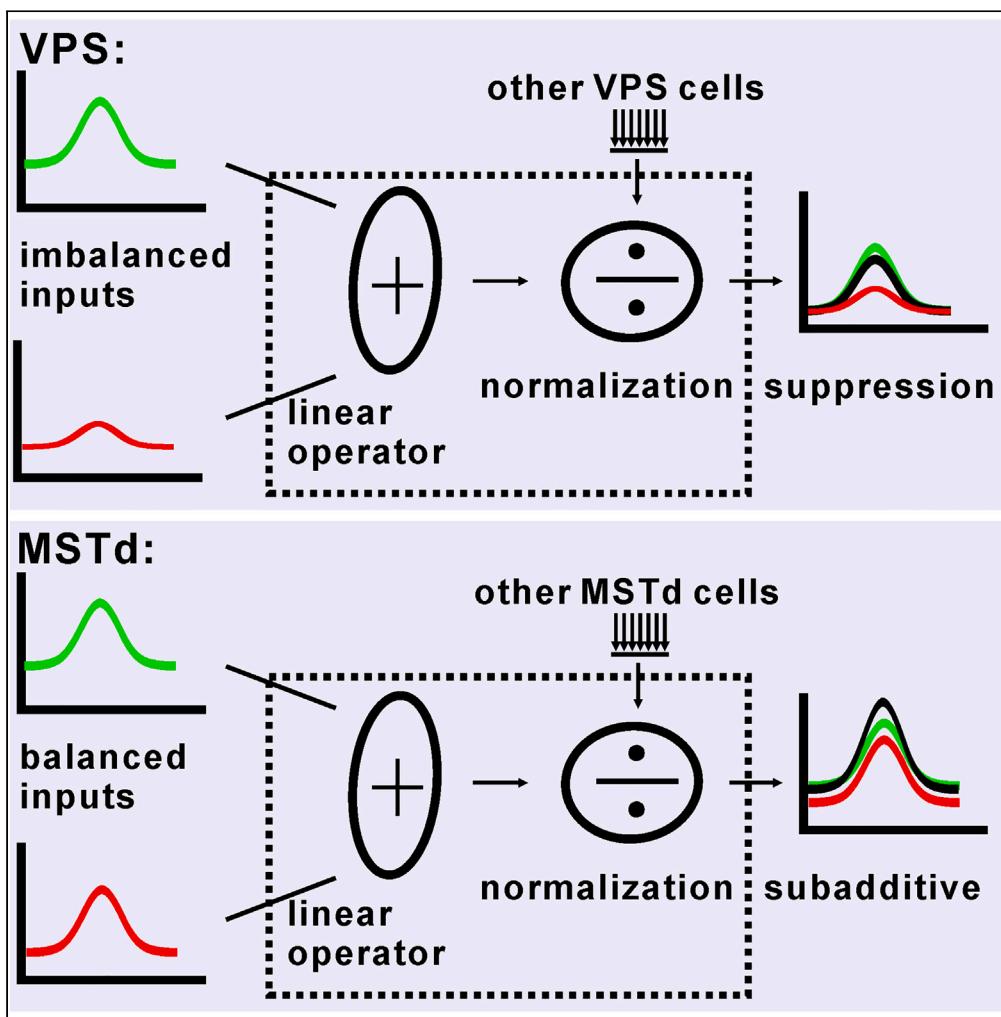


Article

The computational rules of cross-modality suppression in the visual posterior sylvian area



Bin Zhao, Rong Wang, Zhihua Zhu, Qianli Yang, Aihua Chen

qlyang@ion.ac.cn (Q.Y.)
ahchen@brain.ecnu.edu.cn (A.C.)

Highlights
Cross-modal suppression was frequently observed in area VPS

Distinct information was encoded in VPS under different offset conditions

Combined responses in VPS were fit by weighted linear sums of unimodal responses

The normalization model captured most vestibular-visual interaction

Zhao et al., iScience 26,
106973
June 16, 2023 © 2023 The
Author(s).
<https://doi.org/10.1016/j.isci.2023.106973>



Article

The computational rules of cross-modality suppression in the visual posterior sylvian area

Bin Zhao,^{1,3} Rong Wang,^{1,3} Zhihua Zhu,² Qianli Yang,^{2,*} and Aihua Chen^{1,4,5,*}

SUMMARY

The macaque visual posterior sylvian area (VPS) is an area with neurons responding selectively to heading direction in both visual and vestibular modalities, but how VPS neurons combined these two sensory signals is still unknown. In contrast to the subadditive characteristics in the medial superior temporal area (MSTd), responses in VPS were dominated by vestibular signals, with approximately a winner-take-all competition. The conditional Fisher information analysis shows that VPS neural population encodes information from distinct sensory modalities under large and small offset conditions, which differs from MSTd whose neural population contains more information about visual stimuli in both conditions. However, the combined responses of single neurons in both areas can be well fit by weighted linear sums of unimodal responses. Furthermore, a normalization model captured most vestibular and visual interaction characteristics for both VPS and MSTd, indicating the divisive normalization mechanism widely exists in the cortex.

INTRODUCTION

In nature, we are facing abundant multisensory information in the environment. To maintain a unified perception and produce a coherent behavioral response, our brain needs to selectively process these vast amounts of inputs, combining multiple sensory inputs to make good decisions.^{1–4} On the other hand, when stimuli from different modalities exhibit sufficient mismatch, multisensory integration usually breaks down with one stimulus dominating the other.^{5–8} Thus, understanding the neural computational rules underlying multisensory interaction is fundamental to our comprehension of the world.⁹

Visual-vestibular interaction in heading perception provides a good model to characterize how the combined responses related to the unimodal components. Previous studies in the dorsal medial superior temporal area (MSTd), an area known for the integration of visual and vestibular during self-motion perception,^{10,11} demonstrated that combined responses were well fit by a weighted linear summation of unimodal responses.¹² Later studies found the combined responses in MSTd within a narrow heading range followed a critical prediction of the normalization model.^{13,14} Since vestibular and visual signals related to heading are widely distributed in the cortex, several areas have been identified with responses to both vestibular and visual cues.^{10,15–18} It is not clear whether other areas also follow the computational rules observed in MSTd.

The visual posterior sylvian area (VPS) of the macaque brain is an area with neurons responding selectively to heading direction in both visual and vestibular modalities.¹⁷ In contrast to the visual dominance in MSTd, area VPS is vestibular dominant.¹⁷ In particular, opposite heading preferences between vestibular and visual signals were frequently observed,¹⁷ unlike in other vestibular-visual multisensory areas (e.g., MSTd¹⁰ or ventral intraparietal area (VIP)¹⁸) with roughly equal proportions of congruent and opposite neurons. Thus, VPS may not be a key site for combining the perception of self-motion trajectory as MSTd, and the interaction rules between vestibular and visual signals might also be different from MSTd.^{12–14}

To investigate how the visual and vestibular signals interact with each other in VPS, we used an experimental protocol similar to the previous one characterizing the computational rules employed in MSTd.¹² We measured the responses of cells to eight directions of translation (45° apart) in the horizontal plane using visual cues alone (optic flow), vestibular cues alone, and combined stimuli including all 64 (8 × 8) combinations of visual and vestibular headings, both congruent and conflicting. Such an approach had a

¹Key Laboratory of Brain Functional Genomics, East China Normal University, Shanghai 200062, China

²CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Neuroscience, Chinese Academy of Sciences, Shanghai 200031, China

³These authors contributed equally

⁴Senior author

⁵Lead contact

*Correspondence:
qlyang@ion.ac.cn (Q.Y.),
ahchen@brain.ecnu.edu.cn (A.C.)

<https://doi.org/10.1016/j.isci.2023.106973>



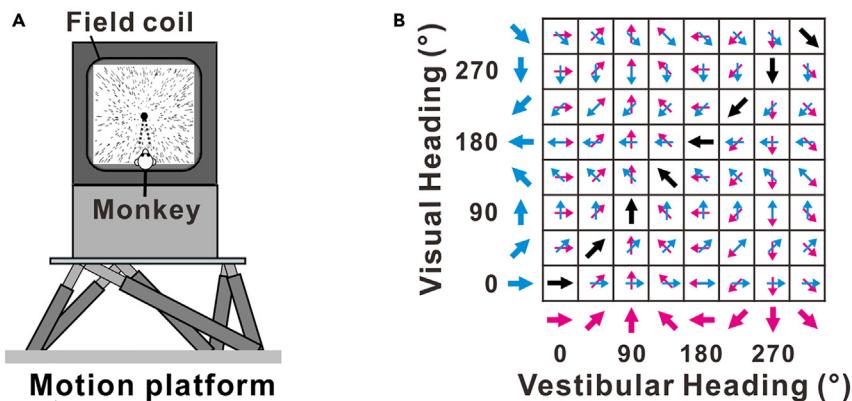


Figure 1. Experimental apparatus and stimuli

(A) Experimental apparatus consists of a display, motion platform, and field coil. The monkey was seated comfortably in a primate chair which was secured to a motion platform. Visual stimuli were generated on the screen placed ~30 cm in front of the monkey.

(B) Sample stimulus set used in our experiments. The stimuli consisted of the matrix populated by all combinations of the tested vestibular and visual headings (from 0° to 315°, in step 45°) in the horizontal plane.

broad range of stimuli with all possible combinations of visual and vestibular headings and thus can evoke widely varying responses, allowing us to comprehensively characterize the neuronal combination rule.

In contrast to the subadditive characteristics in MSTd,¹² cross-modal suppression was typically seen in VPS, and the vestibular weights were much larger than the visual weights,¹⁷ with approximately a winner-take-all competition. To reveal the possible function of VPS, conditional Fisher information analysis was applied to decode vestibular and visual preference. It turns out that VPS neural population contains significant information about visual heading direction under small offset conditions. Whereas under large offset conditions, VPS neural population contains more information about vestibular heading direction.

To further examine the computational principles of VPS neurons, a simple linear model was applied, and the combined responses in single VPS neurons can be well fit by weighted linear sums of unimodal responses. Then, we examined whether the vestibular-visual interaction in population data in VPS can be explained by the divisive normalization model, which is based on a linear summation of inputs and a power-law non-linearity^{19–27} and widely used in V1,²⁸ MT,²⁹ and LIP.³⁰ Our results showed that for both multisensory and unisensory neurons, the normalization approach provides a unified framework for characterizing vestibular-visual interaction in both VPS and MSTd areas, suggesting that normalization is a necessary computation for multisensory perception.

