Head jitter enhances three-dimensional motion perception

Jacqueline M. Fulvio Department of Psychology, University of Wisconsin – Madison, Madison, WI, USA

Huiyuan Miao Department of Psychology, Vanderbilt University, Nashville, TN, USA

Bas Rokers Psychology, New York University – Abu Dhabi, Abu Dhabi, UAE

Motion perception is a critical function of the visual system. In a three-dimensional environment, multiple sensory cues carry information about an object's motion trajectory. Previous work has quantified the contribution of binocular motion cues, such as interocular velocity differences and changing disparities over time, as well as monocular motion cues, such as size and density changes. However, even when these cues are presented in concert, observers will systematically misreport the direction of motion-in-depth. Although in the majority of laboratory experiments head position is held fixed using a chin or head rest, an observer's head position is subject to involuntary small movements under real-world viewing conditions. Here, we considered the potential impact of such "head jitter" on motion-in-depth perception. We presented visual stimuli in a head-mounted virtual reality device that facilitated low latency head tracking and asked observers to judge 3D object motion. We found performance improved when we updated the visual display consistent with the small changes in head position. When we disrupted or delayed head movement–contingent updating of the visual display, the proportion of motion-in-depth misreports again increased, reflected in both a reduction in sensitivity and an increase in bias. Our findings identify a critical function of head jitter in visual motion perception, which has been obscured in most (head-fixed and non-head jitter contingent) laboratory experiments.

Introduction

In a three-dimensional environment, multiple sensory cues carry information about object motion. Previous work has quantified the contribution of monocular motion cues, such as size changes and optic flow, as well as binocular motion cues, such as interocular velocity differences and changing disparities over time. However, even when these cues are present in concert, observers make systematic errors. They frequently report approaching motion as receding and vice versa (Fulvio, Rosen, & Rokers, 2015).

A second hallmark of motion-in-depth studies are considerable, but inconsistent, response biases. Some studies using (monocular) optic flow cues have reported greater sensitivity to expanding rather than contracting flow fields (Edwards & Badcock, 1993; Raymond, 1994), producing a "toward bias" (i.e., observers were more likely to report a receding stimulus as approaching than vice versa). A study of motion-in-depth using binocular cues has reported a "toward bias" as well (Cooper et al., 2016). However, other work using monocular cues has reported an "away bias" instead (Ball & Sekuler, 1980), and the previously mentioned study that presented monocular and binocular cues together also predominantly reported an "away bias" (Fulvio et al., 2015).

These perceptual errors and biases are surprising because they seem at odds with our everyday experience and suggest that we currently have an incomplete understanding of the factors that govern three-dimensional (3D) motion perception. One possibility is that under more naturalistic conditions, observers rely on additional sensory cues that improve perceptual sensitivity and reduce response bias. A critical difference between the laboratory and everyday viewing is that in typical laboratory environments, the presented visual scene does not update according to changes in the observer's head position. Specifically, conventional paradigms that rely on stereoscopic presentation (e.g., haploscopes, shutter glasses) use a fixed-viewpoint arrangement in which head position is restrained. The head restraint is necessary because in the absence of head tracking, any head movement will result in an inappropriately rendered view of the scene, thereby introducing cue conflicts between visual and vestibular cues.

Citation: Fulvio, J. M., Miao, H., & Rokers, B. (2021). Head jitter enhances three-dimensional motion perception. *Journal of Vision*, *21*(3):12, 1–17, https://doi.org/10.1167/jov.21.3.12.

https://doi.org/10.1167/jov.21.3.12 Received April 1, 2020; published March 9, 2021 ISSN 1534-7362

In the real world however, observers might exploit the sensory cues provided by head movement to enhance the perception of 3D motion. Indeed, whole-body movements of an observer relative to a visual scene provide motion parallax cues that signal object depth (Richards, 1985; Rogers & Rogers, 1992; Rogers, 1993; van Damme & van de Grind, 1996; Nadler, Angelaki, & DeAngelis, 2008; Rogers, 2009). These movements provide retinal cues that, in combination with signals that specify eye orientation relative to the scene, contribute to perceptual performance (Rogers & Rogers, 1992; Nawrot, 2003a; Nawrot, 2003b; Naji & Freeman, 2004; Nawrot & Joyce, 2006; Nawrot and Stroyan, 2009). However, the impact of such motion parallax cues on perception is typically studied under conditions where head and body movements are large and under voluntary control, and the objects viewed are stationary.

