The precision of signals encoding active self-movement

Joshua D. Haynes[1], Maria Gallagher[2], John F. Culling[1], Tom C.A. Freeman[1*]

[1]School of Psychology, Cardiff University, 70 Park Place, Cardiff, CF10 3AT

[2]School of Psychology, University of Kent, Canterbury, Kent, CT2 7NP

[*]Corresponding author: freemant@cardiff.ac.uk

ABSTRACT

Everyday actions like moving the head, walking around and grasping objects are typically self-controlled. This presents a problem when studying the signals encoding such actions because active self-movement is difficult to control experimentally. Available techniques demand repeatable trials, but each action is unique, making it difficult to measure fundamental properties like psychophysical thresholds. We present a novel paradigm that recovers both precision and bias of self-movement signals with minimal constraint on the participant. The paradigm relies on linking image motion to previous self-movement, and two experimental phases to extract the signal encoding the latter. The paradigm takes care of a hidden source of external noise not previously accounted for in techniques that link display motion to self-movement in real time (e.g. virtual reality). We use head rotations as an example of self-movement, and show that the precision of the signals encoding head movement depends on whether they are being used to judge visual motion or auditory motion. We find that perceived motion is slowed during head movement in both cases. The 'non-image' signals encoding active head rotation (motor commands, proprioception and vestibular cues) are therefore biased towards lower speeds and/or displacements. In a second experiment, we trained participants to rotate their heads at different rates and found that the imprecision of the head rotation signal rises proportionally with head speed (Weber's Law). We discuss the findings in terms of the different motion cues used by vision and hearing, and the implications they have for Bayesian models of motion perception.

Key words: motion psychophysics, self-movement, vestibular, motor, Weber's law

NEW AND NOTEWORTHY

We present a psychophysical technique for measuring the precision of signals encoding active self-movements. Using head movements, we show that: (1) Precision is greater when active head rotation is performed using visual comparison stimuli versus auditory; (2) Precision decreases with head speed (Weber's Law); (3) Perceived speed is lower during head rotation. The findings may reflect the steps needed to convert different cues into common units, and challenge standard Bayesian models of motion perception.

INTRODUCTION

Bodily movement is a key part of everyday life. Our eyes, head, limbs and torso are seldom at rest. Action therefore sets the backdrop in which perceptual systems normally operate, with many everyday tasks relying on information about current self-movement. This is derived from a number of perceptual signals, some based on images such as retinal flow, and some based on non-image sources including the vestibular and motor systems. Information from non-image sources also plays a role in interpreting images, allowing the observer to differentiate between self-generated movement and movements of external objects.

Success in these active tasks is constrained by two fundamental types of error, namely the precision and accuracy of the underlying perceptual signals. Precision is driven by internal and external noise and corresponds to the width of the distribution of the underlying perceptual signal as it varies across time. Accuracy, on the other hand, corresponds to the distribution's average and is usually referred to as bias. While a lot is known about the precision and accuracy of image signals, especially in vision and hearing, much less is known about the errors accompanying non-image signals. This is especially the case when the self-movement is 'active' (i.e., self-controlled), partly because it is difficult to apply standard psychophysical techniques to spontaneous actions that are under participant control. We therefore developed a new way to measure precision in these circumstances, using head rotation as an example of self-movement. The technique makes no attempt to differentiate between the various sources of non-image information that are used to encode active self-movement. Rather, it assumes that they are combined to provide a single non-image signal, and it is the precision of this composite signal that we measure (see ref 1 for an example non-image signal comprising vestibular cues to passive translation and extra-retinal cues to eye rotation, and ref 2 for an example non-image signal comprising vestibular and extra-retinal cues used to update spatial localisation after self-movement). Our technique relies on using image signals as a comparison[1], which allowed us to compare the results when using vision or hearing in Experiment 1. We used the same technique in Experiment 2 to investigate whether non-image precision obeys Weber's law, that is whether precision scales with magnitude. Investigating Weber's law also allowed us to compare the precision of image and non-image signals for the stimuli we used.

In a typical precision-measuring task, participants are asked to compare a fixed standard stimulus with a range of test stimuli shown over a series of trials (3-5). The values assigned to the test are usually controlled by a method of constant stimuli or a staircase procedure, both of which provide a measure of a just-noticeable-difference that can be used to estimate the precision of the underlying signal. Crucially, these methods rely on the ability to repeat a set of stimuli over trials. Self-controlled self-movements, however, are not

[1] The term 'image' is less intuitive for hearing because space must be recovered indirectly from binaural and monaural cues. Nevertheless, its use is not without precedence (e.g. Middlebrooks & Green, 1991). Here we use it to define the dynamic spatial cues the auditory system uses to recover source movement with respect to the head.

repeatable: every instance of every action is unique. To measure non-image signal precision, this variability must be accounted for, both within and across trials.

One solution is to use ‘passive’ self-movement because the action is then controlled by the experimenter. The best-known examples come from vestibular research, where participants are moved on a chair or platform (6). Notable examples also come from studies of perceived stability during eye movement, where various contraptions and implements have been used to passively rotate the eye (7, 8). But active non-image signals also include efferent sources, such as copies of motor commands (9), not just the vestibular, somatosensory and, in the case of passive rotation of the eye, the proprioceptive cues that passive stimulation generates (10-13). Passive stimulation therefore excludes important non-image sources of self-movement information. Passive and active stimulation also affect activity in the vestibular nucleus in different ways, where the output of some neurons is known to be suppressed during active self-movement (see ref 14 for a review).

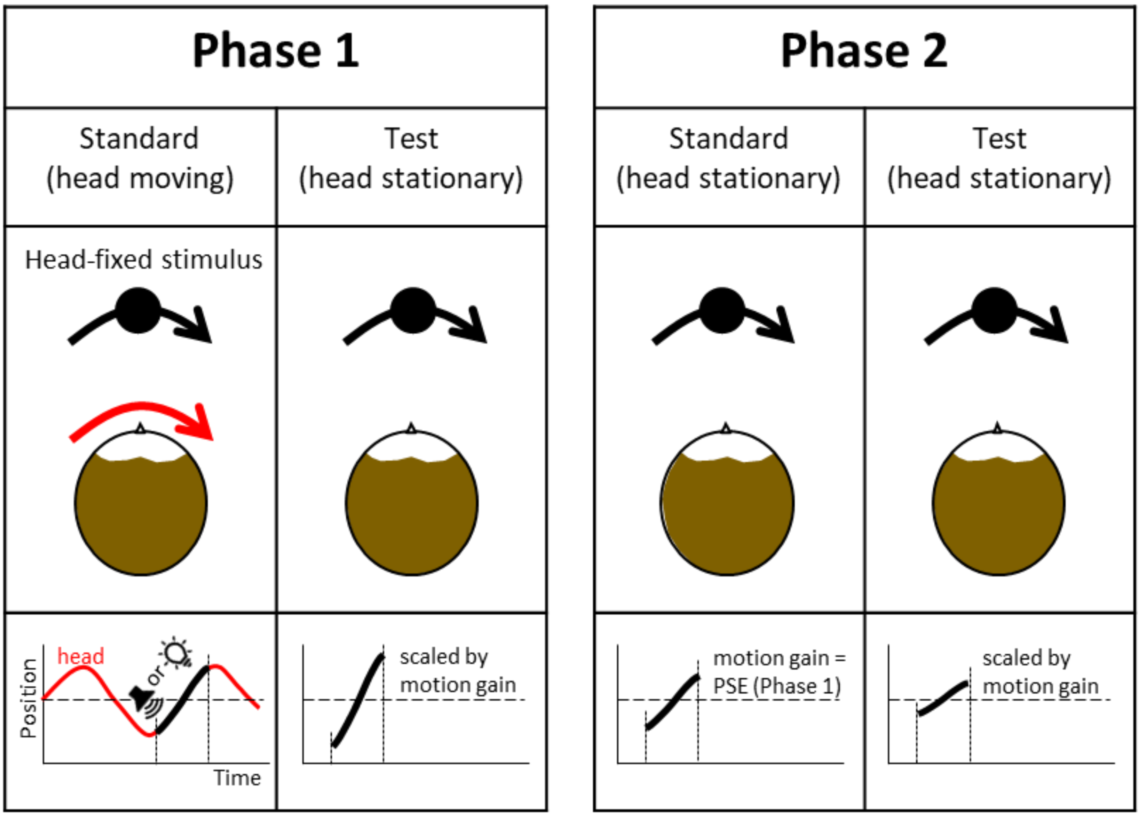

Figure 1 Schematic of the two-phase procedure for measuring the precision of non-image signals encoding active self-movement. We use head rotation as an example. Phase 1 consists of two intervals: a standard interval, in which the head moves, and a head-fixed stimulus (visual or auditory) appears Ψ 2 ý ű K Ŝ ý 2wăeĉ,2 and a test interval, in which the same movement of the stimulus, scaled by the motion gain, is replayed but with head stationary. Phase 2 also consists of two intervals, both with the head stationary. Here, the motion gain used to scale stimulus motion in the standard interval is set to the Point of Subjective Equality found in Phase 1. For both phases, a Method of Constant stimuli is used to manipulate motion gain across trials and construct psychometric functions. These are then used to determine the precision of the image signal in the head-stationary intervals, and the non-image signal in the head-moving interval, based on a model described in the Appendix.

Instead, we focus on active self-movement, specifically head rotation, where non-image sources consist of vestibular cues, motor commands, and proprioceptive feedback, plus any number of somatosensory cues, such as the gliding of hair across the back of the neck. Image-based cues to self-movement were excluded by carrying out the experiments in a completely dark and quiet room, removing important reafferent cues such as retinal flow (15-18). Our method for measuring non-image signal precision uses two experimental phases, combined with a novel analysis that accounts for the variability of self-movement across trials. The paradigm is sketched in Figure 1. Based on two-interval forced-choice, the participant makes self-controlled left and right head rotations in the first interval of each trial of Phase 1, and an auditory or visual stimulus appears in the 3rd sweep (see bottom left panel). This stimulus is head-centred, because it moves with the participant. The participant is instructed to judge the movement of the stimulus (and hence the head movement during the stimulus). We refer to this interval as the 'standard'. In the second 'test' interval of Phase 1, shown in the 2nd column of Figure 1, an auditory or visual stimulus is again shown, but this time with the participant's head stationary. The stimuli move with a trajectory defined by the head movement recorded in the standard interval, but scaled up or down by a multiplicative factor we call 'motion gain'. Hence, the pattern and duration of stimulus movement experienced in the two intervals is the same, apart from overall magnitude, and is provided by different motion cues. In the head stationary interval, the motion cues depend on image signals. In the head-moving interval, they depend on non-image signals, including any extra-retinal contributions related to smooth compensatory eye movements like the vestibulo-ocular reflex (19), or an inhibitory pursuit drive made to keep the eye head-centred (20). We note that the auditory and visual stimuli used to mark the 3rd head sweep do not provide any informative motion cue per se because they are head-fixed and presented in a dark and quiet lab. This remains the case even if their perceived positions shift due to audiogyral (21) and oculogyral illusions (22), which in any case is doubtful given that illusory shifts in position occur over much larger time scales than used in our experiments (23).

