RA
SP:m v 1.30 Prn:13/07/2011; 8:57 sp2346 P.1(1-18)
by:Milda p. 1
SEEING AND PERCEIVING

BRILL *Seeing and Perceiving 0 (2011) 1–18* brill.nl/sp

The Role of Stereo Vision in Visual–Vestibular Integration *

John S. Butler [1,2,**], **Jennifer L. Campos** [1,3,4,5], **Heinrich H. Bülthoff** [1,6,**] **and Stuart T. Smith** [7]

[1] Max-Planck Institute for Biological Cybernetics, Spemannstrasse 38, Tübingen 72076, Germany
[2] Department of Pediatrics, Albert Einstein College of Medicine, Bronx, New York 10461, USA
[3] Toronto Rehabilitation Institute, Toronto, Canada
[4] Department of Psychology, University of Toronto, Toronto, Canada
[5] Centre for Vision Research, York University, Toronto, Canada
[6] Department of Brain and Cognitive Engineering, Korea University, South Korea
[7] Neuroscience Research Australia, Randwick, NSW 2031, Australia

Received 14 January 2011; accepted 9 June 2011

Abstract

Self-motion through an environment stimulates several sensory systems, including the visual system and the vestibular system. Recent work in heading estimation has demonstrated that visual and vestibular cues are typically integrated in a statistically optimal manner, consistent with Maximum Likelihood Estimation predictions. However, there has been some indication that cue integration may be affected by characteristics of the visual stimulus. Therefore, the current experiment evaluated whether presenting optic flow stimuli stereoscopically, or presenting both eyes with the same image (binocularly) affects combined visual–vestibular heading estimates.

Participants performed a two-interval forced-choice task in which they were asked which of two presented movements was more rightward. They were presented with either visual cues alone, vestibular cues alone or both cues combined. Measures of reliability were obtained for both binocular and stereoscopic conditions.

Group level analyses demonstrated that when stereoscopic information was available there was clear evidence of optimal integration, yet when only binocular information was available weaker evidence of cue integration was observed. Exploratory individual analyses demonstrated that for the stereoscopic condition 90% of participants exhibited optimal integration, whereas for the binocular condition only 60% of participants exhibited results consistent with optimal integration. Overall, these findings suggest that stereo vision may be important for self-motion perception, particularly under combined visual–vestibular conditions.

* This article is part of the Multisensorial Perception Collection, guest edited by S. Wuerger, D. Alais and M. Gondan.

** To whom correspondence should be addressed. E-mail: john.butler@einstein.yu.edu; heinrich.buelthoff@tuebingen.mpg.de

 DOI:10.1163/187847511X588070

Keywords
Stereo vision, multi-sensory integration, self-motion perception, maximum likelihood estimation, vestibular

1. Introduction

1.1. Visual–Vestibular Integration for Heading Perception

During self-motion through space several different sensory systems provide information about travelled distance, speed and direction of movement (i.e., heading), including important visual and vestibular information. Optic flow is the stream of retinal information generated during self-movement through space, while vestibular information is provided through the inner ear organs (otoliths and semicircular canals), which provide information about changing velocities. In the context of heading in particular, past research has demonstrated that both optic flow (Lappe *et al*., 1999; Royden *et al*., 1992; Warren and Hannon, 1990) and vestibular information (Butler *et al*., 2010; Fetsch *et al*., 2009; Gu *et al*., 2007, 2008b, 2010; Ohmi, 1996; Telford *et al*., 1995) can be used independently to judge heading. However, until very recently, there has been little understanding of how these two sources of sensory information are integrated in the brain. Based on maximum likelihood estimation (MLE) models, new evidence from both humans and non-human primates demonstrates that visual and vestibular information are typically combined in a 'statistically optimal fashion' (Butler *et al*., 2010; Fetsch *et al*., 2009; Gu *et al*., 2008b). Specifically, both psychophysical measures and neural responses demonstrate a reduction in variance when visual and vestibular cue are combined (i.e., multisensory conditions), compared to the response patterns when either cue is available alone (i.e., unisensory conditions). Conversely, a recent paper by de Winkel *et al*. (2010) reported no reduction in variance for combined cue estimates as would be predicted by MLE. Importantly, however, there were differences between the visual stimulus presentation used by de Winkel *et al*. (2010) compared to past studies, which may account for these discrepant findings (as will be discussed in greater detail below).

Primate neurophysiological studies have shown that neurons in the medial superior temporal (MST) and ventral intraparietal (VIP) areas are tuned to specific patterns of visual motion typical of optic flow and also have response properties appropriate for encoding heading (Bremmer *et al*., 2002a, b; Britten and van Wezel, 1998, 2002; Duffy and Wurtz, 1991; Gu *et al*., 2007, 2008b, 2010; Heuer and Britten, 2004; Page and Duffy, 2003; Perrone and Stone, 1998). While MST has long been associated with responding preferentially to optic flow stimuli, recent groundbreaking studies have now demonstrated functional and behavioral links between MSTd and heading perception based solely on vestibular signals in the absence of vision (Fetsch *et al*., 2009; Gu *et al*., 2007, 2008b, 2010). These intriguing findings demonstrate that there are multisensory properties of heading detection that are observable at the neurophysiological level.

Even though we are beginning to gain an understanding of the extent to which visual and vestibular inputs are integrated, much remains to be investigated with respect to how different characteristics of each sensory input affect the integration process. For instance, it has previously been demonstrated that an introduction of additional depth cues can improve heading estimation when only visual information is available (van den Berg and Brenner, 1994); yet it is currently unclear how the inclusion of different types of visual cues relevant to self-motion perception also affects the way in which visual cues are integrated with non-visual cues, such as vestibular signals.