In this study, we tested the hypothesis that head-free viewing provides critical signals that enhance the perception of an object's 3D motion trajectory by observers who believe they are stationary. Specifically, we investigated the impact of *small*, random, *involuntary* head movements—"head jitter"—which are prevalent under head-free viewing conditions. Previous work has established that the retinal signals produced by head jitter are large enough to exceed perceptual threshold and can be used in principle to enhance perceptual performance (Aytekin & Rucci, 2012). Moreover, small head movements can enhance static slant (Louw, Smeets, & Brenner, 2007) and distance judgements (de la Malla, Buiteman, Otters, Smeets, & Brenner, 2016). However, enhanced performance for these stimuli could be explained on the basis of the observer having access to multiple redundant views of the same static visual scene. Identifying enhancements in perceptual performance for moving stimuli would suggest more sophisticated neural mechanisms.

We used a head-mounted virtual reality device to manipulate the presented visual scene relative to an observer's head position. We used millisecond latency, sub-millimeter precision head tracking to update the view of the virtual environment according to an observer's unrestrained head position. When head-tracking was off, motion-in-depth perception exhibited stereotypical errors—approaching motion was frequently reported as receding and vice versa. Additionally, performance was biased such that observers were more likely to judge motion as receding than approaching. When head-tracking was on, such that the visual display updated according to small involuntary changes in head position, we observed a significant improvement in perceptual sensitivity and a reduction in response bias. In a series of control experiments, we manipulated head jitter-based visual cues, by manipulating the latency of head-tracking contingent updating of the visual display. Under these conditions sensitivity to motion-in-depth was again reduced and response biases returned.

In summary, we identified a critical function of small, involuntary head movement in motion in depth perception. These results advance our understanding of motion perception in 3D environments and the sensory cues used in human visual processing under naturalistic viewing conditions.

Methods

Availability of materials

The data as well as the analysis code needed to recreate the manuscript's figures and statistics are available at https://github.com/rokers/headjitter.

Observers

Fifty members of the University of Wisconsin–Madison community gave informed written consent to participate, and 44 successfully completed all parts of the study (Experiment 1: $n = 24$, and Experiment 2: $n = 20$ observers). Six observers did not complete the study (Experiment 1: $n = 2$, and Experiment 2: $n = 4$) because of either technical issues, difficulty understanding the task, or perceiving depth in the display. Data from an additional six observers who completed all parts of the study were excluded from further analysis due to excessive head tracking errors ($n = 3$ from each experiment; see "Head Jitter Analysis" section below), so that data from a total of 38 observers is reported in the Results section. The experiments were approved by the Institutional Review Board at the University of Wisconsin–Madison. Observers received course credit in exchange for their participation.

Display apparatus

We used the Oculus Rift Development Kit 2 (DK2; www.oculusvr.com), a stereoscopic head-mounted virtual reality system. The experiment was controlled by custom code using MATLAB and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner, Brainard, Pelli, Ingling, Murray, & Broussard, 2007) on a Macintosh computer and projected on the display of the DK2 headset. Embedded in the headset is a Galaxy Note 3 display—a 14.5 cm low-persistence AMOLED screen—providing a resolution of 1920×1080 pixels (960×1080 pixels per eye, and an average resolution of $\sim$9 pixels per degree) with a refresh rate of 75 Hz. The