Following the two intervals, participants indicate which interval appears to 'move more'. We avoided the terminology 'faster' or 'further' because cue preference depends on modality: for vision, participants prefer speed rather than displacement or duration (24, 25), whereas for hearing the reverse is true (26, 27). Motion gain is manipulated across trials using a Method of Constant Stimuli, resulting in a psychometric function that includes two sources of internal noise, one based on the image signal (e.g., visual or auditory motion) and one based on the non-image signal. To tease these two sources of noise apart, Phase 2, shown in the 3rd and 4th columns of Figure 1, isolates the internal noise of the image signal, using the same set of head movement recordings from Phase 1 to move the stimuli in the same trial-by-trial order, but with the head always stationary. Again, the second interval is a scaled version of the first. The precision of the non-image head rotation signal can then be recovered from Phase 1 as described in the Methods, because image signal precision is now known. We assume that the noise associated with the head-centred auditory or visual target is negligible in the first interval, given that the image is not moving. In comparison,

the active self-movement will typically be fast-paced, yielding high levels of internal noise in the non-image signal, as discussed in more detail below.

In Experiment 1, we used this novel technique to measure non-image precision accompanying head rotation, using either auditory or visual stimuli. The modality used to deliver image motion should not affect the measured precision of the non-image signal because the same image signal (with the same underlying noise) is present in all intervals apart from the first standard interval of Phase 1. In Experiment 2, we used auditory stimuli to investigate whether the precision of the non-image signal obeys Weber's law (i.e., scales with magnitude). The new technique requires that eye movements are similar (but not necessarily absent) in the three head-stationary intervals. If this were not the case, then Phase 2 could not be used to estimate the precision of the image signal in the head-stationary test interval of Phase 1. This was tested in Experiment 2.

Phase 1 also provides information about bias, specifically whether the magnitude of perceived motion is the same when image signals and non-image signals are compared. This is interesting in its own right, partly because it is well known that objects pursued by an eye movement appear slower (28, 29). In this case, the non-image 'extra-retinal' signal evidently provides a lower estimate of speed than the image signal. Analogous perceptual slowing has been demonstrated for stimulation of the vestibular system using passive head rotation (30) and active touch (31). It has also been implicated for the auditory system, based on the finding that compensation for head-movement when judging sound-source stability is incomplete (32). But as far as we are aware, whether active head rotation produces the same slowing is currently not known for either vision or hearing.

METHODS

Participants

All observers gave informed consent, and the experimental procedures were approved by the School of Psychology, Cardiff University Ethics Committee (EC.12.04.03.3123GRA2). In Experiment 1, five participants took part in the experiment (2 female, 3 male). Two participants were naïve to the purposes of the experiment and three were experimenters. Participants wore spectacle correction if required. In Experiment 2, three experimenters and seven participants studying psychology at Cardiff University took part in the experiment (2 male, 8 female). Only the experimenters were aware of the aims of the experiment. Participants completed at least two replications of Experiment 2, with eight participants completing three. Eye movements were recorded for nine of the participants. One of these normally wore spectacles, which were removed to allow the eye tracker to operate.

Experiment 1

Stimuli & Materials

Auditory stimuli were played over a 2.4m diameter ring of 48 Cambridge Audio Minx speakers as shown in Figure 2. The room was sound treated using wall and ceiling tiles with absorption coefficients of 0.9, and carpet on the floors, yielding a reverberation time of approx. 60 ms. Data collection was also carried out in complete darkness. The speakers were controlled by two MoTU 24-channel sound cards, each linked to four six-channel Auna amplifiers. Intensity was normalised across individual speakers. The stimuli consisted of white noise spatially windowed by a Gaussian distribution ($\sigma = 5.25°$ in power, equivalent to 0.7 of the speaker spacing i.e., $\sigma = 7.5^{\circ}$ in amplitude). We have previously shown that this value avoids aliasing artifacts in our speaker system that could occur if the Gaussian distribution is undersampled, while at the same time avoiding the sound becoming too diffuse (33). The noise was sampled at a rate of 48 kHz with a peak level of 70 dB. The position of the spatial Gaussian was refreshed at a rate of 240 Hz, a rate set by the motion tracker described below. The result was a 'blob' of noise that could be moved smoothly across the speakers. The actual motion path taken was determined by the measured head movements, using the motion gain parameter to scale its magnitude.

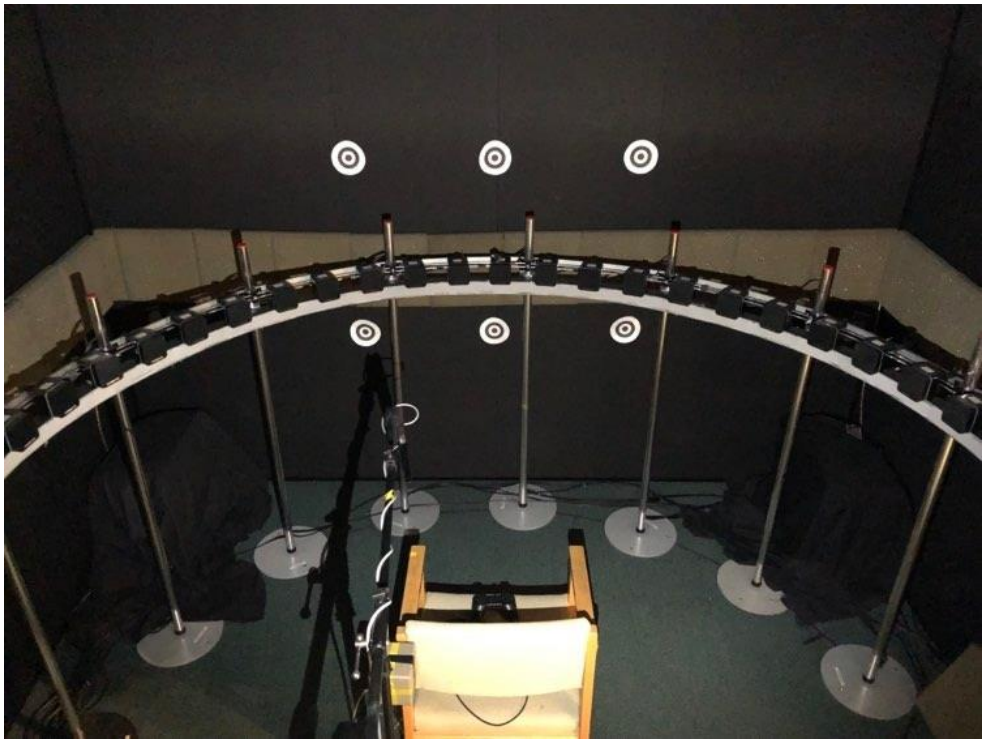

Figure2: Laboratory setup.

Visual stimuli were presented to the participant using an AdaFruit NeoPixel strip of 342 LEDs driven by a single Arduino Uno microcontroller. The LED strip was positioned just below the speakers, as shown in Figure 2, and was driven at a framerate of 40 Hz. The strip subtended 128° either side of straight ahead. This yielded an LED spacing of 0.75°. To ensure that the LEDs presented stimuli at a comfortable brightness, a single layer of 1.2f neutral density filter reduced the intensity of the display. As with the auditory stimuli, smoothly moving

stimuli were created by using a Gaussian distribution that spatially windowed the LED output for each display frame (σ = 1.05°). In order to prevent individual LEDs being visually resolved, the strip was placed in a curved enclosure with one open side that was covered by three layers of diffuser gel at a distance of 35 mm, blurring the image. The overall size of the resulting blob was increased slightly by the diffuser (σ = 1.07˚), which we confirmed using a Minolta LS100 photometer and an array of small apertures. The peak luminance of the blob was ~0.042 cd/m^2.

Head Tracking

Head movement was measured using a Polhemus Liberty tracker that sampled position at a rate of 240 Hz. The tracker was mounted to a head band worn by the participant. For the head-moving standard interval of Phase 1, the head-tracking data were used to detect the 3rd sweep in real-time and keep the subsequent auditory or visual stimulus head-centred (i.e., motion gain = 1). To detect a change in head-movement direction, we convolved the head tracker samples with a finite difference filter to obtain a smoothed derivative. The filter was 13 samples long, meaning there was a 7-frame delay in detecting the head-turn (~30 msec). An example waveform is shown in Figure 3A, with the detected 3rd sweep shown in black and blue.

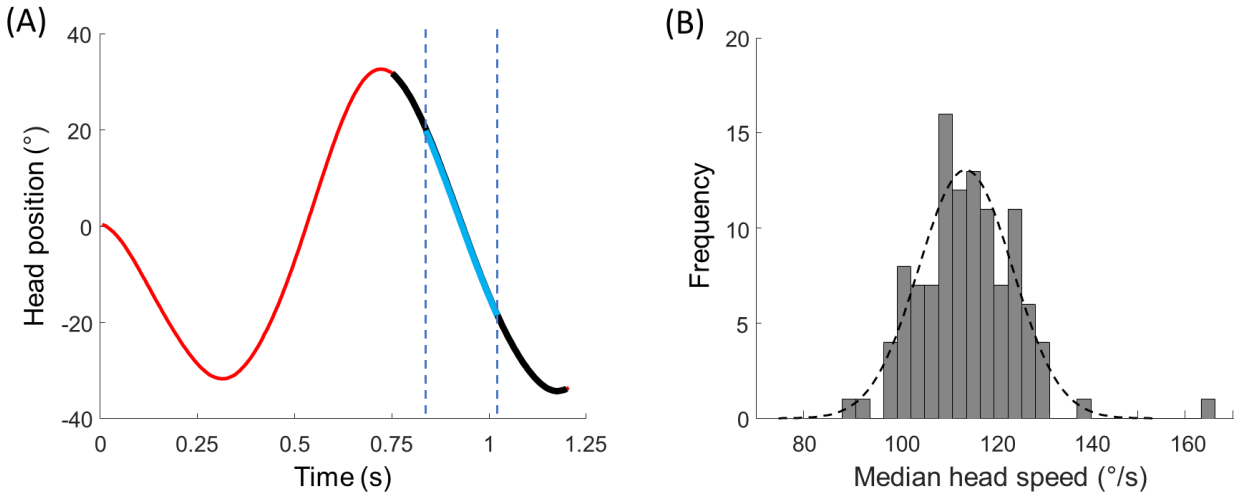

Figure3: (A)Example head movement waveform.The black portion corresponds to the 3rd sweep as detected by the algorithm described in the text. Thevisual or auditory stimulus appeared during this time.Theblue portion defines the region of interest over which median head speed was calculated for analysis. (B) An example distribution of median head speeds for a single repetition of Phase 1 (110 trials) for one participant. The dotted line shows the best fitting Gaussian which was used to determine the mean and standard deviation of the distribution

Procedure

In Phase 1, each trial consisted of a 'head-moving' standard followed by a 'head-stationary' test. The start of the first interval was signalled by a short beep (0.25 s) followed by momentary pause to check the head was centred before the experiment moved on. 'Centred' was defined as 10 consecutive head-tracker samples within ±7.5˚ of the centre of the LED/speaker array. Participants were then instructed to move their heads smoothly left and right, or vice versa, at a pace and amplitude that they felt comfortable with. While it

was their free choice, we found some participants alternated the start direction from trial to trial, while others mostly started in the same direction. The auditory or visual stimulus appeared during the 3rd sweep and moved with the head. Participants were instructed to fixate it. The start of the second head-stationary test interval was signalled using a blue blob that appeared for 0.2 s, with progress again paused to check the head was centred. Note that the motion in the test interval was based on the 3rd sweep recorded in the standard interval only: the dead-time created by the initial 2 sweeps was skipped. Unlike the head-moving interval, participants were instructed not to track the stimulus with their eyes but instead keep looking straight ahead. In Phase 2, the same beep and light were used to identify the start of each interval, with both intervals head-stationary and the initial 2 sweeps skipped. Each replication of Phase 1 and Phase 2 contained the same number of trials, based on the same head movement recordings, shown in the same order.

