found that multisensory normalization can account for analogous phenomena observed in multisensory integration[11].

In summary, empirical principles of multisensory integration have guided the field for many years[4], but a simple computational account of these principles has been lacking. We found that divisive normalization accounts for the classical empirical principles of multisensory integration as well as recent findings regarding the effects of cue reliability on cross-modal integration. The normalization model is appealing for its simplicity and because it invokes a functional operation that has been repeatedly implicated in cortical function. Moreover, the model makes a key prediction, that a non-optimal excitatory input can produce cross-modal suppression, which can be tested experimentally. Although this prediction has not yet been tested systematically, a careful inspection of published data[10,30,46] reveals some examples that may demonstrate cross-modal suppression by a non-optimal excitatory input, although it is generally not clear whether the non-optimal input is significantly excitatory. We systematically examined cross-modal suppression in area MSTd and preliminary results support the model predictions (T.O., D.E.A. & G.C.D., unpublished observations). Thus, normalization may provide a simple and elegant account of many phenomena in multisensory integration.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

We would like to thank R. Jacobs, A. Pouget, J. Drugowitsch, D. Barany, A. Anzai, T. Sanada, R. Sasaki and H. Kim for helpful discussions and comments on the manuscript. This work was supported by US National Institutes of Health R01 grants EY016178 to G.C.D. and EY019087 to D.E.A.

AUTHOR CONTRIBUTIONS

T.O. and G.C.D. conceived the original model design. T.O. performed all model simulations and data analyses. T.O., D.E.A. and G.C.D. refined the model design and its predictions. T.O., D.E.A. and G.C.D. wrote and edited the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Alais, D. & Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* **14**, 257–262 (2004).
2. Ernst, M.O. & Banks, M.S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).
3. Fetsch, C.R., Turner, A.H., DeAngelis, G.C. & Angelaki, D.E. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J. Neurosci.* **29**, 15601–15612 (2009).
4. Stein, B.E. & Stanford, T.R. Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* **9**, 255–266 (2008).
5. Meredith, M.A. & Stein, B.E. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* **56**, 640–662 (1986).
6. Perrault, T.J. Jr., Vaughan, J.W., Stein, B.E. & Wallace, M.T. Neuron-specific response characteristics predict the magnitude of multisensory integration. *J. Neurophysiol.* **90**, 4022–4026 (2003).
7. Stanford, T.R., Quessy, S. & Stein, B.E. Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J. Neurosci.* **25**, 6499–6508 (2005).
8. Meredith, M.A., Nemitz, J.W. & Stein, B.E. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.* **7**, 3215–3229 (1987).
9. Meredith, M.A. & Stein, B.E. Spatial determinants of multisensory integration in cat superior colliculus neurons. *J. Neurophysiol.* **75**, 1843–1857 (1996).

10. Kadunce, D.C., Vaughan, J.W., Wallace, M.T., Benedek, G. & Stein, B.E. Mechanisms of within- and cross-modality suppression in the superior colliculus. *J. Neurophysiol.* **78**, 2834–2847 (1997).
11. Morgan, M.L., DeAngelis, G.C. & Angelaki, D.E. Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* **59**, 662–673 (2008).
12. Duffy, C.J. MST neurons respond to optic flow and translational movement. *J. Neurophysiol.* **80**, 1816–1827 (1998).
13. Gu, Y., Watkins, P.V., Angelaki, D.E. & DeAngelis, G.C. Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J. Neurosci.* **26**, 73–85 (2006).
14. Gu, Y., Angelaki, D.E. & DeAngelis, G.C. Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* **11**, 1201–1210 (2008).
15. Ma, W.J., Beck, J.M., Latham, P.E. & Pouget, A. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**, 1432–1438 (2006).
16. Heeger, D.J. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **9**, 181–197 (1992).
17. Carandini, M., Heeger, D.J. & Movshon, J.A. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* **17**, 8621–8644 (1997).
18. Busse, L., Wade, A.R. & Carandini, M. Representation of concurrent stimuli by population activity in visual cortex. *Neuron* **64**, 931–942 (2009).
19. Simoncelli, E.P. & Heeger, D.J. A model of neuronal responses in visual area MT. *Vision Res.* **38**, 743–761 (1998).
20. Reynolds, J.H. & Heeger, D.J. The normalization model of attention. *Neuron* **61**, 168–185 (2009).
21. Priebe, N.J. & Ferster, D. Mechanisms underlying cross-orientation suppression in cat visual cortex. *Nat. Neurosci.* **9**, 552–561 (2006).
22. Abbott, L.F., Varela, J.A., Sen, K. & Nelson, S.B. Synaptic depression and cortical gain control. *Science* **275**, 220–224 (1997).
23. Priebe, N.J., Mechler, F., Carandini, M. & Ferster, D. The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nat. Neurosci.* **7**, 1113–1122 (2004).
24. Perrault, T.J. Jr., Vaughan, J.W., Stein, B.E. & Wallace, M.T. Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. *J. Neurophysiol.* **93**, 2575–2586 (2005).
25. Heeger, D.J. Half-squaring in responses of cat striate cells. *Vis. Neurosci.* **9**, 427–443 (1992).
26. Alvarado, J.C., Vaughan, J.W., Stanford, T.R. & Stein, B.E. Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *J. Neurophysiol.* **97**, 3193–3205 (2007).
27. Meredith, M.A. & Stein, B.E. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res.* **365**, 350–354 (1986).
28. Magosso, E., Cuppini, C., Serino, A., Di Pellegrino, G. & Ursino, M. A theoretical study of multisensory integration in the superior colliculus by a neural network model. *Neural Netw.* **21**, 817–829 (2008).
29. Ursino, M., Cuppini, C., Magosso, E., Serino, A. & di Pellegrino, G. Multisensory integration in the superior colliculus: a neural network model. *J. Comput. Neurosci.* **26**, 55–73 (2009).
30. Avillac, M., Ben Hamed, S. & Duhamel, J.R. Multisensory integration in the ventral intraparietal area of the macaque monkey. *J. Neurosci.* **27**, 1922–1932 (2007).
31. DeAngelis, G.C., Robson, J.G., Ohzawa, I. & Freeman, R.D. Organization of suppression in receptive fields of neurons in cat visual cortex. *J. Neurophysiol.* **68**, 144–163 (1992).
32. Morrone, M.C., Burr, D.C. & Maffei, L. Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proc. R. Soc. Lond. B Biol. Sci.* **216**, 335–354 (1982).
33. Britten, K.H. & Heuer, H.W. Spatial summation in the receptive fields of MT neurons. *J. Neurosci.* **19**, 5074–5084 (1999).
34. Recanzone, G.H., Wurtz, R.H. & Schwarz, U. Responses of MT and MST neurons to one and two moving objects in the receptive field. *J. Neurophysiol.* **78**, 2904–2915 (1997).
35. Gu, Y., Fetsch, C.R., Adeyemo, B., DeAngelis, G.C. & Angelaki, D.E. Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron* **66**, 596–609 (2010).
36. Kouh, M. & Poggio, T. A canonical neural circuit for cortical nonlinear operations. *Neural Comput.* **20**, 1427–1451 (2008).
37. Alvarado, J.C., Rowland, B.A., Stanford, T.R. & Stein, B.E. A neural network model of multisensory integration also accounts for unisensory integration in superior colliculus. *Brain Res.* **1242**, 13–23 (2008).
38. Cuppini, C., Ursino, M., Magosso, E., Rowland, B.A. & Stein, B.E. An emergent model of multisensory integration in superior colliculus neurons. *Front. Integr. Neurosci.* **4**, 6 (2010).
39. Patton, P.E. & Anastasio, T.J. Modeling cross-modal enhancement and modality-specific suppression in multisensory neurons. *Neural Comput.* **15**, 783–810 (2003).
40. Rowland, B.A., Stanford, T.R. & Stein, B.E. A model of the neural mechanisms underlying multisensory integration in the superior colliculus. *Perception* **36**, 1431–1443 (2007).
41. Wallace, M.T. & Stein, B.E. Cross-modal synthesis in the midbrain depends on input from cortex. *J. Neurophysiol.* **71**, 429–432 (1994).
42. Jiang, W., Wallace, M.T., Jiang, H., Vaughan, J.W. & Stein, B.E. Two cortical areas mediate multisensory integration in superior colliculus neurons. *J. Neurophysiol.* **85**, 506–522 (2001).
43. Alvarado, J.C., Stanford, T.R., Rowland, B.A., Vaughan, J.W. & Stein, B.E. Multisensory integration in the superior colliculus requires synergy among corticocollicular inputs. *J. Neurosci.* **29**, 6580–6592 (2009).
44. Tolhurst, D.J. & Heeger, D.J. Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex. *Vis. Neurosci.* **14**, 293–309 (1997).
45. Olsen, S.R., Bhandawat, V. & Wilson, R.I. Divisive normalization in olfactory population codes. *Neuron* **66**, 287–299 (2010).
46. Sugihara, T., Diltz, M.D., Averbeck, B.B. & Romanski, L.M. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J. Neurosci.* **26**, 11138–11147 (2006).

ONLINE METHODS

Two different versions of the normalization model were simulated: one to model multisensory spatial integration in the superior colliculus and area VIP (**Fig. 1**), and the other to model visual-vestibular integration of heading signals in area MSTd (**Supplementary Figs. 6** and **7**).