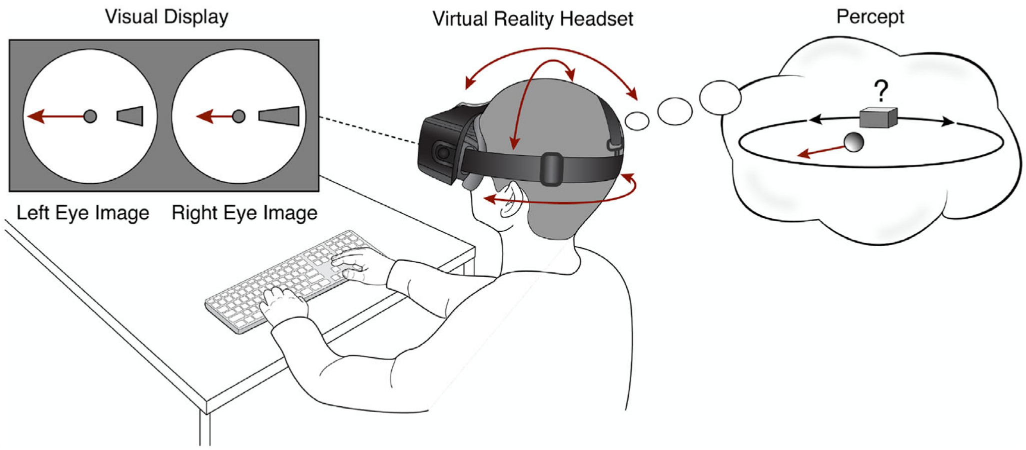

Figure 1. Experimental paradigm. **Percept:** Illustration of perceived visual scene and 3D Pong response paradigm. Observers watched a sphere (gray circle) move through 3D space (red arrow) for 1s before disappearing. Observers subsequently adjusted a paddle (gray rectangle) around a circular orbit (black ellipse, invisible to the observer), so that it would have intercepted the target. This paradigm assesses the observers' perceived target motion. **Virtual Reality Headset:** 3D percepts were generated using a virtual reality (Oculus DK2) head-mounted display, which tracked head movements. Depending on the experimental condition, we updated, delayed, or did not update the virtual scene according to head movement. **Visual Display:** Illustration of left- and right-eye stimulus elements presented in the Headset. The illustration depicts both the target (central sphere), and paddle (left-side trapezoid), although only the target or the paddle was visible at any given time in the actual experiment. The target and paddle were visible within a circular aperture cut out of a "wall" positioned at the focal distance of the display. The target and paddle could move in depth, all other scene elements were stationary relative to the world (but due to head motion not necessarily relative to the observer). See Supplementary Movie S1 for additional details.

horizontal field of view is about 90° (100° diagonal) (Figure 1, "Virtual Reality Headset" panel).

Visual scene

In both experiments, the motion stimuli were presented in the center of a virtual room (3 m × 3.52 m × 3.6 m). The virtual wall, ceiling, and floor were mapped with different tiled textures to facilitate distance judgment throughout the virtual space and judgment of the relative positions of the stimuli. A planar surface mapped with a 1/f noise pattern that was identical in both eyes to aid vergence stood in the center of the virtual room. The observer was located 45 cm from the planar surface in Experiment 1, and 1.2 m (the focal distance of the display) in Experiment 2. All stimulus elements in Experiment 2 were scaled in world size to match the retinal size of the stimuli in Experiment 1 when viewed at the 1.2 m viewing distance. The center of the surface consisted of a gray circular aperture cut-out with a radius of 7.5°. Nonius lines were embedded within a small 1/f noise patch near the center of the aperture to facilitate fixation at the target's starting position on each trial. Observers were asked to fixate the starting position throughout the target's 1s presentation. All stimulus elements were anti-aliased using OpenGL functionality to achieve subpixel resolution.