Psychometric functions were collected based on a Method of Constant Stimuli using 11 motion gain values. For Phase 1, these ranged from 0.2 – 1.2 in steps of 0.1 for the auditory condition, and 0.4 – 1.0 in 0.06 steps for the visual condition. The ranges were based on pilot experiments that showed the visual condition produced steeper psychometric functions than the auditory condition. Each motion gain was repeated 10 times, yielding 110 trials per session. For Phase 2, the same step sizes were used, but the range was centred on the Point of Subjective Equality (PSE) calculated from Phase 1. This ensured that the precision of the image-motion signal we estimated for each replication of Phase 1 and 2 were based on motion gains centred on a comparable value. The PSE was derived by fitting a cumulative Gaussian to the data using the PAL_PFML_Fit function from the Palamedes toolbox (34), with a lapse rate parameter fixed to 0.02 (note the bespoke psychometric function described in the Appendix returns the same PSE as the toolbox).

Participants sat in the centre of the speaker/LED ring and wore the head tracking equipment (in Experiment 2 they also wore an eye tracker). Head position was checked with a laser crosshair mounted above the centre of the ring, which enabled the participant's midline and interaural axis to be aligned with the speaker ring. The head tracker was boresighted with the participant facing forwards and pointing their head towards the central speaker. Boresighting was repeated at the start of each replication of each phase of the experiment.

Each participant repeated three pairs of Phases 1 and 2 for each modality. Psychometric functions were fit to each replication separately. Three out of five participants carried out the auditory condition first.

Head movement analysis

To analyse the head movements after data collection, position samples were first smoothed using MatLab's 'lowpass' function with a passband of 8 Hz. The temporal derivative was then taken and the median velocity calculated over a portion of the 3rd sweep that ranged from 20-60% of the sweep length (shown in blue in the example waveform of Figure 3A). This Region-Of-Interest (ROI) was adopted because it maximised the number of head-movement samples and goodness-of-fit of the psychometric function (see Appendix for evaluation). Figure 3B shows an example for one participant of the distribution of these median velocities for one run of Phase 1. For modelling purposes, the distribution of each

110-trial run was fit with a Gaussian (dotted line in Figure 3B) to extract a mean and standard deviation.

Psychophysical analysis

The distribution shown in Figure 3B emphasises the fact that, as with other self-movements, head rotation varies across trials. Using motion gain therefore seems to be a good way of controlling for this variability because it directly links image motion to the ongoing self-movement in real-time. The patterns of motion are therefore identical, meaning the only difference between signal inputs is speed and displacement – duration is fixed. On the face of it, therefore, motion gain provides the experimenter with a repeatable parameter that can be used to define a psychometric function or drive a staircase. Examples are provided by Serafin et al. (35) and Steinicke et al. (36), who plot psychometric functions defined by changes in motion gain within acoustic and visual virtual reality set-ups, respectively. However, closer inspection of their figures suggests a consistent feature not accounted for by fitting a standard cumulative Gaussian: on occasions, their data appear to asymptote more than a constrained lapse rate parameter would allow (e.g., <6%, as suggested by ref 37). In the Appendix, we construct a model of the psychophysical task that shows why. The model emphasises that motion gain is not always a good shorthand for the actual stimulation experienced by the participant, namely the magnitude of motion (speed or displacement).

In keeping with standard signal detection theory, the model assumes that participants base their judgement on a point estimate of stimulus magnitude (e.g., the peak speed of head and image movement, or average speed, or displacement). Crucially, in addition to internal noise, the point estimates vary across trials due to the external noise introduced by variable self-movement. The external noise produces some surprising effects (see Figure A1). First, the function's true slope is steeper than the best-fitting single cumulative Gaussian. Second, as the variability of the self-movement increases, the function's asymptotes depart markedly from 0 and 100%, much further than a typical constrained lapse rate parameter of 6% would allow.

Following standard practice, we assume that internal and external noise is Gaussian distributed. The precision of a given signal is therefore defined by its standard deviation. If the self-movement did not vary at all, the precision of the non-image signal could be calculated by standard fitting of a cumulative Gaussian to the psychophysical data, and then applying the 'variances sum' law to both phases. Thus, for Phase 1, $\sigma_{G_1}^2 = \sigma_i^2 + \sigma_h^2$, where the subscripts correspond to the cumulative Gaussian fit to the data (G_1), the image signal (auditory or visual, i), and the non-image signal encoding head rotation (h). For Phase 2, $\sigma_{G_2}^2 = 2\sigma_i^2$; hence the precision of the non-image signal (σ_h^2) in Phase 1 can be found by substitution. But, when self-movement varies, this standard approach is an approximation at best. Variable self-movement adds external noise that varies across the psychometric function because it is scaled by motion gain; hence the assumption of a single cumulative Gaussian is not correct. We develop the appropriate formulae in the Appendix and show how these can be used to extract the internal noise of the image signal (σ_i^2) and non-image signal (σ_h^2) from the two phases of our experiment. Similar formulae can be applied to a

more typical motion gain scenario used in virtual reality set-ups, where both self-movement and image motion are shown at the same time (35, 36, 38-40).

Experiment 2

Stimuli & Procedure

The aim of Experiment 2 was to determine signal precision as a function of head and stimulus speed. Before each replication of the main experiment, a training session was therefore run to help participants rotate their heads at one of the five target speeds. The training stimuli were audiovisual, consisting of visual and auditory blobs as used in Experiment 1. These moved in synchrony. The procedure for the main experiment was the same as Experiment 1, consisting of two experimental phases linked by the set of head movements recorded in the first. Unlike Experiment 1, only auditory stimuli were used. Each trained head speed was investigated by completing training and main experiment data collection in pairs. This process was repeated three times, yielding 15 training and main experiment pairs, presented in a random order.

Head Speed Training Sessions

Training sessions consisted of a two-stage process that was run ahead of each replication of the main experiment. In stage 1, participants were asked to track an audiovisual stimulus with their head. The stimulus moved independently along a sinusoidal path at a frequency of 1 Hz at one of five amplitudes: 5°, 12.5°, 20°, 27.5°, 35°. These correspond to median target speeds from 30.0°/s to 209°/s for the ROI defined in Experiment 1. Five and three-quarter periods were shown for each speed to generate 12 sweeps. In stage 2, participants attempted to reproduce the trained head speed, this time using a head-stationary audiovisual fixation target moving with the nose as a guide (motion gain = 1). Again, they completed 12 sweeps, determined by recording the number of head direction reversals detected in the head tracking as described in Experiment 1. The accuracy of head rotation was assessed by calculating the median head speed for each of the final 10 head sweeps as described in Experiment 1. If 7/10 sweeps had a median within 5°/s ±5% of the desired training speed, performance on that training run was deemed sufficiently accurate. If not, the participant was given feedback on how many sweeps were accurate, and how many were too fast and/or slow, and the run repeated. Participants had to complete at least three training runs, with at least one successful run before progressing to each replication of the main experiment.

Eye Tracking and Analysis

Eye movements were tracked using a Pupil Labs Pupil Core head mounted eye tracker. The tracker had a 120 Hz sampling frequency and a front-facing world camera. The camera was used for calibration by having participants look at a 3 by 2 array of calibration points that can be seen in Figure 2. These were used to convert the eye tracker's normalised units into degrees. To analyse, samples with less than 0.6 confidence as defined by the Pupil Labs software were excluded, and the gaps filled using linear interpolation (40). Gaps were more

frequent for the head-moving condition. If the number of dropped samples was 50% or greater, the waveform for that interval was excluded from the analysis.

The remaining waveforms were smoothed using a Gaussian filter (`= 16 Hz in the frequency domain). The 1st, 2nd and 3rd derivatives were then taken numerically, corresponding respectively to velocity, acceleration, and jerk. Saccades were detected using Wyatt's jerk analysis, with a jerk threshold set to $2x10^5$ °/s^3 (41). Saccadic samples were removed from the analysis, along with four samples either side of each detected saccade, plus the initial 20 samples at the start and end of each waveform. Mean velocity and speed were then calculated for the 3rd sweep for each head-movement interval, and the single sweep in each head-stationary interval. For comparison, the VOR needed to maintain stable fixation on a world-stationary point at the same distance as the speakers was calculated, using the approximation described by Leigh and Zee (42, p274): $E = -H(1 + R/D)$, where H is the head velocity, R = 0.1 m (the approximate distance from the eye to centre of head rotation) and D = 1.2 m (the distance from participant to speakers). Note that we assume the eyes were fixating at this distance because a visible fixation point appeared there before the 2nd interval of each trial (for details, see above).

Open science

The code used for fitting the model and analysing the eye movements can be found here (https://osf.io/qcz7w/?view_only=2e2bb846820d4862bcb02a036d3ee815). The link also contains raw psychophysical and head movement data, together with summaries. Eye movement data can be made available on request.

