

# Diverse effects of gaze direction on heading perception in humans

Wei Gao<sup>1,†</sup>, Yipeng Lin<sup>1,†</sup>, Jiangrong Shen<sup>2</sup>, Jianing Han<sup>2</sup>, Xiaoxiao Song<sup>3</sup>, Yukun Lu<sup>1</sup>, Huijia Zhan<sup>1</sup>, Qianbing Li<sup>1</sup>, Haoting Ge<sup>1</sup>, Zheng Lin<sup>6</sup>, Wenlei Shi<sup>4</sup>, Jan Drugowitsch<sup>5</sup>, Huajin Tang<sup>2</sup>, Xiaodong Chen<sup>1,\*</sup>

<sup>1</sup>Department of Neurology and Psychiatry of the Second Affiliated Hospital, College of Biomedical Engineering and Instrument Science, Interdisciplinary Institute of Neuroscience and Technology, School of Medicine, Zhejiang University, 268 Kaixuan Road, Jianggan District, Hangzhou 310029, China,

<sup>2</sup>College of Computer Science and Technology, Zhejiang University, 38 Zheda Road, Xihu District, Hangzhou 310027, China,

<sup>3</sup>Department of Liberal Arts, School of Art Administration and Education, China Academy of Art, 218 Nanshan Road, Shangcheng District, Hangzhou 310002, China,

<sup>4</sup>Center for the Study of the History of Chinese Language and Center for the Study of Language and Cognition, Zhejiang University, 866 Yuhangtang Road, Xihu District, Hangzhou 310058, China,

<sup>5</sup>Department of Neurobiology, Harvard Medical School, Longwood Avenue 220, Boston, MA 02116, United States,

<sup>6</sup>Department of Psychiatry, Second Affiliated Hospital, School of Medicine, Zhejiang University, 88 Jiefang Road, Shangcheng District, Hangzhou 310009, China

\*Corresponding author: Room 301, Kexue Building, Huajiachi Campus, Zhejiang University, Hangzhou, Zhejiang, 310029, China. Email: [xdcrun@zju.edu.cn](mailto:xdcrun@zju.edu.cn)

<sup>†</sup>Wei Gao and Yipeng Lin contributed equally to this work.

Gaze change can misalign spatial reference frames encoding visual and vestibular signals in cortex, which may affect the heading discrimination. Here, by systematically manipulating the eye-in-head and head-on-body positions to change the gaze direction of subjects, the performance of heading discrimination was tested with visual, vestibular, and combined stimuli in a reaction-time task in which the reaction time is under the control of subjects. We found the gaze change induced substantial biases in perceived heading, increased the threshold of discrimination and reaction time of subjects in all stimulus conditions. For the visual stimulus, the gaze effects were induced by changing the eye-in-world position, and the perceived heading was biased in the opposite direction of gaze. In contrast, the vestibular gaze effects were induced by changing the eye-in-head position, and the perceived heading was biased in the same direction of gaze. Although the bias was reduced when the visual and vestibular stimuli were combined, integration of the 2 signals substantially deviated from predictions of an extended diffusion model that accumulates evidence optimally over time and across sensory modalities. These findings reveal diverse gaze effects on the heading discrimination and emphasize that the transformation of spatial reference frames may underlie the effects.

**Key words:** gaze; visual; vestibular; reference frame; diffusion model.

## Introduction

Accurate heading discrimination relative to the body is important for navigation in the environment. Vestibular and visual signals input to the brain provide us with the most important heading information. However, they could dynamically change relative to the body when the eye-in-head (EIH) and/or head-on-body (HOB) positions vary (Weyand and Malpeli 1993; Bremmer et al. 1997; Li and Guo 1997; Shaikh et al. 2004; Hartmann et al. 2011; Ong and Bisley 2011; Chen et al. 2013a), as the vestibular and visual heading signals originating from the retina and otolith organs, respectively are initially coded in eye- and head-centered reference frames. Therefore, they should be first transformed from their original into the body-centered reference frame by combining the EIH and/or HOB signals before they could be used for discriminating heading (Cohen and Andersen 2002). Although many studies reported that both human and macaque subjects can integrate the 2 signals in a statistically optimal way in the brain to facilitate heading perception (Gu et al. 2008; Fetsch et al. 2009; Butler et al. 2011; Fetsch et al. 2011; Gu et al. 2012; Prsa et al. 2012; de Winkel et al. 2013; Chen et al. 2013b; Prsa et al. 2015), the eye-, head-, body-, and world-centered spatial reference frames were aligned in those experiments. Thus, it is still unclear how the change of EIH and/or HOB positions affects the heading discrimination because of the potential spatial reference frame

transformation, and whether this change causes the integration of visual and vestibular signals to deviate from the optimal model.

Physiologically, past work has revealed transformation of reference frames during the heading information processing in the brain. For example, the cortical representation of vestibular heading signals is diverse and flexible: Intermediate between eye- and head-centered in the medial superior temporal area (MST, Fetsch et al. 2007; Chen et al. 2013c), intermediate between head- and body-centered in the parietoinsular vestibular cortex (PIVC, Chen et al. 2013c) and body-centered in the ventral intraparietal area (VIP, Chen et al. 2013c, 2018), suggesting transformations from the original head-centered reference frame, even though the original eye-centered reference frame of visual heading signals is still maintained in the brain as the visual heading tuning of neurons in many cortical areas shifts with the EIH and/or HOB positions (Fetsch et al. 2007; Chen et al. 2013c, 2014; Fan et al. 2015; Yang and Gu 2017). Furthermore, the eye-in-world (EIW) gaze can dynamically shift the encoding reference frame of vestibular signals in VIP from body to world-centered (Chen et al. 2018). Sasaki et al. (2020) has reported that monkeys can switch reference frames from trial to trial between head- and world-centered in a task to discriminate the direction of a moving object during self-motion, and the responses of VIP neurons to visual and vestibular stimuli are significantly modulated by the task

Received: May 7, 2022. Revised: December 24, 2022. Accepted: December 27, 2022

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

reference frame used to compute the object direction. In addition, when visual and vestibular stimuli are presented simultaneously, the 2 signals could be integrated in MST (Gu et al. 2008; Fetsch et al. 2009, 2011) and VIP (Chen et al. 2013a) and the encoding reference frame for the combined signal is transformed and varies as a function of the relative strength of visual and vestibular tuning (Fetsch et al. 2007). However, it is still unknown how the spatial reference frame transformation in the brain finally affects the heading discrimination.

Psychophysically, there are several works studying gaze effects on the heading discrimination. By varying EIH but keeping HOB and body-in-world (BIW) positions constant, Crane (2017) has shown the perceived vestibular and visual heading are biased in the congruent and opposite directions of EIH, respectively, and the bias depends on the relative reliability of visual and vestibular stimuli in the combined condition. By varying both EIH and HOB positions but keeping BIW constant, Ni et al. (2013) has reported the perceived vestibular heading is biased in the congruent direction of EIH or HOB, and Crane (2015) studied the gaze effects on the heading estimation, showing that the gaze biases heading in the visual and combined stimulus conditions toward retina coordinates but do not influence the vestibular heading. However, Crane (2015) used a large sampling step ( $5^\circ$ ) of heading directions prevents it from studying the gaze effects with high precision, especially in the straightforward direction, where subjects can discriminate differences of visual heading that is  $<2^\circ$  (Lappe et al. 1999). And the potential reference frame transformation was not considered specifically in those experiments, for example, to model the heading perception by including the reaction time of subjects or incorporating signals related to EIH and HOB positions of subjects. Therefore, systematically studying gaze effects on the heading discrimination is still needed.

We hypothesize that noise and bias in the heading signals and potential reference frame transformation could affect the heading discrimination, including deviations of the integration of visual and vestibular signals from the optimal modal, increasing reaction time of subjects. Here, the gaze direction of subjects was systematically changed by manipulating the (EIH and HOB) position while keeping the BIW direction straight ahead, and subjects were instructed to report their perceived heading relative to BIW as soon as possible after the stimulus onset. In each gaze direction, the performance of heading discrimination was tested with visual, vestibular, and combined stimuli. Our results showed that both heading accuracy and precision were reduced by the gaze change, and the reaction time of subjects was increased. The visual heading was biased in the opposite direction of EIW gaze, but the vestibular heading was biased in the direction of EIH gaze. Although the bias was reduced when the visual and vestibular signals were combined, integration of the 2 signals was not effectively explained by an extended diffusion model that accumulates evidence optimally over time and across the 2 signals, and trades off accuracy with speed in the decision of subjects.

## Materials and methods

### Subjects

Nine healthy subjects (22–28 years) at least finished each experimental condition in the study. All were right-handed, had normal or corrected-to-normal vision, and no history of musculoskeletal, vestibular, or neurological disorders. Subjects were informed of

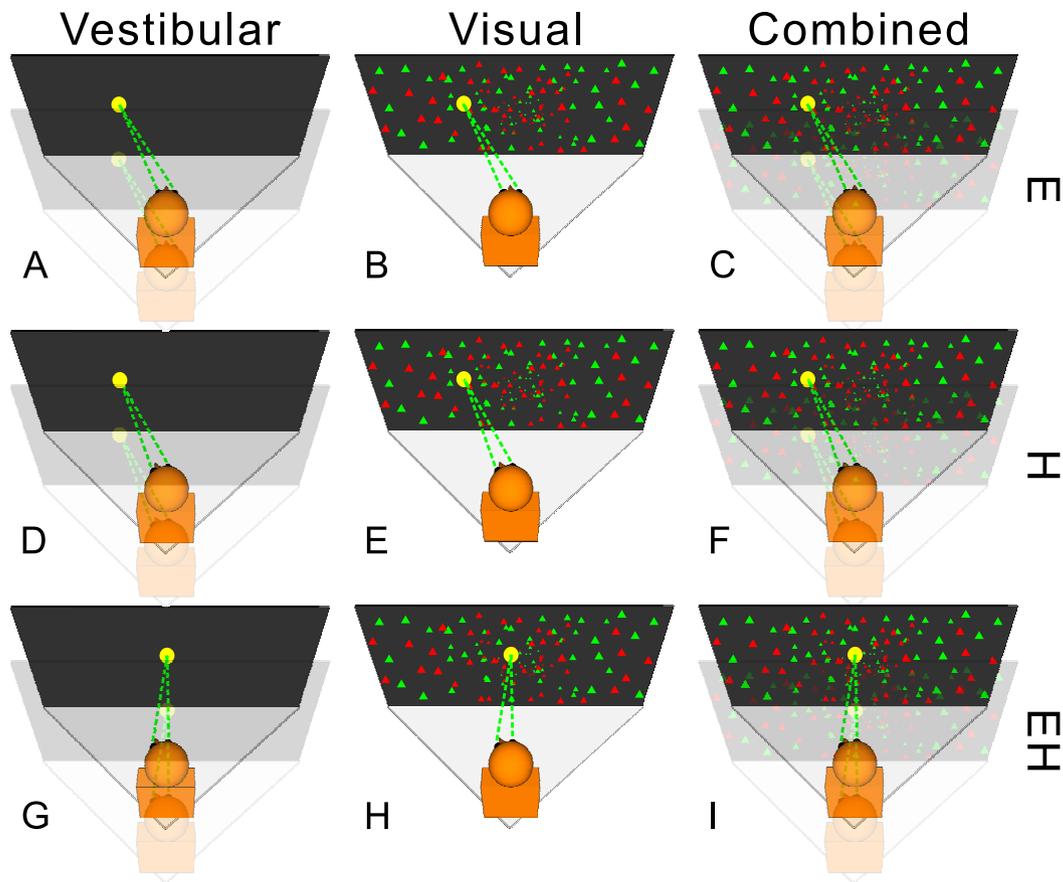
the experimental procedures, and informed written consent was obtained as per the guidelines of the Institutional Review Board.

### Apparatus

Subjects were seated comfortably in a padded racing car seat mounted on a 6 degree-of-freedom motion platform (MB-E-6DOF2000E, Moog). A 5-point safety harness held the subject's body securely in place. A subject-fitted thermoplastic mesh mask secured the head against a metal plate with a spindle shaft fixed on back of the seat, thereby immobilizing the head of subject relative to the platform during movement. A stop pin can be engaged to prevent rotation of the spindle shaft and keep the body direction straight ahead. When the stop pin was removed, the head was free to rotate in the horizontal plane (yaw rotation around the center of the head-body axis) with a maximum amplitude of  $20^\circ$ . A 48-inch LED monitor (SAMSUNG DH48E, 60 Hz, 105.4 cm  $\times$  59.3 cm, 1,440  $\times$  900 pixels) was fixed on top of the platform, 81 cm in front of the subject. Goggles with Wratten filters (Kodak, red no. 29 and green no. 61 for left and right eyes, respectively) were wore to provide stereoscopic depth cues, when subjects watched the visual stimuli on the monitor screen with a  $66^\circ \times 40^\circ$  field of view. Eye positions were tracked at 1,000 Hz (Eyelink 1000 plus, SR Research). The eye tracker was firmly secured to the motion platform by four stainless-steel screws to reduce the vibration impact during movement. As components on top of the platform were all covered within a frame with black nonreflective material, subjects' field of view was restricted to the screen and local environment within the frame. Together, the platform and frame defined a world reference frame, and movements of the platform translated the entire frame relative to the world. Sounds from the platform were masked by playing white noise through headphones. Behavioral responses were collected using 2 buttons (left and right) on a box.

### Stimuli

Subjects were translated by the motion platform to perceive vestibular stimulation. The translation followed a Gaussian velocity profile (duration = 2 s; displacement = 17.7 cm; peak velocity = 0.28 m/s; peak acceleration = 0.68 m/s<sup>2</sup>) and was limited in the horizontal plane. Twelve azimuth directions were tested ( $\pm 0.2^\circ$ ,  $\pm 1.0^\circ$ ,  $\pm 2.0^\circ$ ,  $\pm 4.0^\circ$ ,  $\pm 8.0^\circ$ , and  $\pm 16.0^\circ$ ), where  $0^\circ$  corresponded to straightforward direction in the world, and plus and minus corresponded to the left and right of trunk. Visual optic flow stimuli simulated self-motion through a 3D anaglyphic random dots field (150-cm wide, 150-cm tall, and 100-cm deep) that extended 50 cm from the screen center forward and backward, which were programmed using the C++ plus OpenGL (Open Graphics Library) and generated using an OpenGL accelerator board (NVIDIA Quadro K2000). Dot density of the field was 0.01/cm<sup>3</sup>. All dots in the field moved coherently (100% coherence). Each dot was rendered as a red/green triangle with a side length of 0.15 cm. The simulated visual translation followed the same velocity profile and direction definition as in the vestibular stimuli. Yellow symbols ( $0.2^\circ \times 0.2^\circ$ ) were rendered for the fixation (sphere for eye, and cross for head) or choice (sphere) targets. Three heading stimulus conditions were used: (i) A vestibular condition (Fig. 1A, D, and G) in which only the vestibular stimulus was presented by translating the motion platform, whereas the visual optic flow was not displayed on the screen; (ii) a visual condition (Fig. 1B, E, and H) in which the platform remained stationary, whereas the optic flow was shown to simulate the translation of subjects; (iii) a combined condition



**Fig. 1.** Experimental design. Columns and rows show stimuli and gaze conditions, respectively. The gray triangle indicates the motion platform. Red and green dots displayed on the screen (black rectangle) indicate the optic flow stimulus. The platform together with screen moves from a previous position (transparent) to the current (opaque). The body/trunk direction of a human subject (brass object) remains constant relative to the platform and screen. However, the gaze is varied by manipulating the (EIH and HOB) position to form 3 conditions (Methods): Eye-varied (E, A–C), Head-varied (H, D–F) and Eye&Head-varied (EH, G–I). In the E condition, EIH varies but HOB remains constant, including 3 positions ( $[-20^\circ, 0^\circ]$ ,  $[0^\circ, 0^\circ]$ , and  $[20^\circ, 0^\circ]$ ). In the H condition, EIH remains constant but HOB varies ( $[0^\circ, -20^\circ]$ ,  $[0^\circ, 0^\circ]$ , and  $[0^\circ, 20^\circ]$ ). In the EH condition, EIH and HOB vary together ( $[20^\circ, -20^\circ]$ ,  $[0^\circ, 0^\circ]$ , and  $[-20^\circ, 20^\circ]$ ). The minus sign indicates the direction to the left. In each condition, only the first position is shown here.

