

Article

# Visual-Based Spatial Coordinate Dominates Probabilistic Multisensory Inference in Macaque MST-d Disparity Encoding

Jiawei Zhang <sup>1</sup>, Mingyi Huang <sup>1</sup>, Yong Gu <sup>2</sup>, Aihua Chen <sup>3,\*</sup> and Yuguo Yu <sup>1,\*</sup>

<sup>1</sup> Shanghai Artificial Intelligence Laboratory, Research Institute of Intelligent and Complex Systems and Institute of Science and Technology for Brain-Inspired Intelligence, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Human Phenome Institute, Shanghai 200433, China

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

<sup>3</sup> Key Laboratory of Brain Functional Genomics (Ministry of Education), East China Normal University, 3663 Zhongshan Road N., Shanghai 200062, China

\* Correspondence: ahchen@brain.ecnu.edu.cn (A.C.); yuyuguo@fudan.edu.cn (Y.Y.)

**Abstract:** Numerous studies have demonstrated that animal brains accurately infer whether multisensory stimuli are from a common source or separate sources. Previous work proposed that the multisensory neurons in the dorsal medial superior temporal area (MST-d) serve as integration or separation encoders determined by the tuning–response ratio. However, it remains unclear whether MST-d neurons mainly take a sense input as a spatial coordinate reference for carrying out multisensory integration or separation. Our experimental analysis shows that the preferred tuning response to visual input is generally larger than vestibular according to the Macaque MST-d neuronal recordings. This may be crucial to serving as the base of coordinate reference when the subject perceives moving direction information from two senses. By constructing a flexible Monte-Carlo probabilistic sampling (fMCS) model, we validate this hypothesis that the visual and vestibular cues are more likely to be integrated into a visual-based coordinate rather than vestibular. Furthermore, the property of the tuning gradient also affects decision-making regarding whether the cues should be integrated or not. To a dominant modality, an effective decision is produced by a steep response-tuning gradient of the corresponding neurons, while to a subordinate modality a steep tuning gradient produces a rigid decision with a significant bias to either integration or separation. This work proposes that the tuning response amplitude and tuning gradient jointly modulate which modality serves as the base coordinate for the reference frame and the direction change with which modality is decoded effectively.

**Keywords:** sensory integration and separation; cortical processing hierarchy; MST-d; causal inference



**Citation:** Zhang, J.; Huang, M.; Gu, Y.; Chen, A.; Yu, Y. Visual-Based Spatial Coordinate Dominates Probabilistic Multisensory Inference in Macaque MST-d Disparity Encoding. *Brain Sci.* **2022**, *12*, 1387. <https://doi.org/10.3390/brainsci12101387>

Academic Editor: Fiorenzo Moscatelli

Received: 18 September 2022

Accepted: 7 October 2022

Published: 13 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The primate brain frequently combines multisensory information from different sensory modalities, such as information of visual, vestibular, auditory, and haptic origin, to improve the perception of the external world. Specifically, both visual and vestibular information can be used to infer motion direction, and combining the two is crucial for discriminating self-motion and object motion in a real-world situation. To achieve such discrimination, the neural system integrates or separates multisensory inputs to attribute information from different modalities to either a common source or two independent sources [1,2]. Nevertheless, the mechanisms through which neurons implement multisensory integration and separation have long been debated [3–11]. Although extensive computational studies provide various perspectives, there is a general lack of evidence from physiological records.

On the behavioral level, many researchers have demonstrated that the human brain adopts a discrimination strategy similar to the Bayesian approach [12–17], which makes probabilistic inferences based on the disparity between modalities as well as the changes in cue reliability. By representing the reliability as a prior distribution, the numerical fitting of the Bayesian model provides convincing approximations to the decision probabilities on the behavioral level. However, suffering from a gap between empirical results and mathematical description, the Bayesian strategy faces the challenge of other non-Bayesian strategies such as fixed inference, and even non-causal strategies such as forced fusion [18]. The inference strategy of brain multisensory processing remains elusive, mainly due to the lack of physiological evidence and the resultant lack of parameter constraints.

Previous studies seek neuronal correlates of multisensory processing in the primate dorsal medial superior temporal area (MST-d), which was a classical multisensory area responding to both visual and vestibular modalities [19–21]. Gu et al. found that the subtype of MST-d neurons (neurons preferring congruent and opposite stimuli directions) showed tuning sensitivity change that was analogous to behavioral change [1]. The experiments were designed to present visual and vestibular stimuli from the same direction on the horizontal plane. Fetsch et al. introduced a small conflict between the two directions ( $\pm 4^\circ$ ) and proposed that MST-d neurons could account for weighing cues according to their relative reliabilities. Nevertheless, the MST-d neurons generally have large receptive fields [22], which drew attention to the response property on the  $360^\circ$  horizontal plane. Rideaux et al. trained an artificial network to perform causal inference and showed that a feedforward network can estimate the visual-vestibular motion by reading out the activity from neurons tuned to congruent and opposite directions [23]. WH Zhang proposed that congruent neurons encode integration while opposite neurons encode separation from a vector perspective, which covered the inference case when visual and vestibular directions are opposite [8]. JW Zhang et al. seek a correlation between neuronal preference and response property and proposed that the response strength ratio of MST-d neurons to both modalities determined the neurons serving as integration or separation encoders in a decision trial [24] and triggered the emergence of congruent and opposite neurons. The decision trial was denoted as explicit inference by Acerbi et al. [18], deciding whether the sensory cues originate from the same origin or separate origins. Nevertheless, the strength ratio did not specify visual and vestibular modalities. It is previously reported that the MST-d neurons generally present more significant tuning to visual modality than vestibular [25], but little is known about how such response discrimination affects the multisensory inference at a behavioral level, or whether the two modalities contribute to the perception distinctively at all.

The contribution potentially determines whether one modality serves as the reference frame, based on which the perception coordinates about self-motion are set up. During integration, the reference frame also determines whether the perceived common direction approximates closer to the visual modality or vestibular modality.

In this study, we seek to combine the physiological recordings and computational models and explore the key topic: What is the functional difference between the visual and vestibular signals during multisensory processing in MST-d? How does the response distinction affect the decision-making of integration or separation? We adopt a biophysical model that decides whether the senses should be integrated or separated. The model utilizes physiological recording results and reveals the heterogeneity of visual and vestibular contributions within MST-d.

## 2. Methods

### 2.1. Subjects and Surgery

Two male rhesus monkeys (*Macaca mulatta*) served as subjects. The general procedures followed in this study have been described previously [26,27]. Each animal was outfitted with a circular molded plastic ring anchored to the skull with titanium T-bolts and dental acrylic. To monitor eye movements, a scleral search coil was implanted in

each monkey. Animals were trained to fixate on a central target for fluid rewards using operant conditioning.

### 2.2. Vestibular and Visual Stimuli

A 6-degree-of-freedom motion platform (MOOG 6DOF2000E; Moog, East Aurora, NY, USA) was used to passively translate the animals along one of eight directions in the horizontal plane, spaced  $45^\circ$  apart. A tangent screen was affixed to the front surface of the field coil frame and visual stimuli were projected onto it by a three-chip digital light projector (Mirage 2000; Christie Digital Systems, Cypress, CA, USA). The screen measured  $60 \times 60$  cm and was mounted 30 cm in front of the monkey, thus subtending  $\sim 90^\circ \times 90^\circ$ . The visual stimuli simulated translational movement along the same eight directions through a three-dimensional cloud of stars. Each “star” was a triangle that measured  $0.15 \text{ cm} \times 0.15 \text{ cm}$ ; the cloud measured 100 cm wide by 100 cm tall by 40 cm deep and had a star density of 0.01 per  $\text{cm}^3$ . To provide stereoscopic cues, the cloud was rendered as a red-green anaglyph and viewed through custom red-green goggles. The optic flow field contained naturalistic cues mimicking the translation of the observer in the horizontal plane, including motion parallax, size variations, and binocular disparity.