RESULTS

Experiment 1: Precision of signals encoding active head rotation

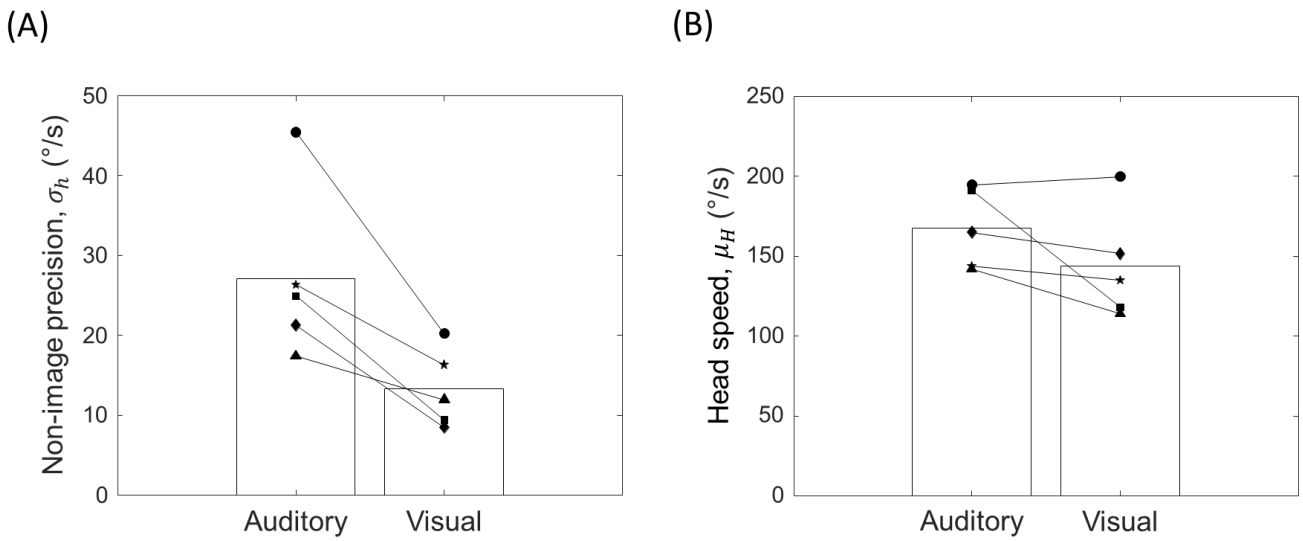

Figure 4. (A) Precision of the non-image signal encoding head rotation for the two stimulus conditions. Precision is defined as the standard deviation of the underlying signal distribution, determined by the model in the Appendix. Note therefore that larger standard deviations correspond to the less precise signals. Bars correspond to the mean of the individual data points, with the latter shown as solid symbols. (B) Head speed, using the same format.

Figure 4A shows the mean precision of the non-image head-rotation signal (bars) across the five participants together with their individual data (solid points). Non-image signals were less precise in the auditory condition, producing a significant increase in the standard deviation of the underlying signal distribution as defined by the model described in the Appendix ($t(4) = 4.19$, $p = .01$). Contrary to our prediction, therefore, modality appears to matter. Figure 4B shows that head speeds were slightly faster on average when an auditory target appeared in the 3rd head sweep compared to a visual one, however this difference was non-significant ($t(4) = 1.76$, $p = .15$). We note that this lack of effect may have been driven by the participant shown with closed circles, who, unlike the other participants, did not show any decrease in mean head speed in the visual condition. It remains possible, therefore, that the difference in non-image signal precision could be explained in part by a scaling of precision with magnitude (i.e. Weber's law). This point is explored further in Experiment 2, where Weber's Law was investigated more thoroughly.

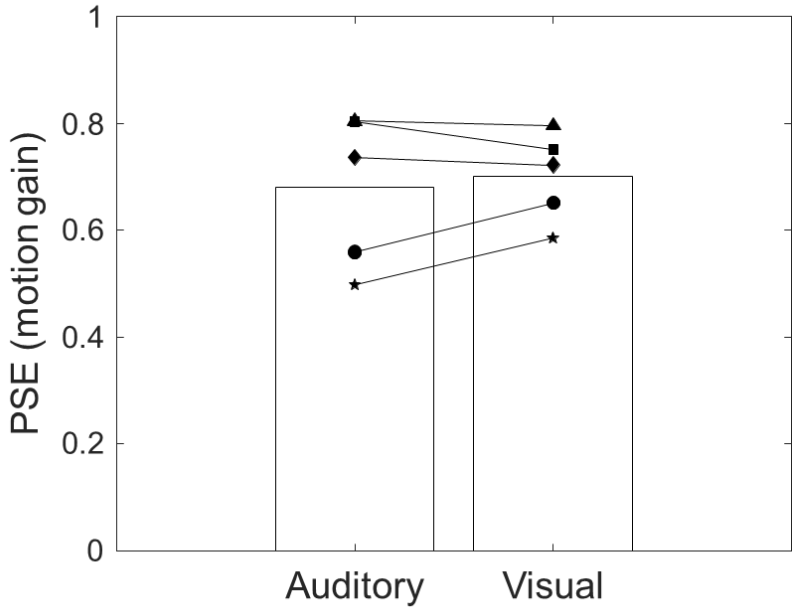

Figure 5: PSEs for the psychometric functions collected in Phase 1. These indicate the relative bias between image and non-image signals for both modalities. A value less than one means head tracked stimuli appear slower. Bars correspond to the mean of the individual data points, with the latter shown as solid symbols

Figure 5 shows the mean PSEs obtained from Phase 1. The PSEs are similar for vision and hearing ($t(4) = -0.71$, $p = .52$). The PSEs are around 0.7, meaning that stimuli had to be slowed by 30% when they moved passed a stationary participant to achieve a perceived speed match with head moving interval. For vision, this finding resembles the Aubert-Fleischl phenomenon (28, 29), where motion appears slower if stimuli are tracked by a smooth eye pursuit. Our data show that perceived slowing occurs for active head rotation too, albeit for auditory and visual stimuli that are linked directly to the self-movement, as opposed to being pursued in a more typical closed-loop manner. It's also worth noting that the biases we found are comparable to those reported for passive head rotation (30).

Experiment 2: Do non-image signals obey Weber's law during active head rotation?

For many sensory systems, discrimination thresholds scale proportionally with stimulus magnitude over a wide range. This is known as Weber's law (43). In Experiment 2, we investigated the extent to which non-image signals accompanying active head rotation adhered to Weber's law by training participants to rotate their heads at different speeds. The same two-phase protocol described in Experiment 1 was used to measure non-image precision at each trained speed, using auditory stimuli only. These data also allowed a more direct comparison between non-image and image precision. This was not possible in Experiment 1 because the two-phase protocol forces the motion gain of the standard in Phase 1 to be different from that used in Phase 2 (i.e., gain = 1 and PSE, respectively). By

manipulating head speed in Experiment 2, signal precision can be described as a function of magnitude, allowing the comparison to be made.

Figure 6A plots the mean head speed made by participants in the main experiment. The dashed line indicates perfect performance with respect to the head speeds they were trained on prior to data collection. Head movements were reasonably accurate in the main experiment, producing a well separated set of rotation speeds that covered a wide range ($F(4,45) = 16.89$, $p < .001$).

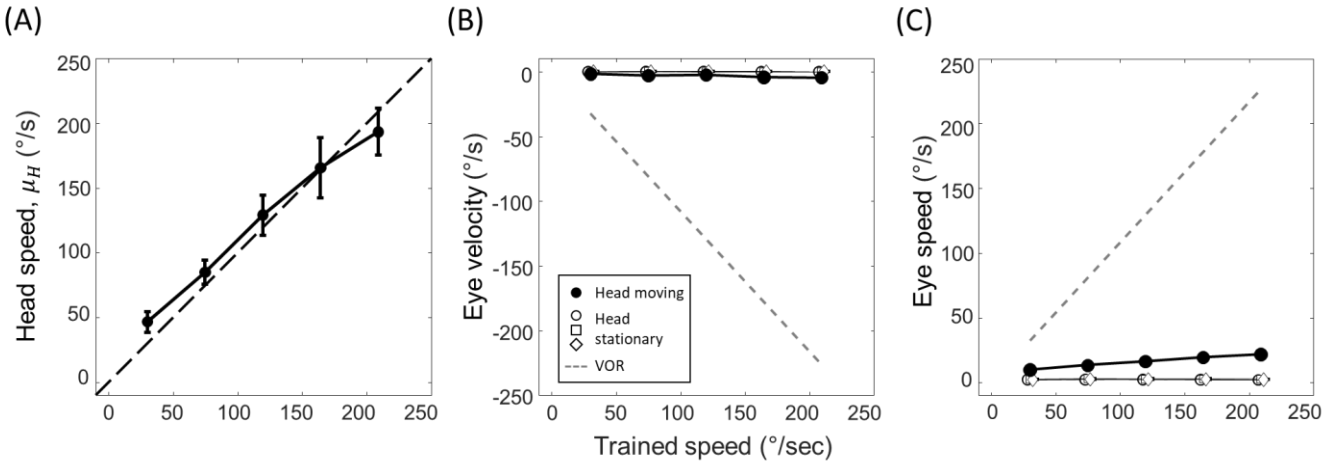

Figure 6: (A) Mean head speed for 10 participants over the ROI defined in Figure 3. The dashed line shows perfect performance with respect to the head speed they were trained on prior to the main experiment. (B) Mean eye velocity over 9 participants for the two intervals in each phase. Open symbols correspond to the three head-stationary intervals, with circles used for the one in Phase 1 and open squares and diamonds for two in Phase 2. Closed circles correspond to the head-moving interval. The dashed line defines the VOR needed to maintain stationary gaze on a world-centred point at the same distance as the speakers. Negative values correspond to an eye movement against the head rotation. (C) Mean eye speed (unsigned average). Error bars are ±1 SE and can be smaller than symbol size.

Figure 6B plots the mean eye velocity for 9/10 participants whom we were able to obtain recordings for. The dashed line indicates the predicted VOR needed to maintain fixation on a world-stationary point at the same distance as the speakers (see Methods for details). Figure 6C plots the mean speed (i.e. unsigned average). In both cases, there was a significant main effect of interval (eye velocity: $F(3,24)=23.07$, $p<.001$; eye speed: $F(3,24)=58.5$, $p<.001$), and a significant interaction with trained speed (eye velocity: $F(12,96)=2.32$, $p=.012$; eye speed: $F(12,96)=17.52$, $p<.001$). Simple effects showed that the interaction was driven by the effect of trained speed on eye speed in the head-moving interval ($F(1,4)=16.98$, $p<.001$), with the same simple effect for eye velocity being close to significant ($F(1,4)=2.38$, $p=.073$). Importantly, however, the eye movement in the head-moving interval was many orders of magnitude smaller than expected from compensatory VOR.

The simple effects for the head-stationary intervals were all non-significant ($p>.254$ or greater). Post hoc comparisons, collapsed across trained speed, showed that none of the head-stationary intervals were significantly different from each other, confirming a key assumption of the paradigm. All three, however, were significantly different from the head-moving interval (Bonferroni-corrected $p<.001$). Whether the relatively small eye movements

found in the head-moving interval could explain the differences between modalities found in Experiment 1 is taken up in the General Discussion.