RESULTS

We recorded 171 well-isolated VPS neurons in four animals (10 in monkey A1, 29 in monkey A2, 78 in monkey B, and 54 in monkey K). Responses in the horizontal plane were obtained during three conditions (vestibular, visual, and combined) by using a virtual-reality system (Figure 1A). For each recorded neuron, we run a block of interleaved trials containing 81 distinct stimulus conditions (8 vestibular only, 8 visual only, 64 combinations, and a “null” condition). The example stimulus set of 64 combinations of vestibular and visual headings was illustrated in Figure 1B, including 8 congruent (marked with black arrows) and 56 cue-conflict presentations. During stimulus presentation, monkeys were simply required to maintain fixation at a central target on the screen.

Vestibular suppression in VPS

Typical responses from a “multisensory”, “vestibular only”, and “visual only” cell in VPS were illustrated in Figure 2A, 2B, and 2C, respectively. Figure 2A shows the average peristimulus time histograms (PSTHs) of a multisensory VPS neuron for all combinations of visual and vestibular heading (8 cue-congruent and 56 cue-conflict conditions). The bottom panels illustrated the responses to 8 vestibular-only conditions (from 0° to 315°), and the left column of PSTHs represented the responses to 8 visual-only conditions. This cell was well tuned to both vestibular and visual signals ($p_{\text{vestibular}} = 1.29 \times 10^{-7}$, $p_{\text{visual}} = 3.48 \times 10^{-4}$, Rayleigh Test), with

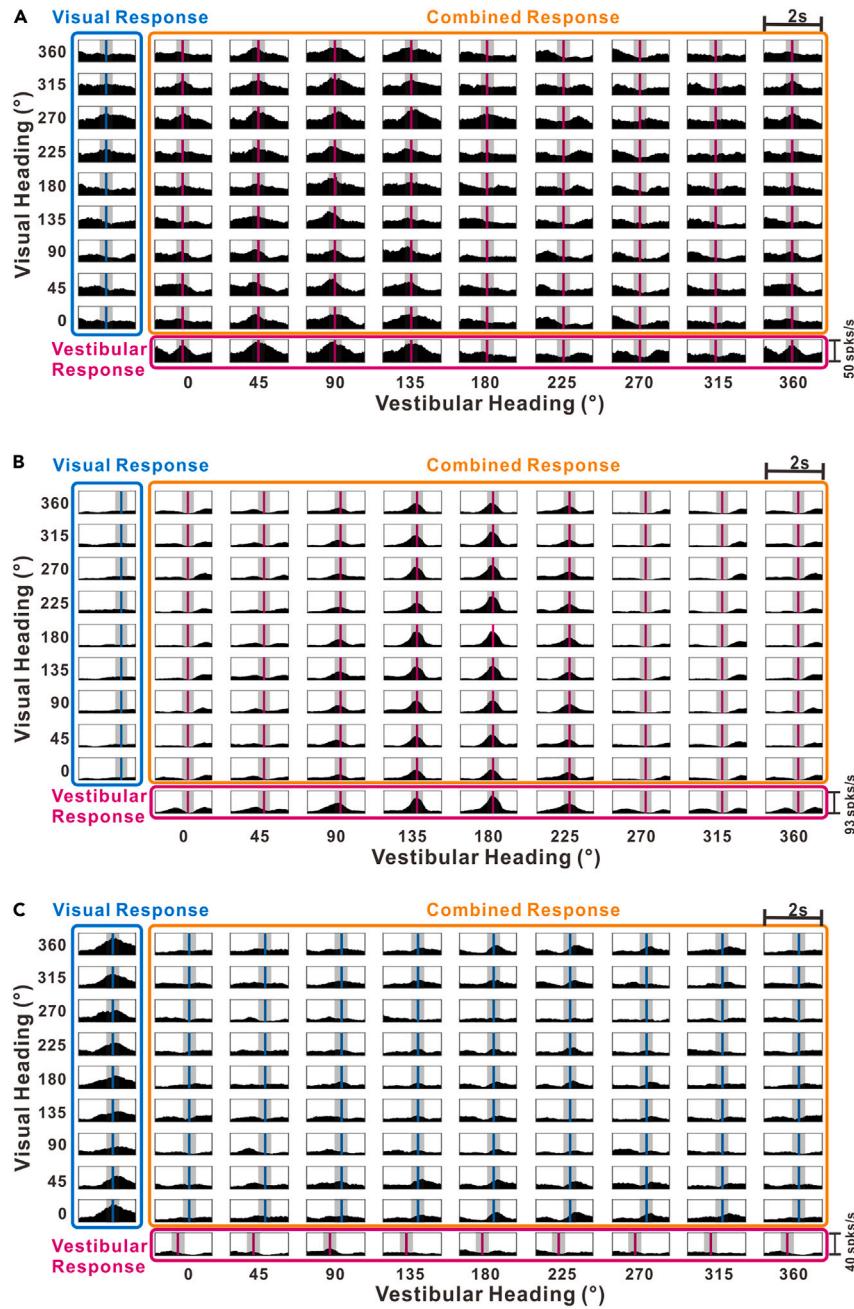


Figure 2. Responses of three example VPS neurons

Vestibular and visual headings are defined in Figure 1B.

(A) Average response PSTHs for a VPS neuron with balanced vestibular and visual responses. The bottom panels in the magenta box show the responses to the vestibular-only condition, with vertical dashed magenta lines indicating the peak time ($t_{\text{vestibular}} = 1.0$ s) when the maximum response across directions occurred during the vestibular-only condition. The left panels in the cyan box show the responses to the visual-only condition, with the cyan line indicating the visual peak time ($t_{\text{visual}} = 1.2$ s). The middle part inside the orange box shows the response PSTHs to the different combinations of vestibular and visual headings.

(B) Average response PSTHs for a VPS neuron with vestibular responses stronger than the visual responses. All the formats are the same as in Figure A.

(C) Average response PSTHs for a VPS neuron with visual responses stronger than the vestibular responses. All the formats are the same as in Figures A and B.

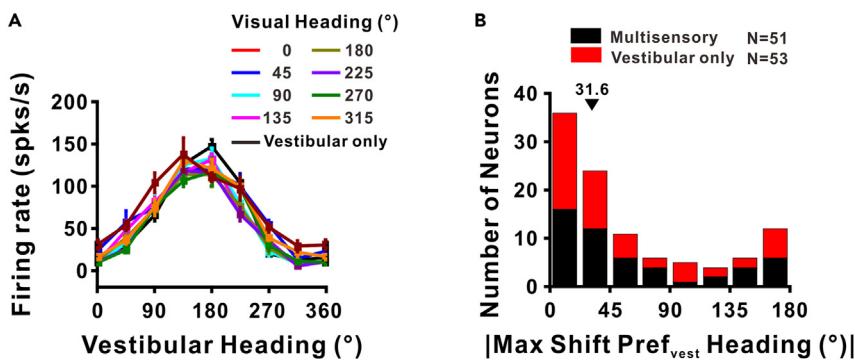


Figure 3. Tuning properties of VPS neurons

(A) The vestibular tuning curves of an example VPS neuron sampled at different visual headings. Black lines and symbols are tuning curves during the vestibular-only condition. Solid lines are vestibular tuning curves in different visual headings indicated by a different color. Error bars denote SE. Although the response magnitude changes as heading values vary away from the preferred, the preferred vestibular heading does not change.

(B) Distribution of max shift of vestibular heading preference caused by visual stimuli. Black bars, cells with significant tuning during both the vestibular and visual conditions; red bars, cells with significant tuning during the vestibular condition only. The black triangle represents the median change in the max shifts of vestibular heading preference.

mismatched visual (270°) and vestibular (45°) heading preference (opposite cells).¹⁰ Opposite neurons occupied most of the multisensory neurons in area VPS (70.6%, 36 of 51), which is consistent with previous reports.¹⁷ Thus, we did not further separate the multisensory neurons into "congruent" and "opposite" subclass as previously characterized in MSTd^{10,11} and VIP,^{18,31} but pooled all the multisensory cells together. For this multisensory neuron, the maximum unimodal responses were 46.5 spikes/sec for vestibular (azimuth_{vestibular} = 45°) and 36.5 spikes/sec for visual (azimuth_{visual} = 270°), respectively. When these two signals were applied together, the combined responses showed a paradoxical cross-modal suppression as reported in MSTd¹⁴: large responses were observed when both visual and vestibular were presented together at the heading preferences of the neuron (maximum responses: 38.5 spikes/sec, azimuth_{vestibular} = 45°, azimuth_{visual} = 270°), although still smaller than the maximum responses during the vestibular-only condition (46.5 spikes/sec, azimuth_{vestibular} = 45°). When either the vestibular or visual cue was presented at a non-preferred heading, the non-preferred cue activated the neuron when presented alone, but suppressed responses when presented together with the other preferred cue. For example, when a vestibular cue was presented at a non-preferred heading (azimuth_{vestibular} = 270°) and the visual cue was presented at a preferred heading (azimuth_{visual} = 270°), the combined responses (azimuth_{vestibular} = 270°, azimuth_{visual} = 270°) were suppressed, with firing rates (19.5 spikes/sec) smaller than the visual alone (36.5 spikes/sec). Similarly, when the vestibular cue was presented at the preferred heading (azimuth_{vestibular} = 45°) but the visual cue was presented at the non-preferred heading (azimuth_{visual} = 90°), the combined responses (azimuth_{vestibular} = 45°, azimuth_{visual} = 90°, 27.5 spikes/sec) were also smaller than the vestibular unimodal response (46.5 spikes/sec).

Cross-modal suppression was also observed in unisensory VPS neurons, as shown in Figure 2B for an example neuron with responses only to vestibular signals and Figure 2C for another example with responses only to visual signals. In Figure 2B, the maximum vestibular response was 72.5 spikes/sec. By adding the non-effective visual signals to the vestibular heading with the maximum vestibular response under vestibular-only condition, the average response during the 8 combined conditions became significantly smaller (55.8 spikes/sec, $p < 0.01$, Wilcoxon's rank-sum test). For the visual-only cell in Figure 2C, the cross-modal suppression effect was also obvious: when the non-effective vestibular signals were added to the effective visual signal, the visual responses were also reduced. As described further below, cross-modal suppression was frequently encountered in VPS.

We first examined whether the cross-modal suppression in VPS was caused by the heading preference changes affected by the vestibular-visual interaction. As most cells tuned to vestibular signals (multisensory: n = 51; vestibular only: n = 53) rather than to visual signals (visual only: n = 21) in VPS, we plotted the vestibular tuning curves when the stimulus was presented at each heading direction, then calculated the vestibular preference by the vector sum for each tuning curve. As shown in Figure 3A, the example neuron had a left-forward

heading (164°) during the vestibular-only condition (black line), and the preferred vestibular heading computed from vector sum was 152° , 162° , 157° , 154° , 155° , 157° , 165° , and 163° for visual heading at 0° , 45° , 90° , 135° , 180° , 225° , 270° , and 315° , respectively. Thus, the max shift of vestibular heading preference between combined responses and vestibular-only conditions was 12° . Across 104 neurons with significant vestibular tuning, the median change in the max shifts of vestibular heading preference was 31.6° (Figure 3B), and there were 57.7% neurons (60/104) whose max shifts of vestibular heading preference were less than 45° . Thus, the cross-modal suppression should not be caused by the tuning preference shift.