### 2.3. Electrophysiological Recordings

We recorded action potentials extracellularly from both hemispheres in each of the two monkeys. For each recording session, a tungsten microelectrode was passed through a transdural guide tube and advanced using a micromanipulator. An amplifier, an eight-pole bandpass filter (400–5000 Hz), and a dual voltage-time window discriminator (BAK Electronics, Mount Airy, MD, USA) were used to isolate action potentials from single neurons. Action potential times and behavioral events were recorded with 1 ms accuracy by a computer. Eye coil signals were processed with a low-pass filter and sampled at 250 Hz.

Magnetic resonance imaging (MRI) scans and Caret software analyses along with physiological criteria were used to guide electrode penetration into the MST-d area [25]. Neurons were isolated while a large field of flickering dots was presented. In some experiments, we further advanced the electrode tip into the lower bank of the superior temporal sulcus to verify the presence of neurons with response characteristics typical of the MT [25]. Receptive field locations changed as expected across guide tube locations based on the known topography of the MT [25].

### 2.4. Experimental Protocol

We measured neural responses to eight heading directions evenly spaced every  $45^\circ$  in the horizontal plane. Neurons were tested under three experimental conditions. (1) In vestibular trials, the monkeys were required to maintain fixation on a central dot on an otherwise blank screen while being translated in one of the eight directions. (2) In visual trials, the monkeys were presented with optic flow simulating self-motion (in the same eight directions), while the platform remained stationary. (3) In bimodal trials, the monkeys experienced both translational motion and optic flow. We paired all eight vestibular headings with all eight visual headings for a total of 64 bimodal stimuli. Eight of these 64 combinations were strictly congruent, meaning that the visual and vestibular cues simulated the same heading. The remaining 56 cases had conflicting cue stimuli. This relative proportion of strictly congruent and conflicting stimuli was adopted to characterize the neuronal combination rule. Each translation followed a Gaussian velocity profile. It had a duration of 2 s, an amplitude of 13 cm, a peak velocity of 30 cm/s, and a peak acceleration of  $0.1 \times g$  ( $981 \text{ cm/s}^2$ ).

These three stimulus conditions were interleaved randomly along with blank trials, which included neither translation nor optic flow. Ideally, five repetitions of each unique stimulus were collected for a total of 405 trials. Experiments with fewer than three repetitions were excluded from the analysis. When isolation remained satisfactory, we ran

additional blocks of trials with the coherence of the visual stimulus reduced to 50% and/or 25%. Motion coherence was lowered by randomly relocating a percentage of the dots on every subsequent video frame. For example, we randomly selected one-quarter of the dots in every frame at 25% coherence and updated their positions to new positions consistent with the simulated motion, while the other three-quarters of the dots were plotted at new, random locations within the 3D cloud. Each block of trials consisted of both unimodal and bimodal stimuli at the corresponding coherence level. When a cell was tested at multiple coherence levels, both unimodal vestibular tuning and unimodal visual tuning were independently assessed in each block.

Trials were initiated by displaying a  $0.2^\circ \times 0.2^\circ$  fixation target on the screen. The monkeys were required to fixate on the target for 200 ms before the stimulus was presented and to maintain fixation within a  $3^\circ \times 3^\circ$  window to receive a liquid reward. Trials in which the monkeys broke fixation were aborted and discarded.

### 2.5. Data Analysis

The neural responses were binned in 100-ms time windows. Mean neural responses were averaged from 5 trials, and the units of measurement were spikes per second. The outliers in the 5 trials were removed, and the mean response was averaged from the remaining 4 trials. Using MATLAB (MathWorks, Natick, MA, USA), we chose the window of 750 ms to 1250 ms to select valid data. We considered a neuron to have discriminative tuning properties to one specific stimulus modality if the maximum response was 5 spikes/s more than the minimum response of the same curve. Tuning curve symmetry was not considered. Neurons that failed to meet this requirement for either the visual or the vestibular unisensory condition were considered unisensory-tuned neurons or poorly tuned neurons and removed from further analysis. Then, we computed the response ratio based on the visual and vestibular unisensory tuning curves of that neuron in the same time window. A threshold of 1.6 was chosen to discriminate between balanced and imbalanced data. We analyzed the tuning curve in the time window from 900 ms to 1250 ms, which corresponds to the maximum movement speed and maximum neural response. The window parameters were first scanned and then selected comprehensively to show the discrimination of the integration probability  $P_{int}$  between the balanced and imbalanced neurons and to be physiologically plausible.

The group  $\Delta\theta$  distribution was rectified by doubling the probability at  $0^\circ$  and  $180^\circ$ , while the probabilities in other directions remained the same. This procedure was followed because of the experimental protocol in which directions were binned in  $45^\circ$  intervals; a  $0^\circ$  preference encompassed preferences from  $-22.5^\circ$  to  $+22.5^\circ$ , and a  $180^\circ$  preference encompassed preferences from  $167.5^\circ$  to  $202.5^\circ$ . However, each of the other preference bins was represented twice because both sides were included (for example, a  $\Delta\theta$  of  $45^\circ$  encompassed directions from  $22.5^\circ$  to  $67.5^\circ$  and from  $-22.5^\circ$  to  $-67.5^\circ$ ). To align the widths of the probability bins, the data counted at  $0^\circ$ , and  $180^\circ$  were included twice.

### 2.6. Multisensory Tuning Curves Averaging

The multisensory tuning curve denoted as  $f_{bal}$  and  $f_{imbal}$  in balanced and imbalanced groups each were direction and modality averaged. The single maximal response in the two-dimensional multisensory response grid is first selected as the maximal response in  $f_{bal}$  and  $f_{imbal}$  ( $R_{mul}^{max} = \max(R(\theta_{vis,mul}, \theta_{vis,mul}))$ ). The other responses were computed as the averaged response of those with the same relative distance ( $\Delta\theta_{rel}$ ) with the maximal one. For simplicity, we considered  $R(\theta_{vis,mul} + \Delta\theta_{rel}, \theta_{vis,mul} + \Delta\theta_{rel})$  as same distant as  $R(\theta_{vis,mul} + \Delta\theta_{rel}, \theta_{vis,mul})$  and  $(\theta_{vis,mul}, \theta_{vis,mul} + \Delta\theta_{rel})$ , thus the averaging is square-wise centered by the maximal response.