Observers judged the motion direction of a sphere ("target") 0.8 cm (Experiment 1) or 2.13 cm (Experiment 2) in diameter (Figure 1, "Visual Display" panel; see also Fulvio & Rokers, 2017). The target appeared at the center of a visual scene, moved along a randomly chosen trajectory for 1 s, and subsequently disappeared. The motion trajectory was defined by independently chosen random speeds in the x (lateral) and the z (motion-in-depth) direction, with no change in y (vertical) direction, resulting in motion trajectories that spanned 360° in the horizontal plane. Velocities in both x and z were chosen from a 2D Gaussian distribution (M = 0 cm/s, $SD = 2$ cm/s for Experiment 1 and $M = 0$ cm/s, $SD = 5.33$ cm/s for Experiment 2) with imposed cut offs at ±6.1 cm/s for Experiment 1 and ±16.3 cm/s for Experiment 2. The target approached and receded from the observer on ∼50% of trials. Because the x- and z- motions were chosen randomly and independently, the amount of perceived lateral movement on each trial did not carry information about the amount of motion-in-depth and vice versa. The target was rendered under perspective projection so that both monocular (size, looming) and binocular cues (disparity, interocular velocity differences) were present.

The display was calibrated using standard gamma calibration procedures. Minimum and maximum display luminance was <0.01 cd/m^{2}and 64.96 cd/m^2, respectively. In Experiment 1, the target was presented at one of three Weber contrast levels (0.75 [high], 0.55

[mid], and .47 -> 0.47 [low]), chosen in a random, counterbalanced order across trials. We note that we did not analyze the effect of contrast here. In Experiment 2, the target was presented at the mid-level (0.55) target contrast only.

Experimental task

The experimental task was designed to assess the observer's perceived 3D motion direction. Observers were asked to keep their head position stationary, but because a chin rest was not used, head position tended to drift slightly during a trial. Depending on the experimental condition, the visual display did or did not update according to head position. Observers indicated the direction of the target's trajectory using a "3D Pong" response paradigm (Fulvio et al., 2015; Fulvio & Rokers, 2017). After the target disappeared, a 3D rectangular block ("paddle") appeared (0.5 cm × 1 cm × 0.5 cm in Experiment 1 and 1.33 cm × 2.67 cm × 1.33 cm in Experiment 2). The paddle was textured with a 1/f noise pattern to provide a fine-grained binocular disparity signal. The paddle was located along an invisible "orbit" about the fixation point in the x-z plane that spanned the edge of the aperture within the planar surface. Observers used the left and right arrow keys to move the paddle around an invisible orbit (recall Figure 1, "Percept" panel) to the location that would have intercepted the target had it continued along its trajectory. Observers were instructed to take their time and ensure the accuracy of the location of their response. Observers were not required to fixate during the paddle adjustment. When satisfied with the paddle setting, observers pressed the spacebar. Response feedback was then initiated—the target reappeared at its last visible location and continued along its trajectory. If the target hit the paddle, observers heard a sound indicating a hit ("cowbell"); otherwise, observers heard a sound indicating a miss ("swish"). Observers then resumed fixation and pressed the up-arrow key to begin the next trial.

In Experiment 1, three head-tracking conditions were tested: on, off, and lagged. In the off condition, the visual display did not update according to head motion—as the observer moved, the entire 3D scene moved with them, as if the scene was yoked to the head. In the on condition, the 3D scene updated immediately in response to head motion. In the lagged condition, the 3D scene updated in response to head movement with a random delay chosen uniformly from 0 to 38 additional image frames on each trial. All observers completed the on, off, and lagged conditions in a randomized, counterbalanced order.

In Experiment 2, we further investigated the impact of small but constant, head tracking lag. Observers completed the task in four blocks. In three blocks, a constant lag was added: 0, 1, or 2 image frames. In the fourth block ("mixed"), lags were randomly chosen from those used in the other three blocks with equal frequency. All observers completed the four blocks in a randomized, counterbalanced order.

Observers carried out 10 to 15 practice trials in the presence of the experimenter with head-tracking turned on to become familiar with the task. Observers then completed all experimental conditions in a single session. Feedback was provided on all trials. To prevent prolonged use of the virtual reality (VR) headset and ensure comfort throughout the session, observers completed 360 experimental trials in three 120-trial blocks (Experiment 1) or four 90-trial blocks (Experiment 2) with short breaks in between.

