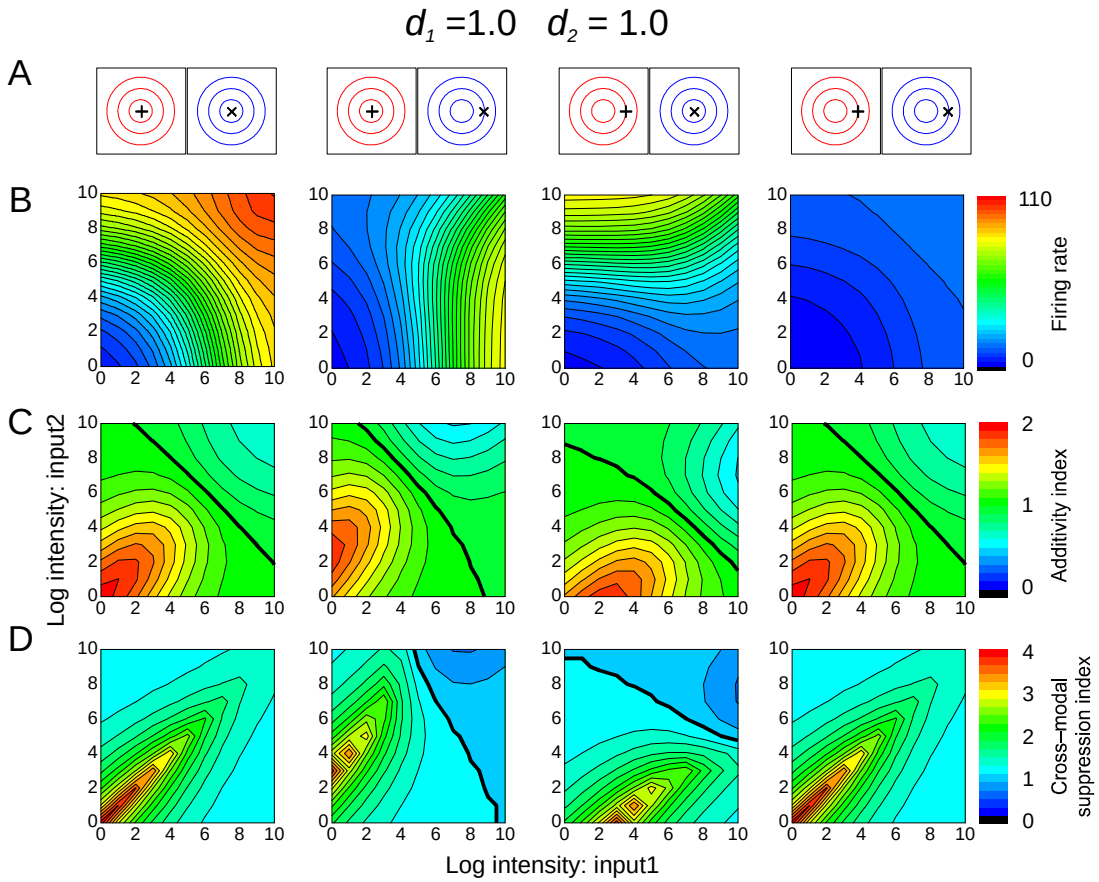
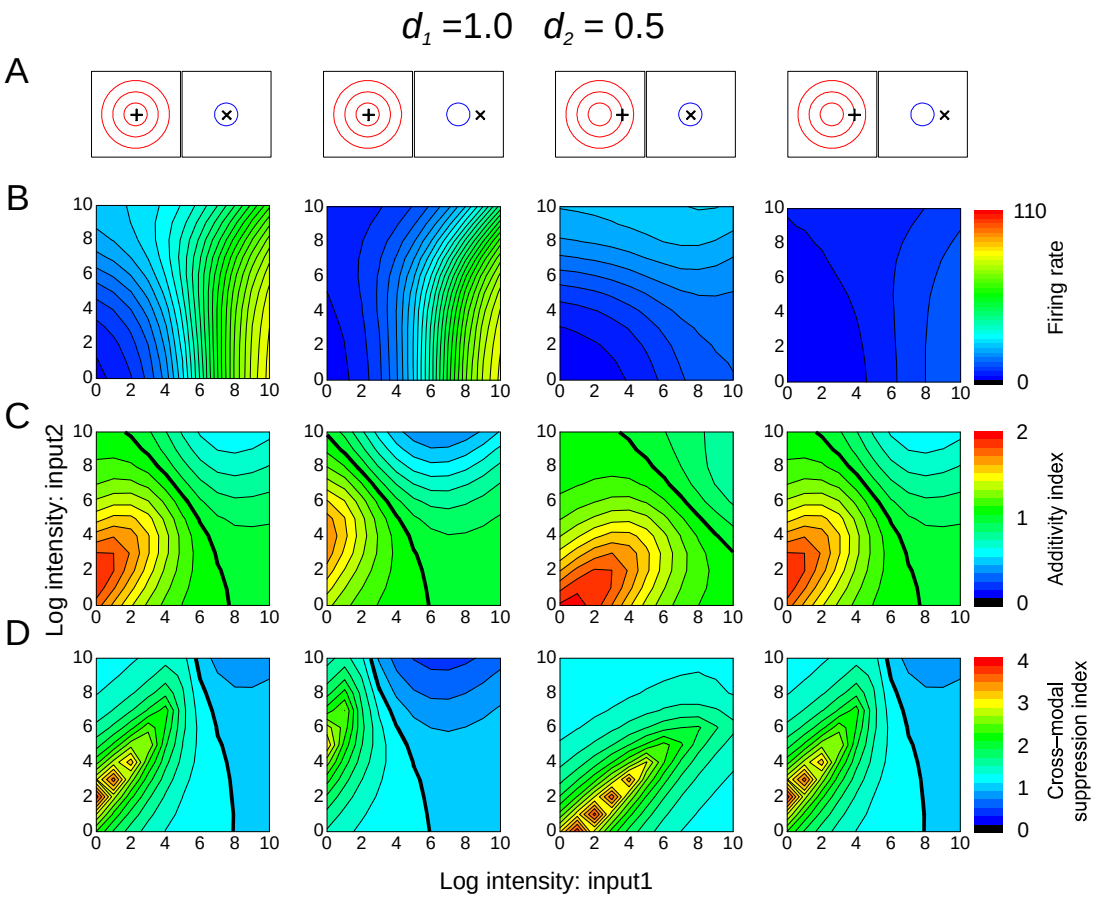
Supplemental Materials for "A Normalization Model of Multisensory Integration" by Ohshiro, Angelaki and DeAngelis