(Fig. 1C, F, and I) in which both the vestibular and visual stimuli were presented congruently and synchronously.

In order to test the delay in the presentation of visual or vestibular stimulus and the lag in the synchronization when they were presented simultaneously in the combined condition (Gu et al. 2006), a validation procedure (Supporting Information, Kim et al. 2019) was performed on our system as soon as it was built before any experiments, and was run again after the current experiments. In brief, an accelerometer was secured on the platform to measure the real-time acceleration in the movement, and a phototransistor was attached on the monitor to detect the real-time refreshing of the visual stimulus (Supplementary Fig. S2, see online supplementary material for a color version of this figure). Analog signals from these 2 sensors and command signal from the GUI (graphical user interface) component were connected into the data acquisition component and sampled simultaneously. Although there were some delays in the presentation of visual ( $19.85 \pm 1.75$  ms) or vestibular ( $19.43 \pm 6.96$  ms) stimuli and lags ( $-0.35 \pm 7.22$  ms) in the synchronization between them in our system (Supplementary Fig. S3, see online supplementary material for a color version of this figure), considering the relative long duration (2,000 ms) of stimuli in our experiments, such delays should not have significant effects on the results.

## Experimental protocol

Gaze directions were manipulated by varying the (EIH and HOB) position while keeping BIW direction straight ahead to include 3 conditions: Eye-varied (E), Head-varied (H), and Eye&Head-varied (EH). Display duration of the eye fixation target was manipulated while keeping EIH, HOB, and BIW aligned to the straightforward direction to include an additional eye target (ET) condition. For each subject, the ET condition was performed first, then E, and finally H and EH conditions that were randomly interleaved. In each gaze condition, the heading discrimination performance of subjects was tested for each stimulus condition. All the experimental protocols were implemented and controlled by customized software.

### Eye-varied gaze condition

The stop pin was set in place to stop head rotation and make HOB and BIW fixed and aligned (Fig. 1A–C). The beginning of each trial was signaled with an audible tone (150 ms), and then an eye fixation target appeared at 1 of 3 positions on the screen:  $-20^\circ$  (left),  $0^\circ$  (straight ahead), or  $+20^\circ$  (right). The subject was required to fixate the target within a  $5^\circ \times 5^\circ$  rectangular window for 500 ms before stimulus onset and to maintain the fixation throughout the following 2-s stimulus duration. The trial ended without

feedback after the subject pressed the button to report the perceived heading direction. Therefore, 3 (EIH and HOB) positions ( $[-20^\circ, 0^\circ]$ ,  $[0^\circ, 0^\circ]$ , and  $[20^\circ, 0^\circ]$ ) were included and 8 directions ( $\pm 0.2^\circ$ ,  $\pm 1.0^\circ$ ,  $\pm 8.0^\circ$ , and  $\pm 16.0^\circ$ ) were tested. There were 2,160 trials (8 directions  $\times$  30 repeats  $\times$  3 positions  $\times$  3 stimuli) at least, which were divided into 2 or 3 blocks and finished in separate days. Trials were randomly interleaved and intertrial interval was 5 s. The subject was set to take a short rest every 20 min. The whole 12 motion directions were used in the ET condition and pilot experiments for the gaze conditions, but the resulting total trials (3,240) were too much to be practical for all subjects. In addition, the response data were noisy. However, we noted the heading discrimination was very accurate when motion directions were  $>1^\circ$  for visual stimuli but reached a high accuracy until  $8^\circ$  for vestibular stimuli as showing in the ET condition (Fig. 6A–C). Therefore, in order to balance the precision and the signal noise ratio (SNR), 4 directions ( $\pm 2.0^\circ$  and  $\pm 4.0^\circ$ ) were not included but a high repeat number was used.

The subject was instructed to fixate the ET while paying attention to the heading direction during stimulation, and to judge whether they were moving leftward or rightward relative to the straightforward direction of his trunk midline. The subject was encouraged to do a best guess when not sure and report his perceived heading direction as soon as possible after the stimulus onset by pressing the corresponding button. If no button was pressed during the stimulation period, the ET disappeared and 2 choice targets were presented  $5^\circ$  to the left and right of the ET after the stimulus offset, informing the subject to do the report within the following 2 s. If the subject still did not respond during this period, the current trial was marked as failed and put back to the trial queue. The reaction time (RT) was defined as the interval from the stimulus onset to button pressing. Before the experiments began, subjects were required to do some practice (30–80 trials) to learn how to do the task and follow the protocol. Feedback (correct choice) was given at the end of each practice trial, and subjects were reminded to judge the heading direction relative to the trunk midline instead of the gaze direction.

Before each block of experiments, the Eyelink system was calibrated with a standard 9-point calibration and validation procedure for each subject, while viewing the points binocularly (de Haan et al. 2018). The eye position was tracked and monitored in real-time during the whole experiment, and data were recorded. If the subject kept breaking the fixation window or the judging accuracy in the last 12 trials was  $<50\%$ , the subject was instructed to focus attention and make fixation through a headphone or by showing a warning message on the screen (red color and flashing). In addition, online eye calibration was done by the drift correction function of Eyelink system if necessary.

### Head- and Eye&Head-varied gaze conditions

The stop pin was removed to free the head to rotate. Each trial started with a 150-ms tone, and then a head target was shown at 1 of 2 positions on the screen:  $-20^\circ$  (left),  $+20^\circ$  (right). The subject was required to rotate head in the corresponding direction until the rotation was stopped by the spindle shaft (reached the maximum amplitude of  $20^\circ$ ). And then, the subject pressed any button to signal the control software to show the ET at 1 of 3 positions on the screen:  $-20^\circ$ ,  $0^\circ$ , or  $+20^\circ$ . The subject was required to keep the head direction constant while fixating the ET, then press any button again to signal the control software to present stimuli. Both the head direction and eye fixation were required to be maintained for 500 ms before the stimulus onset and during

the following 2-s stimulus. Three (EIH and HOB) positions ( $[0^\circ, -20^\circ]$ ,  $[0^\circ, 0^\circ]$ , and  $[0^\circ, 20^\circ]$ ) were included in the H condition (Fig. 1D–F), and another 3 positions ( $[20^\circ, -20^\circ]$ ,  $[0^\circ, 0^\circ]$ , and  $[-20^\circ, 20^\circ]$ ) were included in the EH condition (Fig. 1G–I). Since  $[0^\circ, 0^\circ]$  appeared in all the 3 conditions and was already tested in the experiment set for E condition, it was not tested repeatedly in H and EH conditions. Therefore, there were 4 distinct positions ( $[0^\circ, -20^\circ]$ ,  $[20^\circ, -20^\circ]$ ,  $[-20^\circ, 20^\circ]$ , and  $[0^\circ, 20^\circ]$ ) and 2,880 trials (8 directions  $\times$  30 repeats  $\times$  4 positions  $\times$  3 stimuli) at least, which were divided into 3 or 4 blocks and finished in separate days. Trials were randomly interleaved and intertrial interval was 5 s. All the other settings, procedures, and instructions were same as those for the E condition. The head was fixed by setting the stop pin in place when calibrating the Eyelink system before each block of experiments but was free to rotate during experiments. Therefore, the approximate head direction and eye fixation direction were still be able to be tracked and monitored in real-time during the experiment, although not very accurate.

### ET condition

In addition to the conditions above designed to test the gaze effects, this condition was included to test the effect of eye fixation target as a possible reference for heading discrimination. The stop pin was engaged to make HOB and BIW fixed and aligned, and the ET was shown at the screen center ( $[0^\circ, 0^\circ]$ ) to align EIH with HOB and BIW. Display duration of the ET was manipulated to form 3 conditions: ET-ON, ET-CUE, and ET-OFF. In the ET-ON, after the 150-ms start tone, the ET was shown for 500 ms before stimulus onset and during the following 2-s stimulation period. Subjects were required to fixate the ET during the whole trial. In the ET-CUE, the ET was shown for only 500 ms before the stimulus onset and then disappeared, but subjects were still required to maintain the fixation during the stimulation period even if the ET was not shown. In the ET-OFF, no fixation target was shown during the whole trial, and subjects were not required to do the fixation. A trial ended without feedback after subjects pressed the button to report perceived heading direction as soon as possible after the stimulus onset. There were 3,240 trials (12 directions  $\times$  30 repeats  $\times$  3 conditions  $\times$  3 stimuli), which were divided into 3 or 4 blocks and finished in separate days. Trials were randomly interleaved and intertrial interval was 5 s. All the other settings, procedures, and instructions were same as those in the E condition.

### Data analysis

All data analyses were done in Matlab (Mathworks). In the E condition, a trial was selected as valid if eye data during stimulation period were within a  $5^\circ \times 5^\circ$  rectangular window centered around the fixation point (Supplementary Fig. S1, see online supplementary material for a color version of this figure). The point of subjective equality (PSE) and threshold of psychometric curves fitted with valid trials ( $82 \pm 12\%$ ) were not significantly different from that fitted with all trials across subjects ( $P = 0.9$  for PSE and 0.54 for threshold, Wilcoxon rank sum test). Therefore, all trials were included in the analyses for H and EH conditions. Significant gaze effects were tested with parametric/nonparametric tests, analysis of variance (ANOVA), and linear regression.

### Psychometric function fitting

The analysis was based on psychometric curves representing the proportion of rightward choices as a function of heading (Wichmann and Hill 2001). We calculated a separate psychometric curve for each gaze direction in each stimulus and each gaze condition. The psychometric curves were parameterized by fitting the

probability  $P(\theta)$  of rightward choices with a cumulative Gaussian function (Klein 2001; Ni et al. 2013) using the Palamedes toolbox (Prins and Kingdom 2018):

$$P(\theta; \text{PSE}, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty \exp\left[-\frac{(\tilde{\theta} - (\theta - \text{PSE}))^2}{2\sigma^2}\right] d\tilde{\theta} \quad (1)$$

The  $\theta$  is the heading direction. The PSE characterizes the accuracy of heading discrimination corresponding to the heading which yields 50% rightward and 50% leftward choices. A PSE close to  $0^\circ$  indicates accurate heading discrimination. The  $\sigma$  is the psychophysical threshold representing the inverse precision of heading discrimination. A small threshold indicates highly precise (i.e. sensitive) heading discrimination and steep psychometric curves. PSE and  $\sigma$  were estimated by fitting psychometric function using Bayesian criterion.

$$\text{IPSE}_{ij} = \frac{p_i - p_j}{g_i - g_j}; \text{ITHR}_{ij} \text{ or } \text{IRT}_{ij} = \frac{p_i - p_j}{|g_i - g_j|} \quad (2)$$

Three indices formulated by Equation (2) were computed to quantify the shift of psychometric curves (IPSE), the change of slope or threshold (ITHR), as well as the change of RT of subjects (IRT), relative to the gaze change. The numerator represents the parameter change (p: PSE, threshold, or RT) between a pair of psychometric curves (i and j), and the denominator is the difference between 2 gazes ( $g_i$  and  $g_j$ ) at which the 2 psychometric functions were measured. The magnitude (|IPSE|, |ITHR|, and |IRT|) reflects the extent to which the heading accuracy, precision and RT is affected by the gaze change, with 0 indicating no effects. |IPSE| ranges between 1 (when the psychometric curve shifts by an amount equal to the gaze change) and 0 (when there is no shift). The sign of IPSE indicates whether the shift direction of psychometric curve and gaze change is consistent (plus for congruent and minus for opposite). In each stimulus and each gaze condition (Fig. 2), 3 IPSE values are computed (one for each distinct pair of the 3 psychometric curves), and 2 ITHR or IRT values are computed (one for left versus center and the other for right versus center, not including left versus right) as the threshold and RT were not sensitive to the direction of gaze change. The average index across the 3 or 2 values was grouped and reported according to whether the gaze change was caused by EIW or EIH. In the H condition, the gaze was changed by varying EIW but keeping EIH constant (pure EIW); conversely in the EH condition, subjects changed gaze by varying EIH but keeping EIW constant (pure EIH). In the E condition, EIW and EIH gaze varied together (mixed). Therefore, the average indices were computed only in E and H conditions for EIW, because the EIW position did not change in EH condition, the denominator of Equation (2) is 0 and the result is meaningless. Similarly, the computation was performed only in E and EH conditions for EIH, as the EIH position did not change in H condition.

### Optimal cue integration

In the RT paradigm of heading discrimination like ours, in which subjects choose when to make a decision, Drugowitsch et al. (2014) reported an extended diffusion model (EDM) that accumulates heading evidence over time and across visual and vestibular signals in a statistically optimal way. The EDM fits the data of choices and RTs of subjects simultaneously, and trades off

accuracy with speed in the decision. The EDM was applied to our data from each gaze condition and each subject (Supplemental Information, SI). The unimodal sensitivity (k) in each stimulus condition (visual, vestibular, and combined) was fitted separately to the complete data set from each subject. The sensitivity in the combined condition was then predicted by computing the optimal integration of unimodal visual and vestibular heading sensitivities by Equation (4) in the SI, and the corresponding threshold was defined by Equation (10) in the SI. ITHR values computed with the observed and predicted thresholds in the combined condition were compared with test if subjects still integrate the visual and vestibular heading signals optimally (Fig. 4). Although the EDM included a bias parameter to capture the bias in heading discrimination (i.e. horizontal shift of the psychometric function), the bias was not specifically modeled as the PSE of subjects. Therefore, the bias was not used to predict the IPSE and compared with the IPSE computed with the observed PSE in the combined condition.