To quantify the cross-modal suppression in VPS, we considered the changes in response magnitudes for two particular conditions from the two-dimensional combined stimulus array: one was taken from the cross-section through the joint tuning at the vestibular preference and then aligned with the visual preferences of the cross-section at 0° (Figures 4A–4F); the other was taken from the cross-section through the joint tuning at the visual preference and aligned the vestibular preferences of the cross-section at 0° (Figures 4G–4L). Thus, 0° indicates when both cues are presented together at the visual and vestibular heading preferences of the neuron. As expected, visual responses of this neuron peaked at zero offsets ($\Delta = 0^\circ$) and declined with the absolute value of the offset, $|\Delta|$ (Figure 4A, cyan). All the combined responses (orange) for this example VPS neuron were smaller than the preferred vestibular responses (indicated by a magenta line in Figure 4A). In contrast, the MSTd example showed clear cross-modal enhancement when both cues were presented together at the visual and vestibular heading preferences of the neuron, as shown in Figure 4B. Then, the combined responses also declined when the value of $|\Delta|$ increased, and the diagnostic cross-modal suppression with excitatory visual responses only occurred when visual cues were further from the cell's heading preference, as shown at a relative direction of -90° in Figure 4B for this example. A similar tendency was observed at the population level, as shown in Figure 4C for VPS and Figure 4D for MSTd. For both multisensory neurons in VPS and MSTd, the combined responses declined with the absolute value of the offset. The dynamic range was smaller for vestibular-only neurons (Figures 4C and 4D) than for multisensory neurons, which was analogous to the report by Ohshiro et al.¹⁴

To further examine whether the difference between VPS and MSTd was related to the imbalance of vestibular vs. visual signals, we plotted a measure of multisensory enhancement ($R_{\text{com}}/R_{\text{vest}}$, taken from Δ Heading offset at 0°) against the relative strength of the two unimodal sensory inputs ($R_{\text{vis}}/R_{\text{vest}}$, computed from the average discharge frequency of the 400 ms time window where the peak time is located and represented maximum unimodal response). Most data points from VPS (Figure 4E) were distributed in the bottom-left quadrants, in which visual responses were smaller than the vestibular responses and the maximum combined responses were smaller than the maximum vestibular unimodal responses. For MSTd (Figure 4F), most data points were distributed in the top-right and bottom-left quadrants, indicating that visual and vestibular were more balanced in MSTd and there was a clear multisensory enhancement relative to the vestibular unimodal responses. In addition, the multisensory enhancement and the ratio of visual vs. vestibular responses were positively correlated ($r = 0.86$, $p < 0.0001$, $N = 96$, $df = 94$, Reduced Chi-Sqr test) in MSTd, but the correlation was not significant in VPS ($r = 0.05$, $p = 0.61$, $N = 104$, $df = 102$, Reduced Chi-Sqr test), indicating that the frequent cross-modal suppression in VPS might be related with the weak visual inputs.

As a comparison, Figures 4G–4L showed the results when we fixed the visual cue at the visual preferred heading. For multimodal neurons, adding vestibular input results in cross-modal enhancement when vestibular cues were at the preferred heading (0°), suggesting the visual signals could be amplified by vestibular signals in VPS. Moreover, the enhancement ($R_{\text{com}}/R_{\text{vis}}$) was positively correlated with the ratio of vestibular vs. visual responses in both VPS ($r = 0.78$, $p = 8.89 \times 10^{-16}$, $N = 72$, $df = 70$, Reduced Chi-Sqr test, Figure 4K) and MSTd ($r = 0.48$, $p = 1.24 \times 10^{-8}$, $N = 121$, $df = 119$, Reduced Chi-Sqr test, Figure 4L). This is very different from the vestibular suppression in Figures 4C and 4E, and most data points in Figure 4K were distributed in the top-right rather than in the bottom-left quadrants in Figure 4E, suggesting that the vestibular-visual interaction might be related to the imbalance of inputs. For the visual-only neurons (Figures 4I and 4J), vestibular input evoked no excitation but suppressed the combined responses, which were similar to the vestibular-only neurons in Figures 4C and 4D.

In general, VPS multisensory neurons showed a clear vestibular suppression related to the imbalance of inputs, while subadditive is clearer in area MSTd. On the other hand, the visual signals were enhanced by the vestibular inputs in VPS; in contrast, the vestibular signals were enhanced more obviously by the visual

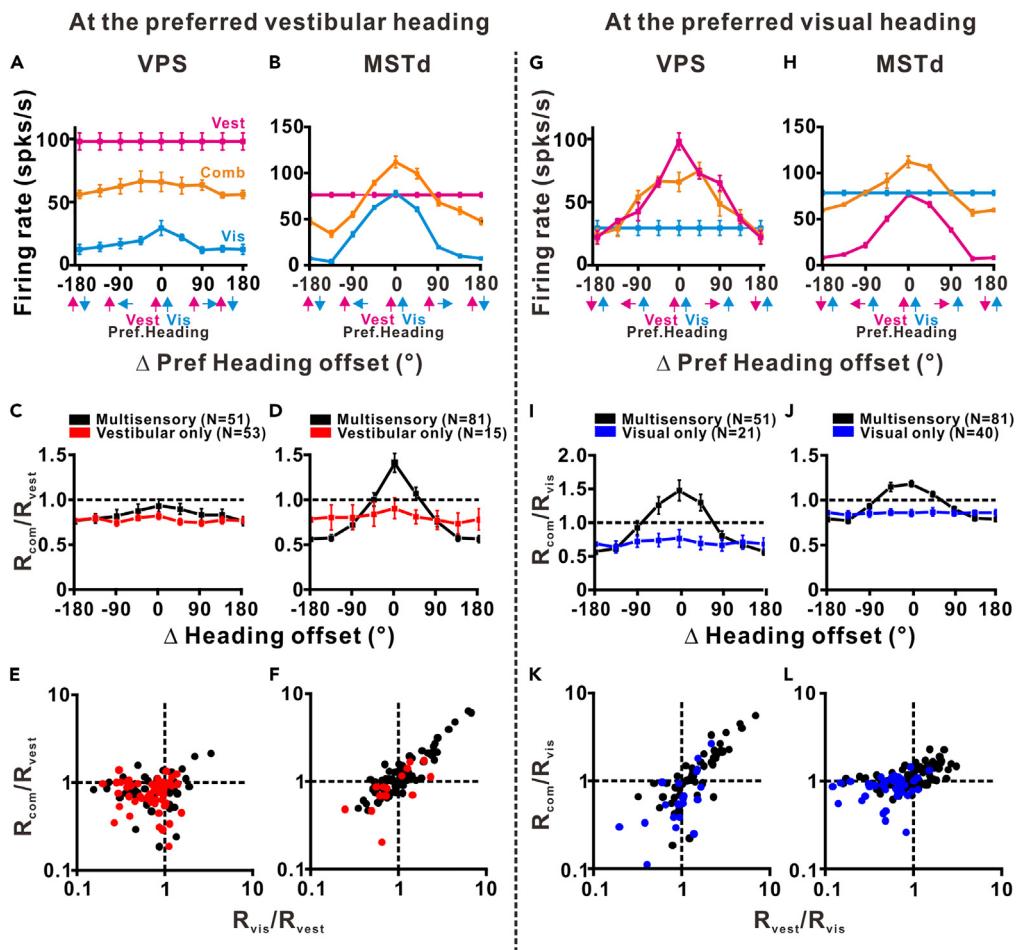


Figure 4. Response magnitude changes caused by vestibular-visual interaction in VPS, compared with MSTd

The left two columns (A–F) show the interaction results when the vestibular stimuli are fixed at the vestibular preferred heading, while the right two columns (G–L) show the interaction results when the visual stimuli are fixed at the visually preferred heading.

(A) An example VPS neuron with combined (orange), vestibular (magenta), and visual (cyan) responses plotted as a function of the offset (Δ°) of the visual heading from heading pref. 0° represents that when visual and vestibular stimuli are presented at the heading preferences of the example neuron. The heading tuning curves were fit with a wrapped Gaussian. Error bars denote SE.

(B) An example MSTd neuron. Formats are the same as in Figure A.

(C and D) The mean ratio of combined response to vestibular response in area VPS (C) and MSTd (D) are plotted as a function of the offset (Δ°). Neurons are separated into two groups: multisensory (black line and symbols) and vestibular-only neurons (red line and symbols).

(E and F) The combined ratio (R_{com}/R_{vest}) is plotted as a function of unimodal ratio (R_{vis}/R_{vest}) in VPS (E) and MSTd (F).

(G) An example VPS neuron with combined (orange), vestibular (magenta), and visual (cyan) responses plotted as a function of the offset (Δ°) of the vestibular heading from heading preference.

(H) An example MSTd neuron. Formats are the same as in Figure G.

(I and J) The mean ratio of combined response to vestibular response in area VPS (I) and MSTd (J) are plotted as a function of the vestibular heading offset (Δ°). Neurons are separated into two groups: multisensory (black line and symbols) and visual-only neurons (blue line and symbols).

(K and L) The combined ratio (R_{com}/R_{vis}) is plotted as a function of unimodal ratio (R_{vest}/R_{vis}) in VPS (K) and MSTd (L).

signals in MSTd. Such symmetry existing between VPS and MSTd seems to suggest that vestibular and visual representations might be distributed in more areas rather than a single area, and VPS and MSTd might play complementary roles during visual-vestibular integration or separation.

Conditional Fisher information analysis

To verify the above hypothesis and explore the possible role of VPS and MSTd, we decoded the vestibular and visual preference during different cue-conflict conditions by computing Fisher information³² based on the responses of multisensory neurons.

