Visual and vestibular perceptual thresholds each demonstrate better precision at specific frequencies and also exhibit optimal integration

Faisal Karmali,
Koeun Lim,
and Daniel M. Merfeld

Jenks Vestibular Physiology Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts; and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts

Submitted 7 May 2013; accepted in final form 24 December 2013

Karmali F, Lim K, Merfeld DM. Visual and vestibular perceptual thresholds each demonstrate better precision at specific frequencies and also exhibit optimal integration. J Neurophysiol 111: 2393–2403, 2014. First published December 26, 2013; doi:10.1152/jn.00332.2013.—Prior studies show that visual motion perception is more precise than vestibular motion perception, but it is unclear whether this is universal or the result of specific experimental conditions. We compared visual and vestibular motion perception over a broad range of temporal frequencies by measuring thresholds for vestibular (subject motion in the dark), visual (visual scene motion) or visual-vestibular (subject motion in the light) stimuli. Specifically, thresholds were measured for motion frequencies spanning a two-decade physiological range (0.05–5 Hz) using single-cycle sinusoidal acceleration roll tilt trajectories (i.e., distinguishing left-side down from right-side down). We found that, while visual and vestibular thresholds were broadly similar between 0.05 and 5.0 Hz, each cue is significantly more precise than the other at certain frequencies. Specifically, we found that 1) visual and vestibular thresholds were indistinguishable at 0.05 Hz and 2 Hz (i.e., similarly precise); 2) visual thresholds were lower (i.e., vision more precise) than vestibular thresholds between 0.1 Hz and 1 Hz; and 3) visual thresholds were higher (i.e., vision less precise) than vestibular thresholds above 2 Hz. This shows that vestibular perception can be more precise than visual perception at physiologically relevant frequencies. We also found that sensory integration of visual and vestibular information is consistent with static Bayesian optimal integration of visual-vestibular cues. In contrast with most prior work that degraded or altered sensory cues, we demonstrated static optimal integration using natural cues.

We perceive self-motion with exquisite precision. Prior studies show that visual perception is more precise than vestibular perception when discriminating linear motion (Butler et al. 2010; Fetsch et al. 2009, 2012; Gu et al. 2008; MacNeilage et al. 2010). However, it is unclear whether visual precision is universally better than vestibular precision, or if these findings are the result of specific experimental conditions (e.g., the motion frequency content, where frequency always refers to temporal, not spatial, frequency).

Motor reflexes suggest that vestibular cues can be more salient than visual cues, particularly at high frequencies. Specifically, gaze stability is better for vestibular than visual disturbances at higher frequencies, and vice versa at lower frequencies (Crawford 1964; Paige 1983; Robinson 1977). Similarly, postural stability in patients with vestibular deficits is more affected at high than low frequencies (Bles et al. 1983; Peterka and Benolken 1995). However, higher gain responses do not dictate higher precision. Furthermore, perception need not mimic reflexive findings, as demonstrated by qualitative differences in sensory processing in the vestibulo-ocular reflex (VOR) and vestibular perception (Merfeld et al. 2005a, 2005b). Therefore, we chose to measure perceptual thresholds to assay perceptual precision [as per conventional definitions, precision varies inversely to threshold and is defined as the inverse of variance i.e., 1/σ², and is analogous to reliability (Faisal et al. 2008)].

By determining thresholds as a function of frequency, this study combines two classic approaches, measuring thresholds and assessing dynamics, to study 1) how the relative precision of sensory modalities varies with stimulus frequency; 2) how the brain combines sensory cues; and 3) dynamic neuroperception. Thresholds have helped understand the relative precision of sensory modalities (e.g., Ernst and Banks 2002; Gu et al. 2008), and finding frequency responses generally enhances our understanding of dynamic neuroprocessing (e.g., Angelaki and Yakusheva 2009; Cullen 2011; Merfeld et al. 2005a; Valko et al. 2012). To our knowledge, visual and vestibular thresholds have not previously been compared across frequencies with the same motion axis, methods and subjects.

In this study, we compare the frequency response of the visual and vestibular perceptual systems by measuring thresholds across a broad range (0.05–5 Hz) of physiological motion frequencies. Specifically, we characterized the frequency response of the perceptual systems that discriminate leftward from rightward roll tilt in three conditions: “visual” (stationary subjects watched a rotating visual scene), “vestibular” (subjects rotated in the dark) and “visual-vestibular” (subjects rotated in the light while viewing an earth-stationary scene). We found frequency ranges where vestibular thresholds were lower than visual thresholds, demonstrating that visual precision is not universally better than vestibular precision.

