Behavioral/Cognitive

Conflict-Dependent Gain Control in Area VIP during Visual-Vestibular Self-Motion Processing

Huizhe Sun, Rong Wang, Fu Zeng, and Aihua Chen
Key Laboratory of Brain Functional Genomics (Ministry of Education), East China Normal University, Shanghai 200062, China

Visual and vestibular signals are continuously integrated to estimate self-motion, yet these sensory cues are often inconsistent under natural conditions. How cortical circuits regulate multisensory interactions under cue conflict while preserving stable heading representations remains poorly understood. Here, we recorded neurons in male macaque ventral intraparietal area (VIP) during passive movement stimulation while systematically varying visual-vestibular heading offsets. VIP neurons exhibited strong conflict-sensitive gain modulation: congruent cues enhanced neuronal responses, whereas increasing cue offsets progressively suppressed response gain. This effect was observed not only in multisensory neurons but also in neurons classified as predominantly visual or vestibular, indicating widespread cross-modal interactions across the population. In contrast, preferred-heading shifts remain modest, suggesting that cue conflict primarily modulated response gain rather than altering heading tuning. Despite substantial suppression at the single-neuron level under large cue offsets, population Fisher information was largely preserved, indicating robust heading discriminability. Further analyses revealed flexible sensory weighting in VIP, with visual and vestibular contributions shifting as cue conflict increased. Finally, a feedforward-gated normalization model substantially improved the characterization of VIP responses by allowing cue conflict to dynamically regulate the effective normalization pool, thereby capturing both multisensory enhancement at small visual-vestibular heading offset and suppression under large cue conflict. Together, these findings show that VIP exhibits conflict-dependent regulation of multisensory gain, preserving robust self-motion representations under sensory conflict.

Key words: **cue conflict; divisive normalization; self-motion perception; ventral intraparietal area; visual-vestibular integration**

Significance Statement

Visual and vestibular cues are often misaligned in natural environments, creating a challenge for accurate self-motion perception. We demonstrate that neurons in the macaque ventral intraparietal area (VIP) dynamically adjust multisensory gain according to the degree of visual-vestibular conflict. As the visual-vestibular heading offset increases, neuronal responses are progressively enhanced or suppressed, while population coding of heading direction remains largely preserved. A conflict-sensitive normalization model accounts for these response dynamics and outperforms traditional multisensory integration models, suggesting that VIP implements adaptive gain-control computations to regulate interactions between visual and vestibular signals. These results identify a cortical mechanism that links sensory conflict detection to robust perceptual coding, advancing our understanding of multisensory processing in the brain.

Introduction

Accurate self-motion perception relies on visual and vestibular cues, together with proprioception and motor-related cues (Gibson, 1950; Dichgans and Brandt, 1978; Duffy, 1998; Schlack et al., 2002; Wallace et al., 2004; Angelaki and Cullen, 2008; Gu et al., 2008; Chen et al., 2011a; DeAngelis and Angelaki, 2012; Stein et al., 2014; Laurens and Angelaki, 2017; Raiser et al., 2020; Keshavarzi et al., 2023; Graf, 2025). Consistent visual-vestibular cues improve heading perception through multisensory integration (Gu et al., 2008; Fetsch et al., 2009; Butler et al., 2010). In natural environments, however, dynamic scenes and sensory noise frequently produce cue conflicts, requiring flexible integration strategies that prevent inappropriate cue fusion while maintaining accurate estimates of self-motion (Warren and Hannon, 1990; Dokka et al., 2019; Keshavarzi et al., 2023; Lacquaniti et al., 2023).

Previous studies have shown that cortical self-motion areas differ substantially in how they represent visual and vestibular information. Neurons in dorsal medial superior temporal area (MSTd) exhibit strong heading selectivity for optic flow and also integrate vestibular inputs during self-motion processing (Lappe et al., 1996; Duffy, 1998; Bremmer et al., 1999; Gu et al.,

Received June 10, 2026; revised Aug. 29, 2026; accepted Sept. 2, 2026.

Author contributions: A.C. designed research; H.S. and R.W. performed research; H.S. and F.Z. analyzed data; H.S., F.Z., and A.C. wrote the paper.

This work was supported by grants from the "technology innovation 2030–major projects" on brain science and brain-like computing of the Ministry of Science and Technology of China (No. 2021ZD0202600), the National Basic Research Program of China (No. 32171034) to A.C., and the ISF-NSFC joint research program to A.C. (No. 32061143003), and National Natural Science Foundation of China (No. 32400826) to F.Z.

The authors declare no competing financial interests.

Correspondence should be addressed to Aihua Chen at ahchen@brain.ecnu.edu.cn or Fu Zeng at fzeng@bio.ecnu.edu.cn.

This paper contains supplemental material available at: https://doi.org/10.1523/JNEUROSCI.0975-26.2026
https://doi.org/10.1523/JNEUROSCI.0975-26.2026

Copyright © 2026 the authors

2006; Gu et al., 2008). Under cue conflict, MSTd responses remain largely dominated by visual signals, consistent with a relatively stable visual-weighted strategy (Gu et al., 2008; Morgan et al., 2008). In contrast, neurons in visual posterior sylvian area (VPS) demonstrated vestibular-dominant winner-take-all-like responses under visual-vestibular conflict (Chen et al., 2011c; Zhao et al., 2023). These findings indicate that cortical self-motion regions differ substantially in their sensory weighting strategies, raising the possibility that additional areas may implement more flexible mechanisms for regulating multisensory interactions.

The ventral intraparietal area (VIP) is a strong candidate for such flexible computations. VIP receives convergent visual, vestibular, somatosensory, and motor-related inputs and contributes to heading perception, multisensory integration, and spatial processing (Schaafsma and Duysens, 1996; Duhamel et al., 1998; Bremmer et al., 2001; Bremmer et al., 2002a,b; Chen et al., 2011a,b, 2013). VIP neurons exhibit linear visual and vestibular self-motion responses, with preferred directions that can be approximately congruent or opposite (Schlack et al., 2002; Chen et al., 2011a, 2013). In addition, VIP neurons are known to be strongly modulated by spatial context, action, task demands, and multisensory congruency (Zaidel et al., 2017; Chen et al., 2021; Churan et al., 2021; Zaidel et al., 2021; Zeng et al., 2023). VIP may therefore regulate visual and vestibular influence according to sensory consistency rather than relying on a fixed modality-dominant hierarchy.

Divisive normalization provides a candidate mechanism for such regulation. It is a canonical cortical computation that combines multiple inputs and generates nonlinear interactions, such as multisensory enhancement, cross-modal suppression, and response competition (Carandini and Heeger, 2011; Ohshiro et al., 2011). Related normalization frameworks have also been used to describe visual-vestibular interactions in cortical self-motion areas (Ohshiro et al., 2017; Zhao et al., 2023). However, conventional normalization frameworks typically assume a fixed normalization pool and static gain control across sensory conditions. Whether normalization mechanisms can flexibly adjust sensory weighting according to the degree of visual-vestibular conflict remains unknown.

Here, we recorded single-neuron activity from macaque VIP during passive movement while systematically varying visual-vestibular heading offsets. We asked whether cue conflict alters heading tuning or instead modulates the gain of an otherwise stable representation. To place the computational properties of VIP within the broader cortical self-motion network, we compared VIP with MSTd and VPS, which provide visual-dominant and vestibular-dominant reference profiles, respectively. These comparisons allowed us to assess whether VIP exhibits an intermediate or distinct response profile under the same experimental paradigm. We further tested whether a normalization-based computational framework could account for the response patterns observed in VIP and distinguish them from those in MSTd and VPS. Our results show that VIP dynamically regulates multisensory gain according to the degree of visual-vestibular cue conflict while preserving robust population coding of heading direction.

Materials and Methods

Subjects and surgery. Two male rhesus monkeys (*Macaca mulatta*, P and R), weighing 8–10 kg, were used in this experiment. Before electrophysiological recordings, the monkeys were gradually trained to adapt to the laboratory environment and maintain a stable sitting posture in a specially designed primate chair. Subsequently, under sterile conditions, both monkeys were implanted with a head fixation device and a scleral coil in the left eye for recording eye movement signals. After the animals had fully recovered postsurgery, they were trained to complete experimental tasks using standard operating conditional reflexes. All animal surgeries and experimental procedures were performed in strict accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and were approved by the Experimental Animal Ethics Committee of East China Normal University.

Equipment setup and motion stimuli. In the experiment, macaques passively experienced vestibular and visual stimuli representing linear self-motion in the horizontal plane. Vestibular stimulation consisted of physical linear translations delivered by a six-degree-of-freedom motion platform (MOOG; MBE-6DOF/12/1,000Kg; Part#170-140A-1-C-1; Fig. 1*A*). Visual stimulation consisted of random-dot optic flow simulating linear translation of the observer through a virtual three-dimensional environment. This virtual 3D space measured 100 cm × 100 cm × 40 cm and was generated using OpenGL programming. The visual stimuli were displayed on a large monitor (SONY KD-50X85K, Sony Corporation; screen size 88 × 88 cm) positioned 30 cm in front of the monkeys. The monitor was mounted on the motion platform and moved together with the animal, thereby remaining fixed relative to the animal during platform translation. The random point density was 0.01/cm^3, with each point being a 0.15 cm × 0.15 cm triangle.

Visual stimuli were presented as red/green anaglyph stereo images on the display. Monkeys viewed the stimuli through Kodak Wratten filters (red 29; green 61), which separated the red and green images between the two eyes. The display included various depth cues, including horizontal parallax, motion parallax, and size cues. A near clipping plane of 5 cm was applied to prevent dots from being rendered closer than 5 cm to the animal's eyes.

Both vestibular and visual self-motion stimuli followed a Gaussian velocity profile. The stimulus duration was 2 s with a total displacement of 13 cm, a peak velocity of 0.31 m/s, and a peak acceleration of 1.13 m/s^2 (Fig. 1*B*). On each trial, the platform started from a central home position and translated 13 cm along one of the eight tested headings, coming to rest at the endpoint after 2 s. Thus, each trial contained a single outward translation rather than a back-and-forth movement. After stimulus completion, the platform returned to the central starting position before the next trial. The Gaussian velocity profile included an acceleration phase followed by a deceleration phase; the reported peak acceleration of 1.13 m/s^2 refers to the maximum acceleration magnitude. The same temporal velocity and acceleration profiles were used for all eight headings, differing only in movement direction.

The stimuli varied in a horizontal plane, with eight evenly spaced directions (0–315°, step size 45°). Three stimulus conditions were presented. In the vestibular-only condition, the platform translated along one of the eight headings while the display showed only the central fixation point; no random-dot pattern or optic flow was presented. In the visual-only condition, the platform remained stationary while random-dot optic flow simulated translation along one of the same eight headings. In the combined visual-vestibular condition, platform translation and optic flow were presented synchronously. All eight vestibular headings were factorially combined with all eight visual headings, yielding 64 combinations that were randomly interleaved across trials. Cue combinations were defined as congruent when the visual and vestibular cues specified the same observer self-motion heading, despite the opposite physical directions of the corresponding optic flow and vestibular motion signals (8 combinations), and as incongruent when the two cues specified different self-motion headings (56 combinations; Fig. 1*C*).

Behavioral task. During electrophysiological recordings, the animals performed a fixation task. At the start of each trial, the macaques fixated on a central target (0.3° × 0.3°) for 200 ms and then maintained their fixation within a 3° × 3° window throughout the subsequent 2 s stimulus period. Macaques that successfully completed the fixation received a liquid reward at the end of the trial; trials that failed to complete the fixation were immediately terminated and discarded.

Although the fixation window was set to 3° × 3° online, we performed additional offline analysis of eye position traces. Trials containing

microsaccades or slow drifts exceeding 1° from the fixation point during the stimulus period were detected and excluded from further analysis to ensure that neural responses were not contaminated by eye movements. Each complete stimulus set contained 81 unique conditions (8 vestibular-only + 8 visual-only + 64 combined + 1 blank), typically repeated five times per condition (~405 trials/neuron); neurons with fewer than three repetitions were not included in subsequent analyses.

Electrophysiological recordings. We used tungsten microelectrodes (Frederick Haer; tip diameter ~3 μm, impedance 1–2 MΩ at 1 kHz) to perform extracellular recordings of single neurons in the VIP brain region of rhesus monkeys. The microelectrodes were implanted into the cortex via a dural catheter using a hydraulic microactuator (Frederick Haer). The raw neural signals were amplified and bandpass filtered (400–5,000 Hz) before being digitized at a sampling rate of 25 kHz using an Alpha Omega system (Alpha Omega Instruments, Nazareth Elite). Online action potential classification was performed using a Tempo system, with pulse firing times and all behavioral events recorded synchronously at a 1 ms time resolution; offline reclassification was performed if online classification was unsatisfactory. The localization of the target brain region VIP was determined based on gray–white matter transition characteristics, magnetic resonance imaging (MRI) scans, stereotactic coordinates, and neuronal physiological response characteristics, consistent with previous studies (Chen et al., 2013). No postmortem histological verification of the recording sites was performed.

Datasets for cross-area comparisons. For the cross-area comparisons, we additionally analyzed previously published neuronal datasets recorded in VPS and MSTd, as reported previously (Morgan et al., 2008; Zhao et al., 2023). No new electrophysiological recordings from VPS or MSTd were performed in the present study. These datasets were collected using the same visual-vestibular heading paradigm used for the VIP recordings in the present study. The datasets comprised 171 VPS neurons recorded from four monkeys and 158 MSTd neurons recorded from two monkeys.

Neural-response analysis of VIP and neuronal classification. We first calculated the poststimulus time histogram (PSTH) for each direction at 25 ms intervals and smoothed it using a 400 ms square wave filter. Then, we extracted the maximum response for each stimulus direction within a window of 0.5–1.5 s after the start of motion. The significance of neuronal direction tuning under vestibular and visual conditions was assessed using the Rayleigh test (single-modality), and neurons were classified into four categories based on this (Fig. S1*A–C*):

1. Multisensory neuron: showed significant direction tuning to both vestibular and visual stimuli ($p_{\text{vestibular}} < 0.05$ and $p_{\text{visual}} < 0.05$, Rayleigh test)
2. Vestibular-only neuron: showed significant direction tuning only to vestibular stimuli ($p_{\text{vestibular}} < 0.05$ and $p_{\text{visual}} > 0.05$, Rayleigh test)
3. Visual-only neuron: showed significant direction tuning only to visual stimuli ($p_{\text{vestibular}} > 0.05$ and $p_{\text{visual}} < 0.05$, Rayleigh test)
4. Untuned neuron: showed no significant direction tuning to either vestibular or visual stimuli ($p_{\text{vestibular}} > 0.05$). And $p_{\text{visual}} > 0.05$ (Rayleigh test)

Under combined stimulation conditions, vestibular peak time was used to analyze responses of vestibular and multisensory neurons, whereas visual peak time was used to analyze responses of visual neurons. For neurons with significant tuning, their vestibular and visual preferred headings were calculated from the vector sum of the average responses. The difference between vestibular and visual preferred headings in VIP were broadly distributed, rather than forming clearly separable populations with congruent (<90°) or opposite (>90°) preferred headings, as defined in previous work (Chen et al., 2011a; Fig. S1*E,F*). Therefore, multisensory neurons were analyzed as a single population in the main analyses.