Spatial model: primary sensory neurons. Each unimodal input to the spatial integration model is specified by its intensity c and its spatial position in Cartesian coordinates, $\theta = (x_\theta, y_\theta)$. The spatial receptive field of each primary sensory neuron is modeled as a two-dimensional Gaussian

$$G(\hat{\theta};\theta) = \exp\frac{-(\hat{\theta}-\theta)^2}{2\sigma^2} \quad (5)$$

where $\hat{\theta} = (x_{\hat{\theta}}, y_{\hat{\theta}})$ represents the center location of the receptive field. Arbitrarily, $x_{\hat{\theta}}, y_{\hat{\theta}}$ take integer values between 1 and 29, such that there are 29 × 29 = 841 sensory neurons with distinct $\hat{\theta}$ in each primary sensory layer. The size of the receptive field, given by σ, was chosen to be 2 (arbitrary units).

The response of each primary sensory neuron was assumed to scale linearly with stimulus intensity, c, such that the response can be expressed as

$$c \cdot G(\hat{\theta};\theta) \quad (6)$$

In addition, we assume that two inputs of the same sensory modality, presented at spatial positions θ_{1a}, θ_{1b}, interact linearly such that the net response is given by

$$c_{1a} \cdot G(\hat{\theta};\theta_{1a}) + c_{1b} \cdot G(\hat{\theta};\theta_{1b}) \quad (7)$$

where c_{1a}, c_{1b} represent the intensities of the two inputs.

We further assume that the linear response in each unisensory pathway is transformed by a nonlinearity, $h(x)$, such that the unisensory input is given by

$$h(c \cdot G(\hat{\theta};\theta)) \quad (8)$$

We used a sublinearly increasing function, $h(x) = \sqrt{x}$ to model this nonlinearity, although other monotonic functions such as $\log(x+1)$ or $\frac{x}{x+1}$ appear to work equally well. This nonlinearity models the sublinear intensity response functions often seen in sensory neurons[47]. It might reflect synaptic depression[22] at the synapse to the multisensory neuron or normalization operating in the unisensory pathways. This input nonlinearity, $h(x)$, has little effect on the multisensory integration properties of model neurons, but it is important for responses to multiple unisensory inputs (**Fig. 5**).

Spatial model: multisensory neurons. Each multisensory neuron in the model receives inputs from primary neurons of each sensory modality, as denoted by a subscript (1 or 2). The multisensory neuron performs a weighted linear sum of the unisensory inputs

$$E(d_1,\hat{\theta}_1,d_2,\hat{\theta}_2;c_1,\theta_1,c_2,\theta_2) = d_1 \cdot h(c_1 \cdot G_1(\hat{\theta}_1;\theta_1)) + d_2 \cdot h(c_2 \cdot G_2(\hat{\theta}_2;\theta_2)) \quad (9)$$

The modality dominance weights, d_1 and d_2, are fixed parameters of each multisensory neuron and each weight takes one of five values: 1.0, 0.75, 0.5, 0.25 or 0. Thus, 5 × 5 = 25 multisensory neurons with distinct combinations of modality dominance weights are included for each set of unisensory inputs. The linear response of the i^{th} neuron E_i (equation (9)) is then subjected to an expansive power-law output nonlinearity and divisively normalized by the net response of all other units, to obtain the final output (R_i) of each neuron

$$R_i = \frac{E_i^n}{\alpha^n + \left(\frac{1}{N}\right)\sum_{j=1}^{N} E_j^n} \quad (10)$$

Here, α is a semi-saturation constant (fixed at 1.0), N is the total number of multisensory neurons, and n is the exponent of the power-law nonlinearity that represents the relationship between membrane potential and firing rate[17,23], The exponent, n, was assumed to be 2.0 in our simulations, except where noted (**Fig. 2d**). Model responses were simulated for the following stimulus intensities: c_1, c_2 = 0, 1, 2, 4, 8, 16, 32, 64, 128, 256, 512 and 1,024.

To compare model responses to the physiological literature, an additivity index (AI) was computed as the ratio of the bimodal response to the sum of the unimodal responses

$$AI = \frac{R_{\text{bimodal}}}{R_{\text{unimodal 1}} + R_{\text{unimodal 2}}} \quad (11)$$

$R_{\text{unimodal 1}}$, $R_{\text{unimodal 2}}$ are obtained by setting one of the stimulus intensities, c_1 or c_2, to zero.

In the simulations of **Figures 2–5**, the receptive fields of the two primary sensory neurons projecting to a multisensory neuron were assumed to be spatially congruent ($\hat{\theta}_1 = \hat{\theta}_2$). In total, there were 841 (receptive field locations) × 25 (modality dominance weight combinations) = 21,025 distinct units in the multisensory layer.

For the simulations of **Figure 2f**, the exponent (n) was fixed at 1.5, and responses were generated for all nine combinations of three dominance weights: d_1, d_2 = 0.50, 0.75 or 1.00. Five neurons having each combination of dominance weights were simulated, for a total population of 45 neurons (similar to that recorded previously[26]), Responses were computed for five stimulus intensities (4, 16, 64, 256 and 1,024), Poisson noise was added and eight repetitions of each stimulus were simulated. A z score metric of response additivity was then computed using a bootstrap method[7,26]. Within-modal responses were also simulated (**Fig. 2f**) for pairs of stimuli of the same sensory modality, one of which was presented in the receptive field center while the other was offset from the receptive field center by 1σ.

For the simulations of **Figure 4d**, responses were generated for model neurons with all possible combinations of five dominance weights (d_1,d_2 = 0, 0.25, 0.50, 0.75 or 1.00) except $d_1 = d_2 = 0$, and semi-saturation constants (α) taking values of 1, 2, 4, 8 or 16. The exponent (n) was fixed at 2.5 for this simulation. Two cross-modal stimuli with intensity = 1,024 were presented at the receptive field center to generate responses.

To replot published data (**Figs. 2e** and **4d**), a high-resolution scan of the original figure was acquired and the data were digitized using software (Engauge, http://digitizer.sourceforge.net/). To replot the experimental data in **Figure 3d**, peristimulus time histograms from the original figure were converted into binary matrices using the 'imread' function in MATLAB (MathWorks). Spike counts were tallied from the digitized peristimulus time histograms and used to compute the enhancement index.

Visual-vestibular heading model. To simulate multisensory integration of heading signals in macaque area MSTd (**Supplementary Fig. 6**), the normalization model was modified to capture the basic physiological properties of MSTd neurons. Because heading is a circular variable in three dimensions[13], responses of the unisensory neurons (visual, vestibular) were modeled as

$$c \cdot \frac{1}{100} \cdot \frac{1+\cos(\phi)}{2} + \xi \cdot \frac{100-c}{100} \quad (12)$$

Here, c represents stimulus intensity (for example, the coherence of visual motion) ranging from 0–100, and ϕ represents the angle between the heading preference of the neuron and the stimulus heading. ϕ can be expressed in terms of the azimuth ($\hat{\varphi}$) and the elevation ($\hat{\theta}$) components of the heading preference, as well as azimuth (φ) and elevation (θ) components of the stimulus

$$\Phi = \arccos(\hat{\mathbf{H}} \cdot \mathbf{H}) \quad (13)$$

where $\hat{\mathbf{H}} = [\cos\hat{\theta}\cdot\cos\hat{\varphi}, \cos\hat{\theta}\cdot\sin\hat{\varphi}, \sin\hat{\theta}]$ and $\mathbf{H} = [\cos\theta\cdot\cos\varphi, \cos\theta\cdot\sin\varphi, \sin\theta]$. The dot operator denotes the inner product of the two vectors.

In the spatial model (equations (5–10)), we assumed that stimulus intensity multiplicatively scales the receptive field of model neurons. However, in areas MT and MST, motion coherence has a different effect on directional tuning[48,49]. With increasing coherence, the amplitude of the tuning curve scales roughly linearly, but the baseline response decreases[48]. To model this effect (see **Supplementary Fig. 7d**), we included the right-hand term in equation (12), with ξ in the range from 0 to 0.5 (typically 0.1), such that total population activity is an increasing function of c. However, our conclusions regarding changes in mixing weights with coherence (**Fig. 6d–f**) do not depend appreciably on the value of ξ.

Each model MSTd neuron performs a linear summation of its visual and vestibular inputs

$$E = d_{vest} \cdot (c_{vest} \cdot \frac{1+\cos(\phi_{vest})}{2} + \xi \cdot \frac{100 - c_{vest}}{100}) + d_{vis} \cdot (c_{vis} \cdot \frac{1+\cos(\phi_{vis})}{2} + \xi \cdot \frac{100 - c_{vis}}{100}) \quad (14)$$

where d_{vest}, d_{vis} are the modality dominance weights which take values from the array [1.0, 0.75, 0.5, 0.25, 0] In equation (14), the input nonlinearity $h(x)$ was omitted for simplicity, because it has little effect on the multisensory integration properties of model neurons, including the simulations of **Figure 6** (data not shown).