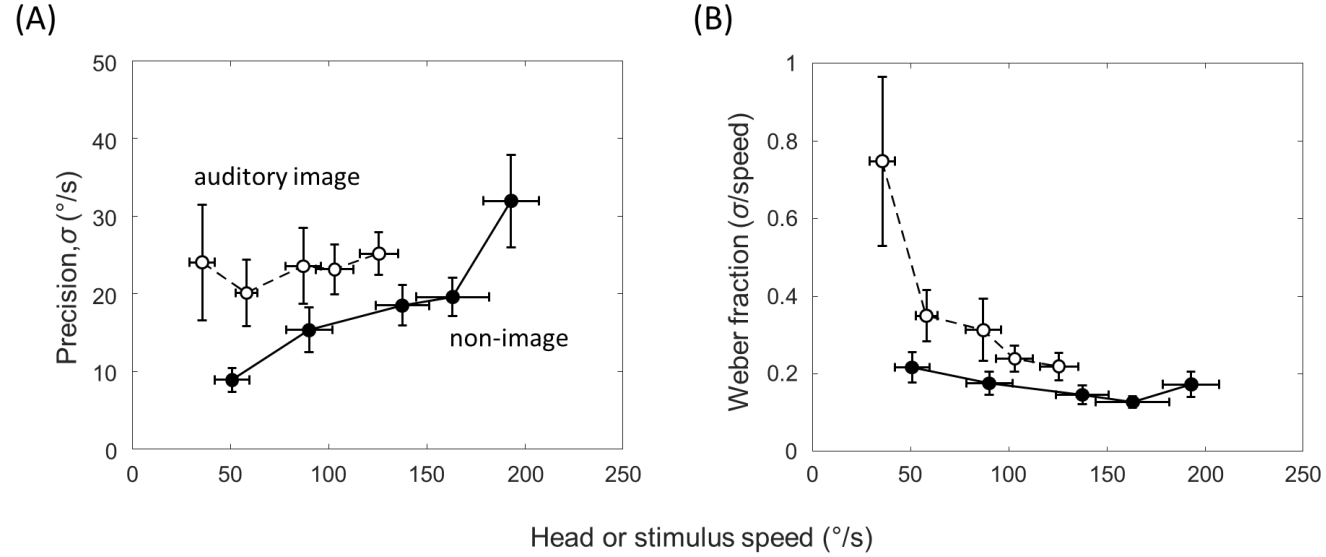

Figure 7: (A) Precision as a function of stimulus speed for the auditory image signal (open symbols) and non-image signals (closed symbols). Precision is defined as the standard deviation of the underlying signal distribution determined by the model in the Appendix. Larger standard deviations correspond to less precise signals. (B) The same data expressed as Weber fractions (i.e. standard deviation / speed). Error bars are ±1SE.

Figure 7A plots the non-image signal precision (filled circles) and auditory image signal precision (open circles) as a function of the mean head or stimulus speed, respectively. The latter compresses horizontally because the speeds are set by the PSE obtained from Phase 2 of the main experiment. This corresponds to a motion gain of around 0.7 (see Figure 8 for the PSEs at each training speed). The horizontal compression is therefore around 30% compared to the closed circles.

For the auditory signal, precision did not vary with stimulus speed ($F(4,45) = 0.154$, $p=.96$). However, for the non-image signal, precision decreased with head speed, such that the standard deviation of the underlying signal distribution significantly increased ($F(4,45) = 6.035$, $p<0.001$). Also evident is the fact that the auditory image signal is less precise than the non-image signal over the range of stimulus speeds tested. Figure 7B plots the same data as Weber fractions (i.e., standard deviation divided by head or stimulus speed). For both types of signal, precision adheres to Weber's law for medium to high speeds. Thus, the Weber fractions are approximately constant over much of the range of speeds tested. At lower speeds, however, the two functions differ, with Weber fractions starting to rise steeply for hearing. This rise is reminiscent of other studies of Weber's law in the perception of auditory motion (4). The same is not true for the non-image signal, where Weber's law appears to hold reasonably well across all head speeds investigated. This finding is similar to previous work using passive stimulation of the vestibular system (3).

Figure 8 plots the mean PSEs from Phase 1 as a function of head speed. They appear similar for all head speeds experienced ($F(4,45) = 0.57$, $p = .69$). Hence the same proportional reduction in image speed was needed to match the perceived motion in the head movement interval. This value was around 0.7, replicating the findings in Experiment 1. Over a wide range of head speeds, therefore, moving auditory stimuli appear slower during head movement, akin to the Aubert-Fleischl phenomenon in vision. However, unlike vision and pursuit eye movement (44, 45), the non-image signal appears more precise than the auditory image signal, which could have important implications for the interpretation of the bias shown in Figure 8. This point is taken up in more detail in the General Discussion.

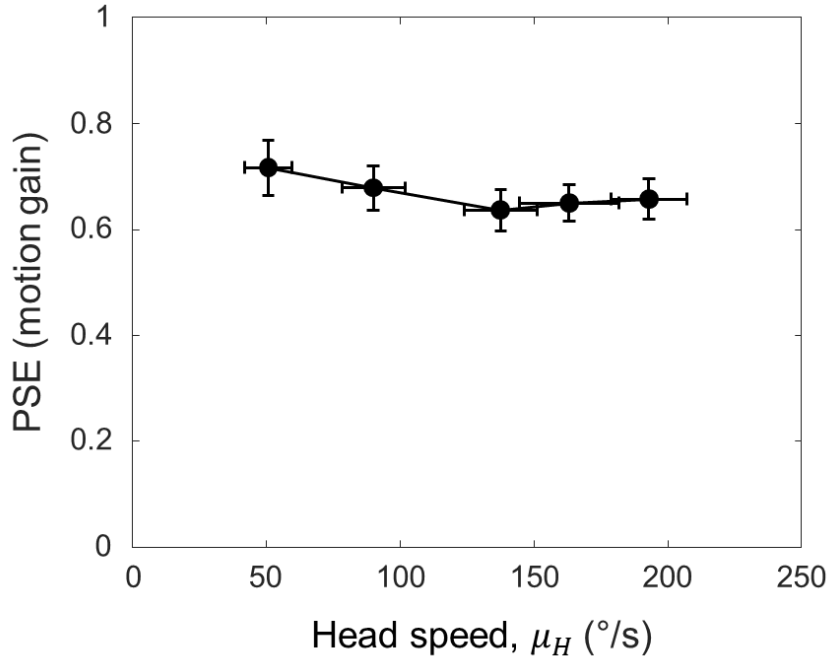

Figure 8: PSEs from Phase 1 as a function of head speed. Error bars are ±1SE.

In Experiment 1, we found non-image signal precision was higher using visual stimuli compared to auditory stimuli. To investigate further, we fit a regression line to the non-image precisions in Figure 7A using Deming's technique, a procedure that is used when both X and Y values are dependent measures with error (46). The result is shown in Figure 9, together with the two non-image signal precision values found in Experiment 1. The regression analysis shows good agreement between Experiments 1 and 2 for auditory stimuli. The precision value from Experiment 1 (open circle) falls very close to the regression line determined by Experiment 2, indicating good replicability of our technique. At the same time, however, the analysis casts further doubt on whether Weber's law can explain the better non-image signal precision found using visual stimuli (open triangle). If Weber's law were to account for the discrepancy in precision, the head speeds in the visual condition of Experiment 1 would need to be roughly halved in order to shift the open triangle

horizontally onto the regression line. Reasons why modality might affect non-image signal precision are taken up in the General Discussion.

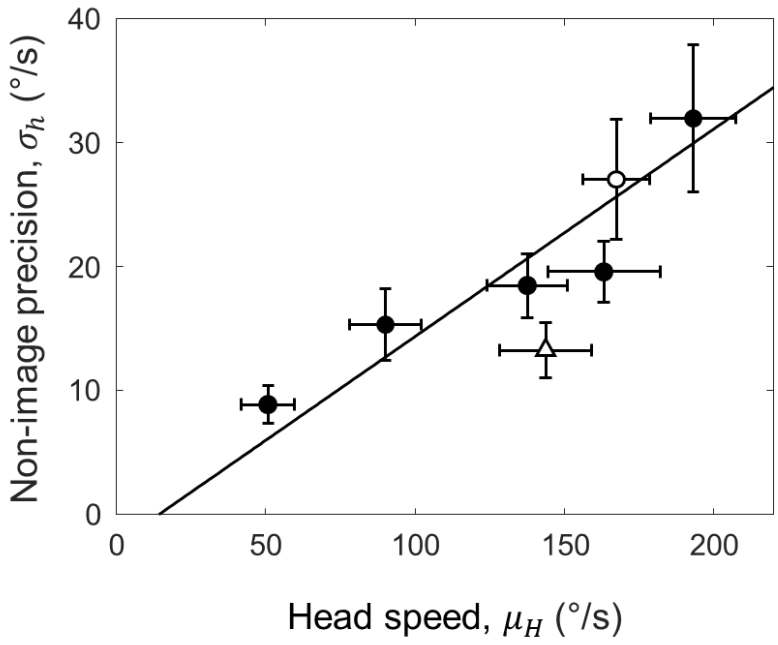

Figure 9: Deming regression (solid line) for the non-image signal precision data from Experiment 2 (filled circles). The open circle is the non-image signal precision from the auditory condition of Experiment 1, and the open triangle the visual condition. Error bars are ±1SE.

General Discussion

We have proposed a novel technique for measuring the combined precision of the non-image signals that encode active self-movement (e.g. vestibular cues, proprioception and motor commands). Traditional psychophysical techniques are difficult to use in these situations because spontaneous self-movement is under participant control. Stimulation is therefore not repeatable across trials. The new technique relies on three factors: (1) linking image motion to prior self-movement using a motion gain parameter that can be manipulated in a consistent fashion across trials; (2) the generation of two psychometric functions limited by identical sources of noise, apart from the internal noise related to non-image signals encoding active self-movement; (3) a model that yields the internal noise sources, while controlling for the external noise created by self-movement as it varies across trials. The technique could easily be adapted for other examples of active self-movement, such as walking and active touch, and situations where non-image signals and image signals are experienced simultaneously (e.g. virtual reality).

Assumptions of the model

The model assumes that eye movements made in all head-stationary test intervals are similar. If this were not the case, then the image noise in Phase 2 could be different from the noise in the test interval of Phase 1. For instance, if observers pursued the stimuli in the test interval of Phase 1, but not Phase 2, then motor cues related to the pursuit system would be present in the first phase but not the second. However, the eye movement recordings in Experiment 2 showed that fixation accuracy was similar in all head-stationary test intervals, with the evidence suggesting that observers were able to keep their eyes stationary. In the head-moving interval, however, we found small eye movements against the head rotation, but these were many orders of magnitude less than would be expected from compensatory VOR. While we did not measure eye movements in Experiment 1, it is unlikely that fixation accuracy explains the difference in non-image precision found when using auditory and visual stimuli. This is because eye fixation would likely be more variable using head-fixed auditory stimuli compared to visual stimuli, and yet fixation accuracy was already very good when using sounds in Experiment 2.

The model assumes that internal noise is fixed. At first sight, this seems at loggerheads with the findings of Experiment 2, which show that the precision of the non-image signal declined as head speed increased. However, while the internal noises were assumed to be fixed across the psychometric function describing each head-speed condition, they were free to vary across conditions. We view the fixed noise assumption as a reasonable approximation for the range of stimuli used to recover a given psychometric function. Indeed, the fixed noise assumption is implicit when fitting a single cumulative Gaussian to the data, unless stimulus values are logged.