In Figures 5A and 5B, we plot the average tuning curves for VPS and MSTd multisensory neural population, conditioned on the large or small offset conditions. The red and magenta curves were aligned at preferred visual heading directions under small offset conditions ($\delta_- = \{0^\circ, 45^\circ, 315^\circ\}$, for details, see STAR Methods) and large offset conditions ($\delta_+ = \{90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ\}$), respectively. The blue and cyan curves were aligned at preferred vestibular heading directions under small offset conditions (δ_-) and large offset conditions (δ_+), respectively. From the figures, we can see there were max firing rates at 0° of reference heading direction and decreasing away from 0° . In Figures 5C and 5D, we plot the average population Fisher information for VPS and MSTd multisensory neural population at different reference heading directions. There were min Fisher information at 0° with max firing rates. The Fisher information is the corresponding neuron's derivative of the tuning curve with respect to reference heading direction. In Figures 5E and 5F, we further calculated the average population Fisher information for VPS (E) and MSTd (F) multisensory neural population and the average is taken over all heading direction conditions. Interestingly, under small offset conditions (red bar and blue bar in Figure 5E), VPS neural population contains significant information about visual heading direction φ ($p < 0.01$, T-test). Whereas under large offset conditions (magenta bar and cyan bar in Figure 5E), VPS neural population contains more information about vestibular heading direction θ ($p < 0.01$, T-test). On the contrary, the MSTd neural population contains more information about visual heading direction φ under both small (red bar and blue bar in Figure 5F, $p < 0.01$, T-test) and large offset conditions (magenta bar and cyan bar in Figure 5F, $p < 0.01$, T-test). This result suggests that VPS neural population prefers to encode information from distinct sensory modalities under large and small offset conditions. When stimuli in two modalities are congruent (offset is small, δ_-), information about visual stimuli in VPS neural population is enhanced. When stimuli are incongruent (offset is large, δ_+), information about vestibular stimuli in VPS neural population is enhanced. MSTd neural population contains more information about visual stimuli in both congruent and incongruent stimulus conditions.

The linear model fits

The above analysis revealed the difference between VPS and MSTd in encoding and decoding characteristics, given the same visual and vestibular stimuli. One possible explanation is that the vestibular and visual signals were weighted differently in these two areas. To examine this hypothesis, we first fitted the combined response of a single VPS neuron by a simple model with a linear sum of responses from the vestibular and visual conditions; for details, see the study by Morgan et al.¹² For simplicity, responses from PSTHs in Figure 2A were transformed into a color contour map with the vestibular heading along the abscissa and the visual heading along the ordinate (Figure 6A). For this example, the linear model (left panel in Figure 6B) predicted the combined response profile very well, with 86.3% of the variance accounted for. As a comparison, we also examined whether a nonlinear component was required (see Figure S1).

We further analyzed the weights from the linear model fits (w_{visual} and $w_{\text{vestibular}}$, Equation 3), which describe the strength of the contributions of each unimodal input to the combined response. As shown in Figure 6C, the visual and vestibular weights of VPS from the linear model are typically less than 1. In addition, the vestibular weights (mean \pm SD: 0.75 ± 0.22 , $N = 50$) are significantly larger than the visual weights (mean \pm SD: 0.22 ± 0.28 , $N = 50$) ($p = 1.11 \times 10^{-13}$, Wilcoxon's rank-sum test), indicating that the vestibular signals appear to be weighted much more strongly than the visual signals in VPS. Also, the vestibular signals in VPS were weighted stronger than in MSTd, with the normalized vestibular weights (mean \pm SD: 0.82 ± 0.32 , $N = 50$, left panel in Figure 6D) substantially larger than that for MSTd (mean \pm SD: 0.27 ± 0.27 , $N = 96$). In contrast, the normalized visual weights in VPS (mean \pm SD: 0.18 ± 0.32 , $N = 50$, right panel in Figure 6D) were substantially less than that for MSTd (mean \pm SD: 0.73 ± 0.27 , $N = 96$). Overall, the vestibular signals overwhelm the visual signals in area VPS.

Since the linear model uses unimodal responses to fit the data, how does the multisensory neuron in VPS weigh the vestibular signals much more strongly than the visual signals? One possible resolution is divisive normalization, where each cell in the circuit performs a linear weighted summation of its inputs, and

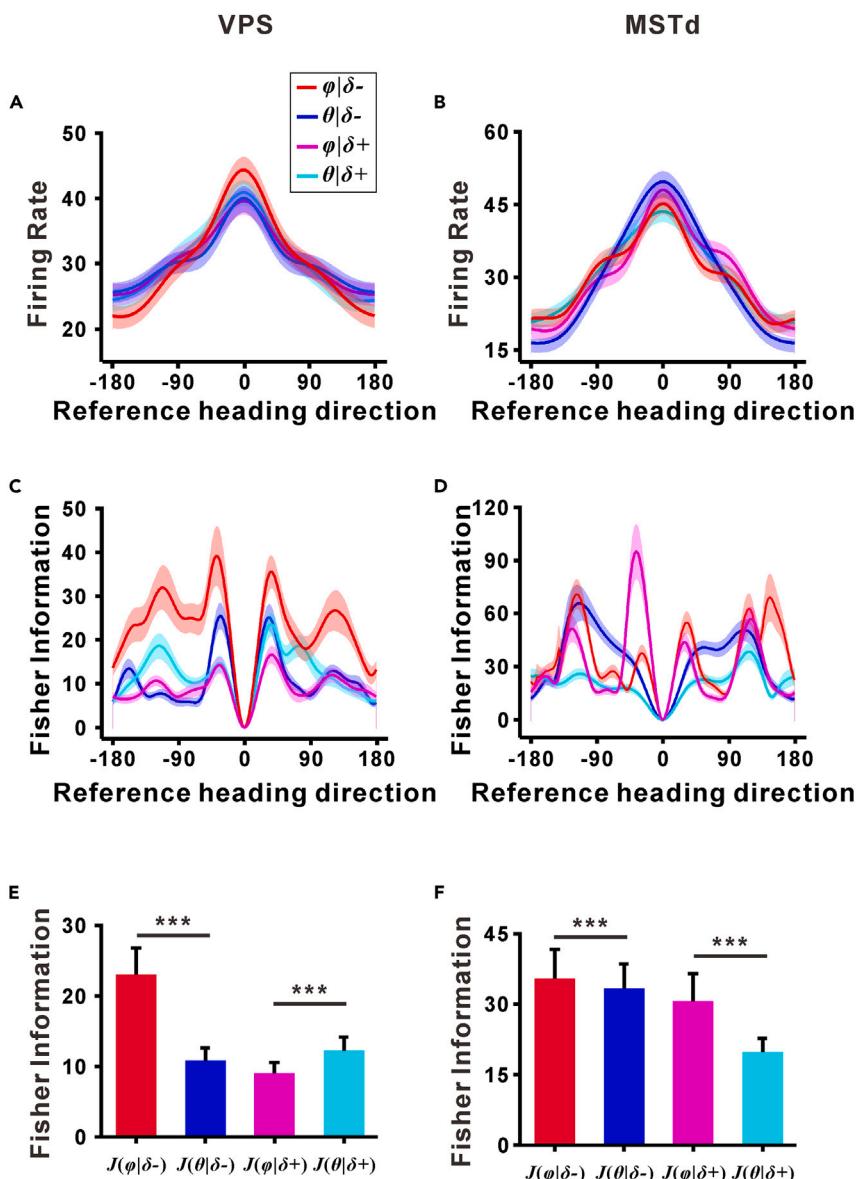


Figure 5. Population data of conditional Fisher information analysis

(A and B) Average tuning curves for VPS (A) and MSTd (B) multisensory neural population, conditioned on offset conditions. The firing rate data averaged from the multisensory neural population of preferred visual stimuli under small offset conditions($\varphi|\delta_-$, red), vestibular stimuli under small offset conditions($\theta|\delta_-$, blue), visual stimuli under large offset conditions($\varphi|\delta_+$, magenta), and vestibular stimuli under large offset conditions($\theta|\delta_+$, cyan). These tuning curves for each neuron were aligned at their preferred heading directions. Shaded indicate standard error.

(C and D) Conditional Fisher information for VPS (C) and MSTd (D) multisensory neural population. These curves for each neuron were aligned at their preferred visual (red) and vestibular (blue) heading directions under small offset conditions, while visual (magenta) and vestibular (cyan) heading directions under large offset conditions. Shaded indicate standard error.

(E and F) Average population Fisher information for VPS (E) and MSTd (F) multisensory neural population. The average is taken over all heading direction conditions. The error bar indicates standard error. Under small offset conditions (red bar and blue bar), VPS (E) and MSTd (F) neural population contain significant information about visual heading direction φ ($p < 0.01$, T-test). Under large offset conditions (magenta bar and cyan bar), VPS (E) and MSTd (F) neural population contain more information about vestibular heading direction θ ($p < 0.01$, T-test).

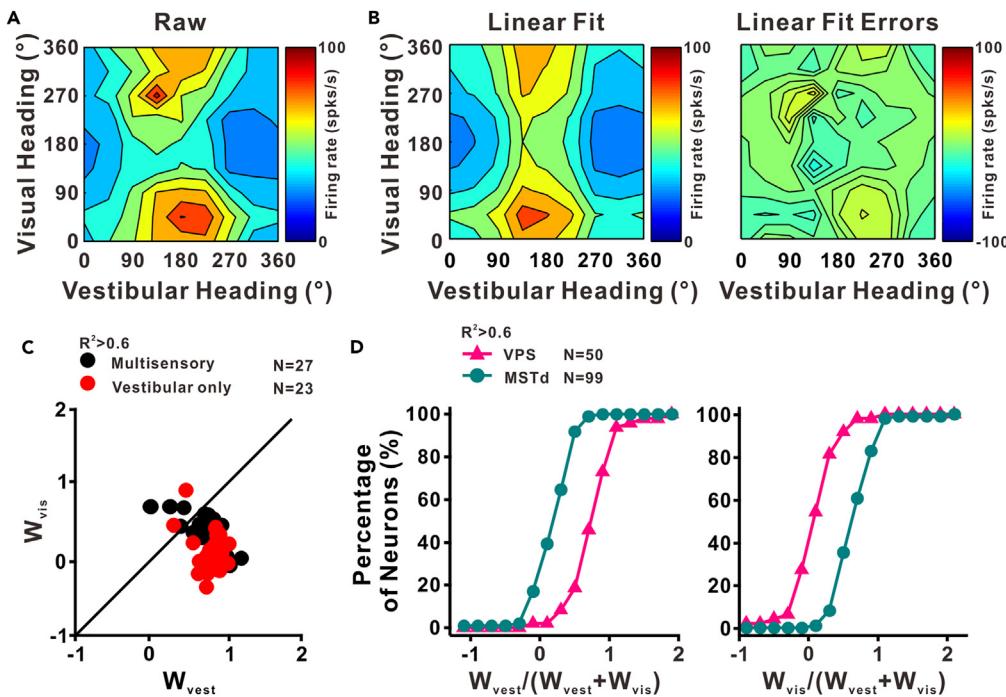


Figure 6. Example fitting data and population summary of linear models

(A) Contour map of raw combined response.

(B) Left panel: contour map of linear fitting combined response. $R^2 = 0.67$. Right panel: contour map of the errors that raw data minus linear fitting data.

(C) Scatterplot of visual weight against vestibular weight in VPS. Only cells with good fits of the linear model ($R^2 > 0.6$) are included in this analysis, and the resulting VPS neurons are 27 for multisensory (black symbols) and 23 for vestibular only neurons (red symbols).