### 2.7. Bayesian Modeling

We adopted Bayesian optimal inference strategy as proposed in [17]. Here we explain the theory briefly. The strategy determines whether the sensory measurements  $X_{vis}$  and

$X_{ves}$  originates from a common source ( $C = 1$ ) or separate sources ( $C = 2$ ). If there is a common source, draw a position  $s$  from a normal prior distribution  $\mathcal{N}(0, \sigma_p)$ , which stands for a normal distribution with 0 mean and standard deviation  $\sigma_p$ . Since it's a common source, we denote  $S_{vis} = S_{ves} = S_{common}$ . If there are two sources, then the position  $S_{vis}$  and  $S_{ves}$  are independent of  $\mathcal{N}(0, \sigma_p)$ . We assumed that visual and vestibular signals are corrupted by gaussian noise with standard deviation  $\sigma_{vis}$  and  $\sigma_{ves}$ . Thus, the sensory measurements attribute to  $X_{vis} \sim \mathcal{N}(S_{vis}, \sigma_{vis})$  and  $X_{ves} \sim \mathcal{N}(S_{ves}, \sigma_{ves})$ . The noise is independent across modalities. Given the statistical parameters, the posterior likelihood of a common source is expressed as,

$$p(C = 1|X_{vis}, X_{ves}) = \frac{p(X_{vis}, X_{ves}|C = 1)p(C = 1)}{p(X_{vis}, X_{ves})} \quad (1)$$

According to Bayes theorem,

$$p(C = 1|X_{vis}, X_{ves}) = \frac{p(X_{vis}, X_{ves}|C = 1)p(C = 1)}{p(X_{vis}, X_{ves}|C = 1)p_{common} + (X_{vis}, X_{ves}|C = 2)(1 - p_{common})}$$

To  $p(X_{vis}, X_{ves}|C = 1)$ ,

$$p(X_{vis}, X_{ves}|C = 1) = \int p(X_{vis}, X_{ves}|s)p(s)ds \quad (2)$$

Based on conditional independence hypothesis,

$$\int p(X_{vis}, X_{ves}|s)p(s)ds = \int p(X_{vis}|s)p(X_{ves}|s)p(s)ds \quad (3)$$

All the factors are Gaussians, then we obtain the analytical solution:

$$p(X_{vis}, X_{ves}|C = 1) = \frac{1}{2\pi\sqrt{\sigma_{vis}^2\sigma_p^2 + \sigma_{ves}^2\sigma_p^2 + \sigma_{vis}^2\sigma_{ves}^2}} \exp\left[-\frac{1}{2}\frac{(X_{vis}-X_{ves})^2\sigma_p^2 + (X_{ves}-\mu_p)^2\sigma_{vis}^2 + (X_{vis}-\mu_p)^2\sigma_{ves}^2}{\sigma_{vis}^2\sigma_p^2 + \sigma_{ves}^2\sigma_p^2 + \sigma_{vis}^2\sigma_{ves}^2}\right] \quad (4)$$

where  $\mu_p = 0$  is the mean of prior.

Similarly, to  $p(X_{vis}, X_{ves}|C = 2)$ ,

$$p(X_{vis}, X_{ves}|C = 2) = \iint p(X_{vis}, X_{ves}|S_{vis}, S_{ves})p(S_{vis}, S_{ves})dS_{vis}dS_{ves} \quad (5)$$

We can derive the analytical solution as,

$$p(X_{vis}, X_{ves}|C = 2) = \frac{1}{2\pi\sqrt{(\sigma_{vis}^2 + \sigma_p^2)(\sigma_{ves}^2 + \sigma_p^2)}} \exp\left[-\frac{1}{2}\left(\frac{(X_{vis} - \mu_p)^2}{\sigma_{vis}^2 + \sigma_p^2} + \frac{(X_{ves} - \mu_p)^2}{\sigma_{ves}^2 + \sigma_p^2}\right)\right] \quad (6)$$

To reach a binary decision ( $C = 1$  or  $C = 2$ ), we assume an optimal Bayesian observer that reports a common source ( $C = 1$ ) when  $p(C = 1|X_{vis}, X_{ves}) > 0.5$ , and vice versa.

In summary, the generative Bayesian strategy is featured by free parameters  $\sigma_p$ ,  $\sigma_{vis}$ ,  $\sigma_{ves}$  and the prior  $p_{common}$ . We assume that  $\sigma_{vis} = \sigma_{ves}$ , reducing the free parameters to 3. The external cue disparity condition is simulated by the hyper-parameter  $0^\circ \leq |S_{vis} - S_{ves}| \leq 180^\circ$ . The fitting of fMCS model follows the least-squares estimate (LSE).

### 3. Results

#### 3.1. Quantification of MST-d Neuronal Reliability Weightings Based on Tuning Curves

We began with experimental data analysis. The dataset includes physiological recordings of 158 MST-d neurons (as in [24]). In the experiments, monkeys were trained to actively perceive visual and vestibular stimuli with various directions ( $45^\circ$  intervals) on the horizontal plane but were not required to give behavioral reports. Each neuron was recorded in three conditions: the visual and vestibular unisensory stimulus conditions

and the visual-and-vestibular multisensory stimuli condition. Previous work proposed that the response balance between the two unisensory conditions determined neuronal function during the inference [24]. In this work, we similarly categorize the dataset as balanced and imbalanced neurons are discriminated by a ratio threshold of 1.6, which specifies only the balance between the dominant and subordinate unisensory responses and does not determine the visual or vestibular origin. The ratio in this work slightly deviates from previous work (1.7 in [24]) in choosing a larger time window, rendering the neuronal activity more moderate,

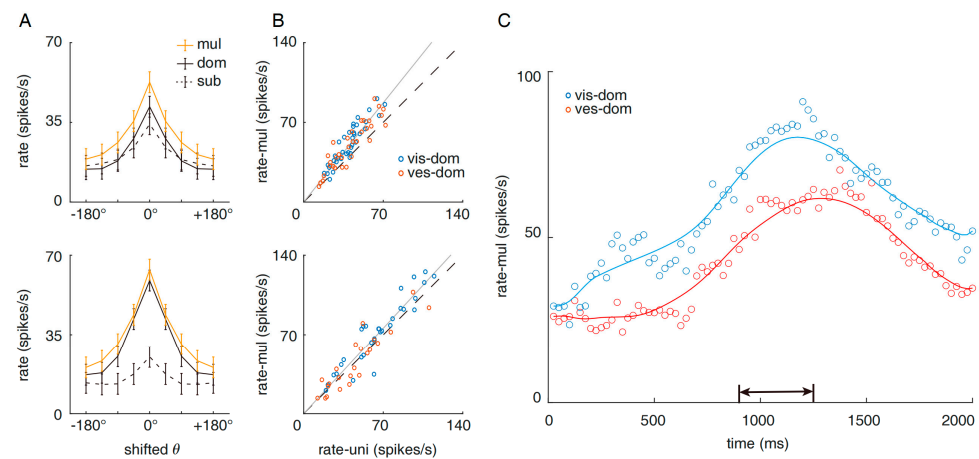
$$r = \frac{\max(R_{vis}^{max}, R_{ves}^{max})}{\min(R_{vis}^{max}, R_{ves}^{max})} \quad (7)$$

where  $r$  is the response ratio, and  $R_{vis}^{max}$  and  $R_{ves}^{max}$  are the maximal values of preferred tuning curves of MST neurons to the visual and vestibular uni-sensory inputs.

$$R_{vis}^{max} = \max[f_0(\theta_{vis})] \quad (8)$$

$$R_{ves}^{max} = \max[f_0(\theta_{ves})] \quad (9)$$

where  $\theta_{vis}$  and  $\theta_{ves}$  represent the visual and vestibular input directions respectively, and  $f_0$  is the neuronal spatial response function. As defined, it is obvious that the balanced neurons respond to both modalities with approximately equal amplitude (1~1.6) and the imbalanced neurons are characterized by the dominance of one modality, i.e.,  $r \geq 1.6$  (Figure 1A; note that the maximal responses are shifted to align at 0 degrees).