In area MSTd, many neurons have heading tuning that is not matched for visual and vestibular inputs[13]. Thus, the model included multisensory neurons with both congruent and mismatched heading preferences. Our model also incorporated the fact that there are more neurons tuned to the lateral self-motion than fore-aft motion[13,35]. Specifically, two random vector variables, $(\hat{\varphi}_{vest}, \hat{\theta}_{vest})$ and $(\hat{\varphi}_{vis}, \hat{\theta}_{vis})$, were generated to mimic the experimentally observed distributions of heading preferences (**Supplementary Fig. 7a**,**b**) and visual and vestibular heading preferences were then paired randomly (200 pairs). To mimic the finding that more neurons have congruent or opposite heading preferences than expected by chance[13], we added 28 units with congruent heading preferences, and another 28 with opposite preferences (**Supplementary Fig. 7c**). Combining these factors, a population of 256 (heading preference combinations) × 25 (dominance weight combinations) = 6,400 units comprised our MSTd model. The linear response of each unit (equation (14)) was squared and normalized by net population activity, as in equation (10). The semi-saturation constant was fixed at 0.05.

In the simulations of **Figure 6**, motion coherence took on values of 25, 50 and 100%, while the intensity of the vestibular input was fixed at 50, reflecting the fact that vestibular responses are generally weaker than visual responses in MSTd[50]. In equation (3), baseline activity ($c_{vis} = c_{vest} = 0$) was subtracted from each response before the linear model was fit. In MSTd[11], the effect of coherence on the mixing weights, w_{vest} and w_{vis}, did not depend on the congruency of visual and vestibular heading preferences. To examine this in our model (**Fig. 6d**–**f**), we present results for three particular combinations of heading preferences, congruent ($\hat{\varphi}_{vest} = 90°, \hat{\varphi}_{vis} = 90°$), opposite ($\hat{\varphi}_{vest} = 90°, \hat{\varphi}_{vis} = 270°$) and intermediate ($\hat{\varphi}_{vest} = 90°, \hat{\varphi}_{vis} = 180°$), all in the horizontal plane ($\hat{\theta}_{vest} = \hat{\theta}_{vis} = 0°$). Note, however, that all possible congruencies are present in the model population (**Supplementary Fig. 7c**). For each congruency type, all possible combinations of modality dominance weights from the set (d_1, d_2 = 1.0, 0.75, 0.5, 0.25) were used. Thus, we simulated responses for a total of 3 (congruency types) × 16 (dominance weight combinations) = 48 model neurons (**Fig. 6d**–**f**). For each congruency type, simulation results were sorted into three groups according to the ratio of dominance weights: $d_{vest}/d_{vis} \le 0.5$, $0.5 < d_{vest}/d_{vis} < 2.0$, or $2.0 \le d_{vest}/d_{vis}$. Data were averaged within each group. Thus, results are presented (**Fig. 6d**–**f**) as nine curves, corresponding to all combinations of three congruency types and three weight ratio groups.

Derivation of the cue-reweighting effect. Here, we derive the expression for the effective mixing weights of the model MSTd neurons (equation (4)) through some simplification and rearrangement of the equations for the normalization model. Assuming that the exponent (n) is 1 and that tuning curves simply scale with stimulus intensity ($\xi = 0$), the net population activity in response to a vestibular heading (φ_{vest}) with intensity c_{vest} and a visual heading (φ_{vis}) with intensity φ_{vis}, can be expressed as

$$\sum_{d_{vest}} \sum_{\hat{\varphi}_{vest}} \sum_{d_{vis}} \sum_{\hat{\varphi}_{vis}} (d_{vest} \cdot c_{vest} \cdot F_{vest}(\hat{\varphi}_{vest}; \varphi_{vest}) + d_{vis} \cdot c_{vis} \cdot F_{vis}(\hat{\varphi}_{vis}; \varphi_{vis})) = \sum_{d_{vest}} \sum_{\hat{\varphi}_{vest}} \sum_{d_{vis}} \sum_{\hat{\varphi}_{vis}} (d_{vest} \cdot c_{vest} \cdot F_{vest}(\hat{\varphi}_{vest}; \varphi_{vest})) + \sum_{d_{vest}} \sum_{\hat{\varphi}_{vest}} \sum_{d_{vis}} \sum_{\hat{\varphi}_{vis}} (d_{vis} \cdot c_{vis} \cdot F_{vis}(\hat{\varphi}_{vis}; \varphi_{vis})) \quad (15)$$

where $F_{vest}(\hat{\varphi}_{vest}; \varphi_{vest})$, $F_{vis}(\hat{\varphi}_{vis}; \varphi_{vis})$ represent vestibular and visual tuning curves with heading preferences at $\hat{\varphi}_{vest}$ and $\hat{\varphi}_{vis}$, respectively. Because heading preferences and dominance weights are randomly crossed in the model population, we can assume that this expression can be factorized

$$c_{vest} \cdot \left(\sum_{d_{vis}} 1\right) \cdot \left(\sum_{\hat{\varphi}_{vis}} 1\right) \cdot \left(\sum_{d_{vest}} d_{vest}\right) \cdot \left(\sum_{\hat{\varphi}_{vest}} F_{vest}(\hat{\varphi}_{vest}; \varphi_{vest})\right) + c_{vis} \cdot \left(\sum_{d_{vest}} 1\right) \cdot \left(\sum_{\hat{\varphi}_{vest}} 1\right) \cdot \left(\sum_{d_{vis}} d_{vis}\right) \cdot \left(\sum_{\hat{\varphi}_{vis}} F_{vis}(\hat{\varphi}_{vis}; \varphi_{vis})\right) \quad (16)$$

If we make the additional simplification that heading preferences, $\hat{\varphi}_{vest}$ and $\hat{\varphi}_{vis}$, are uniformly distributed, then terms involving sums of tuning curves, $\sum_{\hat{\varphi}_{vest}} F_{vest}(\hat{\varphi}_{vest}; \varphi_{vest})$ and $\sum_{\hat{\varphi}_{vis}} F_{vis}(\hat{\varphi}_{vis}; \varphi_{vis})$ become constants. Moreover, the summations $\sum_{d_{vest}} d_{vest}$ and $\sum_{d_{vis}} d_{vis}$ are also constants because d_{vest}, d_{vis} are fixed parameters in the model. With these simplifications, equation (15) can be expressed as

$$k \cdot (c_{vest} + c_{vis}) \quad (17)$$

where k is a constant that incorporates the sums of tuning curves and dominance weights. The bimodal response of a model neuron can now be expressed as

$$R_{bimodal}(\varphi_{vest}, \varphi_{vis}) = \frac{d_{vest} \cdot c_{vest} \cdot F_{vest}(\hat{\varphi}_{vest}; \varphi_{vest}) + d_{vis} \cdot c_{vis} \cdot F_{vis}(\hat{\varphi}_{vis}; \varphi_{vis})}{\alpha + k \cdot (c_{vest} + c_{vis})} \quad (18)$$

Unimodal responses can be obtained by setting one of the stimulus intensities, c_{vest} or c_{vis}, to zero

$$R_{vest}(\varphi_{vest}) = \frac{d_{vest} \cdot c_{vest} \cdot F_{vest}(\hat{\varphi}_{vest}; \varphi_{vest})}{\alpha + k \cdot c_{vest}}$$
$$R_{vis}(\varphi_{vis}) = \frac{d_{vis} \cdot c_{vis} \cdot F_{vis}(\hat{\varphi}_{vis}; \varphi_{vis})}{\alpha + k \cdot c_{vis}} \quad (19)$$

With these simplifications, the bimodal response (equation (18) can be expressed as a weighted linear sum of the unimodal responses, with weights that depend on stimulus intensity

$$R_{bimodal}(\varphi_{vest}, \varphi_{vis}) = \frac{\alpha + k \cdot c_{vest}}{\alpha + k \cdot (c_{vest} + c_{vis})} \cdot R_{vest}(\varphi_{vest}) + \frac{\alpha + k \cdot c_{vis}}{\alpha + k \cdot (c_{vest} + c_{vis})} \cdot R_{vis}(\varphi_{vis}) \quad (20)$$

Comparing equations (3) and (20), the closed forms of the mixing weights (equation (4)) are obtained.

47. Albrecht, D.G. & Hamilton, D.B. Striate cortex of monkey and cat: contrast response function. *J. Neurophysiol.* **48**, 217–237 (1982).
48. Britten, K.H. & Newsome, W.T. Tuning bandwidths for near-threshold stimuli in area MT. *J. Neurophysiol.* **80**, 762–770 (1998).
49. Heuer, H.W. & Britten, K.H. Linear responses to stochastic motion signals in area MST. *J. Neurophysiol.* **98**, 1115–1124 (2007).
50. Gu, Y., DeAngelis, G.C. & Angelaki, D.E. A functional link between area MSTd and heading perception based on vestibular signals. *Nat. Neurosci.* **10**, 1038–1047 (2007).

doi:10.1038/nn.2815

it is therefore likely that both ARC and AVPV neurons were affected by a treatment at day 20. There are other lines of investigation worthy of pursuit. What are the neurobiological mechanisms for compensation during development and are these similar to or distinct from those that maintain some degree of fertility after adult-induced GPR54 ablation? What is the contribution to phenotype of ablating kisspeptin- and GPR54-expressing cells outside of the nervous system? Does adult ablation of GPR54 signaling alter the response of the reproductive system to steroid feedback or to other cues that have been controversially proposed to be conveyed to GnRH neurons by kisspeptin, such as metabolic cues[14,15]? Can the roles of ARC and AVPV kisspeptin be more clearly parsed? What is the effect of these targeted neuronal ablations on puberty and fertility in male mice? And, of course, does this finding translate to humans?

