Supplementary information

Neural correlates of multi-sensory cue integration in macaque area MSTd

YONG GU, DORA E. ANGELAKI, AND GREGORY C. DEANGELIS

Supplemental Methods

Local vs. global measures of congruency. In addition to the Congruency Index (CI) used throughout this study, we also examined the congruency of visual and vestibular responses using a measure based on the global heading tuning curves (Fig. 2a,d). Specifically, we computed the difference in heading preference between visual and vestibular conditions (Suppl. Fig. 11), and we found this measure to be bimodal, similar to our previous findings[29]. By comparison, the distribution of CI was not bimodal because any tuning curve that is flat around straight-ahead will lead to a CI near zero. Example neurons with CI values near zero ('intermediate' cells) are shown in Suppl. Fig. 7. For most neurons, there was good agreement between our local measure of congruency (CI) and this alternative global measurement ($R = -0.52$, $p << 0.001$, Spearman rank correlation; Suppl. Fig. 11). However, local and global measures of congruency were widely discrepant for some neurons (examples in Suppl. Fig. 11), and this typically arose when responses were noisy or when one tuning curve had a shallow slope near straight ahead. The local measure is a more reliable predictor of involvement in our task because it is based on many more stimulus repetitions than the global measure and because it identifies neurons as CI-congruent or CI-opposite that have substantial tuning slope around straight ahead. Nevertheless, all of our main results and conclusions hold when we base the

analyses on the global congruency measure, provided that we focus on the subset of neurons that have greater than median sensitivity in the visual and vestibular conditions (red symbols in Suppl. Fig. 12).

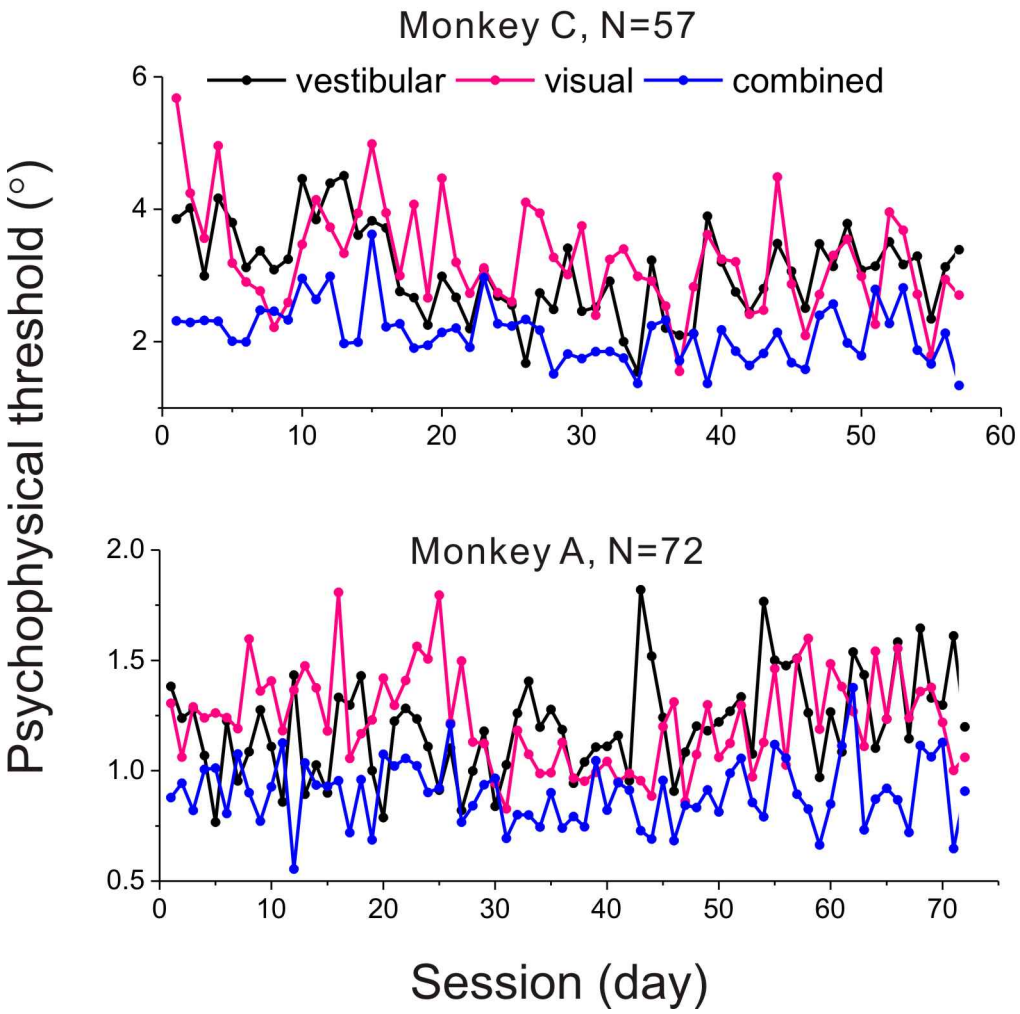
Alternative linear summation models. We evaluated three variants of the linear summation model that was used to predict combined condition responses from responses measured in the visual and vestibular conditions (see also Suppl. Fig. 6):

$$R_{prediction} = w \times (R_{vestibular} + R_{visual}) \quad (4)$$

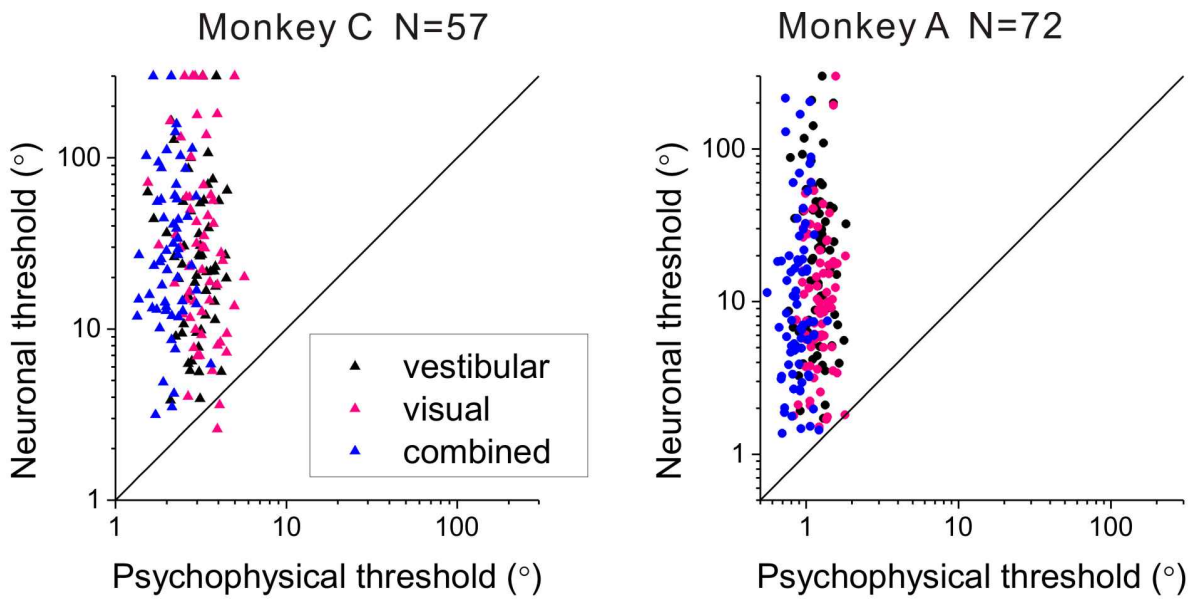
$$R_{prediction} = R_{vestibular} + R_{visual} + w \quad (5)$$

