Learning to integrate contradictory multisensory self-motion cue pairings

Mariia Kaliuzhna
Center for Neuroprosthetics, School of Life Science, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Laboratory of Cognitive Neuroscience, Brain Mind Institute, School of Life Science, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Mario Prsa
Center for Neuroprosthetics, School of Life Science, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Laboratory of Cognitive Neuroscience, Brain Mind Institute, School of Life Science, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Steven Gale
Center for Neuroprosthetics, School of Life Science, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Laboratory of Cognitive Neuroscience, Brain Mind Institute, School of Life Science, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Stella J. Lee
Harvard–MIT Division of Health Sciences and Technology, Harvard Medical School, Cambridge, MA, USA

Olaf Blanke
Center for Neuroprosthetics, School of Life Science, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Laboratory of Cognitive Neuroscience, Brain Mind Institute, School of Life Science, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Department of Neurology, University Hospital, Geneva, Switzerland
Humans integrate multisensory information to reduce perceptual uncertainty when perceiving the world and self. Integration fails, however, if a common causality is not attributed to the sensory signals, as would occur in conditions of spatiotemporal discrepancies. In the case of passive self-motion, visual and vestibular cues are integrated according to statistical optimality, yet the extent of cue conflicts that do not compromise this optimality is currently underexplored. Here, we investigate whether human subjects can learn to integrate two arbitrary, but co-occurring, visual and vestibular cues of self-motion. Participants made size comparisons between two successive whole-body rotations using only visual, only vestibular, and both modalities together. The vestibular stimulus provided a yaw self-rotation cue, the visual a roll (Experiment 1) or pitch (Experiment 2) rotation cue. Experimentally measured thresholds in the bimodal condition were compared with theoretical predictions derived from the single-cue thresholds. Our results show that human subjects combine and optimally integrate vestibular and visual information, each signaling self-motion around a different rotation axis (yaw vs. roll and yaw vs. pitch). This finding suggests that the experience of two temporally co-occurring but spatially unrelated self-motion cues leads to inferring a common cause for these two initially unrelated sources of information about self-motion. We discuss our results in terms of specific task demands, cross-modal adaptation, and spatial compatibility. The importance of these results for the understanding of bodily illusions is also discussed.

Introduction

During passive self-motion (be it heading or rotation), an observer derives the direction, speed, and distance traveled from an optimal combination of redundant information provided by different sensory cues: visual, vestibular, auditory, and tactile sensations (Gibson, 1950; Kapralos, Zikovitz, Jenkin, & Harris, 2004; Warren & Wertheim, 2014). To date, in order to account for such processes, Bayesian statistics of sensory cue combination have been used, which describe how the perceptual uncertainty associated with the sensory cues (due to noisy sensory and neural processing) is reduced according to statistical optimality when multiple uncertain estimates of the same physical property are probabilistically combined (Ernst & Banks, 2002).

Computational theories of multisensory integration posit integration when a common cause is inferred for the sensory cues (Körding et al., 2007; Parise, Spence, & Ernst, 2012; Shams & Beierholm, 2010). The probability of a common cause depends on how similar or correlated these cues are and on the observer’s prior beliefs or knowledge about the possible common cause. In everyday natural settings, visual and vestibular stimuli signaling self-motion are most likely congruent and highly correlated, which results in their integration and mandatory fusion even in the presence of slight cue conflicts (Butler, Smith, Campos, & Bülthoff, 2010; Fetsch, Turner, DeAngelis, & Angelaki, 2009; Jürgens & Becker, 2006; Prsa, Gale, & Blanke, 2012). Accordingly, it is also assumed that integration should break down if the conflict between the stimuli is too large, implying that they relate to different sources. Such latter conflicts could also mean that there is no correlation between stimuli. In the present study we address the limits of visuo-vestibular integration by testing spatially conflicting multisensory (visuo-vestibular) cues.

Previous work has shown that large degrees of consciously detectable directional conflict do not lead to integration of visual, vestibular, and proprioceptive self-motion cues (Ohmi, 1996). Similar results have been obtained for conflicting cues pertaining to external objects (Bertelson & Radeau, 1981; Gebhard & Mowbray, 1959; Gepshtein, Burge, Ernst, & Banks, 2005; Lunghi, Morrone, & Alais, 2014; Pick, Warren, & Hay, 1969; Recanzone, 2003; Welch & Warren, 1980). Specifically, it has been found that integration occurs for small degrees of conflict, whereas for larger degrees of conflict such interactions between the two modalities are no longer observed (Gepshtein et al., 2005; Roach, Heron, & McGraw, 2006; Wallace et al., 2004).

It has, however, also been shown that multisensory cues can influence one another despite the absence of perceptual unification. Participants who initially do not integrate arbitrary cue pairings may learn to combine them when these stimuli temporally co-occur over time (i.e., arbitrary but correlated stimuli; Bresciani et al., 2005; Ernst, 2007; Wozny & Shams, 2011). Additionally, the particular demands of the task an observer has to perform might influence the integration process (Roach et al., 2006). It therefore appears possible that multiple parameters of conflicting stimuli could be taken into account to determine whether integration occurs: the amount of conflict, the task at hand, and the amount of correlated stimulus characteristics. The exact contribution of each of these parameters to the integration process remains unknown.