Multisensory gain modulation relative to preferred heading. To characterize how the response to one sensory cue was modulated by the other cue, combined responses were compared with the corresponding unimodal responses as a function of the deviation of the nonreference cue from its own preferred heading. In the vestibular-referenced condition, vestibular heading was fixed at the neuron's vestibular preferred heading while visual heading varied relative to the neuron's visual preferred heading. This angular difference was termed the visual preferred-heading deviation. In the visual-reference condition, visual heading was fixed at the neuron's visual preferred heading, whereas vestibular heading was varied relative to the neuron's vestibular preferred heading, termed the vestibular preferred-heading deviation. Negative and positive values indicate counterclockwise and clockwise deviations from the corresponding preferred heading, respectively, and 0° indicates that the stimulus was presented at the preferred heading.

Multisensory gain was quantified as $R_{\text{com}}/R_{\text{vest}}$ in the vestibular-reference condition and $R_{\text{com}}/R_{\text{vis}}$ in the visual-reference condition. R_{com} denotes the combined response, R_{vest} denotes the vestibular-only response, and R_{vis} denotes the visual-only response. Response ratios >1 indicated multisensory enhancement, whereas values below 1 indicated suppression relative to the unimodal response.

To examine whether multisensory gain depended on relative unimodal strength, vestibular-referenced gain ratios ($R_{\text{com}}/R_{\text{vest}}$) were plotted against visual-to-vestibular response ratios ($R_{\text{vis}}/R_{\text{vest}}$) across different visual preferred-heading deviations, and visual-referenced gain ratios ($R_{\text{com}}/R_{\text{vis}}$) were plotted against vestibular-to-visual response ratios ($R_{\text{vest}}/R_{\text{vis}}$) across different vestibular preferred-heading deviations.

Population Fisher information analysis. To quantify the precision of population heading coding, Fisher information (FI) was computed from population tuning curves under visual-only, vestibular-only, and combined conditions. Assuming independent Poisson variability across neurons, the population FI with respect to heading direction θ can be written as follows:

$$J(\theta) = \sum_{i=1}^{N} \frac{f_i'(\theta)^2}{f_i(\theta)}, \tag{1}$$

where $f_i(\theta)$ denotes the tuning curve of the i-th neuron (mean firing rate) as a function of θ and $f_i'(\theta)$ is the derivative of the tuning curve with respect to θ. FI provides an upper bound on the sensitivity of an unbiased estimator (Cramér–Rao bound), such that the variance of an efficient decoder is bounded by $\sigma_\theta^2 \geq 1/J(\theta)$. Since neurons were recorded sequentially, noise correlations between neurons could not be directly computed. Thus, our FI estimates represent the information available in the "shuffled" population (ignoring correlations), which corresponds to the upper bound of encoding accuracy accessible by an optimal linear decoder (Gu et al., 2010).

Under combined visual-vestibular stimulation, responses depended on both visual and vestibular headings. To examine how cue conflict affected population discriminability, we computed FI separately for combined responses grouped by the angular offset between visual and vestibular headings, defined as $\Delta = \min(|\theta_{\text{vis}} - \theta_{\text{vest}}|, 360^\circ - |\theta_{\text{vis}} - \theta_{\text{vest}}|)$. Here, Δ represents the absolute angular difference between the visual and vestibular headings, folded into the range of 0–180°. Accordingly, the zero-offset condition ($\Delta = 0°$) corresponded to the previously defined congruent condition, whereas increasing values of Δ indicated progressively greater visual-vestibular cue conflict. Because stimulus headings were sampled at 45° intervals, Δ took one of five values: 0°, 45°, 90°, 135°, or 180°. For each offset condition, clockwise and counterclockwise stimulus pairs with the same absolute visual-vestibular angular difference were pooled.

For comparison with combined responses, FI was also computed for visual-only and vestibular-only tuning curves using the same procedure. Therefore, for each brain area, FI was estimated for five conditions: visual-only, vestibular-only, combined offset 0°, combined offset 90°, and combined offset 180°.

FI was calculated at each interpolated heading direction according to the following:

$$J(\theta|\Delta) = \sum_{i=1}^{N} \frac{f_i'(\theta|\Delta)^2}{f_i(\theta|\Delta)}, \tag{2}$$

where $f_i(\theta|\Delta)$ denotes the tuning curve of the i-th neuron under a given offset condition Δ. To avoid numerical instability caused by near-zero firing rates, a minimum firing-rate floor was applied before division. FI was then averaged across heading directions to obtain a mean FI value for each condition.

FI (Eq. 2) requires both the tuning curves and their derivatives. Because headings were sampled at eight discrete directions (45° spacing), tuning curves were interpolated using cubic spline interpolation, and derivatives were obtained from the spatial derivative of the spline fits. To prevent numerical instability due to near-zero firing rates, a minimum firing-rate floor was imposed: when $f_i(\theta|\Delta) < 0.5$ spikes/s, firing rates were clipped to 0.5 spikes/s. For tuning curves affected by clipping that produced artifactual peaks in FI, we further smoothed the tuning curves by convolving with a Gaussian kernel (SD = 10°) to remove clipping-induced artifacts.

Confidence intervals for population FI were obtained using a bootstrap resampling procedure. Neurons were resampled with replacement to generate pseudopopulations, and this procedure was repeated 100 times. For each bootstrap sample, population tuning curves and FI were computed at each reference heading direction, and the standard errors were used to generate the shaded error regions shown in the figures.

Offset-specific sensory weighting model. To quantify how visual and vestibular signals contribute to the combined response, we modeled the multisensory interaction using a linear weighted model. For each neuron, the combined responses were first arranged into a two-dimensional response matrix indexed by vestibular heading θ_{vest} and visual heading θ_{vis} at the peak time of the dominant unimodal response (as in Fig. 7*B*). These matrices were visualized as color-contour maps to illustrate response structure across stimulus combinations.

The presence of vestibular and visual main effects, as well as their interaction, was first assessed using a two-way ANOVA on the combined responses. A significant interaction term ($p < 0.05$) indicated nonlinear integration between the two modalities.

To quantify the relative contributions of sensory cues, we modeled the combined response as a weighted sum of unimodal responses. Unlike previous approaches that assume fixed weights, we estimated weights separately for each of the five offset conditions defined above ($\Delta = 0°, 45°, 90°, 135°, 180°$). This offset-specific fitting approach allowed us to examine whether the relative contributions of visual and vestibular inputs changed with cue conflict, without assuming a single fixed pair of weights across all stimulus combinations.

For a given neuron and offset condition Δ only stimulus pairs satisfying $\theta_{vis} = \theta_{vest} + \Delta \bmod 360°$ were included. The combined responses under this constraint were fit using the following:

$$r_{combined}(\theta_{vest}, \theta_{vis}|\Delta) = w_{vest}(\Delta) r_{vest}(\theta_{vest}) + w_{vis}(\Delta) r_{vis}(\theta_{vis}) + C\,, \tag{3}$$

where $r_{combined}$ is the measured firing rate under the combined condition, r_{vest} and r_{vis} are responses under the firing rates under the vestibular-only or visual-only conditions, $w_{vest}(\Delta)$ and $w_{vis}(\Delta)$ are the vestibular and visual weights for that offset, and C is a constant term. Spontaneous firing rates measured from blank trials were subtracted before all fittings. Model parameters were obtained by minimizing the sum of squared errors (SSE) between the predicted values and the measured combined responses.

The goodness of fit was indicated by the coefficient of determination R^2:

$$R^2 = 1 - \frac{SSE}{SST}, \tag{4}$$

where SST is the total sum of squares of the data relative to the mean. We estimated $w_{vest}(\Delta)$ and $w_{vis}(\Delta)$ separately for each neuron under the five offset conditions and then calculated the mean and standard error of the weights at the population level for VIP, VPS, and MSTd (Fig. 7*D–F*). This analysis reveals the continuous variation pattern of visual and vestibular weights with offset across different brain regions under sensory conflict conditions.

Feedforward-gated normalization model. To describe multisensory interactions under visual-vestibular cue conflict, we adapted the divisive normalization framework proposed by Ohshiro et al. (2011). Because the standard normalization model has been described previously, we focused here on the extension introduced in the present study. In the standard framework, the combined response is determined by a weighted sum of visual and vestibular inputs followed by divisive normalization. This formulation assumes that the effective normalization strength is fixed across cue-offset condition.

For the j-th neuron, the excitatory drive under vestibular heading v and visual heading s was defined as follows:

$$L_j(v, s) = b_j + d_{vest,j} S_{vest,j}(v) + d_{vis,j} S_{vis,j}(s), \tag{5}$$

where b_j is a baseline input term and $d_{vest,j}$ and $d_{vis,j}$ represent the vestibular and visual input weights.$S_{vest,j}(v)$ and $S_{vis,j}(s)$ denote the vestibular and visual tuning templates. The vestibular template was modeled as follows:

$$S_{vest,j}(v) = \left(\frac{1 + \cos\phi_{vest,j}(v)}{2}\right)^2, \tag{6}$$

$S_{vest,j}(v)$ is a bell-shaped function, symmetric around its peak at $\phi_{vest,j} = 0.\phi_{vest,j}$can be expressed in terms of azimuth ($\phi_{vest,j}$) components of the heading preference, as well as azimuth $\hat{\phi}_{vest,j}$ components of the stimulus. $S_{vis,j}(s)$ was defined analogously.

To account for the offset-dependent gain modulation observed in VIP, we introduced a feedforward conflict-dependent gate into the normalization term. The gate was defined as follows:

$$G(\Delta) = \frac{1 - \cos\Delta}{2}, \tag{7}$$

where Δ is the absolute visual-vestibular angular offset defined above. The predicted response of the feedforward-gated normalization (FGN) model was as follows:

$$R_j^{FGN}(v, s) = R_{max,j}(v, s) \frac{L_j(v, s)}{\alpha_j + [1 + \lambda G(\Delta)] P_{norm}}, \tag{8}$$

where α_j is the semisaturation constant, P_{norm} is the normalization pool defined as in the standard normalization model, and λ is a fitted parameter that controls the strength of conflict-dependent normalization. When $\lambda = 0$, the FGN model reduces to the original normalization model. When $\lambda > 0$, the effective normalization strength increases with visual-vestibular heading disparity, because the normalization pool is scaled by $1 + \lambda G(\Delta)$. Therefore, congruent cues are modeled with the original normalization strength, whereas larger cue discrepancies produce stronger normalization-like suppression.

For model fitting, we included neurons that met two criteria: first, neurons showed significant tuning to either vestibular or visual stimuli ($p_{vest} < 0.05$ or $p_{vis} < 0.05$, Rayleigh test); second, trial-to-trial response reliability was larger than 0.4. These criteria were met by 20 VIP neurons, including 12 neurons from Monkey P and 8 neurons from Monkey R, 30 VPS neurons, and 74 MSTd neurons. The same criteria were applied across all three areas before model comparison and independently of which model was favored. Therefore, the model-based conclusions are restricted to the tuned and reliable neurons included in the analysis.

Both the standard normalization model and the FGN model were fitted to the responses of every included neuron. For each neuron, the fitting used responses to the 64 visual-vestibular heading combinations. Thus, the model was fit to 1,280 data points for VIP ($20 \times 8 \times 8$), 1,920 data points for VPS ($30 \times 8 \times 8$), and 4,736 data points for MSTd ($74 \times 8 \times 8$).

The code for the normalization model was written in Python, and the optimization was performed using Python scipy.optimize function with

SLSQP (Sequential Least-Squares Programming) method with parameter boundary constraint options (b_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: b_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.

Model parameters were estimated by minimizing the squared error between predicted and observed responses across all visual-vestibular heading combinations. Model performance was evaluated using R^2, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). Lower AIC and BIC values indicate a better balance between goodness of fit and model complexity. Model differences between the FGN and standard normalization models were calculated as $\Delta\text{AIC} = \text{AIC}_{\text{FGN}} - \text{AIC}_{\text{standard}}$ and $\Delta\text{BIC} = \text{BIC}_{\text{FGN}} - \text{BIC}_{\text{standard}}$. Therefore, negative values of ΔAIC or ΔBIC values favor the FGN model, whereas positive values favor the standard normalization model.

Results

We recorded 114 VIP neurons from the right hemisphere of two animals (93 in Monkey P and 21 in Monkey R), including 45 multisensory neurons, 25 vestibular-only neurons, 27 visual-only neurons, and 17 untuned neurons. Neural responses were measured during passive movement stimulation in three conditions: vestibular-only, visual-only, and combined visual-vestibular stimulation. Vestibular stimulation consisted of physical linear translation delivered by a motion platform, whereas visual stimulation consisted of optic flow presented on a frontoparallel screen to simulate linear translation (Fig. 1*A*,*B*). Each neuron was tested with interleaved trials including 8 vestibular headings, 8 visual headings, 64 visual-vestibular heading combinations, and a null condition. Combined condition included 8 congruent and 56 incongruent stimuli (Fig. 1*C*). During stimulus presentation, monkeys were simply required to maintain fixation at a central target on the screen.

For cross-area comparisons, we additionally reanalyzed previously published datasets from VPS (Zhao et al., 2023) and MSTd (Morgan et al., 2008); no new recordings were performed in either area as part of the present study. Using the same analytical framework applied to the VIP dataset, we analyzed single-neuron gain modulation, population FI, offset-specific sensory weighting, and normalization model performance in all three areas. Comparing these complementary regions—visual-dominant MSTd, vestibular-dominant VPS, and multisensory VIP—allowed us to determine whether the conflict-dependent computations identified in VIP reflect a general property of the cortical visual-vestibular network or a distinctive computational strategy of VIP.

VIP neurons exhibit robust cross-modal gain modulation

Representative VIP neurons are shown in Figure 2*A*,*B*, including a multisensory neuron, a vestibular-only neuron, and a visual-only neuron. The multisensory neuron (Fig. 2*A*) exhibited significant tuning to both vestibular and visual stimuli ($p_{\text{vest}} = 1.75 \times 10^{-10}$; $p_{\text{vis}} = 1.08 \times 10^{-10}$, one-way ANOVA), with the same preferred heading of 315° for both modalities. The peak unimodal responses were 46.9 spk/s for vestibular stimulation and 46.4 spk/s for visual stimulation. When both cues were presented near the preferred heading, the combined response was strongly enhanced (62.9 spk/s). In contrast, pairing a preferred cue with a nonpreferred cue reduced the response. For example, when the visual cue was presented at its preferred heading of 315°, adding a nonpreferred vestibular cue at 180° reduced the combined response to 34.9 spk/s. Similarly, when the vestibular cue was presented at its preferred heading of 315°, adding a nonpreferred visual cue at 90° reduced the combined response to 31.9 spk/s. Thus, the same multisensory neuron exhibited enhancement when visual and vestibular cues were both near the preferred direction, but strong suppression when the two cues became substantially mismatched.