$$R_{prediction} = w_1 \times (R_{vestibular} + R_{visual}) + w_2 \quad (6)$$

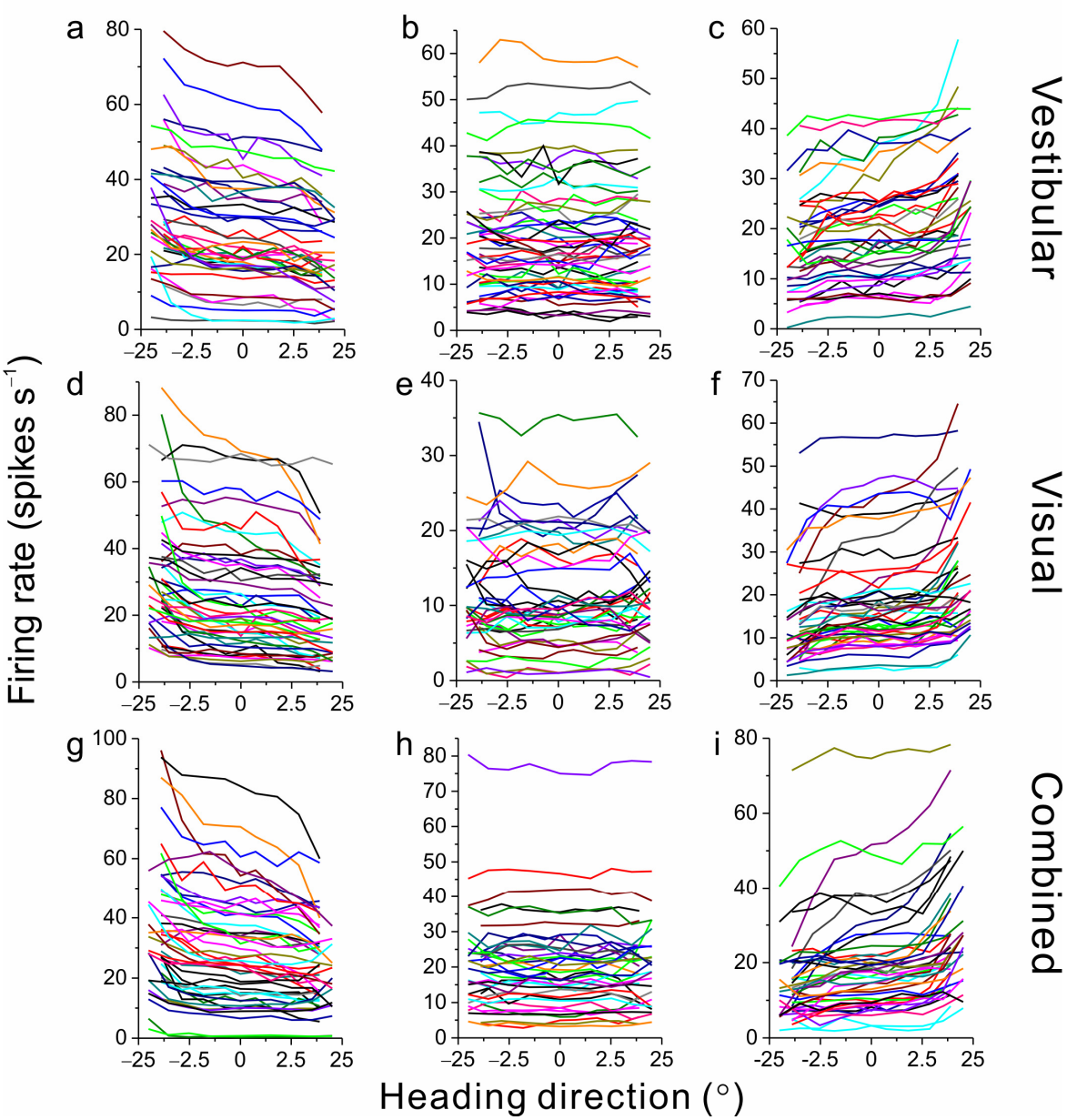
In Eqn. 4, predicted responses are a weighted sum of the single cue inputs but the weight is the same for each single cue, hence there is one less free parameter than in the model used in the main text. In Eqn. 5, the predicted responses are given by a straight arithmetic sum of the two single cue responses and the one free parameter is a DC offset. In Eqn. 6, the predicted responses are derived from equally weighted single-cue responses plus a DC offset. Root mean square (RMS) error was computed between the predicted responses from each linear model and the measured combined responses. RMS errors were then compared between different variants of the linear model to assess which free parameters had the largest impact (see Suppl. Fig. 6).



Supplementary Figure 1. Psychophysical performance for all 57 recording sessions in monkey C (top panel) and 72 sessions in monkey A (bottom panel). Each curve shows psychophysical thresholds plotted as a function of session number. Black, pink, and blue data are from the vestibular, visual, and combined conditions, respectively. Note that thresholds in the combined condition (blue) are consistently smaller than thresholds in the two single-cue conditions.



Supplementary Figure 2. Comparison of neuronal and psychophysical thresholds for each individual experiment performed with monkey C (left panel) and monkey A (right panel). Each datum represents one recording session, with black, pink, and blue symbols denoting the vestibular, visual, and combined conditions, respectively. Most data points lie well above the diagonal, indicating that neurons are less sensitive than the monkeys. Only the most sensitive neurons have thresholds comparable to the animal.