Quantifying perceptual performance

We summarized observer performance in the task based on three behavioral measures. The first behavioral measure, percentage of target interceptions, captured the overall accuracy with which the target's 3D motion direction was estimated and reported across trials. The observer's response was determined an interception if the target hit any portion of the paddle, which comprised an 8° range, yielding a chance interception rate of 2.2%. The percentage of trials intercepted was computed for each observer in each head-tracking condition and across binned trajectory angles. Successful target interceptions imply accurate estimates of both the target's motion-in-depth, as well as its lateral motion.

Because we were especially interested in the potential impact of head jitter on perceptual errors, we calculated motion-in-depth sensitivity (d') and bias (c). We characterized an observer's response as a hit, miss, correct rejection, or false alarm according to the relationship between the presented and reported direction of motion-in-depth on each trial, where we defined a hit to mean that both the presented and reported motion was toward the observer by convention. We then computed motion-in-depth sensitivity as

$$d' = Z\,(Hit\ Rate) - Z\,(False\ Alarm\ Rate)\,, \quad (1)$$

where Z(p), $p \in [0,1]$, is the inverse of the cumulative distribution function of the Gaussian distribution. Finally, we computed motion-in-depth bias as

$$c = -\,(Z\,(Hit\ Rate) + Z\,(False\ Alarm\ Rate))\,/2 \quad (2)$$

such that $c = 0$ indicates that the target's motion-in-depth estimates are unbiased; $c > 0$ indicates a bias to report motion as receding and $c < 0$ indicates a bias to report motion as approaching. We note that this measure is unaffected by changes in d' (Stanislaw & Todorov, 1999). To calculate d' and c averages across observers, we sorted each trial into 15°-wide bins on the basis of the presented motion-in-depth and then further collapsed the four quadrants of the 360° space down to 90°, so that the edges of the 90° space corresponded to lateral motion (left/right) and motion-in-depth (toward/away), respectively.

Head-tracking

Spatial tracking performance for the DK2 HMD provides sub-millimeter translation accuracy (Kotaru & Katti, 2017) via an external camera with near-infrared CMOS sensor taking measurements at 60 Hz. Moreover, the DK2 HMD provides high rotation tracking accuracy (<0.5°) and precision (Chang, Hsu, Hsu, & Chen, 2016) via an accelerometer, gyroscope, and magnetometer embedded in the headset, taking measurements at 1000Hz. Reports of motion-to-photon lag in the DK2 HMD are generally excellent. They vary from ∼1 to 10 ms for small and predictable periodic movements (Kijima & Miyajima, 2016; Zhao, Allison, Vinnikov, & Jennings, 2017) to ∼41 to 48 ms for sudden large movements (Raaen & Kjellmo, 2015; Chang et al. 2016).

To confirm that head tracking in the DK2 headset was sufficiently precise for our purposes, we measured tracking error independent of head jitter. We mounted the DK2 on a platform at the position of a typical observer's head while we simulated an otherwise normal experimental session. Analysis of the head tracking data revealed that tracking error between two subsequent 75 Hz–sampled time points was in the submillimeter and sub-arcmin range. Median translation error was 0.0062 mm (interquartile range [IQR] = 0.009 mm)—horizontal: 0.004 mm (IQR = .005 mm), vertical: 0.006 mm (IQR = 0.008 mm), and in depth: 0.011 mm (IQR = 0.018 mm). Median rotation error was 0.063 arcmin (IQR = .084 arcmin)—pitch: 0.106 armin (IQR = 0.142 arcmin), yaw: 0.045 arcmin (IQR = 0.057), roll: 0.057 arcmin (IQR = 0.070 arcmin). Over the course of the 1s target motion, the median translation error was 0.422 mm (IQR = 0.424 mm), and the median rotation error was 3.931 arcmin (IQR = 4.263 arcmin), which were substantially smaller than observers' head jitter (see Supplementary Figure S1 for an example comparison).