Cross-modal suppression was also observed in neurons that were significantly tuned to only one modality. The vestibular-only neuron in Figure 2*B* showed robust vestibular tuning but no significant visual tuning ($p_{\text{vest}} = 1.82 \times 10^{-10}$; $p_{\text{vis}} = 0.11$, one-way ANOVA). When vestibular heading was fixed at the neuron's preferred direction (0°) and visual heading was varied,

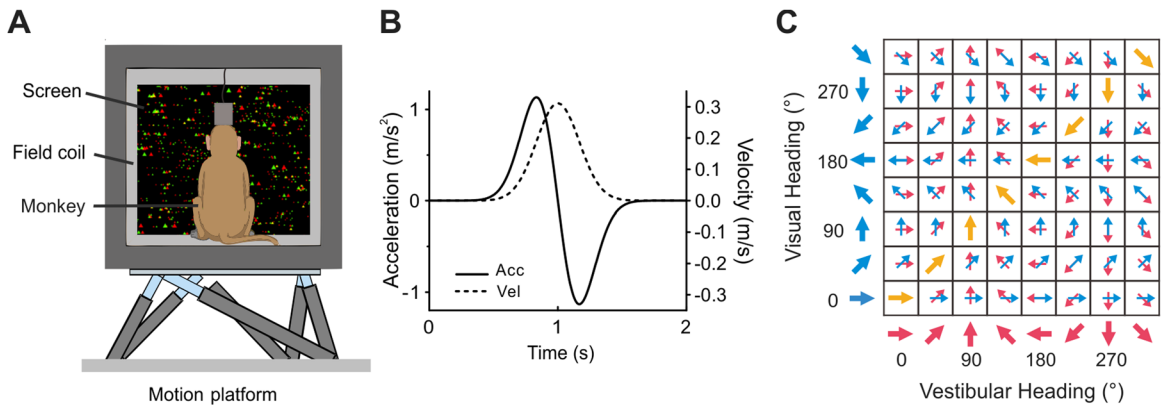


Figure 1. Experimental setup and visual-vestibular cue-conflict paradigm. ***A***, Experimental setup. Monkeys were seated on a motion platform and maintained fixation on a central target. Vestibular stimulation was delivered by physical translation of the platform, whereas visual self-motion was simulated by random-dot optic flow displayed on a monitor mounted on the platform. The monitor moved together with the animal and therefore remained fixed relative to the animal during platform translation. ***B***, Temporal profiles of the self-motion stimulus. Both vestibular platform motion and visual optic-flow stimuli followed the same Gaussian velocity profile. The stimulus duration was 2 s, with a total displacement of 0.13 m. The dashed curve indicates the velocity profile (peak velocity: 0.31 m/s), and the solid curve indicates the corresponding acceleration profile (peak acceleration magnitude: 1.13 m/s^2). Positive and negative phases represent acceleration and deceleration during the translational movement. ***C***, Visual-vestibular cue-conflict stimulus matrix. Vestibular headings (red arrows) and visual (blue arrows) headings were independently varied across eight directions in the horizontal plane, generating 64 combined visual-vestibular heading pairs. The eight congruent combinations, in which the visual and vestibular cues specified the same inferred self-motion heading, are indicated by orange arrows along the diagonal; the remaining 56 combinations, in which the two cues specified different inferred self-motion headings, were incongruent.

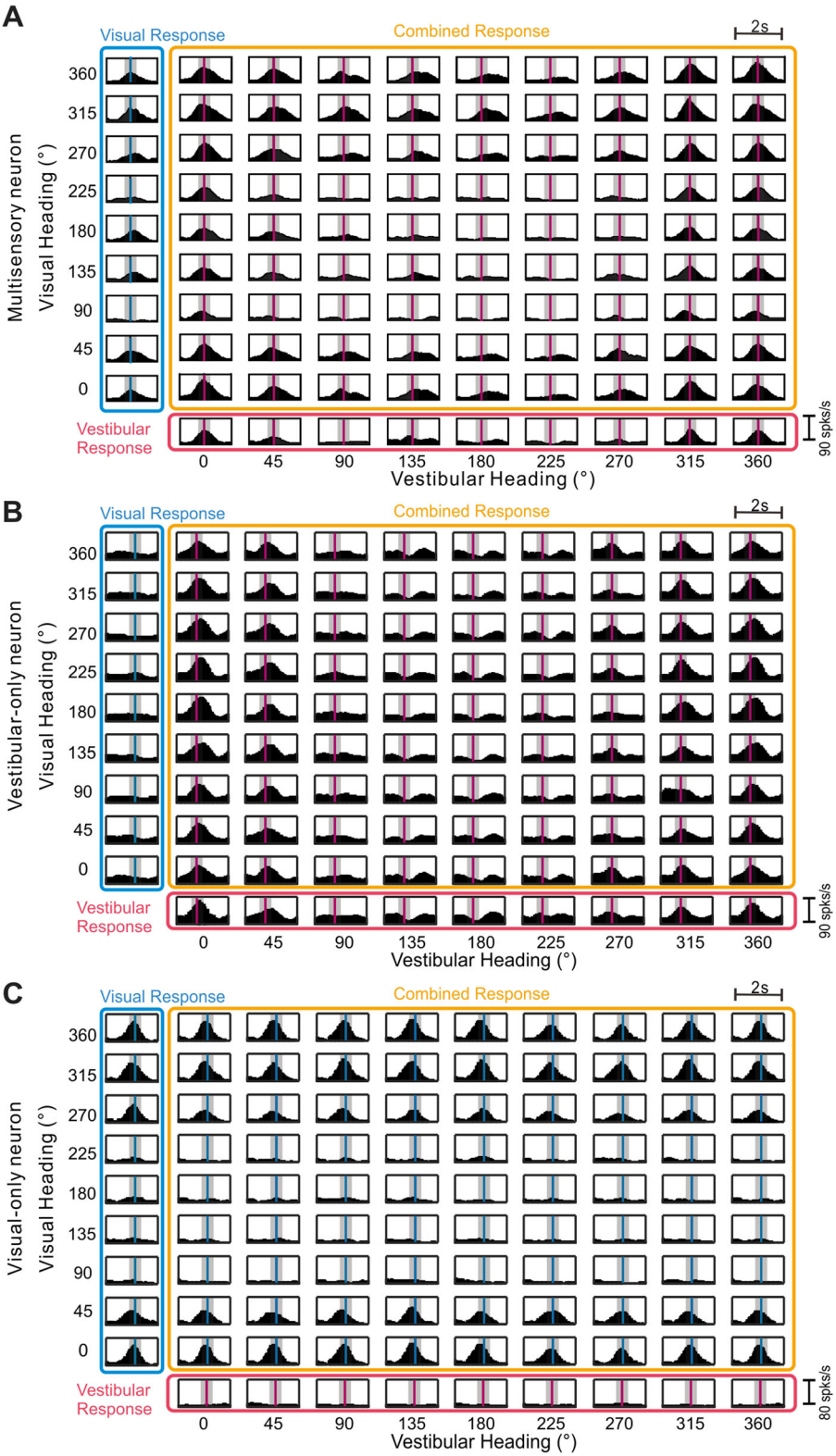


Figure 2. Representative response patterns of VIP neurons. ***A***, Responses of a representative multisensory VIP neuron exhibiting significant tuning to both vestibular and visual stimuli ($p_{vest} = 1.75 \times 10^{-10}$; $p_{vis} = 1.08 \times 10^{-10}$, one-way ANOVA). Bottom panels (magenta box) show vestibular-only responses, left panels (cyan box) show visual-only responses, and center panels (orange box) show responses to all combined cue conditions. Vertical dashed lines indicate the peak response times used for subsequent analyses ($t_{vestibular} = 1.1$ s; $t_{visual} = 1.3$ s). ***B***, Responses of a representative vestibular-only VIP neuron showing significant vestibular tuning but weak visual modulation ($p_{vest} = 1.82 \times 10^{-10}$; $p_{vis} = 0.11$, one-way ANOVA). Format as in ***A***. ***C***, Responses of a representative visual-only VIP neuron showing significant visual tuning but weak vestibular modulation ($p_{vest} = 0.25$; $p_{vis} = 6.48 \times 10^{-21}$, one-way ANOVA). Format as in ***A***.

the combined responses were consistently lower than the vestibular-only response at the preferred heading (vestibular-only: 83.4 spk/s; combined: 43.9–56.4 spk/s), indicating that visual input suppressed vestibular response gain even in the absence of significant visual tuning. Similarly, the visual-only neuron in Figure 2*C* showed strong visual tuning but no

significant vestibular tuning ($p_{vest} = 0.25$; $p_{vis} = 6.48 \times 10^{-21}$, one-way ANOVA). When visual heading was fixed at the preferred direction (0°) and vestibular heading was varied, most combined responses were lower than the visual-only response (visual-only: 64.8 spk/s; combined: 51.4–64.3 spk/s), demonstrating that vestibular input can suppress visual response gain even in neurons lacking significant vestibular tuning.

Together, these results show that VIP neurons exhibit robust cross-modal gain modulation even in unisensory neurons. Importantly, response modulation depends on visual-vestibular cue congruence: congruent cues enhance neuronal responses, whereas incongruent cues suppress them, revealing nonadditive multisensory interactions.

Multisensory interactions reflect gain modulation rather than tuning shifts

We next asked whether the cross-modal interactions observed above arose from changes in preferred heading or from modulation of response gain. These alternatives have distinct implications for self-motion coding. Large tuning shifts would indicate a reorganization of heading representations across sensory contexts, whereas gain modulation would allow sensory signals to be reweighted while maintaining a stable tuning structure.

For the example neuron in Figure 3*A*, visual input strongly modulated vestibular-referenced response amplitude. At the unimodal vestibular preferred heading (320.5°), responses varied substantially across visual heading conditions (58.35 to 111.72 spk/s). In contrast, vestibular preferred heading remained relatively stable, ranging from 309.4 to 338.3° across combined conditions, with a maximum shift of only 17.8°. A similar pattern was observed for visual-referenced tuning in the same neuron (Fig. 3*B*): vestibular input markedly altered response amplitude (69.33–111.72 spk/s at the unimodal visual preferred heading of 165.7°), while preferred visual heading changed only modestly (168.6–182.7°; maximum shift = 16.7°). Thus, cross-modal inputs strongly influenced response amplitude but produced only limited changes in tuning preference.

To quantify this effect across the population, we computed, for each neuron, the maximum absolute deviation of preferred heading across combined conditions relative to its unimodal preferred heading. Among the 70 VIP neurons with significant vestibular tuning, the median maximum vestibular preferred-heading shift was 28.8°, and 60.0% of neurons showed a maximum shift of <45° (42/70 neurons; Fig. 3*C*). Similarly, among the 72 neurons with significant visual tuning, the median maximum visual preferred-heading shift was 30.2°, and 62.5% of neurons showed a maximum shift of <45° (45/72 neurons; Fig. 3*D*).

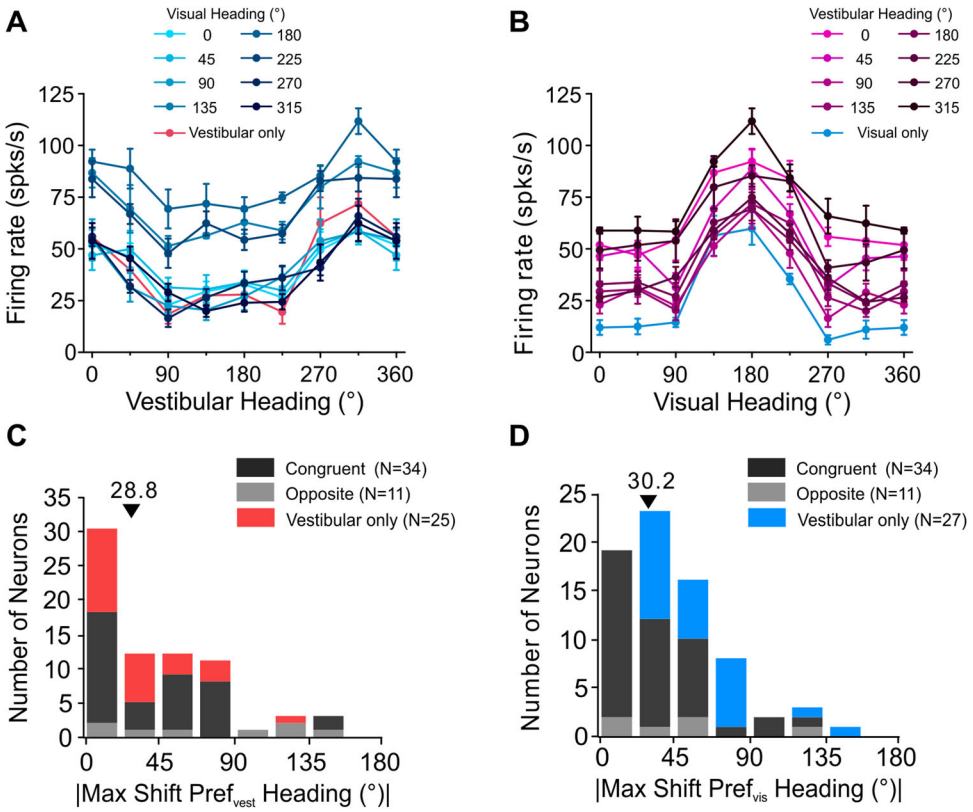


Figure 3. Visual-vestibular interactions primarily modulate response gain while preserving heading tuning. ***A***, Vestibular tuning curves of an example VIP neuron measured under different visual heading contexts. Curves ranging from light to dark cyan show vestibular tuning during combined stimulation for different fixed visual headings, whereas the red curve indicates the vestibular-only response. Although response magnitude varied across visual contexts, vestibular preferred heading remained relatively stable. Error bars denote SE. ***B***, Visual tuning curves of the same neuron measured under different vestibular heading contexts. Curves ranging from light to dark magenta show visual tuning during combined stimulation for different fixed vestibular headings, whereas the blue curve indicates the visual-only response. Response gain varied substantially across vestibular contexts, whereas preferred visual heading changed little. Error bars denote SE. ***C***, Distribution of the maximum shift in vestibular preferred heading induced by visual stimuli. Black bars denote multisensory neurons with congruent visual and vestibular preferences (preferred-heading difference < 90°, $N = 34$), gray bars denote multisensory neurons with opposite visual and vestibular preferences (preferred-heading difference > 90°, $N = 11$), and red bars denote vestibular-only neurons ($N = 25$). The inverted triangle indicates the population median (28.8°). ***D***, Distribution of the maximum shift in visual preferred heading induced by vestibular stimuli. Black bars denote multisensory neurons with congruent visual and vestibular preferences ($N = 34$), gray bars denote multisensory neurons with opposite visual and vestibular preferences ($N = 11$), and blue bars denote visual-only neurons ($N = 27$). The inverted triangle indicates the population median (30.2°).