In addition, we used this frequency-response approach to investigate the integration of visual and vestibular tilt cues. Previous threshold studies have reported static Bayesian optimal cue integration for rotation (Jurgens and Becker 2006), translation (Butler et al. 2010; Fetsch et al. 2009, 2010, 2012; Gu et al. 2008), and static roll orientation (Clemens et al. 2011; De Vrijer et al. 2009) but visual-vestibular cue integration using thresholds and dynamic tilt has not been reported. Furthermore, unlike the earlier studies that degraded visual cues (Butler et al. 2010; Fetsch et al. 2009, 2012; Gu et al. 2008) or put two sensory cues in conflict (Butler et al. 2010; Fetsch et al. 2009, 2012; Jurgens and Becker 2006), we used natural cue
METHODS AND MATERIALS

General methods mimicked recent studies (Grabberr et al. 2008; Valko et al. 2012). Data collection was divided into two studies. In study A, we measured and compared thresholds between 0.05 Hz and 5 Hz; our primary goal in this study was to compare visual precision with vestibular precision across a broad range of frequencies. In study B, we made more meticulous threshold measurements between 1 and 5 Hz, a frequency range that included frequencies (e.g., 2 Hz) where the visual and vestibular thresholds were similar; our primary goals in study were to test the optimal integration hypothesis and to test the hypothesis that vestibular thresholds were lower than visual thresholds above 2 Hz.

Subjects. Thirteen healthy volunteers (7 men, 6 women; 36 ± 11 yr) participated. Six subjects participated in study A, and five in study B. One subject withdrew before completing testing. We also omitted one subject from our data analysis. Our hypothesis was that vestibular thresholds were lower than visual thresholds. All subjects underwent a clinical vestibular diagnostic exam to screen for undiagnosed vestibular disorders, which consisted of Hallpike testing, caloric electroneystagmography, angular VOR evoked via rotation and postural control measures. Subjects also answered a questionnaire to indicate any history of dizziness or vertigo; back/neck problems; cardiovascular, neurological and other physical problems; as well as motion sickness history. All subjects underwent a clinical vestibular diagnostic exam to screen for undiagnosed vestibular disorders, which consisted of Hallpike testing, caloric electroneystagmography, angular VOR evoked via rotation and postural control measures. Subjects also answered a questionnaire to indicate any history of dizziness or vertigo; back/neck problems; cardiovascular, neurological and other physical problems; as well as motion sickness history. Informed consent was obtained from all subjects prior to participation in the study. The study was approved by the local ethics committee and has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Testing conditions and motion stimuli. Motion was in roll tilt about a nasooccipital axis through head center at the level of the vestibular organs and was provided by a motion platform (MOOG GDF02000E) controlled by a computer in real time.

Each subject was tested in each of three conditions, vestibular, visual or visual-vestibular, during separate sessions. In the vestibular condition, subjects sat on the motion platform and were tilted in roll to the left or right in the dark. In the visual condition, subjects sat in a stationary chair and viewed the roll motion of a visual scene mounted on the motion platform. In the visual-vestibular condition, subjects sat on the motion platform and were tilted in roll while viewing the earth-stationary visual scene. Our vestibular stimuli activated both the semicircular canals (SCC) and otolith organs, providing both motion and displacement cues. Similarly, our visual stimuli provided both motion and displacement cues. Using the same platform to provide motion for the subject or the visual scene ensured that the motion dynamics were matched across conditions.

We desired to determine and compare thresholds that were relevant to real-life, physiological situations, so we chose to make the visual scene as realistic as possible, instead of providing a computer-displayed view. Of course this realism limits our ability to manipulate visual cues experimentally, but realistic visual scenes are likely to provide near-maximal information compared with sinusoidal gratings or random dot patterns. The realistic view (Fig. 1) was dominated by

a large poster of a forest and also included other cues typical of our experiment rooms such as a vertical wire bundle. The forest poster (cropped version of “Ashley, Aspens NF” by Alain Thomas, 1.0 m × 0.4 m) was mounted directly in the subject’s line of sight and attached 0.05 m in front of the background. The edges of the poster and the trees in the forest provided strong vertical and horizontal cues. The subjects saw the view through a round aperture (0.30 m diameter) made from a block of black plastic placed near the normal visual near point; the peripheral visual scene consisted of a low-contrast black fabric that prevented use of body or motion device landmarks to judge motion of the visual scene (e.g., by visually aligning their knee with an object). In study A, for the visual-vestibular condition, the poster was mounted on the wall of the experiment room and the view also included vertical wire bundles. For all of study B and for the visual condition in study A, the forest poster was mounted on a life-sized realistic print of a photograph of the real wall scene described above that was used in study A for the visual-vestibular condition. The print included the vertical wire bundles, and the forest poster was attached 0.05 m in front of the print. The photograph was taken from the perspective of the subject sitting on the motion platform using a Nikon Digital SLR camera. Two subjects independently reported during visual-only testing that they felt that they were looking at the actual wall of the experiment room and doing the visual-vestibular testing, rather than looking at the large print and doing visual-only testing. This anecdotally suggests that the large print was compelling.