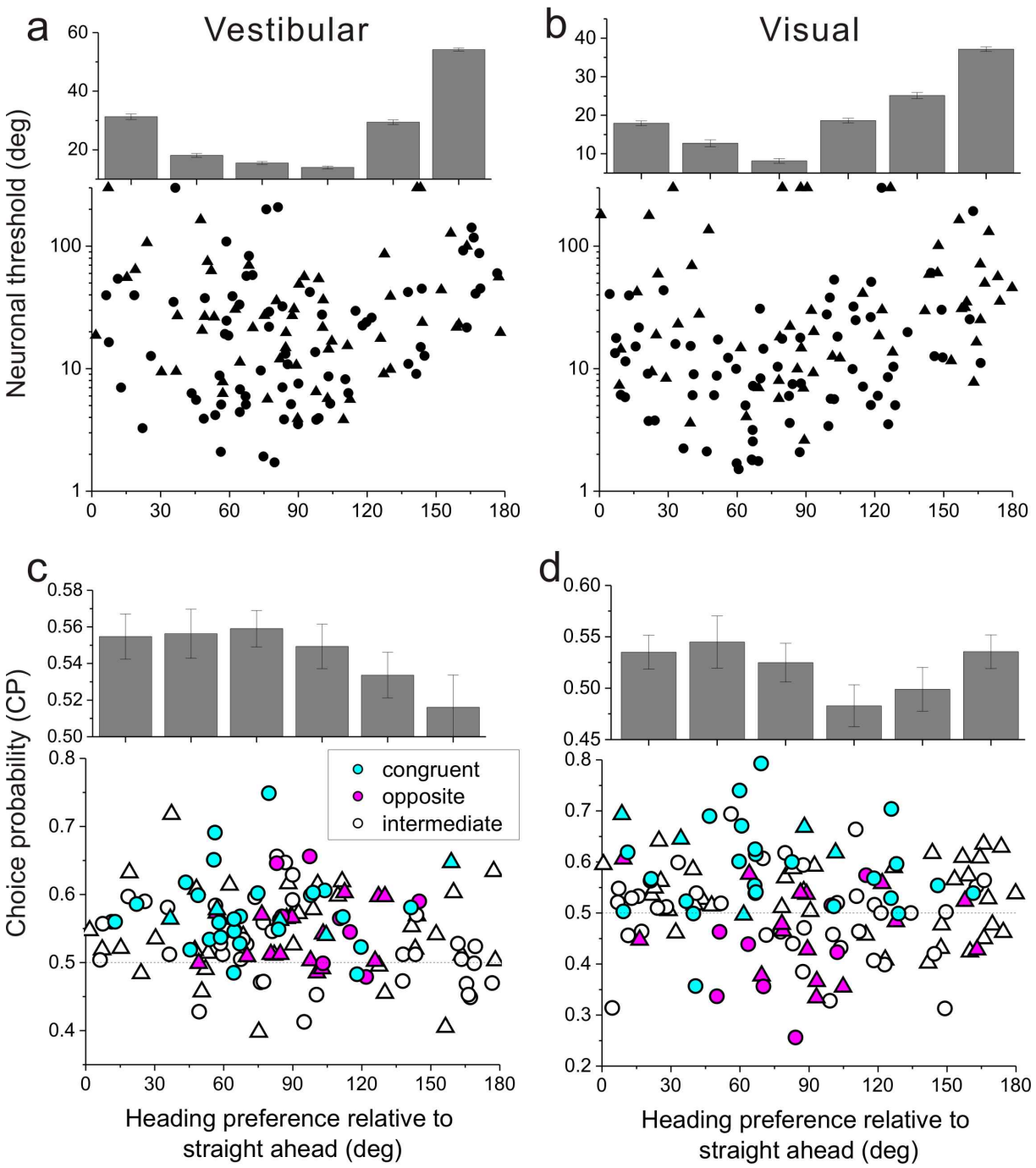
Supplementary Figure 3: Tuning curves measured during the heading discrimination task for all 129 neurons. Top row: tuning curves in the vestibular condition. Middle row: tuning curves in the visual condition. Bottom row: tuning curves in the combined condition. To quantify and summarize the shapes of these tuning curves, each was fit with a second-order polynomial of the form:

$$R(\theta) = a\,\theta^2 + b\,\theta + c,$$

where R(θ) represents the response of each neuron as a function of heading (θ), *a* is the coefficient of the quadratic term, *b* is the coefficient of the linear term, and *c* is the constant offset. Based on the significance ($p < 0.05$) of the coefficients *a* and *b*, each tuning curve was classified into one of four groups. Flat tuning: neither *a* nor *b* significantly different from zero. U-shaped tuning: *a* is significantly different from zero but *b* is not. Linear tuning: *b* is significantly different from zero but *a* is not. Non-linear tuning: both *a* and *b* significantly different from zero. The middle column represents neurons with either Flat or U-shaped tuning. The left and right columns include neurons with linear or non-linear tuning for which *b* is significantly negative or positive, respectively.

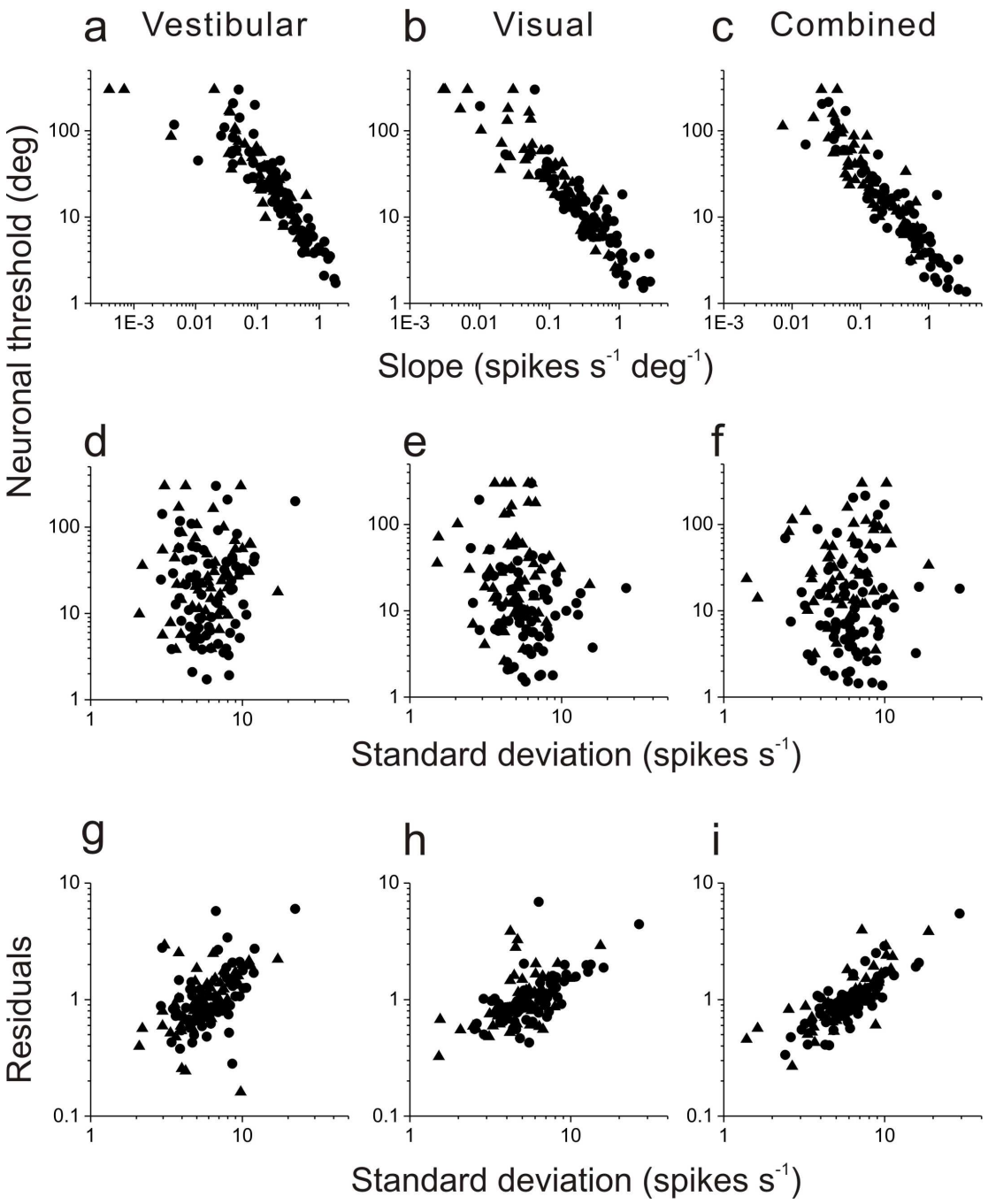
The table below summarizes data from the four classes of neurons defined above, including the percentages of each type for each stimulus condition, the median R^2 values (goodness of fit), the geometric mean neuronal thresholds, and the average choice probabilities. For the combined condition, three sets of values are given as follows: All cells (CI-congruent cells/CI-opposite cells).