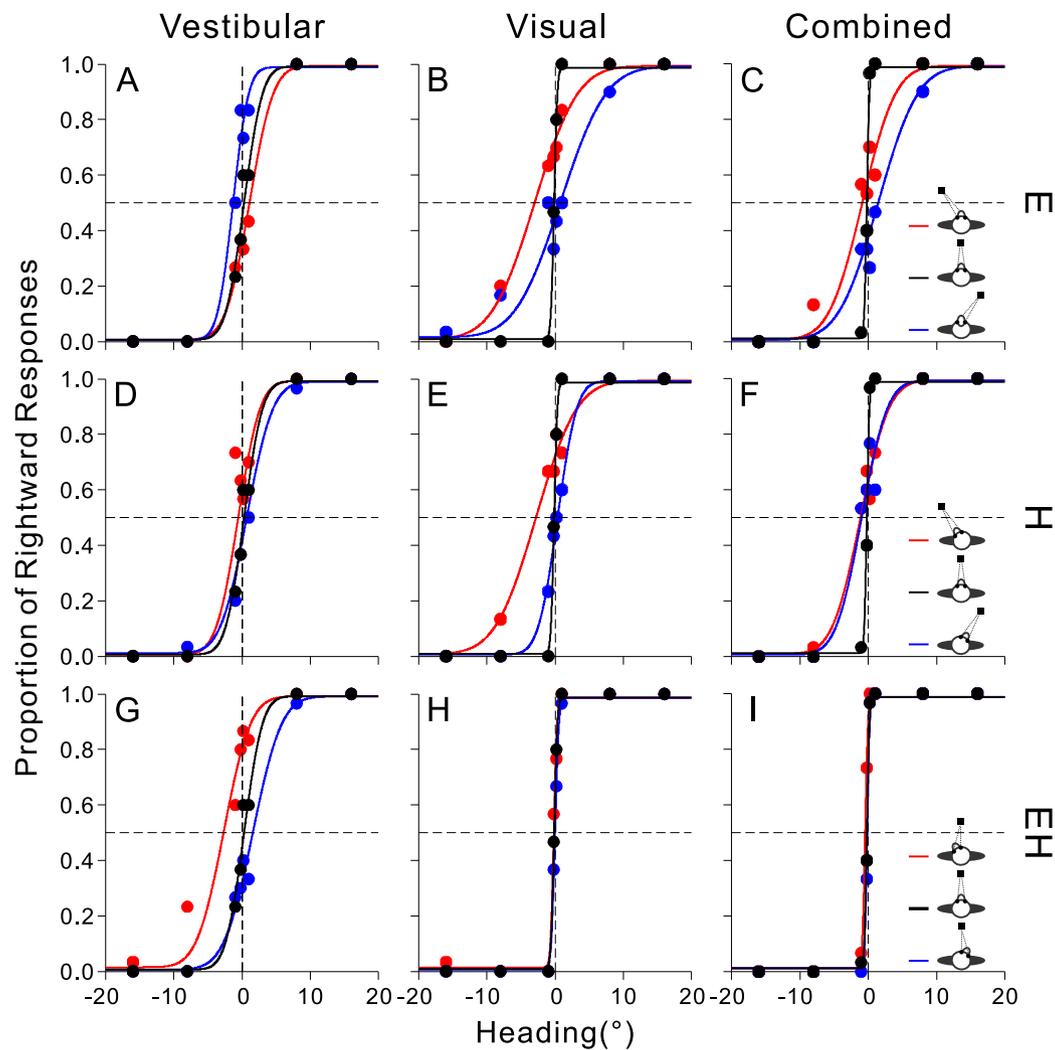
## Results

The gaze direction of subjects was systematically changed by manipulating the (EIH and HOB) position while keeping the BIW direction straight ahead to form 3 conditions (E, H, and EH in Figs. 1 and 2, Methods). In each gaze condition, the performance of heading discrimination was tested with 3 translational self-motion stimuli: Vestibular, visual, and combined. Our results showed that the performance was affected by the EIW and EIH gaze.

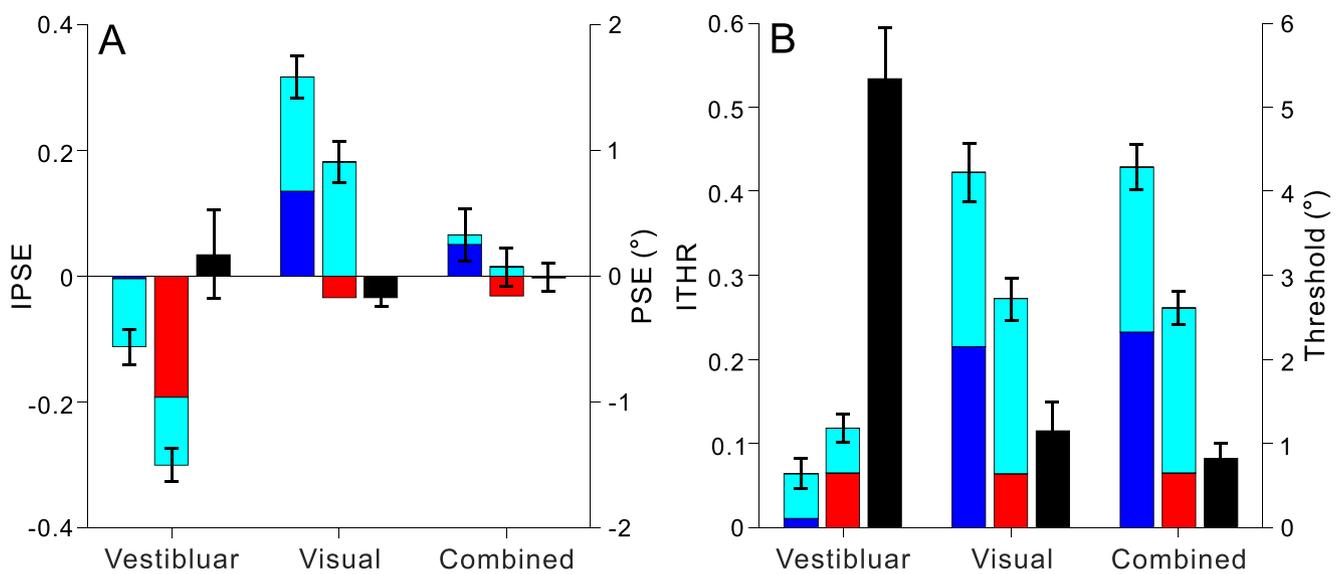
### EIW and EIH gaze affects visual and vestibular heading discrimination, respectively

Data from an example subject are shown in the Fig. 2. The black psychometric curves were measured at  $[0^\circ, 0^\circ]$  position, where (EIH and HOB) aligned with the BIW direction in the straightforward direction and there was no gaze change. The subject perceived heading accurately in all stimulus conditions as PSEs of the black curves were all close to  $0^\circ$  ( $-0.14^\circ$ ,  $0.29^\circ$ , and  $-0.17^\circ$  for visual, vestibular, and combined stimuli, respectively). The average PSE across subjects was also close to and not significantly different from  $0^\circ$  (black bars in Fig. 3A;  $P=0.4$  for visual,  $0.8$  for vestibular and  $>0.9$  for combined, Wilcoxon rank sum test). Although amplitude of the value in the combined condition was minimal, there was no significant difference across stimuli ( $P=0.78$ , 1-way ANOVA). In contrast, the heading precision was highest in combined but lowest in vestibular condition. The black curves for combined ( $\sigma=0.80^\circ$ ) and visual ( $\sigma=0.43^\circ$ ) stimuli were much steeper than the curve for vestibular stimulus ( $\sigma=2.18^\circ$ ), and their average thresholds were significantly less than that for vestibular (black bars in Fig. 3B;  $P<0.001$ , Wilcoxon rank sum test). The average threshold for combined stimulus was not significantly different from that for visual ( $P=0.81$ ). Therefore, the heading discrimination was accurate without gaze change, and the accuracy and precision increased when visual and vestibular stimuli were combined.

However, when gaze deviated from the straightforward direction, the accuracy and precision were affected by EIW and EIH. For the visual stimulus, pure EIW shifted psychometric curves in the direction of gaze and increased thresholds, whereas pure EIH neither significantly shifted curves nor significantly increased thresholds. In H condition (Fig. 2E, pure EIW), relative to the straightforward gaze (black curve), the leftward and rightward gaze shifted psychometric curves substantially to the left (red,



**Fig. 2.** Example data. Data from a representative subject are shown for gaze conditions (rows) and stimuli (columns). Each curve in each panel shows a psychometric function measured at each of the 3 (EIH and HOB) positions ( $[-20^\circ, 0^\circ]$ ,  $[0^\circ, 0^\circ]$ ,  $[20^\circ, 0^\circ]$  in E;  $[0^\circ, -20^\circ]$ ,  $[0^\circ, 0^\circ]$ ,  $[0^\circ, 20^\circ]$  in H; and  $[20^\circ, -20^\circ]$ ,  $[0^\circ, 0^\circ]$ ,  $[-20^\circ, 20^\circ]$  in EH), as indicated by the red, black, and blue curves, respectively and color coded in the insets of C, F, and I.



**Fig. 3.** Summary effects of EIW and EIH gaze on the PSE and threshold. A) The average PSE at  $[0^\circ, 0^\circ]$  gaze position (black psychometric curves in Fig. 2) in the 3 stimulus conditions is shown by the right axis (black bars). The average IPSE in each stimulus and gaze condition is shown by the left axis. Cyan, blue, and red indicate E, H, and EH conditions, respectively. The first bar in each stimulus condition (cyan and blue) is for EIW gaze and the second bar (cyan and red) is for EIH gaze. The error bar indicates SEM of all data for EIW or EIH. B) The average threshold and ITHR are shown. Format is as in A.

PSE =  $-2.79^\circ$ ) and right (blue, PSE =  $0.36^\circ$ ), respectively, and greatly reduced the curve slope ( $\sigma = 4.72^\circ$  for red,  $2.24^\circ$  for blue). However, in EH condition (Fig. 2H, pure EIH), the 3 curves almost overlapped, showing no obvious shift and threshold change (PSE =  $-0.17^\circ$ ,  $\sigma = 0.38^\circ$  for red; and PSE =  $-0.01^\circ$ ,  $\sigma = 0.45^\circ$  for blue). The IPSE ( $-0.004$ ) and ITHR ( $0.002$ ) values for pure EIH were very small and substantially less than that for pure EIW ( $0.079$  and  $0.16$ ). The 2 average indices across subjects for pure EIH (red bars for the visual stimulus in Fig. 3A and B) were not significantly different from 0 ( $P = 0.20$  for IPSE and  $0.22$  for ITHR, Wilcoxon rank sum test), but significantly less ( $P = 0.002$  for IPSE and  $0.001$  for ITHR) than that for pure EIW (blue bars). In E condition (Fig. 2B, mixed), the data were consistent with pure EIW instead of pure EIH. The PSE also shifted in the direction of EIW gaze (PSE =  $-2.99^\circ$  for red and  $0.8^\circ$  for blue; IPSE =  $0.09$ ), and the threshold increased ( $\sigma = 4.96^\circ$  for red,  $5.23^\circ$  for blue; ITHR =  $0.24$ ). The 2 average indices (cyan bars for visual stimulus in Fig. 3A and B) were significantly greater than that for pure EIH ( $P < 0.001$  for both IPSE and ITHR) but not significantly different from that for pure EIW ( $P = 0.76$  for IPSE,  $0.78$  for ITHR). In summary, EIW instead of EIH reduced the accuracy of visual heading discrimination by shifting the PSE in the direction of gaze and reduced the precision by increasing the threshold.

For the vestibular stimulus, pure EIH shifted psychometric curves in the opposite direction of gaze and increased thresholds, whereas pure EIW neither significantly shifted curves nor significantly increased thresholds. In EH condition (Fig. 2G, pure EIH), leftward gaze substantially shifted the psychometric curve to the right (blue, PSE =  $1.57^\circ$ ), and rightward gaze shifted the curve to the left (red, PSE =  $-2.68^\circ$ ). However, in H condition (Fig. 2D, pure EIW), there was only a slight shift between curves (PSE =  $-0.56^\circ$  for red,  $0.68^\circ$  for blue). The IPSE for pure EIW ( $0.031$ ) was close to 0 and substantially greater than that for pure EIH ( $-0.11$ ). Negative IPSE values indicated opposite shifting relative the gaze. The average IPSE for pure EIW (the blue bar for vestibular stimulus in Fig. 3A) was not significantly different from 0 ( $P > 0.9$ , Wilcoxon rank sum test), but significantly greater ( $P < 0.001$ ) than that for pure EIH (red bar). Although there was a visible increase of threshold in the EH ( $\sigma = 3.11^\circ$  for red,  $3.34^\circ$  for blue; ITHR =  $0.055$ ) and H ( $\sigma = 2.53^\circ$  and  $3.0^\circ$ ; ITHR =  $0.032$ ) conditions, the average ITHR for pure EIW (the blue bar for vestibular stimulus in Fig. 3B) was not significantly different from 0 ( $P = 0.8$ ). The value for pure EIH (red bar) was greater than 0 and that for pure EIW but was not significant ( $P = 0.2$  and  $0.11$ ). In E condition (Fig. 2A, mixed), the data were more consistent with pure EIH instead of pure EIW. The PSE also shifted in the opposite direction of EIH gaze (PSE =  $1.1^\circ$  for red,  $-1.33^\circ$  for blue; IPSE =  $-0.06$ ), and the threshold increased slightly ( $\sigma = 2.74^\circ$  for red,  $1.78^\circ$  for blue; ITHR =  $0.007$ ). The average IPSE (the cyan bar for vestibular stimulus in Fig. 3A) was significantly less than the value for pure EIW ( $P = 0.03$ ), but not significantly different from that for pure EIH ( $P = 0.08$ ). The average ITHR (the cyan bar for vestibular stimulus in Fig. 3B) was not significantly different from that for pure EIW ( $P = 0.86$ ) and pure EIH ( $P = 0.27$ ). In summary, EIH instead of EIW reduced the accuracy of vestibular heading discrimination by shifting the PSE in the opposite direction of gaze, which was opposite to the gaze effect for visual stimulus. And only EIH gaze obviously reduced the precision by increasing the threshold.

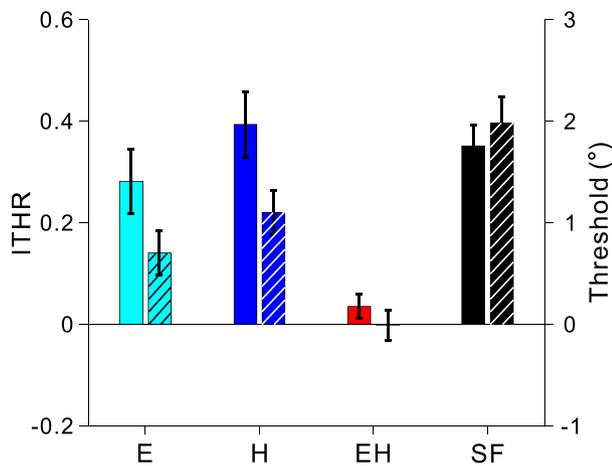
For the combined stimulus, neither pure EIW nor pure EIH gaze significantly shifted psychometric curves, whereas pure EIW instead of pure EIH significantly increased the threshold. There was no obvious PSE shift by pure EIW (Fig. 2F; PSE =  $-1.1^\circ$  for red,  $-0.96^\circ$  for blue; IPSE =  $0.004$ ) and pure EIH gaze (Fig. 2I;

PSE =  $-0.45^\circ$  for red,  $-0.15^\circ$  for blue; IPSE =  $-0.008$ ). The average IPSE for pure EIW (the blue bar for combined stimulus in Fig. 3A) and pure EIH (red bar) were not significantly different from 0 ( $P > 0.9$ , Wilcoxon rank sum test), and not significantly different from each other ( $P = 0.34$ ). However, the threshold was substantially increased by pure EIW ( $\sigma = 3.26^\circ$  for red,  $2.97^\circ$  for blue), but not by pure EIH ( $\sigma = 0.33^\circ$  and  $0.28^\circ$  in Fig. 2I). The ITHR for pure EIH ( $0.0002$ ) was very small and substantially less than that for pure EIW ( $0.14$ ). The average ITHR for pure EIH (the red bar for combined stimulus in Fig. 3B) was not significantly different from 0 ( $P = 0.2$ ), but significantly less ( $P < 0.001$ ) than that for pure EIW (blue bar). In E condition (Fig. 2C, mixed), the data were consistent with pure EIW instead of pure EIH. There was a slight shift of PSE (PSE =  $-0.85^\circ$  for red,  $1.45^\circ$  for blue; IPSE =  $0.058$ ), but a substantial increase of threshold ( $\sigma = 3.83^\circ$  for red,  $4.56^\circ$  for blue; ITHR =  $0.195$ ). The average IPSE (the cyan bar for combined stimulus in Fig. 3A) was not significantly different from 0 ( $P > 0.9$ ), the value for pure EIW ( $P = 0.61$ ) and pure EIH ( $P = 0.93$ ). The average ITHR (the cyan bar for combined stimulus in Fig. 3B) was not significantly different from that for pure EIH ( $P = 0.47$ ) and pure EIW ( $P > 0.9$ ). In summary, neither EIW nor EIH gaze significantly affected the accuracy of heading discrimination when the visual and vestibular signals were combined, and EIW instead of EIH gaze reduced the precision.