1.2. Role of Stereo Vision in Visual Self-Motion Perception

Most of the work investigating observers' abilities to estimate heading based on optic flow alone have used monocularly or binocularly viewed random dot flow fields. However, in order to properly interpret the magnitude of self-motion using optic flow alone, scaling must often be provided *via* additional depth cues (Frenz and Lappe, 2005, 2006; Lappe *et al*., 1999). There has also been a suggestion that the inclusion of depth information could help dissociate retinal motion associated with movements of the head, from motion associated with eye movements in order to accurately perceive visual self-motion information (Warren and Rushton, 2009). In one of the first and only demonstrations of the effect of stereo vision on heading perception, van den Berg and Brenner (1994) reported that by presenting a random dot optic flow stimulus in stereo, heading estimates were improved. Specifically, stereoscopic conditions were shown to be far more robust to decreases in the signal-to-noise ratio than were binocular conditions (both eyes were presented with the same image). Pictorial depth cues also improved performance but not to the extent observed with the introduction of stereoscopic cues.

Vection is the illusory sensation of physical self-motion induced by moving visual patterns and has also been shown to be affected by the presence or absence of stereoscopic information. Specifically, Palmisano (1996) reported that when viewing random dot optic flow displays depicting linear self-motion, earlier vection onset times and longer vection durations were observed for stereoscopic conditions compared to binocular or monocular conditions without stereo cues. The magnitude of this effect also appeared to be contingent on speed, with the stereoscopic advantage decreasing as a function of increasing optic flow speed.

Neurophysiological findings have also shown that neurons in MST (known to be associated with visual and vestibular heading perception) have a stereo sensitivity that could play a role in signaling the direction of self-motion (Roy and Wurtz, 1990; Roy *et al*., 1992). Further, heading tuning in MST neurons is improved when depth information is added to the visual scene (Upadhyay *et al*., 2000). These lines of evidence strongly suggest that the inclusion of stereoscopic visual information is important for self-motion perception in general, and heading specifically.

Despite the fact that there is now evidence to suggest that stereo is important for visual heading perception, it is not clear whether the absence of stereoscopic infor-

mation impacts performance when additional *non-visual information* is available. It is possible, for instance, that vestibular inputs could help scale the magnitude of optic flow, thus, reducing the importance of adding visual depth cues. If this were the case, it would be predicted that there would be no difference in heading estimates under combined visual–vestibular conditions when comparing visual stimuli with or without stereoscopic information. Alternatively, it is also possible that if the visual information does not reliably provide a compelling sense of self-motion that is consistent with vestibular inputs, the brain might not interpret these two sensory cues as originating from the same event and, therefore, integration may not occur (Kording *et al*., 2007; Sato *et al*., 2007; Wallace *et al*., 2004). Consequently, if stereo cues (or other depth cues) are needed to provide reliable visual information consistent with self-motion, visual–vestibular integration may not occur under non-stereoscopic (binocular) conditions. Interestingly, the only study to date that has failed to report the optimal integration of visual and vestibular cues for heading estimates (based on MLE), presented participants with non-stereoscopic visual input (de Winkel *et al*., 2010). Therefore, in the current experiment we evaluated whether presenting a random dot optic flow display stereoscopically compared to binocularly would affect visual–vestibular heading estimates. MLE models were used to make predictions about the optimal reduction in variance in combined cue estimates using the individual variances of the unisensory estimates.

2. Methods

2.1. Participants

Ten participants (six male) with normal or corrected-to-normal vision, including normal stereo vision (tested using the stereo fly test; http://www.stereooptical.com/html/stereo-test.html) completed the experiment for payment. Half of the participants completed the stereoscopic condition first, while the second half of participants completed the binocular condition first. All participants apart from two of the authors were naïve to the purpose of the experiment. The average age was 26 years (range 21–40). Participants gave their informed consent before taking part in the experiment, which was performed in accordance with the ethical standards specified by the 1964 Declaration of Helsinki.

2.2. Apparatus

This experiment was conducted in the Motion Lab at the Max Planck Institute for Biological Cybernetics which consists of a Maxcue 600, six-degree-of-freedom Stewart motion platform manufactured by Motion-Base PLC, UK (Fig. 1; see also von der Heyde, 2001, for a complete description). All visual motion information was displayed on a projection screen, with a field of view of $86^\circ \times 65^\circ$ and a resolution of 1400×1050 pixels with a refresh rate of 60 frames per second. Participants viewed the projection screen through an aperture, which reduced the field of view

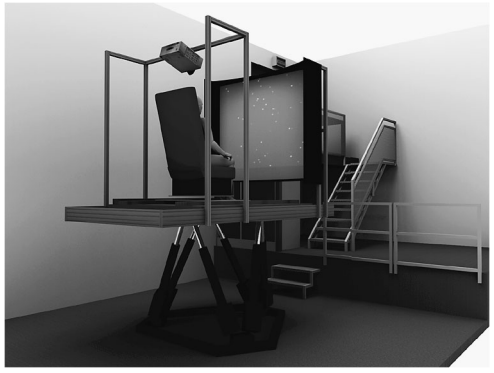

Figure 1. Apparatus. Participants were seated on the MPI Stewart motion platform and viewed the projection screen through an aperture, which reduced the field of view to 50° × 50°. Participants responded using a button box. The platform was surrounded by a black curtain to ensure that no cues relating to the spatial configuration of the surrounding laboratory space were available.

to 50° × 50°. This ensured that the edges of the screen were not visible, thereby increasing immersion and avoiding conflicting information provided by the stability of the frame and the visual motion being projected on the screen. The stereoscopic image was generated using red-cyan anaglyphs.

Participants wore noise-cancellation headphones with two-way communication capability and white noise was played to mask the noise of the platform. Subwoofers installed underneath the seat and foot plate were used to produce somatosensory vibrations to mask the platform motors. To keep head motion to a minimum, a foam head rest was used. Participants responded using a simple four-button response box. The entire experiment was coded using a graphical real-time interactive programming language (Virtools™, France).