The pubertal onset of proper GnRH neuron function is vital to the survival of vertebrate species. Given this, the persistent failure to find a unique initiator of puberty is not surprising. Multiple GnRH subpopulations with different activating afferent networks would be advantageous to survival; traits that can enhance fertility are often selected for. Redundancy in the control of GnRH neurons does not take away from the importance of any of the mediators involved, including kisspeptin. All are parts of an intricate network that weighs many factors, both proximal and distal, to determine whether fertility is appropriate for an individual under the particular conditions to which that individual is exposed; this network ultimately produces appropriately patterned GnRH release. The next phase of this journey is to understand how all of these factors interact with the intrinsic properties of GnRH neurons to determine their output.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Mayer, C. & Boehm, U. *Nat. Neurosci.* **14**, 704–710 (2011).
2. Moenter, S.M., Brand, R.M., Midgley, A.R. & Karsch, F.J. *Endocrinology* **130**, 503–510 (1992).
3. Seminara, S.B. *et al. N. Engl. J. Med.* [see comment] **349**, 1614–1627 (2003).
4. de Roux, N. *et al. Proc. Natl. Acad. Sci. USA* **100**, 10972–10976 (2003).
5 Oakley, A.E., Clifton, D.K. & Steiner, R.A. *Endocr. Rev.* **30**, 713–743 (2009).
6. Lapatto, R. *et al. Endocrinology* **148**, 4927–4936 (2007).
7. d'Anglemont de Tassigny, X. *et al. Proc. Natl. Acad. Sci. USA* **104**, 10714–10719 (2007).
8. Chan, Y.M., Broder-Fingert, S., Wong, K.M. & Seminara, S.B. *J. Neuroendocrinol.* **21**, 1015–1023 (2009).
9. Gibson, M.J. *et al. Science* **225**, 949–951 (1984).
10. Raivio, T. *et al. N. Eng. J. Med.* **357**, 863–873 (2007).
11. Cravo, R.M. *et al. Neurosci.* **173**, 37–56 (2011).
12. DeFazio, R.A., Heger, S., Ojeda, S.R. & Moenter, S.M. *Mol. Endocrinol.* **16**, 2872–2891 (2002).
13. Kauffman, A.S., Navarro, V.M., Kim, J., Clifton, D. & Steiner, R.A. *Am. J. Physiol. Endocrinol. Metab.* **297**, 1212–1221 (2009).
14. Louis, G.W. *et al. Endocrinology* published online, doi:10.1210/en.2011-0096 (22 March 2011).
15. Smith, J.T., Acohido, B.V., Clifton, D.K. & Steiner, R.A. *J. Neuroendocrinol.* **18**, 298–303 (2006).

Normalizing relations between the senses

Anne K Churchland

An established framework that describes how visual neurons combine their inputs, divisive normalization, proves valuable in explaining multisensory processing in the superior colliculus and medial superior temporal area.

In the past century, sensory neuroscience has provided a basis for understanding the operation of individual senses, such as vision, hearing and touch. This approach has revealed much about how neurons in sensory areas corresponding to these modalities process information and ultimately drive perception. However, the approach leaves unexplored the very realistic scenario in which an object in the world activates more than one sensory system at the same time. When localizing a moving object, for example, both the sound and light emitted from the object provide cues about where it is (**Fig. 1**). Because the information provided by each sense alone is often noisy or unreliable, combining information across modalities, a process known as multisensory integration, affords the opportunity to improve one's estimate about the object in question. Despite its importance in sensory processing, much about the neural computations that underlie multisensory integration remain unknown. In this issue of *Nature Neuroscience*, Ohshiro *et al.*[1] describe a computational principle, divisive normalization, on which multisensory integration seems to rely[1].

Ohshiro *et al.*'s[1] observations were motivated by a foundation of knowledge about how the responses of single neurons change when they are faced with single-sensory versus multisensory inputs. This understanding developed largely from observations in the superior colliculus, an area in which auditory and visual inputs converge at the level of single neurons (**Fig. 1**). More recent observations come from the authors' own laboratories, where they have examined responses to visual and vestibular stimuli in the medial superior temporal area (MST)[2,3]. In both areas, neurons often respond more to multisensory input than to single-sensory input, a phenomenon known as multisensory enhancement.

These studies have led to a collection of principles that govern multisensory integration at the level of single neurons[4]. The principles include, for instance, the observation that the greatest degree of multisensory enhancement is observed when the stimulus presented to each individual sense is very weak (principle of inverse effectiveness). A second example is known as the spatial principle: an ineffective stimulus in one sensory modality (for example, a visual stimulus placed off the center of a superior colliculus neuron's receptive field) can suppress the response to a highly effective stimulus from the other modality (such as an auditory tone positioned in the optimal spatial location).

A similar observation to the spatial principle, and perhaps one that led Ohshiro *et al.*[1] to their eventual hypothesis, has been extensively studied in the visual system. In primary visual cortex, neurons are tuned for the orientation of lines or bars. Ineffective stimuli, such as bars that are orthogonal to the neuron's preferred orientation, can suppress the response to effective stimuli when the two are presented together[5]. This observation is particularly interesting when the ineffective stimulus is weakly excitatory when it is presented on its own. In that case, a simple model that linearly sums excitatory and inhibitory inputs would not predict any suppression. A straightforward mechanism, divisive normalization, is usually invoked to explain visual effects such as this. According to this mechanism, neurons receive

The author is at the Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA.
e-mail: churchland@cshl.edu

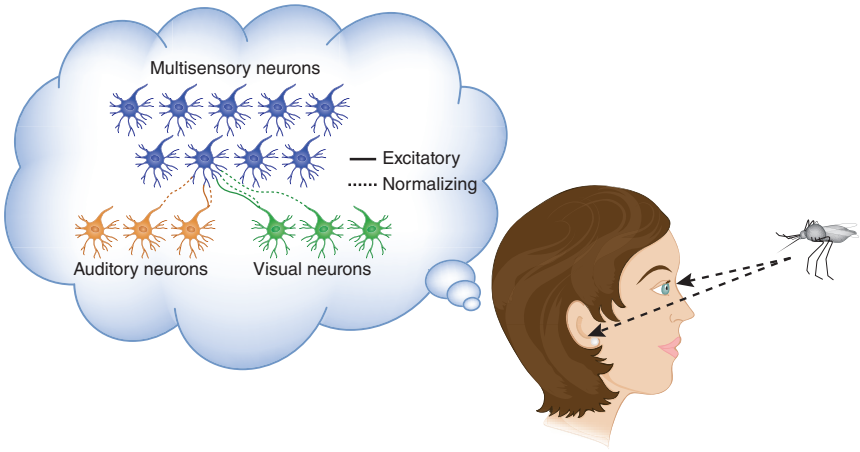


Figure 1 Neural pathways for multisensory integration. Visual and auditory information from a mosquito provide information about its location to an onlooker. Information from visual neurons (green) and auditory neurons (orange) provide excitatory drive (solid lines) to neurons in a multisensory area (such as the superior colliculus). A broader set, which includes both the excitatory neurons and some differently tuned neighbors, constitutes the 'normalization pool' that serves to mitigate the excitatory drive (dashed lines).

excitatory drive from a narrow range of inputs; this excitatory drive is mitigated (through normalization) by a broader pool of inputs (**Fig. 1**). In other words, a normalization pool containing neurons driven by all sorts of stimuli is used to divide a more narrowly tuned excitatory drive. Notably, diminishing a neuron's response through normalization is rather different from doing so through traditional inhibitory mechanisms such as subtraction. The normalization step serves to scale existing signals, rather than introduce a new signal in the way that inhibition does. This mechanism has been an essential component of models of visual processing for some time and has even been used to unite seemingly conflicting reports about visual attention[6]. In spite of its ubiquitous use in the visual system, however, it has been notably absent as a candidate mechanism for understanding interactions between vision and other sensory modalities.

Ohshiro *et al.*[1] used divisive normalization to explain a variety of effects seen in multisensory neurons in the superior colliculus and MST. The broadly tuned normalization pool turned out to be critical for explaining multisensory effects such as the spatial principle. In our example above, the visual stimulus was ineffective because it was not optimally positioned in the receptive field of the neuron being studied. Thus, it didn't elicit much response on its own. However, the same stimulus provided an optimal input to different neurons that, crucially, are included in the larger group of neurons that constitutes the normalization pool for the neuron being studied. These neurons serve to suppress the response, even when they supply only a weak excitatory drive. The same mechanism explains why neurons in multisensory areas that are almost entirely driven by one sensory stimulus can nevertheless show suppression when the two sensory modalities are presented together; the non-optimal modality drives neurons in the normalization pool that serve to suppress the excitatory inputs.

Previous models that have attempted to explain the spatial principle[7] have relied on the idea that the suppression results from the second modality eliciting an inhibitory response. However, this type of mechanism fails to explain how single-sensory inputs that are weakly excitatory on their own still serve to suppress the response to an optimal stimulus. Because previous experiments have not set out to directly test divisive normalization against simple inhibitory models, it is not clear whether cross-modal suppression by weakly excitatory stimuli is a reliable or robust effect. Future experiments that make this comparison will directly test whether divisive normalization is a necessary feature of multisensory integration.