**Figure 1.** Visual and vestibular response distinction in balanced and imbalanced MST-d neurons. (A) Averaged tuning curves of balanced (top) and imbalanced (bottom) neurons in the monkey brain dorsal medial superior temporal area (MST-d). Black solid lines denote the curves corresponding to the modality of  $\max(R_{vis}^{max}, R_{ves}^{max})$ , and black dashed lines denote the curves corresponding to the modality of  $\min(R_{vis}^{max}, R_{ves}^{max})$ . Gold solid lines denote the multisensory curve, which is direction-averaged and modality-averaged. Error bars denote the mean standard error. (B) Comparison of the unisensory and multisensory maximal fire-rate amplitudes of balanced (top) and imbalanced (bottom) neurons. Blue: visual-dominant ( $R_{vis}^{max} > R_{ves}^{max}$ ). Red: vestibular-dominant ( $R_{vis}^{max} < R_{ves}^{max}$ ). Each circle denotes a neuron. The dominance is determined by the unisensory responses. (C) Mean dynamics in the time domain for MST-d visual-dominant (blue) and vestibular-dominant (red) neurons. Each circle denotes the averaged response of corresponding neuron group at a particular time step. The area indicated by the two-headed arrow is the time window in which the averaged responses are computed.

We further examined the multisensory condition and then compared the responses of these two types of neurons with the unisensory condition. When both stimuli are presented

simultaneously as a multisensory condition, the response is a function of both the visual and vestibular directions,

$$R_{multi} = f_{multi}(\theta_{vis}, \theta_{ves}) \quad (10)$$

Intuitively, the responses formed a two-dimensional grid along the two modalities. To unify the dimension of the tuning curves, the multisensory responses were averaged based on the spatial relative distance from the single maximal response ( $R_{multi}^{max}$ ), and the averaged multisensory curves are presented in Figure 1A (gold lines). It is clear that the multisensory condition generally enhanced the responses of the balanced neurons ( $df = 70, p = 5.30 \times 10^{-16}$ , paired  $t$ -test, mean response gain = 26%), but the response from the imbalanced neurons was less enhanced ( $df = 55, p = 0.018$ , paired  $t$ -test, mean response gain = 8%).

Figure 1B demonstrates  $\max(R_{vis}^{max}, R_{ves}^{max})$  versus  $R_{multi}^{max}$  in each neuron (top: balanced neurons, bottom: imbalanced neurons). Neurons with visual and vestibular unisensory dominance are presented with different colors. In balanced neurons, the visual and vestibular response distributions were not discriminated in either condition, suggesting that the visual and vestibular modalities share equal weights or contributions (Uni-sensory:  $p = 0.79, df = 67$ . Multisensory:  $p = 0.70, df = 67$ , two-sample  $t$ -test). However, the imbalanced neurons showed clear discrimination: both the unisensory and multisensory maximal responses of visual-dominant neurons ( $R_{vis}^{max} > R_{ves}^{max}$ ) exceeded those of vestibular-dominant neurons ( $R_{ves}^{max} > R_{vis}^{max}$ ) (Unisensory:  $p = 0.0086, df = 52$ . Multisensory:  $p = 0.0017, df = 52$ , two-sample  $t$ -test). Figure 1C shows in the time domain that the visual-dominant neurons generally respond more strongly than the vestibular-dominant neurons after stimulus onset (time = 0 s), suggesting an overall dominance of the visual modality in MST-d imbalanced neurons. Assuming that the response amplitude indicates modality weight in the psychophysical process, the visual information thus has greater potential to outweigh the vestibular information during the process of neural decision-making and alter the decision-making to present a bias towards visual information. In the next section, we validate this hypothesis by simulating decision-making with a biophysical model.

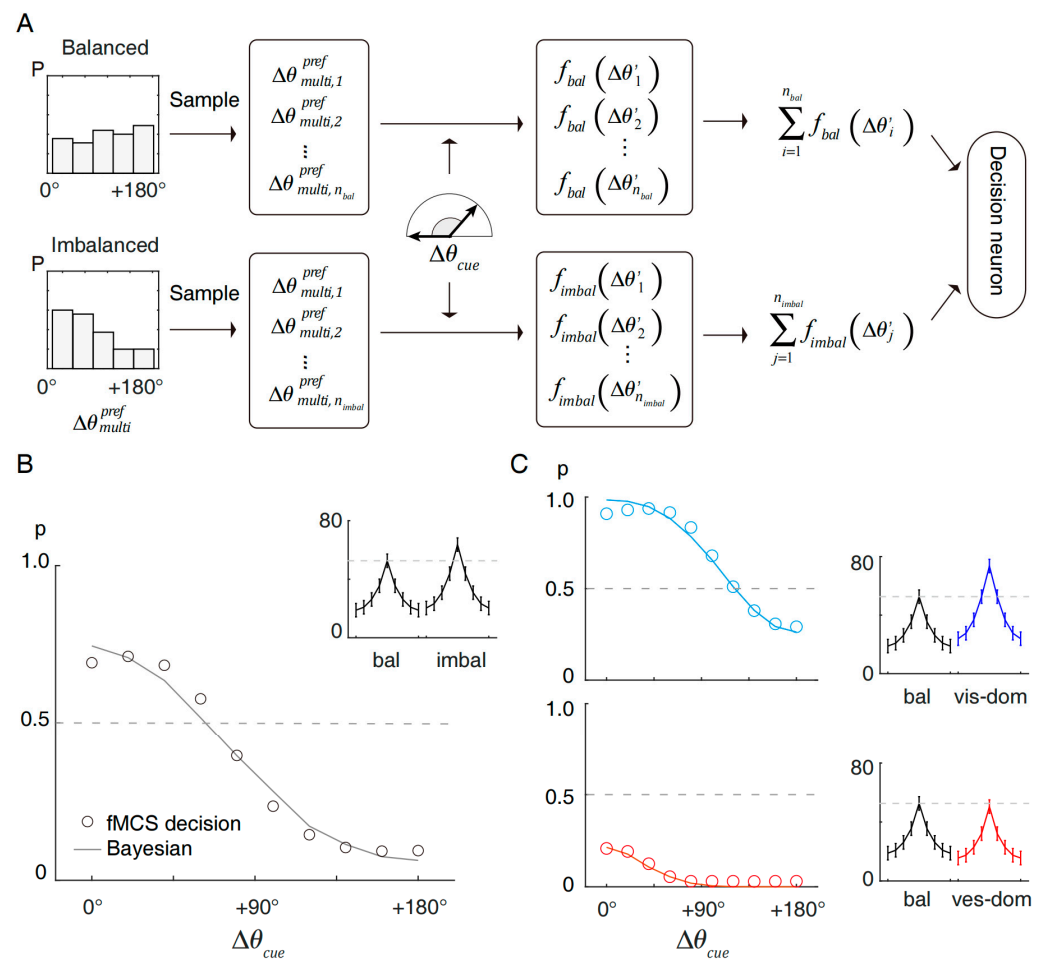
### 3.2. Balanced and Imbalanced MST-d Neurons Comprise Encoding Bases

Our model is a functional Monte-Carlo sampling (fMCS) model, using a plausible biophysical process to determine whether to integrate or separate the visual and vestibular external cues based on the balanced and imbalanced neuronal responses. As presented in Figure 2A, the fMCS model is mainly composed of a sampling module and a decision module (one decision neuron). The sampling module randomly chooses  $n_{bal}$  and  $n_{imbal}$  neurons with a preference according to the probability distribution, which is observed from the data (Figure 2A left). Approximated to the data observations ( $n_{bal} = 71, n_{imbal} = 56$ ),  $n_{bal} : n_{imbal}$  in the MCS model is set to 9:7 (72:56). To find a minimal requirement for reaching effective decisions, we set the numbers as 9 and 7. In previous works, we discussed that when the ratio does not meet this standard (for example, 1:1 or 1:2), the prior will deviate from 0.5 [24].