The model also assumes that the noise associated with an image signal at rest is negligible. Accordingly, in the head moving interval, the noise that limits performance is assumed to be driven entirely by the non-image signal. For active self-movement this is a reasonable approximation because the magnitudes are typically high, yielding noise far greater than any other signal at rest. An example is provided by Clemens et al, who made a similar

assumption when modelling cue combination between vestibular and extra-retinal cues (1). We note, however, that this approximation is less appealing at slow speeds. A case in point is smooth pursuit, where at least one previous model of perceived speed includes noise terms for the resting state of both retinal and extra-retinal signals (44), in part because the speeds considered were relatively slow (2-12°/s).

Measurement noise

The psychometric functions derived from Phase 1 and 2 are based on the same measurements of head rotation. This controls for any effect of measurement noise introduced by the head tracker because the (head-stationary) test intervals are based on the same set of recordings, presented in identical order. The effect on the precision of the image signal in the two phases is therefore the same, such that any influence is cancelled out.

This does not mean that all of our conclusions are therefore immune to the effect of measurement noise. A case in point is the comparison carried out in Experiment 2 because measurement noise could differentially affect the estimates of auditory signal precision, as these estimates are based on one phase only. However, the effect of measurement noise is probably small. We measured the standard deviation of position samples output by our head tracker as ~0.55 deg. This is considerably lower than the positional noise needed to produce significant changes in speed discrimination thresholds previously reported for visual stimuli (47, 48). For instance, in the 'high noise' condition of Bentvelzen et al, positional noise was added to their LED system with a standard deviation of 7.4° and an update rate of 25 Hz. This produced thresholds that doubled compared to baseline. If the stimuli used in Experiment 2 had been visual, we would therefore expect thresholds to change around 7.4% (i.e. 0.55/7.4 x 100%). Bentvelzen et al used a two-interval technique, so in terms of the precision we report, this equates to a 5.25% change in the standard deviation of the underlying signal. This is a much smaller difference than found between the auditory image signal and non-image signal in Experiment 2. Moreover, spatial hearing is considerably less precise than perifoveal vision, suggesting the effect would be smaller still.

Vision versus Hearing

Experiment 1 showed that non-image signal precision was lower when using visual stimuli than auditory stimuli in the head-stationary test intervals. We also found that head speeds were slightly lower too, but the suggestion that Weber's law might explain the difference in precision was not supported by the regression analysis of Experiment 2. Here we discuss two possible reasons for this difference.

The first concerns extra-retinal signals originating from the eye movement system during head movement. These can be produced by both reflexive (19) and deliberate eye movements (e.g., smooth pursuit – see ref 49 for a review). They can also arise when compensatory VOR is inhibited to fixate head-stationary targets during head rotation (20). Hence the relationship between extra-retinal input and eye movement recordings is not straightforward. Halow, Liu, Folmer and MacNeilage (40) make a similar point when trying to account for changes in perceived scene stability during active and passive vestibular stimulation, combined with different fixation strategies including the use of head-fixed

targets. Nevertheless, in Experiment 2, we found that eye movements were small, indicating good adherence to the instruction to fixate the head-dependent sound. We assume that fixation of a head-fixed visual target would have been the same or slightly better (e.g., 1, 40). On this basis, we doubt that extra-retinal input could explain the difference, given that the change in non-image precision was almost a factor of 2 between modalities in Experiment 1.

A second possible reason for the large difference in non-image precision that we found is the need to convert perceptual signals into common units. The non-image signal combines vestibular and motor cues, both of which are likely to be based on speed. In the case of the vestibular system, the canals mechanically integrate rotary acceleration for all but very low rates (50), leading to canal afferents that encode speed (14, 51). In the case of motor signals, Freeman, Cucu and Smith (24) showed that observers prefer speed versus displacement and duration cues when judging the motion of a pursued target, and that the speed cue was based on extra-retinal signals (i.e. motor commands and/or proprioception). Assuming the same is true for head rotation, the motor signals in our experiments were also in speed units. Hence the non-image signal likely encodes speed, making the comparison with visual motion signals relatively straightforward because vision prefers speed over displacement (24, 25). However, the same cannot be said for hearing, which prefers displacement cues over speed (26, 27). The mismatch for hearing can be resolved by integrating the non-image signal. A transform like this may add noise, leading to a non-image signal that is less precise when moving sounds are used.

An alternative, suggested by one of the reviewers, is that 'silent' visual and auditory cues to self-movement, such as retinal flow, could have played a part in explaining the difference in precision. The idea is that image-signals related to the background are always on and integrated with other self-movement cues, even though our experiments were carried out in a dark and sound-deadened room. This could lead to apparent differences in estimates of 'non-image' precision, and also perceived motion that is lower during head movement, assuming that these undetectable cues are integrated in the same way that detectable cues to self-movement are (52), and that different undetected cues were used in the two conditions. However, we are not aware of any models of cue combination that include undetected cues. An example would be Hillis, Watt, Landy and Banks (53), who identified three additional cues not included in their model. They resolved this issue like us, by making these cues undetectable in their experiments.

Weber's law

In Experiment 2, we found that Weber's law described the precision of both image and non-image signals for medium to high speeds. However, at low speeds Weber fractions for the auditory image signal rose steeply, unlike those for non-image signal precision. Both findings echo previous reports in the literature. For auditory motion based on ITDs, Altman and Viskov (4) found Weber fractions were roughly constant from around 60-140°/s but rose steeply at lower speeds. For vision, the same rise at slower speeds is found but matched by a similar rise at faster speeds (5). In the case of the non-image signal, passive vestibular stimulation reveals good adherence to Weber's law similar to the results we found for active

self-movement, although further analysis by Mallery, Olomu, Uchanski, Militchin and Hullar (3) showed that a power law with an exponent around 0.4 is better description of the raw thresholds than the straight line predicted by Weber's law. Similar behaviour has been reported for the variability in the vestibulo-ocular reflex, an eye movement controlled by the vestibular system (54). One implication is that the non-image signal we measured is dominated by the vestibular system, despite the fact that some neurons in the vestibular nucleus are suppressed during active self-movement (14). This implication only holds if the precision of motor signals encoding head rotation behave differently to the vestibular signals potentially dominating performance in our experiments, but we are unaware of any studies that have isolated the precision of motor cues.

Bayesian models of motion perception

In both Experiments 1 and 2 we found that perceived speed was lower when the head rotated. The bias was very consistent across modalities (Experiment 1) and stimulus speed (Experiment 2), adding to a large body of evidence showing that non-image signals based on eye rotation, head rotation, and hand/arm movement typically provide lower estimates of motion magnitude than signals encoding image motion in vision, hearing and touch (see below). On the face of it, the bias between non-image signals and image signals is puzzling because one might expect this type of constant error to be calibrated out by the perceptual system. One possible explanation is that the bias results from a Bayesian observer balancing unbiased yet imprecise sensory information against prior expectations about the state of the world. The result is a posterior distribution that has greater precision than the original measurements, but not necessarily greater accuracy. As signals become noisier, the position of the posterior is increasingly pulled towards the prior distribution such that accuracy shifts. For motion, the claim is that the prior peaks at 0 because most objects are at rest (55). Hence, as motion signals become noisier, speed estimates reduce.

The Bayesian framework has been used to explain why perceived visual speed slows at low contrast (56), why pursued objects appear slower (44, 45), why moving sounds appear slower when presented against background noise (57) and why tactile stimuli appear slower when made noisier or 'pursued' by an hand/arm movement (31). It can also be used to account for individual differences in motion perception (45). Nevertheless, the overarching theory is not without its detractors (58-60). One simple test is to correlate measures of precision (e.g. thresholds) with bias – the Bayesian hypothesis predicts that as precision declines, perceived speed should slow. Many of the papers cited above show this to be case. However, there are a growing number of reports that this isn't always true. Some recent studies in vision, hearing and vestibular research have shown changes in bias with little evidence for changes in precision (30, 32, 59, 61). At least one paper reports the opposite (48). The findings of Experiment 2 seem to add to these 'non-Bayesian' set of results. They show that auditory motion signals are less precise than non-image signals, even though the latter produce substantially lower estimates of motion magnitude.

Conclusions

We have presented a novel technique for the measurement of the precision of non-image signals encoding active self-movement. We used head rotation as an example of self-

movement, and showed that the precision measured was different using auditory and visual stimuli. One possible explanation is that more steps are needed to get non-image signals and auditory image signals into the same perceptual units. In agreement with current literature, we found that the non-image signal obeys Weber's law over a wide range of stimulus speeds, unlike its image-based counterpart. We also found that the magnitude of perceived motion is reduced during head movement for both vision and hearing. This finding is difficult to explain within a Bayesian framework because we found that image signal precision was lower than non-image signal precision over the wide range of stimulus speeds investigated.

Acknowledgements

The work here was supported by a Leverhulme Trust project grant (RPG-2018-151). We would also like to thank the anonymous reviewers for the improvements they suggested to the paper.

APPENDIX

Phase 1 consists of a head-movement interval followed by an image-motion interval. In the first, there is no image motion as the object is spatially linked to the movement of the participant. Perceived motion therefore depends on a point estimate (h) of the non-image signal encoding head rotation. We assume that h is corrupted by fixed additive Gaussian noise across trials. Using $\mathrm{N}(\mu, \sigma)$ to denote a normal distribution with mean μ and standard deviation σ, the non-image signal is therefore distributed as $h = \mu_h + \mathrm{N}(0, \sigma_h)$. The mean μ_h depends on the head movement magnitude (H), which we also assume is normally distributed across trials (see Figure 3B in the main text). Perceived motion in interval 1 (M_1) is therefore:

$$M_1 = b\ N(\mu_H, \sigma_H) + N(0, \sigma_h) \quad (1)$$

where b is a linear bias term that sets the gain of the head-movement signal relative to its input i.e. $h = b\ H$ Note that either speed or displacement could be used to characterise the distributions of head rotation and signals (to reiterate a point made in the main text, displacement and speed are perfectly correlated when manipulating motion gain because duration is fixed). The model is ambivalent. Swapping between speed and displacement changes the units but not the relative differences found for a chosen parameter across conditions.

In the second interval, image motion (I) moves as a fixed proportion (g) of the head movements recorded in interval 1: $I(t) = g\ H(t)$. We refer to g as the 'motion gain'. As the head and eyes are stationary, sensed movement depends on an image signal (i). Following similar logic to interval 1, the perceived motion in interval 2 is therefore:

$$M_2 = gN(\mu_H, \sigma_H) + N(0, \sigma_i) \quad (2)$$

Note that Equation (2) assumes that the image signal is unbiased. Hence b in Equation 1 defines the *relative* bias between h and i, such that b < 1 means that the non-image signal registers a lower magnitude than the image signal. Following standard signal detection theory (e.g. 62), we assume observers base their choice on an internal decision variable (d) that depends on the difference between the perceived motion in the two intervals:

$$d = M_2 - M_1 \quad (3)$$

The choice 'Interval 2 appears to move more' corresponds to $d > 0$. From signal detection theory we define

$$d' = \frac{\mu_d}{\sigma_d} \quad (4)$$

such that the probability of choosing Interval 2 is given by:

$$P = \frac{\lambda}{2} + (1 - \lambda)\Phi\left(\frac{d'}{\sqrt{2}}\right) \quad (5)$$

where λ is the lapse rate and Φ is the cumulative distribution function of the standard normal distribution.