Previous models that have attempted to explain the principle of inverse effectiveness[8,9] share some common features with the model proposed by Ohshiro *et al.*[1]. Specifically, these models explain the effect of using a squaring nonlinearity that causes weak inputs to have a superadditive effect. However, because previous models mainly rely on interactions of inputs in a single neuron, they can't explain the spatial principle or cross-modal suppression, both of which rely on interactions between populations of neurons with different tuning properties. Furthermore, the single-neuron models, at least in their present form, cannot account for recent data regarding changes in the weighting of sensory inputs in relation to their reliability[10]. The divisive normalization model accounts nicely for these findings; specifically, the model accurately predicts a complex pattern of weight changes that was observed[10] when a visual stimulus gradually changes in reliability alongside a vestibular stimulus of constant reliability.

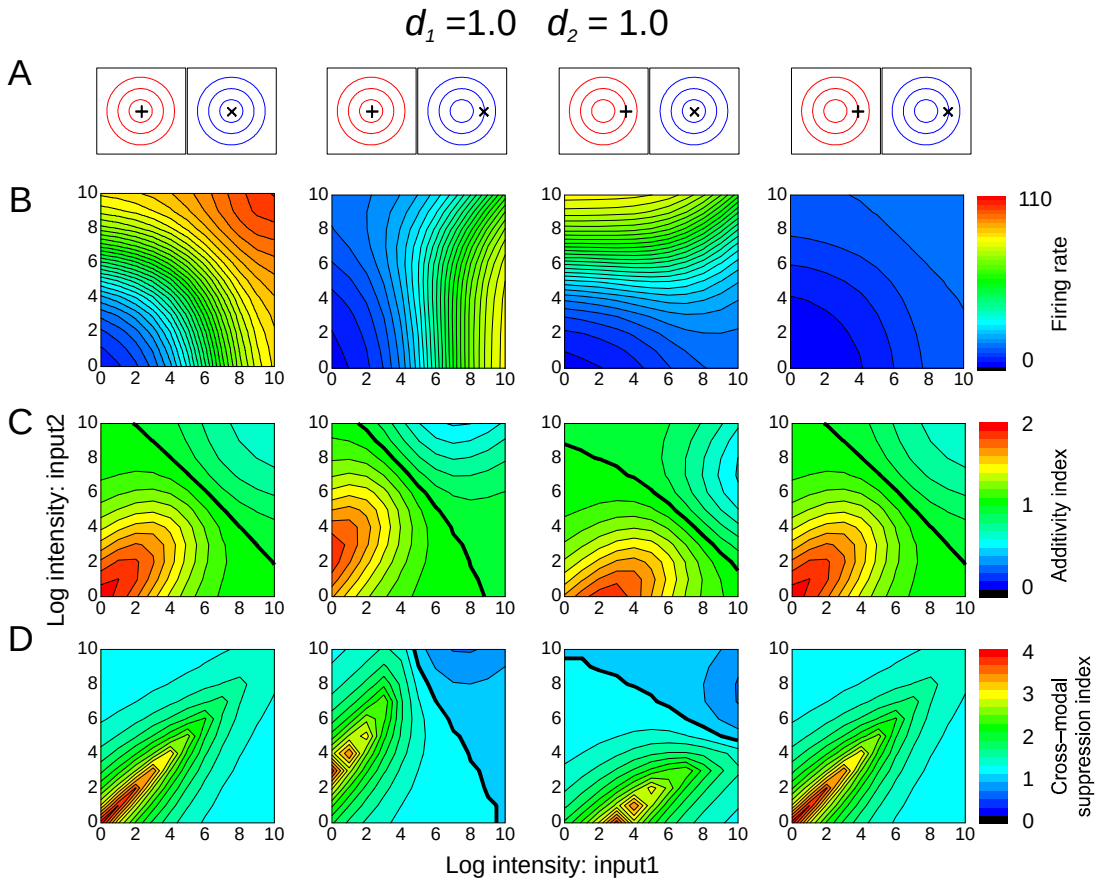
Insights into how the brain combines inputs from different modalities constitute a major advance toward a full understanding of perception. Ultimately, when we perceive the world, we not only combine information across sensory modalities, but also integrate that sensory information with our existing knowledge about the structure of the world, for example, our memories about specific objects and our predictions for their future behavior. An appealing possibility raised by the new findings is that signals such as these might be combined with sensory inputs using straightforward computational principles similar to those that have been uncovered by existing studies of sensory processing. Perhaps a prior belief about the statistics of the world (for example, an idea garnered from a life of experience about how fast mosquitoes tend to fly) contributes to both the excitatory inputs and the normalization pool and can be thought of as just another cue that is combined with the others to form a coherent percept of the world.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Ohshiro, T., Angelaki, D.E. & DeAngelis, G.C. *Nat. Neurosci.* **14**, 775–782 (2011).
2. Fetsch, C.R., Wang, S., Gu, Y., DeAngelis, G.C. & Angelaki, D.E. *J. Neurosci.* **27**, 700–712 (2007).
3. Gu, Y., Fetsch, C.R., Adeyemo, B., DeAngelis, G.C. & Angelaki, D.E. *Neuron* **66**, 596–609 (2010).
4. Stein, B.E. & Stanford, T.R. *Nat. Rev. Neurosci.* **9**, 255–266 (2008).
5. Carandini, M., Heeger, D.J. & Movshon, J.A. *J. Neurosci.* **17**, 8621–8644 (1997).
6. Reynolds, J.H. & Heeger, D.J. *Neuron* **61**, 168–185 (2009).
7. Cuppini, C., Magosso, E. & Ursino, M. *Biosystems* **96**, 195–205 (2009).
8. Alvarado, J.C., Rowland, B.A., Stanford, T.R. & Stein, B.E. *Brain Res.* **1242**, 13–23 (2008).
9. Rowland, B.A., Stanford, T.R. & Stein, B.E. *Perception* **36**, 1431–1443 (2007).
10. Morgan, M.L., DeAngelis, G.C. & Angelaki, D.E. *Neuron* **59**, 662–673 (2008).

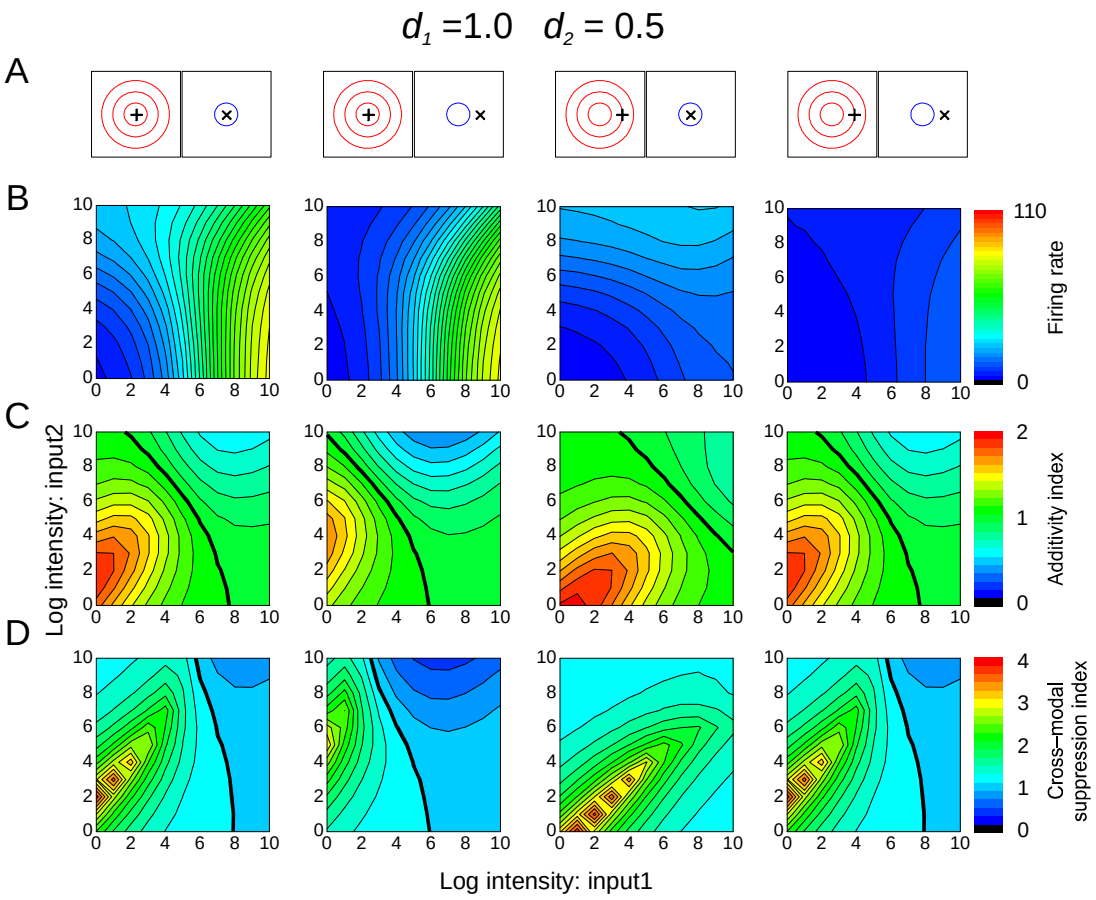
Supplemental Materials for "A Normalization Model of Multisensory Integration" by Ohshiro, Angelaki and DeAngelis

Supplementary Figures and Legends



Supplementary Figure 1. Inverse effectiveness and cross-modal suppression for stimuli that are centered or mis-centered on the unimodal receptive fields. (A) Schematic illustration of the locations of Input 1 ('+') and Input2 ('x') relative to the receptive fields (red and blue contours). Conventions as in Figure 3A. (B) Bimodal response of a model neuron with balanced dominance weights (d_1 = d_2 = 1.0) is plotted as a function of the stimulus intensities of Input1 and Input2. (C) AI is plotted as a function of stimulus intensity. A gradual change in additivity from super- to sub-additivity (inverse effectiveness) can be observed for all four combinations of input locations. Thick contour lines indicate AI=1. (D) The ratio of the bimodal response to the larger of the two unimodal responses (cross-modal suppression index) is plotted as a function of stimulus intensity. Thick contour lines indicate that the cross-modal suppression index = 1. Regions with ratios <1 denote portions of the stimulus space in which cross-modal suppression takes place. Note that the cross-modal suppression occurs when the effectiveness of the two unisensory inputs is substantially different (second and third columns) but not when the inputs are balanced (first and fourth columns).