The neuronal preference determines when the neuron responds maximally (the peak response shown at  $0^\circ$  in Figure 1A) and characterizes how the neuron responds to specific external inputs. For simplicity, only the preference for disparity ( $\Delta\theta_{multi}^{pref}$ ) is considered, and the absolute direction is omitted.

$$\left[ \theta_{vis,multi}^{pref}, \theta_{ves,multi}^{pref} \right] = \operatorname{argmax}(R_{multi}) \quad (11)$$

$$\Delta\theta_{multi}^{pref} = \min\left( \left| \theta_{vis,multi}^{pref} - \theta_{ves,multi}^{pref} \right|, 360^\circ - \left| \theta_{vis,multi}^{pref} - \theta_{ves,multi}^{pref} \right| \right) \quad (12)$$



**Figure 2.** Functional Monte-Carlo sampling model that indicates the distinct roles of visual and vestibular neurons during decision-making. **(A)** Schematics of the MCS model for monkey MST-d. The model input is the stimulus disparity ( $\Delta\theta_{cue}$ ). The neuronal preference is sampled from each distribution of balanced and imbalanced groups and transformed into the fire rate by a response function ( $f$ ). The decision of integration or separation is reached by comparing the sampled group response from balanced and imbalanced neurons. **(B)** Probabilistic decision functions produced by the MCS model (circles) and fitted by the Bayesian model (solid curves). The decision probability  $p$  of whether to integrate the cues (reporting cues from the same source) is a function of the external cue disparity ( $\Delta\theta_{cue}$ ). The side panel demonstrates the averaged response functions in the simulation, which are shared by all sampled neurons. **(C)** The decision by the fMCS model with visual-dominant imbalanced neurons only (top) or vestibular-dominant imbalanced neurons only (bottom). The visual-dominant neuronal response curves are presented in the top side panel (blue curve) and the vestibular-dominant neuronal response curves are presented in the bottom side panel (red curve).

By definition,  $\Delta\theta_{multi}^{pref}$  ranges from  $0^\circ$  to  $180^\circ$ . The external inputs take the form of disparity as well, denoted as  $\Delta\theta_{cue}$  ( $0^\circ \leq \Delta\theta' \leq 180^\circ$ ). When  $\Delta\theta_{cue}$  is unaligned with  $\Delta\theta_{multi}^{pref}$ , the neuron response is weaker than the maximal response. In this case, the response is a function of the relative disparity  $\Delta\theta'$ ,

$$\Delta\theta' = \left| \Delta\theta_{cue} - \Delta\theta_{multi}^{pref} \right| \quad (13)$$

By averaging the data, the response under specific inputs is further quantified by the averaged multisensory tuning curves  $f_{bal}$  and  $f_{imbal}$ , which are also data derived (Figure 2B



side panel, same as the gold lines in Figure 1A).  $f_{bal}$  and  $f_{imbal}$  are direction-averaged and modality-averaged.

$$R_{bal} = f_{bal}(\Delta\theta) \quad (14)$$

$$R_{imbal} = f_{imbal}(\Delta\theta) \quad (15)$$

For example, if  $\Delta\theta_{cue}$  is  $30^\circ$ , the neurons with  $\Delta\theta_{multi}^{pref}=30^\circ$  show the maximal response because  $f(|\Delta\theta_{cue} - \Delta\theta_{multi}^{pref}|) = f(\Delta\theta = 0^\circ) = R_{multi}^{max}$ . However, the neurons with  $\Delta\theta_{multi}^{pref}=180^\circ$  have a low response because  $f(\Delta\theta = 150^\circ)$ , which assigns a low response according to the curve. Different neurons respond distinctively according to Equations (14) and (15) because of different preferences. Therefore, sampling preferences represent sampling the response of a particular MST-d neuron, which is attributed to the various neuronal activation thresholds in biophysical conditions.

After sampling, a hypothetical decision neuron (decision module) receives the summed response outputs of the balanced and imbalanced MST-d groups ( $R_{bal}^{tot}$  and  $R_{imbal}^{tot}$ ) and decodes the source nature ( $\emptyset$ ) by comparing response amplitudes. Note that the balanced neurons generally serve as the separation basis, while the imbalanced neurons generally serve as the integration basis [24]. The growing imbalance of the responses between the two modalities pushed the neuron to have a single representation of the external cues. If the balanced group responds more strongly, then the decision separates the modalities and attributes the inputs to separate sources ( $\emptyset(\text{common source}) = 0$ ). Otherwise, they are attributed to a common source ( $\emptyset(\text{common source}) = 1$ ). The decision-making process is described as follows.

$$R_{bal}^{tot} = \sum_{i=1}^{n_{bal}} R_{bal,i} \quad (16)$$

$$R_{imbal}^{tot} = \sum_{j=1}^{n_{imbal}} R_{imbal,j} \quad (17)$$

$$\begin{cases} \emptyset(\text{common source}) = 0 & \text{if } R_{bal}^{tot} > R_{imbal}^{tot} \\ \emptyset(\text{common source}) = 1 & \text{if } R_{bal}^{tot} \leq R_{imbal}^{tot} \end{cases} \quad (18)$$

$R_{bal,i}$  denotes the response from the  $i^{th}$  sampled balanced neuron (Equation (14)), and  $R_{imbal,j}$  denotes the  $j^{th}$  sampled imbalanced neuron (Equation (15)). In conclusion, the input to the fMCS model is the hyper-parameter  $\Delta\theta_{cue}$  ( $0^\circ \leq \Delta\theta_{cue} \leq 180^\circ$ ) representing the external cue disparity. The output of each decision trial is either integration or separation and the decisions are repeated for 100,000 trials to obtain the probability of reporting a common source ( $p_{int} = n_{int} / (n_{int} + n_{sep})$ ).

As presented in Figure 2B, the fMCS model generated decision functions with a reverse sigmoidal shape, which had a high similarity to the behavioral results [18], especially in that the decision was prone to integrate at a small cue disparity and separate at a large cue disparity. Critically, the decision in any cue disparity condition was not determined but probabilistic due to the sampling nature.

We compared the decision function with the classical Bayesian causal inference strategy [17]. The Bayesian strategy fit closely with the fMCS decisions and further revealed that the best-fit prior was 0.51. Since we stated previously that imbalanced neurons with visual dominance have a higher response than those with vestibular dominance, it is evident that visual and vestibular dominance contribute to decisions differently. In the form of an averaged curve (Figure 2C, side panel), the visual-dominant imbalanced neurons responded with 73.55 spikes/s on average, while the vestibular-dominant imbalanced neurons responded with 50.45 spikes/s. Compared with the averaged balanced neuronal response (52.61 spikes/s), the visual-dominant neurons had an advantage in triggering an integration decision, while the vestibular-dominant neurons were subordinate to the balanced neurons and were prone to resulting in a separation decision. This was further validated by the fMCS model, which showed that the integration probability ( $p$ ) was biased

toward 1 if the fMCS model only contained visual-dominant neurons in the imbalanced group (Figure 2C, top) but biased toward 0 with vestibular-dominant imbalanced neurons only (Figure 2C, bottom). The classical Bayesian strategy predicted that the former case would be matched with an integration prior of 0.83, while the latter would be matched with 0.34.