Together, these results indicate that visual-vestibular interactions in VIP primarily modulate neuronal gain rather than reorganizing heading selectivity. The enhancement and suppression observed under combined stimulation therefore reflect adaptive scaling of response strength while preserving a largely stable representation of heading direction.

Cross-modal gain modulation differs across cortical areas

Having established that combined stimulation primarily affected response gain, we next examined how preference-referenced cross-modal modulation varied with the preferred-heading deviation of the non-reference cue. Figure 4, *A* and *B*, illustrates this analysis for an example multisensory VIP neuron. In the

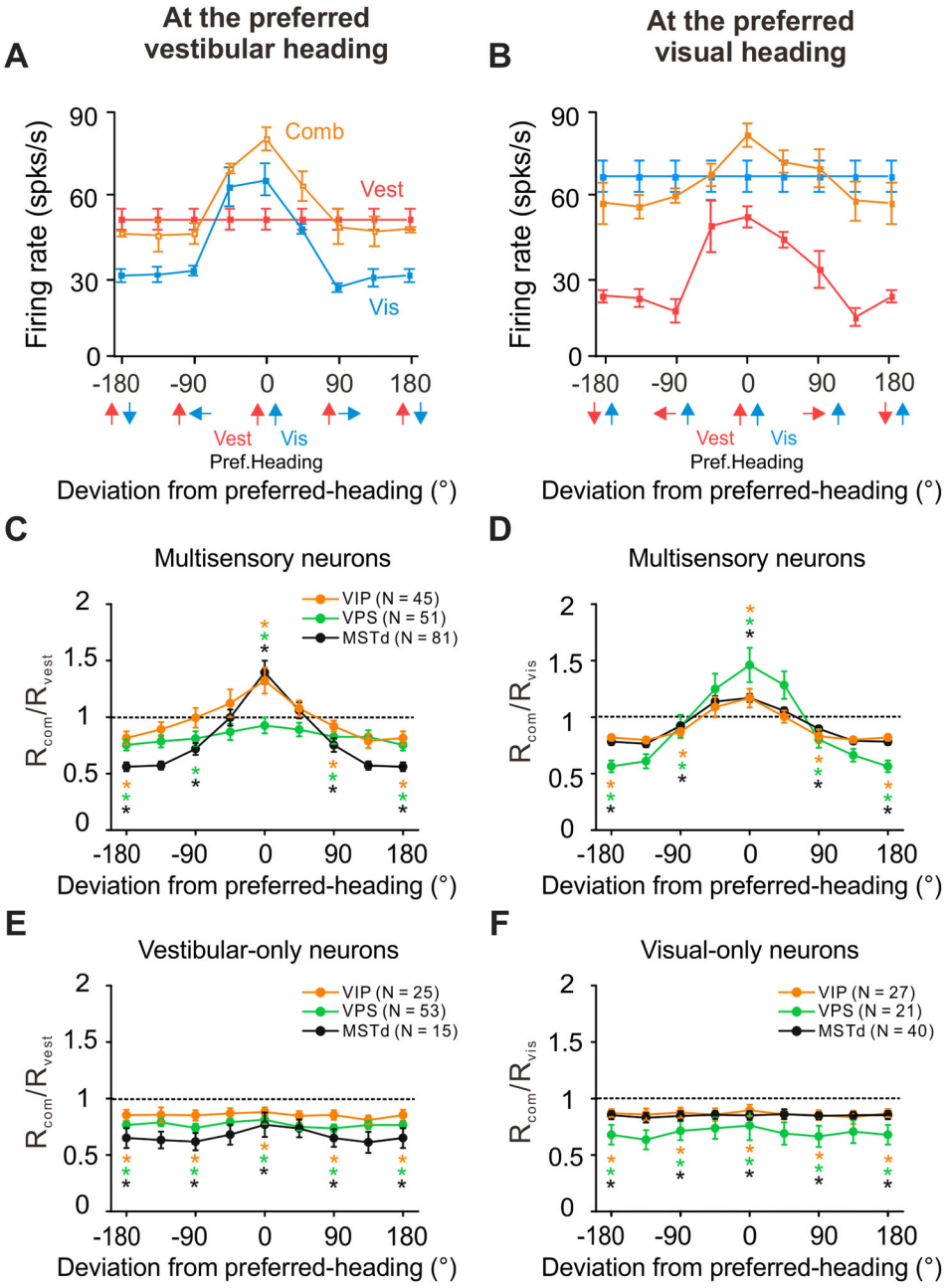


Figure 4. Conflict-dependent modulation of response gain in VIP, VPS, and MSTd. ***A***, Responses of an example VIP neuron in the vestibular-reference condition. Vestibular heading was fixed at the neuron's vestibular preferred heading, and visual heading was varied relative to the neuron's visual preferred heading. The *x*-axis indicates the offset of the visual heading from its preferred heading; 0° indicates that both vestibular and visual cues were presented at their respective preferred headings. Orange, magenta, and cyan curves indicate combined, vestibular-only, and visual-only responses, respectively. Error bars denote SE. ***B***, Responses of the same neuron in the visual-reference condition. Visual heading was fixed at the neuron's visual preferred heading, and vestibular heading was varied relative to the neuron's vestibular preferred heading. The *x*-axis indicates the offset of the vestibular heading from its preferred heading. Formats are the same as in ***A***. ***C***, Mean ratio of combined response to vestibular response, R_{com}/R_{vest}, plotted as a function of visual preferred-heading deviation for multisensory neurons. VIP ($N = 45$), VPS ($N = 51$), and MSTd ($N = 81$) are shown in orange, green, and black, respectively. Values greater than 1 indicate multisensory enhancement relative to the vestibular-only response, whereas values less than 1 indicate suppression. ***D***, Mean ratio of combined response to visual response, R_{com}/R_{vis}, plotted as a function of vestibular preferred-heading deviation for multisensory neurons. Values greater than 1 indicate enhancement relative to the visual-only response, whereas values less than 1 indicate suppression. ***E***, ***F***, Population summary for vestibular-only neurons (***E***) and visual-only neurons (***F***). Response ratios are plotted as a function of the preferred-heading deviation of the nonreference stimulus. Error bars denote SE. Asterisks indicate significant deviation from unity ($*p < 0.05$). Across all three areas, response ratios generally decreased as the nonreference stimulus deviated further from its preferred heading.

vestibular-referenced analysis, the vestibular heading was fixed at the neuron's vestibular preferred heading, whereas the visual heading was varied relative to its own preferred heading (Fig. 4*A*). Conversely, in the visual-referenced analysis, the visual heading was fixed at its preferred heading, whereas the vestibular heading was varied relative to its own preferred heading (Fig. 4*B*). As expected from directional tuning, combined responses were greatest when the nonreference cue was presented at its preferred heading and decreased as the magnitude of its preferred-heading deviation increased. We therefore quantified cross-modal gain relative to the corresponding unimodal response and compared the resulting gain profiles across VIP, VPS, and MSTd.

We quantified this effect at the population level by defining multisensory gain as the ratio between the combined response and the corresponding unimodal response. In the vestibular-referenced analysis, gain was quantified as R_{com}/R_{vest}; in the visual-referenced analysis, gain was quantified as R_{com}/R_{vis}. Values greater than 1 indicated enhancement relative to the unimodal response, whereas values below 1 indicated suppression.

In the vestibular-reference analysis (Fig. 4*C*, Table 1), R_{com}/R_{vest} in VIP was significantly >1 at 0° visual preferred-heading deviation ($p < 0.01$, Wilcoxon signed-rank test vs 1). The ratio decreased as the magnitude of the visual preferred-heading deviation increased. At the 180° visual preferred-heading deviation, response ratios differed significantly across cortical areas ($F_{(2,49)} = 50.40$, $p = 1.12 \times 10^{-12}$, Welch's ANOVA). VIP exhibited higher R_{com}/R_{vest} values than MSTd (Games–Howell post hoc test: $p = 2.70 \times 10^{-7}$) but did not differ significantly from VPS ($p = 0.408$). Thus, vestibular-referenced gain was less strongly suppressed in VIP than in MSTd when the visual cue was presented opposite its preferred heading.

In the visual-referenced analysis (Fig. 4*D*), R_{com}/R_{vis} similarly decreased as the magnitude of the vestibular preferred-heading deviation increased and fell below unity under large offsets ($p < 0.01$, Wilcoxon signed-rank test vs 1). At the 180° vestibular preferred-heading deviation, response ratios differed significantly across areas ($F_{(2,37)} = 3.69$, $p = 0.034$, Welch's ANOVA). In this condition, VIP maintained a significantly higher R_{com}/R_{vis} than VPS (Games–Howell post hoc test: $p = 0.023$), whereas it did not differ from MSTd (Games–Howell post hoc test: $p = 0.62$). Thus, visual-referenced gain was less strongly suppressed in VIP than in VPS when the vestibular cue was presented opposite its preferred heading.

We further asked whether this pattern extends to neurons dominated by a single modality. Separate analyses of vestibular-only and visual-only neurons revealed that the nondominant cue generally exerted suppressive effects, with response ratios significantly below unity across areas ($p < 0.01$, Wilcoxon signed-rank test vs 1; Fig. 4*E*,*F*; Table 1). For vestibular-only neurons, suppression at the 180° visual preferred-heading deviation was weaker in VIP than in MSTd ($p = 0.045$, Games–Howell post hoc test; Fig. 4*E*). For visual-only neurons, suppression at the 180° vestibular preferred-heading deviation was weaker in VIP than in VPS ($p = 0.036$, Games–Howell post hoc test; Fig. 4*F*). Thus, area-specific cross-modal gain modulation was also present in neurons lacking significant tuning to one of the two modalities.

Together, these results show that cross-modal gain depended on the position of the nonreference cue relative to its own preferred heading and differed across cortical areas. At the 180° preferred-heading deviation, VIP retained more vestibular-referenced gain than MSTd and more visual-referenced gain than VPS, consistent with a more balanced cross-modal gain profile in VIP.

Cue conflict reshapes unimodal drive-gain relationship

We next examined whether preference-referenced cross-modal gain covaried with relative unimodal response strength. In the vestibular-referenced analysis, the combined-to-vestibular response ratio (R_{com}/R_{vest}) was plotted against the visual-to-vestibular response ratio (R_{vis}/R_{vest}) at selected visual preferred-heading deviations (Fig. 5, Table 2). At 0° visual preferred-heading deviation, neurons with stronger visual drive relative to vestibular responses tended to exhibit larger multisensory gain in VIP ($r = 0.52$, $p = 3.62 \times 10^{-6}$, $N = 70$, Pearson's correlation; Fig. 5*A*). This relationship was even stronger in MSTd ($r = 0.86$, $p < 0.001$, $N = 96$, Pearson's correlation; Fig. 5*B*), whereas VPS showed no significant dependence ($r = 0.05$, $p = 0.61$, $N = 104$, Pearsons correlation; Fig. 5*C*). At 90 and 180° visual preferred-heading deviation, the positive correlations in VIP and MSTd became progressively weaker but remained significant, whereas VPS showed only a weak relationship at 90° and no significant relationship at 180°. Thus, visual response strength remained associated with vestibular-referenced gain across preferred-heading deviations in VIP, although this association was consistently stronger in the visual-dominant MSTd and weaker in the vestibular-dominant VPS.

In the visual-referenced analysis, R_{com}/R_{vis} was plotted against R_{vest}/R_{vis} (Fig. 6, Table 2). At 0° vestibular preferred-heading

Table 1. Cross-modal gain modulation across preferred-heading deviations in VIP, VPS, and MSTd

				Preferred-heading deviation				
Neuron class	Response ratio	Region	*N*	0°	45°	90°	135°	180°
Multisensory neurons	R_{com}/R_{vest}	VIP	45	1.32 ± 0.11	1.07 ± 0.07	0.91 ± 0.04	0.78 ± 0.06	0.86 ± 0.06
		VPS	51	0.92 ± 0.06	0.88 ± 0.06	0.82 ± 0.06	0.82 ± 0.05	0.82 ± 0.16
		MSTd	81	1.39 ± 0.10	1.05 ± 0.07	0.75 ± 0.06	0.57 ± 0.33	0.67 ± 0.08
	R_{com}/R_{vis}	VIP	45	1.16 ± 0.08	1.00 ± 0.05	0.83 ± 0.06	0.79 ± 0.02	0.88 ± 0.07
		VPS	51	1.46 ± 0.15	1.28 ± 0.12	0.80 ± 0.07	0.66 ± 0.05	0.53 ± 0.03
		MSTd	81	1.16 ± 0.03	1.05 ± 0.03	0.89 ± 0.02	0.79 ± 0.02	0.83 ± 0.17
Vest-only neurons	R_{com}/R_{vest}	VIP	25	0.88 ± 0.04	0.84 ± 0.04	0.85 ± 0.04	0.81 ± 0.03	0.85 ± 0.04
		VPS	53	0.81 ± 0.03	0.74 ± 0.03	0.73 ± 0.03	0.76 ± 0.04	0.76 ± 0.03
		MSTd	15	0.76 ± 0.10	0.73 ± 0.07	0.65 ± 0.07	0.61 ± 0.09	0.65 ± 0.08
Vis-only neurons	R_{com}/R_{vis}	VIP	27	0.89 ± 0.04	0.85 ± 0.04	0.85 ± 0.04	0.83 ± 0.05	0.87 ± 0.03
		VPS	21	0.76 ± 0.12	0.69 ± 0.09	0.66 ± 0.09	0.70 ± 0.10	0.68 ± 0.09
		MSTd	40	0.85 ± 0.03	0.86 ± 0.04	0.84 ± 0.03	0.85 ± 0.02	0.85 ± 0.03

Values are mean ± SEM. For R_{com}/R_{vest}, preferred-heading deviation refers to the deviation of the visual stimulus from the neuron's visual preferred heading while the vestibular stimulus was fixed at the vestibular preferred heading. For R_{com}/R_{vis}, it refers to the deviation of the vestibular stimulus from the neuron's vestibular preferred heading while the visual stimulus was fixed at the visual preferred heading.