In study A, the aperture was mounted 0.24 m from the eyes, the view was 1.3 m from the eyes, and the field of view (FOV) was 63°. The forest poster filled more than one-half of the horizontal FOV and one-fourth of the vertical. In study B the aperture was mounted 0.28 m from the eyes, the view was 0.95 m from the eyes and the FOV was 56°. The forest poster filled most of the horizontal FOV and one-third of the vertical. The difference in display and overall FOV between the two studies is not large. More fundamentally, the differences do not affect our conclusions, because we do not compare data across the two
We conducted a number of verifications to ensure that actual motion matched desired motion, since mechanical dynamics could result in a mismatch, especially at higher frequencies. Before beginning the studies presented herein, we performed a detailed calibration between 0.01 Hz and 5 Hz using a 6 degree-of-freedom sensor (Analog Devices ADIS16360) attached directly to the Moog platform. This calibration procedure included the application of single-cycle sinusoidal acceleration like those used herein, as well as standard multicycle steady-state sinusoidal oscillations. During this calibration process, we designed a simple first-order compensator to make device motion closer to the desired performance. The compensators had the form \((b_1 + b_2 z)/(1 + a_1 z)\), where \(z\) represents the standard discrete-time \(z\)-transform for a Moog 60-Hz update and sampling rate. With this compensation, actual roll tilt motion was within 0.3 dB (i.e., accurate to within 4%) of desired motion at frequencies of 0.01 Hz to 5 Hz. We also investigated whether there were any nonlinearities whereby motion characteristics varied with amplitude and found that our device is capable of producing the desired motions for both small and large displacements (Fig. 3A). We also confirmed that the Moog feedback signal accurately measured motion by independently measuring motion using an accelerometer mounted to the helmet support structure (Fig. 3B). We note that small platform motion errors would affect visual and vestibular conditions equally, so it would not impact the visual-vestibular threshold comparison, which is our primary finding. Another consideration is that, although we used a split helmet “vice” that firmly held the head in roll, it is still possible that inertia would cause the head to slip slightly, resulting in slower head motions during the vestibular condition at high frequencies. This would result in an artifact of higher vestibular thresholds. In contrast, this artifact would be unlikely to affect visual thresholds because of the secure attachment between the visual scene and platform. Thus, if this artifact were occurring and were corrected, it would actually increase

![Diagram](http://jn.physiology.org/ Downloaded from)

Fig. 2. Examples of ideal motion stimuli for 2 Hz and 5 Hz. Although both have the same peak velocity of 2°/s, they differ in peak velocity and acceleration (A). Actual device position (gray x) from a 60-Hz feedback signal closely matches the theoretical motion. f, frequency; T, period.

Fig. 3. A: actual device position with (•) and without (x) a human subject closely matches the theoretical motion (line) for various displacement magnitudes. Device position from a 60-Hz Moog position feedback signal is shown. Displacements of 0.04°, 0.05° and 0.10° are shown because they span the thresholds at 5 Hz for the three testing conditions, as well as the displacements for trials at fixed ratios (1.30, 1.55, 1.80) of an initial estimate of threshold. B: an accelerometer used to measure tilt verified that the Moog feedback signal accurately reported device displacement at 5 Hz for motions near threshold and larger.
the difference between vestibular and visual thresholds at 5 Hz, and thus our primary hypothesis would be more strongly supported. Similar reasoning applies to other artifacts that would result in inflated vestibular thresholds. For example, there is evidence that external “noise” may raise thresholds for nonvestibular perception (e.g., Watan- 

**Experimental procedure.** Subjects were seated upright in a racing-style chair and secured with a five-point harness. The subject’s head was fixed relative to the chair and platform with a foam-lined helmet that could be tightened until snug. We attempted to minimize the influence of cues other than visual and vestibular. Wind cues were reduced by covering all skin surfaces (gloves, long sleeves, clear plastic visor over face). Tactile cues were reduced to the extent possible using foam padding. Auditory cues were masked with noise-canceling headphones playing white noise (~60 dB). Vestibular-only trials were performed in the dark in a light-tight room. Relative success of these techniques to reduce task-relevant cues is demonstrated by elevated thresholds measured in patients suffering total vestibular loss (Valko et al. 2012). After trials were required to push a button in their left hand if they perceived leftward motion or a button in their right hand if they perceived rightward motion. Subjects were instructed to make their best guess if they were unsure. Each trial started from upright, and a subthreshold motion returned the platform to upright after each trial. A computerized voice indicated to subjects that motion was commencing (“start”), and that motion was complete and that they needed to respond (“respond”). Before each condition, practice trials spanning a range of stimuli magnitudes were administered to assure that subjects understood the task. For the visual and visual-vestibular conditions, a computer-controlled light turned on 3 s before motion and turned off after the subject responded, and thus the subjects had no visual cues while they moved back to the origin.

For study A, seven frequencies spanning 2 decades, 0.05, 0.1, 0.2, 0.5, 1, 2 and 5 Hz, were tested. Motion magnitudes were selected by an adaptive procedure, a three-down, one-up staircase paradigm (Leek 2001), meaning that stimuli magnitude reduced after three consecutive correct responses, and increased after one incorrect response. Testing started at stimuli magnitudes well above typical thresholds (0.76°/s at 0.05 Hz, 1.53°/s at 0.1 Hz, 3.18°/s at 0.2 Hz, 2.55°/s at 0.5, 1 and 2 Hz, 2.04°/s at 5 Hz). Testing ended after the adaptive track reached the third minimum direction reversal, which occurs when the subject responds incorrectly following at least three correct responses. Each frequency was tested on two separate occasions to reduce measurement error, and thresholds were determined using the pooled data across both blocks. With three conditions this resulted in a total of 42 blocks per subject. On average there were four blocks per test-session. Study A included a total of 9,545 trials for all subjects, with each threshold (i.e., for each subject and frequency, with two blocks combined) based on between 46 and 120 trials, and a mean of 76 trials.