2.3. Stimuli

The visual stimulus consisted of a limited lifetime starfield. Each star was a Gaussian blob and had a limited lifetime in the range of 0.5–1.0 s. The maximum number of Gaussian blobs on the screen at any one time was 200 and the minimum was 150. The participants were seated 100 cm from the screen. All blobs subtended angles ranging from 0.1° to 0.2°, which depended on their virtual distance ranging from 2 to 2.5 m. The starfield was presented either with or without stereoscopic depth cues (i.e., by viewing the starfield with or without the red-cyan passive stereo glasses). In the binocular condition, white blobs were presented on a black background, whereas in the stereo conditions a grey background was used to facilitate the fusing of the red and cyan blobs by minimizing ghost images. The vestibular stimuli were presented *via* the movement of the motion simulator on which participants were seated. For conditions in which only vestibular cues were available, passive movements were experienced in the complete absence of visual inputs to motion (i.e., in a completely darkened space covered in black cloth).

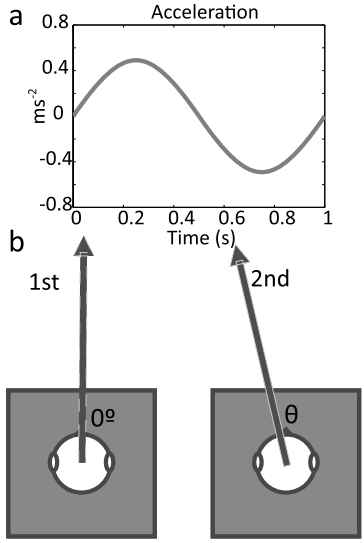

Figure 2. Participants were presented with two short movement intervals in different directions on each trial and were asked to judge in which interval they moved more to the right. (a) The sinusoidal acceleration profile for all motions. (b) Example trial: In the first interval, participants were moved along the standard heading of straight ahead (0°) and in the second interval they were moved in a leftwards direction.

The visual and vestibular heading motion profile (m) was

$$s(t) = 0.49\frac{(2\pi t - \sin(2\pi t))}{4\pi^2}, \quad 0 \leqslant t \leqslant 1 \text{ s}, \tag{1}$$

where t is time (Fig. 2(a)). All motion profiles had the same maximum forward displacement, velocity and acceleration of 0.078 m, 0.156 m/s and 0.49 m/s^2, respectively, which is above the detection threshold for blindfolded linear accelerations (Benson *et al.*, 1986). In pilot studies conducted with two participants, we manipulated the maximum acceleration to find a value such that the unisensory vestibular reliability was approximately the same as the unisensory visual reliability in order to most effectively reveal any effects of cue integration.

The linear direction of motion was defined by the angle of heading, θ, which was kept constant during each individual movement interval. Hence, the motion in the horizontal plane was defined as

$$\begin{aligned} x(t) &= s(t)\sin(\theta) \\ y(t) &= s(t)\cos(\theta), \end{aligned} \tag{2}$$

where $x(t)$ is the fore-aft direction and $y(t)$ is the lateral direction (Fig. 2(b)).

2.4. General Procedure

Participants performed a 2-interval forced choice task (2IFC) in which they were asked to judge in which of two movement intervals they moved more to the right

(see Fig. 2(b)). Each trial consisted of two linear heading motions, one being the standard heading angle and the other being one of several comparison headings angles. The standard and comparison heading angles were counterbalanced for order across trials (i.e., for 50% of trials the standard was first). The standard angle was always fixed at 0° (straight ahead), while there were eight comparison angles (−20°, −10°, −5°, −2°, 2°, 5°, 10°, 20°). The comparison angles were presented using the method of constant stimuli. All trials were initiated with a short auditory beep played over the headphones to indicate to the participant that they could start the trial with a button press. After pressing the start button, there was a 0.75 s pause before the onset of the motion. Between intervals, there was a 1 s pause, followed by a second auditory signal indicating the commencement of the second interval. After the second interval, the participants responded *via* the button box. A left button press indicated that they judged the first motion to be more to the right and a right button press indicated that they judged the second motion to be more to the right. Each participant completed two conditions — a stereoscopic condition and a binocular condition (blocked and counterbalanced). Within each of these conditions there were three trial types, including, vision alone trials (VIS), vestibular alone trials (VEST) and visual–vestibular cues combined trials (VIS–VEST) (blocked and counterbalanced).

For the VIS and VIS–VEST trials, the limited life-time Gaussian starfield appeared and remained static for 0.75 s before the onset of the motion. In the VEST trials there was a 0.75 s pause before the onset of the motion to ensure that the length of each trial was equal across all cue conditions. In the VEST and VIS–VEST trials, after responding, the participants were moved back to the start position in darkness at a constant, sub-threshold velocity of 0.025 m/s (Benson *et al*., 1986) for approximately six seconds before the next trial was initiated.

In total, participants completed at least 30 repetitions of each of the 8 comparison heading stimuli for each condition (240 trials). These trials were divided into blocks of 80 and each participant completed fifteen blocks over six days (approx. 2.5 blocks per day). Each block took approximately 20 min to complete. A preliminary experimental session was used to familiarize the participants with the stimuli and setup and these data are not reported here.

2.5. Data Analysis

The proportion of rightward responses made by participants were plotted as a function of heading angle, and cumulative Gaussian psychometric functions were fitted using the psignifit toolbox (Wichmann and Hill, 2001a, b; see Fig. 3 for a representative example of data from one participant in the stereo condition). The just noticeable difference (JND) was calculated, which is proportional to the standard deviation, σ, of the probability density function

$$JND = \sigma\sqrt{2}. \tag{3}$$

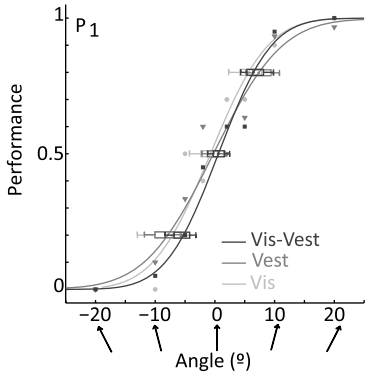

Figure 3. Data for visual alone (VIS, light grey circles), vestibular alone (VEST, medium grey inverted triangles) and visual–vestibular conditions (VIS–VEST, dark grey squares) for participants 1 (P_1) for stereoscopic condition. The data show the proportion of perceived 'more rightward' responses as a function of heading angle. Solid lines represent the cumulative Gaussian functions that were fitted to the data. Box plots whiskers indicate the confidence intervals at -2, -1, 1, 2 standard deviations.