In our previous work, we have demonstrated that human observers integrate congruent yaw visual and yaw vestibular rotation cues according to statistical optimality (Prsa et al., 2012). Here, with an identical experimental apparatus and task design, we address whether repeated exposure to overtly spatially incongruent (but correlated on other dimensions) multisensory self-motion stimuli (visual and vestibular) results in optimal cue integration. Our participants were exposed to whole-body yaw rotations in conjunction with temporally synchronized optic flow rotation around a different axis (roll, Experiment 1). We asked
them to compare the sizes of two successively experienced rotation angles, and determined probabilistic descriptions of their perceptual estimates. Obtained results reveal that the variance associated with these incongruent visual–vestibular cue pairings decreased over time and progressively approached the statistically optimal predictions derived from the variances of the single-cue estimates. We argue that our results can be accounted for by a progressively increasing probability of attributing a common cause to the two incongruent stimuli. Further evidence for such integration was found when simultaneously exposing participants to whole-body yaw rotations in conjunction with synchronized optic flow rotation around the pitch axis (Experiment 2), which resulted in integration optimality right from the onset. We discuss the implications of our finding in the context of multisensory integration, own-body perception, and vestibular symptoms in neurological patients.

Materials and methods

Participants

Eight healthy adults naïve to the purpose of the study with normal or corrected vision and no history of inner-ear disease participated in each experiment (Experiment 1: two women, mean age = 24 ± 2.7 years; Experiment 2: three women, mean age = 22 ± 3.9 years). Three participants from Experiment 1 also participated in Experiment 2. All participants gave informed consent and received monetary compensation at 20 CHF/h. The studies were approved by a local ethics committee and were conducted in accordance with the Declaration of Helsinki.

Optimal Bayesian Estimator (OBE) model

In perceiving a whole-body rotation of size $S$, each sensory modality provides an independent estimate of $S$. Perceptual uncertainty is naturally associated with each of the unimodal estimates, the visual and the vestibular in our case, and can be measured as their trial-by-trial variance $\sigma_{vi}^2$ and $\sigma_{ve}^2$, respectively. Maximum likelihood estimation (derived from Bayes’s rule) dictates that if the two unimodal cues are integrated according to statistical optimality, the uncertainty associated with the bimodal estimate $\sigma_{bi}^2$ is reduced relative to unimodal uncertainties according to

$$\sigma_{bi}^2 = \frac{\sigma_{vi}^2 \sigma_{ve}^2}{\sigma_{vi}^2 + \sigma_{ve}^2}. \quad (1)$$

Experimental setup

Subjects were seated in a centrifuge cockpit-style chair which delivered passive whole-body rotational stimuli (Figure 1). After adopting a comfortable position, the subjects were restrained by a five-point racing harness, foot straps, and extra cushioning. To prevent the subject’s head from moving, a chin rest and a head fixation at the forehead were used.

The chair was digitally servo controlled (PCI-7352) and had very precise positioning of around 0.1°. The chair always rotated in the yaw plane and was centered on the rotation axis, restricting the vestibular stimuli to angular accelerations only. The rotation profiles of the chair were preset and designated the immediate angular position of the chair at a rate of 100 Hz. The velocity profile $v(t)$ of the rotations was a single cycle of a 0.77-Hz raised cosine function:

$$v(t) = \frac{A}{T} \left[1 - \cos\left(\frac{2\pi t}{T}\right)\right], \quad (2)$$

where $A$ is rotation size and $T$ is duration ($T = 1.3$ s in this case). Instantaneous angular position $p(t)$ was then specified as

$$p(t) = \frac{A}{T} \left[\frac{t}{T} - \frac{1}{2\pi} \sin\left(\frac{2\pi t}{T}\right)\right]. \quad (3)$$

The visual stimuli were presented on a 22-in. display, which was fixed to the chair in front of the subject at a distance of about 29 cm. The limited visual field covered ~80° of horizontal and 56° of vertical visual angle. The visual image consisted of a stereoscopic pattern of randomly distributed moving dots of different sizes. The dots were two-dimensional symmetric grayscale Gaussian blobs with a minimum and maximum standard deviation of 0.5 and 3 pixels, respectively. For each blob, the standard deviation was drawn from an exponential distribution with a rate parameter of 2, and the peak pixel intensity from a uniform distribution between 0.1 and 0.3 (1 denotes maximum intensity, i.e., white). The binocular disparity was a linear function of blob standard deviation and yielded minimum and maximum values of 0 (for the maximum-sized dots) and 50 (for the minimum-sized dots) pixels, respectively. All dots had therefore zero or positive stereoscopic depth. The dot density was set to 0.002 dots/pixel, producing roughly 3,500 dots in any given frame of the $1680 \times 1050$ resolution display. Their lifetime was not limited and their initial position reset at the start of every trial.

Rotation was simulated by placing the subject’s viewpoint in the middle of the scene and rotating it around either the roll axis (Experiment 1) or the pitch axis (Experiment 2). These patterns simulated the actual optic flow that would result from physically...
rotating the subject around these axes in congruent directions. During the rotations in all conditions, subjects were instructed to fixate a stationary central point, which was a filled red circle with a radius of 3 pixels and an intensity level of 0.5 presented at zero binocular disparity. The stereoscopic stimulus was generated by the Nvidia Quadro FX 3800 graphics card using the OpenGL quad-buffer mechanism. The stimulus was programmed with the Python language and viewed with the Nvidia 3D Vision kit (active shutter glasses) paired with a Samsung SyncMaster 2233RZ display (120-Hz refresh rate) via an infrared transmitter. The velocity of the optic flow matched that of the rotating chair. While subjects performed the task, masking white noise was presented over headphones.