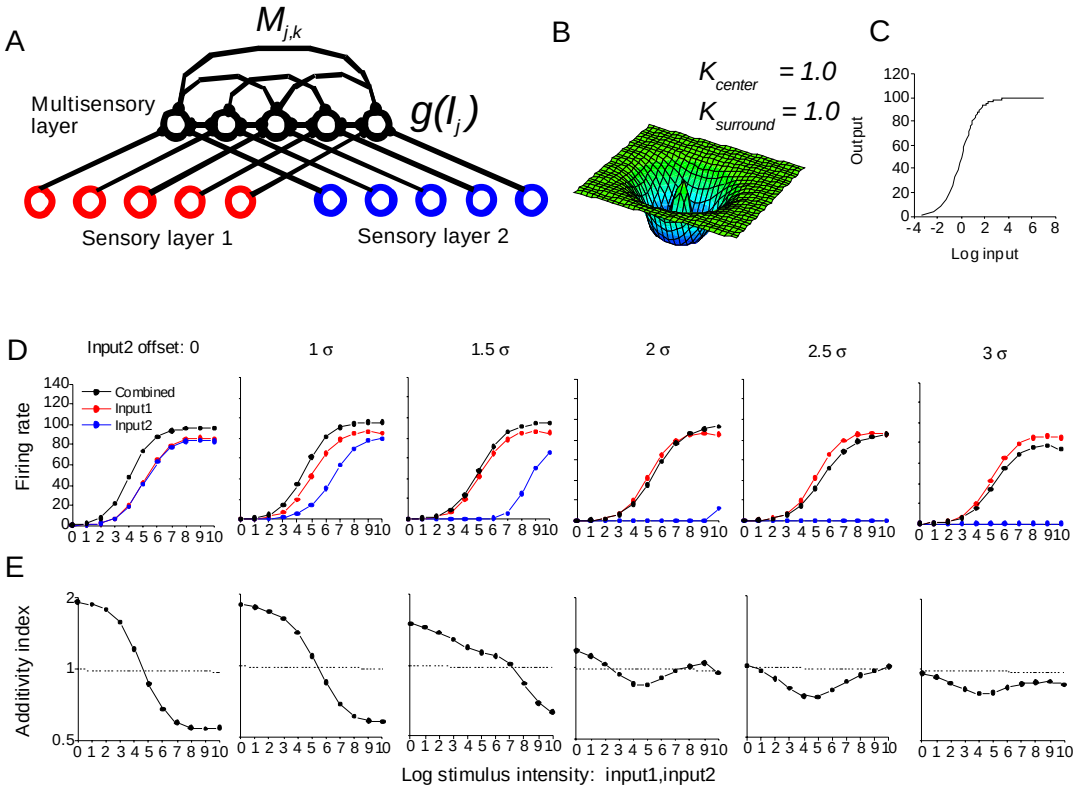
Nature Neuroscience: doi:10.1038/nn.2815



Supplementary Figure 2. Inverse effectiveness and cross-modal suppression for a model neuron with imbalanced modality dominance weights (d_1 = 1.0, d_2 = 0.5).

(A) The receptive fields and the stimulus locations relative to the receptive field are illustrated schematically. Format as in Supplementary Figure 1. (B) Bimodal response is plotted as a function of the stimulus intensities of Input1 and Input2. (C) AI is plotted as a function of stimulus intensities. Super-additive interactions and inverse effectiveness can be observed for weak stimulus intensities in all configurations. (D) Cross-modal suppression is exaggerated when the modality dominance weights are not balanced (compare the first two columns with those of Supplementary Fig. 1D). Note that cross-modal suppression is not observed in the third column, where the reduced dominance of Input 2 (blue contour) is matched by Input 1 being displaced from the center of the receptive field.

Nature Neuroscience: doi:10.1038/nn.2815



Supplementary Figure 3. Alternative multisensory integration model with a recurrent network architecture and subtractive inhibition

(A) Schematic illustration of the alternative model architecture that captures the basic design features of the multisensory integration models proposed by Magosso et al. [1] and Ursino et al. [2]. Each multisensory neuron (black circles) performs a weighted linear sum of inputs from primary sensory neurons (red, blue circles). Synaptic weights for the lateral connections between multisensory neurons are described by the weight matrix, $M_{k,j}$. (B) Illustration of the "Mexican hat" shaped weight matrix, $M_{k,j}$, which provides subtractive lateral inhibition. (C) Total synaptic input to the *j* th neuron (I_j) is transformed into firing rate by a sigmoidal activation function, *g*(). This nonlinearity, which can produce super-additivity for weak intensities and sub-additivity for strong intensities, is therefore built into each neuron in this model. (D) Responses of a model neuron to multisensory inputs with different degrees of offset between Input1 and Input2. Format as in Fig. 3B. The combined response (black) is suppressed below the unimodal response to Input1 (red) only when the Input2 is offset from the RF center by more than 2.5 x σ (columns 5 and 6), such that Input2 becomes inhibitory on its own. Note that crossmodal suppression does not occur while Input2 (blue) is excitatory. (E) Additivity Index (AI) is plotted as a function of stimulus intensity.

Methods: The temporal dynamics of the response are given by:

$$\tau * \frac{dI_j}{dt} = -I_j + E_j + \beta * \sum_k M_{k,j} * g(I_k) \quad \ldots\ldots \text{(S1)}$$

In this equation, I_j represents the synaptic current in the *j* th multisensory neuron; E_j, the linear sum of the two unisensory inputs as defined in Eqn. (9); $M_{k,j}$, the synaptic weights for lateral interactions between the *k* th, and *j* th multisensory units; $g(\,)$, a

Nature Neuroscience: doi:10.1038/nn.2815

sigmoidal activation function which transforms the synaptic current to the firing rate; and β, the coupling constant for the lateral interactions.

This alternative model and the normalization model both have primary sensory neurons with the same response properties, and both incorporate weighted linear summation of inputs (i.e., the term E_j) to the multisensory neurons. For simplicity, RFs for both modalities are spatially congruent, and the dominance weights (d_1, d_2) are set to 1.0 in the simulations of Supplementary Figs. 3 and 4. As the non-linear function $h()$ in E_j, we used a hyperbolic ratio function, $h(x) = 5 * x/(x + 128)$ to be consistent with the models of Ursino and colleagues.

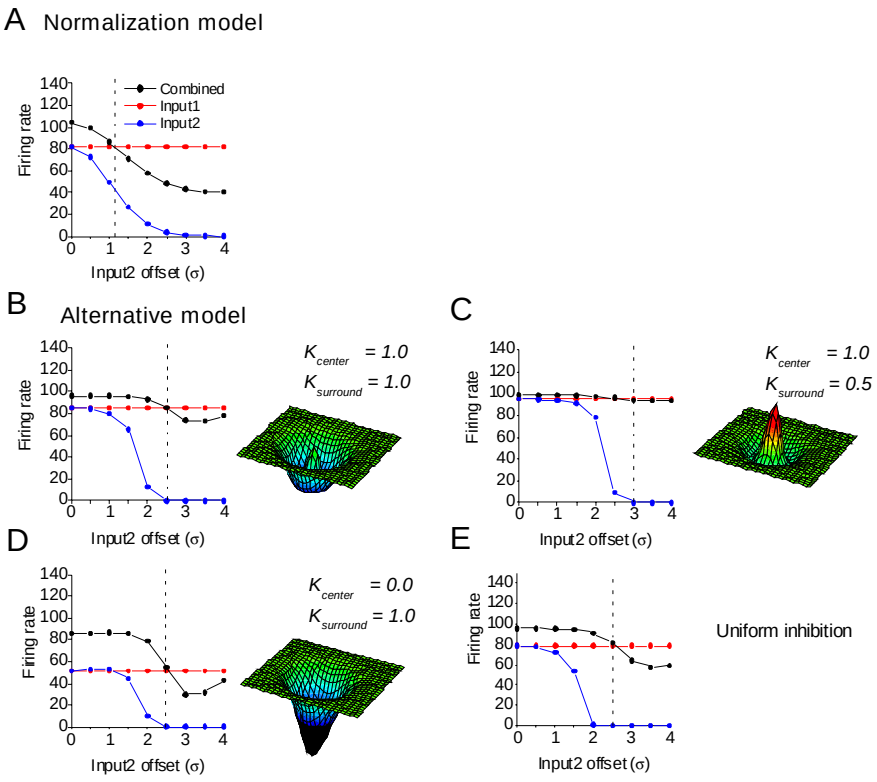
We assume that the synaptic weights for the lateral interactions between multisensory neurons follow a Mexican hat profile [1, 2] of the following form:

$$M_{k,j} = K_{center} * exp\{-D_{k,j}^2/(2 * \sigma_{center}^2)\} - K_{surround} * exp\{-D_{k,j}^2/(2 * \sigma_{surround}^2)\}$$