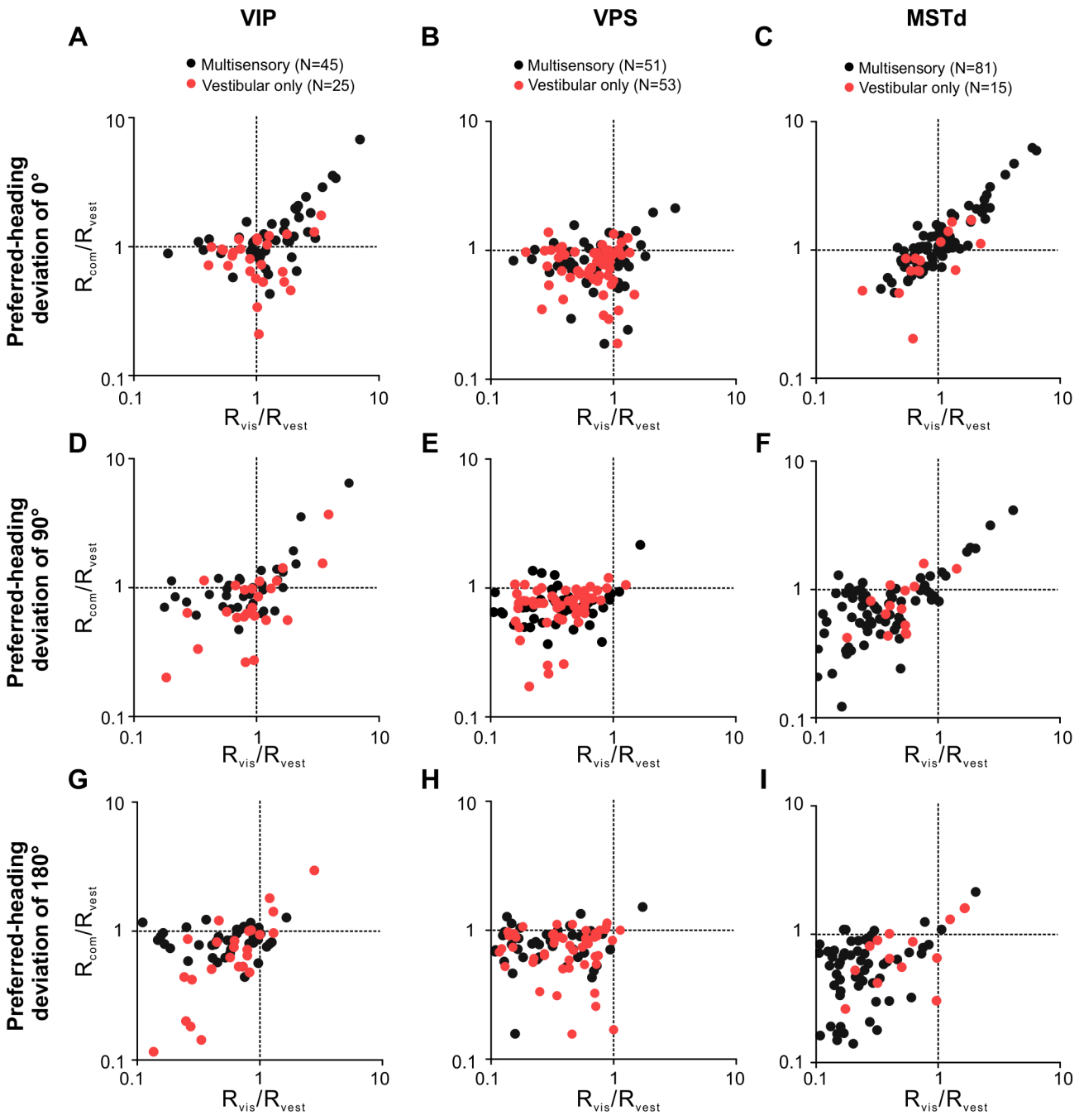


Figure 5. Visual modulation of vestibular-referenced gain as a function of visual-vestibular preferred-heading difference. ***A–C***, Relationship between the combined-to-vestibular response ratio, R_{com}/R_{vest} (*y*-axis), and the visual-to-vestibular response ratio, R_{vis}/R_{vest} (*x*-axis), at a visual preferred-heading deviation of 0° in VIP (***A***), VPS (***B***), and MSTd (***C***). Black symbols indicate multisensory neurons, and red symbols indicate vestibular-only neurons. Included neurons were VIP multisensory ($N = 45$) and vestibular-only ($N = 25$), VPS multisensory ($N = 51$) and vestibular-only ($N = 53$), and MSTd multisensory ($N = 81$) and vestibular-only ($N = 15$). Dashed horizontal and vertical lines mark ratio values of 1. ***D–F***, Same analysis at a visual preferred-heading deviation of 90° in VIP (***D***), VPS (***E***), and MSTd (***F***). ***G–I***, Same analysis at a visual preferred-heading deviation of 180° in VIP (***G***), VPS (***H***), and MSTd (***I***). Values above the horizontal unity line indicate enhancement relative to the vestibular-only response, whereas values below unity indicate suppression. Positive correlations were strongest in MSTd, intermediate in VIP, and weak in VPS and generally decreased as visual preferred-heading deviation increased.

deviation, positive correlations were observed in all three areas, with the strongest relationship in VPS, followed by VIP and MSTd. VPS maintained strong positive correlations at both 90 and 180° vestibular preferred-heading deviation, whereas the relationship in MSTd weakened with increasing vestibular preferred-heading deviation and was no longer significant at 180°. In VIP, the correlation remained significant at both 90 and 180°, indicating that relative vestibular response strength continued to covary with visual-referenced gain across the tested preferred-heading deviations.

The reciprocal patterns in MSTd and VPS were consistent with their established visual and vestibular dominance, respectively, and therefore provided modality-dominant reference profiles. In contrast, VIP showed significant relationships in both reference analyses and at all tested preferred-heading deviations. Thus, relative visual response strength predicted

Table 2. Correlations between relative unimodal drive and cross-modal gain across preferred-heading deviations

Reference analysis	Correlated response ratios	Preferred-heading deviation	VIP	VPS	MSTd
Vestibular	R_{com}/R_{vest} versus R_{vis}/R_{vest}	0°	$r = 0.52$, $p < 0.001$	$r = 0.05$, $p = 0.61$	$r = 0.86$, $p < 0.001$
		90°	$r = 0.46$, $p < 0.001$	$r = 0.22$, $p < 0.05$	$r = 0.73$, $p < 0.001$
		180°	$r = 0.31$, $p < 0.05$	$r = 0.25$, $p = 0.07$	$r = 0.46$, $p < 0.001$
Visual	R_{com}/R_{vis} versus R_{vest}/R_{vis}	0°	$r = 0.69$, $p < 0.001$	$r = 0.78$, $p < 0.001$	$r = 0.50$, $p < 0.001$
		90°	$r = 0.49$, $p < 0.05$	$r = 0.72$, $p < 0.001$	$r = 0.23$, $p < 0.05$
		180°	$r = 0.54$, $p < 0.01$	$r = 0.70$, $p < 0.001$	$r = 0.19$, $p = 0.08$

Values are Pearson's correlation coefficients. In the vestibular-reference analysis, preferred-heading deviation refers to the deviation of the visual stimulus from the neuron's visual preferred heading. In the visual-reference analysis, it refers to the deviation of the vestibular stimulus from the neuron's vestibular preferred heading.

vestibular-referenced gain, while relative vestibular response strength also predicted visual-referenced gain in VIP. This bidirectional dependence distinguishes VIP from the modality-dominant profiles of MSTd and VPS and is consistent with a more balanced contribution of visual and vestibular signals to combined-response gain.

Population coding of heading remains robust in VIP

The preceding analyses showed that visual-vestibular conflict strongly modulated neuronal gain in VIP. We next asked whether these changes affect the fidelity of population heading coding. To address this question, we computed population tuning curves and FI under visual-only, vestibular-only, and combined conditions grouped according to the physical visual-vestibular heading offsets ($\Delta = 0°$, 90°, or 180°; Fig. 7). The averaged population tuning curves provide descriptive summaries of the population responses, whereas FI was calculated from the condition-specific tuning curve of each neuron before being summed across the population.

In VIP, population tuning curves were broadly similar across stimulus conditions (Fig. 7*A*). Vestibular-only stimulation produced relatively weak responses, whereas visual-only and all three combined conditions generated stronger and largely overlapping tuning profiles. Although cue conflict strongly affected single-neuron responses, its influence on the population representation was much less pronounced, with only modest differences among the combined conditions. This observation suggests that population-level heading representations remained relatively stable despite substantial conflict-dependent gain modulation at the single-neuron level.

To quantify coding precision, we computed FI as a function of heading directions (Fig. 7*B*) and averaged FI across headings for each condition (Fig. 7*C*). FI differed significantly across conditions (repeated-measures ANOVA, $F_{(4,176)} = 10.66$, $p = 9.29 \times 10^{-8}$; Fig. 7*C*). Congruent multisensory stimulation ($\Delta = 0°$) significantly increased FI (mean ± SEM: 18.32 ± 0.48) relative to both visual-only (14.92 ± 0.42, $p = 4.52 \times 10^{-3}$, Holm-corrected paired *t* test) and vestibular-only conditions (13.28 ± 0.43, $p = 3.22 \times 10^{-5}$, Holm-corrected paired *t* test). Importantly, FI under maximal conflict ($\Delta = 180°$: 18.09 ± 0.46) remained significantly higher than either unimodal condition ($p = 8.47 \times 10^{-3}$ for visual-only, $p = 1.35 \times 10^{-3}$ for vestibular-only, Holm-corrected paired *t* test) and did not differ significantly from the congruent condition ($p = 0.38$, Holm-corrected paired *t* test). Thus, despite strong cue-conflict effects on neuronal gain, population heading information remained robust in VIP.

VPS exhibited a distinct vestibular-dominant response profile (Fig. 7*D*). Vestibular-only stimulation produced strong heading tuning, whereas visual-only responses were substantially weaker. Notably, population tuning curves obtained under all three combined conditions closely resemble the vestibular-only profile and varied little with visual-vestibular offset. Consistent with this expected vestibular dominance, FI was highest under vestibular-only stimulation and significantly lower in all combined conditions (Fig. 6*E*,*F*; all $p < 10^{-7}$, Holm-corrected paired *t* tests). Thus, visual inputs failed to improve coding precision beyond vestibular performance, suggesting a largely vestibular-dominant coding strategy in VPS.

MSTd served as the corresponding visual-dominant comparison area (Fig. 7*G*). As expected, visual-only stimulation generated relatively narrow tuning curves with high peak responses near straight-ahead headings, whereas vestibular-only stimulation produced broader but lower-amplitude tuning. Under combined stimulation, population responses retained the strong peak activity associated with visual input while exhibiting broader tuning than the visual-only condition. Congruent stimulation ($\Delta = 0°$) produced the strongest population responses, whereas increasing cue conflict progressively reduced response amplitude. These effects were reflected in the FI analysis (Fig. 7*H*,*I*). Congruent stimulation significantly increased FI (28.39 ± 0.88) relative to visual-only (24.88 ± 0.87) stimulation ($p = 0.003$, Holm-corrected paired *t* test). In contrast, both $\Delta = 90°$ (21.16 ± 0.54) and $\Delta = 180°$ (22.50 ± 0.72) conditions significantly reduced FI relative to the congruent condition (both $p < 0.001$, Holm-corrected paired *t* test). These results were consistent with the established visual specialization of MSTd and provided a visual-dominant benchmark for comparison with VIP.

Together, these findings reveal distinct population coding profiles across cortical self-motion areas under the same fixation task and analytical framework. VIP preserved robust heading information across congruent and incongruent conditions, whereas VPS maintained a vestibular-dominant coding regime and MSTd exhibited a visual-dominant yet integration-enhanced coding strategy.

Dynamic sensory reweighting distinguishes VIP from modality-dominant cortical areas

The population FI analysis suggested that VIP differs fundamentally from VPS and MSTd in how sensory conflict influences multisensory coding. We therefore asked whether these differences arise from distinct sensory-weighting strategies. To quantify the relative contributions of visual and vestibular inputs, we fitted an offset-specific linear weighting model, in which combined response were approximated as weighted sums of unimodal visual and vestibular responses, with weights estimated separately for each cue-offset condition (see Materials and Methods for further details).

Figure 8, *A* and *B*, illustrates this analysis for an example VIP neuron. The raw response matrix (Fig. 8*A*) revealed a localized response hotspot centered on a specific combination of visual and vestibular headings. Responses were strongest when the two cues were approximately congruent near their preferred directions and gradually weakened as cue discrepancy increased. Importantly, the response pattern was neither organized as horizontal bands, which would indicate purely visual coding, nor as

