Integration of visual and inertial cues in perceived heading of self-motion

Ksander N. de Winkel
Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands, & TNO Defence, Security and Safety, Soesterberg, The Netherlands

Jeroen Weesie
Department of Social Sciences, Utrecht University, Utrecht, The Netherlands

Peter J. Werkhoven
TNO Defence, Security and Safety, Soesterberg, The Netherlands

Eric L. Groen
TNO Defence, Security and Safety, Soesterberg, The Netherlands

In the present study, we investigated whether the perception of heading of linear self-motion can be explained by Maximum Likelihood Integration (MLI) of visual and non-visual sensory cues. MLI predicts smaller variance for multisensory judgments compared to unisensory judgments. Nine participants were exposed to visual, inertial, or visual–inertial motion conditions in a moving base simulator, capable of accelerating along a horizontal linear track with variable heading. Visual random-dot motion stimuli were projected on a display with a 40° horizontal × 32° vertical field of view (FoV). All motion profiles consisted of a raised cosine bell in velocity. Stimulus heading was varied between 0 and 20°. After each stimulus, participants indicated whether perceived self-motion was straight-ahead or not. We fitted cumulative normal distribution functions to the data as a psychometric model and compared this model to a nested model in which the slope of the multisensory condition was subject to the MLI hypothesis. Based on likelihood ratio tests, the MLI model had to be rejected. It seems that the imprecise inertial estimate was weighed relatively more than the precise visual estimate, compared to the MLI predictions. Possibly, this can be attributed to low realism of the visual stimulus. The present results concur with other findings of overweighing of inertial cues in synthetic environments.

Keywords: maximum likelihood integration, multisensory integration, visual, vestibular, detection thresholds


Introduction

Perception of self-motion and -orientation in the environment is based on neural integration of inputs from the visual, vestibular, kinesthetic, and tactile senses. There has been extensive research on this topic, in particular on the contribution of the vestibular and visual systems (e.g., Bischof, 1974; Guedry, 1974; Howard, 1982). The visual system provides multidimensional information on self-motion by means of optic flow and on spatial orientation by means of visual frame and polarity information (Howard & Childerson, 1994). The vestibular system in the inner ear detects linear motion by means of the otolith organs. Furthermore, there is evidence for extra-vestibular sensory neurons dedicated to the perception of gravity, the so-called “graviceptors” (Mittelstaedt, 1996; Zaichik, Rodchenko, Rufov, Yashin, & White, 1999). The dynamics of the different sensory systems and interactions between them have been represented in mathematical models to explain fundamental psychophysical characteristics of self-motion and -orientation (e.g., Bos & Bles, 2002; Zupan, Merfeld, & Darlot, 2002). For example, models have successfully described the multisensory processes in the onset of perceived self-motion (e.g., Henn, Cohen, & Young, 1980; Young, Dichgans, Murphy, & Brandt, 1973), the neural disambiguation of the gravito-inertial force into gravity and linear acceleration by means of visual information (MacNeillage, Banks, Berger, & Bültlhoft, 2007), and visual–vestibular interaction in the perception of self-tilt (Vingerhoets, De Vrijer, Van Gisbergen, & Medendorp, 2009). However, heading perception, i.e., the direction of linear motion along the naso-occipital axis, has mainly been studied as a visual task with non-moving observers and has not received much attention in the literature on multisensory perception. Furthermore, although multisensory
perception has been associated with higher precision than unisensory perception for several perceptual tasks, it is not yet known whether this principle also holds for the perception of heading. In this paper, we describe an experiment in which we measured variability in visual, inertial, and combined visual–inertial perceptions of heading in order to test whether multisensory stimuli yielded more precise heading judgments than their unisensory constituents.

Visual heading perception

During self-motion, heading is specified by the point in the optic array from which the surrounding image radially expands (the focus of expansion, FOE). Heading can be estimated by localizing this point (Gibson, 1950; Warren & Hannon, 1988). Warren, Morris, and Kalish (1988) presented to participants moving horizontal random-dot planes and instructed the participants to report whether it looked as if they moved to the left or right of a target; 75% correct thresholds were in the range of 0.65° to 1.2°, depending somewhat on the density of the dots and stimulus speed. In a paper by Telford, Howard, and Ohmi (1995), a 75% correct detection threshold of 5.5° was reported, averaged over participants. Accuracy of heading judgments depends upon the part of the fovea stimulated, eccentricity of the FOE (Crowell & Banks, 1993; Warren & Kurtz, 1992), and the coherence of motion of particles in the optic flow field (Gu, Angelaki, & DeAngelis, 2008).