The JND value is inversely proportional to reliability and, thus, the higher the JND the less reliable the cue (see Ernst and Bülthoff, 2004). For all tests the type-I error rate was set at 0.05.

2.6. Maximum Likelihood Estimation (MLE)

Using a simplified form of MLE (Alais and Burr, 2004; Bülthoff and Yuille, 1991; Ernst and Banks, 2002; Ernst *et al*., 2000; Yuille and Bülthoff, 1996), we used the JND to estimate a Gaussian likelihood distribution for each of the unimodal cues (visual, $\hat{S}_{\text{Vis}}$ and vestibular, $\hat{S}_{\text{Vest}}$). If visual and vestibular information combine in an optimal fashion, we can predict the visual–vestibular likelihood, $\hat{S}_{\text{Vis-Vest}}$, using the equation

$$\hat{S}_{\text{Vis-Vest}} = w_{\text{Vis}}\hat{S}_{\text{Vis}} + w_{\text{Vest}}\hat{S}_{\text{Vest}}, \tag{4}$$

where w_{Vis}, w_{Vest} are the weights corresponding to the reliability of the unimodal cues. From equation (4) we can predict the JND of the visual–vestibular combined estimates

$$JND_{\text{Vis-Vest}} = \sqrt{\frac{JND^2_{\text{Vis}} JND^2_{\text{Vest}}}{JND^2_{\text{Vis}} + JND^2_{\text{Vest}}}}. \tag{5}$$

Based on these assumptions, the greatest reduction in the combined JND should be observed when the unimodal cues are of equal reliability $JND_{\text{Vis}} = JND_{\text{Vest}}$ which yields a $\sqrt{2}$ reduction in the JND for the combined trials. From the VIS–VEST condition we can extract the observed JND and compare it to the predicted JND, calculated from the unimodal JNDs using equation (5). Finally, based on the MLE

account, the combined, $JND_{\text{Vis-Vest}}$ should always be less than or equal to the unimodal JNDs

$$JND_{\text{Vis-Vest}} \leqslant \min(JND_{\text{Vis}}, JND_{\text{Vest}}). \quad (6)$$

In order to evaluate whether the observed data were consistent with MLE predictions, both group and individual analyses were conducted. For the group analyses, the JNDs were submitted to a one-way repeated-measures analysis of variance (ANOVA) with factors VIS, VEST and VIS–VEST for both the binocular and stereoscopic conditions. The observed VIS–VEST JNDs for each condition were also compared to the MLE predicted VIS–VEST JNDs using a paired t-test.

For the individual participant analyses, differences between the observed pattern of responding and that predicted using MLE were assessed in two ways. First, each participant's observed VIS–VEST JND and unimodal JNDs were submitted to equation (6). If violated, this would suggest that the most reliable unimodal cue was more reliable than the combined cue estimate. Second, the 95% confidence intervals for the unimodal and combined cue JND were calculated by 1999 repetitions of a bootstrap procedure (for details see Wichmann and Hill, 2001b). From the unimodal bootstrapped 95% confidence intervals (VIS and VEST), the predicted 95% confidence interval was calculated using propagation of error (Taylor, 1997). The 95% confidence intervals were used to determine if the observed combined JND was statistically different from the predicted JND and *vice versa*. If a participant's data failed both types of analyses this would suggest their VIS–VEST data was not consistent with the MLE model.

Using these criteria, participants were divided into two groups as a function of their individual bootstrapped JND distributions; one group consisted of participants who *did* demonstrate the predicted optimal reduction in variance for the combined cue condition as defined above, and the other group consisted of participants who *did not* demonstrate an optimal reduction in variance. The importance of analyzing individual data when assessing behavioral and neurophysiological measures related to multisensory processing has recently been emphasized (e.g., Bentvelzen *et al*., 2009; Werner and Noppeney, 2010).

3. Results

3.1. Binocular Vision

The JNDs and standard errors for the binocular condition averaged across all participants were $5.9° \pm 0.7°$ (VIS), $6.0° \pm 0.65°$ (VEST) and $4.8° \pm 0.42°$ (VIS–VEST). These values are plotted in Fig. 4(a) along with the predicted visual–vestibular JNDs calculated using MLE. To determine whether the unimodal cues were significantly different from the combined cues, we performed a one-way, repeated-measures ANOVA on the VIS, VEST and VIS–VEST trials. No significant main effect was observed ($F(2, 18) = 2.285$, MSE $= 2.31$, $p = 0.1$). This finding is *not consistent* with an MLE model, which predicts a significant reduction in variance

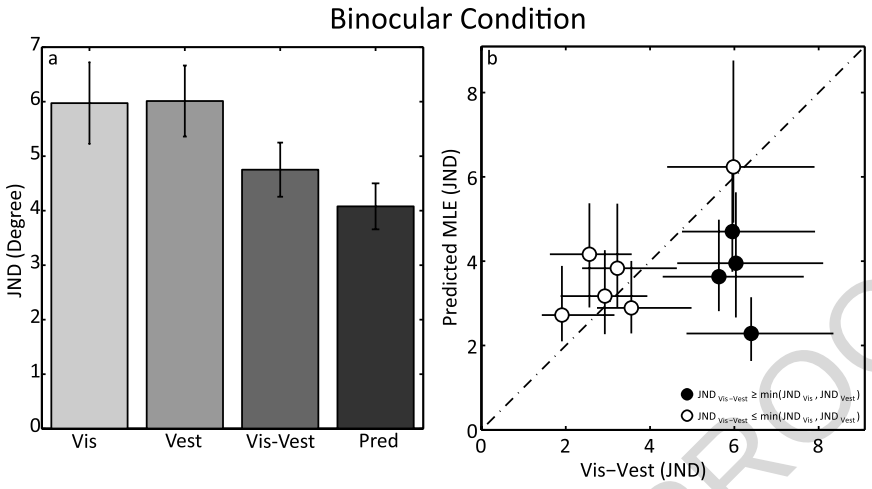

Figure 4. Results of binocular condition. (a) The different bars represent each of the experimental conditions and the predicted visual–vestibular data. Error bars represent standard error of the mean across ten participants. (b) Scatterplot of predicted JND's as a function of observed visual–vestibular trials. The black circles represent participants whose observed visual–vestibular JND is less than their unimodal JNDs. The filled circles represent participants whose observed visual–vestibular JND is greater than their unimodals. Error bars represent 95% bootstrapped confidence intervals.