.....(S2)

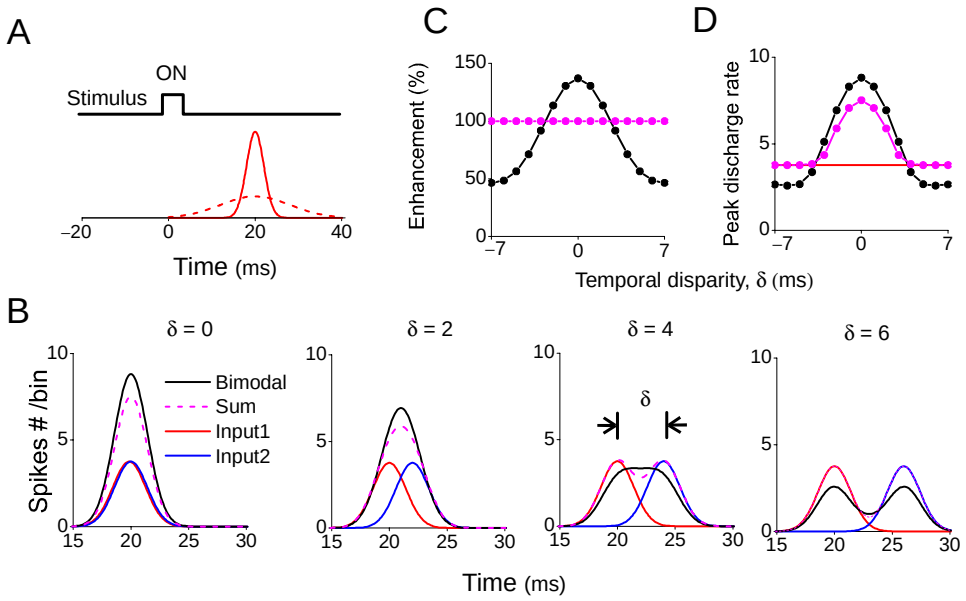
where, $D_{k,j}$ represents the topological distance between the *k* th, and *j* th unit. σ_{center} and $\sigma_{surround}$ are set to values of 2 and 4, respectively. K_{center} and $K_{surround}$ represent the strengths of the excitatory center component and the inhibitory surround component of $M_{k,j}$, respectively. Both are set to 1 as a default (but see Supplementary Figure 4). With these settings, the input arising through the lateral connections ($\sum_k M_{k,j} * g(I_k)$) is always negative (because $M_{k,j} \leq 0,\ g(I_k) \geq 0$); thus, it has an inhibitory influence on the unit's response. The coupling constant β was set to 0.2. We used a hyperbolic ratio function as the sigmoidal activation function, $g(x) = 100 * x^2/(x^2 + 1^2)\ for\ x \geq 0,\ and\ g(x) = 0\ otherwise$, although other formulations of a sigmoid would produce similar results. Note that, as Eqn. S1 shows, inhibition mediated through the lateral connections has a subtractive form. The excitatory input from primary neurons (E_j) is reduced by an amount proportional to the activity of surrounding neurons ($\sum_k M_{k,j} * g(I_k)$). This key difference in the mode of lateral interactions (subtractive versus divisive) results in the distinct behavior of the alternative model and the normalization model with respect to cross-modal suppression.

Nature Neuroscience: doi:10.1038/nn.2815



Supplementary Figure 4. Summary of crossmodal suppression results, comparing the normalization model and the alternative model.
Unimodal responses to Input1 (red) or Input2 (blue), along with combined responses (black), are plotted as a function of the offset of Input2 relative to the RF center. Input1 is always presented at the center of the RF. Stimulus intensity for both inputs is fixed at 1024. (A) Prediction of the normalization model. The combined response (black) clearly starts to be suppressed below the response to Input1 (red) when Input2 is offset from the RF center by more than roughly 1*σ (dashed vertical line). The unimodal response to Input2 is still clearly excitatory until the offset reaches roughly 3*σ. (B) Predictions of the alternative model. The combined response is suppressed below the response to Input1 only when the offset is more than 2.5*σ (dashed vertical line). Input2 is no longer excitatory for offsets > 2.5*σ. Thus, cross-modal suppression in the alternative model only occurs when the non-optimal Input2 becomes inhibitory as a result of the subtractive lateral inhibition. The profile of connection weights $M_{k,j}$ used in this simulation (same as in Supplementary Fig. 3) is depicted on the right. (C-E) Predictions of the alternative model for different forms of the lateral inhibition connection weights, $M_{k,j}$: one case with stronger excitatory interactions between nearby neurons (C), one case with exclusively inhibitory interactions between neurons (D), and one case with spatially uniform inhibitory interactions among all neurons in the population (E). Note that the behavior of the alternative model, with respect to cross-modal suppression, depends little on the particular form of $M_{k,j}$

Nature Neuroscience: doi:10.1038/nn.2815



Supplementary Figure 5. Normalization can account for the temporal principle of multisensory integration.

(A) Temporal profiles of the excitatory input to a model neuron (solid red curve) and the normalization signal (dashed red curve). The scaling of the time axis is arbitrary. We assume that response latency and duration vary among neurons, such that the normalization pool has a broader temporal response than individual units. (B) Model responses to bimodal stimuli presented at four temporal disparities (0, 2, 4, 6 ms) between the onset of the two stimuli. Red and blue curves show unimodal responses to Inputs 1 and 2, respectively. The black curve represents the bimodal response, and the dashed magenta curve shows the sum of the two unimodal responses. (C) The index of multisensory enhancement (defined below) declines with the temporal disparity of the two stimuli for model neurons (black curve), whereas the sum of the two unimodal responses is constant (magenta). (D) The peak discharge rate of the bimodal response (black) is suppressed below that of the unimodal response (red line) for large temporal disparities, whereas the peak discharge of the sum of unisensory responses (magenta) never falls below the red line.

Methods. The normalized bimodal response is modeled as:

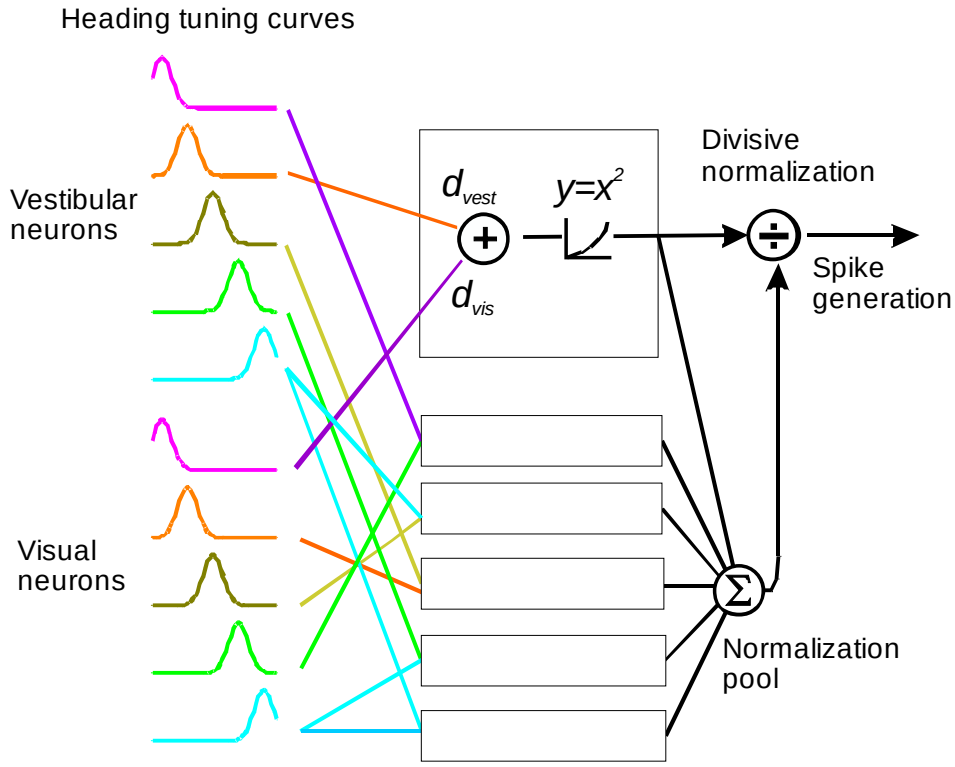
$$r(t) = \frac{\left(c_1 * d_1 * G_{T_1,\sigma_1} + c_2 * d_2 * G_{T_2,\sigma_2}\right)^2}{(0.09)^2 + \left(c_1 * G_{T_1,\sigma_{N1}} + c_2 * G_{T_2,\sigma_{N2}}\right)^2}$$

where c_1, c_2 represent the stimulus intensities of input1 and 2, taking values of 0 or 1. d_1, d_2 are dominance weights of the model neuron, both set at 1.0. $G_{T,\sigma}$ represents a Gaussian function with a standard deviation of σ and a peak at T. $\sigma_1, \sigma_2,$ σ_{N1} and σ_{N2} were assumed to be 2.0, 2,0, 8.0, 8.0, respectively. T_1, T_2 are defined as: $T_1 = l - t,$ $T_2 = l - t + \delta,$ where t represents time after the onset of stimulus 1 and l represents the peak time of the excitatory input, assumed to be 20 ms for both modalities. δ represents the temporal disparity between the two sensory inputs, ranging from -7 to 7. The multisensory enhancement index (panel C) is given by:

Enhancement Index (%) = [(CM – SMmax) / SMmax] x 100

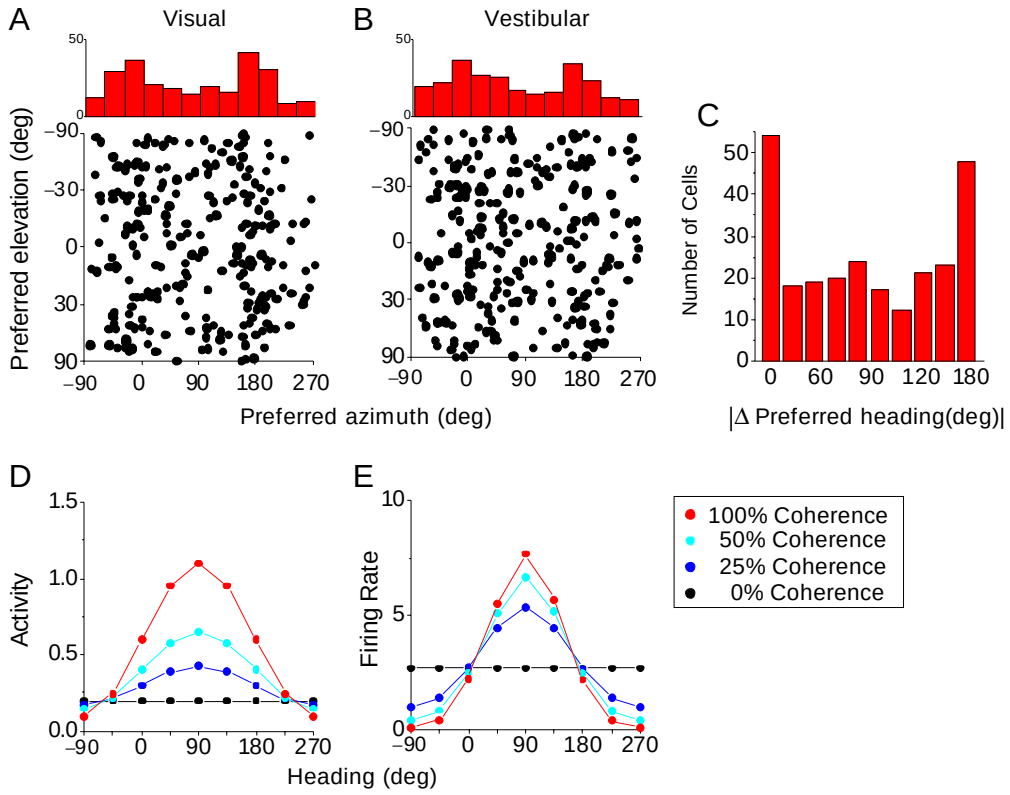
where CM is the total spike count evoked by the bimodal stimulus and SMmax is the total spike count evoked by the most effective unimodal stimulus.

Nature Neuroscience: doi:10.1038/nn.2815



Supplementary Figure 6. Schematic illustration of the version of the normalization model used to simulate visual-vestibular integration by multisensory neurons in area MSTd. Heading tuning curves are shown for a subset of (unisensory) visual and vestibular input neurons (left). Each neuron was tuned for heading in three dimensions, but a two-dimensional cross-section (e.g., the horizontal plane) is shown for clarity. The basic structure of the model is the same as Fig. 1, except that the input non-linearity, *h(x)*, is omitted, and multisensory neurons can combine visual and vestibular inputs that have different heading preferences. This was done to model the distribution of visual-vestibular congruencies seen in MSTd (Supplementary Fig. 7C). Visual and vestibular modality dominance weights are denoted by d_{vis}, d_{vest}, respectively.

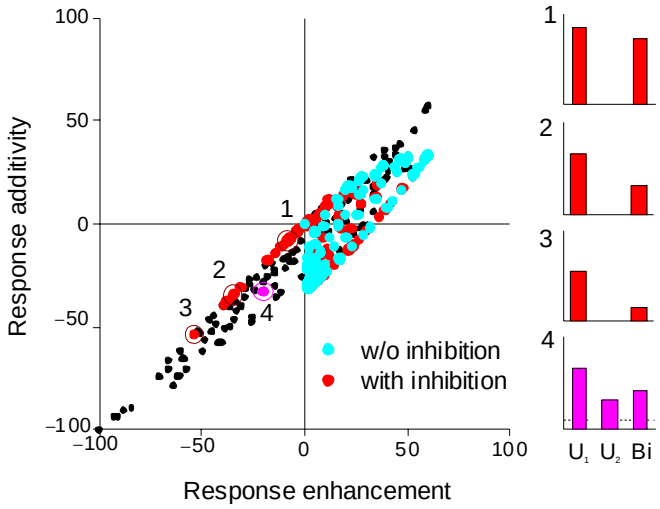
Nature Neuroscience: doi:10.1038/nn.2815



Supplementary Figure 7. Details of response properties for the visual-vestibular heading (MSTd) model.

(A,B) Distribution of 3D heading preferences of model MSTd neurons for the visual ($\hat{\varphi}_{vis}, \hat{\theta}_{vis}$) and vestibular stimulus conditions ($\hat{\varphi}_{vest}, \hat{\theta}_{vest}$). Each data point in the scatter plot corresponds to the preferred azimuth (φ, abscissa) and elevation (θ, ordinate) of a model neuron (total n =256 neurons). The data are plotted on Cartesian axes that represent the Lambert cylindrical equal-area projection of the spherical stimulus space. The histograms at the top of the scatter plots show marginal distributions. These distributions of heading preference were generated to roughly mimic data from area MSTd [3, 4]. (C) Distribution of the (absolute) difference in heading preference between the visual and vestibular conditions, denoted |Δ preferred heading|. (D) The linear component of the response of a model neuron under the visual condition (Eqn. 14) is plotted as a function of heading direction for four different visual stimulus intensities (i.e., coherences). (E) The divisively normalized, final output of a model neuron, in the unimodal visual condition, for the four different visual stimulus intensities.

Nature Neuroscience: doi:10.1038/nn.2815



Supplementary Figure 8. The alternative model less successfully accounts for multisensory interactions in macaque VIP neurons.
Predictions of the alternative model (colored symbols) are compared to the dataset of Avillac et al. [5] from macaque area VIP (black symbols). Format as in Figure 4D. Responses of the alternative model were simulated with (red symbols) or without (cyan symbols) subtractive lateral inhibition. Without inhibition (cyan), the alternative model only predicts data that lie in the right half of the plot, such that all cells show cross-modal enhancement. With inhibition, the alternative model only produces cross-modal suppression (red symbols in the lower left quadrant) when the less effective input (U_2) is not excitatory any more, such that the data lie on the unity slope diagonal in the lower left quadrant. Examples of such model neurons are illustrated to the right (units #1-3). In contrast, data from some real VIP neurons lie in the lower left quadrant but off the diagonal (example unit #4; replotted from Fig.1E of Avillac et al, 2007). The normalization model can account for such neurons that lie off the diagonal (see example # 3 in Figure 4D).

Methods. In this simulation, model neurons take on all possible combinations of dominance weights among the set (d_1, d_2 = 0.00, 0.25, 0.50, 0.75 or 1.00). Responses were computed in response to congruent stimuli presented in the RF center with the following range of stimulus intensities: 16, 32, 64, 128, 256, 512, and 1024. For this simulation, the lateral inhibition was uniform (e.g., Suppl. Fig. 4E), but similar results are obtained with other profiles of inhibition. β was set to either -0.002 (red symbols) or zero (cyan symbols).

Nature Neuroscience: doi:10.1038/nn.2815

References Cited

1. Magosso, E., Cuppini, C., Serino, A., Di Pellegrino, G. & Ursino, M. A theoretical study of multisensory integration in the superior colliculus by a neural network model. *Neural Netw* **21**, 817-829 (2008).
2. Ursino, M., Cuppini, C., Magosso, E., Serino, A. & di Pellegrino, G. Multisensory integration in the superior colliculus: a neural network model. *J Comput Neurosci* **26**, 55-73 (2009).
3. Gu, Y., Fetsch, C.R., Adeyemo, B., Deangelis, G.C. & Angelaki, D.E. Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron* **66**, 596-609 (2010).
4. Gu, Y., Watkins, P.V., Angelaki, D.E. & DeAngelis, G.C. Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J Neurosci* **26**, 73-85 (2006).
5. Avillac, M., Ben Hamed, S. & Duhamel, J.R. Multisensory integration in the ventral intraparietal area of the macaque monkey. *J Neurosci* **27**, 1922-1932 (2007).

Nature Neuroscience: doi:10.1038/nn.2815