**Study B** focused on a more narrow range of 1, 2, 3, 4 and 5 Hz. For each frequency, all three conditions were tested on the same day to reduce possible day-to-day variation in thresholds, and subjects were required to take breaks between conditions. For study B, we desired to have lower measurement variability than in study A to be able to test the optimal integration hypothesis. To more precisely measure thresholds, we used a two-stage approach to adaptively select stimuli magnitude. In the first stage, a three-down, one-up staircase, starting at 4°/s, was used. After each trial, a real-time psychometric fit was performed, and the coefficient of variation (CV) of the threshold was computed as a measure of goodness of fit. The first stage ended when CV was less than 0.25. In the second stage, the method of constant stimuli was utilized: testing occurred at three magnitudes that were provided at fixed ratios (1:30, 1:55, 1:80) of the threshold calculated at the end of the first phase, with motion in both directions. These ratios were selected because they tailored the stimuli for each subject to yield a large amount of information for each trial (Wetherill 1963). Testing continued until CV was less than 0.10. Each block had between 183 and 578 trials with a mean of 240 trials per block. Study B included a total of 18,027 trials for all subjects.

Two small supporting studies are described in the discussion. The first measured visual thresholds using the methods described for study A, but tested a single frequency only (1 Hz), and varied the size of the round aperture to test the effects of changing FOV. The second measured visual-vestibular thresholds using the methods described for study A, but compared thresholds with and without the round aperture. In each frequency was tested in a block of contiguous trials. The order of frequencies and conditions (vestibular, visual, visual-vestibular) was randomized across subjects. In each block, an adaptive one-interval two-alternative categorical forced choice procedure (Leek 2001; Treutwein 1995) was used.

**Threshold determination.** Thresholds were determined using a psychometric curve fit. A Gaussian cumulative distribution psychometric function defined by two parameters, sigma and bias, was fit to the data of each block, using a maximum likelihood fit (MATLAB Statistics Toolbox version 7.0) determined using a generalized linear model and probit link function (Habarucakova et al. 2012; Zupan and Merfeld 2008). The Gaussian distribution was chosen (Wichmann and Hill 2001), as in every similar study of which we are aware (e.g., Butler et al. 2010; MacNeilage et al. 2010; Roditi and Crane 2012b; Soyka et al. 2011), because it correctly fits the second-order statistics (i.e., variance of the underlying distribution). Although a logistic distribution may also provide an appropriate fit, the differences between the logistic and Gaussian distributions are small (Dobson and Barnett 2008) and inconsequential, and the most appropriate distribution cannot be determined without extensive data collection. The underlying Gaussian probability distribution is described by “bias,” which represents an offset from zero, and “sigma” which represents the noise standard deviation (Dobson and Barnett 2008; Hubarucakova et al. 2012; Merfeld 2011). We report all thresholds as “1-sigma thresholds” i.e., , which can be converted to the commonly reported 79.4% threshold targeted by a three-down/one-up staircase, using .

**Theoretical framework.** Signal detection theory relates the variability of an underlying noisy sensory estimate to the threshold for perceptual responses (Green and Swets 1966; Merfeld 2011). Figure 4 shows the assumed relationship between sensory stimuli and noisy sensory estimates as defined by a standard deviation , which, in the absence of any vestibular bias, equals the stimulus level where 84% of motion directions will be correctly recognized (Green and Swets 1966). In this hypothetical example, vestibular has the largest variability, with , visual has a slightly lower variability with and visual-vestibular has lower variability than either cue individually with a standard deviation , which is 0.65. Since precision is defined as the inverse of variance, i.e., , and since we use 1-sigma thresholds, threshold , which can be improved in precision is indicated by a smaller threshold. The term reliability, used by some related studies, is analogous to precision (Faisal et al. 2008).

Bayesian optimal integration results in a sensory estimate that has the highest possible precision for the linear weighting of static individual cues (e.g., Ernst and Banks 2002; Gelb 1992; Gu et al. 2008; Landy et al. 1995). The predicted optimal (e.g., maximum likelihood) integration occurs when the weights are inversely proportional to the precision of each cue, if one assumes that the sensory cues have Gaussian distributions, independent noise sources, and uniform priors. For two cues, the predicted optimal visual-vestibular threshold occurs when the vestibular weight , visual weight , and can be determined using the
because the stimulus amplitude equals the noise standard deviation. In these examples no bias is assumed.

Lines show the threshold, i.e., the stimulus amplitude at which subjects will report rightward motion 84.1% of the time. The psychometric curve is fit to these data, and the predicted threshold is smaller for the vestibular condition (Fig. 5A–C) than visual condition (Fig. 5D–F). In this illustrative example, the threshold for the visual-vestibular condition (Fig. 5C) is less than for either visual or vestibular, suggesting that, statistically, the experiment was well-powered. We used a single statistical test for each hypothesis. Therefore, multiple-comparison correction is not applicable.

To determine whether static optimal integration occurred, four hypotheses were tested using planned comparisons to explain how vestibular and visual thresholds combine to give the visual-vestibular thresholds: 1) they depend only on vestibular thresholds, 2) they depend only on visual thresholds, 3) they are the minimum of vestibular or visual thresholds at each frequency, 4) they depend on the linear optimal maximum likelihood estimate for integration of vestibular and visual cues. For each of the hypotheses, a visual-vestibular predicted threshold was determined and compared with the visual-vestibular actual threshold using ANOVA. The optimal integration hypothesis would be supported if the prediction differed significantly from the data for all but the fourth hypothesis.