**Experimental paradigm**

Subjects were seated in the rotating chair with a computer screen in front of them. On every trial they experienced two successive rotations (a standard and a test) and had to judge their relative size. The rotations were either delivered by the chair alone, simulated by the motion of the visual field on the display, or a combination of both (on every trial, the two rotations were of the same kind). The size of one of the two rotations was always 20° (i.e., the standard rotation), and the size of the second (i.e., the test rotation) was one of second equally spaced angles in the interval of 12°–27° tested using the method of constant stimuli (see Prsa et al., 2012, for a similar procedure). The two rotations were preceded, followed, and separated by an interval of 0.5 s. A 2-s period followed during which the subjects had to answer, via a button press, whether the second rotation was bigger or smaller than the first. The standard rotation was randomly assigned to come either first or second.

Subjects came in on two different days. On the first day we determined their unimodal discrimination thresholds (see Data analysis) for the vestibular and the visual modalities separately. In Experiment 1, participants performed the same task described previously for vestibular yaw rotations and visual roll rotations. In Experiment 2, the same task was performed, but with
vestibular yaw and visual pitch rotations. Participants first completed several blocks of vestibular-only stimuli, where the different test angles were presented in a randomized order. Each block contained 35 trials and lasted for about 5 min. Participants performed a minimum of 280 trials (in total), amounting to 40 trials per test angle. After extracting participants’ thresholds for the vestibular modality, we used the same procedure for the visual modality. In order to match participants’ visual thresholds to their vestibular thresholds, we manipulated the reliability of the visual stimulus by changing the coherence of the visual motion (number of dots simulating rotation or moving randomly). The random dots moved in a straight line with the identical displacement velocity profile as the rotation. The overall displacement size was limited to 200 pixels in horizontal and vertical directions and was drawn from a uniform distribution. The radial motion direction was randomly chosen for each dot between $-180^\circ$ and $180^\circ$ (uniform distribution). The random dots were also Gaussian blobs with identical parameters and therefore were visually indistinguishable (when stationary) from the blobs simulating rotation. Their initial positions were also reset at the start of each trial and their binocular disparity remained constant. Participants performed a minimum of four visual-only blocks (140 trials, 20 trials per test angle) with a given level of coherence. If their performance matched that on the vestibular modality, this level of coherence was retained; otherwise it was changed and the procedure continued as described until a matched level was obtained. The experimentally established levels of visual coherence corresponding to matched discrimination thresholds in the two single modalities were then used for Experiment 1; they were 100% for four subjects, 95% for one subject, 85% for two subjects, and 80% for one subject. Analogously, for Experiment 2 the levels used were 100% for three subjects, 95% for three subjects, and 85% for two subjects. Overall, subjects performed a mean of 557.5 trials for each modality (about 40 repetitions of each test angle for each modality).

On the second day, subjects performed the task but were now exposed to the three conditions: unimodal vestibular, unimodal visual, and both modalities together. For bimodal comparisons, visual and vestibular stimuli were temporally synchronized and occurred simultaneously (e.g., Prsa et al., 2012). The experiment was divided into sessions of approximately 5 min that we grouped into six blocks (the comparisons between these blocks allowed us to test for progressive learning of the visuo-vestibular association). Every session contained trials of each of the three conditions presented in a randomized order, which made it impossible to predict which condition would occur next. Subjects performed a total of 420 trials for each of the three conditions: 70 trials per block per condition, 10 trials per test angle per condition (i.e., 60 trials per test angle for each of the conditions). The direction of rotation (left or right for vestibular yaw and visual roll; up or down for pitch) was randomly chosen on each trial. In Experiment 1, left yaw rotations were always arbitrarily paired with right visual roll rotations (i.e., simulating a left roll self-rotation) and right yaw rotations with left visual roll rotations. In Experiment 2, left yaw was always arbitrarily paired with down visual pitch rotations (and right yaw rotations with up pitch rotations).

Data analysis

The data analysis was done using custom programs compiled in MATLAB (MathWorks). During the pretest, in order to match performance between the two modalities, we pooled the answers obtained for each test angle separately for the vestibular and visual conditions. For the analysis, the test angle was always compared to the standard angle, regardless of their order of occurrence in a trial. The proportion of “bigger” responses was calculated and fitted with a cumulative Gaussian function. From this fit we obtained discrimination thresholds for the two modalities. In order to match these thresholds, we manipulated the reliability of the visual cue as described earlier.