(D) Left panel: Cumulative distribution of the vestibular contribution in VPS and MSTd. Right panel: Cumulative distribution of the visual contribution in VPS and MSTd. Related to Figure S1.

the output of each neuron is further divided by the summed activity of all neurons in the circuit. The weights can depend on the specific combination of stimulus strengths,²⁸ ranging from equal summation (where the response to summed stimuli resembles the scaled sum of the responses) to winner-take-all competition (where the response to summed stimuli resembles the response to the preferred stimulus alone).

Normalization model fits VPS population responses

To explore whether the cross-modal interaction with vestibular dominance in VPS was caused by the normalization of the population neurons' activities in the circuit, we applied the same version of the normalization model described previously in MSTd.¹⁴ There were 54 VPS neurons (3 in monkey A1, 12 in monkey A2, 17 in monkey B, and 22 in monkey K) and 108 MSTd neurons for comparison, with significant tuning either to vestibular or visual signals ($p_{vestibular} < 0.05$, $p_{visual} < 0.05$, Rayleigh Test), and trial-to-trial correlations larger than 0.4 (for details, see STAR Methods). The 216 parameters (54 neurons \times 4 free parameters) for VPS with 3456 data points (54 neurons \times 8 vestibular headings \times 8 visual headings) and 432 parameters (108 neurons \times 4 free parameters) for MSTd neurons with 6912 (108 neurons \times 8 vestibular headings \times 8 visual headings) data points were simultaneously optimized at each fitting process (for details, see STAR Methods). Since the number of parameters in the normalization model is more than that of the linear model, we calculated the Akaike information criterion (AIC) value to evaluate how well the normalization model fits the data, compared with the linear model. Usually, the smaller the AIC value, the more concise and accurate the model. Since the responses were fitted to a single cell in the linear model while the responses were fitted to the cell population in the normalization model, we first averaged the AIC values of all the cells from the linear model, and then compared them with the AIC values with the normalization model. It turns out that the AIC value of the linear model and normalization model was 25.94 and 17.96,

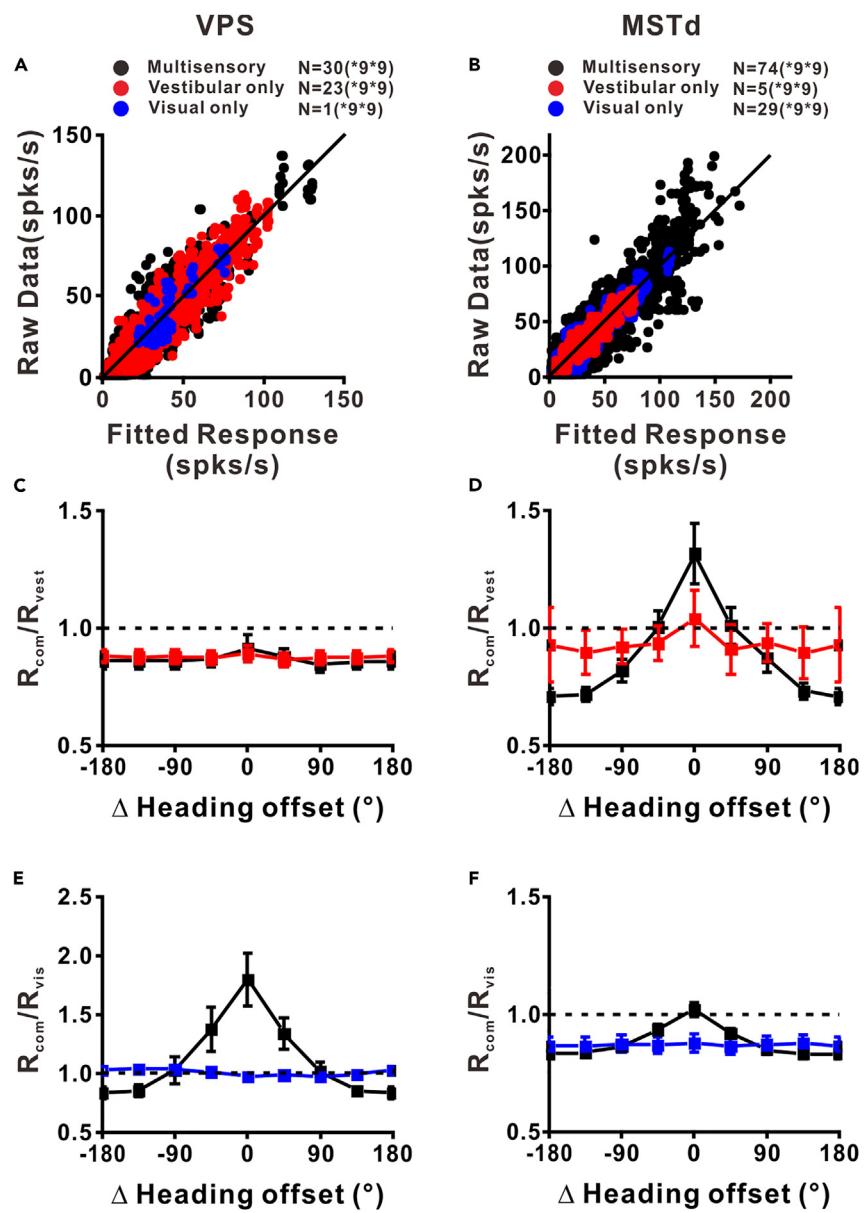


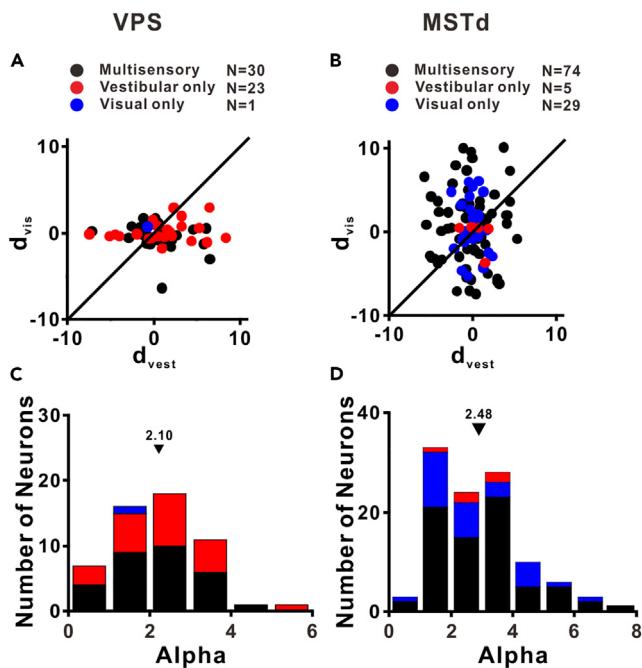
Figure 7. Summary of the goodness of fit for the normalization model

(A and B) Goodness of normalization model fit in VPS (A) and MSTd (B). The response of raw data is plotted as a function of the response of fitted data. Each neuron has 81 data points (9 vestibular conditions * 9 visual conditions). All data points are evenly distributed on both sides of the diagonal. The goodness of fit measured with R^2 is 0.90 for VPS and 0.90 for MSTd.

(C–F) Responses of normalization fitting are drawn in the same way as in Figures 4C, 4D, 4I, and 4J. The characteristics of VPS and MSTd are consistent with the original data results. Only cells with significant spatial tuning to vestibular or visual stimuli are included in this analysis.

respectively, indicating that the normalization model was more concise and accurate and can better display the characteristics of VPS neurons.

Overall, the normalization model fitted the wide range of responses rather well for both VPS and MSTd, as shown in Figure 7. Figures 7A and 7B showed the comparison of the combined responses between the data from the model fit and the observed responses for VPS and MSTd, respectively. It shows that the predicted responses (mean \pm SD: 29.3 ± 21.4 , N = $54 \times 9 \times 9$ for VPS; mean \pm SD: 30.2 ± 24.3 , N = $108 \times 9 \times 9$ for

**Figure 8. Summary of parameters from the normalization model**

(A and B) The visual weight of the normalization model is plotted as a function of vestibular weight in VPS (A) and MSTd (B). (C and D) Distribution of Alpha of the normalization model in VPS (C) and MSTd (D). Only cells with significant spatial tuning to vestibular or visual stimuli are included in this analysis.

MSTd) were not significantly different from raw data (mean \pm SD: 29.4 ± 22.9 , N = $54 \times 9 \times 9$ for VPS; mean \pm SD: 30.2 ± 26.1 , N = $108 \times 9 \times 9$ for MSTd) ($p = 0.91$ for VPS; $p = 0.94$ for MSTd, one-way ANOVA), with most data points falling along the diagonal. Importantly, the model predicted the cross-modal suppression in population responses for both multisensory and vestibular-only neurons in VPS as shown in Figure 7C. It should be noted that here we took the cross-section through the joint tuning along with the preferred vestibular heading from the model fit and aligned the visual preferences across neurons. Similar to the results in Figure 4C, the maximum responses from the model are smaller than the preferred vestibular responses ($R_{com}/R_{vest} < 1$). Besides, the model also predicted the cross-modal enhancement for multisensory neurons in MSTd when both cues are at the preferred heading (0°), and the cross-modal suppression when the visual headings are further from the preferred heading (Figure 7D). This also agreed with the raw data in MSTd (Figure 4D) very well. The normalization model predicted the enhancement of visual signals for multisensory neurons in VPS (Figure 7E) and MSTd (Figure 7F), respectively, which also agreed with the raw data in Figures 4I and 4J.

Figures 8A–8D further summarized the parameters from the normalization model fit in VPS and MSTd. For VPS, the responses were contributed mostly from the vestibular signals, as observed by the horizontal spread of the vestibular weights in Figure 8A, and the variance of vestibular weights (3.25 , N = 54 , $d_{vest,j}$ in Equation 6) in VPS were significantly larger than visual weights (1.40 , N = 54 , $d_{vis,j}$ in Equation 6) ($p = 3.19 \times 10^{-5}$, Levene's test). Thus, the population responses in VPS were mostly dominated by vestibular signals with a winner-take-all competition. In contrast, the vestibular and visual contribution was more balanced in MSTd (Figure 8B), with the variance of visual weights (4.20 , N = 108) close to the variance of vestibular weights (2.24 , N = 108) ($p = 7.92 \times 10^{-7}$, Levene's test). We also examined whether the difference in cross-modal suppression between MSTd and VPS was related to saturation. We observed the difference in semi-saturation constant α from best-fit between VPS (median: 2.10 , N = 54 , Figure 8C) and MSTd (median: 2.48 , N = 108 , Figure 8D) was significant ($p < 0.01$, Wilcoxon's rank-sum test). Thus, the stronger suppression in VPS might be related to saturation.