Inertial heading perception

The vestibular system responds to rotational and linear accelerations of the head associated with changes of self-motion. It is less useful for the perception of the actual speed and trajectory of self-motion. For these aspects, we rely on the visual system. However, the otolith organs do provide information on the direction of linear acceleration (Gu et al., 2008; Gu, DeAngelis, & Angelaki, 2007; Ohmi, 1996; Telford et al., 1995), as do several extra-vestibular graviceptors throughout the body (Mittelstaedt, 1996; Zaichik et al., 1999). For humans, a 75% correct detection threshold of non-zero heading (i.e., not straight ahead) for inertial stimuli has been reported to be 11.4°, averaged over participants (Telford et al., 1995). Thresholds as small as 1.3° have been reported for extensively trained macaque monkeys, suggesting that inertial signals can indeed contribute to precise heading judgments (Gu et al., 2007). Since extra-vestibular sensors respond to the same environmental stimuli as the vestibular system and are difficult to isolate from the vestibular system, we will treat them as a single system, here collectively designated inertial heading sensors.

Multisensory heading perception

It is likely that the brain combines estimates of heading made by multiple senses in a way that allows us to benefit from having multiple sources of information. Such a benefit can be expressed by improvements in precision (i.e., a reduction of variance) of the estimate. A well-known integration scheme is Bayesian Integration (BI). BI models have been applied to describe the influence of prior knowledge on sensory integration (e.g., Bresciani, Dammeyer, & Ernst, 2006; Jürgens & Becker, 2006; Laurens & Droulez, 2006; MacNeilage et al., 2007). In its most simple form, the Bayesian scheme is essentially Maximum Likelihood Integration (MLI; Ernst & Banks, 2002). This is a statistically optimal strategy to combine multiple cues (“observations”). Assuming normality and independence of noises in internal representations, MLI effectively states that multisensory estimates of the same environmental property are combined as a weighted average, with weights proportional to each estimate’s reliability (i.e., inverse of its variance; Howard, 1997). MLI yields minimal variance in the integrated estimate among all weighted averages. Thus, MLI predicts how the (parameters of the) multisensory condition relate to the (parameters of) unisensory conditions. It has been shown that the brain acts according to MLI for several psychophysical phenomena, such as integration of visual and haptic information on an object’s size (Ernst & Banks, 2002) and for integration of multisensory information on sequences of events (e.g., Andersen, Tiipana, & Sams, 2005; Shams, Ma, & Beierholm, 2005). In these studies, multisensory estimates were more precise than unisensory estimates. In a recent study, Gu et al. (2008) investigated whether this integration strategy also holds for the perception of heading. Macaque monkeys were trained to perform a discrimination task in which they were passively moved along a linear track in the horizontal plane with a certain heading. The monkeys indicated whether the experienced motion was to the left or right of straight-ahead heading. The results suggested that the monkeys were more precise in their judgments when multisensory visual–inertial cues were presented than when either cue was presented in isolation, consistent with the MLI hypothesis. Furthermore, Fetsch, Turner, DeAngelis, and Angelaki (2009) recently showed that cue weighting is a dynamic process. They presented macaque monkeys and humans with multisensory heading stimuli. Reliability of the visual heading cue was manipulated between successive trials by varying coherence of the direction of moving dots in the visual field between 25% and 70%. It was observed that the weight attributed to each cue was updated on a trial-to-trial basis.

Our interest in the present study was to investigate whether multisensory presentation yields a more precise estimate compared to its unisensory constituents, when the objects in the optical array move in a completely coherent fashion. We hypothesized that the variance of the estimate would be smallest when both visual and inertial cues were
available, compared to when either cue was presented in isolation. We focused on MLI theory, which gives a precise prediction of the extent to which the variance is reduced. The study was performed in a moving base simulator that allowed us to independently manipulate visual and inertial motion cues.

**Methods**

**Participants**

Nine paid volunteers took part in this experiment (six men, three women, mean age 28.7, standard deviation 6.9). All participants reported normal vestibular function and normal or corrected-to-normal vision. After receiving general instructions of the experimental goals and procedures, all participants signed an informed consent.

**Apparatus**

The experiment was performed using the DESDEMONA simulator at the TNO Institute in Soesterberg, The Netherlands (Bles & Groen, 2009). This centrifuge-based simulator features a motion platform with six degrees of freedom. For this study, only two degrees of freedom were used: rotation about the cabin’s vertical yaw axis and linear motion along the 8-m horizontal track. Participants were seated on a padded seat inside the cabin and secured by a five-point safety harness. Foam cushions were placed between a headrest and the left and right sides of the head to minimize head movements.