The analysis for the actual experiments was run in a similar fashion. Answers obtained for each test angle were pooled across all subjects to obtain a probabilistic measure and create a sufficient sample set for statistical comparisons. From the variance of the Gaussian fits to the proportion of “bigger” answers we obtained the measure of the discrimination threshold for each of the three conditions. We next conducted a bootstrap analysis to compare the predictions of the optimal observer model to the experimentally obtained values. To this end, we repeated the data fit for each condition 9,999 times, using a different subset of responses every time. The different subsets were formed by taking at random, with replacement, $N$ trials from the total set of $N$ for each test angle. The standard deviation of 9,999 repeated measures is then the standard error of the measure obtained using the original data set. Statistical tests were made by assessing the amount of overlap between the bootstrap iterations of two measures. If the measure of interest is $\sigma$ and $\sigma_{ex}$ and $\sigma_{pr}$ are its experimentally and predicted estimates obtained from the $j$th bootstrap sample, then the one-tailed bootstrap probability of $\sigma_{ex} > \sigma_{pr}$ is

$$p = \frac{1}{B} \sum_{j=1}^{B} I(\sigma_{ex}^j - \sigma_{pr}^j > 0),$$

(4)

where $B=9,999$ and $I()$ is the indicator function, which is equal to 1 when its argument is true and 0 otherwise. The
inequality would be reversed for the probability of $\sigma_{ex} < \sigma_{pr}$. The one-tailed bootstrap $p$ value is therefore simply the proportion of values of $\sigma_{ex} - \sigma_{pr}$ that are more extreme than 0. We prefer this approach to parametric testing because it provides a direct computation of the cumulative distribution of a test statistic instead of requiring the use of an asymptotic approximation.

The threshold values obtained through the bootstrapping were also analyzed using repeated-measures ANOVAs. In the group analysis (see Results), the threshold values pooled across subjects were used in a $6 \times 3$ ANOVA (with six blocks and three conditions as factors). For the single-subject analysis, the ANOVAs were performed on the bootstrapped values from single subjects for each block and condition. Bonferroni post hoc comparisons were used to explore the result of the interactions.

Results

For both experiments, the same analyses were performed. We first explored the results of our subjects as a group and then conducted more detailed analyses on single-subject data.

Experiment 1. Vestibular yaw and visual roll: Group analysis. A $6 \times 3$ repeated-measures ANOVA with blocks (six) and conditions (three) as factors yielded significant main effects and interaction (all $p$s < 0.0001). All Bonferroni post hoc comparisons were significant (all $p$s < 0.0001, except for Block 5 unimodal visual threshold not being different from Block 2 unimodal vestibular threshold). Next, we performed a one-tailed bootstrap analysis collapsing all blocks for all participants, which revealed a significant difference between the two single cues (vestibular threshold = 6.0, visual threshold = 5.3; $p = 0.004$). The bimodal threshold was significantly different from the best single cue, i.e., visual (bimodal threshold = 4.6; $p = 0.0001$), and from the threshold predicted by the OBE (predicted threshold = 3.97; $p = 0$).

We further divided the data (pooled across subjects, as described previously under Data analysis) according to the six experimental blocks that our participants performed. Figure 2 summarizes the experimentally obtained and predicted discrimination thresholds for each of the six experimental blocks. In the first four blocks, the difference between the best single-cue thresholds and the bimodal thresholds did not differ significantly (all $p$s > 0.1). Bimodal thresholds—those measured experimentally and those predicted by the OBE model—also differed significantly (one-tailed bootstrap test, $p < 0.03$), consistent with the absence of statistically optimal visual–vestibular integration. A significant difference between the best single-cue thresholds and the empirically measured bimodal thresholds emerged only in the last two blocks ($p < 0.02$), signaling integration. In these last two blocks, participants’ bimodal thresholds also became not statistically different from the predicted thresholds ($p > 0.05$; Figure 2).

We next conducted a more detailed analysis of each condition separately (visual, vestibular, and visual-vestibular) over the six blocks. Table 1 contains the results of the linear regression analysis for each condition over the experimental blocks as well as the $p$ values of the bootstrap test comparing the thresholds of the first and last blocks. For the unimodal vestibular condition, no significant changes of the discrimination thresholds were observed over time ($r^2 = 0.31$, $p = 0.25$). The same analysis for the unimodal visual condition showed a significant difference between the first and last experimental blocks ($p = 0.005$) due to a linear
increase of the threshold values ($r^2 = 0.89, p = 0.005$).
Finally, the bootstrap test for the bimodal condition showed that the thresholds did not significantly change between the initial and final blocks ($p = 0.19$; a linear change was not observed, $r^2 = 0.03, p = 0.73$). Contrary to the ANOVA, there was no significant difference between the unimodal vestibular and visual threshold values in any of the blocks (although a trend for a difference was found in Blocks 2, 3, and 5, $p = 0.06, 0.06, and 0.07$, respectively), indicating well-matched single-cue reliabilities between the two sensory modalities. This suggests that the emergence of optimal multisensory integration for incongruent but temporally co-occurring visual–vestibular stimuli is in our case revealed by a stable bimodal threshold which did not accompany a progressive increase in the visual threshold, thereby exposing a relative reduction of perceptual variance when cues are combined.

### Experiment 1. Vestibular yaw and visual roll: Single-subject analysis

We performed $6 \times 3$ (block $\times$ condition) repeated-measures ANOVAs on the values for each subject generated by the bootstrap procedure. Bonferroni post hoc comparisons between the bimodal and the best unimodal cue showed integration for four subjects: Subject 3 in Block 6, Subject 4 in all blocks except Block 4, Subject 5 in Block 5, and Subject 6 in all blocks except Blocks 3 and 6. To quantify the extent of overall integration for each subject, we performed a within-subject test comparing the bimodal threshold to the predicted threshold values (Figure 4; Table 2). The tests were performed by pooling all blocks together per subject in order to yield enough data points for a statistical comparison. Four subjects out of eight in Experiment 1 showed optimal integration. Out of the four remaining subjects for whom this analysis showed no optimal integration, Subject 1 showed integration in the last two blocks (optimal in Block 6), Subject 3 showed integration only in Block 6, Subject 5 only in Block 5, and Subject 8 in Blocks 4 and 6 (Figure 4). The differences between this analysis and the results of the ANOVA are due to the latter results being skewed by high threshold values in one of the conditions in one of the blocks for some subjects.