Comparing the average IPSE across stimulus conditions (Fig. 3A), for combined versus visual, the comparison was significantly less for the overall IPSE values (EIW plus EIH,  $P = 0.02$ ), overall EIW (pure plus mixed,  $P = 0.02$ ) and EIH (pure plus mixed,  $P = 0.03$ ), mixed ( $P = 0.03$ ), but not for pure EIW ( $P = 0.19$ ) and pure EIH ( $P = 0.73$ ). For combined versus vestibular, the comparison was significantly greater for overall IPSE ( $P = 0.004$ ), overall EIH ( $P < 0.001$ ) and pure EIH ( $P < 0.001$ ), but not for overall EIW ( $P = 0.11$ ), pure EIW ( $P = 0.27$ ), and mixed ( $P = 0.15$ ). For visual versus vestibular, all comparisons were significantly greater ( $P < 0.05$ ). Comparing the average ITHR across stimulus conditions (Fig. 3B), for combined versus visual, all comparisons were not significantly different ( $P > 0.05$ ). For combined versus vestibular, all comparisons were significantly greater ( $P < 0.01$ ) except for pure EIH ( $P = 0.47$ ). For visual versus vestibular, all comparisons were significantly greater ( $P < 0.01$ ) except for pure EIH ( $P = 0.44$ ). Therefore, compared with individual stimulus conditions, the gaze effects on bias of the heading discrimination were reduced when the visual and vestibular stimuli were combined, whereas the effects on threshold were closer to that in the visual condition and not significantly reduced. In addition, the overall effect of gaze on threshold was lowest in the vestibular condition, suggesting vestibular heading was less susceptible to the gaze change, although it was more sensitive to the EIH gaze.

### Gaze substantially affects the integration of visual and vestibular signals

Many studies reported that visual and vestibular signals were integrated in an optimal mode to facilitate heading perception (Gu et al. 2008; Fetsch et al. 2009; Butler et al. 2011; Fetsch et al. 2011; Gu et al. 2012; Prsa et al. 2012; de Winkel et al. 2013; Chen et al. 2013a; Prsa et al. 2015). According to our results above, in the combined stimulus condition, accuracy, and precision of the heading discrimination increased when the EIW and EIH gaze was not changed ( $[0^\circ, 0^\circ]$  position in Figs. 2 and 3). However, the increase was only observed in accuracy but not precision when the EIW and/or EIH gaze deviated from the straightforward direction. Therefore, we tested if the optimal integration model still held when gaze varied. However, it is reported that the integration



**Fig. 4.** Gaze effects on the optimal cue integration. The average threshold measured (filled black bars) and predicted by the EDM (hatched black bars) at  $[0^\circ, 0^\circ]$  gaze position in the combined stimulus condition is shown by the right y axis (SF: straightforward, black psychometric curves in Fig. 2). The average ITHR measured (filled color bars) and predicted (hatched color bars) in the combined condition is shown by the left y axis. Cyan, blue, and red indicate E, H, and EH conditions, respectively. The error bar indicates SEM. Stars indicate significance ( $P < 0.05$ ).

is suboptimal when quantified with traditional optimality metrics that ignore RTs in RT paradigm of the heading discrimination like ours where the RT is under the control of subjects (Drugowitsch et al. 2014). To take into account how a change in RT might impact heading discrimination performance, the data of choices and RTs of our subjects were fitted simultaneously with the EDM that accumulates heading evidence optimally over time and across visual and vestibular signals (Methods, Drugowitsch et al. 2014). ITHR values based on the fitted and predicted thresholds in the combined condition were computed and compared (Fig. 4). For comparison, the ITHR and IPSE values were still quantified with the traditional optimal integration model (TOIM, in the Supplementary Fig. S4 of Supplemental Information, see online supplementary material for a color version of this figure).

When the EIW and EIH gaze aligned in the straightforward direction without change, the measured threshold was not significantly different from the threshold predicted with the EDM ( $P=0.72$  in Fig. 4, Wilcoxon rank sum test) or TOIM ( $P=0.5$  in Supplementary Fig. S4B, see online supplementary material for a color version of this figure), which indicated that subjects still implemented an optimal integration strategy in the heading discrimination that considered the reliability of unimodal visual and vestibular signals as reported previously (Fetsch et al. 2009; Drugowitsch et al. 2014; Dokka et al. 2015). However, when the EIW and/or EIH gaze deviated from the straightforward direction, the predicted ITHR with TOIM was significantly smaller than that measured across gaze conditions ( $P < 0.001$ , 1-way ANOVA), and individually in the E ( $P < 0.01$ , Wilcoxon rank sum test), H ( $P = 0.01$ ), and EH ( $P < 0.05$ ) conditions (Supplementary Fig. S4B, see online supplementary material for a color version of this figure). For the EDM, the difference in the ITHR between predicted and measured became significantly smaller than that for the TOIM across gaze conditions ( $P = 0.03$  in Fig. 4 and Supplementary Fig. S4B, see online supplementary material for a color version of this figure; 1-way ANOVA). However, the predicted ITHR with EDM was still substantially smaller than that measured (Fig. 4), although our data do not allow us to reject the null hypothesis that the comparison was not

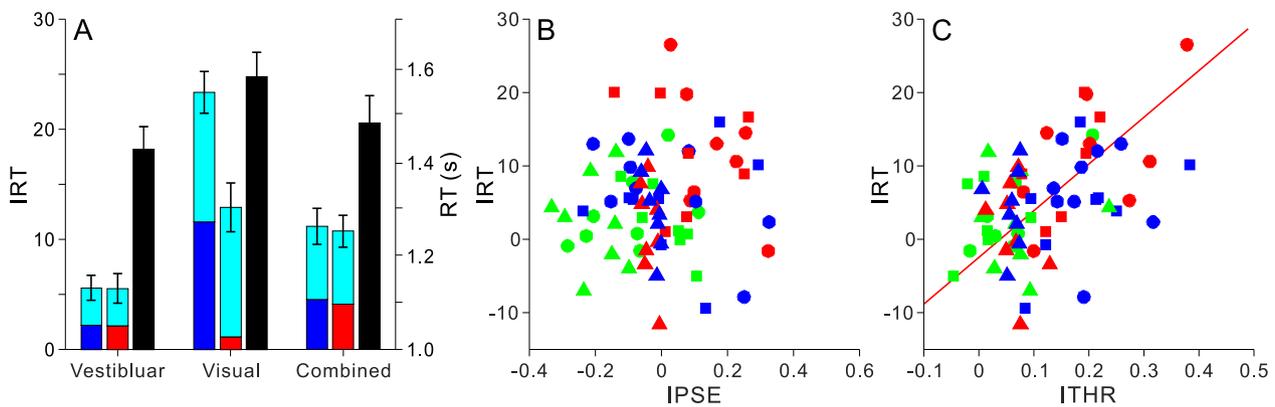
significantly smaller in each gaze condition ( $P = 0.16$ ,  $0.13$ , and  $0.9$  for E, H, and EH, respectively, Wilcoxon rank sum test), but the comparison indicated a significant difference across the gaze conditions ( $P = 0.01$ , 1-way ANOVA). In addition, the average PSE predicted with TOIM was also not significantly different from that measured when the EIW and EIH gaze aligned in the straightforward direction ( $P = 0.69$  in Supplementary Fig. S4A). Although the magnitude of predicted IPSE with TOIM was substantially larger than that measured when the gaze deviated from the straightforward direction, our data do not allow us to reject the null hypothesis that the comparison was not significantly greater in each and across gaze conditions ( $P = 0.49$  in Supplementary Fig. S4A, see online supplementary material for a color version of this figure, 1-way ANOVA). Therefore, the integration of visual and vestibular signals significantly deviated from the TOIM by gaze change, and also substantially deviated from the EDM even if the EDM accumulates evidence optimally across both signals and over time.

### Gaze increases the time for heading discrimination

RT could reflect how much cognitive effort it took subjects to do the heading discrimination in different gaze conditions. When EIW and EIH aligned with the BIW direction in the straightforward direction (black bars in Fig. 5A), the average RT across subjects for the vestibular stimulus was significantly less than that for visual ( $P = 0.02$ , Wilcoxon rank sum test), but there was no significant difference between vestibular versus combined ( $P = 0.22$ ) and visual versus combined ( $P = 0.11$ ). Therefore, the cognitive effort spent on discriminating heading was most in the visual stimulus condition, which was reduced when combined with vestibular stimulus.

However, the RT increased when the EIW and/or EIH gaze deviated from the straightforward direction. Comparing the average IRT with 0 (Fig. 5A), for vestibular stimulus, the comparison was significantly greater for overall IRT ( $P = 0.01$ , t-test), overall EIW ( $P = 0.02$ ), and overall EIH ( $P = 0.03$ ). For visual stimulus, all comparisons were significantly greater ( $P < 0.01$ ), and the average IRT for pure EIH was significantly less than that for pure EIW ( $P = 0.003$ ) and mixed ( $P < 0.01$ ). Therefore, the RT of visual heading discrimination was increased by EIW but not EIH gaze. For combined stimulus, all comparisons were significantly greater ( $P < 0.002$ ). Comparing the average IRT across stimulus conditions (Fig. 5A), for vestibular versus visual, the comparison was significantly less for overall IRT ( $P = 0.01$ , Wilcoxon rank sum test), overall EIW ( $P = 0.001$ ), pure EIW ( $P = 0.008$ ), and mixed ( $P = 0.03$ ), but not for overall EIH ( $P = 0.1$ ) and pure EIH ( $P = 0.52$ ). For vestibular versus combined, the comparison was significantly less for overall IRT ( $P = 0.03$ ) and overall EIW ( $P = 0.04$ ), but not for others ( $P > 0.05$ ). For visual versus combined, the comparison was significantly greater for overall EIW ( $P = 0.02$ ), but not for others ( $P > 0.05$ ). In summary, gaze change increased RT in all stimulus and gaze conditions, suggesting it took subjects more time to discriminate the heading.

In addition, we tested whether the increase of RT was synchronous with the change of PSE and threshold. Across gaze conditions, ITHR and IRT were significantly correlated in the visual stimulus condition (Pearson correlation  $R = 0.6$ ,  $P < 0.001$ , Fig. 5C), but the correlation was not significant in the vestibular ( $R = 0.27$ ,  $P = 0.2$ ) and combined ( $R = 0.29$ ,  $P = 0.2$ ) conditions. However, the IPSE and IRT were not significantly correlated in all stimulus conditions.



**Fig. 5.** Summary effects of EIW and EIH gaze on the reaction time. A) The average reaction time (RT) of subjects at  $[0^\circ, 0^\circ]$  gaze position in the 3 stimulus conditions is shown by the right axis (black bars). The average IRT in each stimulus and gaze condition is shown by the left axis. Color indicates gaze conditions: Cyan, blue, and red for E, H, and EH conditions, respectively. The first bar in each stimulus condition (cyan and blue) is for EIW gaze and the second bar (cyan and red) is for EIH gaze. The error bar indicates SEM of all data for EIW or EIH. The correlation between IRT with IPSE (B) and ITHR (C) are shown. Each symbol represents data from one subject. Shape indicates gaze conditions: Circle, square, and triangle for E, H, and EH, respectively. Color indicates stimuli: Red, green, and blue for visual, vestibular, and combined, respectively. The solid lines show regression fits.

### Gaze reference affects heading precision and reaction time but not accuracy

As an ET was used in our experiments to control the eye fixation positions, so it may potentially be used as a reference by subjects in discriminating heading. In order to test this possibility, the eye fixation target was manipulated differently in 3 ET conditions (Methods): ET-ON (always turned on in a trial), ET-CUE (only shown for 500 ms before stimulus onset) and ET-OFF (no ET in a trial). The subject perceived heading accurately in all the ET and stimulus conditions (Fig. 6A–C): The PSE in [ET-ON, ET-CUE, and ET-OFF] conditions was  $[-0.10, -0.39, \text{ and } -0.30]$  for the visual,  $[1.19, 1.17, \text{ and } 0.53]$  for vestibular and  $[-0.17, -0.39, -0.29]$  for combined stimuli. The average PSE across subjects (Fig. 6D) was not significantly different from  $0^\circ$  in all the ET and stimulus conditions ( $P > 0.05$ , Wilcoxon rank sum test), and not significantly different between ET conditions ( $P = 0.51$ ) and stimuli ( $P = 0.08$ , 2-way ANOVA). However, the heading precision was slightly different between ET conditions (Fig. 6A–C): The threshold was  $[0.40, 0.87, \text{ and } 1.02]$  for the visual,  $[3.03, 4.60, \text{ and } 5.03]$  for vestibular and  $[0.29, 0.91, \text{ and } 1.17]$  for combined stimuli. The average thresholds in ET-OFF and ET-CUE conditions were not significantly different for all stimuli ( $P > 0.05$  in Fig. 6E), but they were significantly greater than the value in ET-ON condition for visual and combined stimuli ( $P < 0.01$ ) but not for vestibular. The average threshold for vestibular was significantly greater than that for visual and combined ( $P < 0.001$ ), but there was no significant difference between visual and combined ( $P = 0.49$ ). The average RT (Fig. 6F) was significantly different between ET conditions ( $P = 0.02$ ) and stimuli ( $P = 0.004$ , 2-way ANOVA). In summary, turning off the gaze reference during stimulation period reduced the heading precision and increased RT, but did not affect the heading accuracy. In addition, the vestibular heading was fastest and insensitive to the gaze reference although its precision was lowest.