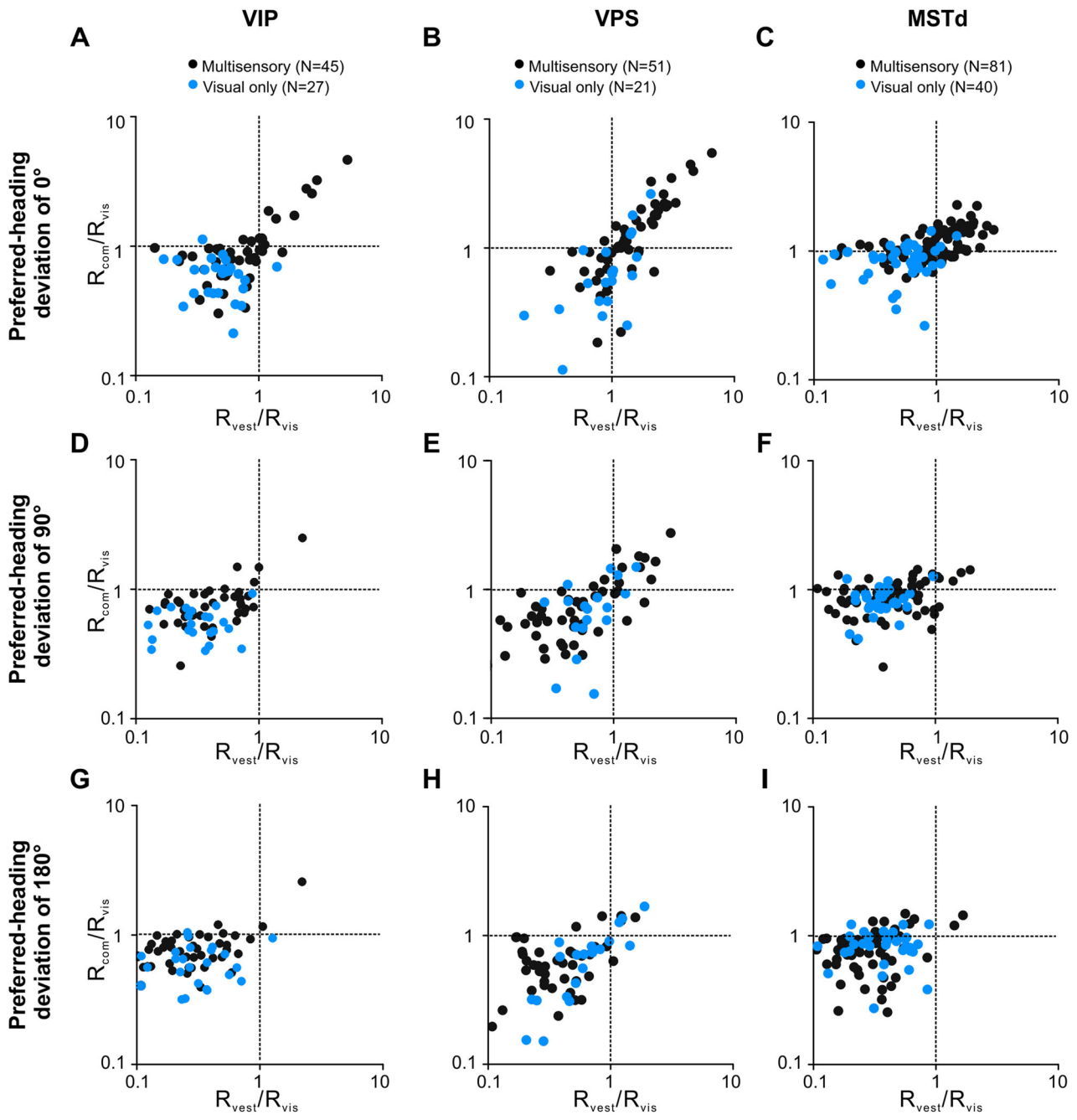


Figure 6. Vestibular modulation of vestibular-referenced gain as a function of visual-vestibular preferred-heading difference. ***A–C***, Relationship between the combined-to-visual response ratio, R_{com}/R_{vis} (*y*-axis), and the vestibular-to-visual response ratio, R_{vest}/R_{vis} (*x*-axis), at a vestibular preferred-heading deviation of 0° in VIP (***A***), VPS (***B***), and MSTd (***C***). Black symbols indicate multisensory neurons, and blue symbols indicate visual-only neurons. Included neurons were VIP multisensory ($N = 45$) and visual-only ($N = 27$), VPS multisensory ($N = 51$) and visual-only ($N = 21$), and MSTd multisensory ($N = 81$) and visual-only ($N = 40$). Dashed horizontal and vertical lines mark ratio values of 1. ***D–F***, Same analysis at a vestibular preferred-heading deviation of 90° in VIP (***D***), VPS (***E***), and MSTd (***F***). ***G–I***, Same analysis at a vestibular preferred-heading deviation of 180° in VIP (***G***), VPS (***H***), and MSTd (***I***). Values above the horizontal unity line indicate enhancement relative to the visual-only response, whereas values below unity indicate suppression. VPS showed strong positive correlations across all three deviations, whereas the relationship weakened with increasing deviation in MSTd. VIP showed significant positive correlations at all three deviations.

vertical bands, which would indicate purely vestibular coding. Instead, firing rates depended jointly on both visual and vestibular headings, suggesting substantial contributions from both modalities. The oblique bands corresponding to constant visual-vestibular heading offsets further suggested that neuronal responses were strongly modulated by the degree of visual-vestibular conflict.

To quantify these contributions, we estimated visual and vestibular weights separately for each offset condition (Fig. 8*B*). At small visual-vestibular heading offsets (0–45° offset), visual weights exceeded vestibular weights ($w_{vis} = 0.48$ vs $w_{vest} = 0.34$), indicating stronger visual influence when the two cues were spatially consistent. As cue offset increased, however, visual weights progressively decreased whereas vestibular weights increased. Near the 90° offset condition, the two weights converged ($w_{vis} = 0.38$, $w_{vest} = 0.41$), indicating approximately balanced contributions from the two modalities. At large visual-vestibular heading offsets (135–180° offset), vestibular weights became

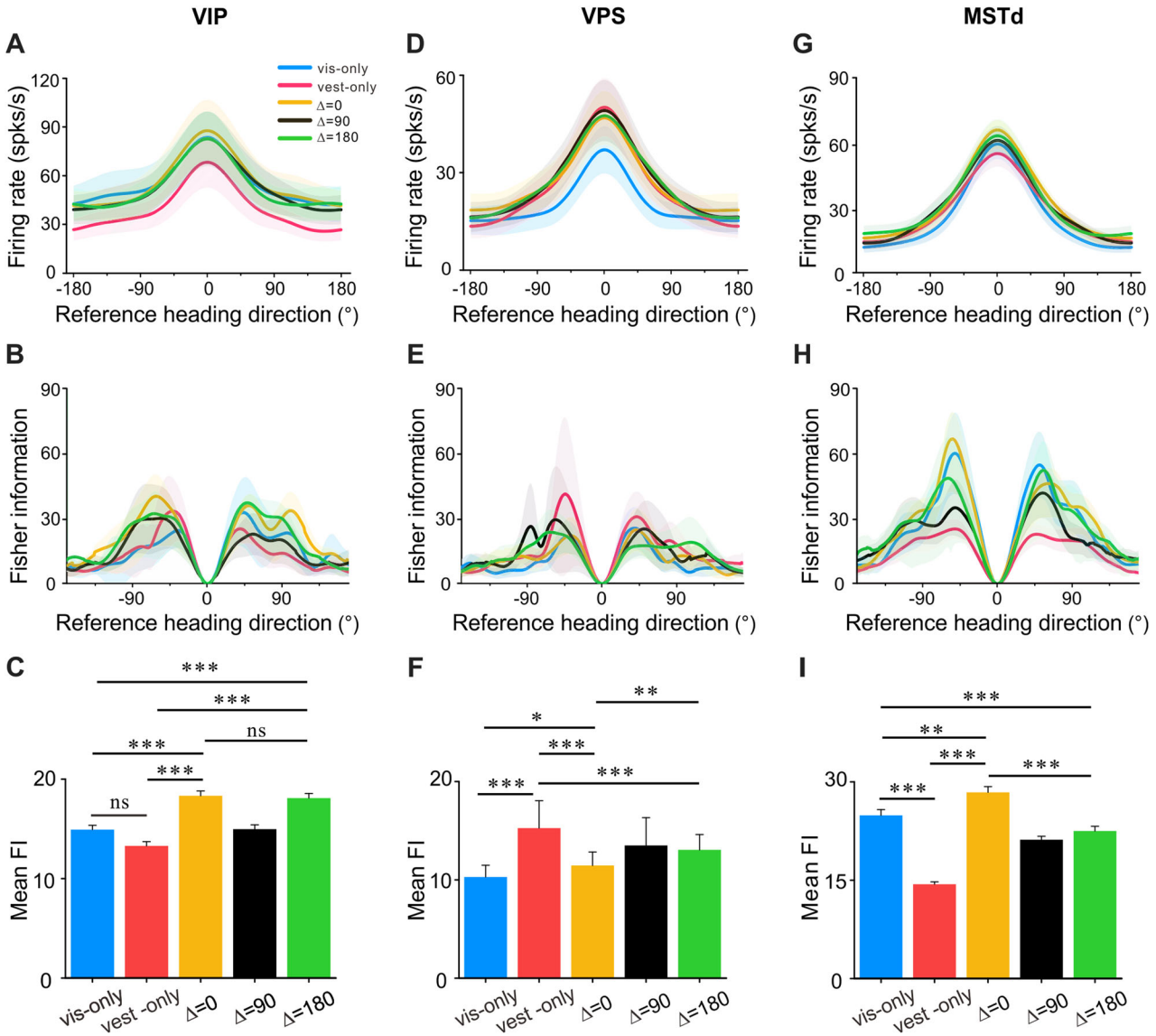


Figure 7. Population FI remains stable despite visual-vestibular conflict. ***A–C***, Population tuning and FI analyses for VIP. ***A***, Mean population tuning curves under visual-only, vestibular-only, and combined conditions with visual-vestibular offsets of 0, 90, and 180°. Tuning curves were circularly aligned such that, for each condition, the heading evoking the maximal population response was assigned to 0° on the *x*-axis. Thus, 0° denotes the condition-specific peak response heading rather than the physical stimulus direction. Shaded regions indicate SE. ***B***, Population FI as a function of heading for the same five conditions. FI was calculated from the condition-specific tuning curve of each neuron and then summed across neurons. ***C***, Mean FI averaged across headings for each condition. VIP maintained relatively high FI under both 0 and 180° combined conditions. ***D–F***, Corresponding population tuning curves (***D***), heading-resolved FI profiles (***E***), and mean FI values (***F***) for VPS. Vestibular-only stimulation produced the highest FI in VPS. ***G–I***, Corresponding population tuning curves (***G***), heading-resolved FI profiles (***H***), and mean FI values (***I***) for MSTd. FI was highest under the 0° combined condition and decreased under the 90 and 180° conditions. Statistical comparisons were performed using repeated-measures ANOVA followed by Holm-corrected post hoc paired comparisons (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$; ns, not significant.).

dominant ($w_{vest} = 0.46$ vs $w_{vis} = 0.32$). Thus, this neuron illustrated a gradual shift from stronger visual weighting at small cue offsets to stronger vestibular weighting at large offsets. We next examined whether this pattern was also present at the population level.

To characterize the overall weighting profile of each neuron, visual and vestibular weights were averaged across the five offset conditions (0, 45, 90, 135, and 180°). Across well-fit VIP neurons ($R^2 > 0.6$, $N = 25$), these offset-averaged visual and vestibular weights were broadly distributed around the diagonal (Fig. 8*C*), indicating balanced but heterogeneous multisensory contributions. Some neurons showed stronger visual weighting, whereas others were more vestibular-weighted, spanning a continuum rather than clustering into distinct modality-dominant groups. Consistent with this heterogeneity, mean visual and vestibular weights did not differ significantly when averaged across all offsets (w_{vest}: 0.58 ± 0.22; w_{vis}: 0.55 ± 0.28; $N = 25$; Wilcoxon signed-rank test, $p = 0.82$). However, weighting changed systematically with cue offset (Fig. 8*D*). Visual weights were significantly larger than vestibular weights under cue congruency (0° offset; $p < 0.05$, *t* test), whereas vestibular weights became significantly larger than visual weights under maximal conflict (180° offset; $p < 0.05$, *t* test; Fig. 8*D*). Thus, VIP did not exhibit a fixed sensory hierarchy. Instead, the relative visual and vestibular weights varied systematically with cue offset.

In contrast, VPS exhibited a stable vestibular-dominant weighting profile (Fig. 8*E*). Vestibular weights remained significantly larger than visual weights across all cue offsets (main

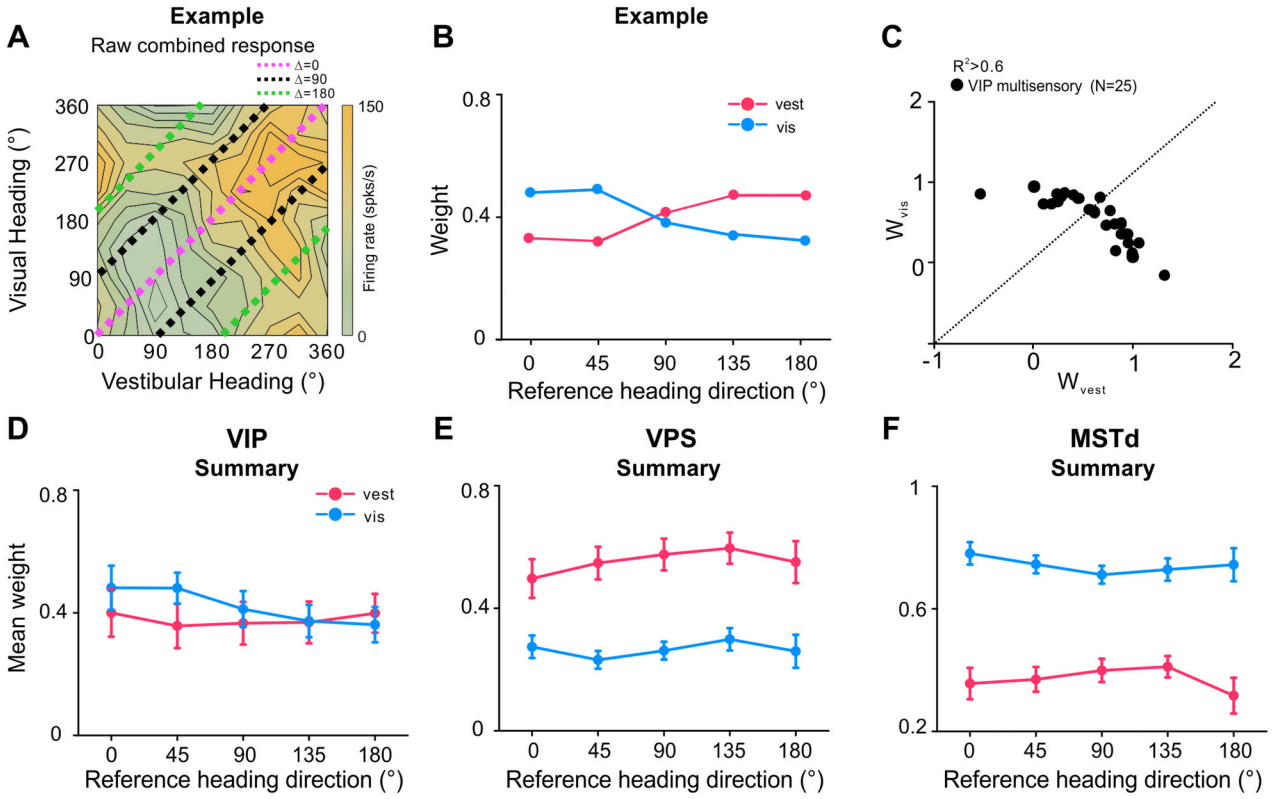


Figure 8. Example fitting data and population summary of linear weighting models. ***A***, Contour map of the raw combined response of an example VIP neuron. Firing rate is plotted as a function of visual heading and vestibular heading. Dashed lines indicate three representative visual-vestibular offset conditions: $\Delta = 0°$ is shown in red, $\Delta = 90°$ in black, and $\Delta = 180°$ in blue. ***B***, Visual and vestibular coefficients estimated separately for the example neuron at each of the five visual-vestibular heading offsets. The red curve indicates the fitted visual coefficient, and the blue curve indicates the fitted vestibular coefficient. For this neuron, the visual coefficient was larger at small offsets and decreased as cue offset increased, whereas the vestibular coefficient increased. ***C***, Scatterplot of visual weight against vestibular weight in VIP multisensory neurons. For each neuron, visual and vestibular weights were first estimated separately for the five offset conditions (0, 45, 90, 135, and 180°) and then averaged across offsets. Each symbol represents the offset-averaged visual and vestibular weights of one neuron. Only neurons for which the linear model achieved $R^2 > 0.6$ at each of the five offset conditions were included in this analysis ($N = 25$). This criterion was used because visual and vestibular weights cannot be reliably interpreted when the corresponding model fit is poor. Therefore, this analysis characterizes the subset of neurons whose responses were consistently captured by the linear weighting model, rather than the entire recorded population. ***D–F***, Population summary of visual and vestibular weights as a function of visual-vestibular heading offset in VIP (***D***), VPS (***E***), and MSTd (***F***). The red line indicates visual weight, and the blue line indicates vestibular weight. Error bars denote SE. This analysis included VIP ($N = 25$), VPS ($N = 46$), and MSTd ($N = 63$) neurons.

effect of modality: $F_{(1,45)} = 19.17$, $p = 7.05 \times 10^{-5}$, two-way repeated-measures ANOVA). Neither the main effect of offset ($F_{(2,131)} = 0.49$, $p = 0.614$, Greenhouse–Geisser correction) nor the modality-by-offset interaction ($F_{(2,109)} = 0.21$, $p = 0.853$, Greenhouse–Geisser correction) was significant. Thus, sensory weighting in VPS remained largely invariant to visual-vestibular conflict.