Vestibular condition	Percentage	Median R^2	Avg. threshold (°)	Avg. CP
Sig. *a* & *b* (Nonlinear)	17.8	0.62	12.7	0.54
Sig. *b* only (Linear)	43.4	0.78	10.0	0.57
Sig. *a* only (U-shaped)	7.0	0.08	71.4	0.51
Neither sig. (Flat)	31.8	0.20	57.2	0.53
Visual condition				
Sig. *a* & *b* (Nonlinear)	37.2	0.74	7.7	0.52
Sig. *b* only (Linear)	31	0.79	9.8	0.53
Sig. *a* only (U-shaped)	14	0.06	80.6	0.49
Neither sig. (Flat)	17.8	0.22	55.8	0.52
Combined condition				
Sig. *a* & *b* (Nonlinear)	43.4 (13.2/10.8)	0.65 (0.85/0.44)	9.1 (4.2/13.5)	0.52 (0.57/0.49)
Sig. *b* only (Linear)	25.6 (10.1/1.6)	0.83 (0.94/0.86)	7.2 (3.4/13.8)	0.56 (0.59/0.44)
Sig. *a* only (U-shaped)	8.5 (0.0, 2.3)	0.05 (NA/0.04)	62.1 (NA/51.0)	0.49 (NA/0.48)
Neither sig. (Flat)	22.5 (0.0/3.9)	0.18 (NA/0.16)	67.0 (NA/ 63.4)	0.51 (NA/0.45)



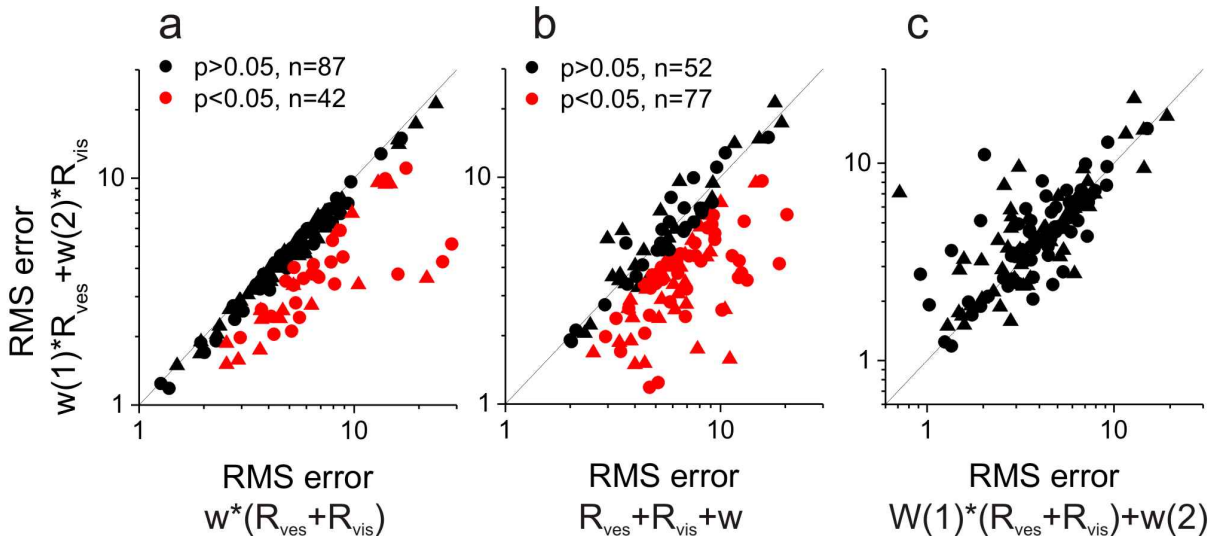
Supplementary Figure 4. Relationships between neuronal thresholds, choice probabilities (CPs) and heading preferences. **(a)** Neuronal threshold in the vestibular condition is plotted as a function of each cell's heading preference relative to straight forward (0 degrees on the abscissa corresponds to neurons that prefer forward motion). The marginal histogram (top) shows the

geometric mean (± geometric s.e.) of the neuronal threshold as a function of heading preference. Thresholds tend to be lowest for neurons with heading preferences 60-90 degrees away from straight ahead. Circles: monkey A; triangles: monkey C. **(b)** Neuronal thresholds in the visual condition as a function of heading preference; same format as in (a). **(c)** Relationship between CP in the vestibular condition and heading preference. Cyan, magenta, and unfilled symbols represent CI-congruent, CI-opposite, and intermediate cells, respectively. Marginal histogram (top) shows the average CP (± s.e.) as a function of heading preference. **(d)** Relationship between visual CP and heading preference; format as in (c). Each global heading tuning curve was fit with a wrapped Gaussian function and tuning width was measured at half-maximum response. The median half-width of heading tuning is 60.5° and 54.5° in the vestibular and visual conditions, respectively. These tuning widths are broadly consistent with the fact that neuronal thresholds are smallest around 60-90 degrees away from straight ahead in panels (a) and (b).