### Discussion

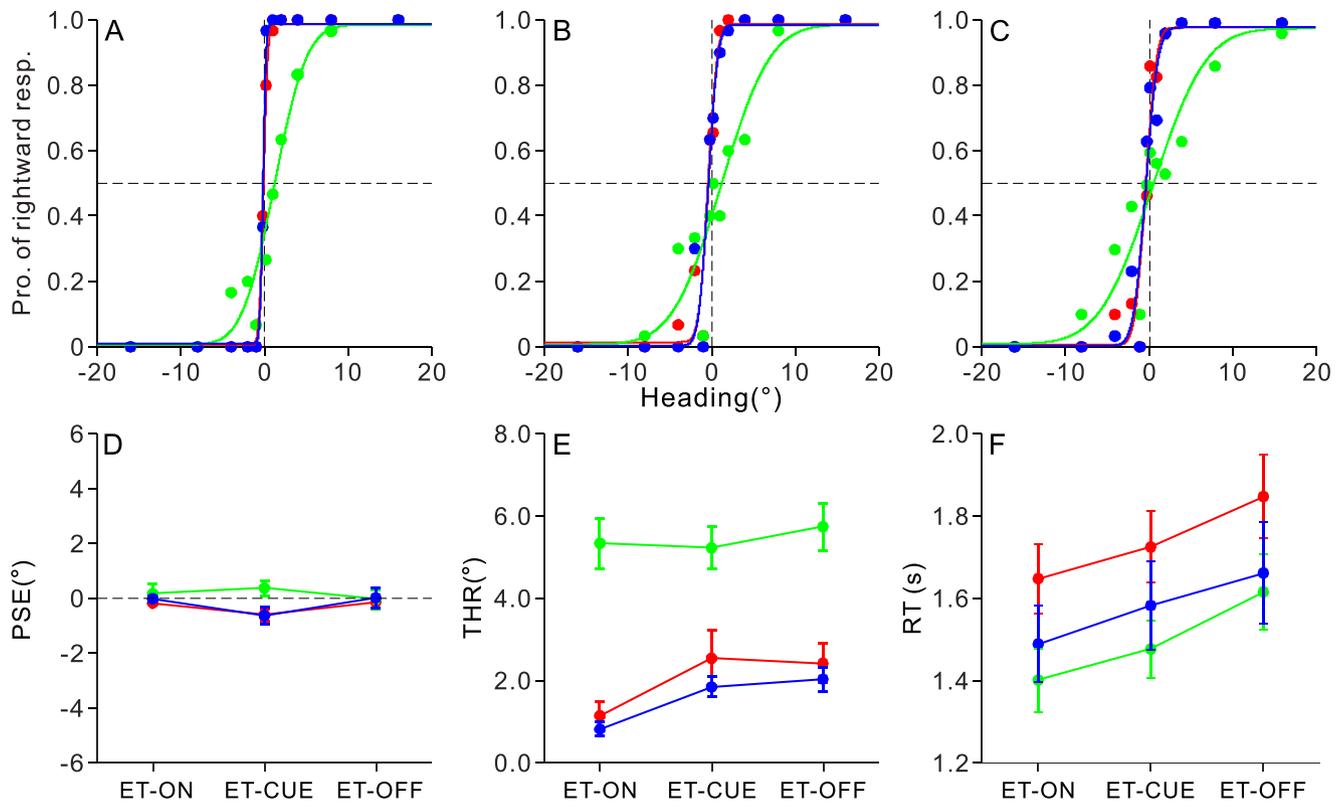
In order to study how the heading discrimination was affected by gaze, the (EIH and HOB) position was systematically varied while keeping the BIW direction straight ahead. In each gaze condition, the performance of subjects was tested with 3 translational self-motion stimuli: Vestibular, visual, and combined. When the (EIH and HOB) aligned with the BIW direction and there was

no gaze change, the discrimination was accurate in all stimulus conditions. In vestibular condition, the precision was lowest, and RT was shortest. The optimal integration of visual and vestibular signals enabled subjects to have highest heading precision in the combined condition. However, when gaze deviated from the straightforward direction, the accuracy, precision and RT were substantially affected by EIW and EIH gaze.

### The gaze effects on visual heading

In the visual stimulus condition, when the gaze was varied by manipulating EIW, the PSE of psychometric curves shifted in the direction of gaze (Figs. 2 and 3). The first possible reason for this result was that subjects discriminated the visual heading using the eye fixation target as reference instead of the straightforward direction of their trunk midline as instructed (Methods). If this was true, the PSE should shift same magnitude as the EIW gaze, and the IPSE value should be close to 1. However, the average IPSE for EIW gaze were very small ( $0.16 \pm 0.03$  in Fig. 3) and significantly less than 1 ( $P < 0.001$ ). So, subjects were unlikely to do the discrimination relative to the EIW gaze. The second possible reason was that the eye fixation positions were tracked but not restricted in the H condition (pure EIW). Then, the actual eccentric fixation positions were likely to be  $< 20^\circ$  as designed, and the IPSE values could be underestimated. However, the eye positions were tracked and restricted in the E condition, and most eye data were within a  $5^\circ \times 5^\circ$  rectangular window centered around the fixation point (Supplementary Fig. S1, see online supplementary material for a color version of this figure, Methods). So, even in extreme cases, the minimum fixation eccentricity would be  $17.5^\circ$ , and actual value of the underestimated IPSE was still much  $< 1$ . Moreover, our results in the ET condition (Fig. 6) showed that even without restriction of eye movement, the PSE was not significant different from that with restriction. Therefore, our results were unlikely to be interpreted by no restriction of eye movement.

Ideally, subjects should had discriminated the visual heading using body as the reference. As the BIW direction did not change in our experiments, no matter how the (EIH and HOB) position varied, PSE should not change, and IPSE should be close to 0. However, we found the IPSE was significantly deviated away from 0 and toward 1 by the EIW gaze (Figs. 2 and 3). Crane (2017) reported similar EIW gaze effects and found that the threshold increased when the noise in the visual stimulus was increased by decreasing the visual coherence. Yang and Gu (2017) reported



**Fig. 6.** Summary effects of gaze reference to heading discrimination. Psychometric curves from a representative subject are shown for ET-ON A), ET-CUE B), and ET-OFF C) conditions. The PSE, threshold and reaction time of subjects are summarized in D-F). Color indicates stimuli: Red, green, and blue for visual, vestibular, and combined, respectively. The error bar indicates SEM.

monkeys could do visual heading estimation when the EIW gaze changed, but the performance had a small bias in the direction of gaze. We assume these results are related to the transformation of spatial reference frames. The initial reference frame encoding visual signals is the eye, and almost all cortical areas encoding the optic flow stimulus are eye-centered (Fetsch et al. 2007; Yang et al. 2011; Chen et al. 2013c, 2014; Fan et al. 2015; Yang and Gu 2017). So, when the EIW gaze deviates from the straightforward direction and misaligns the eye- and body-centered spatial reference frames, it is impossible that the eye-centered visual information could be used directly for heading discrimination. Otherwise, it is equivalent to use an eye-centered reference to do the discrimination, and the IPSE should be close to 1. Therefore, there should be a transformation of the heading information from eye-centered to body-centered in the brain, by integrating the eye-centered sensory and EIW gaze signals. Our results also showed that the threshold and RT were increased by the EIW gaze and changed synchronously (Figs. 3 and 5). This suggested that the change of EIW gaze reduced the sensitivity of subjects to heading, and they may have to spend additional time doing the potential spatial reference frame transformation. However, the transformation was not complete as the IPSE was not 0 but biased in the direction of EIW gaze. This may be caused by noise in the perceived visual and EIW gaze signals, which increased the uncertainty in the transformation (Schlicht and Schrater 2007).

### The gaze effects on vestibular heading

Although the initial reference frame encoding vestibular signals is the head, the cortical representation of vestibular heading

signals are diverse (Fetsch et al. 2007; Chen et al. 2013a; Fan et al. 2015; Yang and Gu 2017; Chen et al. 2018). For example, vestibular signals follow a representation that is intermediate between head- and eye-centered in the dorsal medial superior temporal area (MSTd, Fetsch et al. 2007; Chen et al. 2013a; Yang and Gu 2017), but intermediate between head- and body-centered in the parietoinsular vestibular cortex (PIVC, Chen et al. 2013b). Most remarkably, if the BIW direction did not change, vestibular heading signals in VIP were shown to remain stable with respect to the body despite variations in (EIH and HOB) position, suggesting a body-centered spatial reference frame (Chen et al. 2013c, 2018). So, the VIP signals in the brain could be used for vestibular heading discrimination. However, subjects in our experiments did not use the body as an absolute reference to discriminate the vestibular heading, as the PSE was biased in the direction of EIH gaze (Figs. 2 and 3). Similar results were also reported (Ni et al. 2013; Crane 2017). This suggested that subjects may use a biased body-centered reference frame to do the discrimination. Indeed, Chen et al. (2018) reported that VIP neurons shifted the reference frame encoding vestibular heading signals from body-centered to world-centered when the body rotated but the gaze remained fixed in the world. Moreover, when monkeys switched reference frames from trial to trial between head- and world-centered in a task to discriminate the direction of a moving object during self-motion, Sasaki et al. (2020) reported that the neural responses in VIP were modulated by the task reference frame to represent the object direction in either reference frame, and vestibular signals was needed to compute the object motion. Therefore, VIP neurons may flexibly and dynamically change the encoding reference frame depending on gaze and task demands, which may underlie our behavioral results.

## The gaze effects on integration of visual and vestibular signals

Many studies reported that both human and macaque subjects could integrate visual and vestibular signals in a statistically optimal way to facilitate heading perception (Gu et al. 2008; Fetsch et al. 2009; Butler et al. 2011; Fetsch et al. 2011; Gu et al. 2012; Prsa et al. 2012; de Winkel et al. 2013; Chen et al. 2013a; Prsa et al. 2015; Crane 2017). In our experiments, when the (EIH and HOB) aligned with the BIW direction and there was no gaze change, the accuracy and precision of heading discrimination increased in the combined stimulus condition (Figs. 2 and 3), and the PSE and threshold predicted with the TOIM matched that measured (Supplementary Fig. S4). This indicated that subjects still implemented an optimal way in judging heading. However, when gaze deviated from the straightforward direction, the predicted ITHR was significantly less than that measured, and the magnitude of predicted IPSE was substantially greater than that measured (Supplementary Fig. S4, see online supplementary material for a color version of this figure). This indicated that gaze change deviated the integration of visual and vestibular signals from the optimal model.

The first possible reason for this deviation may come from overestimation of the visual heading signals in the integration, as gaze effects in the combined condition were closer to that in the visual condition (Fig. 3). As the optic flow motion coherence was 100%, the visual stimulus is relatively more reliable than the vestibular stimulus (Fetsch et al. 2009, 2011; Dokka et al. 2015; Crane 2017), and subjects may rely more on visual than vestibular signals in judging heading, although the vestibular sensitivity (threshold) was less susceptible to the gaze change (Fig. 3B). Therefore, the visual component in the combined stimulus may be weighted more heavily by subjects than that would be predicted from the TOIM, which made the predicted threshold lower than that measured. The second possible reason is noise introduced by the sensory information and potential transformation of spatial reference frames during the integration. Visual signals extracted from the optic flow stimulus could be noisy and with errors about the heading. Such noisy signals are encoded in an eye-centered reference frame in the cortex and need to be transformed to a body-centered reference frame by combining noisy gaze signals before integrated with the vestibular signals. This process may introduce additional coordinate transformation uncertainty (Schlicht and Schrater 2007), which could deviate the integration from TOIM. In addition, the threshold and RT increased in the combined stimulus condition (Figs. 3 and 5), which indicated the uncertainty was increased by the gaze change and subjects spent more time to discriminate the heading.

By incorporating RT, the TOIM was extended to form the EDM that accumulates evidence optimally over time and across visual and vestibular signals, and trades off accuracy with speed in the decision of subjects performing heading discrimination. Like TOIM, the threshold predicted by EDM in the straightforward direction matched that measured (Fig. 4), indicating optimal integration. Furthermore, compared with TOIM, differences in the ITHR between predicted and measured were significantly reduced in EDM when gaze deviated from the straightforward direction, which indicated that EDM explained more variance in the behavior data than that explained by TOIM. However, the predicted ITHR with EDM was still substantially smaller than that measured, indicating there was still a lot of variance that cannot be explained by EDM and may come from the spatial reference frame transformation during the heading discrimination of subjects. In order to test this, the TOIM and EDM should be extended to model

the process of potential transformation by incorporating signals related to EIH and HOB positions of subjects (Cohen and Andersen 2002), the reference frame representation of visual and vestibular signals in cortical neurons (Fetsch et al. 2007; Chen et al. 2013a, 2013b, 2014; Fan et al. 2015; Yang and Gu 2017; Chen et al. 2018), and considering the relative reliability of the 2 signals (Fetsch et al. 2009, 2011; Dokka et al. 2015; Crane 2017) and the modulation of possibly attentional or top-down demands (Chen et al. 2018).

In conclusion, our behavioral results indicate that gaze change substantially affects the heading discrimination and integration of visual and vestibular signals, suggesting the transformation of spatial reference frames encoding heading signals in cortex underlies the results. Indeed, flexibility, diversity, and dynamicity in spatial reference frame representations have been reported in a number of sensory signals and brain areas (Snyder et al. 1998; Metzger et al. 2004; Avillac et al. 2005; Mulette-Gillman et al. 2005; Schlack et al. 2005; Pesaran et al. 2006; Sereno and Huang 2006; Fetsch et al. 2007; Crowe et al. 2008; Mulette-Gillman et al. 2009; Bernier and Grafton 2010; McGuire and Sabes 2011; Rosenberg et al. 2013; Chen et al. 2013a, 2013b; Bremner and Andersen 2014; Chen et al. 2014; Leone et al. 2015; Chen et al. 2018), and many multisensory areas like VIP encode different sensory signals in different reference frames. Although all these suggest prevalence of the transformation of spatial reference frames in cortex, our results show that such transformation may not be very accurate and precise. Further experiments are needed to explore how neural responses represent the spatial reference frame transformation. And current theories about the integration of visual and vestibular signals may need to be extended by studying how such transformation affects heading perception in future. In addition, the difference in experimental procedures may also need to be considered, as the difference may affect the size of gaze effect and thus signal integration. For example, the adaptive (staircase) procedure (Crane 2017) seems to show larger effective size than the RT procedure used in our experiments.

## Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

## Funding

This study was supported by National Key Research and Development Program of China under grant no. 2020AAA0105900 to H.J.T. Science and Technology Innovation 2030—major project of Brain Science and Brain-Like Research (2021ZD0202602), National Natural Science Foundation of China (31471047 and 32171028), and the MOE Frontier Science Center for Brain Science & Brain-Machine Integration, Zhejiang University, Fundamental Research Funds for the Central Universities (2019XZZX001-01-19) to X.D.C.

*Conflict of interest statement:* The authors declare no conflict of interest.

## Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## References

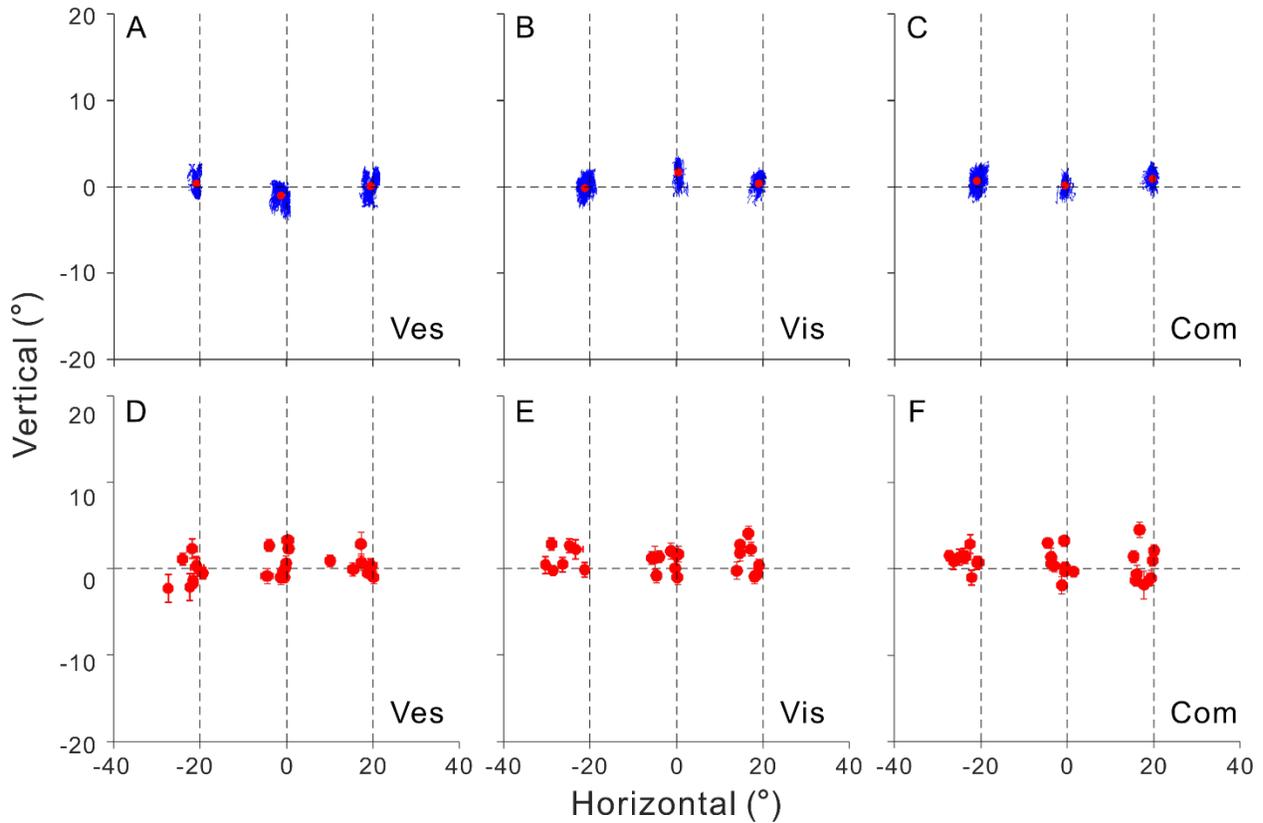
Avillac M, Deneve S, Olivier E, Pouget A, Duhamel JR. Reference frames for representing visual and tactile locations in parietal cortex. *Nat Neurosci*. 2005;8:941–949.

- Bernier PM, Grafton ST. Human posterior parietal cortex flexibly determines reference frames for reaching based on sensory context. *Neuron*. 2010;68:776–788.
- Bremmer F, Ilg UJ, Thiele A, Distler C, Hoffmann KP. Eye position effects in monkey cortex. I. Visual and pursuit-related activity in extrastriate areas MT and MST. *J Neurophysiol*. 1997;77:944–961.
- Bremner LR, Andersen RA. Temporal analysis of reference frames in parietal cortex area 5d during reach planning. *J Neurosci*. 2014;34:5273–5284.
- Butler JS, Campos JL, Bulthoff HH, Smith ST. The role of stereo vision in visual-vestibular integration. *Seeing Perceiving*. 2011;24:453–470.
- Chen A, DeAngelis GC, Angelaki DE. Functional specializations of the ventral intraparietal area for multisensory heading discrimination. *J Neurosci*. 2013a;33:3567–3581.
- Chen X, DeAngelis GC, Angelaki DE. Diverse spatial reference frames of vestibular signals in parietal cortex. *Neuron*. 2013b;80:1310–1321.
- Chen X, DeAngelis GC, Angelaki DE. Eye-centered representation of optic flow tuning in the ventral intraparietal area. *J Neurosci*. 2013c;33:18574–18582.
- Chen X, DeAngelis GC, Angelaki DE. Eye-centered visual receptive fields in the ventral intraparietal area. *J Neurophysiol*. 2014;112:353–361.
- Chen X, DeAngelis GC, Angelaki DE. Flexible egocentric and allocentric representations of heading signals in parietal cortex. *Proc Natl Acad Sci U S A*. 2018;115:E3305–E3312.
- Cohen YE, Andersen RA. A common reference frame for movement plans in the posterior parietal cortex. *Nat Rev Neurosci*. 2002;3:553–562.
- Crane BT. Coordinates of human visual and inertial heading perception. *PLoS One*. 2015;10:e0135539.
- Crane BT. Effect of eye position during human visual-vestibular integration of heading perception. *J Neurophysiol*. 2017;118:1609–1621.
- Crowe DA, Averbeck BB, Chafee MV. Neural ensemble decoding reveals a correlate of viewer- to object-centered spatial transformation in monkey parietal cortex. *J Neurosci*. 2008;28:5218–5228.
- de Winkel KN, Soyka F, Barnett-Cowan M, Bulthoff HH, Groen EL, Werkhoven PJ. Integration of visual and inertial cues in the perception of angular self-motion. *Exp Brain Res*. 2013;231:209–218.
- de Haan MJ, Brochier T, Grun S, Riehle A, Barthelemy FV. Real-time visuomotor behavior and electrophysiology recording setup for use with humans and monkeys. *J Neurophysiol*. 2018;120:539–552.
- Dokka K, DeAngelis GC, Angelaki DE. Multisensory integration of visual and vestibular signals improves heading discrimination in the presence of a moving object. *J Neurosci*. 2015;35:13599–13607.
- Drugowitsch J, DeAngelis GC, Klier EM, Angelaki DE, Pouget A. Optimal multisensory decision-making in a reaction-time task. *Elife*. 2014;3:e03005.
- Fan RH, Liu S, DeAngelis GC, Angelaki DE. Heading tuning in macaque area V6. *J Neurosci*. 2015;35:16303–16314.
- Fetsch CR, Wang S, Gu Y, DeAngelis GC, Angelaki DE. Spatial reference frames of visual, vestibular, and multimodal heading signals in the dorsal subdivision of the medial superior temporal area. *J Neurosci*. 2007;27:700–712.
- Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci*. 2009;29:15601–15612.
- Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. Neural correlates of reliability-based cue weighting during multisensory integration. *Nat Neurosci*. 2011;15:146–154.
- Gu Y, Watkins PV, Angelaki DE, DeAngelis GC. Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *J Neurosci*. 2006;26:73–85.
- Gu Y, Angelaki DE, DeAngelis GC. Neural correlates of multisensory cue integration in macaque MSTd. *Nat Neurosci*. 2008;11:1201–1210.
- Gu Y, DeAngelis GC, Angelaki DE. Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. *J Neurosci*. 2012;32:2299–2313.
- Hartmann TS, Bremmer F, Albright TD, Kregelberg B. Receptive field positions in area MT during slow eye movements. *J Neurosci*. 2011;31:10437–10444.
- Kim B, Kenchappa SC, Sunkara A, Chang TY, Thompson L, Doudlah R, Rosenberg A. Real-time experimental control using network-based parallel processing. *Elife*. 2019;8.
- Klein SA. Measuring, estimating, and understanding the psychometric function: a commentary. *Percept Psychophys*. 2001;63:1421–1455.
- Lappe M, Bremmer F, van den Berg AV. Perception of self-motion from visual flow. *Trends Cogn Sci*. 1999;3:329–336.
- Leone FT, Monaco S, Henriques DY, Toni I, Medendorp WP. Flexible reference frames for grasp planning in human parietofrontal cortex(1,2,3). *eNeuro*. 2015;2:e0008–0015.
- Li CY, Guo K. Gaze-position-dependent activities of striate cortex (V1) neurones of awake macaque monkeys. *Sheng Li Xue Bao*. 1997;49:299–306.
- McGuire LM, Sabes PN. Heterogeneous representations in the superior parietal lobule are common across reaches to visual and proprioceptive targets. *J Neurosci*. 2011;31:6661–6673.
- Metzger RR, Mullette-Gillman OA, Underhill AM, Cohen YE, Groh JM. Auditory saccades from different eye positions in the monkey: implications for coordinate transformations. *J Neurophysiol*. 2004;92:2622–2627.
- Mullette-Gillman OA, Cohen YE, Groh JM. Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *J Neurophysiol*. 2005;94:2331–2352.
- Mullette-Gillman OA, Cohen YE, Groh JM. Motor-related signals in the intraparietal cortex encode locations in a hybrid, rather than eye-centered reference frame. *Cereb Cortex*. 2009;19:1761–1775.
- Ni J, Tatalovic M, Straumann D, Olasagasti I. Gaze direction affects linear self-motion heading discrimination in humans. *Eur J Neurosci*. 2013;38:3248–3260.
- Ong WS, Bisley JW. A lack of anticipatory remapping of retinotopic receptive fields in the middle temporal area. *J Neurosci*. 2011;31:10432–10436.
- Pesaran B, Nelson MJ, Andersen RA. Dorsal premotor neurons encode the relative position of the hand, eye, and goal during reach planning. *Neuron*. 2006;51:125–134.
- Prins N, Kingdom FAA. Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Front Psychol*. 2018;9:1250.
- Prsa M, Gale S, Blanke O. Self-motion leads to mandatory cue fusion across sensory modalities. *J Neurophysiol*. 2012;108:2282–2291.
- Prsa M, Jimenez-Rezende D, Blanke O. Inference of perceptual priors from path dynamics of passive self-motion. *J Neurophysiol*. 2015;113:1400–1413.
- Rosenberg A, Cowan NJ, Angelaki DE. The visual representation of 3D object orientation in parietal cortex. *J Neurosci*. 2013;33:19352–19361.

- Sasaki R, Anzai A, Angelaki DE, DeAngelis GC. Flexible coding of object motion in multiple reference frames by parietal cortex neurons. *Nat Neurosci*. 2020;23:1004–1015.
- Schlack A, Sterbing-D'Angelo SJ, Hartung K, Hoffmann KP, Bremmer F. Multisensory space representations in the macaque ventral intraparietal area. *J Neurosci*. 2005;25:4616–4625.
- Schlicht EJ, Schrater PR. Impact of coordinate transformation uncertainty on human sensorimotor control. *J Neurophysiol*. 2007;97:4203–4214.
- Sereno MI, Huang RS. A human parietal face area contains aligned head-centered visual and tactile maps. *Nat Neurosci*. 2006;9:1337–1343.
- Shaikh AG, Meng H, Angelaki DE. Multiple reference frames for motion in the primate cerebellum. *J Neurosci*. 2004;24:4491–4497.
- Snyder LH, Grieve KL, Brotchie P, Andersen RA. Separate body- and world-referenced representations of visual space in parietal cortex. *Nature*. 1998;394:887–891.
- Weyand TG, Malpeli JG. Responses of neurons in primary visual cortex are modulated by eye position. *J Neurophysiol*. 1993;69:2258–2260.
- Wichmann FA, Hill NJ. The psychometric function: I. fitting, sampling, and goodness of fit. *Percept Psychophys*. 2001;63:1293–1313.
- Yang L, Gu Y. Distinct spatial coordinate of visual and vestibular heading signals in macaque FEFsem and MSTd. *Elife*. 2017;6:e29809.
- Yang Y, Liu S, Chowdhury SA, DeAngelis GC, Angelaki DE. Binocular disparity tuning and visual-vestibular congruency of multisensory neurons in macaque parietal cortex. *J Neurosci*. 2011;31:17905–17916.

# 1 Supporting Information

## 2 Eye fixation positions



3

### 4 **Figure S1. Eye position data during the stimulation period in the gaze condition of E. (A-C)**

5 Data (blue) from a representative subject are shown at three eye fixation positions (intersections

6 of the dashed lines) in the horizontal plane ( $-20^\circ$ ,  $0^\circ$  and  $20^\circ$ ). Each red dot indicates the mean

7 center of data at each fixation position. **(D-F)** The mean center of data is shown for all subjects.

8 The error bar indicates SEM. Columns are for stimuli: vestibular (Ves, A and D), visual (Vis, B

9 and E) and combined (Com, C and F). As shown here, the eye data basically fell within a  $5^\circ \times 5^\circ$

10 rectangular window in all the three stimulus conditions, although there is some offset between

11 the mean center of data and the fixation position, which is possibly due to systematic bias in the

12 eye calibration and drifting during the recording. Therefore, the eye tracking was not obviously

13 affected by vibration from the moving platform in the vestibular and combined conditions. As

14 the eye tracker was firmly secured to the platform, and vibrated simultaneously with the subject

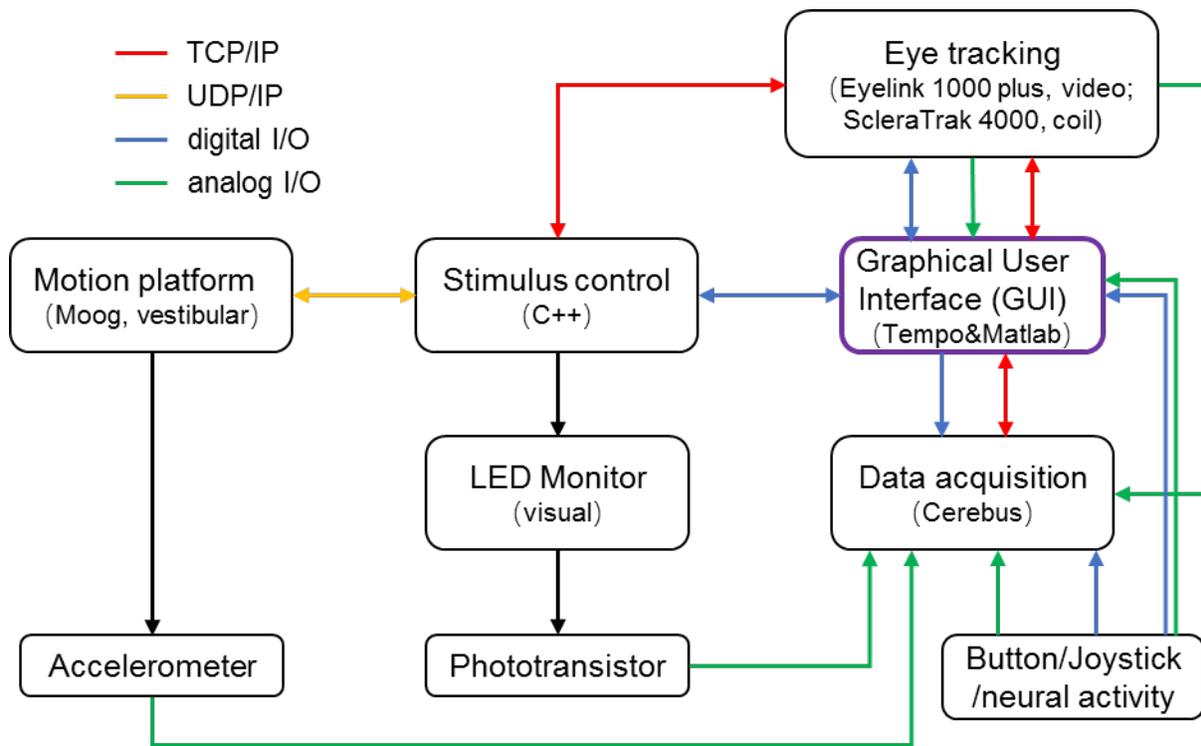
15 during movement, the resulting relative displacement between them should be small.

16

17

18 **Calibration and validation of experimental setup**

19



20

21 **Figure S2. The configuration of experimental setup.** Experimental control and monitoring are  
 22 achieved using a real-time GUI system (Tempo, Reflective Computing; purple box) that  
 23 coordinates components including: (1) stimulus presentation and control (visual and/or  
 24 vestibular); (2) data acquisition; (3) external input/output devices. Arrows indicate the direction  
 25 of information flow.