in combined cue trials compared to unisensory trials. To test whether the observed VIS–VEST data and the predicted VIS–VEST data were statistically different, a *post-hoc* paired Student t-test was performed on the averaged data (Fig. 4), which revealed that they were not statistically different ($p = 0.195$). Therefore, taken together, at the group level these results provide only weak evidence of optimal visual and vestibular cues when visual stimuli are presenting binocularly.

3.2. Stereoscopic Vision

The JNDs and standard errors for the stereoscopic condition averaged across all participants were $5.8° \pm 0.5°$ (VIS), $6.0° \pm 0.65°$ (VEST) and $4.2° \pm 0.46°$ (VIS–VEST) (Fig. 5(a)). To determine whether the unimodal trial estimates were significantly different from the combined cue trial estimates, a one-way, repeated-measures ANOVA was performed on the VIS, VEST and VIS–VEST trials, which revealed a significant main effect ($F(2, 18) = 5.430$, MSE $= 1.816$, $p < 0.05$). The significant main effect of condition was analyzed by single degree of freedom, 'repeated' contrasts. Effect sizes were computed as partial eta-squared values. The contrasts indicate that there was no significant difference between the VIS and VEST trials $F(1, 9) = 0.06$, $p = 0.815$. There was, however, a significant difference between the average of the two unimodal conditions and the combined condition ($F(1, 9) = 40.3$, MSE $= 0.71$, $p < 0.05$). This effect accounted for 82% of the variability in the JND scores. A paired t-test performed on the observed VIS–

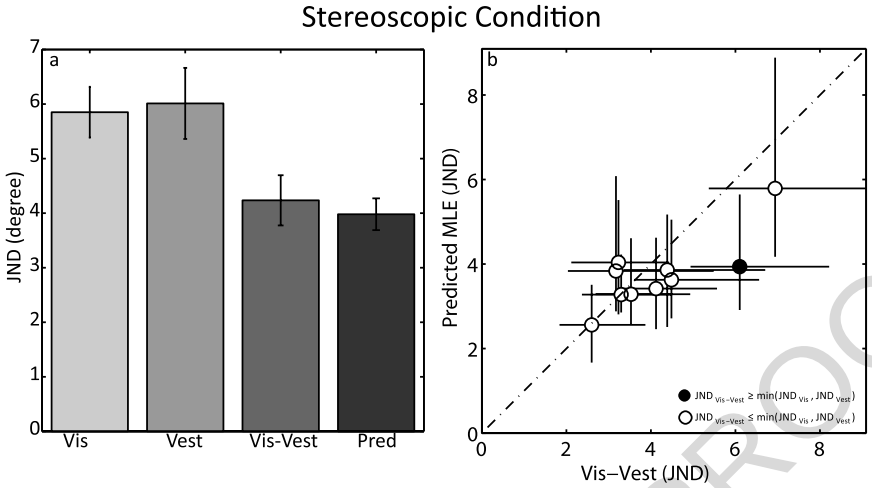

Figure 5. Results of stereoscopic condition. (a) The different bars represent each of the experimental conditions and the predicted visual–vestibular data calculated from the unimodal data. Error bars represent standard error of the mean across ten participants. (b) Scatterplot of predicted JND's as a function of observed visual–vestibular trials. The black circles represent participants whose observed visual–vestibular JND is less than their unimodal JNDs. The filled circles represent one participant whose observed visual–vestibular JND is greater than their unimodal JND. Error bars represent 95% bootstrapped confidence intervals.

VEST trials and the predicted VIS–VEST trials revealed no statistical difference ($p = 0.35$).

Therefore, taken together at the group level, these results provide stronger and more consistent evidence in support of optimal visual–vestibular integration when visual stimuli were presented stereoscopically.

In order to evaluate whether stereoscopic cues affected the JND of heading estimates when only visual information was available, the VIS trials in the stereoscopic condition were compared to the VIS trials in the binocular condition. Interestingly, the addition of stereoscopic cues to the optic flow field did not significantly affect average threshold values under unisensory visual conditions ($5.9° \pm 0.7°$ for the binocular condition and $5.8° \pm 0.5°$ for the stereoscopic condition ($p = 0.658$)). However, when visual information was combined with vestibular cues, differences were observed for binocular and stereoscopic conditions.

3.3. Comparing Binocular and Stereoscopic Conditions for Individual Participants

To gain further insight the differences between the characteristics of cue integration in stereo *versus* binocular visual conditions, individual participant data was analyzed independently. In Figs 4(b) and 5(b) the open circles represent participants whose VIS–VEST JND was lower than either of their VIS and VEST JNDs. The

filled circles represent participants whose VIS–VEST JND was larger than either of the unimodal JNDs (violating equation (6)) and whose bootstrapped 95% confidence intervals of the observed data did not overlap with the predicted JND.