### 3.3. Discriminated Multisensory Tuning of Visual and Vestibular Modalities

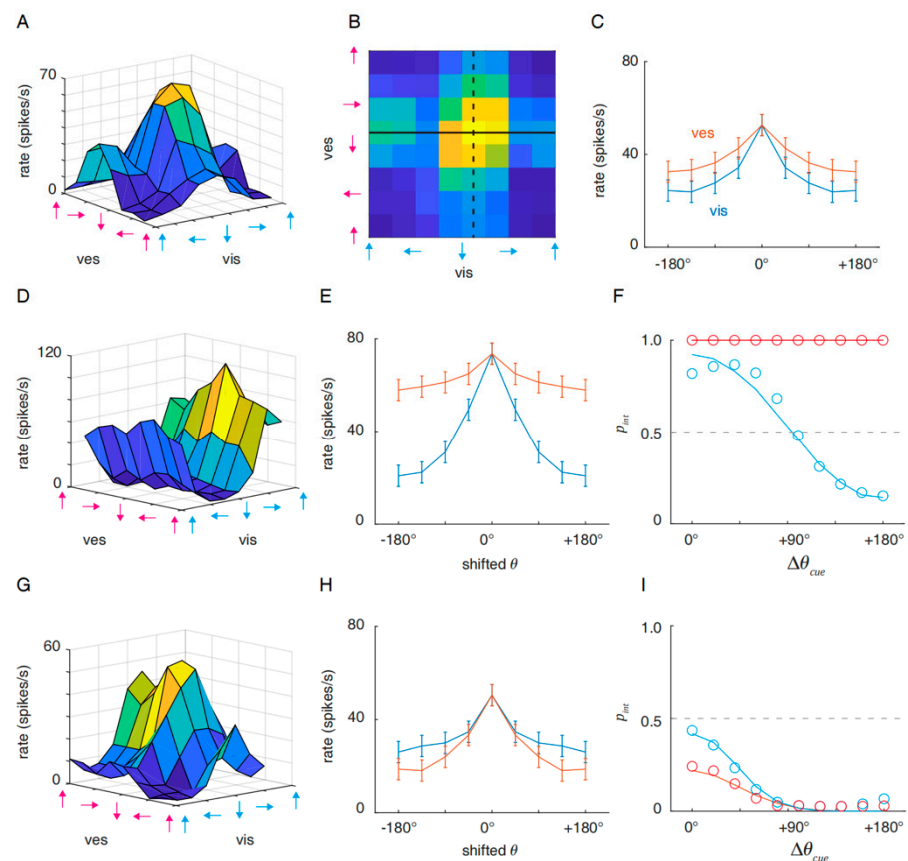
From the analysis above, we conclude that the response amplitude is crucial to the final decision. Despite the distinction between the visual and vestibular dominance of the imbalanced neurons, the response function of each neuron was modality-averaged in the previous section. However, in realistic multisensory tuning, this may not be the case. As proposed previously, multisensory tuning is a two-dimensional grid (exemplified in Figure 3A), one dimension of which is the visual input direction and the other of which is the vestibular direction. The tuning on both dimensions may have different slopes (gradients), which determine the neuronal response at non-preferred inputs. A steep tuning gradient indicates that the neuron is more sensitive to the change in one modality [1] and vice versa. To study the modality-specific gradients individually, we denote the multisensory visual tuning curve (Figure 3B, black solid line) as

$$R_{vis,multi} = f_{multi}(\theta_{vis,multi} | \theta_{ves,multi}^{pref}) \quad (19)$$

where  $\theta_{ves,multi}^{pref}$  is the preferred vestibular direction correlated with multisensory maximal response  $R_{multi}^{max}$  and  $\theta_{vis,multi}$  is the varying visual direction. Similarly, the multisensory vestibular curve (Figure 3B, black dashed line) is denoted as

$$R_{ves,multi} = f_{multi}(\theta_{ves,multi} | \theta_{vis,multi}^{pref}) \quad (20)$$

For each neuron, the multisensory visual or vestibular tuning curves intersect at  $R_{multi}^{max}$ . The responses at  $\theta_{multi}^{pref} \pm \theta$  ( $0^\circ \leq \theta \leq 180^\circ$ ) are averaged to obtain symmetric curves. The modality-specific tuning curves are further averaged across neurons in the same group. Figure 3C shows that the neurons were tuned more steeply to the visual modality than the vestibular modality in the balanced group, but the tuning gradients were proximate in general. In the imbalanced group, the visual-dominant neurons had a steep tuning gradient to the visual modality, while the responses to the vestibular modality were slightly tuned (Figure 3D,E). This suggests that, when the vestibular input deviates from the preference, the response remains constrained by the visual input. We simulated decision-making with the averaged balanced neurons as before but replaced the visual-dominant imbalanced neuronal tuning curve with the visual curve in Figure 3E. This is equivalent to simulating an external disparity ( $\Delta\theta_{cue}$ ) with a visual-direction change only. Figure 3F shows that the visual direction change altered the decision from integration to separation following the growing disparity with the vestibular direction. This suggests that the decoded tuning of the visual modality is effective in representing external cue changes. The classical Bayesian strategy determined the best-fit prior to be 0.64, which is relatively close to a flat prior. However, when replacing the tuning curve with the vestibular curve in Figure 3E, the decision is robust integration across the vestibular direction change, in which case the Bayesian prior is obviously 1.



**Figure 3.** Distinct multisensory visual and vestibular tuning affects the decision. **(A)** Multisensory tuning grid for a balanced neuron example. **(B)** The 2-dimensional contour of the response grid in A. The black solid line denotes  $f_{multi}(\theta_{vis,multi} | \theta_{ves,multi}^{pref})$ , which is the multisensory visual curve, and the black dashed line denotes  $f_{multi}(\theta_{ves,multi} | \theta_{vis,multi}^{pref})$ , which is the multisensory vestibular curve. The two curves intersect at  $R_{multi}^{max}$ . **(C)** Visual (blue) and vestibular (red) multisensory response curves for balanced neurons,  $n = 71$ . The curves are averaged to be symmetric. **(D)** Multisensory tuning grid for a visual-dominant imbalanced neuron example. **(E)** Averaged multisensory curve of visual-dominant imbalanced neurons,  $n = 32$ . **(F)** The decision inference determined by the fMCS model, with the averaged balanced curve as in Figure 2 and the visual (blue) and vestibular (red) curves of the visual-dominant imbalanced neurons. **(G–I)** Same as **(D–F)**, for vestibular-dominant imbalanced neurons.

Now let us focus on the vestibular-dominant imbalanced neurons. Compared with the visual-dominant imbalanced neurons, vestibular dominance did not lead to significant sharp vestibular tuning and blunt visual tuning; instead, the two curves presented close tuning gradients similar to those of the balanced neurons (Figure 3G–H), except that the vestibular tuning curve was slightly steeper than that of visual tuning. The decisions were simulated by an fMCS model similar to visual-dominance cases. Surprisingly, we found that such dominance did not produce an advantageous decision decoding to a vestibular direction change (Figure 3I). Rather, more flexible decisions were made with visual direction change again. The Bayesian strategy captured both modality changes with priors smaller than 0.5 (visual change: prior = 0.363, vestibular change: prior = 0.360). In conclusion, the vestibular modality was subordinate to the visual modality in MST-d both at the reference frame level and the effective decoding level.

#### 4. Discussion

This work demonstrated the computational role of distinct MST neuronal response properties to either visual or vestibular inputs in the decision by using a response-based computational framework.

We chose the functional Monte-Carlo sampling method in the framework because the sampling of response provides a concrete neuronal implementation to represent the randomization of perceiving the external world state. It was proposed that the probability of neuronal spiking is indeed computing the posterior probability of perception through response sampling [28–31]. As the main encoding components, different probability distributions of preference were observed empirically but the explicit mathematical description of the distribution was not accessible. By enough repetitions, the fMCS model carries out random sampling to approximate the probability density function, thus estimating the decision inference probability of the neural network. In our fMCS model, we sampled the neuronal response rate to represent the randomization of neuron activation as well as the perceived multisensory state of this neuronal unit.