**RESULTS**

Figure 5, A–C, shows results for one subject at 5 Hz in the vestibular, visual, and visual-vestibular conditions, respectively. The experimental data show the proportion of trials at each stimulus amplitude in which the subject responded that motion was rightward. In all conditions, when there are large, positive (rightward) motions, the subject selected rightward in close to 100% of trials. When there are large negative (leftward) motions, the subject selected rightward in close to 0% of trials. For tiny motions, the subject selected rightward in close to 50% of trials, indicating that they were guessing. The data are fit by a Gaussian cumulative distribution function to give the psychometric curve fit (black curve). The thresholds $\sigma$ at the $P = 84.1\%$ level are shown with black lines. For this subject and frequency, the threshold is smaller for the vestibular condition (Fig. 5A) than visual condition (Fig. 5B). In this illustrative example, the threshold for the visual-vestibular condition (Fig. 5C) is less than for either visual or vestibular, suggesting that,
when the brain has both cues, it recognizes motion direction more precisely.

Frequency responses of vestibular, visual and visual-vestibular thresholds. In study A, we measured vestibular, visual and visual-vestibular thresholds over 2 decades of temporal motion frequencies from 0.05 to 5 Hz (Fig. 6A). We plot both displacement (Fig. 6A) and velocity (Fig. 6C) at threshold to show that there are threshold variations in both domains. Each data point is the geometric mean of five subjects.

Vestibular thresholds (triangle) are lower than visual thresholds (square). As frequency increases, the displacement threshold decreases, reaching 0.047° at 5 Hz, i.e., a 0.2-s motion. The visual thresholds (squares) also decrease with frequency, decreasing from 1.40° at 0.05 Hz to 0.088° at 5 Hz.

Visual-vestibular thresholds (circles) follow the same overall pattern as the individual cues, gradually decreasing from 1.27° at 0.05 Hz to 0.040° at 5 Hz. Visual thresholds (squares) decrease more gradually from 1.398° at 1 Hz to 0.16° at 5 Hz. Visual-vestibular thresholds (circles) decreased more gradually from 2.02° at 1 Hz to 0.072° at 5 Hz. Visual-vestibular thresholds (circles) decrease from 1.27° at 0.05 Hz to 0.040° at 5 Hz.

In study B we made more precise measurements (Fig. 6, B and D) at five closely-spaced frequencies (1, 2, 3, 4, 5 Hz) after finding that visual and vestibular thresholds are similar around 2 Hz in study A. Each data point is the geometric mean of five subjects. Vestibular thresholds (triangles) decreased from 0.28° at 1 Hz to 0.043° at 5 Hz. Visual thresholds (squares) decreased more gradually from 0.20° at 1 Hz to 0.072° at 5 Hz. Visual-vestibular thresholds (circles) decreased significantly from 0.16° at 1 Hz to 0.034° at 5 Hz.

Differences in the frequency response of visual and vestibular perceptual thresholds. We examined the dynamics of visual (circles) and visual-vestibular (squares) thresholds in both studies A and B (Fig. 6, A and B), and found that:

1) Vestibular and visual thresholds are very close at 0.05 Hz.

2) There is a cross-over at 2 Hz where vestibular and visual thresholds are very close.

3) Vestibular thresholds were lower than visual thresholds above the 2 Hz cross-over (ANOVA, \( P = 0.0013 \)).

4) Visual thresholds were lower than vestibular thresholds for frequencies ranging from 0.1 Hz to 1 Hz (ANOVA, \( P < 0.001 \)).

Optimal static vestibular-visual integration in roll tilt using two natural cues. The optimal predictions for the visual-vestibular thresholds (Fig. 6, A and B, dotted lines) were computed based on the individual vestibular and visual thresholds at each frequency. The visual-vestibular thresholds (thin solid lines) closely followed the optimal prediction for visual-vestibular thresholds, suggesting that optimal or near-optimal visual-vestibular integration was occurring. There is not a significant difference between the experimental results and optimal predictions (ANOVA, \( P = 0.15 \)). To further support the optimal hypothesis, we ruled out three other potential hypotheses. Specifically, we found that the visual-vestibular thresholds were significantly lower than visual thresholds alone (ANOVA, \( P = 0.005 \)), vestibular thresholds alone (ANOVA, \( P < 0.001 \)), and the best of these two individual thresholds (ANOVA, \( P < 0.001 \)).

DISCUSSION

Dynamics of vestibular and visual perceptual thresholds. To our knowledge, this is the first comparison of vestibular and visual motion perception thresholds across a broad range of frequencies. It shows for the first time that vestibular perceptual thresholds can be lower than visual thresholds for a range of frequencies that we regularly experience and that are behaviorally important. Specifically, we report that: 1) visual and vestibular thresholds were indistinguishable at 0.05 Hz and 2 Hz; 2) visual thresholds were lower than vestibular thresholds...
for frequencies ranging from 0.1 Hz to 1 Hz; and 3) visual thresholds were higher than vestibular thresholds above 2 Hz. For comparison, previous studies using linear motion found lower visual than vestibular thresholds, except when visual cues were intentionally degraded (Butler et al. 2010; Fetsch et al. 2009, 2012; Gu et al. 2008; MacNeilage et al. 2010). These earlier findings may, in part, be due to the frequency content (~0.5–1 Hz) of the stimuli; we also found lower visual than vestibular thresholds in this frequency range.