Whereas in the binocular condition four of the ten participants' data (40%) were not consistent with the MLE predictions (Fig. 4(a)), in the stereo condition only one participant's data (10%) were not consistent with the MLE predictions (Fig. 5(a)). The results of a paired t-test performed on this subgroup of 4 participants comparing their observed binocular VIS–VEST trials and the observed stereoscopic VIS–VEST trials approached significance ($p = 0.06$). To ensure that this result was replicable, we ran the same four participants who did not optimally combine with only binocular cues in another binocular control experiment which yielded the same results as in the main experiment (see the Appendix). The results of this control experiment suggest that these individual differences are not simply due to features of the experimental design (e.g., order of conditions) or transient contextual effects, but rather reflect a consistently identified characteristic of individual participant responding. Finally, to ensure that these results were not due to a false assumption that the data could be fit to a Gaussian function, we examined the fits of the individual participants data using the R^2 value and Monte Carlo cumulative probability estimates (CPE) of the model (Wichman *et al.*, 2001a). The mean R^2 for the unimodal and bimodal fits were 0.943 and 0.97, respectively. The mean CPE for unimodal and bimodal fits were 0.49 with a standard deviation of 0.27 and 0.623 with a standard deviation of 0.26, thus, verifying that the Gaussian was an appropriate function to use here.

4. General Discussion

4.1. The Role of Stereo in Visual–Vestibular Integration

Overall, the group data showed strong evidence of optimal visual–vestibular integration when the visual stimuli were presented stereoscopically. In contrast, there was only weak evidence of optimal integration when visual stimuli were presented binocularly. Furthermore, exploratory analyses of individual participants' data showed that, when stereoscopic cues were available, 90% of participants optimally combined cues in a manner consistent with MLE predictions, whereas only 60% of participants optimally combined when only binocular cues were available. The data from the four participants who did not optimally combine in the binocular condition are consistent with the single previous study that did not include stereoscopic information and that also reported a lack of optimal visual–vestibular integration during heading estimation (de Winkel *et al.*, 2010). The results of the stereoscopic condition, on the other hand, are in agreement with the conclusions of most previous studies that have consistently demonstrated an optimal integration of visual and vestibular cues for heading and that also presented optic flow stimuli stereoscopically (Butler *et al.*, 2010; Fetsch *et al.*, 2009; Gu *et al.*, 2008a).

The fact that in the current study stereoscopic visual information was more strongly associated with optimal cue integration, whereas non-stereoscopic visual information provided weak evidence of optimal cue integration, could potentially relate to the fact that the scale of the optic flow is more ambiguous without the additional depth information provided by stereoscopic cues. In our binocular display, because the optic flow stimuli contained no absolute size cues, this may have led to multiple interpretations of the movement profile, thereby increasing uncertainty regarding whether the two sources of sensory information originated from the same event. This would lead to a violation of an underlying assumption of MLE (Kording *et al*., 2007). We postulate that the stereoscopic display may have thereby reduced this uncertainty.

It should be noted that there are many visual depth cues that could help to scale optic flow and, therefore, more research will be required to assess whether the effects observed here are attributable to stereoscopic cues specifically or whether the addition of other cues to depth would result in similar effects. For instance, one strategy that has been used in other studies has been to include familiar size cues in the visual scene, such as human avatars (MacNeilage *et al*., 2007), or other monocular depth cues (e.g., ground texture gradient, linear perspective, etc.). Therefore, future work will consider whether stereoscopic information is unique, or whether the inclusion of other depth cues is also more likely to result in optimal visual–vestibular integration.

Fetsch *et al*. (2009) recently evaluated the effect of stereo cues on combined cue heading estimation in non-human primates that was motivated by anecdotal observations from previous studies in their laboratory suggesting that stereo cues appeared to be important for observing statistically optimal reductions in variance. However, when they later compared stereo and binocular conditions in two well-trained monkeys, no differences were observed. Specifically, even without stereo cues a statistically optimal reduction in variance was observed. The authors suggested that stereo cues might be important when first performing the task, but may become less important with continued extensive training. In a control experiment following from the current experiment, we investigated whether the results would be replicated for the same participants in subsequent testing sessions (see the Appendix). Specifically, this study evaluated whether having some additional experience with both stereoscopic and binocular conditions would change participants' performance. The results demonstrated that the responses in the control experiment replicated the behavioral results reported in the current study. Specifically, participants who combined binocular visual and vestibular cues in the main experiment also exhibited optimal integration in the control experiment, while those who did not demonstrate optimal cue integration in the main experiment also did not exhibit optimal integration in the control binocular experiment. These results demonstrate that the effects reported for the current experiment were robust and replicable for individual participants. Future work will be needed to evaluate whether having even

more extensive experience with the task and stimuli would change this pattern of responding.

4.2. No Effect of Stereo on Unisensory Visual Conditions

When comparing the reliability of heading estimates for the binocular and stereo conditions for unisensory trials when only visual information was presented, surprisingly no significant differences were observed. These results are not directly consistent with those of van den Berg and Brenner (1994) who reported improved heading estimation during stereoscopic unisensory visual conditions. One possible reason for this discrepancy may relate to the signal-to-noise ratio in the optic flow stimuli used in each experiment. Specifically, van den Berg and Brenner (1994) reported that the benefits provided through stereoscopic cues (over binocular cues) were only observed under conditions of low signal-to-noise; however, this was not controlled for in the current study. The trend of the measured variances in binocular and stereoscopic conditions, however, indicates that the binocular estimates, on average were associated with a higher JND (i.e., lower reliability) compared to stereoscopic conditions. These results highlight the need to more carefully evaluate the mechanisms by which stereoscopic cues contribute to the interpretation of heading from optic flow.

5. Summary

In summary, it appears as though the presence or absence of stereoscopic visual information can impact the extent to which visual and vestibular cues are integrated during heading perception. Specifically, the presence of stereoscopic cues is associated with stronger evidence of optimal integration compared to conditions in which no stereoscopic cues are available. These results have implications for research areas focused on understanding the contributions of particular visual and non-visual cues in self-motion perception and of the principles underlying multisensory integration in general. There could also be considerable applied implications, including whether the incorporation of stereoscopic displays might improve motion simulation technologies for training and evaluation (e.g., the development of driving and flight simulators). By integrating specific depth cues into the visual display, this may provide a more realistic experience of self-motion and could possibly reduce motion sickness associated with visual–vestibular cue conflicts.