Inside the cabin, an out-the-window (OTW) visual stimulus was projected on a screen at about 1.5 m in front of the participant. Participants wore a mask that restricted their field of view to 40° (horizontal) × 32° (vertical) of the OTW display, blocking stationary visual cues from the cabin interior. The mask served as a substitute for a lacking cockpit canopy and made the OTW scenery appear in the background. The latter is important, since it is known thatvection (the visually induced sense of self-motion) is induced more effectively when a visual motion stimulus is presented in the perceptual background (Howard & Heckmann, 1989). Participants wore an audio-headset that allowed for constant contact with the experimenters. Responses were given verbally and were noted by the experimenters.

**Stimuli**

Stimuli consisted of visual, inertial, or combined visual–inertial linear horizontal motion with different headings. Heading is defined as the direction of motion with respect to the median plane of the body: A heading of 0° corresponds to linear forward motion along the participants’ naso-occipital axis; a 90° heading corresponds to linear rightward motion along the inter-aural axis. We did not expect consistent differences between perceptions of left- and rightward motions. Therefore, only rightward motion was presented, which reduced the number of trials.

Visual stimuli consisted of linear horizontal motion through a star field. Different angles of visual heading were achieved by shifting the focus of expansion (FOE) sideways. The star field consisted of a cloud of solid white circles, placed in random positions on a three-dimensional grid in a dark surrounding environment. This stimulus was generated at random with each trial (Figure 1). The objects never appeared at the FOE. Displacement of the visual objects was coupled linearly with the inertial motion. Absolute velocity of the visual motion was arbitrary since participants neither had objective information about the distance between objects, nor could they determine their size. Velocity of movement through the star field was determined in a pilot study; the stimulus amplitude was chosen such that it subjectively matched the velocity of the inertial motion. Participants never reported a feeling of discrepancy between the visual and inertial velocities.

Inertial stimuli consisted of motion along the linear horizontal track of the simulator over a total length of 7 m. The velocity profile was a raised cosine bell with maximum velocity of 1.5 m/s. Maximum acceleration was 0.5 m/s². Each motion profile lasted 9.3 s (see additional material: HeadingI.wmv). Since the vestibular system is responsive to acceleration and not to constant velocity, a constantly changing velocity profile was used to ensure vestibular reactivity.

Participants reported whether or not they perceived a heading of 0°. Psychometric curves (cumulative normal distribution functions) of the probability that a participant perceived the stimulus as being not straight ahead as a
function of heading angle were determined based on six different fixed stimulus headings. When using fixed stimuli, it is important that the data points fall as much as possible on the steepest part of the psychometric curve. Based on previous reports, we knew that the detection thresholds differ between the different sensory modalities. For example, visual detection thresholds fall in the range of 0.65°–5.5° (Telford et al., 1995; Warren et al., 1988, respectively) and an inertial detection threshold of 11.4° has been reported (Telford et al., 1995). The range of stimulus values for which the percentage correct responses gradually increases thus varies between sensory modalities. Therefore, we used different ranges of stimulus values for each condition, which we verified in a pilot experiment (Table 1).

### Procedure

Stimuli were presented in separate simulator runs of about 30 s. At the beginning of each run, the cabin was positioned at one end of the linear track. The run started with rotation of the cabin about its yaw axis to orient the participant at the desired stimulus heading relative to the linear track. The cabin always rotated the longest distance (i.e., around 180°) with an angular velocity between 12 and 13.33°/s, depending on the stimulus angle. The duration of this rotation was kept constant to eliminate it as a possible cue to stimulus heading. To allow the aftereffects of the response of the semi-circular canals for this yaw rotation to wash out, a 6-s pause was implemented before the actual stimulus started. After this pause, the cabin was moved over 7 m to the other end of the linear track in the inertial and visual–inertial conditions. In the visual-only condition, the cabin remained stationary at the end of the linear track, and only visual motion was presented. Linear motion stimuli always lasted 9.3 s. Following stimulus presentation, participants gave their verbal response in a 1-s pause before the next run started with the reorientation of the cabin. Participants were asked to judge the direction of perceived self-motion by means of a two-alternative forced-choice (2AFC) task. In other words, they indicated after each stimulus whether they perceived the motion as “straight ahead” or “not straight ahead.” Participants were instructed to use all sensory information on self-motion, irrespective whether this was inertial, visual, or both.