MSTd showed the opposite pattern (Fig. 8*F*). Visual weights were consistently larger than vestibular weights across all cue offset (main effect of modality: $F_{(1,62)} = 40.08$, $p = 3.05 \times 10^{-8}$, two-way repeated-measures ANOVA), with no significant main effect of offset ($F_{(2,124)} = 0.78$, $p = 0.461$, Greenhouse–Geisser correction) and no significant modality-by-offset interaction ($F_{(2,124)} = 1.23$, $p = 0.296$, Greenhouse–Geisser correction). Therefore, MSTd maintained a stable visual-dominant weighting strategy regardless of cue conflict.

Together, these findings reveal fundamentally different multisensory weighting strategies across cortical self-motion areas. VPS and MSTd rely on relatively fixed vestibular- and visual-dominant weighting schemes, respectively. In contrast, VIP flexibly reallocated sensory contributions as cue conflict increases, shifting from visual-biased weighting under cue agreement toward increased vestibular weighting under conflict. This dynamic reweighting provides a mechanistic explanation for the robust population coding observed in VIP and suggests that flexible sensory weighting is a key component of conflict-resistant self-motion representations.

Feedforward-gated normalization explains VIP computations

Divisive normalization has been proposed as a canonical mechanism for computing underlying multisensory interactions. In this framework, weighted vestibular and visual inputs are combined and subsequently scaled by a population normalization pool. We first tested whether this standard normalization model, previously used to describe multisensory responses in VPS and MSTd, could account for VIP activity under visual-vestibular cue conflict.

Although the canonical model captured the overall structure of VIP responses, it assumes a fixed normalization pool and therefore cannot explicitly account for the offset-dependent gain modulation observed in Figures 4 and 5. We therefore extended the model by introducing a feedforward-gated term that allows the effective normalization pool to vary as a function of visual-vestibular heading offset (see Materials and Methods). This FGN model retained the excitatory drive and divisive normalization structure of the original model while adding a single parameter, λ, that controls the strength of offset-dependent modulation. When $\lambda = 0$, the model reduced exactly to the canonical normalization model.

The FGN model provided an excellent account of VIP responses. Predicted responses closely matched measured firing rates across all cue combination (Fig. 9*A*; $R^2 = 0.96$). Importantly, model comparison strongly favored the FGN model despite the additional parameter [$\Delta AIC = -117.14$, $\Delta BIC = -111.85$ relative to the canonical model (Fig. 9*B*)]. The fitted gating parameter was consistently positive in VIP ($\lambda = 0.154$), indicating that the effective normalization pool systematically varied with cue offset. In contrast, applying the same model extension to VPS and MSTd yielded virtually no improvement (Fig. 9*C*). In both areas, λ approached zero (VPS: $\lambda = 0.7 \times 10^{-25}$; MSTd: $\lambda = 0.8 \times 10^{-27}$; Fig. 9*C*), a model-selection metrics favored the standard normalization model. Thus, offset-dependent gating was not a generic consequence of model complexity but rather a computation selectively expressed in VIP.

A key question is whether the FGN model captured the central empirical signature of VIP coding. To address this issue, we repeated the response-ratio analyses using model-predicted responses. The FGN model successfully reproduced the hallmark gain pattern observed experimentally. When responses were aligned to the preferred vestibular heading, model-predicted gain was maximal when the visual cue was presented near its preferred heading and progressively decreased as visual heading deviated further from the preferred heading (Fig. 9*D*). A similar

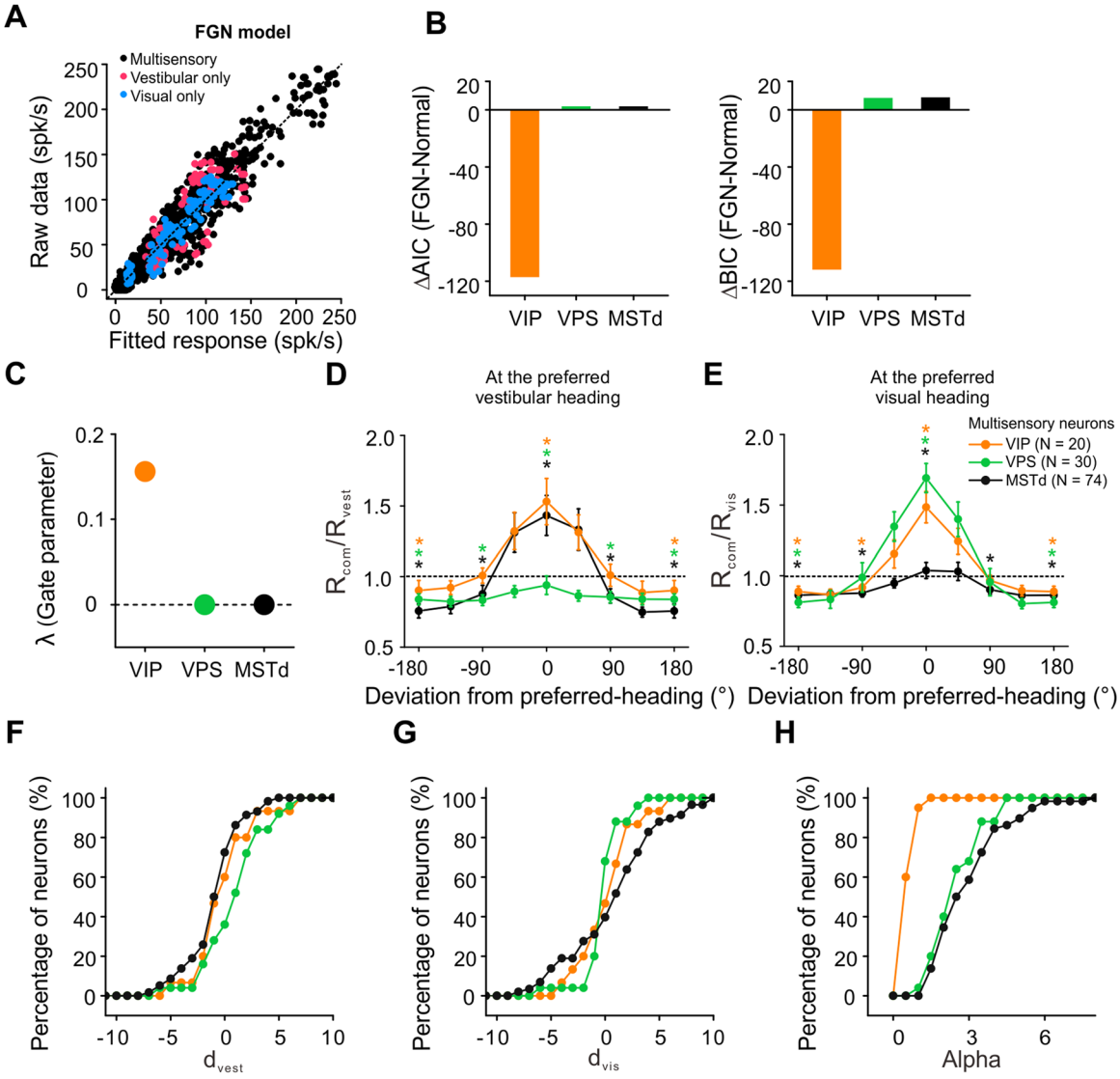


Figure 9. An FGN model accounts for conflict-dependent gain modulation. ***A***, Goodness of fit for the FGN model in VIP. Raw responses are plotted against fitted responses. Each neuron contributed 64 data points, corresponding to 8 vestibular heading conditions × 8 visual heading conditions. The diagonal line indicates perfect agreement between observed and predicted responses. Symbols denote multisensory, vestibular-only, and visual-only neurons included in the model fitting. Neurons were included based on significant heading tuning and trial-to-trial response reliability greater than 0.4; these criteria were applied before model comparison and independently of which model was favored. ***B***, Comparison of model-selection criteria between the FGN model and the standard normalization model across VIP, VPS, and MSTd. Left, ΔAIC; right, ΔBIC, calculated as the value for the FGN model minus that for the standard normalization model. Negative values favor the FGN model, whereas positive values favor the standard normalization model, after accounting for model complexity. Sample sizes were VIP ($N = 20$), VPS ($N = 30$), and MSTd ($N = 74$). ***C***, Fitted gating parameter λ in VIP, VPS, and MSTd. Positive λ values indicate conflict-dependent strengthening of normalization, whereas $\lambda \approx 0$ corresponds to conventional fixed normalization. VIP exhibited substantially larger λ values than VPS and MSTd, consistent with stronger conflict-sensitive modulation. ***D***, Model-predicted responses in the vestibular-reference condition. R_{com}/R_{vest} is plotted as a function of heading offset when vestibular heading was fixed at the preferred vestibular heading, following the format of Figure 3*C*. ***E***, Model-predicted responses in the visual-reference condition. R_{com}/R_{vis} is plotted as a function of visual-vestibular heading offset. ***F***, ***G***, Distributions of the vestibular input weight d_{vest} (***F***) and visual input weight d_{vis} (***G***) estimated from the model. ***H***, Distribution of the semisaturation constant α estimated from the model. The same neuronal subsets were included in all model comparison and parameter analyses. Sample sizes were VIP ($N = 20$), VPS ($N = 30$), and MSTd ($N = 74$; $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

pattern emerged in the visual-reference analysis (Fig. 9*E*). Thus, the model reproduced both multisensory enhancement at small preferred-heading deviations and gain suppression at larger deviations, providing a mechanistic explanation for the conflict-dependent gain control observed in VIP.

We next examined whether the fitted model parameters could explain the distinct sensory-weighting regimes observed across cortical areas. Vestibular input weights (d_{vest}) followed the hierarchy MSTd < VIP < VPS (Fig. 9*F*), whereas visual input weights (d_{vis}) showed the opposite ordering VPS < VIP < MSTd (Fig. 9*G*). These complementary gradients closely mirror the weighting profiles revealed in Figure 8, confirming that VPS and MSTd are biased toward vestibular and visual inputs, respectively, whereas VIP receives comparatively balanced sensory drive.

The semisaturation constant α revealed a second key distinction. In normalization models, smaller α values indicate stronger influence of the normalization pool on neuronal responses. VIP exhibited substantially smaller α values (VIP: 0.62 ± 0.13, $N = 20$) than either VPS (2.41 ± 0.20, $N = 30$; $p = 2.69 \times 10^{-6}$, Mann–Whitney test) or MSTd (2.88 ± 0.18, $N = 74$; $p = 1.20 \times 10^{-8}$, Mann–Whitney test; Fig. 9*H*), indicating that VIP activity is more strongly governed by divisive normalization. Consequently, changes in the effective normalization pool exert a larger influence on response gain in VIP than in the modality-dominant areas.

Together, these results identify FGN as a computational mechanism for conflict-dependent multisensory processing in VIP. Compared with VPS and MSTd, VIP combines more balanced visual and vestibular inputs with stronger normalization sensitivity. By allowing the visual-vestibular heading relationship to dynamically regulate the normalization pool, the FGN model explains how VIP enhances responses when the two cues specify similar self-motion headings, suppresses responses as their heading offsets increases, and simultaneously preserves robust population-level representations of heading direction.

Discussion

The present study demonstrates that visual-vestibular conflict in macaque VIP primarily regulates multisensory gain while preserving the underlying structure of heading representations. As expected, combined stimulation enhanced neuronal responses when visual and vestibular cues were congruent and reduced response gain as cue offset increased. Importantly, these effects occurred with relatively small shifts in preferred heading, indicating that cue conflict modulated response amplitude more strongly than tuning geometry. Despite substantial suppression at the single-neuron level, population FI remained robust across conflict conditions. Furthermore, VIP dynamically redistributed visual and vestibular contributions as cue conflict increased, a property not observed in the modality-dominant areas VPS and MSTd. Finally, a FGN model accounted for both the response patterns and the conflict-dependent gain modulation observed in VIP. Together, these findings characterize offset-dependent multisensory response modulation in VIP under controlled passive-stimulation conditions.

Conflict-dependent gain control preserves heading structure

Across cue-offset conditions, neuronal firing rates changed substantially, whereas preferred-heading shifts remained relatively small. This stability is consistent with previous findings that visual-motion direction preferences in VIP remain largely preserved across stimulus depths, despite disparity-dependent changes in response magnitude (Bremmer et al., 2013) and that motion-direction signals can remain separable from disparity and vergence signals (Caziot et al., 2025). Large tuning shifts would distort the geometry of the population code and potentially destabilize downstream heading decoding (Seung and Sompolinsky, 1993; Pouget et al., 2000; Jazayeri and Movshon, 2006; Gu et al., 2010). In contrast, gain modulation preserves the representational structure of the population while allowing flexible adjustment of sensory influence (Salinas and Sejnowski, 2001; Fetsch et al., 2011; Ferguson and Cardin, 2020).

Such a strategy may be particularly advantageous under natural conditions, where sensory signals frequently vary in reliability and consistency due to environmental dynamics, self-generated movements, and sensory noise (Dokka et al., 2019; Keshavarzi et al., 2023). Consistent with this interpretation, population FI in VIP remained remarkably robust even at large cue offsets. However, robust FI should not be interpreted as evidence that the brain perceptually fused discrepant visual and vestibular cues. Rather, FI quantifies the discriminability supported by a given population response pattern. The present findings therefore suggest that VIP maintains reliable heading coding across multisensory contexts while dynamically regulating the influence of conflicting sensory inputs.

Area-specific multisensory weighting profiles in VIP, VPS, and MSTd

Comparison with VPS and MSTd provides modality-dominant reference profiles for interpreting VIP. VPS and MSTd represent vestibular-dominant and visual-dominant self-motion areas, respectively. Under the same fixation task and analytical framework, VIP did not simply exhibit an intermediate profile between these areas; instead, its relative visual and vestibular weights showed a stronger dependence on cue offset.

Consistent with previous studies, MSTd maintained a predominantly visual weighting profile across cue offsets, reflecting its well-established role in optic-flow processing (Gu et al., 2008; Morgan et al., 2008). VPS exhibited the opposite pattern, with stable vestibular dominance and winner-take-all-like weighting under cue conflict (Zhao et al., 2023). In contrast, VIP dynamically redistributed visual and vestibular contributions according to cue consistency. Visual influences were strongest when the two modalities were congruent, whereas vestibular contributions increased progressively as cue conflict grew. The correlation analyses likewise showed significant relationships in both vestibular- and visual-reference analyses across the tested preferred-heading deviations, in contrast to the more modality-specific patterns in MSTd and VPS. These observations indicate that VIP expressed a more balanced and bidirectional dependence on visual and vestibular drive than the two reference areas.