Substituting (2) and (3) into (4):

$$d = (g - b)\mathrm{N}(\mu_H, \sigma_H) + \mathrm{N}(0, \sigma_i) + \mathrm{N}(0, \sigma_h) \quad (6)$$

By inspection:

$$\mu_d = (g - b)\mu_H \quad (7)$$

Note that the PSE occurs when $\mu_d = 0$. At this point $g = b$; hence the relative bias between h and i can be read directly from the psychometric function. If the bias $b < 1$, then the PSE occurs when image motion is slower than head-movement. This is analogous to the Aubert-Fleischl phenomenon (28, 29), in which moving objects appear slower when pursued. Conversely, if $b > 1$, then image motion must be faster to achieve the PSE.

To obtain σ_d, we sum the variances of the three distributions defined by equation (7) and take their square root:

$$\sigma_d = \sqrt{(g - b)^2\sigma_H^2 + \sigma_h^2 + \sigma_i^2} \quad (8)$$

If the head movement did not vary across trials ($\sigma_H^2 = 0$), then the square root of the sum $\sigma_h^2 + \sigma_i^2$ is the slope of the best-fitting cumulative Gaussian. The precision of the non-image signal (σ_h) could then be obtained by measuring σ_i^2 in Phase 2 and subtracting it from the sum. However, $\sigma_H^2 \neq 0$. Variable head movements make the recovery of σ_h more complicated because they act as an external source of noise that varies with motion gain across the psychometric function.

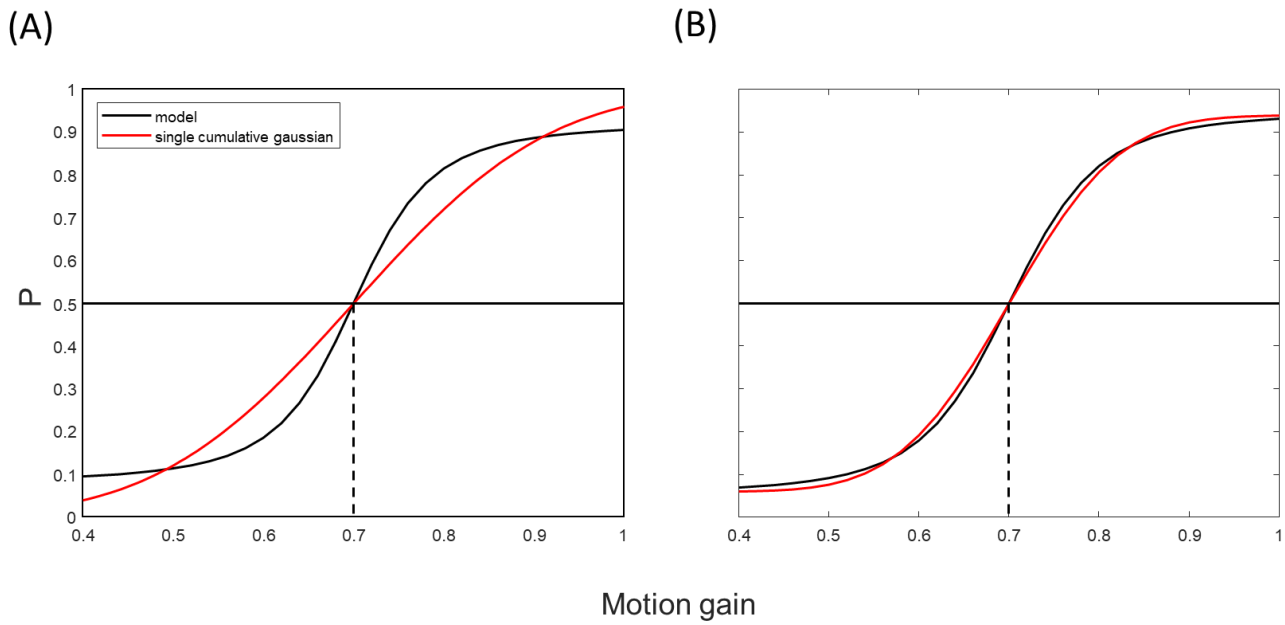

Figure A1: (A) The black curve shows a psychometric function based on the gain-dependent noise with $[\mu_H, \sigma_H^2, \sigma_h^2, \sigma_i^2, b, \lambda] = [2, 0.1, 0.2, 0.1, 0.7, 0]$. *The red curve is the best fitting cumulative Gaussian as determined by the Palamedes toolbox, with lapse rate* $\lambda = 0$. *(B) Same curves but with lapse rate* $\lambda = 0.06$ *for the gain-dependent noise psychometric function, and free to vary for the single cumulative Gaussian (*$\lambda \leq 0.06$, *a standard constraint suggested by Wichman & Hill, 2001).*

Figure A1 shows that fitting a single cumulative Gaussian is an approximation at best. The black curves show example psychometric functions based on the formulae above (see legend for parameter values) while the red curves show the best-fitting single cumulative Gaussian. The difference between the two panels is whether a lapse-rate is included or not. The external noise has two effects: (1) the asymptotes of the psychometric function move away from P= 0 and 1; (2) the slope becomes steeper and is not well fit by a single cumulative Gaussian. The degree to which the external noise causes substantial departures from the standard fit depends on the relationship between the values of $\mu_H, \sigma_H^2, \sigma_h^2, \sigma_i^2, b$ and whether lapse-rate is allowed to vary in the standard fit.

Fitting procedure

We fit psychometric functions to our data based on the formulae above, using the measured head movements to estimate the mean and standard deviation of H. A matlab function for doing this can be found in the Supplementary Material ("fitSMmodel"). Phase 2 data were fit first, with σ_i^2 and λ free to vary, and μ_H and σ_H^2 fixed, using the Gaussian distributions that we fit to the obtained head movement speeds (see Figure 3B in the main text). Phase 1 was then fit, with σ_h^2, b and λ free to vary and σ_i^2, μ_H and σ_H^2 fixed. To avoid local minima in the fit, each parameter was cycled through a search space of 20 values and the best fit chosen. This yielded 20^n separate cycles of the fitting routine, where n is the number of free parameters which was different for the two phases.

We did not find much difference between fitting the new psychometric function and fitting a single cumulative Gaussian. One likely explanation for this similarity was that the head

movements were relatively consistent (σ_H^2 low) given the repetitive nature of the task. It may also be the case that including a constrained lapse-rate parameter soaked up a proportion of the asymptotic effect of the external noise. This can be seen by comparing Figure A1A (no lapse rate) with Figure A1B (constrained lapse rate <= 6%). The lapse rate mimics the asymptotic behaviour produced by the external noise.

Region-of-interest for calculating head rotation speed

The analysis depends on mean and variance of the head movements made in interval 1 (see equations 8 and 9). The mean and variance were estimated from histograms of average speeds in the 3rd sweep as described in the main text. To determine the region-of-interest (ROI), we compared the goodness-of-fit of psychometric functions from three ROIs: 20-80%, 20-60% or 40-60% of the sweep length. The psychometric functions were fit using MLE, so the appropriate measure of goodness-of-fit is the deviance (Wichman & Hill, 2001). Figure A2A shows that deviance in Experiment 1 did not change with the different ROIs used (the deviance has been averaged across conditions, phases and participants). The same was true for Experiment 2 (not shown). However, Figure A2B shows that an ROI of 20-80% produced a slower estimate of head movement speed than the other two ROIs, which was also more variable due to the inclusion of salient periods of acceleration and deceleration. Again the same was true for Experiment 2 (not shown). We therefore opted for an ROI of 20-60%.

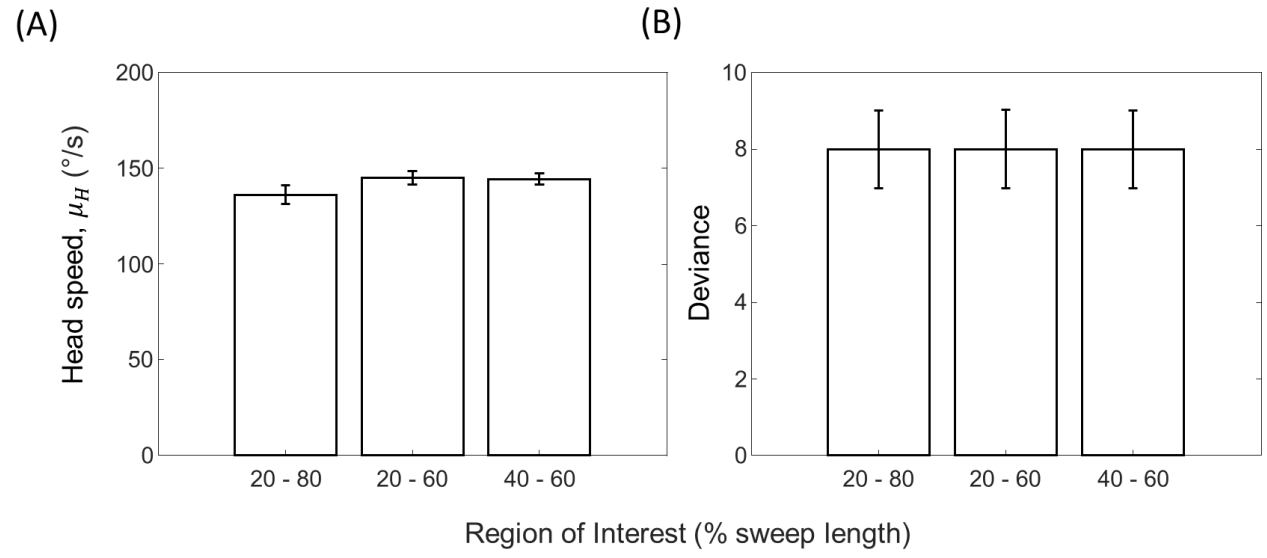

Figure A2: (A) Head speed and (B) model goodness-of-fit for each ROI used to analyse the head movements in Experiment 1. Error bars are ±1SE.

FOOTNOTES

1: The term ‘image’ is less intuitive for hearing because space must be recovered indirectly from binaural and monaural cues. Nevertheless, its use is not without precedence (63). Here we use it to define the dynamic spatial cues the auditory system uses to recover source movement with respect to the head.