To assess intersubject variability, we analyzed the performance of individual subjects in each of the six experimental blocks (Figure 5). To quantify the extent of integration, we subtracted the bimodal threshold values (white bars) and the values predicted by the OBE model (red lines) from the best single-cue thresholds.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Linear regression $R^2$</th>
<th>Linear regression $p$ value</th>
<th>Bootstrap $p$ value</th>
<th>Blocks 1 to 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1 (yaw + roll)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vestibular</td>
<td>0.31</td>
<td>0.25</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>visual</td>
<td>0.89</td>
<td><strong>0.01</strong></td>
<td><strong>0.01</strong></td>
<td></td>
</tr>
<tr>
<td>bimodal</td>
<td>0.03</td>
<td>0.73</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>Experiment 2 (yaw + pitch)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vestibular</td>
<td>0.90</td>
<td><strong>0.00</strong></td>
<td><strong>0.02</strong></td>
<td></td>
</tr>
<tr>
<td>visual</td>
<td>0.01</td>
<td>0.88</td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>bimodal</td>
<td>0.04</td>
<td>0.71</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Between-block comparisons for each condition in both experiments. Notes: $R^2$ and $p$ values of the linear regression and bootstrap analysis within each condition for the two experiments. Bold values represent a significant change in threshold values across blocks.
Positive values represent cue integration, which may or may not be optimal (this can be assessed by the proximity of the white bars and the red lines: White bars at the level of or above the red lines suggest optimality). Subject 1 is an example participant who learns to integrate the two cues only in the last two experimental blocks (negative white bars in all blocks but 5 and 6); the integration moreover seems to reach optimality in Block 6.

The results of Experiment 1 indicate that it is possible to learn to integrate visuo-vestibular cue pairings, which each signal self-motion around a different axis. In order to further corroborate this finding, we decided to use another set of stimuli that also have a high degree of disparity: vestibular yaw and visual pitch rotations.

**Experiment 2. Vestibular yaw and visual pitch: Group analysis**

A $6 \times 3$ repeated-measures ANOVA with blocks (six) and conditions (three) as factors yielded significant main effects and interaction (all $p < 0.0001$). All Bonferroni post hoc comparisons were significant (all $p < 0.0001$, except for Block 6 unimodal vestibular threshold and Block 4 unimodal visual threshold not being different from Block 2 unimodal visual threshold). Next, a general one-tailed bootstrap analysis collapsing across the experimental blocks and participants revealed the following results: There was a significant difference between the visual and vestibular thresholds (vestibular threshold $= 5.2$, visual threshold $= 4.7$; $p = 0.01$), and the bimodal threshold differed significantly from the best single cue, i.e., visual (bimodal threshold $= 3.7$; $p = 0$), but not from the threshold predicted by the OBE model (predicted threshold $= 3.5$; $p = 0.06$).

For further analysis, we divided the data (pooled across subjects, as described under Data analysis and done in Experiment 1) according to the six blocks performed. Figure 3 summarizes the experimentally obtained and predicted discrimination thresholds for each of the six experimental blocks. For the yaw–pitch combinations, participants’ bimodal thresholds were not significantly different from the optimal prediction in all but the second experimental block (one-tailed bootstrap test, $p > 0.05$), and in the same blocks, the

<table>
<thead>
<tr>
<th>Subject</th>
<th>Experiment 1 (yaw + roll)</th>
<th>Experiment 2 (yaw + pitch)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bimodal to optimal</td>
<td>Bimodal to best single cue</td>
</tr>
<tr>
<td>1</td>
<td>0.05</td>
<td>0.27</td>
</tr>
<tr>
<td>2</td>
<td>0.39</td>
<td>0.05</td>
</tr>
<tr>
<td>3</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>4</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>6</td>
<td>0.39</td>
<td>0.01</td>
</tr>
<tr>
<td>7</td>
<td>0.44</td>
<td>0.02</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Table 2. Single-subject analysis. $P$ values of the bootstrap comparison. Notes: $P$ values of the one-tailed bootstrap analysis testing whether the bimodal thresholds are greater than the theoretically predicted values and whether the bimodal thresholds are lower than the best single cue for each subjects for all experimental blocks pooled together. Numbers in bold indicate the subjects who integrated (optimally).
experimentally measured bimodal thresholds were significantly lower than the best single-cue estimates ($p < 0.05$). In the case of yaw–pitch pairings, subjects therefore optimally integrated the two incongruent cues from the onset and throughout the tested experimental blocks (except Block 2).

As for Experiment 1, we performed a detailed analysis of the discrimination thresholds in each condition over the six blocks. The regression and the bootstrap test results are summarized in Table 1. For the unimodal vestibular condition, a significant difference was observed between the first and last experimental blocks ($p = 0.022$) due to a linear decrease of the threshold values ($r^2 = 0.9$, $p = 0.004$). For the unimodal visual condition, there was no change in the discrimination thresholds over time ($r^2 = 0.006$, $p = 0.88$). There was a significant difference between the two unimodal conditions, but only in the first two experimental blocks, once again indicating well-matched single-cue discrimination thresholds. In the bimodal condition, the threshold values showed no overall linear change ($r^2 = 0.04$, $p = 0.7$) and no significant difference between the start and end of the experimental session ($p = 0.25$).