Acknowledgements

This work was supported by the Max Planck Society, Enterprise Ireland and by the WCU (World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10008-0). The authors would like to thank Daniel Berger, John Foxe, Marc Ernst and Martin Banks for their invaluable advice and guidance. We would like to

thank Julian Hofmann, Michael Weyel and the participants for help with the data collection.

References

Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* **14**, 257–262.

Benson, A. J., Spencer, M. B. and Stott, J. R. (1986). Thresholds for the detection of the direction of whole-body, linear movement in the horizontal plane, *Aviat. Space Environ. Med.* **57**, 1088–1096.

Bentvelzen, A., Leung, J. and Alais, D. (2009). Discriminating audiovisual speed: optimal integration of speed defaults to probability summation when component reliabilities diverge, *Perception* **38**, 966–987.

Bremmer, F., Duhamel, J. R., Ben Hamed, S. and Graf, W. (2002a). Heading encoding in the macaque ventral intraparietal area (VIP), *Eur. J. Neurosci.* **16**, 1554–1568.

Bremmer, F., Klam, F., Duhamel, J. R., Ben Hamed, S. and Graf, W. (2002b). Visual–vestibular interactive responses in the macaque ventral intraparietal area (VIP), *Eur. J. Neurosci.* **16**, 1569–1586.

Britten, K. H. and Van Wezel, R. J. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys, *Nat. Neurosci.* **1**, 59–63.

Britten, K. H. and Van Wezel, R. J. (2002). Area MST and heading perception in macaque monkeys, *Cerebral Cortex* **12**, 692–701.

Bülthoff, H. H. and Yuille, A. (1991). Bayesian models for seeing shapes and depth, *Comm. Theoret. Biol.* **2**, 283–314.

Butler, J. S., Smith, S. T., Campos, J. L. and Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading, *J. Vision* **10**, 23.

De Winkel, K. N., Weesie, J., Werkhoven, P. J. and Groen, E. L. (2010). Integration of visual and inertial cues in perceived heading of self-motion, *J. Vision* **10**, 1.

Duffy, C. J. and Wurtz, R. H. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli, *J. Neurophysiol.* **65**, 1329–1345.

Ernst, M. O. and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion, *Nature* **415**, 429–433.

Ernst, M. O., and Bülthoff, H. H. (2004). Merging the senses into a robust percept, *Trends Cognit. Sci.* **8**, 162–169.

Ernst, M. O., Banks, M. S. and Bülthoff, H. H. (2000). Touch can change visual slant perception, *Nat. Neurosci.* **3**, 69–73.

Fetsch, C. R., Turner, A. H., Deangelis, G. C. and Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception, *J. Neurosci.* **29**, 15601–15612.

Frenz, H. and Lappe, M. (2005). Absolute travel distance from optic flow, *Vision Research* **45**, 1679–1692.

Frenz, H. and Lappe, M. (2006). Visual distance estimation in static compared to moving virtual scenes, *Spanish J. Psychol.* **9**, 321–331.

Gu, Y., Deangelis, G. C. and Angelaki, D. E. (2007). A functional link between area MSTd and heading perception based on vestibular signals, *Nat. Neurosci.* **10**, 1038–1047.

Gu, Y., Angelaki, D. E. and Deangelis, G. C. (2008a). Neural correlates of multisensory cue integration in macaque MSTd.

Gu, Y., Angelaki, D. E. and Deangelis, G. C. (2008b). Neural correlates of multisensory cue integration in macaque MSTd, *Nat. Neurosci.* **11**, 1201–1210.

Gu, Y., Fetsch, C. R., Adeyemo, B., Deangelis, G. C. and Angelaki, D. E. (2010). Decoding of MSTd population activity accounts for variations in the precision of heading perception, *Neuron* **66**, 596–609.

Heuer, H. W. and Britten, K. H. (2004). Optic flow signals in extrastriate area MST: comparison of perceptual and neuronal sensitivity, *J. Neurophysiol.* **91**, 1314–1326.

Kording, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B. and Shams, L. (2007). Causal inference in multisensory perception, *PLoS One* **2**, e943.

Lappe, M., Bremmer, F. and Van Den Berg, A. V. (1999). Perception of self-motion from visual flow, *Trends Cognit. Sci.* **3**, 329–336.

MacNeilage, P. R., Banks, M. S., Berger, D. R. and Bülthoff, H. H. (2007). A Bayesian model of the disambiguation of gravitoinertial force by visual cues, *Exper. Brain Res.* **179**, 263–290.

Ohmi, M. (1996). Egocentric perception through interaction among many sensory systems, *Brain Res. Cognit. Brain Res.* **5**, 87–96.

Page, W. K. and Duffy, C. J. (2003). Heading representation in MST: sensory interactions and population encoding, *J. Neurophysiol.* **89**, 1994–2013.

Palmisano, S. (1996). Perceiving self-motion in depth: the role of stereoscopic motion and changing-size cues, *Percept. Psychophys.* **58**, 1168–1176.

Perrone, J. A. and Stone, L. S. (1998). Emulating the visual receptive-field properties of MST neurons with a template model of heading estimation, *J. Neurosci.* **18**, 5958–5975.

Roy, J. P. and Wurtz, R. H. (1990). The role of disparity-sensitive cortical neurons in signalling the direction of self-motion, *Nature* **348**, 160–162.

Roy, J. P., Komatsu, H. and Wurtz, R. H. (1992). Disparity sensitivity of neurons in monkey extrastriate area MST, *J. Neurosci.* **12**, 2478–2492.

Royden, C. S., Banks, M. S. and Crowell, J. A. (1992). The perception of heading during eye movements, *Nature* **360**, 583–585.

Sato, Y., Toyoizumi, T. and Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli, *Neural Comput.* **19**, 3335–3355.

Taylor, J. R. (1997). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, Sausalito, CA, USA.

Telford, L., Howard, I. P. and Ohmi, M. (1995). Heading judgments during active and passive self-motion, *Exper. Brain Res.* **104**, 502–510.