The analysis from the Bayesian perspective predicted well the results from the fMCS model, indicating that the multisensory computation in the physical cortex implemented causal inference with a nearly flat prior. Decision-making with flat prior means that the mature MST-d circuit is neither biased towards integration nor separation, which is most efficient for the neural system to address unknown cues. This further indicated that matured MST-d balanced and imbalanced neurons reached an ‘equilibrium’ state in causal inference, which was achieved through the counterbalance between the neuron number and response amplitude, combined with the preference distributions. Since the fMCS model in this work utilized neuronal response data from matured monkeys, it characterized the local-cortical decision properties after synaptic learning.

Concerning the response amplitude, we present in Figure 1 that the relationship between unisensory and multisensory response matched with the additive effect found in previous works [25,32], and the impaired additive effect with growing response imbalance was predicted by the divisive normalization theory [33]. The division normalization is mainly attributed to intra-cortical shunting inhibition. Therefore, such passive modulation may not be the main site of synaptic learning discussed here. The learning process of multisensory inference is more likely to locate at the synaptic learning along the hierarchy [24,34], such as the connections from unisensory area MT (middle temporal) and PIVC (parietal insular vestibular cortex) to multisensory area MST-d, instead of the synaptic learning within the MST-d circuit. In other words, the learning is projected to unisensory responses rather than multisensory ones.

We further propose that effective decision decoding requires not only matching the maximal response amplitude, but also the response gradients. The latter is a measurement of response selectivity, which is crucial to shifting the decisions flexibly between integration and separation based on the external cue disparity.

In Figure 3, we demonstrate that the visual-dominant neurons respond selectively to the visual direction, but unselectively to the vestibular direction. As a result, the vestibular direction change did not affect the final decision and such rigid decoding has limited information capacity for vestibular-cue information. In addition, in vestibular-dominant neurons, the dominance did not produce an advantageous decision decoding to the vestibular direction change. We reason that vestibular-dominant imbalanced neurons generally respond at lower rates (mean maximal response is 50.45 spikes/s) than visual-dominant neurons (73.56 spikes/s) and balanced neurons (52.63 spikes/s). In this case, the weak tuning of the visual modality in turn enhanced the overall responses and pulled the decision closer to the equilibrium state.

The results indicated that the visual and vestibular modalities were weighed differently in MST-d, potentially leading to a biased reference frame for local decision-making [35]. The modality-specific neuronal response is the dynamic origin of reference frame selection, so the two modalities are prone to integration into a visual frame due to visual-response

dominance but less likely to integrate into a vestibular frame. Furthermore, effective decision encoding mostly emerged from a visual direction change, while a vestibular direction change produced either robust integration or robust separation, making the decision rigid and inflexible. The analysis is compatible with previous experimental findings that MST-d was dominated by the visual frame when performing sensory integration [35]. It was previously debated whether an accurate decision requires a common frame, intermediate, or even distinct frames [36,37]. This work proposes that, in MST-d, the visual and vestibular frames are pre-coded concurrently by visual and vestibular-dominant neurons. Due to the prevalence of visual signals in MST-d [25], the integration is mostly based on the visual frame. Nevertheless, we reason that the vestibular frame serves as a fundamental role during the decision, because it also provides a base coordinate about self-motion for the vestibular-related signal if the decision is to separate the cues.

## 5. Conclusions

This study has combined the methods of computational modeling and analysis of physiological recordings and examined the effect of functional heterogeneity between visual and vestibular modalities in MST-d. The heterogeneity stemmed from distinct tuning selectivity, which denotes two aspects: the response amplitude, and the tuning gradient (the tuning curve's steepness). Previously, related works mainly explored the multisensory causal inference on two levels: one is the mathematical principle of multisensory behaviors [12,17,18], and the other is biophysical computation modeling at a neuronal circuit level [1,2,8,33]. Nevertheless, few works have bridged the gap between the two levels and studied the decision property stemming from neuronal architecture. This work initialized from the neuronal response and cortical processing hierarchy and reproduced decisions with Bayesian strategy. Importantly, we reported that the modality which produces larger neuronal response amplitude and steeper tuning is more likely to have a larger weight in multisensory decisions. Furthermore, we predicted that the response dominance of one modality determined the base coordinate to set up the reference frame for motion detection. In conclusion, the neuronal network response patterns compose the internal representation of external motion relationships, which may serve as a ubiquitous probabilistic-inference base for the merging of senses.

While the present work focused on the primate MST-d network, past studies demonstrated that both human and primate behavioral properties exhibited the Bayesian inference approach [1,12,34]. Despite the lack of physiological evidence of human subjects, it was reported that human brain contains the homolog of macaque MST area, which locates in human MT+ complex and compiles both visual and vestibular signals [38–40]. As Zhang et al. proposed [24], the macaque MST-d provided essential components for Bayesian population coding, thus future works may need to examine whether human follows similar principle to make a multisensory inference.

## 6. Limitations

As we seek to provide a general principle for biophysical sampling-based processes, the present work suffers from several limitations. It has been demonstrated that real decision processes in the brain also involve sampling of neuronal responses [31]. In MST-d, the neuronal response shows complex modulation of preferences and inputs, and the response decay may not be symmetrical when the inputs deviate from the preference of each input modality. Furthermore, the preference from unisensory to multisensory conditions may change, potentially causing a sharper decrease in neural responses if one input is unreliable. Both conditions add more variability to the decision process, which induces more flexibility. A high-dimensional sampling process would be required to include this variability.

**Author Contributions:** Y.Y. and A.C. supervised the research, Y.Y., A.C., Y.G. and J.Z. designed the research, J.Z., M.H. performed the research, J.Z., M.H. wrote the analysis tools and analyzed the data, J.Z. and Y.Y. wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project is supported by the Science and Technology Innovation 2030—Brain Science and Brain-Inspired Intelligence Project (2021ZD0200204, 2021ZD0202600 and 2021ZD0201301), the National Natural Science Foundation of China (U20A20221), the Shanghai Municipal Science and Technology Major Project (2018SHZDZX01 and 2021SHZDZX0103) and ZJLab, Shanghai Municipal Science and Technology Committee of Shanghai outstanding academic leaders plan (21XD1400400).

**Institutional Review Board Statement:** The Institutional Animal Care and Use Committee at Washington University approved all animal surgeries and experimental procedures, which were performed following National Institutes of Health guidelines.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The datasets analyzed in this study are available from the corresponding author for reasonable request. The key analysis codes are available for reasonable request.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## References