Supplementary Figures and Legends

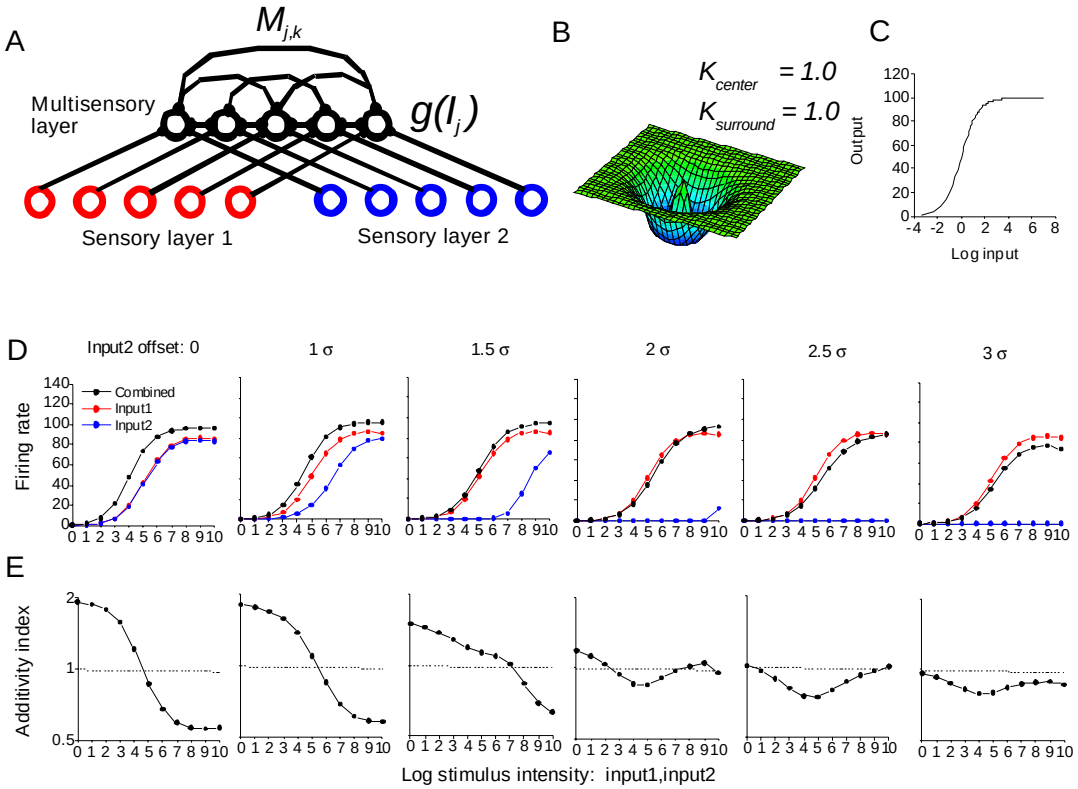


Supplementary Figure 1. Inverse effectiveness and cross-modal suppression for stimuli that are centered or mis-centered on the unimodal receptive fields. (A) Schematic illustration of the locations of Input 1 ('+') and Input2 ('x') relative to the receptive fields (red and blue contours). Conventions 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).



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.



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

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

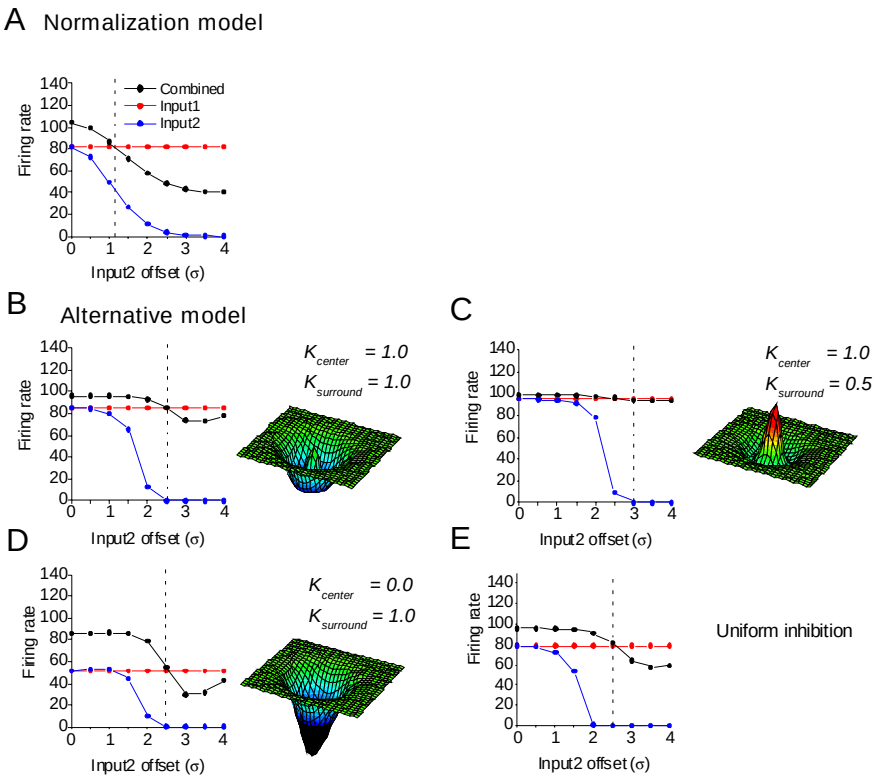
This alternative model and the normalization model both have primary sensory neurons with the same response properties, and both incorporate weighted linear summation of inputs (i.e., the term E_j) to the multisensory neurons. For simplicity, RFs for both modalities are spatially congruent, and the dominance 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:

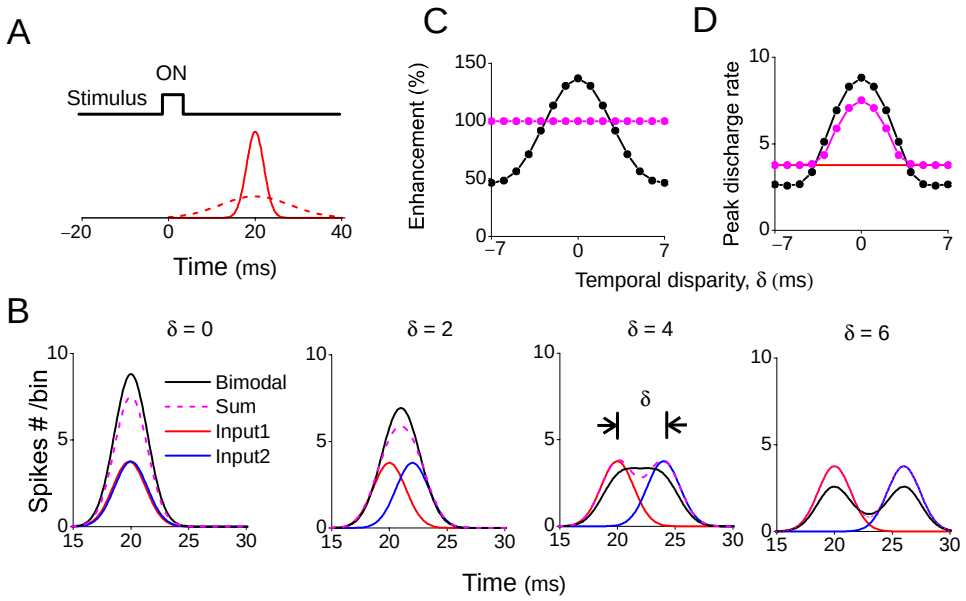
$$M_{k,j} = K_{center} * exp\{-D_{k,j}^2/(2 * \sigma_{center}^2)\} - K_{surround} * exp\{-D_{k,j}^2/(2 * \sigma_{surround}^2)\}$$

.....(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.



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 $M_{k,j}$



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.

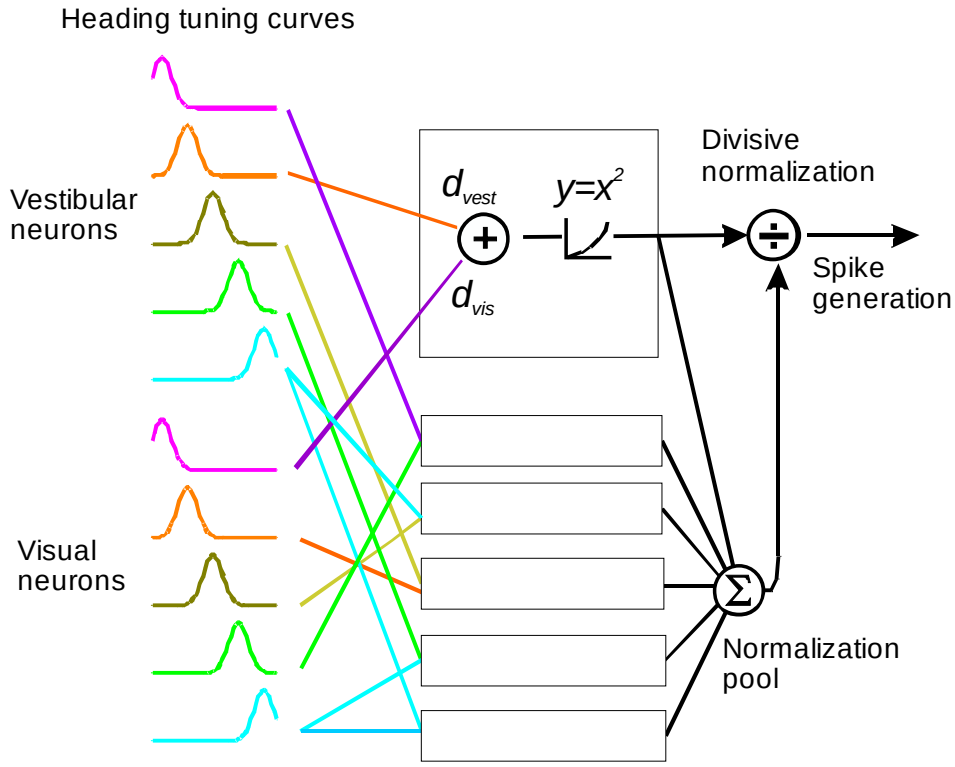
Methods. The normalized bimodal response is modeled as:

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

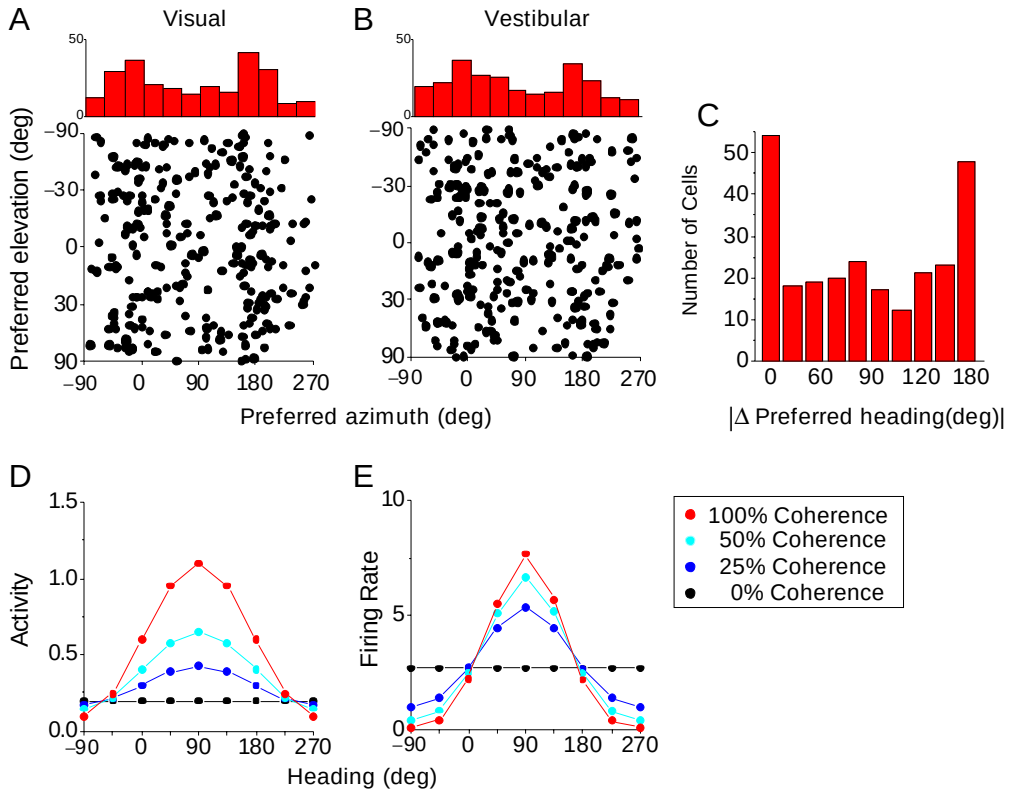
where c_1, c_2 represent the stimulus intensities of input1 and 2, taking values of 0 or 1. d_1, d_2 are dominance weights of the model neuron, both set at 1.0. $G_{T,\sigma}$ represents a Gaussian function with a standard deviation of σ and a peak at T. $\sigma_1, \sigma_2,\ \sigma_{N1}$ and σ_{N2} 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

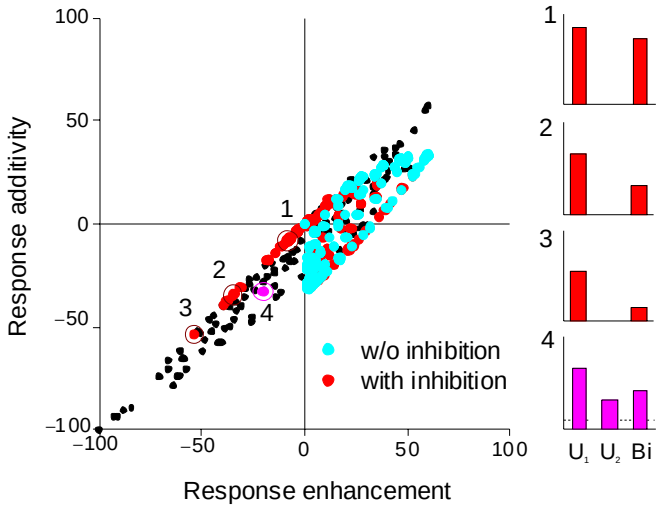
where CM is the total spike count evoked by the bimodal stimulus and SMmax is the total spike count evoked by the most effective unimodal stimulus.



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.



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.



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).

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).