26

27

28

29

30

31

32

33

34

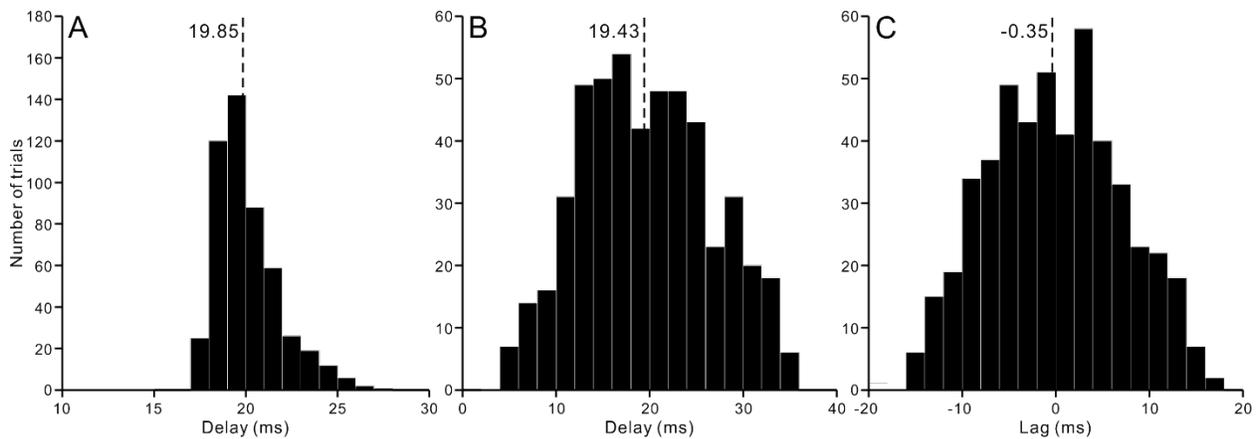
35

36

As shown in Figure S2, our experimental setup was controlled by Tempo (Reflective Computing), a hard real-time system executing each operation in our experimental protocols precisely within predefined time limits by coordinating the other components. The “Stimulus control” component was implemented using C++ and OpenGL (Open Graphics Library) based on the code that has already been validated (Gu et al. 2006) and widely used in many previous publications (Chen et al. 2013, 2013, 2014, 2018). In the component, the visual stimulus was rendered and synchronized with the vestibular stimulus (movement of the motion platform) in a specific thread, whose priority was set to the highest level in the C++ code to minimize delay and variation in the presentation and synchronization of stimuli. The motion platform (MB-E-6DOF2000E, Moog) was controlled by a real-time operation system communicating with the “Stimulus control” component at 1000 Hz, so positions of the platform were updated with 1ms

37 time accuracy in theory. The “Stimulus control” component also had input and/or output  
 38 connections with external devices for eye tracking, data acquisition and responses from subjects.  
 39 In addition, an accelerometer (ADXL354C, Analog Devices) was secured on the platform to  
 40 measure the real-time acceleration during the movement, and a phototransistor circuit (OPT101,  
 41 Texas Instruments) was attached on the monitor to detect the real-time refreshing of the visual  
 42 stimulus by detecting a small bright patch showing on the monitor screen. Analog signals from  
 43 these two sensors or neural activity, digital signals from button/joystick and command signal  
 44 from the GUI component were all connected to the data acquisition component and sampled  
 45 simultaneously.

46



47

48 **Figure S3. Quantification of the fidelity in presentation of the stimuli. (A)** The distribution of  
 49 delay between the command signaling the platform to move to the detected movement by the  
 50 accelerometer in the vestibular condition (n=500 trials). **(B)** The distribution of delay between  
 51 the command signaling the presentation of visual stimulus to the detected stimulus shown on the  
 52 monitor by the phototransistor (n=500) in the visual condition. **(C)** The distribution of lag  
 53 between the detected movement of platform to the detected visual stimulus (n=500) in the  
 54 combined condition.

55 As reported previously (Kim et al. 2019), before any experiments, the fidelity of stimulus  
 56 presentation in our system was assessed when the visual and vestibular stimuli were presented  
 57 individually and simultaneously. In the vestibular only condition, when the command of “Move”  
 58 sent from Tempo was received in the component of “Stimulus control”, the specific thread with  
 59 the highest priority began to update the position of motion platform in a pre-calculated trajectory  
 60 at 1000 Hz, and the platform started to move after receiving a new position. Delay from the  
 61 command to starting point of the movement (SPM) was calculated as the difference of their time

62 in the data recorded by the data acquisition component. The SPM was defined as the first time  
63 point in the recorded accelerometer data after the command, where the acceleration was 5  
64 standard deviation above the average ( $-0.41+5\times 1.6\text{ mm/s}^2$ ) of baseline (500ms before the  
65 command). A distribution of the delay was constructed from 500 trials in the vestibular condition  
66 of current experiments, as the accelerometer was secured on the platform since the setup was  
67 built, so the acceleration data were always recorded during experiments. There was a short  
68 average delay ( $19.85\pm 1.75\text{ms}$ , [17, 27] for [min, max]; Fig. S3A). However, considering there  
69 was a theoretical delay of the accelerometer and its peripheral circuit, the measured delay should  
70 be even small.

71 In the visual only condition, in the specific thread with the highest priority, the optic flow  
72 was rendered with native APIs of OpenGL and “flipped” to the front buffer in the professional  
73 graphics card (NVIDIA Quadro K2000) for presenting on the monitor after receiving the “Move”  
74 command. A distribution of the delay between the command and the detected visual presentation  
75 (DVP) by the phototransistor was constructed from 500 trials. There was a short average delay  
76 ( $19.43\pm 6.96\text{ms}$ , [4, 35.5] for [min, max]; Fig. S3B). There were several possible factors  
77 contributing to this delay. One was from refresh rate of the monitor (60 Hz), which could  
78 introduce a maximum delay of 16.7ms. Second was from the time for rendering the optic flow  
79 stimuli. Other factors include the dispatch of processes and threads by the operating system, as  
80 well as communication between the components.

81 In order synchronize the visual and vestibular stimuli when they were presented  
82 simultaneously in the combined condition, as reported previously (Gu *et al.* 2006), the platform  
83 motion was predicted and synchronized with the visual motion in the specific thread. A  
84 distribution of the lags between DVP and SPM was constructed from 500 trials. The average lag  
85 ( $-0.35\pm 7.22\text{ms}$ , [-15.5, 16.5] for [min, max]; Fig. S3C) was less than 1ms, although there was  
86 still a variation within 20ms. Actually, we have already tried to use the phototransistor signal to  
87 synchronize the movement of platform, but the results were not good, including breaking  
88 communications with Moog, although the delay and its variation became smaller.

89 Therefore, although there were delays and lags in our system, considering the relative long  
90 duration (2000ms) of stimuli in our experiments, such delays should not have significant effects  
91 on the results.

92

93 **The extended diffusion model**

94 **Model description.** For each subject, the data of choices and reaction time (RT) in each gaze  
 95 condition were fitted simultaneously with the extended diffusion model (EDM) reported  
 96 previously (Drugowitsch et al. 2014). For the RT paradigm of our heading discrimination task,  
 97 the EDM would operate as a drifting and diffusing “particle”, whose dynamics represent a noisy  
 98 sensory signal of heading, called the momentary evidence and denoted by  $\dot{x} = b(t)k\sin(h) +$   
 99  $\eta(t)$ , where  $b(t)$  indicates a physical quantity encoding how the reliability of this heading signal  
 100 changes over time (acceleration  $a(t)$  for the vestibular modality, velocity  $v(t)$  for the visual  
 101 modality),  $k$  determines how effectively/sensitively each subject can make use of the incoming  
 102 information,  $h$  is the heading direction,  $\eta(t)$  is a unit variance Gaussian white noise process.  
 103 The momentary evidence in the combined condition ( $\dot{x}_{com}$ ) is computed by optimally integrating  
 104 those in the visual ( $\dot{x}_{vis}$ ) and vestibular ( $\dot{x}_{ves}$ ) conditions by weighting them in proportion to  
 105 their relative sensitivities ( $k_{vis}$  and  $k_{ves}$ ):

106 
$$\dot{x}_{vis} = v(t)k_{vis}\sin(h) + \eta_{vis}(t) \quad (1)$$

107 
$$\dot{x}_{ves} = a(t)k_{ves}\sin(h) + \eta_{ves}(t) \quad (2)$$

108 
$$\dot{x}_{com} = \sqrt{\frac{k_{vis}^2}{k_{vis}^2+k_{ves}^2}} \dot{x}_{vis} + \sqrt{\frac{k_{ves}^2}{k_{vis}^2+k_{ves}^2}} \dot{x}_{ves} \quad (3)$$

109 
$$k_{com} = \sqrt{k_{vis}^2 + k_{ves}^2} \quad (4)$$

110 
$$\dot{x}_{com} = d(t)k_{com}\sin(h) + \eta_{com}(t) \quad (5)$$

111 The momentary visual and vestibular evidence were temporally weighted by the time course of  
 112  $v(t)$  and  $a(t)$  respectively:  $\dot{X}_{vis} = v(t)\dot{x}_{vis}$  and  $\dot{X}_{ves} = a(t)\dot{x}_{ves}$ , and optimally integrated in the  
 113 combined condition:  $\dot{X}_{com} = d(t)\dot{x}_{com}$ , where the sensitivity profile  $d(t)$  is a weighted  
 114 combination of the unimodal sensitivity profiles:

115 
$$d(t) = \sqrt{\frac{k_{vis}^2}{k_{com}^2} v^2(t) + \frac{k_{ves}^2}{k_{com}^2} a^2(t)} \quad (6)$$

116 The EDM assumes that the particle starts at  $x(0)=0$ , drifts according to the heading evidence  
 117 accumulated optimally over time, and diffuses with an average slope given by  $b(t)k\sin(h)$   
 118 until it hits either the upper bound  $\theta$  or the lower bound  $-\theta$ , corresponding to rightward and  
 119 leftward choices, respectively. The decision time is the time when the particle hits a bound. The  
 120 RT is the sum of the decision time and the non-decision time (a model parameter, see below).

121 As the heading discrimination task requires identifying the sign of  $h$ , which, for the small  $h$   
 122 used in the task, equals the sign of  $\sin(h)$ , optimal decisions can be performed by computing the  
 123 posterior over  $\sin(h)$ , given all available information. After observing the stimulus for  $t$   
 124 seconds, and under the assumption of a uniform prior, this posterior is given by Bayes rule:

$$125 \quad p(\sin(h) | \delta x_{vis,1:t}) = p(\sin(h) | X_{vis}(t), V(t)) = N \left\{ \sin(h) \mid \frac{X_{vis}(t)}{k_{vis}V(t)}, \frac{1}{k_{vis}^2 V(t)} \right\} \quad (7)$$

126 where  $\delta x_{vis,1:t}$  are the momentary visual evidences across all time steps up to time  $t$ ,  
 127  $X_{vis}(t) = \sum_{n \in 1:t} v_n \delta x_{vis,n}$  is the velocity-weighted evidence,  $v_n$  is the velocity at time step  $n$ ,  
 128  $\delta x_{vis,n} \sim N(v_n k_{vis} \sin(h) \Delta, \Delta)$  is Gaussian with mean  $v_n k_{vis} \sin(h) \Delta$  and variance  $\Delta$ , and  
 129  $V(t) = \sum_{n \in 1:t} v_n^2 \Delta$ . Consequently, the belief about ‘rightward’ being correct can be fully  
 130 expressed by  $X_{vis}(t)$  and  $V(t)$ . This shows that, in the visual condition, the optimal  
 131 accumulation of visual heading evidence with a single-particle EDM with time-varying evidence  
 132 sensitivity requires the momentary evidence to be weighted by its momentary sensitivity. In the  
 133 vestibular condition, a similar formulation holds for the posterior over heading, but the vestibular  
 134 signal is assumed to be weighted by the temporal profile of stimulus acceleration, instead of  
 135 velocity. In the combined condition, a two-particle EDM optimally integrate the visual and  
 136 vestibular heading signals, and the posterior probability of  $\sin(h)$  is given by:

$$137 \quad p(\sin(h) | \delta x_{com,1:t}) = p(\sin(h) | \delta x_{vis,1:t}, \delta x_{ves,1:t}) \quad (8)$$

138 where  $\delta x_{com,1:t}$  is the sequence of momentary evidence in the combined condition, following  
 139  $\delta x_{com,n} \sim N(d_n k_{com} \sin(h) \Delta, \Delta)$ .  $k_{com}$  and  $d(t)$  are defined by equations (4 and 6)  
 140 respectively.

141 ***The psychometric function and discrimination threshold of EDM.*** For the EDM, the  
 142 psychometric function is formed by plotting how the fraction of choosing one of the two options  
 143 changes as a function of heading. This fraction is the probability given by the logistic sigmoid  
 144 (Drugowitsch *et al.* 2014):

$$145 \quad p(x(T) = \theta | \sin(h) = H, T, X(T) = \pm\theta) = \frac{1}{1 + e^{-2kH\theta}} \quad (9)$$

146 for some heading  $H$ , and at some decision time  $T$  at which a boundary has been reached (i.e.  
 147  $X(T) = \pm\theta$ ), that the particle has reached the upper boundary,  $X(T) = \theta$ . By fitting a  
 148 cumulative Gaussian  $\Phi(\frac{H}{\sigma})$  with threshold  $\sigma$  to the psychometric function, the heading  
 149 discrimination threshold is approximately determined:

150 
$$\sigma \approx \frac{\pi}{\sqrt{12k\theta}} \quad (10)$$

151 which is inversely proportional to the sensitivity  $k$  and the bound height  $\theta$ . The  $\sigma$  was  
 152 multiplied by 100 to match the dimension of threshold computed with the traditional optimal  
 153 integration model.

154 **Model parameterization.** In the visual condition, the behavior is modeled by an EDM with  
 155 sensitivity  $k_{vis}$ , time-course  $v(t)$ , and bound  $\theta_{vis}$ , and the model is parameterized by  
 156  $\{k_{vis}, \theta_{vis}\}$ .  $v(t)$  followed a Gaussian profile (Methods in the main text). In the vestibular  
 157 condition, an EDM models the behavior with sensitivity  $k_{ves}$ , time-course  $a(t)$ , and bound  
 158  $\theta_{ves}$ , and the model is parameterized by  $\{k_{ves}, \theta_{ves}\}$ .  $a(t)$  was given by taking the derivative of  
 159  $v(t)$ . In the combined condition, given that visual and vestibular signals are integrated optimally,  
 160 and the sensitivity to the evidence  $k_{com}$  and its time-course  $d(t)$  are determined by the  
 161 equations (4 and 6) respectively, its EDM is parameterized solely by the bound  $\{\theta_{com}\}$ .

162 The EDM assumes that RTs featured by the subjects are composed of the decision time as  
 163 predicted by the diffusion model, and a non-decision time that captures the initial stimulus  
 164 processing delay and the motor preparation time. The non-decision time is captured by on  
 165 parameter per stimulus condition (visual, vestibular and combined), three in total:  
 166  $\{t_{nd,vis}, t_{nd,ves}, t_{nd,com}\}$ . A lapse probability parameter  $\{p_{lapse}\}$  was introduced to account for  
 167 random choices (with probability  $\frac{1}{2}$  for each motion direction) due to accidental button presses  
 168 or lapses of attention. A potential bias in heading perception (i.e., horizontal shift of the  
 169 psychometric function) is captured by the bias parameter  $\{\tilde{h}\}$  for each gaze position and  
 170 stimulus condition.