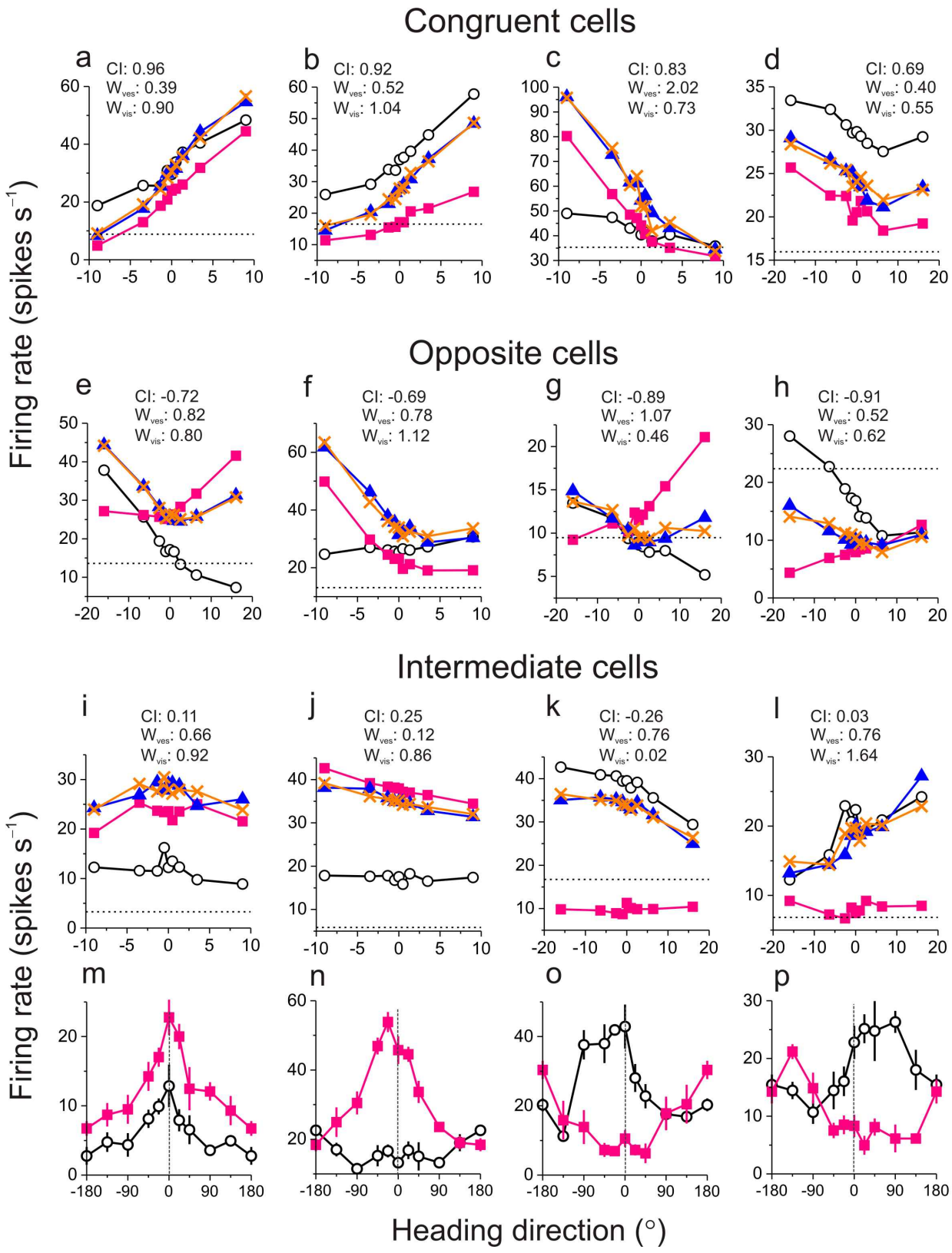


Supplementary Figure 5. Dependence of neuronal threshold on tuning slope and standard deviation (SD) of responses in the vestibular (left), visual (middle) and combined (right) conditions. (**a-c**) Neuronal threshold plotted as a function of the slope of the heading tuning curve. The slope was computed from a linear fit to the firing rates across the 9 headings tested in

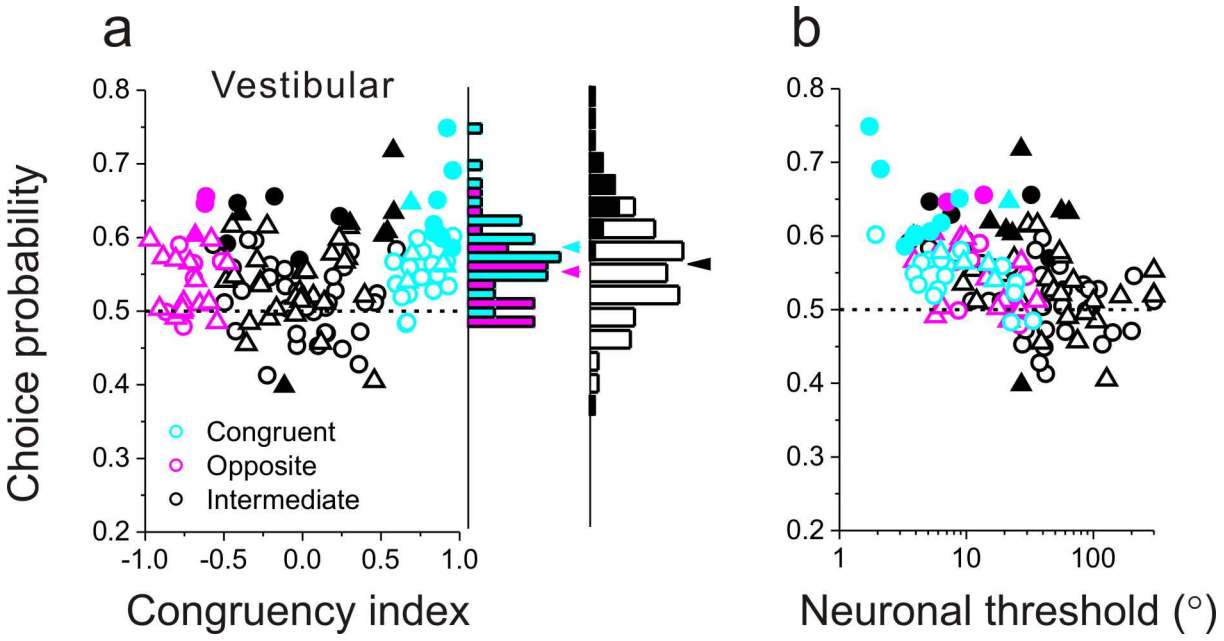
the discrimination task. The simple correlations between threshold and slope were strong and highly significant ($p << 0.001$; vestibular: $R = -0.87$; visual, $R = -0.92$; combined, $R = -0.92$). Since we did not tailor the range of headings to the steep slope of each cell's tuning curve, many neurons have high thresholds because tuning was flat around straight ahead. (**d-f**) Neuronal thresholds plotted as a function of SD (averaged over the 9 headings; results were similar when computed only from the ambiguous 0° heading). The simple correlation between threshold and SD did not reach significance ($p > 0.06$) for any condition (vestibular: $R = 0.07$; visual, $R = -0.16$; combined, $R = 0.02$). Thus, the influence of SD on threshold was much smaller than the influence of slope. (**g–i**) Residuals from the linear fit of threshold vs. slope are plotted as a function of response SD. A positive correlation is now revealed, such that neurons with higher SDs tend to have larger residual thresholds. These relationships were verified by a multiple regression analysis. Partial correlations between threshold and slope were all significant (vestibular: $R = -0.90$; visual, $R = -0.96$; combined, $R = -0.98$; $p << 0.001$). Partial correlations between threshold and SD were weaker (vestibular: $R = 0.47$; visual, $R = 0.67$; combined, $R = 0.83$) but were still highly significant ($p < 0.001$ for all conditions). Thus, both tuning slope and response SD contribute to neuronal sensitivity, but the effect of slope is larger than the effect of SD. Triangles and circles denote data from monkeys C and A, respectively.