The runs were presented in three separate blocks, or conditions: a condition with visual-only (V) runs, a condition with inertial (I) runs, and a condition with combined visual–inertial (C) runs. In each condition, all six stimulus angles were presented 10 times, totaling 60 runs per condition. Participants also experienced an extra set of 60 visual–inertial stimuli to answer another research question, which will be reported elsewhere. The order of stimuli within a condition was randomized as much as possible using Latin squares. Each participant performed a total of 240 runs in four 30-min blocks. After each block, they had a 15-min break outside the simulator. Including instruction, the whole experiment lasted about 4 h for each participant.

### Theoretical model

We assume that participants have an internal noisy but “continuous” representation $X^*$ of their heading angle $\alpha$, with normal distributed noise, $X^* \sim N(\alpha, \sigma)$, where the standard deviation $\sigma$ reflects the size of the noise in the random variables. We expect that participants experience to move “not straight ahead” and respond accordingly (i.e., binary response; $X = 1$) when a certain internal threshold is exceeded: $X^* > \tau$ (Long, 1997, Section 3.2). When the heading angle increases, the probability of responding “not straight ahead” also increases.

### Maximum likelihood integration hypothesis

MLI hypothesizes how the standard deviation parameters for the multisensory (combined) condition ($\sigma_c$) depend on the standard deviation parameters for the unisensory inertial ($\sigma_i$) and visual ($\sigma_v$) conditions. The value of an environmental property, such as an assessment of heading, can be represented by an “internal” random variable $X^*_i$. When we have two assessments, $X^*_i$ and $X^*_v$, of a single environmental property, as is the case with multiple senses in the combined condition, the value of that property can be estimated by a convex (i.e., coefficients sum to one) combination of the unisensory representations

$$X^*_c = w_i X^*_i + w_v X^*_v, w_i, w_v > 0, w_i + w_v = 1, \quad (1)$$

<table>
<thead>
<tr>
<th>Heading angle $\alpha$</th>
<th>0°</th>
<th>1°</th>
<th>3°</th>
<th>5°</th>
<th>7°</th>
<th>10°</th>
<th>15°</th>
<th>20°</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

Table 1. Heading angles used in the three conditions. V, I, and C stand for the visual, inertial, and combined visual–inertial condition, respectively.
where the \(w_i\) are weights. Assuming unbiased unisensory noisy representations of the true heading angle \(\alpha\), the linear combination is also an unbiased noisy representation of \(\alpha\). Since a linear combination of normal variates is itself normal distributed, the noise in the combined estimate \(X^*_j\) is also normal distributed. Assuming that the random noises are stochastically independent, the variance \(\sigma^2_c\) of \(X^*_j\) is

\[
\sigma^2_c = w_i^2 \sigma^2_i + w_v^2 \sigma^2_v.
\]  

(2)

The statistical method of Maximum Likelihood (ML) can be used to give a prediction about the weights \((w_i\) and \(w_v)\). Since we assumed normal distributed noise, the likelihood \(L_j\) for the internal representation \(X^*_j\) for the sensory condition \(j\) is

\[
L_j(\alpha; X^*_j, \sigma_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left( -\frac{1}{2} \left( \frac{X^*_j - \alpha}{\sigma_j} \right)^2 \right). 
\]  

(3)

In the multisensory condition, the likelihood function of \((X^*_s, X^*_v)\) is given by the product of the likelihoods of the unisensory variables \(X^*_s\) because we assume that the noises are independent across senses. Treating the \(\sigma^*_j\) as knowns, the maximum of this function yields the ML estimate of the true angle \(\alpha\) in terms of these parameters. It can be derived mathematically that the values of the expression

\[
\frac{\pi^*_{ja}}{\pi^*_{ja} + \pi^*_{ja}} = \Phi\left( \frac{\tau_{j^*} - \alpha}{\sigma_j} \right)
\]  

(6)

where \(\Phi\) denotes the cumulative standard normal distribution. The model is a dichotomous analogue to the familiar ANCOVA for continuous responses, where we regress the binary dependent variable “response” on the independent variable “sensory modality tested,” in the presence of the continuous independent variable “heading angle.”