171 Overall, given that the EDM predicts mean decision times represented by  
 172  $t_{DM,corr}(h, g, cond, \varphi)$  and  $t_{DM,incorr}(h, g, cond, \varphi)$  for correct and incorrect decisions,  
 173 respectively, with model parameters  $\varphi$ , and given that the probability of choosing ‘rightward’ for  
 174 each combination of heading direction  $h$ , gaze position  $g \in \{-20^\circ, 0^\circ, 20^\circ\}$  and stimulus  
 175 condition  $cond \in \{vis, vest, comb\}$  is represented by  $p_{DM,r}(h, g, cond)$ , we assumed that the  
 176 subject would feature mean RTs and choice probabilities given by:

177 
$$\left. \begin{aligned} t_{corr}(h, g, cond, \varphi) &= t_{DM,corr}(h + \tilde{h}, g, cond, \varphi) + t_{nd,cond}, \\ t_{incorr}(h, g, cond, \varphi) &= t_{DM,incorr}(h + \tilde{h}, g, cond, \varphi) + t_{nd,cond}, \\ p_r(h, g, cond, \varphi) &= (1 - p_{lapse})p_{DM,r}(h + \tilde{h}, g, cond, \varphi) + \mathbb{1}_{lapse} \frac{1}{2} \end{aligned} \right\} \quad (11)$$

178 The EDM itself and the non-decision times were parameterized by 8 parameters  
 179  $\{k_{vis}, \theta_{vis}, k_{ves}, \theta_{ves}, \theta_{com}, t_{nd,vis}, t_{nd,ves}, t_{nd,com}\}$ , and an additional 10 parameters captured the  
 180 biases and lapse rates.

181 **Model fitting.** The EDM assumes that the fraction of correct choices follows a binomial  
 182 distribution, and likelihood of the EDM parameters  $\varphi$  to describe the fraction of rightward  
 183 choices by:

$$184 \quad L_{r,h,g,cond}(\varphi) = Bin(\hat{p}_r(h, g, cond)n_{h,g,cond} | n_{h,g,cond}, p_r(h, g, cond, \varphi)) \quad (12)$$

185 where  $\hat{p}_r(h, g, cond)$  is the observed number of rightward choices,  $n_{h,g,cond}$  is the number of  
 186 trials,  $p_r(h, g, cond, \varphi)$  is the model prediction. The EDM assumes that the RTs of correct and  
 187 incorrect choices were distributed according to a Gaussian centered on the empirical mean and  
 188 spread according to the standard error. The likelihood describing the RTs for correct choices was  
 189 given by the Gaussian:

$$190 \quad L_{corr,h,g,cond}(\varphi) = N\left\{\hat{t}_{corr}(h, g, cond) | t_{corr}(h, g, cond, \varphi), \frac{var_{corr}(h, g, cond, \varphi)}{n_{corr,h,g,cond}}\right\} \quad (13)$$

191 where  $\hat{t}_{corr}(h, g, cond)$  is the observed mean RT over the  $n_{h,g,cond}$  trials,  
 192  $t_{corr}(h, g, cond, \varphi)$  is the mean RT predicted by the EDM,  $var_{corr}(h, g, cond, \varphi)$  is the  
 193 variance of the prediction.  $L_{incorr,h,g,cond}(\varphi)$  is an analogous term for those corresponding to  
 194 incorrect choices. Model predictions were found by numerically evaluating the equation (11),  
 195 and the RT distributions were evaluated in steps of 5ms (Drugowitsch *et al.* 2014). Based on  
 196 these distributions, the probability of a choosing ‘rightward’ and the mean and variance of the  
 197 RTs were computed.

198 Overall, the complete likelihood was given by:

$$199 \quad L(\varphi) \prod_{h,g,cond} L_{r,h,g,cond}(\varphi) L_{corr,h,g,cond}(\varphi) L_{incorr,h,g,cond}(\varphi) \quad (14)$$

200 The model was fitted to the behavior data (choices and RT) in each gaze condition and each  
 201 subject by finding the model parameters  $\varphi$  that maximized the likelihood given the observed  
 202 behavior. The parameter vector found by gradient ascent on the log-likelihood was used as initial  
 203 sample for taking 44000 samples from the Bayesian parameter posterior by Markov Chain  
 204 Monte Carlo methods (assuming a bounded uniform parameter prior), and then the highest-  
 205 likelihood sample was as a starting point for another gradient ascent step to find the posterior’s  
 206 mode, which was used as the maximum likelihood parameter vector (Drugowitsch *et al.* 2014).

207 The overall goodness of fit was estimated by  $R^2(\varphi) = \frac{1}{2}(R_{psych}^2(\varphi) + R_{chron}^2(\varphi))$ , where  
 208  $R_{psych}^2(\varphi)$  and  $R_{chron}^2(\varphi)$  are the adjusted coefficients of determination for the psychometric  
 209 and the chronometric curves, respectively:

$$210 \left. \begin{aligned} R_{psych}^2(\varphi) &= \tilde{R}_{psych}^2(\varphi) - \left(1 - \tilde{R}_{psych}^2(\varphi)\right) \frac{k}{N_s - k - 1} \\ \tilde{R}_{psych}^2(\varphi) &= 1 - \frac{\sum_{h,g,cond} w_{h,g,cond} (\hat{p}_r(h,g,cond) - p_r(h,g,cond,\varphi))^2}{\sum_{h,g,cond} w_{h,g,cond} (\hat{p}_r(h,g,cond) - \bar{p}_r)^2} \end{aligned} \right\} \quad (15)$$

211 where  $\hat{p}_r(h, g, cond)$  and  $p_r(h, g, cond, \varphi)$  are the same terms as in the above likelihood,  $\bar{p}_r$   
 212 is the mean probability of choosing right over all trials,  $w_{h,g,cond}$  is the fraction of trials with  
 213 heading  $h$ , gaze position  $g$  and stimulus condition  $cond$ ,  $k$  is the number of model parameters,  
 214 and  $N_s$  is the number of trials performed by subject  $s$ .

$$215 \left. \begin{aligned} R_{chron}^2(\varphi) &= \tilde{R}_{chron}^2(\varphi) - \left(1 - \tilde{R}_{chron}^2(\varphi)\right) \frac{k}{N_s - k - 1} \\ \tilde{R}_{chron}^2(\varphi) &= 1 - \frac{\sum_{h,g,cond} \left( w_{corr,h,g,cond} (\hat{t}_{corr}(h,g,cond) - t_{corr}(h,g,cond,\varphi))^2 + w_{incorr,h,g,cond} (\hat{t}_{incorr}(h,g,cond) - t_{incorr}(h,g,cond,\varphi))^2 \right)}{\sum_{h,g,cond} \left( w_{corr,h,g,cond} (\hat{t}_{corr}(h,g,cond) - \bar{t})^2 + w_{incorr,h,g,cond} (\hat{t}_{incorr}(h,g,cond) - \bar{t})^2 \right)} \end{aligned} \right\} \quad (16)$$

216 where  $\hat{t}_{incorr}(h, g, cond)$  and  $t_{incorr}(h, g, cond, \varphi)$  as well as their counterpart for incorrect  
 217 trials are again the same terms as in the likelihood,  $\bar{t}$  is the mean RT over all trials,  $w_{h,g,cond}$   
 218 and  $w_{incorr,h,g,cond}$  are the fractions of correct and incorrect trials.

219

220

221 **Traditional optimal integration model**

222 Based on the optimal cue integration theory (Landy et al. 2011), the linear ideal-observer  
 223 model postulated that heading performance in the combined stimulus condition was decided by  
 224 an internal heading signal  $S_{com}$  that is a weighted sum of vestibular and visual heading signals  
 225  $S_{ves}$  and  $S_{vis}$  (Fetsch et al. 2011; Crane 2017):

226 
$$S_{com} = w_{ves} * S_{ves} + w_{vis} * S_{vis} \quad (17)$$

227 
$$w_{vis} = 1 - w_{ves} \quad (18)$$

228 Assuming each  $S$  is a Gaussian random variable with mean  $\mu$  and variance  $\sigma^2$ , weights of the  
 229 optimal model are estimated by the cue reliability or precision:

230 
$$w_{ves} = \frac{1/\sigma_{ves}^2}{1/\sigma_{ves}^2 + 1/\sigma_{vis}^2}; w_{vis} = \frac{1/\sigma_{vis}^2}{1/\sigma_{ves}^2 + 1/\sigma_{vis}^2} \quad (19)$$

231 
$$\sigma_{com} = \sqrt{\frac{\sigma_{ves}^2 \sigma_{vis}^2}{\sigma_{ves}^2 + \sigma_{vis}^2}} \quad (20)$$

232 where  $\sigma_{ves}$  and  $\sigma_{vis}$  are thresholds of psychometric functions in the visual and vestibular  
 233 stimulus conditions respectively. To compare behavior data to optimal predictions, the optimal  
 234 threshold  $\sigma_{com}$  is calculated by single-cue thresholds in equation (20).

235 
$$\mu_{ves} = \theta + b_{ves}; \mu_{vis} = \theta + b_{vis}; \mu_{com} = \theta + b_{com} \quad (21)$$

236 
$$\mu_{com} = w_{ves} * \mu_{ves} + w_{vis} * \mu_{vis} \quad (22)$$

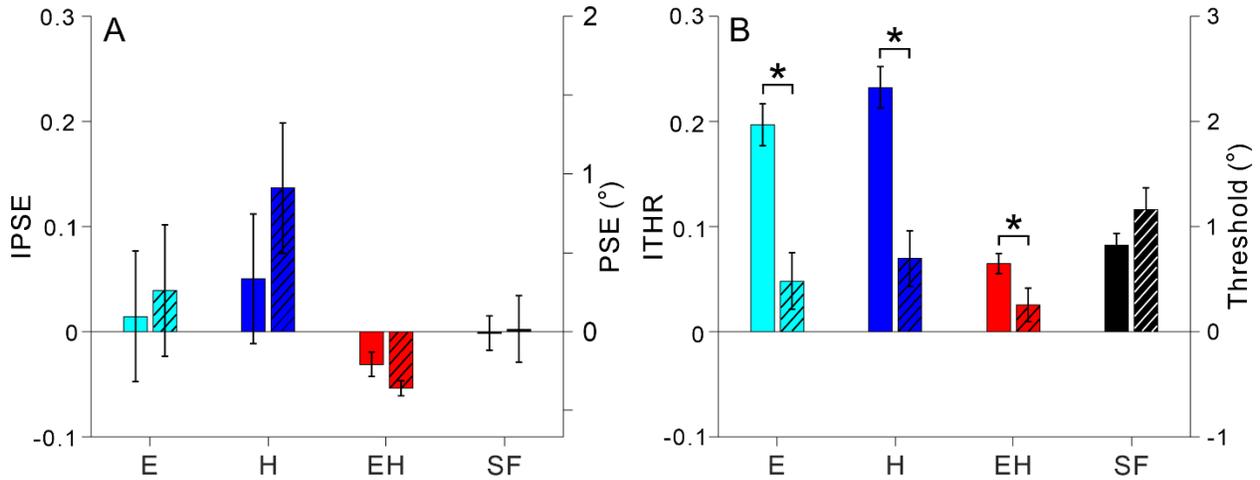
237 
$$PSE_{com} = w_{ves} * PSE_{ves} + w_{vis} * PSE_{vis} \quad (23)$$

238 Under Gaussian distribution, the internal signal  $S$  can be estimated from behavior data by  
 239 equation (21), where  $b_{ves}, b_{vis}$  and  $b_{com}$  are bias terms (assumed to be independent of the  
 240 heading direction  $\theta$ ) (Fetsch *et al.* 2011; Crane 2017). These biases, when estimated from  
 241 behavioral data, are equal to  $-PSE_{ves}, -PSE_{vis}$  and  $-PSE_{com}$  respectively. Taking the mean of  
 242 both sides in equation (18) yields equation (22). Then, substituting equation (18 and 21) into  
 243 equation (22), we have equation (23) used to calculate the optimal  $PSE_{com}$  in the combined  
 244 condition.

245 Therefore, the predicted PSE ( $PSE_{com}$ ) and threshold ( $\sigma_{com}$ ) in the combined condition  
 246 using unimodal PSE and threshold in the visual and vestibular conditions were used to calculate  
 247 the predicted IPSE and ITHR values separately. These predicted values were then compared to  
 248 the IPSE and ITHR values calculated using the observed PSE and threshold of psychometric  
 249 functions in the combined condition to test if subjects integrate the visual and vestibular heading

250 signals optimally (Fig. S4).

251



252  
253

254 **Figure S4. Deviation from traditional optimal cue integration model by gaze change. (A)**  
255 The average PSE measured (filled black bars) and predicted by the optimal model (hatched black  
256 bars) at  $[0^\circ, 0^\circ]$  gaze position in the combined stimulus condition is shown by the right axis (SF:  
257 straightforward, black psychometric curves in Fig. 2). The average IPSE measured (filled color  
258 bars) and predicted (hatched color bars) in the combined condition is shown by the left axis.  
259 Cyan, blue and red indicate E, H and EH conditions respectively. The error bar indicates SEM.  
260 Stars indicate significance ( $p < 0.05$ ). **(B)** The average threshold and ITHR (measured and  
261 predicted) are shown. Format is as in A.

262

263

264 **References**

265 Chen X, DeAngelis GC, Angelaki DE. 2013. Diverse spatial reference frames of vestibular signals in parietal  
266 cortex. *Neuron*. 80:1310-1321.

267 Chen X, DeAngelis GC, Angelaki DE. 2013. Eye-centered representation of optic flow tuning in the ventral  
268 intraparietal area. *J Neurosci*. 33:18574-18582.

269 Chen X, DeAngelis GC, Angelaki DE. 2014. Eye-centered visual receptive fields in the ventral intraparietal  
270 area. *J Neurophysiol*. 112:353-361.

271 Chen X, DeAngelis GC, Angelaki DE. 2018. Flexible egocentric and allocentric representations of heading  
272 signals in parietal cortex. *Proc Natl Acad Sci U S A*. 115:E3305-E3312.

273 Crane BT. 2017. Effect of eye position during human visual-vestibular integration of heading perception. *J*  
274 *Neurophysiol*. 118:1609-1621.

275 Drugowitsch J, DeAngelis GC, Klier EM, Angelaki DE, Pouget A. 2014. Optimal multisensory decision-making  
276 in a reaction-time task. *Elife*. 3.

277 Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE. 2011. Neural correlates of reliability-based cue weighting  
278 during multisensory integration. *Nat Neurosci*. 15:146-154.

279 Gu Y, Watkins PV, Angelaki DE, DeAngelis GC. 2006. Visual and nonvisual contributions to three-dimensional  
280 heading selectivity in the medial superior temporal area. *J Neurosci*. 26:73-85.

281 Kim B, Kenchappa SC, Sunkara A, Chang TY, Thompson L, Doudlah R, Rosenberg A. 2019. Real-time  
282 experimental control using network-based parallel processing. *Elife*. 8.

283 Landy MS, Banks MS, Knill DC. 2011. Ideal-observer models of cue integration. In: *Sensory Cue Integration*  
284 (eds. Trommershäuser, J., Kording, K.P. & Landy, M.S.) Oxford University Press, New York.

285

286