**Experiment 2. Vestibular yaw and visual pitch: Single-subject analysis**

With $6 \times 3$ (block $\times$ condition) repeated-measures ANOVAs on the values for each subject generated by the bootstrap procedure and subsequent Bonferroni post hoc comparisons between the bimodal and the best unimodal cue, we found integration for five subjects: Subject 2 in all blocks but 4 and 6; Subject 3 in all blocks but Block 4; Subject 4 in Blocks 1, 3, and 4; Subject 5 in all blocks but 1 and 4; and Subject 6 in all blocks but 2 and 3. The results of single-subject bootstrap analysis from Experiment 2 are shown in Figures 4 and 5 (see also Table 2). This analysis revealed that seven out of eight subjects in Experiment 2 showed optimal integration. The difference between the bootstrap test and the ANOVA is again due to high variance in one block and condition skewing the results of the ANOVA. An example of a participant who integrates throughout Experiment 2 is Subject 5 (white bars have a positive value in each experimental block), and this integration is optimal (red lines at the level of or overlapping with the white bars). This is different from Subject 4, whose bimodal thresholds are only

![Figure 5. Individual subject data. Difference between the lowest single-cue standard deviation and the measured (white bars) and predicted (red lines) bimodal standard deviations across the six experimental blocks. Positive values indicate integration. (A) Experiment 1 (yaw + roll). (B) Experiment 2 (yaw + pitch).](https://journalofvision.org/15/10/9/article/1)}
lower than the best unimodal thresholds in half of the
blocks (Blocks 1, 3, and 4) and whose integration is
optimal only in Blocks 1 and 4.

### Discussion

The main finding of the present study is that human
observers can optimally integrate or learn to integrate
coccurring multisensory self-motion stimuli when
those stimuli imply rotations around different axes. In
two experiments, we showed that participants’ perfor-
man ce was better for directionally conflicting bimodal
visuo-vestibular cues than for either vestibular or visual
cues alone. In the yaw–roll experiment (Experiment 1),
such optimal integration improved over successive
blocks, whereas in the yaw–pitch experiment (Experi-
m ent 2) it was observed in all but one experimental
block. These results extend the previous literature on
self-motion perception, demonstrating that optimal
integration also occurs for stimuli with large, con-
sciously detectable discrepancies.

It has been proposed that sensory integration would
only occur for stimuli attributed to the same causal
event (Körding et al., 2007; Parise et al., 2012; Shams &
Beierholm 2010). The present results would mean that
in the bimodal condition, instead of experiencing two
distinct rotational stimuli as provided by the incon-
gruent visual and vestibular modalities, subjects per-
cieve one single self-displacement, possibly going in an
intermediate direction with respect to the two cues.
Compatible with such a proposal are data from
previous studies showing that selected visuo-vestibular
conflicts are perceived as one single motion (Ishida,
Fushiki, Nishida, & Watanabe, 2008; Wright, DiZio, &
Lackner, 2005). Thus, incongruent visual and vestibu-
lar cues along the same yaw axis but indicating yaw
rotations in the same direction (an ecological conflict—
e.g., clockwise vestibular yaw and clockwise visual yaw)
are perceived as rotations that depend more strongly on
the direction of the visual stimulus, compatible with
visual dominance. Findings with the vestibular-ocular
reflex (VOR) are also compatible with such visual
dominance. The VOR occurs when the head moves or
when vection is perceived, and keeps the image stable
on the retina by moving the eyes in the opposite
direction. Ishida et al. (2008) found that the direction of
the VOR in cases of nonecological visual-vestibular
stimulus combinations in the yaw axis also reveals such
visual dominance and is congruent with the visual
stimulus. Also of relevance are VOR studies of cross-
modal adaptation between visual and vestibular cues
(indicating motion in different directions) that have
been performed in several species (Baker, Wickland, &
Peterson, 1987; Schultheis & Robinson, 1981; Trilren-
berg, Shelhamer, Roberts, & Zee, 2003). Subjects were
exposed for a certain period of time to simultaneously
presented conflicting stimuli (different axes from those
employed in the present study), and adaptation was
reflected in the change of the direction and gain of the
VOR. These studies indicate that a visuo-vestibular
conflict, in certain conditions, is formulated by the
nervous system into one single percept (i.e., conscious
perception of self-motion in one particular direction).

Based on our own finding of optimal integration in
Experiments 1 and 2 and these previous findings, we
argue that in the incongruent visual–vestibular combi-
nations (bimodal conditions), subjects might perceive a
single self-displacement. Unfortunately, no subjective
reports of the perceived direction of motion have been
collected in the present study or in the previous
literature. In our study, only one subject (in Experiment
2) spontaneously reported a single illusory diagonal
displacement during the bimodal (yaw + pitch)
condition. We propose that future work could extend
the present paradigms and additionally record eye
movements in order to investigate whether the two
conflicting motion directions are combined into a single
representation and whether the resulting integrated
percept depends more on the vestibular or the visual
cue (as reported in related work by Ishida et al., 2008).
Although the direction of perceived motion might be
inferred from the direction of fixating microsaccades,
the use of a stereoscopic visual stimulus necessitating
shutter glasses prevented us from recording eye
movements by means of video tracking. In previous
studies (Ishida et al., 2008; Trillenberg et al., 2003),
subjects were exposed to the same continuous stimu-
lation for a prolonged amount of time (one to several
hours) before the adaptation could be objectified with
the eye-movement recording. The present findings,
however, suggest that such motion integration for
incompatible directions may occur much faster than
previously thought.