We estimated the parameters \((\tau_j, \sigma_j)\) of the psychometric curve for condition \(j\) by maximum likelihood estimation (MLE) of the psychometric curves, so that we finally fitted the model (Equation 6) with the MLI-induced constraint (Equation 5). This restricted model has five parameters (three \(\tau\)s and two \(\sigma\)s), as the slope in the combined condition was predicted by MLI of the parameters of unisensory conditions. Since the \(\tau\) parameter is free to vary for each condition, a comparison of the unrestricted and restricted models is only affected by the slopes \((\sigma)\) of the fitted functions. A comparison of the fit of the models with and without this constraint, using likelihood ratio tests with one degree of freedom, allowed us to test whether our data support the MLI hypothesis.

In other work on testing MLI in cue integration (e.g., Fetsch et al., 2009; Helbig & Ernst, 2007), the standard deviations \(\sigma\) of the underlying Gaussians for the unisensory and multisensory conditions are often derived from the slope of fitted cumulative Gaussians:

\[
slope = \frac{\sigma}{\sqrt{2}}
\]  

(7)

The standard deviation of the multisensory condition is subsequently compared to the value that is predicted using the standard deviations of the unisensory conditions (Equation 5). Our approach essentially does not differ but is a more direct evaluation of the MLI hypothesis that is statistically more efficient.

**Data analysis**

Since we assume normally distributed unbiased internal representations of heading, this results in a probit regression of the binomially distributed binary response \(X_{ja}\) on the condition \(j = (i, v, c)\) and angle \(\alpha\), and their interaction. Here \(i, v,\) and \(c\) represent the inertial, visual, and combined condition, respectively. More specifically,

\[
\pi_{ja} = Pr(X_{ja} = 1) = \Phi\left( \frac{\tau_j - \alpha}{\sigma_j} \right),
\]  

(6)

Results

Figure 2 shows the results for each individual participant. In each panel, the fitted probability that a stimulus
was not perceived as “straight ahead” (SA) is plotted against heading angle for each condition (visual, inertial, and combined visual–inertial). The curves plotted through the data points are cumulative normal distribution functions. The shaded areas represent pointwise 95% confidence intervals. Table 2 describes the two fitted models and two associated tests for each participant. More specifically, the table lists the log likelihood of the unrestricted model, Pearson’s goodness-of-fit test for the unrestricted model, the log likelihood of the MLI-restricted model, and a likelihood ratio test for the MLI restriction.

The fit of the probit model is based on a Pearson’s $\chi^2$ test for goodness of fit with 180 observations equally distributed over 18 heading angle x condition combinations. In general, the model fits the data well, and hence we may proceed below to test whether the 6 parameters of this model satisfy the MLI constraint. The violation for participant number three is caused by a poor fit of the model in the combined condition. The results reported below do not differ substantially if we include or exclude participant three.

A likelihood ratio test showed large differences in the parameters of the fitted cumulative normal distributions ($\tau$, $\sigma$) between participants ($\chi^2(48) = 168.92$, $p < 0.001$). Differences between the curves were also assessed for each participant individually using Wald $\chi^2$ tests (Table 3).

According to the MLI hypothesis, the variance of the multisensory estimate should be lower than the variance of either unisensory estimate. We compared an unrestricted model (see Data analysis section) to a model in which the standard deviation of the multisensory condition was constrained to the value predicted by MLI (Equation 5). The average standard deviations for the visual, inertial, and combined visual–inertial conditions were: $1.07^\circ$ ($SD = 0.68$), $7.20^\circ$ ($SD = 3.94$), and $3.56^\circ$ ($SD = 1.04$), respectively. The model likelihoods and likelihood ratios,
as well as a measure of the model goodness of fit are given in Table 2. Associated observed and predicted standard deviations are presented in Figure 3.

Combined over all participants, we have strong evidence against MLI ($\chi^2(9) = 156.19$, $p < 0.001$). For all but one participant, we have strong evidence that their heading perceptions conflict with MLI. To allow comparison with previous studies on heading perception, we also calculated the group’s average 75% detection thresholds for the three conditions, using the estimated model parameters and an inverse CDF function. These values amounted to 4.6° ($SD = 1.3$), 16.1° ($SD = 5.2$), and 9.4° ($SD = 1.94$) for the visual, inertial, and combined visual–inertial conditions, respectively. Note that these thresholds indicate the shift of a subjective judgment from “straight ahead” to “not straight ahead.” These thresholds should be seen as a “Point of Subjective Straight Heading.”