Our findings appear consistent with other relevant studies. Our vestibular thresholds as a function of frequency in the dark are very similar to those recently reported for roll tilt by Valko et al. (2012), and also quite similar to those recently reported by other groups at 2 Hz (Crane 2012b) and for low-velocity tilts using a detection task (Janssen et al. 2011). Another study (Lagace-Nadon et al. 2009) reported that visual roll tilt thresholds decreased with increasing temporal frequency, consistent with our findings. Their thresholds were approximately 8× higher than ours, which can probably be attributed to visual display differences, as we chose to use a broader realistic scene as opposed to their narrower low-spatial-frequency scene. Beyond this, it is difficult to compare vestibular and visual precision using the results of these two studies because they used different subjects and different methods.

Our results complement previous studies reporting frequency-dependent differences between visual and vestibular reflexive responses. For example, gaze stability studies (Crawford 1964; Paige 1983; Robinson 1977) suggest that ocular stability is more vestibular based at high frequencies and visual dependent at low frequencies, since these studies report that the VOR gain acts like a high-pass filter, and optokinetic gain as a low-pass filter. Similar findings have been reported for postural stability (Bles et al. 1983; Peterka and Benolken 1995) and when controlling a vehicle (Zacharias and Young 1981). In both reflexes and perception, dynamics originate from the sensory periphery and central processing (as well as motor responses for reflexes), but these dynamics are not necessarily the same. For example, there are qualitative differences in sensory processing between the VOR and vestibular perception (Merfeld et al. 2005a, 2005b), and differences in the frequency response between VOR thresholds and vestibular perceptual thresholds (Haburcakova et al. 2012). In addition, perception usually involves a host of processes, including cognitive, that could be the source of our measured perceptual dynamics and differences between vestibular and visual perceptual precision.

Given the complex nature of visual processing, an important question is how broadly applicable are our findings. For several reasons, we think that our findings are broadly applicable. First, we made additional measurements (Fig. 7) of visual thresholds with different FOVs and found that visual thresholds were relatively invariant over a large range. Specifically, we found a small, but statistically insignificant, increase in visual threshold below 30° FOV. More importantly, there were no consistent threshold variations measured for FOVs between 34° and 63°, which includes the FOVs for the primary studies. Second, since our scene was realistic, the spatial frequencies in our visual scene were, by definition, representative of a typical indoor scene. Third, the physical images used for our visual scene avoided discrete pixel effects associated with digital displays. Finally, we eliminated relative motion cues and/or relative displacement cues, those in which space-fixed and body-fixed landmarks could be readily compared, by using a circular aperture. To demonstrate the importance of eliminating relative visual cues, we measured thresholds without the aperture present (Fig. 8) and found that the thresholds measured without the aperture were 30% lower than with the

---

**Fig. 7.** Field of view (FOV) above 30° shows little effect on visual thresholds. Although visual thresholds seem to be somewhat higher when FOV was much smaller (20°) than that used in our studies (63° in study A, 56° in study B, as indicated by arrows), none of the differences observed, nor the overall trend, are statistically significant. FOV was modified by changing the diameter of the circular aperture through which subjects viewed the visual scene, as demonstrated in the illustrations (a larger picture of the visual scene appears in Fig. 1). The visual scene was slightly closer in study B than study A, as shown by the larger forest poster size in the illustration, and thus 56° was not sampled for this FOV study. Subjects always viewed the scene binocularly. The stated FOV is for all areas that subjects could see, even if by only one eye, which became challenging at smaller diameters. At 20° FOV, the FOV with binocular coverage was between 3° and 6°, depending on the interpupillary distance. Each point is the geometric average across four subjects, with the bars indicating standard error of the mean. The forest poster “Aspens, Ashley NF” is replicated with permission of photographer Alain Thomas.
thresholds by changing the motion frequency. This contrasts
with prior approaches in which one cue is artificially degraded
to alter its precision, as in some self-motion perception studies
(Butler et al. 2010; Fetsch et al. 2009, 2012; Gu et al. 2008; Jurgens and Becker 2006). We highlight
important similarities and differences below.

First, few studies of optimal integration have been per-
formed using only natural cues. Since at least some studies
cautions that the way visual cues are presented can disrupt
optimal integration (Butler et al. 2011; de Winkel et al. 2010),
it was important to utilize a natural visual scene and manipulate
thresholds by changing the motion frequency. This contrasts
with prior approaches in which one cue is artificially degraded
to alter its precision, as in some self-motion perception studies
(Butler et al. 2010; Fetsch et al. 2009, 2012; Gu et al. 2008) and
other multisensory integration studies (e.g., Ernst and Banks
2002; Landy et al. 1995). In these earlier self-motion studies,
the visual cue was artificially manipulated using a 3D comput-
er-generated scene with varying dot movement coherence. In
contrast, our approach confirmed optimal static integration,
mimicking the earlier findings, but did so without degrading
either the visual or vestibular stimuli.