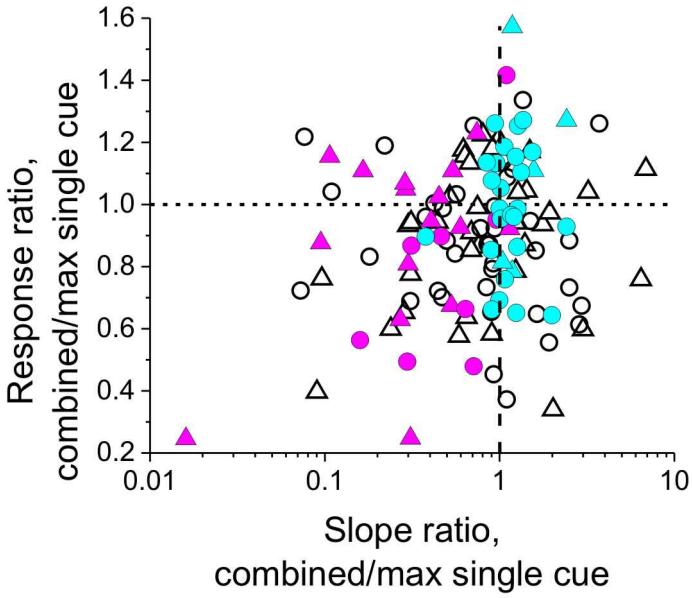
Supplementary Figure 6. Comparison of the weighted linear sum model with alternative linear models. RMS errors were computed according to: $RMS_{error} = \sqrt{\sum (R_{predict} - R_{actual})^2}$. In each panel, the RMS error of the weighted linear sum model used in the main text is compared with the RMS error for another variant of the linear model. (**a**) For a model where weights were forced to be identical for vestibular and visual inputs, the median RMS error was significantly larger than the median for the weighted linear sum model ($p<<0.001$, sign test). However, the model with a common weight on visual and vestibular inputs produced significantly worse fits (sequential F-test, $p < 0.05$) for only 32% of MSTd neurons (red data points). (**b**) For a model in which the weights for vestibular and visual inputs were constrained to be unity but the DC offset was varied, the median RMS error was again significantly larger than that for the weighted linear sum model ($p<<0.001$, sign test). In this case, 60% of MSTd neurons had significantly worse fits than with the linear weighted sum model (red data points). (**c**) For a model in which vestibular and visual inputs shared a common weight and the DC offset was varied, the median RMS error was not significantly different from that obtained using the weighted linear sum model ($p>0.4$, sign test). Circles: monkey A; Triangles: monkey C.


Congruent cells
a
CI: 0.96
W_ves: 0.39
W_vis: 0.90
b
CI: 0.92
W_ves: 0.52
W_vis: 1.04
c
CI: 0.83
W_ves: 2.02
W_vis: 0.73
d
CI: 0.69
W_ves: 0.40
W_vis: 0.55
Opposite cells
e
CI: -0.72
W_ves: 0.82
W_vis: 0.80
f
CI: -0.69
W_ves: 0.78
W_vis: 1.12
g
CI: -0.89
W_ves: 1.07
W_vis: 0.46
h
CI: -0.91
W_ves: 0.52
W_vis: 0.62
Intermediate cells
i
CI: 0.11
W_ves: 0.66
W_vis: 0.92
j
CI: 0.25
W_ves: 0.12
W_vis: 0.86
k
CI: -0.26
W_ves: 0.76
W_vis: 0.02
l
CI: 0.03
W_ves: 0.76
W_vis: 1.64
m
n
o
p
Firing rate (spikes s⁻¹)
Firing rate (spikes s⁻¹)
Heading direction (°)

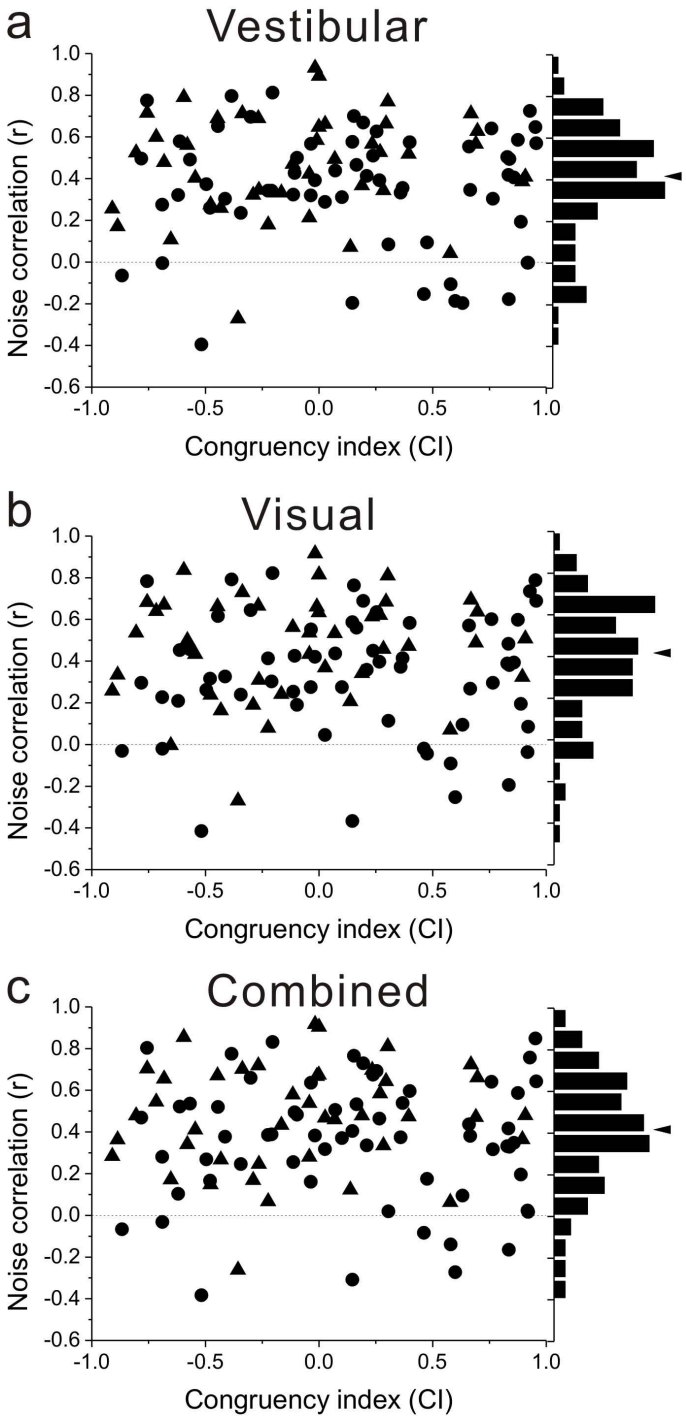
Supplementary Figure 7. Heading tuning curves obtained during the discrimination task for 12 additional MSTd neurons. (**a-d**) Congruent neurons. (**e-h**) Opposite neurons. (**i-l**) Intermediate neurons. Vestibular condition: black circles/lines; Visual condition: pink squares/lines; Combined condition: blue triangles/lines. Orange crosses/lines represent best fits of the weighted linear sum model. Dashed line: spontaneous activity. w_{ves}: vestibular weight; w_{vis}: visual weight. CI: congruency index. (**m-p**) Global heading tuning curves for the intermediate cells, obtained in separate tests while the monkey maintained fixation.