In summary, the same normalization operation accounts for VPS and MSTd data, suggesting that there may be a common set of operations going on across these areas.

DISCUSSION

In the current study, we recorded the responses of VPS cells to unimodal as well as congruent and conflicting combined stimuli, and then compared them with MSTd. One interesting difference between VPS and MSTd is that VPS neural population encodes distinct sensory information under large and small offset conditions but the MSTd neural population consistently contains more information about visual heading direction under both small and large offset conditions. However, the combined responses of a single neuron in area VPS can be captured very well by a linear model as MSTd, and the vestibular and visual interaction at the population level in both areas can be explained mostly by a normalization model.

Distributed representations for causal inference

The conditional Fisher information analysis shows distinct modulation effects of sensory offsets onto neural population information concerning visual and vestibular stimulus between two brain areas (Figure 5). VPS neural population can selectively choose to readout visual modal in small offset condition and represent vestibular modal in large offset condition. This segregation strategy is optimal for a task requiring subjects to report estimates about visual or vestibular stimuli according to the reliability of cues. MSTd neural population contains more information about visual stimuli in both congruent and incongruent stimulus conditions. This forced fusion strategy is optimal for a task requiring subjects to report estimates about vestibular stimuli regardless of sensory offset conditions. Interestingly, similar findings were reported in recent work in causal inference tasks with visual and proprioceptive sensory cues, with the posterior parietal cortex showing asymmetrical selective encoding (similar to VPS) and the premotor cortex showing consistent encoding (similar to MSTd).³³ In the future study, it would be interesting to verify this hypothesis by requiring subjects to report their estimates about either stimulus at a single-trial level and recording both VPS and MSTd brain areas simultaneously. This can reveal the neural computation underlying causal inference behavior.

The computational role of this asymmetrical selective encoding about different sensory inputs can be related to causal inference with multisensory evidence. In a natural environment, multiple sensory sources can arise from the same source. The process of inferring the causes of sensory inputs for perception is known as causal inference.³⁴ A typical Bayesian causal inference model assumes a binary variable to denote two different causal structures: a common cause or separate causes. In common cause conditions, stimuli are congruent across multiple sensory modalities, suggesting small offsets δ_- . In separate causes conditions, stimuli are incongruent across multiple sensory modalities, suggesting large offsets δ_+ . In common cause conditions, visual evidence is usually more reliable than vestibular evidence. Efficient perception of self-movements can be achieved by preferably encoding information about visual stimuli. However, under separate causes, visual evidence can be ambiguated by the effect of self-movements or outside stimuli movements, leading vestibular evidence to be a more trustworthy source for perceiving self-movements. Our finding suggests that VPS may take an important role in causal perception from vestibular and visual sensory modalities.

Normalization model for vestibular-visual interaction

Although there are some major differences between VPS and MSTd, most vestibular and visual interaction characteristics for these two areas can be explained by the normalization model. Normalization correctly predicts that the VPS population exhibits strong winner-take-all competition in response to sums of stimuli with different stimuli. When the visual signal is added to vestibular input, the population responses mostly reflect vestibular responses, even though the visual signal was perfectly able to elicit strong responses when presented alone. Normalization provides winner-take-all competition because the presence of multiple stimuli effectively raises the constant in the denominator, reducing the sensitivity of the neurons to the point that the weaker stimuli become unable to drive them. These results seem to agree with a previous study about the contrast normalization model to capture population activity in response regimes ranging from equal summation to winner-take-all competition.²⁸ Our study extended these previous findings to multisensory integration and established that normalization could capture the effects of vestibular-visual interaction. The summation of responses to the individual components can be weighted by stimulus strength, with a single set of parameters. Thus, the normalization model predicts that the same neural circuitry can operate across all stimulus conditions, and there may be a common set of operations going on across these areas.

Limitations of the study

Here, we demonstrated that the neural population in a vestibular dominant area VPS shows a distinct sensory preference for vestibular or visual under large and small offset conditions, which might be related to the neural computation underlying causal inference behavior. However, these speculations are based on a passive fixation task, in which the visual and vestibular stimuli are not directly related to heading judgment. To verify the exact roles of these areas, we are carrying out causal inference tasks in these areas.

On the other hand, the visual-vestibular interaction multisensory responses in both VPS and MSTd can also be explained by the divisive normalization model. However, the circuit mechanisms in VPS underlying normalization are not well understood. Further studies need to use new methodological approaches to offer an unprecedented ability to precisely determine the functional properties of distinct inhibitory circuits. The ability to selectively target and perturb specific inhibitory circuits will lead to a better mechanistic understanding of their exact role in cortical function and help reveal the biological advantage of such a variety of inhibitory processes. Furthermore, identifying the specific roles of cortical inhibitory interneurons will help us understand their contribution to neurological or cognitive disorders.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
 - Apparatus and motion stimuli
 - Experimental protocol and behavioral tasks
 - Electrophysiological recordings
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1101/23.106973>.

ACKNOWLEDGMENTS

We would like to thank Minhu Chen for the software development. This work was supported by grants from the “STI2030-major projects” (No. 2021ZD0202600), the National Basic Research Program of China (No. 32171034), the Project supported by Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX) and the Youth Innovation Promotion Association of the Chinese Academy of Sciences (to Q.Y.). We thank Prof. Dora Angelaki and Prof. Greg DeAngelis for the helpful comments.

AUTHOR CONTRIBUTIONS

A.C. designed the experiments. B.Z. performed data collection. B.Z., R.W., Q.Y., and Z.Z. analyzed the data. A.C., B.Z., R.W., and Q.Y. drafted the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 6, 2022

Revised: March 13, 2023

Accepted: May 23, 2023

Published: May 26, 2023

REFERENCES

1. Kayser, C., and Logothetis, N.K. (2007). Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.* 212, 121–132. <https://doi.org/10.1007/s00429-007-0154-0>.
2. Jones, E.G., and Powell, T.P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* 93, 793–820. <https://doi.org/10.1093/brain/93.4.793>.
3. Ghazanfar, A.A., and Schroeder, C.E. (2006). Is neocortex essentially multisensory? *Trends Cognit. Sci.* 10, 278–285. <https://doi.org/10.1016/j.tics.2006.04.008>.
4. Dokka, K., MacNeilage, P.R., DeAngelis, G.C., and Angelaki, D.E. (2015). Multisensory self-motion compensation during object trajectory judgments. *Cerebr. Cortex* 25, 619–630. <https://doi.org/10.1093/cercor/bht247>.
5. Alais, D., and Burr, D. (2004). No direction-specific bimodal facilitation for audiovisual motion detection. *Brain Res. Cogn. Brain Res.* 19, 185–194. <https://doi.org/10.1016/j.cogbrainres.2003.11.011>.
6. McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. <https://doi.org/10.1038/264746a0>.
7. Gliner, C.R., Pick, A.D., Pick, H.L., Jr., and Hales, J.J. (1969). A developmental investigation of visual and haptic preferences for shape and texture. *Monogr. Soc. Res. Child Dev.* 34, 1–40.
8. Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature* 408, 788. <https://doi.org/10.1038/35048669>.
9. Miller, R.L., Pluta, S.R., Stein, B.E., and Rowland, B.A. (2015). Relative unisensory strength and timing predict their multisensory product. *J. Neurosci.* 35, 5213–5220. <https://doi.org/10.1523/jneurosci.4771-14.2015>.
10. Gu, Y., Watkins, P.V., Angelaki, D.E., and DeAngelis, G.C. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J. Neurosci.* 26, 73–85. <https://doi.org/10.1523/jneurosci.2356-05.2006>.
11. Gu, Y., Angelaki, D.E., and DeAngelis, G.C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* 11, 1201–1210. <https://doi.org/10.1038/nn.2191>.
12. Morgan, M.L., DeAngelis, G.C., and Angelaki, D.E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron* 59, 662–673. <https://doi.org/10.1016/j.neuron.2008.06.024>.
13. Ohshiro, T., Angelaki, D.E., and DeAngelis, G.C. (2011). A normalization model of multisensory integration. *Nat. Neurosci.* 14, 775–782. <https://doi.org/10.1038/nn.2815>.
14. Ohshiro, T., Angelaki, D.E., and DeAngelis, G.C. (2017). A neural signature of divisive normalization at the level of multisensory integration in primate cortex. *Neuron* 95, 399–411.e8. <https://doi.org/10.1016/j.neuron.2017.06.043>.
15. Gu, Y., Cheng, Z., Yang, L., DeAngelis, G.C., and Angelaki, D.E. (2016). Multisensory convergence of visual and vestibular heading cues in the pursuit area of the frontal eye field. *Cerebr. Cortex* 26, 3785–3801. <https://doi.org/10.1093/cercor/bhv183>.
16. Avila, E., Lakshminarasimhan, K.J., DeAngelis, G.C., and Angelaki, D.E. (2019). Visual and vestibular selectivity for self-motion in macaque posterior parietal area 7a. *Cerebr. Cortex* 29, 3932–3947. <https://doi.org/10.1093/cercor/bhy272>.
17. Chen, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Convergence of vestibular and visual self-motion signals in an area of the posterior sylvian fissure. *J. Neurosci.* 31, 11617–11627. <https://doi.org/10.1523/jneurosci.1266-11.2011>.
18. Chen, A., DeAngelis, G.C., and Angelaki, D.E. (2011). Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex. *J. Neurosci.* 31, 12036–12052. <https://doi.org/10.1523/jneurosci.0395-11.2011>.
19. Ferster, D. (1994). Linearity of synaptic interactions in the assembly of receptive fields in cat visual cortex. *Curr. Opin. Neurobiol.* 4, 563–568. [https://doi.org/10.1016/0959-4388\(94\)90058-2](https://doi.org/10.1016/0959-4388(94)90058-2).
20. Jagadeesh, B., Wheat, H.S., and Ferster, D. (1993). Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex. *Science* 262, 1901–1904. <https://doi.org/10.1126/science.8266083>.
21. Skaliola, I., Doubell, T.P., Holmes, N.P., Nodal, F.R., and King, A.J. (2004). Functional topography of converging visual and auditory inputs to neurons in the rat superior colliculus. *J. Neurophysiol.* 92, 2933–2946. <https://doi.org/10.1152/jn.00450.2004>.
22. Carandini, M., and Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *J. Neurosci.* 20, 470–484. <https://doi.org/10.1523/jneurosci.20-01-00470.2000>.
23. Carandini, M., Heeger, D.J., and Movshon, J.A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* 17, 8621–8644. <https://doi.org/10.1523/jneurosci.17-21-08621.1997>.
24. Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9, 181–197. <https://doi.org/10.1017/s0952523800009640>.
25. Miller, K.D., and Troyer, T.W. (2002). Neural noise can explain expansive, power-law nonlinearities in neural response functions. *J. Neurophysiol.* 87, 653–659. <https://doi.org/10.1152/jn.00425.2001>.
26. Priebe, N.J., and Ferster, D. (2008). Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. *Neuron* 57, 482–497. <https://doi.org/10.1016/j.neuron.2008.02.005>.
27. Heeger, D.J. (1992). Half-squaring in responses of cat striate cells. *Vis. Neurosci.* 9, 427–443. <https://doi.org/10.1017/s095252380001124x>.
28. Busse, L., Wade, A.R., and Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron* 64, 931–942. <https://doi.org/10.1016/j.neuron.2009.11.004>.
29. Britten, K.H., and Heuer, H.W. (1999). Spatial summation in the receptive fields of MT neurons. *J. Neurosci.* 19, 5074–5084. <https://doi.org/10.1523/jneurosci.19-12-05074.1999>.
30. Louie, K., Grattan, L.E., and Glimcher, P.W. (2011). Reward value-based gain control: divisive normalization in parietal cortex. *J. Neurosci.* 31, 10627–10639. <https://doi.org/10.1523/jneurosci.1237-11.2011>.
31. Chen, A., DeAngelis, G.C., and Angelaki, D.E. (2013). Functional specializations of the ventral intraparietal area for multisensory heading discrimination. *J. Neurosci.* 33, 3567–3581. <https://doi.org/10.1523/jneurosci.4522-12.2013>.
32. Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* 7, 358–366. <https://doi.org/10.1038/nrn1888>.
33. Qi, G., Fang, W., Li, S., Li, J., and Wang, L. (2022). Neural dynamics of causal inference in the macaque frontoparietal circuit. *Elife* 11, e76145. <https://doi.org/10.7554/elife.76145>.
34. French, R.L., and DeAngelis, G.C. (2020). Multisensory neural processing: from cue integration to causal inference. *Curr. Opin. Physiol.* 16, 8–13. <https://doi.org/10.1016/j.cophys.2020.04.004>.
35. Chen, A., DeAngelis, G.C., and Angelaki, D.E. (2010). Macaque parieto-insular vestibular cortex: responses to self-motion and optic