The observed visual–vestibular integration may also
depend on the demands of the experimental task and
the particular stimuli chosen. Thus, multisensory
integration in our study could have occurred as a result
of the specific task demands, which may have made the
directional conflict irrelevant to the task. Thus, in
previous experiments on multisensory conflicts (includ-
ing visual–vestibular stimulation but also other multi-
sensory stimulus combinations), the response of the
subject (e.g., localizing a stimulus, judging the number
of stimuli, judging perceptual qualities of stimuli) was
found to depend on the amount of conflict between the
two stimuli. For instance, in a task where subjects are
asked to localize a visual stimulus in the presence of an
auditory cue from a conflicting location, perceptual
unification breaks down and the two modalities bias
each other less when the distance between the visual and
the auditory cues is too large (e.g., Roach et al., 2006; Wallace et al., 2004). In the present study, however, the direction of motion was irrelevant to the task we asked our participants to perform, because estimating rotation size is independent of rotation direction. As the amount of rotation provided by the visual and the vestibular cues was always the same in the present bimodal conditions, one could expect that the extraction of this feature alone could lead to the observed visual–vestibular patterns of integration. This extraction could further have been facilitated by the fact that other stimulus features were matched for the visual and vestibular stimuli despite their directional conflict. Thus the motion onset, the duration, and the spatiotemporal motion profile of both stimuli were matched. In the same vein, a recently published work reports, for instance, that simultaneous visuo-vestibular stimuli indicating the same heading direction but having a different acceleration profile are optimally integrated (Butler, Campos, & Bülthoff, 2014). Another unpublished work reports that stimuli designating the same amount of motion and having the same motion profiles (but being temporally offset) can also be learned to be integrated (Campos et al., 2009). These data indicate that, despite the fact that some properties of the two stimuli are not matched, integration occurs when the task-relevant features of these stimuli are not in conflict (i.e., most features of the stimuli are correlated). Further work is needed to disentangle the different contributions of stimulus attributes and task demands for the integration process.

Our results show that integration in the bimodal condition occurred from the beginning in Experiment 2 (vestibular yaw + visual pitch), whereas such integration only appeared during the later phases of Experiment 1 (vestibular yaw + visual roll). That is, integration of pitch with yaw was present throughout the experiment, whereas integration of roll with yaw was learned over time. To our knowledge, there exists no anatomical or functional evidence for a facilitated integration of pitch with yaw versus roll with yaw stimuli. Neither at the level of the vestibular nuclei (Büttner-Ennever, 1992; Hightstein & Holstein, 2006; Naito, Newman, Lee, Beykirch, & Honrubia, 1995) nor in the cortex (Arnoldussen, Goossens, & van den Berg, 2013) can the pattern of projections from the semicircular canals account for our findings. Although recordings of neural responses to vertical rotations reveal that roll neurons outnumber pitch neurons in the brain stem (Baker, Goldberg, Hermann, & Peterson, 1984; Bolton et al., 1992; Endo, Thomson, Wilson, Yamaguchi, & Yates, 1995; Kasper, Schor, & Wilson, 1988; Wilson, Yamagata, Yates, Schor, & Nonaka, 1990), optimal activations of cortical vestibular neurons are uniformly distributed over all possible rotation planes (Akbarian et al., 1988; Grüsser, Pause, & Schreiter, 1990). Other studies looking at conflicting visuo-vestibular stimuli (e.g., Bockisch, Straumann, & Haslwanter, 2003; Waespe & Henn, 1978) have failed to provide the comparisons relevant to our study and findings (different stimulation axes and parameters of stimulation).

We do not think that our results can be attributed to the fact that participants performed in general better in Experiment 2 (lower thresholds for the unimodal conditions). Nor could this result be attributed to the fact that three subjects participated in both experiments (i.e., learned integration from Experiment 1 influenced thresholds in Experiment 2): Their performance was comparable in both experiments, and all three subjects showed integration in the majority of experimental blocks in both experiments. Accordingly, we propose that the difference between yaw–roll and yaw–pitch integration may be caused by supravestibular directional influences that have been observed previously in cognitive neuroscience. In Experiment 1, vestibular clockwise (i.e., rightward) yaw was always paired with visual leftward roll, and in Experiment 2, vestibular rightward yaw was always paired with upward pitch. Previous research on spatial compatibility (e.g., Simon effect, spatial Stroop, mental number line) has shown a facilitation effect for stimuli occurring in the same spatial plane (e.g., right + right) and for left–down/ right–up pairings (Cho & Proctor, 2003; Nicoletti & Umlit& auml; , 1984; Nishimura & Yokosawa, 2006). For instance, participants are faster to respond to visual stimuli presented on the left or at the bottom of the screen with their left hand and to rightward and upward stimuli with their right hand (for a review, see Proctor & Cho, 2006). Similarly, if vocal responses “right” and “left” are attributed to stimuli presented above or below the midline of the screen, such a random pairing yields faster responses than pairing “right” with below and “left” with above (Weeks & Proctor, 1990).