Supplementary Figure 8. Choice probabilities (CP) for the vestibular condition. **(a)** The scatter plot shows CP plotted as a function of congruency index (CI) for each of the 129 MSTd neurons tested under the vestibular condition (triangles: monkey C; circles: monkey A). Cyan and magenta data points represent neurons classified as 'CI-congruent' and 'CI-opposite', respectively, based on the sign and significance of the CI. Filled symbols denote neurons for which the CP was significantly different from 0.5. The rightmost marginal histogram shows the distribution of CP values for all neurons, with filled bars denoting CPs significantly different from 0.5. The adjacent marginal histogram shows distributions of CP values for neurons classified as CI-congruent (cyan) and CI-opposite (magenta). Note that, unlike in the visual and combined conditions (Fig. 6), there is no clear relationship between CI and CP under the vestibular condition. MSTd responses are consistently correlated with perceptual choices regardless of congruency. **(b)** CP is significantly anti-correlated with neuronal threshold ($R = -0.48$, $p \ll 0.001$, Spearman rank correlation). Neurons with lower thresholds tend to have larger CPs, regardless of congruency.

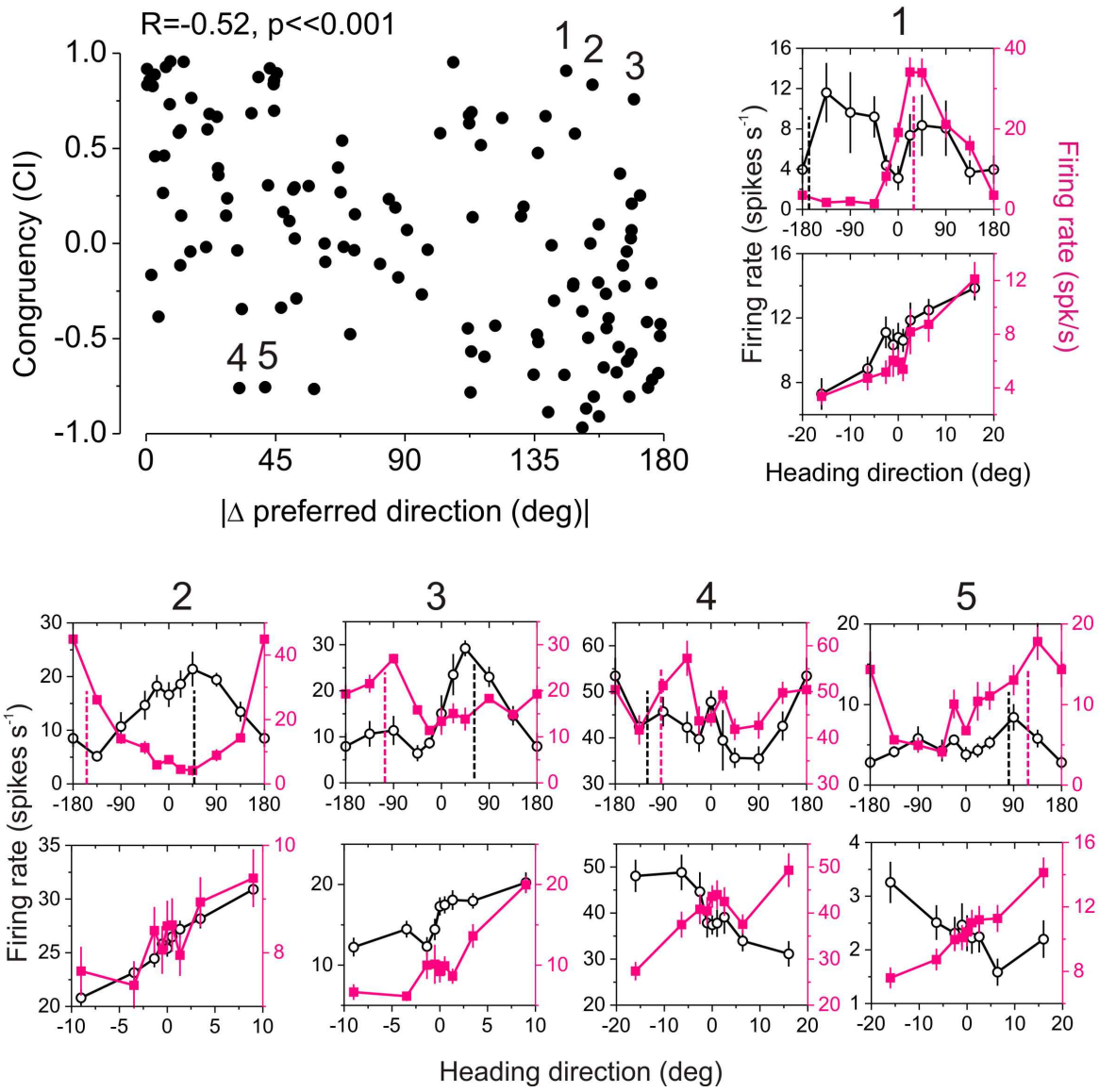


Supplementary Figure 9. Relationship between the slope change in the combined condition and the ratio of firing rates between the combined and single-cue conditions. Response ratio (ordinate) is computed as the average response across all 9 headings in the combined condition divided by the average response elicited by the most effective single cue. Response ratios larger than 1 indicate that combined responses were larger than both single cue responses. Ratios smaller than 1 correspond to combined responses that were smaller than at least one of the single cue responses. Slope ratio (abscissa) is the ratio of the tuning curve slope in the combined condition to the largest slope from the two single-cue conditions. The correlation between the two ratios was not significant ($R=0.15$, $p>0.09$, Spearman rank correlation). For CI-congruent cells (cyan), combined condition slopes are steeper, yet firing rates can be either greater or less than those elicited by the most effective single cue. For CI-opposite cells (magenta), the ratio of combined/single-cue firing rates is roughly similar, yet slopes in the combined condition are consistently reduced. Unfilled symbols illustrate intermediate cells (i.e., neither CI-congruent nor CI-opposite). Circles: monkey A; Triangles: monkey C.