Upadhyay, U. D., Page, W. K. and Duffy, C. J. (2000). MST responses to pursuit across optic flow with motion parallax, *J. Neurophysiol.* **84**, 818–826.

Van Den Berg, A. V. and Brenner, E. (1994). Why two eyes are better than one for judgements of heading, *Nature* **371**, 700–702.

Von Der Heyde, M. (2001). *A Distributed Virtual Reality System for Spatial Updating: Concepts, Implementation, and Experiments*. Universität Bielefeld, Germany.

Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W. and Schirillo, J. A. (2004). Unifying multisensory signals across time and space, *Exper. Brain Res.* **158**, 252–258.

Warren, W. H., Jr. and Hannon, D. J. (1990). Eye movements and optical flow, *J. Optic. Soc. Amer. A* **7**, 160–169.

Warren, P. A. and Rushton, S. K. (2009). Perception of scene-relative object movement: optic flow parsing and the contribution of monocular depth cues, *Vision Research* **49**, 1406–1419.

Werner, S. and Noppeney, U. (2010). Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization, *Cereb. Cortex* **20**, 1829–1842.

Wichmann, F. A. and Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit, *Percept. Psychophys.* **63**, 1293–1313.

Wichmann, F. A. and Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling, *Percept. Psychophys.* **63**, 1314–1329.

Yuille, A. and Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics, in: *Perception as Bayesian Inference*, D. Knill and W. Richards (Eds). Cambridge University Press, Cambridge, UK.

Appendix: Experiment on Binocular Replication

Participants

Four of the original ten participants completed the control experiment without any additional knowledge of the purpose and manipulations in the main experiment. Three of the four participants (P8–P10) who exhibit non-optimal integration of visual and vestibular information in the binocular condition of the main experiment were included in this group.

General Procedure

The apparatus and stimuli were identical to those in the main experiment. In this case participants performed a 2-alternative forced choice task (2AFC) in which they were asked to judge whether they moved to the left or the right. Each trial consisted of one linear heading motion chosen from eight angles ranging from -10° to 10° on a log scale centered around 0°. After the motion ended, the participants indicated if they moved left or right *via* a button box. Participants completed one binocular experimental session consisting of three experimental blocks; VEST, VIS and VIS–VEST. Each block contained 96 trials, participants completed 12 repetitions of each of the 8 heading stimuli for each condition.

Results

The average JNDs and their standard errors for the binocular condition were $3.3^\circ \pm 0.4^\circ$ (VIS), $5.33^\circ \pm 1.2^\circ$ (VEST) and $4.4^\circ \pm 0.86^\circ$ (VIS–VEST). Figure A1(a) shows the average JND values for the binocular condition. To determine whether the responses in the unimodal cue conditions were different from the combined cue condition, we performed a one-way, repeated-measures ANOVA on the binocular VIS, VEST and VIS–VEST conditions. The analysis revealed no significant difference between the unimodal conditions and the combined cue condition ($F(2, 6) = 1.012$, MSE $= 4.039$, $p = 0.42$). A *post-hoc* t-test performed on the observed binocular VIS–VEST values and the predicted VIS–VEST values revealed no statistical difference ($p = 0.15$).

Figure A1(b) shows the scatterplot for the observed VIS–VEST values *versus* the predicted VIS–VEST values for each participant, with the dotted line representing the ideal. The open circles indicate participants whose VIS–VEST JND was lower than at least one of their unimodal JNDs. The three filled circles indicate the participants whose VIS–VEST JND was larger than both of their unimodal JNDs. The

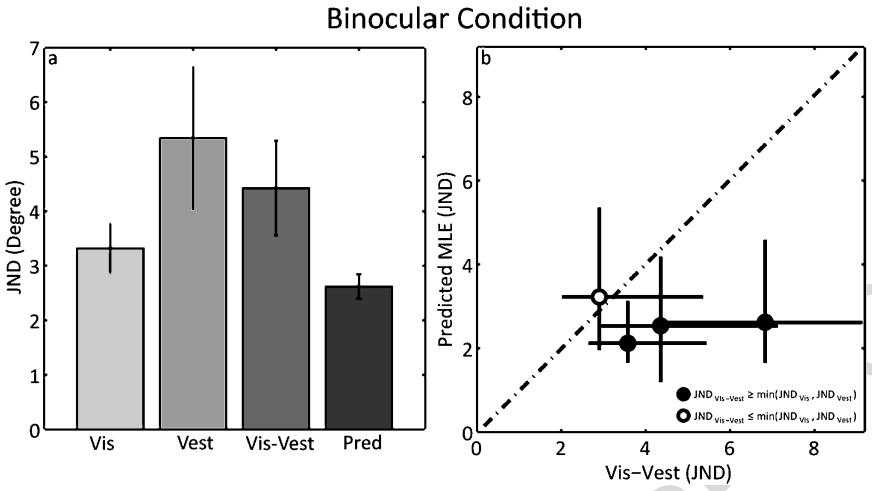

Figure A1. Results of the binocular condition. (a) Averaged observed JND values for each of the binocular conditions and the predicted combined cue data calculated from the unimodal data. Error bars denote standard error of the mean across four participants. (b) Scatterplot of observed *vs*. predicted JND values for the binocular combine cue condition. The open circle represents the participant whose observed binocular VIS–VEST JND was less than their unimodal JNDs. The filled circles represent the three participants who had observed binocular VIS–VEST JNDs that were greater than their unimodal JNDs. Error bars represent 95% bootstrapped confidence intervals.

vertical and horizontal bars denote 95% bootstrapped confidence intervals for the observed and predicted bimodal JNDs.

Summary

The results of the control binocular experiment are consistent with the results presented in the main text. This demonstrates that, for a subset of participants, binocular visual and vestibular cues do not combine in an optimal fashion and that this is a robust and replicable result that is not dependent on individual participant prior experience with the different types of visual stimuli (i.e., not due to experimental carry-over effects).