We also measured the head set's motion-to-photon latency using the Oculus Debug Tool, which displays performance statistics on-screen when the device is in use. This reported a motion-to-photon latency of 14 ms. Because this measurement does not take into account the true time of physical device movement, but rather only the measured time, we carried out a separate measurement in which we video-recorded the display of the DK2 in slow-motion (120 Hz). We compared the movement of a glint captured on the screen during the recording to the corresponding movement in the experimental display and noted a lag of 67 to 83 ms, which corresponds to ∼5-6 frames at the device's 75 Hz framerate. In addition, some smoothing of the head-motion signal was evident in these recordings as well. However, we subjected the device to brief, fairly rapid movements for these measurements.

To minimize motion-to-photon latency and smoothing, our experiment was designed to rely on gradual small drifts in head position, not sudden large excursions. We also took advantage of the temporal prediction features of the headset. Furthermore, we optimized our code and monitored the Psychtoolbox output to make sure no frames were dropped during the experiment. Finally, as we show in the results section below, performance decreases as lag in display updating is increased in single frame increments, suggesting that the device's temporal properties are sufficiently granular at this scale to impact behavior.

Head jitter analysis

We analyzed head translations and rotations based on the headset's six degrees of freedom head tracking. A single continuous trace was acquired for each block, composed of the model-view matrix for each eye at every screen refresh (∼13.33 ms). We inverted the model-view matrix and determined the "cyclopean" view matrix, M at each time point based on the midpoint between the two eyes' views. We extracted the time points corresponding to the 1s target presentation on each trial. For visualization purposes (see, for example, Figure 3), we shifted the individual trial traces so that they shared the same origin in 3D space (i.e., at 0, 0, 0). No additional transformations were applied. To characterize observer head movement, we performed a discrete Fourier transform on each of the three movement direction vectors (i.e., horizontal, vertical, and movement in depth) obtained for each experimental block using the fast Fourier transform function in Matlab, resulting in a periodogram for each movement direction (see examples from representative observers in Supplementary Figure S2).

To quantify translation, we computed the head's path length through 3D space ("translation jitter") during the 1s target presentation intervals. We path-integrated head translation by summing the Euclidean distance between consecutive head positions obtained from the X, Y, and Z components of M. Point-to-point estimates ≥ 0.002 m (corresponding to a velocity ≥ 0.15 m/s) were excluded because they were