Such compatibility effects have also been shown for multisensory stimuli. For example, a high-frequency tone and a tactile stimulus at a higher location presented together (and a low-frequency tone plus a tactile stimulus at a lower location) are more strongly associated than, e.g., a high-frequency tone with a tactile stimulus at a lower location (Occhielli, Spence, & Zampini, 2009). Such cross-modal mechanisms have also been shown for the vestibular system: Active head turns (Loetscher, Schwarz, Schubiger, & Brugger, 2008) as well as passive displacements (Hartmann, Grabherr, & Mast, 2012) to the right and left have been found to influence numerical cognition in a magnitude-specific way. Leftward movements facilitate the generation of smaller numbers and rightward motion that of larger numbers. A related supravestibular directional mechanism could have influenced responses in the present two experiments, meaning that processing the association of
the bimodal stimuli we have chosen is a priori facilitated due to the correspondence of spatial representations or dimensional overlap of their directions (Kornblum, Hasbroucq, & Osman, 1990; Kosslyn & Kosslyn, 1996; Li, Nan, Wang, & Liu, 2014). Given the strong implication of vestibular signals for space perception, the existence of such automatic associations seems highly plausible. Indeed, the pairing of visual (simulated) rightward roll with rightward vestibular yaw and rightward vestibular yaw with upward pitch might represent a preferred direction for integration, potentially providing a partial explanation for our results.

Alternatively, other characteristics shared between yaw and pitch stimuli might contribute to the faster integration observed in Experiment 2. For instance the gain of the VOR elicited by both yaw and pitch movement or optic flow is generally close to 1.0, whereas the gain of the torsional VOR elicited by roll movement is generally more limited (Tweed et al., 1994). The optic flow resulting from yaw and pitch movements involves the motion of the whole visual field in one direction, whereas in the roll plane the visual scene rotates around a central point (Duffy & Wurtz, 1995). Finally, it may be that in daily life, combined yaw–pitch head motion (i.e., up and to the right) is more frequent than combined yaw–roll motion, although this speculation has yet to be confirmed by empirical research. These natural constraints may separately or in combination with the mentioned supravestibular mechanism result in a multisensory system that is more tolerant to conflicts between stimuli sharing a larger number of common characteristics.

Although of no direct relevance for the integration results we report, it should be noted that the fluctuations of unimodal thresholds were not the same in the two experiments. During Experiment 1, the unimodal visual threshold became progressively more elevated over consecutive sessions, which was not observed in Experiment 2. We speculate that this change can be attributed to fatigue or decreased attention over time as the subjects took part in a prolonged experiment requiring a continued high level of visual attention. In Experiment 2, however, the vestibular threshold was reduced over time. We presume that this reduction can be attributed to perceptual learning or improvement as subjects repeatedly experience the same stimulus. It is safe to assume that these idiosyncratic phenomena are also present and affect perception in the same way when the unimodal stimuli are paired with one another in the bimodal condition. The increasing visual thresholds in Experiment 1 and the decreasing vestibular thresholds in Experiment 2 therefore impact the bimodal prediction as well. Despite these changes, the experimental bimodal threshold still closely matched this prediction, thus showing optimal integration.

Finally, our result may be of relevance for visual–vestibular integration in neurological patients. It has been suggested that illusory own-body perceptions such as room-tilt illusions, inversion illusions, and out-of-body experiences are related to abnormal multisensory integration involving the visual and the vestibular senses (Blanke, 2012; Ionta et al., 2011; Lopez, Halje, & Blanke, 2008). Our experiments show that conflicting information from these two modalities is optimally integrated by the brain, possibly to produce a single percept and thus merge contradictory visual and vestibular self-motion cues into one coherent representation. We argue that such a single representation may also account for illusory own-body perceptions related to self-location and self-motion during such neurological conditions. Abnormal perception in these illusory states could be due to a statistically optimal integration of abnormal visuo-vestibular cue pairings instead of, as often claimed, a failure to integrate multisensory cues which may be providing conflicting information about self-motion and self-location. Neurological patients with out-of-body experiences caused by cortical damage (Ionta et al., 2011) and healthy subjects who are prone to out-of-body experiences (e.g., Murray & Fox, 2005) may integrate visual and vestibular stimuli across a larger range of stimulus incompatibilities than subjects without such experiences. This interpretation, however, is to be taken with caution, as it is yet to be supported by experimental evidence and the illusory own-body perceptions are related to abnormal processing of gravitational information (i.e., depend on the otolith organs), whereas our experimental manipulations involved vestibular yaw rotations and only partly involved visual gravitational stimulation.

**Keywords:** multisensory integration, multisensory conflict, self-motion, vestibular

### Acknowledgments

Support for this work was provided by the VERE project (FP7-ICT-2009-5, Project 257695) and the Bertarelli foundation.

Commercial relationships: none.
Corresponding author: Olaf Blanke.
Email: olaf.blanke@epfl.ch.
Address: Center for Neuroprosthetics and Brain Mind Institute, Faculty of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
References


Hartmann, M., Grabbherr, L., & Mast, F. W. (2012). Moving along the mental number line: Interactions...