Furthermore, we did not provide conflicting vestibular and
visual cues. Cue conflict is common in many self-motion percep-
tion studies (Butler et al. 2010; Fetsch et al. 2009, 2012;
Jurgens and Becker 2006) and other multisensory integration
studies (e.g., Ernst and Banks 2002; Landy et al. 1995). Cue
conflict is not necessary to demonstrate optimal integration
(e.g., Gu et al. 2008), although it allows the brain’s weights for
the two cues to be determined.

This study complements a limited number of studies that
have investigated visual-vestibular integration in roll tilt.
Bayesian integration for static roll orientation has been dem-
onstrated (Clemens et al. 2011; De Vrijer et al. 2009), and
perceptual visual-vestibular interaction has been studied in roll
tilt (e.g., Dichgans et al. 1972; Zacharias and Young 1981), but
this is the first experimental study to examine optimal integra-
tion using thresholds determined with dynamic roll tilt stimuli.

Our results are consistent with static linear optimal integra-
tion, but other hypotheses could also explain the data. For
example, cue integration could occur in a dynamic nonlinear
manner. This possibility also exists for related studies (e.g.,
Butler et al. 2010; Ernst and Banks 2002; Fetsch et al. 2009,
2012; Gu et al. 2008).

Recent work has suggested that individual subjects may use
different strategies for integration of self-motion yaw-rotation
cues, and that some subjects behave in a manner inconsistent
with optimal integration (de Winkel et al. 2013). Most of our
individual subject responses seem consistent with optimal integra-
tion (Fig. 9), although our measurement variability for
each subject was larger than that in the de Winkel et al. study,
precluding the use of statistical tests to test individual subject
responses.

Contribution of other sensory modalities. When tilted, the
component of body weight supported by the side of the chair
results in a force that provides tactile cues (Valko et al. 2012).
In addition (or alternatively), somatic graviception (e.g., Mit-
telstaedt 1996) could contribute. Although thresholds mea-
sured during motion may include nonvestibular sensory con-
tributions, this would not fundamentally alter any of our
conclusions regarding sensory integration, because, even if
measured “vestibular” thresholds included significant tactile/
somatosensory contributions, visual-vestibular thresholds mea-
sured herein would include all of the same vestibular and
nonvestibular contributions.

Furthermore, a recent report (Valko et al. 2012) suggests that
vestibular cues predominate in the dark. Patients with total
dorsal vestibular loss were reported to have higher thresholds
in the dark than normal subjects for several different motion
directions, including roll tilt. For roll tilt in the dark, the
average thresholds for total vestibular loss patients were sig-
ificantly greater than normal for the frequency range tested
herein, which suggests that, for this task in the dark, vestibular
cues predominate over all other cues combined. Finally,
another recent report suggests that motion direction recognition
thresholds are relatively unaffected by device vibration, pro-
viding further support for the importance of vestibular cues in
motion direction recognition (Chaudhuri et al. 2013).

Reasons for using dynamic roll tilts. We chose to investigate
dynamic roll tilts because this allowed us to study how the
brain combines both static (displacement) and dynamic (mo-
tion) cues. Such dynamic stimuli are particularly relevant and
interesting because both the visual and vestibular systems
provide dynamic time-varying motion and displacement sig-
als. Specifically, motion of the visual scene provides both
velocity and displacement cues (e.g., Nakayama and Tyler
1981). Similarly, the SCC provide a velocity signal, while the
otolith organs provide a tilt (or displacement) signal. Roll tilt

Fig. 8. Thresholds are lower when relative visual cues are provided in addition
to absolute visual cues. We define relative cues as those in which the motion
of space-fixed and body-fixed landmarks can be compared, while absolute cues
only provide motion of the visual scene relative to the retina. Open circles are
the same condition presented for visual-vestibular (study A) thresholds in
which the circular aperture is used to prevent subjects from using relative
visual cues. Solid circles are visual-vestibular thresholds for which the aperture
was not used, allowing subjects to look for relative cues; for example, one
subject reported using their knee as a landmark to judge motion of the visual
scene. Averaged across frequencies, thresholds are 30% lower when relative
cues are provided. Although the effect of relative visual cues could have been
more directly assessed using visual thresholds rather than visual-vestibular
thresholds, this was not possible because of the limited space to mount the
visual scene on the motion device. While a possible confound is that FOV also
increased without the aperture, this is unlikely because visual thresholds are
relatively insensitive to FOV, especially above 30° (Fig. 7). Thresholds are
gemetric means of five subjects, with the bars indicating standard error of the
mean.

**Optimal multisensory integration in roll tilt with natural
cues.** Our study is consistent with other studies demonstrating
static optimal integration of visual and vestibular cues (Butler
et al. 2010, 2011; De Vrijer et al. 2009; Fetsch et al. 2009,
2012; Gu et al. 2008; Jurgens and Becker 2006). We highlight
important similarities and differences below.