The observed weighting differences may partly reflect differences in the relative strengths of visual and vestibular inputs (Allum et al., 1976). They may also depend on stimulus configuration. Peripheral visual stimulation preferentially supports global or egocentric self-motion perception, whereas central stimulation is more strongly associated with exocentric object-motion perception (Brandt et al., 1973; Dichgans and Brandt, 1978). Because retinal eccentricity was not manipulated independently here, its contribution to visual-vestibular weighting in VIP remains unresolved.

These findings support a view of the cortical self-motion network as a distributed system in which different areas perform complementary computations. MSTd contributes to the intermediate cortical processing of optic-flow signals related to self-motion (Lappe et al., 1996; Duffy, 1998; Gu et al., 2008; Morgan et al., 2008), and VPS may provide a vestibular anchor

during sensory conflict (Chen et al., 2011c; Zhao et al., 2021). VIP may regulate the relative influence of both modalities according to cue consistency (Bremmer et al., 2002b; Chen et al., 2011a, 2013; Zaidel et al., 2017; Zaidel et al., 2021). This interpretation does not imply that such response properties are unique to VIP. Neighboring parietal areas were not examined and may show related computations.

Dynamic sensory weighting and adaptive multisensory inference

The conflict-dependent gain modulation observed in VIP is consistent with a neural operation that could support adaptive multisensory inference (Kording et al., 2007; Shams and Beierholm, 2010; Rohe and Noppeney, 2015a). When cues are consistent, greater combined-response gain could enhance mutually supportive evidence; when they diverge, reduced gain could limit the influence of inconsistent inputs. In this sense, VIP responses covaried with sensory consistency in a manner compatible with flexible regulation of multisensory influence.

Importantly, the present results do not demonstrate explicit causal inference or perceptual segregation. Animals performed a fixation task without reporting perceived heading, and thus the neural activity cannot be directly linked to trial-by-trial perceptual decisions. Nevertheless, the observed computations resemble mechanisms that would be useful for regulating multisensory integration under uncertainty (Kording et al., 2007; Shams and Beierholm, 2010; Rohe and Noppeney, 2015b).

Previous studies have shown that VIP contains robust visual and vestibular self-motion signals, choice-related activity, and multisensory recalibration signals (Bremmer et al., 2002b; Chen et al., 2011b, 2013; Zaidel et al., 2017; Zaidel et al., 2021). Together with these findings, the present results are consistent with a contribution of VIP to cue-dependent modulation of multisensory responses, but they do not establish its causal role in perception, recalibration, or action selection.

Conflict-sensitive normalization as a computational mechanism

The FGN model provides a compact computational account of the multisensory interactions observed in VIP. Divisive normalization has been proposed as a canonical cortical computation for nonlinear sensory interactions, competition, and gain control (Ohshiro et al., 2011). The standard normalization model captured the overall structure of VIP responses reasonably well. However, it failed to account for the systematic dependence of gain modulation on cue conflicts. The FGN extension allowed cue offset to modulate the effective normalization pool and reproduced enhancement near congruency together with stronger suppression at larger offsets.

Importantly, the model improvement was selective for VIP. The gating parameter was effectively absent in VPS and MSTd, suggesting that conflict-sensitive normalization is not a generic property of cortical self-motion areas but instead represents a distinctive computational feature of VIP.

The fitted parameters further revealed two characteristics that distinguish VIP from the modality-dominant areas. First, VIP received relatively balanced visual and vestibular inputs, whereas VPS and MSTd were biased toward vestibular and visual drive, respectively. Second, VIP exhibited substantially stronger normalization sensitivity, as reflected by smaller semisaturation constants. Together, these properties make VIP particularly well suited for conflict-dependent modulation of multisensory gain.

Although the model does not specify the underlying circuit implementation, it demonstrates that dynamic regulation of normalization is sufficient to explain the principal features of VIP responses. One possibility is that cue conflict alters the recruitment of shared inhibitory populations or recurrent normalization circuits (Carandini and Heeger, 2011; Heeger and Zemlianova, 2020). Alternatively, top-down contextual signals may modulate normalization strength according to sensory consistency (Macaluso, 2009; Reynolds and Heeger, 2009). Distinguishing among these possibilities will require future experiments combining simultaneous population recordings, circuit perturbations, and causal manipulations.

Limitations and future directions

Several limitations should be considered. First, the fixation task did not provide behavioral reports, so the present study cannot directly relate VIP activity to perceptual reports, subjective cue weighting, or causal-inference judgments. Moreover, it remains unclear whether the response relationships observed here generalize to natural self-motion, in which visual and vestibular signals are continuous and accompanied by proprioceptive, motor-related, and behavioral signals. Future experiments combining behavioral heading discrimination with neural recordings will be important for establishing the relationship between conflict-dependent gain modulation and perception.

Second, cue reliability was not manipulated independently of cue conflict. Thus, the observed changes in sensory weighting may reflect cue discrepancy, relative cue reliability, or interactions between the two. Future studies independently varying these factors will be necessary to dissociate their respective contributions.

Third, neurons were recorded sequentially rather than simultaneously. The FI analysis therefore did not incorporate correlated variability, which can substantially influence population coding. Large-scale simultaneous recordings will be required to determine how noise correlations affect heading information under cue conflict.

Finally, although the FGN model captured the principal response patterns in VIP, it remains a descriptive model at the algorithmic level. Future work integrating anatomical, physiological, and computational approaches will be needed to identify the circuit mechanisms that implement conflict-sensitive normalization.

In summary, under controlled passive-stimulation conditions, VIP exhibited cue-offset-dependent modulation of visual-vestibular response gain, systematic variation in fitted visual and vestibular weights, and relatively stable population heading information. Rather than relying on a fixed dominant modality, VIP adjusted multisensory gain and sensory weighting according to cue consistency. These findings characterize conflict-sensitive response modulation in VIP while leaving open whether similar computations operate during natural behavior or in neighboring cortical areas.

References

Allum JH, Graf W, Dichgans J, Schmidt CL (1976) Visual-vestibular interactions in the vestibular nuclei of the goldfish. Exp Brain Res 26:463–485.

Angelaki DE, Cullen KE (2008) Vestibular system: the many facets of a multimodal sense. Annu Rev Neurosci 31:125–150.

Brandt T, Dichgans J, Koenig E (1973) Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. Exp Brain Res 16:476–491.

Bremmer F, Kubischik M, Pekel M, Lappe M, Hoffmann KP (1999) Linear vestibular self-motion signals in monkey medial superior temporal area. Ann N Y Acad Sci 871:272–281.

Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann K, Zilles K, Fink GR (2001) Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. Neuron 29:287–296.
Bremmer F, Duhamel JR, Ben Hamed S, Graf W (2002a) Heading encoding in the macaque ventral intraparietal area (VIP). Eur J Neurosci 16:1554–1568.
Bremmer F, Klam F, Duhamel JR, Ben Hamed S, Graf W (2002b) Visual–vestibular interactive responses in the macaque ventral intraparietal area (VIP). Eur J Neurosci 16:1569–1586.
Bremmer F, Schlack A, Kaminiarz A, Hoffmann KP (2013) Encoding of movement in near extrapersonal space in primate area VIP. Front Behav Neurosci 7:8.
Butler JS, Smith ST, Campos JL, Bulthoff HH (2010) Bayesian integration of visual and vestibular signals for heading. J Vis 10:1–13.
Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51–62.
Caziot B, Fathkhani S, Bremmer F (2025) Coding of egocentric distance in the macaque ventral intraparietal area. iScience 28:111993.
Chen A, DeAngelis GC, Angelaki DE (2011a) Representation of vestibular and visual cues to self-motion in ventral intraparietal cortex. J Neurosci 31:12036–12052.
Chen A, DeAngelis GC, Angelaki DE (2011b) A comparison of vestibular spatiotemporal tuning in macaque parietoinsular vestibular cortex, ventral intraparietal area, and medial superior temporal area. J Neurosci 31: 3082–3094.
Chen A, DeAngelis GC, Angelaki DE (2011c) Convergence of vestibular and visual self-motion signals in an area of the posterior Sylvian fissure. J Neurosci 31:11617–11627.
Chen A, DeAngelis GC, Angelaki DE (2013) Functional specializations of the ventral intraparietal area for multisensory heading discrimination. J Neurosci 33:3567–3581.
Chen A, Zeng F, DeAngelis GC, Angelaki DE (2021) Dynamics of heading and choice-related signals in the parieto-insular vestibular cortex of macaque monkeys. J Neurosci 41:3254–3265.
Churan J, Kaminiarz A, Schwenk JCB, Bremmer F (2021) Action-dependent processing of self-motion in parietal cortex of macaque monkeys. J Neurophysiol 125:2432–2443.
DeAngelis GC, Angelaki DE (2012) Visual-Vestibular integration for self-motion perception. In: *The neural bases of multisensory processes* (Murray MM, Wallace MT, eds), pp 629–649. Boca Raton, FL: CRC Press.
Dichgans J, Brandt T (1978) Visual-vestibular interaction: effects on self-motion perception and postural control. In: *Perception* (Held R, Leibowitz HW, Teuber H-L, eds), pp 755–804. Berlin, Heidelberg: Springer.
Dokka K, Park H, Jansen M, DeAngelis GC, Angelaki DE (2019) Causal inference accounts for heading perception in the presence of object motion. Proc Natl Acad Sci U S A 116:9060–9065.
Duffy CJ (1998) MST neurons respond to optic flow and translational movement. J Neurophysiol 80:1816–1827.
Duhamel JR, Colby CL, Goldberg ME (1998) Ventral intraparietal area of the macaque: congruent visual and somatic response properties. J Neurophysiol 79:126–136.
Ferguson KA, Cardin JA (2020) Mechanisms underlying gain modulation in the cortex. Nat Rev Neurosci 21:80–92.
Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE (2009) Dynamic reweighting of visual and vestibular cues during self-motion perception. J Neurosci 29:15601–15612.
Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE (2011) Neural correlates of reliability-based cue weighting during multisensory integration. Nat Neurosci 15:146–154.
Gibson JJ (1950) *The perception of the visual world*. Boston, MA: Houghton Mifflin.
Graf W (2025) Visual-vestibular guided control of posture, movement and self-motion perception. In: *Encyclopedia of the human brain* (Grafman JH, ed), Ed 2, pp 582–611. London: Elsevier.
Gu Y, Watkins PV, Angelaki DE, DeAngelis GC (2006) Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. J Neurosci 26:73–85.
Gu Y, Angelaki DE, DeAngelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. Nat Neurosci 11:1201–1210.
Gu Y, Fetsch CR, Adeyemo B, DeAngelis GC, Angelaki DE (2010) Decoding of MSTd population activity accounts for variations in the precision of heading perception. Neuron 66:596–609.
Heeger DJ, Zemlianova KO (2020) A recurrent circuit implements normalization, simulating the dynamics of V1 activity. Proc Natl Acad Sci U S A 117:22494–22505.
Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. Nat Neurosci 9:690–696.
Keshavarzi S, Velez-Fort M, Margrie TW (2023) Cortical integration of vestibular and visual cues for navigation, visual processing, and perception. Annu Rev Neurosci 46:301–320.
Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L (2007) Causal inference in multisensory perception. PLoS One 2:e943.
Lacquaniti F, La Scaleia B, Zago M (2023) Noise and vestibular perception of passive self-motion. Front Neurol 14:1159242.
Lappe M, Bremmer F, Pekel M, Thiele A, Hoffmann KP (1996) Optic flow processing in monkey STS: a theoretical and experimental approach. J Neurosci 16:6265–6285.
Laurens J, Angelaki DE (2017) A unified internal model theory to resolve the paradox of active versus passive self-motion sensation. Elife 6:e28074.
Macaluso E (2009) Orienting of spatial attention and the interplay between the senses. Cortex 46:282–297.
Morgan ML, DeAngelis GC, Angelaki DE (2008) Multisensory integration in macaque visual cortex depends on cue reliability. Neuron 59:662–673.
Ohshiro T, Angelaki DE, DeAngelis GC (2011) A normalization model of multisensory integration. Nat Neurosci 14:775–782.
Ohshiro T, Angelaki DE, DeAngelis GC (2017) A neural signature of divisive normalization at the level of multisensory integration in primate cortex. Neuron 95:399–411.
Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. Nat Rev Neurosci 1:125–132.
Raiser TM, Flanagin VL, Duering M, van Ombergen A, Ruehl RM, Zu Eulenburg P (2020) The human corticocortical vestibular network. Neuroimage 223:117362.
Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168–185.
Rohe T, Noppeney U (2015a) Sensory reliability shapes perceptual inference via two mechanisms. J Vis 15:22.
Rohe T, Noppeney U (2015b) Cortical hierarchies perform Bayesian causal inference in multisensory perception. PLoS Biol 13:e1002073.
Salinas E, Sejnowski TJ (2001) Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. Neuroscientist 7:430–440.
Schaafsma SJ, Duysens J (1996) Neurons in the ventral intraparietal area of awake macaque monkey closely resemble neurons in the dorsal part of the medial superior temporal area in their responses to optic flow patterns. J Neurophysiol 76:4056–4068.
Schlack A, Hoffmann KP, Bremmer F (2002) Interaction of linear vestibular and visual stimulation in the macaque ventral intraparietal area (VIP). Eur J Neurosci 16:1877–1886.
Seung HS, Sompolinsky H (1993) Simple models for reading neuronal population codes. Proc Natl Acad Sci U S A 90:10749–10753.
Shams L, Beierholm UR (2010) Causal inference in perception. Trends Cogn Sci 14:425–432.
Stein BE, Stanford TR, Rowland BA (2014) Development of multisensory integration from the perspective of the individual neuron. Nat Rev Neurosci 15:520–535.
Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA (2004) Unifying multisensory signals across time and space. Exp Brain Res 158:252–258.
Warren WHJ, Hannon DJ (1990) Eye movements and optical flow. J Opt Soc Am A 7:160–169.
Zaidel A, DeAngelis GC, Angelaki DE (2017) Decoupled choice-driven and stimulus-related activity in parietal neurons may be misrepresented by choice probabilities. Nat Commun 8:715.
Zaidel A, Laurens J, DeAngelis GC, Angelaki DE (2021) Supervised multisensory calibration signals are evident in VIP but not MSTd. J Neurosci 41: 10108–10119.
Zeng F, Zaidel A, Chen A (2023) Contrary neuronal recalibration in different multisensory cortical areas. Elife 12:e82895.
Zhao B, Zhang Y, Chen A (2021) Encoding of vestibular and optic flow cues to self-motion in the posterior superior temporal polysensory area. J Physiol 599:3937–3954.
Zhao B, Wang R, Zhu Z, Yang Q, Chen A (2023) The computational rules of cross-modality suppression in the visual posterior Sylvian area. iScience 26:106973.