- Gu, Y.; Angelaki, D.E.; DeAngelis, G.C. Neural correlates of multisensory cue integration in macaque MSTd. *Nat. Neurosci.* **2008**, *11*, 1201–1210. [[CrossRef](#)] [[PubMed](#)]
- Fetsch, C.R.; Pouget, A.; DeAngelis, G.C.; Angelaki, D.E. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat. Neurosci.* **2011**, *15*, 146–154. [[CrossRef](#)] [[PubMed](#)]
- Binns, K.E.; Salt, T.E. Importance of NMDA receptors for multimodal integration in the deep layers of the cat superior colliculus. *J. Neurophysiol.* **1996**, *75*, 920–930. [[CrossRef](#)]
- Meredith, M.A. On the neuronal basis for multisensory convergence: A brief overview. *Cogn. Brain Res.* **2002**, *14*, 31–40. [[CrossRef](#)]
- Meredith, M.A.; Stein, B.E. Spatial determinants of multisensory integration in cat superior colliculus neurons. *J. Neurophysiol.* **1996**, *75*, 1843–1857. [[CrossRef](#)] [[PubMed](#)]
- Parise, C.V.; Ernst, M.O. Correlation detection as a general mechanism for multisensory integration. *Nat. Commun.* **2016**, *7*, 11543. [[CrossRef](#)]
- Truskowski, T.L.S.; Carrillo, O.A.; Bleier, J.; Ramirez-Vizcarrondo, C.M.; Felch, D.L.; McQuillan, M.; Truskowski, C.P.; Khakhalin, A.S.; Aizenman, C.D. A cellular mechanism for inverse effectiveness in multisensory integration. *Elife* **2017**, *6*, e25392. [[CrossRef](#)] [[PubMed](#)]
- Zhang, W.H.; Wang, H.; Chen, A.; Gu, Y.; Lee, T.S.; Wong, K.M.; Wu, S. Complementary congruent and opposite neurons achieve concurrent multisensory integration and segregation. *Elife* **2019**, *8*, e43753. [[CrossRef](#)] [[PubMed](#)]
- Driver, J.; Spence, C. Multisensory perception: Beyond modularity and convergence. *Curr. Biol.* **2000**, *10*, R731–R735. [[CrossRef](#)]
- Schroeder, C.E.; Foxe, J. Multisensory contributions to low-level, ‘unisensory’ processing. *Curr. Opin. Neurobiol.* **2005**, *15*, 454–458. [[CrossRef](#)]
- Kaysers, C.; Logothetis, N.K. Do early sensory cortices integrate cross-modal information? *Brain Struct. Funct.* **2007**, *212*, 121–132. [[CrossRef](#)] [[PubMed](#)]
- Knill, D.C.; Pouget, A. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci.* **2004**, *27*, 712–719. [[CrossRef](#)] [[PubMed](#)]
- Ma, W.J.; Beck, J.M.; Latham, P.E.; Pouget, A. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **2006**, *9*, 1432–1438. [[CrossRef](#)] [[PubMed](#)]
- Pouget, A.; Beck, J.M.; Ma, W.J.; Latham, P.E. Probabilistic brains: Knowns and unknowns. *Nat. Neurosci.* **2013**, *16*, 1170–1178. [[CrossRef](#)]
- Kording, K.P.; Wolpert, D.M. Bayesian integration in sensorimotor learning. *Nature* **2004**, *427*, 244–247. [[CrossRef](#)]
- Wallace, M.T.; Roberson, G.; Hairston, W.D.; Stein, B.E.; Vaughan, J.W.; Schirillo, J.A. Unifying multisensory signals across time and space. *Exp. Brain Res.* **2004**, *158*, 252–258. [[CrossRef](#)]
- Kording, K.P.; Beierholm, U.; Ma, W.J.; Quartz, S.; Tenenbaum, J.B.; Shams, L. Causal inference in multisensory perception. *PLoS ONE* **2007**, *2*, e943. [[CrossRef](#)]
- Acerbi, L.; Dokka, K.; Angelaki, D.E.; Ma, W.J. Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. *PLoS Comput. Biol.* **2018**, *14*, e1006110. [[CrossRef](#)]
- Duffy, C.J.; Wurtz, R.H. Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *J. Neurosci.* **1995**, *15*, 5192–5208. [[CrossRef](#)]
- Duffy, C.J. MST neurons respond to optic flow and translational movement. *J. Neurophysiol.* **1998**, *80*, 1816–1827. [[CrossRef](#)]

21. Bremmer, F.; Kubischik, M.; Pekel, M.; Lappe, M.; Hoffmann, K.P. Linear vestibular self-motion signals in monkey medial superior temporal area. *Ann. N. Y. Acad. Sci.* **1999**, *871*, 272–281. [[CrossRef](#)] [[PubMed](#)]
22. Tanaka, K.; Fukada, Y.; Saito, H. Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **1989**, *62*, 642–656. [[CrossRef](#)] [[PubMed](#)]
23. Rideaux, R.; Storrs, K.R.; Maiello, G.; Welchman, A.E. How multisensory neurons solve causal inference. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2106235118. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, J.; Gu, Y.; Chen, A.; Yu, Y. Unveiling Dynamic System Strategies for Multisensory Processing: From Neuronal Fixed-Criterion Integration to Population Bayesian Inference. *Research* **2022**, *2022*, 9787040. [[CrossRef](#)] [[PubMed](#)]
25. Gu, Y.; Watkins, P.V.; Angelaki, D.E.; DeAngelis, G.C. Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J. Neurosci. Off. J. Soc. Neurosci.* **2006**, *26*, 73–85. [[CrossRef](#)]
26. Shulman, G.L.; Corbetta, M.; Buckner, R.L.; Raichle, M.E.; Fiez, J.A.; Miezin, F.M.; Petersen, S.E. Top-down modulation of early sensory cortex. *Cereb Cortex* **1997**, *7*, 193–206. [[CrossRef](#)]
27. Pessoa, L.; Kastner, S.; Ungerleider, L.G. Neuroimaging studies of attention: From modulation of sensory processing to top-down control. *J. Neurosci.* **2003**, *23*, 3990–3998. [[CrossRef](#)]
28. Hoyer, P.; Hyvärinen, A. Interpreting neural response variability as Monte Carlo sampling of the posterior. *Adv. Neural Inf. Process. Syst.* **2002**, *15*.
29. Buesing, L.; Bill, J.; Nessler, B.; Maass, W. Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons. *PLoS Comput. Biol.* **2011**, *7*, e1002211. [[CrossRef](#)]
30. Huang, Y.; Rao, R.P. Neurons as Monte Carlo Samplers: Bayesian Inference and Learning in Spiking Networks. *Adv. Neural Inf. Process. Syst.* **2014**, *27*.
31. Zhang, W.; Wu, S.; Josic, K.; Doiron, B. Sampling-based Bayesian inference in recurrent circuits of stochastic spiking neurons. *bioRxiv* **2022**. [[CrossRef](#)]
32. Ferrè, E.R.; Walther, L.E.; Haggard, P. Multisensory interactions between vestibular, visual and somatosensory signals. *PLoS ONE* **2015**, *10*, e0124573. [[CrossRef](#)] [[PubMed](#)]
33. Ohshiro, T.; Angelaki, D.E.; DeAngelis, G.C. A normalization model of multisensory integration. *Nat. Neurosci.* **2011**, *14*, 775–782. [[CrossRef](#)] [[PubMed](#)]
34. Rohe, T.; Noppeney, U. Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol.* **2015**, *13*, e1002073. [[CrossRef](#)]
35. Yang, L.; Gu, Y. Distinct spatial coordinate of visual and vestibular heading signals in macaque FEFsem and MSTd. *Elife* **2017**, *6*, e29809. [[CrossRef](#)]
36. Groh, J.M. Converting neural signals from place codes to rate codes. *Biol. Cybern.* **2001**, *85*, 159–165. [[CrossRef](#)]
37. Deneve, S.; Latham, P.E.; Pouget, A. Efficient computation and cue integration with noisy population codes. *Nat. Neurosci.* **2001**, *4*, 826–831. [[CrossRef](#)]
38. Vaina, L.M.; Cowey, A.; Eskew, R.T., Jr.; LeMay, M.; Kemper, T. Regional cerebral correlates of global motion perception: Evidence from unilateral cerebral brain damage. *Brain* **2001**, *124*, 310–321. [[CrossRef](#)]
39. Huk, A.C.; Dougherty, R.F.; Heeger, D.J. Retinotopy and functional subdivision of human areas MT and MST. *J. Neurosci.* **2002**, *22*, 7195–7205. [[CrossRef](#)]
40. Bottini, G.; Karnath, H.-O.; Vallar, G.; Sterzi, R.; Frith, C.D.; Frackowiak, R.S.; Paulesu, E. Cerebral representations for egocentric space: Functional-anatomical evidence from caloric vestibular stimulation and neck vibration. *Brain* **2001**, *124*, 1182–1196. [[CrossRef](#)]