First, few studies of optimal integration have been per-
formed using only natural cues. Since at least some studies
cautions that the way visual cues are presented can disrupt
optimal integration (Butler et al. 2011; de Winkel et al. 2010),
it was important to utilize a natural visual scene and manipulate
thresholds by changing the motion frequency. This contrasts
with prior approaches in which one cue is artificially degraded
to alter its precision, as in some self-motion perception studies
(Butler et al. 2010; Fetsch et al. 2009, 2012; Gu et al. 2008) and
other multisensory integration studies (e.g., Ernst and Banks
2002; Landy et al. 1995). In these earlier self-motion studies,
the visual cue was artificially manipulated using a 3D comput-
er-generated scene with varying dot movement coherence. In
contrast, our approach confirmed optimal static integration,
mimicking the earlier findings, but did so without degrading
either the visual or vestibular stimuli.

Furthermore, we did not provide conflicting vestibular and
visual cues. Cue conflict is common in many self-motion percep-
tion studies (Butler et al. 2010; Fetsch et al. 2009, 2012;
Jurgens and Becker 2006) and other multisensory integration
studies (e.g., Ernst and Banks 2002; Landy et al. 1995). Cue
thresholds depend on both SCC and otolith cues (Lewis et al. 2011), which was determined by comparing roll tilt thresholds with roll rotation thresholds collected with the subject supine and reliant on SCC cues. Furthermore, preliminary evidence suggests that the brain integrates SCC and otolith cues to determine self-motion more precisely than with either cue individually, as demonstrated by control data collected from one of our subjects in Fig. 10. [This investigation of SCC-otolith integration is described in more detail in a conference abstract (Lim et al. 2009)].

Another advantage of roll-tilt visual motion is that, by definition, it occurs in a plane perpendicular to the line of sight, minimizing potential three-dimensional perspective changes and parallax cues. Furthermore, our design minimizes potential artifacts from the VOR, even if uncompensated centrally (Wade and Curthoys 1997), because torsional VOR gains are $-0.15$ (Bockisch and Haslwanter 2001), much lower than horizontal and vertical VOR gains (Bockisch et al. 2005).

While any such artifact would be small, it is also important to note that any such torsional VOR would reduce visual motion on the fovea during the visual-vestibular condition; thus any resulting artifact would make the combined thresholds larger and more suboptimal, rather than leading to a false finding of optimality. It is also unlikely that the oculogyral effect contributes to our results, since it has only been demonstrated for a head-fixed visual target and yaw rotation (e.g., Clark and Stewart 1969), while we used a space-fixed visual scene and roll tilt. While we are unaware of any published studies demonstrating this effect for roll tilt, an abstract has reported that the oculogyral effect did not occur for roll tilt (Kolev and Merfeld 2010).

**Rationale for experimental design.** Despite some overlap and differences between studies A and B, each was important to support the primary findings. First, only a single frequency in study A is above the cross-over frequency of 2 Hz, where the data show that vestibular perception can be more precise than visual perception; this necessitated additional data collection above 2 Hz in study B. Second, since optimal integration yields the largest improvement when the individual cues have similar precision (e.g., Ernst and Banks 2002), we chose to do additional testing at frequencies near the cross-over frequency of 2 Hz. While other studies have yielded similar cue precision by degrading visual cues (e.g., Ernst and Banks 2002; Gu et al. 2008), we chose to search for the frequency range where the cues naturally had similar precision. Specifically, if two cues have equal precision, optimal integration yields a combined precision...
precision 29% lower than the individual cues’ precisions, while, if one cue is three times as precise as the other, optimal integration yields a combined precision that is only 5% lower than the better cue’s precision, making statistical evaluation more challenging. Furthermore, statistical testing was aided by making more meticulous thresholds measurements in study B by asking subjects to perform approximately three times as many trials at each frequency. Since most subjects from study A were unavailable or unwilling to participate in further testing, we designed study B as a stand-alone companion study to focus on the range around 2 Hz. Retesting new volunteers at all frequencies would have been a significant burden on volunteers; since total testing time was ~180 h. Small differences in experimental parameters (e.g., FOV), which were the result of our equipment moving to a different room, had little or no effect. Since we do not directly compare results from studies A and B, any potential subtle differences have no impact on the reported findings.

ACKNOWLEDGMENTS

We appreciate the participation of our anonymous subjects. Data collection was possible only with the assistance of Adil Adatia, Romain Claret, Vartan Mardirianossian, Keyvan Nicoucar, Adrian Priessol and Yulila Valko. We thank John Pack for providing photography expertise; Bob Grimes and Wangsong Mardirossian, Keyvan Nicoucar, Adrian Priesol and Yulia Valko for advice on statistics; Michelle R. Greene, Rick Lewis and Yulia Valko for technical support; Alan Natapoff, Wei Wang and Shomesh Chaudhuri for providing photography expertise; Bob Grimes and Wangsong Mardirossian, Keyvan Nicoucar, Adrian Priesol and Yulia Valko. We thank the National Science Foundation, the National Institute of Deafness and Other Communications Disorders Grant R01-DC04158. The role of stereo vision in 3-D perception ambiguity. Comparison of brainstem, cerebellum, and thalamus. Annu N Y Acad Sci 1164: 19–28, 2009.


Kolev O, Merfeld DM. Thresholds for self-motion perception in roll about earth-vertical and earth-horizontal axes with and without fixation targets (in


J Neurophysiol • doi:10.1152/jn.00332.2013 • www.jn.org