REFERENCES

1. **Clemens IAH, Selen LP, Pomante A, MacNeilage PR, and Medendorp WP**. Eye movements in darkness modulate self-motion perception. *ENeuro* 4: 2017.
2. **Genzel D, Firzlaff U, Wiegrebe L, and MacNeilage PR**. Dependence of auditory spatial updating on vestibular, proprioceptive, and efference copy signals. *Journal of neurophysiology* 116: 765-775, 2016.
3. **Mallery RM, Olomu OU, Uchanski RM, Militchin VA, and Hullar TE**. Human discrimination of rotational velocities. *Experimental brain research* 204: 11-20, 2010.
4. **Altman J, and Viskov O**. Discrimination of perceived movement velocity for fused auditory image in dichotic stimulation. *The Journal of the Acoustical Society of America* 61: 816-819, 1977.
5. **De Bruyn B, and Orban GA**. Human velocity and direction discrimination measured with random dot patterns. *Vision research* 28: 1323-1335, 1988.
6. **Brooks JX, and Cullen KE**. Predictive sensing: the role of motor signals in sensory processing. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 4: 842-850, 2019.
7. **Skavenski AA**. Inflow as a source of extraretinal eye position information. *Vision research* 12: 221-IN227, 1972.
8. **Merton P**. Human position sense and sense of effort. In: *Symposia of the Society for Experimental Biology*1964, p. 387-400.
9. **Von Holst E**. Relations between the central nervous system and the peripheral organs. *British Journal of Animal Behaviour* 1954.
10. **Tuthill JC, and Azim E**. Proprioception. *Current Biology* 28: R194-R203, 2018.
11. **St George RJ, and Fitzpatrick RC**. The sense of self-motion, orientation and balance explored by vestibular stimulation. *The Journal of physiology* 589: 807-813, 2011.
12. **Cullen KE, and Zobeiri OA**. Proprioception and the predictive sensing of active self-motion. *Current opinion in physiology* 20: 29-38, 2021.
13. **Israel I, and Warren WH**. Vestibular, proprioceptive, and visual influences on the perception of orientation and self-motion in humans. *Head direction cells and the neural mechanisms of spatial orientation*347-381, 2005.
14. **Angelaki DE, and Cullen KE**. Vestibular system: the many facets of a multimodal sense. *Annu Rev Neurosci* 31: 125-150, 2008.
15. **Jörges B, and Harris LR**. Object speed perception during lateral visual self-motion. *Attention, Perception, & Psychophysics*1-22, 2022.
16. **Warren PA, and Rushton SK**. Optic flow processing for the assessment of object movement during ego movement. *Current Biology* 19: 1555-1560, 2009.
17. **Warren Jr WH, and Hannon DJ**. Direction of self-motion is perceived from optical flow. *Nature* 336: 162-163, 1988.
18. **Royden CS, Banks MS, and Crowell JA**. The perception of heading during eye movements. *Nature* 360: 583-585, 1992.
19. **Bedell HE, Klopfenstein JF, and Yuan N**. Extraretinal information about eye position during involuntary eye movement: Optokinetic afternystagmus. *Perception & Psychophysics* 45: 579-586, 1989.
20. **Barnes G**. Head-eye co-ordination: visual and nonvisual mechanisms of vestibulo-ocular reflex slow-phase modification. *Progress in brain research* 76: 319-328, 1988.
21. **Clark B, and Graybiel A**. The effect of angular acceleration on sound localization: the audiogyral illusion. *The Journal of Psychology* 28: 235-244, 1949.

22. **Graybiel A, and DI H**. The oculo-gyral illusion; a form of apparent motion which may be observed following stimulation of the semicircular canals. *The Journal of Aviation Medicine* 17: 3-27, 1946.
23. **Carriot J, Bryan A, DiZio P, and Lackner J**. The oculogyral illusion: retinal and oculomotor factors. *Experimental brain research* 209: 415-423, 2011.
24. **Freeman TC, Cucu MO, and Smith L**. A preference for visual speed during smooth pursuit eye movement. *Journal of Experimental Psychology: Human Perception and Performance* 44: 162-179, 2018.
25. **Reisbeck TE, and Gegenfurtner KR**. Velocity tuned mechanisms in human motion processing. *Vision research* 39: 3267-3286, 1999.
26. **Freeman TC, Leung J, Wufong E, Orchard-Mills E, Carlile S, and Alais D**. Discrimination contours for moving sounds reveal duration and distance cues dominate auditory speed perception. *PloS one* 9: e102864, 2014.
27. **Carlile S, and Best V**. Discrimination of sound source velocity in human listeners. *The Journal of the Acoustical Society of America* 111: 1026-1035, 2002.
28. **Aubert H**. Die Bewegungsempfindung: Zweite Mittheilung. *Archiv für die gesamte Physiologie des Menschen und der Tiere* 40: 459-480, 1887.
29. **Fleischl vE**. Physiologisch-optische Notizen [Physiological-optical notes], 2. Mitteilung. *Sitzung Wiener Bereich der Akademie der Wissenschaften* 3, 1882.
30. **Garzorz IT, Freeman TC, Ernst MO, and MacNeilage PR**. Insufficient compensation for self-motion during perception of object speed: The vestibular Aubert-Fleischl phenomenon. *Journal of Vision* 18: 9-9, 2018.
31. **Moscatelli A, Scotto CR, and Ernst MO**. Illusory changes in the perceived speed of motion derived from proprioception and touch. *Journal of Neurophysiology* 2019.
32. **Freeman TC, Culling JF, Akeroyd MA, and Brimijoin WO**. Auditory compensation for head rotation is incomplete. *Journal of Experimental Psychology: Human Perception and Performance* 43: 371, 2017.
33. **Stevenson-Hoare JO, Freeman TC, and Culling JF**. The pinna enhances angular discrimination in the frontal hemifield. *The Journal of the Acoustical Society of America* 152: 2140-2149, 2022.
34. **Prins N, and Kingdom FA**. Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers in psychology* 9: 1250, 2018.
35. **Serafin S, Nilsson NC, Sikstrom E, De Goetzen A, and Nordahl R**. Estimation of detection thresholds for acoustic based redirected walking techniques. In: *2013 IEEE Virtual Reality (VR)*: IEEE, 2013, p. 161-162.
36. **Steinicke F, Bruder G, Jerald J, Frenz H, and Lappe M**. Estimation of detection thresholds for redirected walking techniques. *IEEE transactions on visualization and computer graphics* 16: 17-27, 2009.
37. **Wichmann FA, and Hill NJ**. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & psychophysics* 63: 1293-1313, 2001.
38. **Nilsson NC, Peck T, Bruder G, Hodgson E, Serafin S, Whitton M, Steinicke F, and Rosenberg ES**. 15 years of research on redirected walking in immersive virtual environments. *IEEE computer graphics and applications* 38: 44-56, 2018.
39. **Cherni H, Métayer N, and Souliman N**. Literature review of locomotion techniques in virtual reality. *International Journal of Virtual Reality* 2020.
40. **Halow S, Liu J, Folmer E, and MacNeilage PR**. Motor Signals Mediate Stationarity Perception. *Multisensory Research* 36: 703-724, 2023.
41. **Wyatt HJ**. Detecting saccades with jerk. *Vision research* 38: 2147-2153, 1998.
42. **Leigh RJ, and Zee DS** editors. *The neurology of eye movements (3rd edition)*: Oxford: Contemporary Neurology, 1999.

43. **Laming D**. Weber's law. In: *Insid Psychol a Sci over 50 years*: Oxford University Press, 2009.
44. **Freeman TC, Champion RA, and Warren PA**. A Bayesian model of perceived head-centered velocity during smooth pursuit eye movement. *Current Biology* 20: 757-762, 2010.
45. **Powell G, Meredith Z, McMillin R, and Freeman TC**. Bayesian models of individual differences: Combining autistic traits and sensory thresholds to predict motion perception. *Psychological science* 27: 1562-1572, 2016.
46. **Harrison JJ, Freeman TC, and Sumner P**. Saccadic compensation for reflexive optokinetic nystagmus just as good as compensation for volitional pursuit. *Journal of Vision* 15: 24-24, 2015.
47. **Bentvelzen A, Leung J, and Alais D**. Discriminating audiovisual speed: optimal integration of speed defaults to probability summation when component reliabilities diverge. *Perception* 38: 966-987, 2009.
48. **Rideaux R, and Welchman AE**. But still it moves: static image statistics underlie how we see motion. *Journal of Neuroscience* 40: 2538-2552, 2020.
49. **Furman M, and Gur M**. And yet it moves: Perceptual illusions and neural mechanisms of pursuit compensation during smooth pursuit eye movements. *Neuroscience & Biobehavioral Reviews* 36: 143-151, 2012.
50. **Rabbitt RD**. Semicircular canal biomechanics in health and disease. *Journal of Neurophysiology* 121: 732-755, 2019.
51. **Cullen KE**. Vestibular processing during natural self-motion: implications for perception and action. *Nature Reviews Neuroscience* 20: 346-363, 2019.
52. **Prsa M, Gale S, and Blanke O**. Self-motion leads to mandatory cue fusion across sensory modalities. *Journal of neurophysiology* 108: 2282-2291, 2012.
53. **Hillis JM, Watt SJ, Landy MS, and Banks MS**. Slant from texture and disparity cues: Optimal cue combination. *Journal of vision* 4: 1-1, 2004.
54. **Nouri S, and Karmali F**. Variability in the vestibulo-ocular reflex and vestibular perception. *Neuroscience* 393: 350-365, 2018.
55. **Weiss Y, Simoncelli EP, and Adelson EH**. Motion illusions as optimal percepts. *Nature neuroscience* 5: 598-604, 2002.
56. **Stocker AA, and Simoncelli EP**. Noise characteristics and prior expectations in human visual speed perception. *Nature neuroscience* 9: 578-585, 2006.
57. **Senna I, Parise CV, and Ernst MO**. Hearing in slow-motion: Humans underestimate the speed of moving sounds. *Scientific reports* 5: 14054, 2015.
58. **Thompson P, Brooks K, and Hammett ST**. Speed can go up as well as down at low contrast: Implications for models of motion perception. *Vision research* 46: 782-786, 2006.
59. **Hassan O, and Hammett ST**. Perceptual biases are inconsistent with Bayesian encoding of speed in the human visual system. *Journal of vision* 15: 9-9, 2015.
60. **Hammett ST, Champion RA, Thompson PG, and Morland AB**. Perceptual distortions of speed at low luminance: Evidence inconsistent with a Bayesian account of speed encoding. *Vision research* 47: 564-568, 2007.
61. **Freeman TC, and Powell G**. Perceived speed at low luminance: Lights out for the Bayesian observer? *Vision Research* 201: 108124, 2022.
62. **Jones PR**. A tutorial on cue combination and Signal Detection Theory: Using changes in sensitivity to evaluate how observers integrate sensory information. *Journal of Mathematical Psychology* 73: 117-139, 2016.
63. **Middlebrooks J, and Green D**. Sound localization by human listeners. *Annual review of psychology* 42: 135-159, 1991.