- flow. *J. Neurosci.* 30, 3022–3042. <https://doi.org/10.1523/jneurosci.4029-09.2010>.
36. Chen, A., DeAngelis, G.C., and Angelaki, D.E. (2011). A comparison of vestibular spatiotemporal tuning in macaque parietoinsular vestibular cortex, ventral intraparietal area, and medial superior temporal area. *J. Neurosci.* 31, 3082–3094. <https://doi.org/10.1523/jneurosci.4476-10.2011>.
37. Gu, Y., DeAngelis, G.C., and Angelaki, D.E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nat. Neurosci.* 10, 1038–1047. <https://doi.org/10.1038/nn1935>.
38. Chen, A., Gu, Y., Takahashi, K., Angelaki, D.E., and DeAngelis, G.C. (2008). Clustering of self-motion selectivity and visual response properties in macaque area MSTd. *J. Neurophysiol.* 100, 2669–2683. <https://doi.org/10.1152/jn.90705.2008>.
39. Zohary, E., Shadlen, M.N., and Newsome, W.T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370, 140–143. <https://doi.org/10.1038/370140a0>.
40. Sompolinsky, H., Yoon, H., Kang, K., and Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* 64, 051904. <https://doi.org/10.1103/PhysRevE.64.051904>.
41. Shamir, M., and Sompolinsky, H. (2004). Nonlinear population codes. *Neural Comput.* 16, 1105–1136. <https://doi.org/10.1162/0899760477371759>.
42. Paradiso, M.A. (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biol. Cybern.* 58, 35–49. <https://doi.org/10.1007/bf00363954>.
43. Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nat. Neurosci.* 17, 1410–1417. <https://doi.org/10.1038/nn.3807>.
44. Ecker, A.S., Berens, P., Tolias, A.S., and Bethge, M. (2011). The effect of noise correlations in populations of diversely tuned neurons. *J. Neurosci.* 31, 14272–14283. <https://doi.org/10.1523/jneurosci.2539-11.2011>.
45. Gu, Y., Fetsch, C.R., Adeyemo, B., DeAngelis, G.C., and Angelaki, D.E. (2010). Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron* 66, 596–609. <https://doi.org/10.1016/j.neuron.2010.04.026>.
46. Sengupta, S.K. (1995). Fundamentals of statistical signal processing: estimation theory. *Technometrics* 37, 465–466. <https://doi.org/10.1080/00401706.1995.10484391>.
47. Yang, Q., Walker, E., Cotton, R.J., Tolias, A.S., and Pitkow, X. (2021). Revealing nonlinear neural decoding by analyzing choices. *Nat. Commun.* 12, 6557. <https://doi.org/10.1038/s41467-021-26793-9>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw data	This study	https://github.com/RongWang58/VPS
Software and algorithms		
Custom code	This study	https://github.com/RongWang58/VPS
MATLAB	MathWorks	R2020b
Origin	OriginLab	Origin2017

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Aihua Chen (ahchen@brain.ecnu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All raw data have been deposited at GitHub and are publicly available as of the date of publication. Accession information is listed in the [key resources table](#).
- All original code has been deposited at GitHub and is publicly available as of the date of publication. Accession information is listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Physiological experiments were performed in 5 hemispheres of 4 male rhesus monkeys (*Macaca mulatta*) weighing 6–10 kg. The surgical preparation, behavioral training, and electrophysiological recording procedures have been described in detail previously.^{10–12,17,18,35–38} Briefly, each animal was chronically implanted with a lightweight plastic ring for head restraint and with a scleral search coil in at least one eye for monitoring eye movements inside a magnetic field (Riverbend Instruments, Birmingham, AL, USA). After recovery from surgery, animals were trained using standard operant conditioning procedures to perform a fixation task described below. All animal surgical and experimental procedures were approved by the Institutional Animal Care and Use Committee at East China Normal University.

METHOD DETAILS

Apparatus and motion stimuli

Vestibular stimuli were delivered by movements of a six-degree-of-freedom motion platform (MOOG; MB-E-6DOF/12/1000Kg; Part#170-140A-1-C-1, East Aurora, NY, USA, [Figure 1A](#)), as described in detail previously.^{10,38} Visual stimuli simulated the identical translational self-motion through movements of random dots (optical flow) in a virtual 3D space of 100 cm wide, 100 cm tall, and 40 cm deep, was programmed in OpenGL and presented on the large computer screen (PHILIPS BDL4225, Royal Philips, Amsterdam, Netherlands) placed ~30 cm (88 × 88°) in front of the monkey. Star density was 0.01/cm³, with each star being a 0.15 cm × 0.15 cm triangle. Stimuli were presented stereoscopically as red/green anaglyphs and were viewed through Kodak Wratten filters (red no. 29; green no. 61, Barrington, NJ, USA). The display contained a variety of depth cues, including horizontal disparity, motion parallax, and size information. The binocular disparity of the stars ranged from 32° crossed (nearest dots at the clipping plane distance of 5 cm)

to 3° uncrossed. A near-clipping plane prevented stars from being rendered when they were closer than 5 cm to the animal's eyes.

During experiments, the monkey sat in a primate chair which was mounted on top of the motion platform and inside the magnetic field coil frame. The visual screen monitor was mounted on the front surface of the field coil frame. Thus, platform motion (vestibular stimuli) and optic flow stimuli could be presented either together or separately.

Experimental protocol and behavioral tasks

Neuronal responses from VPS were measured to eight heading directions evenly spaced every 45° in the horizontal plane under three experimental conditions. (1) vestibular condition: the monkey was translated along with one of the eight directions and only need to fix a central dot on the screen during movement. (2) visual condition: the monkey experienced optic flow simulating self-motion along with the same eight directions while the platform remained stationary. (3) combined condition: the monkey was translated by motion platform and saw the optic flow. All eight vestibular headings were paired with eight visual headings, resulting in a total of 64 combined stimuli with 8 congruent and 56 cue-conflict stimuli. The duration of the stimulus is 2 s, with an amplitude of 13 cm, a peak velocity of 30 cm/s, and a peak acceleration of ~0.1 G (9.81 m/s²). These three stimulus conditions together with a blank trial (neither translation nor optic flow) were interleaved randomly, resulting in 81 (8 vestibular only, 8 visual only, 64 combinations, and a "null" condition) trials for each repetition.

For each trial, the monkey first fixated at the central target (0.3° × 0.3°) for 200 ms, then maintained fixation within a 3° × 3° window during the 2 s stimulus presentation. The successful fixation will be rewarded with a liquid at the end of the trial. Trials in which the monkey broke fixation were aborted and discarded. We usually required five repetitions of each unique stimulus (a total of 405 trials) during electrophysiological recording. Based on the monkey and cell reactions, there are 92 cells with over five repetitions and 79 cells with over three repetitions but less than five repetitions. All cells were used for further analysis.

Electrophysiological recordings

We recorded extracellularly from single neurons using tungsten microelectrodes (Frederick Haer Company; tip diameter 3 μm; impedance, 1–2 MΩ at 1 kHz). The microelectrode was advanced into the cortex through a transdural guide tube, using a hydraulic microdrive (Frederick Haer Company, Bowdoin, ME, USA). Neural signals were amplified, band-pass filtered (400 – 5000 Hz), digitized, and recorded (AlphaOmega Instruments, Nazareth Illit, Israel). Spike times of single-neuron were isolated either with an online sorting module (AlphaLab SnR, Israel) or with offline sorting software Spike2 V8 (Cambridge Electronic Design Limited, Cambridge, UK), both of which applied template matching algorithms.

Area VPS was identified using a combination of magnetic resonance imaging scans, stereotaxic coordinates, white/gray matter transitions, and physiological response properties, as described in detail previously.¹⁷ Briefly, VPS was located just posterior to PIVC, and the multiunit responses to visual or vestibular motion extended 4~5mm anterior to posterior (for details, see^{17,35}). For medial to lateral extension, we first identified the medial tip of the lateral sulcus, then moved laterally until we no longer encountered directionally selective responses to either visual or vestibular in the multiunit activity.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis

All analyses except the normalization model were performed using custom scripts written in MATLAB (Mathworks, Natick, MA).