### Discussion

The results of this study showed that the 75% correct detection threshold for deviations from subjective straight ahead was much lower for visual motion than for inertial motion in all participants. The average 75% threshold was 4.6° in the visual condition and 16.1° for the inertial condition. These values correspond to earlier findings, e.g., by Telford et al. (1995). However, the present multisensory results differ from earlier observations. We found that the detection threshold in the combined condition was always larger (average 9.4°) than in the visual-only condition, while Telford et al. reported a near identical visual and combined visual–inertial threshold of 5.5° and 5.7°, respectively. More importantly, they reported a slight but non-significant reduction in variance when both visual information and vestibular information were available, compared to when only visual information was available. This is in line with the predictions of a Maximum Likelihood Integration (MLI) model. According to MLI, the variance of a combined estimate should always be smaller than the variance of the best unisensory constituent. Although Telford et al. did not explicitly test the MLI hypothesis, Gu et al. (2008) found supporting evidence for MLI in macaque monkeys. In contrast, our results indicated that the variance of the combined estimate was actually larger than that of the best unisensory estimate. Provided that our assumptions on independent and normally distributed errors in the internal representation of stimuli were met, this implies that the weight assigned to the inertial estimate was too large and that MLI has to be rejected.

Our observation does concur with other findings that inertial cues in a synthetic simulator environment are weighted more heavily relative to visual cues (Groen &
Bles, 2004; Groen, Valenti Clari, & Hosman, 2001; Harris, Jenkin, & Zikovitz, 2000). Furthermore, Fetsch et al. (2009) also observed that some participants tended to overweight inertial cues. Remarkably, this overweighting seemed to occur mostly for multisensory stimuli with the highest level (70%) of visual motion coherence (how many of the dots in the visual stimulus coherently moved in the same direction). In the present experiment, the dots in the visual cue moved in a fully coherent fashion (100%) and we observed overweighting of the inertial cue for every participant. Although the different experimental setup used by Fetsch et al. prohibits direct comparison of the results, it seems that the less reliable inertial information gets more weight when motion in a random-dot pattern becomes more coherent.

The observed overweighting of inertial cues might be explained by a violation of the assumption that sensory estimates are unbiased. Although this assumption is standard for MLI, it has recently been shown by Todd, Christensen, and Guckes (2010) that when this assumption does not hold, a 2AFC experiment can yield biased estimates of the size of the internal noise. Using these biased estimates in an integration scheme will result in erroneous estimates of the sensory weights. Although a 2AFC paradigm, such as used in the present study, does identify sensory noise (variable error), it does not provide information on a possible bias (constant error). Therefore, we cannot validate whether sensory estimates were biased. For this purpose, it would be necessary to measure heading judgments at a continuous scale.

An alternative explanation of this overweighting is that participants did not perceive the visual and inertial motions as associated. In the present study, inertial motion may have given a much stronger and compelling sensation

---

Figure 3. Standard deviations (SD; square root of the variance) for each participant and condition. The dots represent the observed inertial (blue), visual (red), and combined (green) visual/inertial SDs, respectively. The location of the black and green dots in terms of the x-coordinate indicates the associated visual weight. Black dots are the MLI-predicted optimal weights (x-coordinates) and corresponding SDs (y-coordinates). If MLI theory applies, the black and green dots should be close to each other.
of self-motion than the random-dot pattern. It is therefore likely that the visual cue was interpreted as a separate event. Although in the debriefing all participants asserted to have followed the instruction to use all sensory information available, they may instinctively have discarded the visual cue. If so, performance in the multisensory condition would correspond to performance in the inertial condition, which was the case for five out of nine participants. As noted by Fetsch et al. (2009), this is analogous to causal inference models (Körding et al., 2007; Sato, Toyoizumi, & Aihara, 2007). These integration models include a step prior to actual integration of multisensory estimates; it is first evaluated whether two cues arise from a single source or from multiple sources. In case multisensory cues are attributed to separate events, integration does not occur.

The quality of a visual stimulus itself may affect the extent to which it induces self-motion (or vection). It has been suggested that binocular visual cues are more effective in inducing vection than monocular cues, although Fetsch et al. (2009) did not find any effects of stereo vision on cue integration. However, it has been shown that a larger FoV (Allison, Howard, & Zacher, 1999) and photorealistic visual cues (Trutoiu, Mohler, Schulte-Pelkum, & Bültthoff, 2009) enhance vection. Compared to the stimuli used in the present experiment, the stimuli used by Fetsch et al., Gu et al. (2008), and Telford et al. (1995) were displayed with larger FoV; further, the stimulus used by Telford et al. depicted the actual surroundings, which may have increased vection. In future experiments, we plan to investigate the effects of these visual factors on cue integration.

**References**