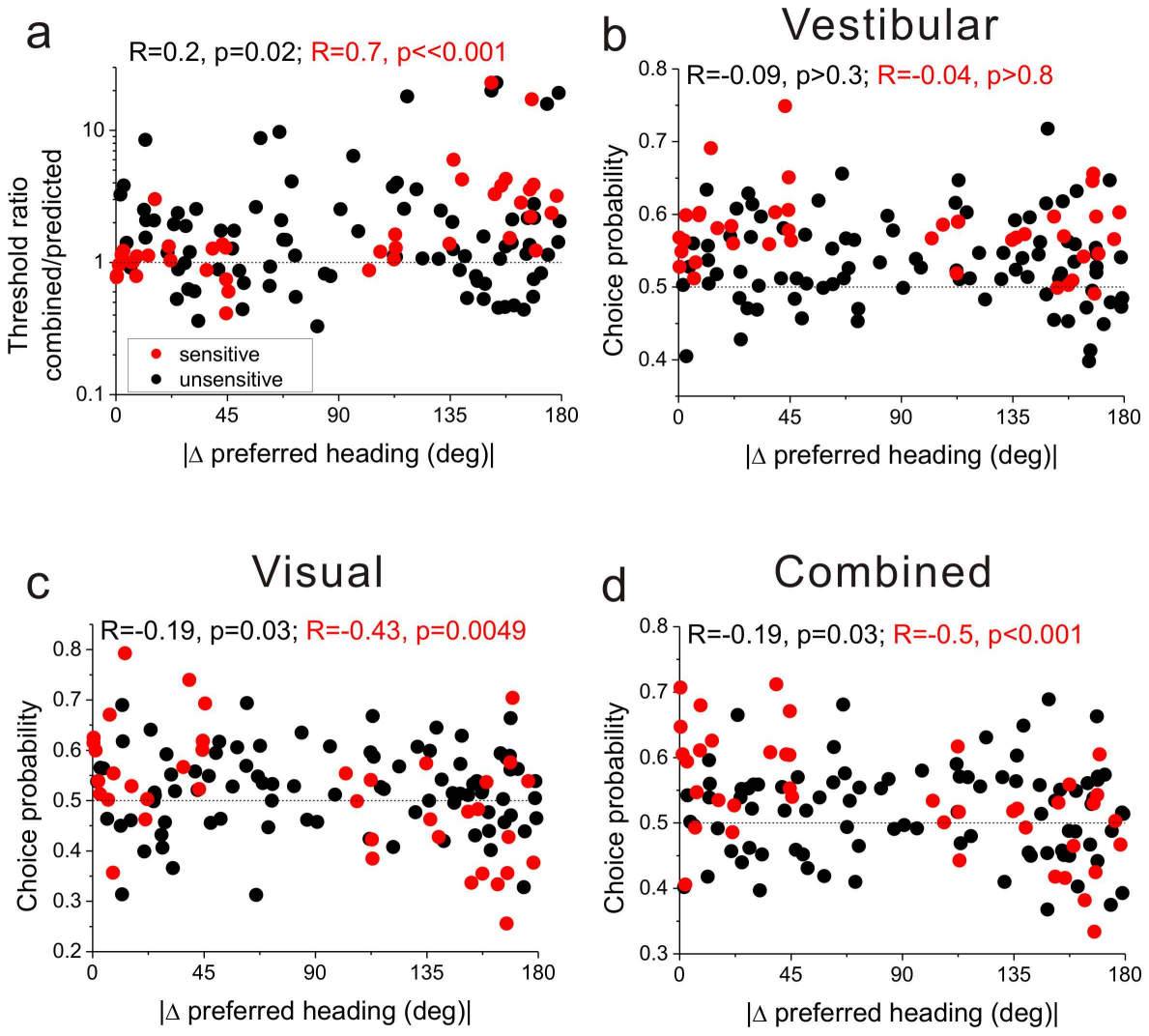
Supplementary Figure 10. Noise correlation (r) between single-unit (SU) and multi-unit (MU) responses was computed for 105 recordings (raw neural signals were not available for the

remaining 24 recordings). MU responses were obtained by thresholding the raw digitized neural signals, and SU spikes were removed from the MU signal (as verified by cross-correlation). Very slow variations in firing rate across time were then removed by high-pass filtering the SU and MU response sequences. To compute noise correlation, responses to each unique heading stimulus were z-scored, and then z-scored responses were combined across heading conditions. The Pearson correlation coefficient (r) was then computed between SU and MU responses. Data are shown for the vestibular **(a)**, visual **(b)**, and combined conditions **(c)**. The mean values of r for the three stimulus conditionsare 0.4, 0.39, and 0.4, respectively. These mean values are not significantly different from each other ($p>0.9$, one-way ANOVA), and all are significantly larger than 0 ($p<<0.001$, t-test). In all three stimulus conditions, there is no significant correlation between r and the congruency index (CI) of the SU response ($p>0.6$, Spearman rank correlation). Noise correlations are highly correlated across stimulus conditions (vestibular vs. visual, $R=0.93$, $p<<0.001$, Spearman rank correlation; vestibular vs. combined, $R=0.93$, $p<<0.001$; visual vs. combined, $R=0.94$, $p<<0.001$), suggesting that, unlike CP, r is largely independent of stimulus condition. Circles: monkey A; triangles: monkey C.



Supplementary Figure 11. Relationships between local and global measures of congruency between visual and vestibular heading tuning. The scatter plot shows the local measure (CI) plotted as a function of the global measure, which is defined as the absolute value of the difference between heading preferences determined from the global heading tuning curve. There is a clear negative correlation between local and global measures in the scatter plot (R = -0.52, p<<0.001, Spearman rank correlation). In some cases, local and global measures do not agree well, as illustrated by the insets (examples 1-5). For each example (inset), the global heading

tuning curve is shown at the top and the local heading tuning curves (measured during the discrimination task) are shown below. Black: vestibular tuning; pink: visual tuning. Dashed vertical lines indicate the global heading preferences as determined from a vector average of responses.



Supplementary Figure 12. Reproduction of main results based on the global measure of visual-vestibular congruency. Red data points represent the subset of neurons (41/129) with above-median sensitivity in both the visual and vestibular conditions. **(a)** Combined/predicted threshold ratio is plotted as a function of the global congruency measure (absolute difference between visual and vestibular heading preferences). For all 129 neurons (black and red symbols together), there is a weak but significant positive correlation (R=0.2, p=0.02, Spearman rank correlation) between threshold ratio and global congruency. For the subset of sensitive neurons (red symbols, N = 41), the correlation between threshold ratio and congruency is much stronger

(R=0.7, p<<0.001, Spearman rank correlation), and produces a result similar to that seen for the CI in Fig. 3a. **(b)** Choice probability for the vestibular condition is plotted as a function of global congruency. There is no correlation between these variables for the entire population (R=-0.09, p>0.3) nor for the subset of sensitive cells (R=-0.04, p>0.8; red symbols). **(c)** In the visual condition, there is a weak negative correlation between choice probability and global congruency for the entire population of neurons (R=-0.19, p=0.03), but a strong negative correlation (R=-0.43, p<0.005) for the subset of sensitive neurons (red symbols). (d) In the combined condition, choice probability again shows a weak correlation with global congruency for the entire sample (R=-0.19, p=0.03) and a strong negative correlation for the sensitive cells (R=-0.5, p<0.001; red symbols). Thus, the data from sensitive neurons recapitulate the main results found using the local measure of congruency (CI).