Spatiotemporal analysis of VPS responses

We first computed the PSTHs for each heading using 25 ms time bins, and were smoothed with a 400 ms boxcar filter.³⁵ Then we calculated the maximum response between 0.5~1.5 s after motion onset across stimulus directions.³⁵ Rayleigh Test in directional statistics was used to assess the significance of tuning in the (unimodal) vestibular and visual conditions, and we divided the neurons into four classes: "Multisensory" neurons with significant tuning to both vestibular and visual conditions ($p_{\text{vestibular}} < 0.05$ & $p_{\text{visual}} < 0.05$, Rayleigh Test); "Vestibular only" neurons with significant tuning only to vestibular ($p_{\text{vestibular}} < 0.05$ &

$p_{\text{visual}} > 0.05$, Rayleigh Test); "Visual only" neurons with significant tuning only to visual ($p_{\text{vestibular}} > 0.05 \& p_{\text{visual}} < 0.05$, Rayleigh Test); "Not-tuned" neurons without significant tuning to vestibular or visual ($p_{\text{vestibular}} > 0.05 \& p_{\text{visual}} > 0.05$, Rayleigh Test). The vestibular peak time was also used for the "Vestibular only" and "Multisensory" neurons to analyze the combined response; the visual peak time was used for the "Visual only" neuron to analyze the combined response. For each neuron with significant tuning, vestibular and visual heading preferences were calculated using the vector sum of mean responses. Since most multisensory neurons in VPS had opposite vestibular and visual preferences¹⁷, we did not further separate these neurons into congruent and opposite neurons.

Conditional Fisher information analysis

Fisher Information measures the local information about one stimulus.^{39–44} Assuming independent Poisson noise for neurons, we can compute population Fisher Information related to stimulus θ as

$$J(\theta) = \sum_{i=1}^N \frac{f'_i(\theta)^2}{f_i(\theta)} \quad (\text{Equation 1})$$

Where $f_i(\theta) = \langle r_i | \theta \rangle$ is i -th neuron's tuning curve with respect to the stimulus θ , $f'_i(\theta)$ is the corresponding neuron's derivative of the tuning curve with respect to θ .⁴⁵ Fisher Information provides an upper bound for the sensitivity of an unbiased estimator without specifying a particular estimator.⁴⁰ The inverse of the Fisher Information is the upper bound of the efficient decoder's variance with $\sigma_\theta^2 \geq 1/J$ according to Cramer-Rao bound.⁴⁶

Under the condition with multiple stimulus inputs (e.g., in our case here, θ represents vestibular stimuli and φ represents visual stimuli), the computation of the Fisher Information with respect to one stimulus requires considering the effect of the other stimuli. One way to approach it is to recompute the neural statistics by conditioning on the other relevant variable and then compute the Fisher Information accordingly.⁴⁷ Here in our experiment, we can compute the neuron's firing rate conditioned on the offset between vestibular and visual stimulus directions, $\delta = \theta - \varphi$, and then compute Fisher Information conditioning on offset δ .

$$J(\theta|\delta) = \sum_{i=1}^N \frac{f'_i(\theta|\delta)^2}{f_i(\theta|\delta)} \quad (\text{Equation 2})$$

We can summarize the effect of the stimuli offsets by categorizing offset variable into binary state, $\delta_- = \{0^\circ, 45^\circ, 315^\circ\}$, denoting state with small offset, and $\delta_+ = \{90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ\}$, denoting state with large offset. In this way, we can compute the Fisher Information of VPS multisensory neural population with respect to vestibular or visual stimuli input under small or large offset conditions: $J_{\text{VPS}}(\theta|\delta_-)$, $J_{\text{VPS}}(\theta|\delta_+)$, $J_{\text{VPS}}(\varphi|\delta_-)$, $J_{\text{VPS}}(\varphi|\delta_+)$. The Fisher Information of the MSTd neural population can be computed similarly.

The computation of conditional Fisher Information (Equation 2) requires each neuron's tuning curve and derivative of the tuning curve. To compute the tuning curve slope, we used a cubic spline function to interpolate among the coarsely sampled data points (45° spacing). The derivative of the tuning curve was obtained as the spatial derivative of the spline fit. To avoid near-zero variances, we placed a floor on firing rates at 0.5 spikes/s. Consequently, for 5/81 (16.47%) MSTd multisensory neurons and 0/51 VPS multisensory neurons, tuning curves were clipped at 0.5 spikes/s and smoothed by convolving with a Gaussian kernel (SD = 10°). This smoothing operation removed artifactual peaks in Fisher Information that resulted from clipping the tuning curve. Confidence intervals on population Fisher information were obtained using a bootstrap procedure in which random samples of neurons were generated by resampling with replacement from the population of recorded neurons. This resampling was repeated 100 times, and the standard errors on the tuning curve and Fisher Information were computed for each reference heading (shades in Figure 5A, B, C, D).

Linear and non-linear model of combined responses

The combined responses were arranged into two-dimensional arrays indexed by the vestibular and visual headings at the peak time of dominant unimodal response and visualized by plotting the data as a color-contour map (Figure 6). The significance of vestibular and visual tuning (main effects) in the combined responses, as well as their interaction, was assessed with two-way ANOVA. A significant interaction ($p < 0.05$) indicates the nonlinearities in the combined responses.

We used linear models to explore the vestibular and visual interaction in area VPS. For the linear model, combined responses (averaged across 5 repetitions) were fit by a linear combination of the weighted corresponding vestibular and visual unimodal responses.

$$r_{\text{combined}}(\theta, \varphi) = w_{\text{vestibular}} r_{\text{vestibular}}(\theta) + w_{\text{visual}} r_{\text{visual}}(\varphi) + C \quad (\text{Equation 3})$$

Here, r_{combined} is the predicted response for the combined condition, and $r_{\text{vestibular}}$ and r_{visual} are the responses in the vestibular and visual unimodal conditions, respectively. It should be noted that the spontaneous responses (measured in the blank trials) were subtracted from the mean responses in these fits. Angles θ and φ represent vestibular and visual stimulus directions. $w_{\text{vestibular}}$ and w_{visual} are the weights, while C is constant. They are chosen to minimize the sum of the squared errors between predicted and measured combined responses.

The nonlinear model includes a multiplicative nonlinearity as in the following equation:

$$r_{\text{combined}}(\theta, \varphi) = w_{\text{vestibular}} r_{\text{vestibular}}(\theta) + w_{\text{visual}} r_{\text{visual}}(\varphi) + w_{\text{product}} r_{\text{vestibular}}(\theta) r_{\text{visual}}(\varphi) + C \quad (\text{Equation 4})$$

Here, w_{product} is the weight on the multiplicative interaction term.

For each fit, the R^2 was computed as

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} \quad (\text{Equation 5})$$

where SSE is the sum squared error between the fit and the data, and SST is the sum of squared differences between the data and the mean of the data. Since the nonlinear model has more parameters than the linear model, we use a sequential F test to assess the statistical significance between the linear and nonlinear fit. If the outcome of the sequential F test is significant ($p < 0.05$), it means that the nonlinear model provides a better characterization of the data than the linear model.

Divisive normalization model

To provide a better description of the vestibular-visual interaction rule in VPS population data across the broad range of stimuli, we used a similar version of the normalization model in MSTd described previously.¹⁴ Briefly, the multisensory input to the j^{th} model VPS neuron is expressed as a weighted linear sum of the vestibular and visual inputs:

$$L_j = d_{\text{vest},j} * S_{\text{vest},j} + d_{\text{vis},j} * S_{\text{vis},j} + \text{base}_j \quad (\text{Equation 6})$$

In this equation, $d_{\text{vest},j}$ and $d_{\text{vis},j}$ represent the modality dominance weights of each neuron. base_j is a positive constant. $S_{\text{vest},j}$ and $S_{\text{vis},j}$ represent the tuning functions of vestibular and visual inputs to the multisensory neuron. The heading tuning function of the vestibular input is modeled as

$$S_{\text{vest},j} = \left(\frac{1 + \cos \varphi_{\text{vest},j}}{2} \right)^2 \quad (\text{Equation 7})$$

where $\varphi_{\text{vest},j}$ represents the angle between the vestibular heading preference of the neuron and the vestibular stimulus heading. $S_{\text{vest},j}$ is a bell-shaped function, symmetric around its peak at $\varphi_{\text{vest},j} = 0^\circ$. $\varphi_{\text{vest},j}$ can be expressed in terms of azimuth ($\hat{\varphi}_{\text{vest},j}$) components of the heading preference, as well as azimuth $\varphi_{\text{vest},j}$ components of the stimulus. $S_{\text{vis},j}$, $\hat{\varphi}_{\text{vis},j}$ and $\hat{\theta}_{\text{vis},j}$ were defined analogously.

The output of each model neuron was a divisively normalized version of its activity,^{14,24,27} given by

$$r_j = R_{\text{max},j} * \frac{L_j}{\alpha_j + \left(\frac{e_j}{N} \right) * \sum_{k=1}^N L_k} \quad (\text{Equation 8})$$

$R_{\text{max},j}$ is the maximum firing rate of the j^{th} model neuron. α_j is the semi-saturation constant. e_j determines how much normalizing effect the population activity has on each model neuron, and here e is fixed at 1.

Accordingly, there are 4 model parameters (base_j , α_j , $d_{\text{vest},j}$, and $d_{\text{vis},j}$) to be estimated in each neuron, and base_j , α_j have boundary constraints (larger than 0) but $d_{\text{vest},j}$ and $d_{\text{vis},j}$ do not have boundary constraints. To better reflect the characteristics of neurons, the cell's response met the following criteria: 1) neurons are either tuned to vestibular or visual signals ($p_{\text{vestibular}} < 0.05$, $p_{\text{visual}} < 0.05$, Rayleigh Test); 2) the trial-to-trial

correlations is larger than 0.4. These criteria were met in 54 VPS neurons (3 in monkey A1, 12 in monkey A2, 17 in monkey B, and 22 in monkey K) and 108 MSTd neurons for comparison. For VPS neurons, there were 4 (free parameters) \times 54 neurons = 216 total parameters, while there are 4 (free parameters) \times 108 neurons = 432 total parameters for MSTd neurons. These parameters need to be simultaneously optimized at each fitting process with a total of 3456 (= 54 neurons \times 8 vestibular headings \times 8 visual headings) VPS data points and 6912 (= 108 neurons \times 8 vestibular headings \times 8 visual headings) MSTd data points.

The code for the normalization model was written in Python (<https://www.python.org>), and the optimization was performed using Python `scipy.optimize` function with SLSQP (Sequential Least-Squares Programming) method with parameter boundary constraint options (base_j , α_j are larger than 0, other parameters have no boundary constraints). We started the optimization with the parameters initialized as the following conditions: base_j sampled from the normal distribution with zero to be the mean and 10 to be the standard deviation; α_j sampled from the normal distribution with zero to be the mean and 50 to be the standard deviation; $d_{\text{vest},j}$ and $d_{\text{vis},j}$ were set to be 0.1. We calculated each optimization process with 100000 iterations until convergence (SSE difference across two iterations is smaller than 0.1). SSE was 89.553 for the MSTd dataset and 69.808 for the VPS dataset. We used a (four-core) computer for these calculations. It took 288 seconds for the MSTd data set and 51 seconds for the VPS data set.