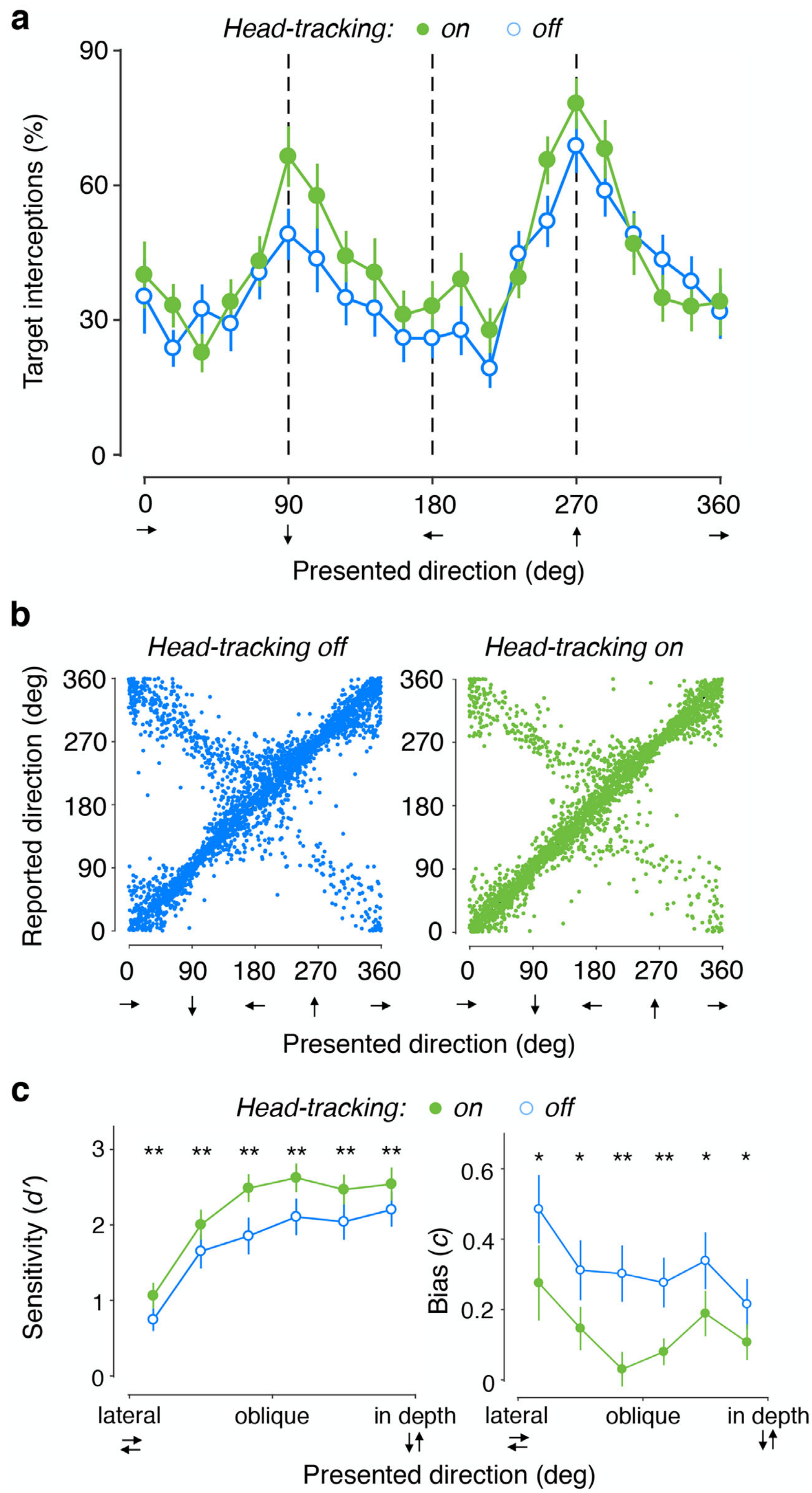

Figure 2. Head-tracking enhances sensitivity to direction of motion-in-depth. (a) Percentage of target interceptions as a function of binned presented direction. Target were intercepted more frequently when head tracking was on, compared to when it was off. The percentage of interceptions was particularly enhanced when targets directly approached or receded from the observer (i.e., presented directions near 90° and 270°). These results suggest that the visual cues produced by head jitter contribute reliable information to sensory estimates of motion-in-depth. (b) Reported direction as a function of the presented direction when head-tracking was off (left panel) and when head-tracking was on (right panel). Each data point represents a single trial for a single participant; each plot depicts 2880 data points. The majority of data points fall along the positive diagonal indicating accurate reports of both the lateral (x) and depth (z) component of the target's motion. However, the prominent negative diagonal corresponds to

→

←
inaccurate reports of the depth (z) component of the target's motion—that is, misreports of motion-in-depth direction. When head-tracking is on, misreports of motion-in-depth direction are reduced, resulting in fewer data points clustered along the negative diagonal. (c) Sensitivity and bias of motion-in-depth perception as a function of presented direction. Left: Sensitivity to motion-in-depth as a function of target direction when head-tracking was on (green) and off (blue) grouped in 15° bins. For all trajectory bins, sensitivity was enhanced when head-tracking was on. Right: Bias in motion-in-depth perception when head-tracking was on (green) and off (blue) grouped in 15° bins. For all bins, bias was reduced when head-tracking was on. This effect was especially robust for oblique target motion trajectories. Data points in (a) and (c) reflect between-subject averages in 15° bins, and error bars correspond to ±1 SEM. Double asterisks (**) in (c) correspond to FDR-corrected *p* values < 0.05 for *t*-tests comparing performance with head-tracking on and off; single asterisks (*) correspond to FDR-corrected *p* values < 0.1.

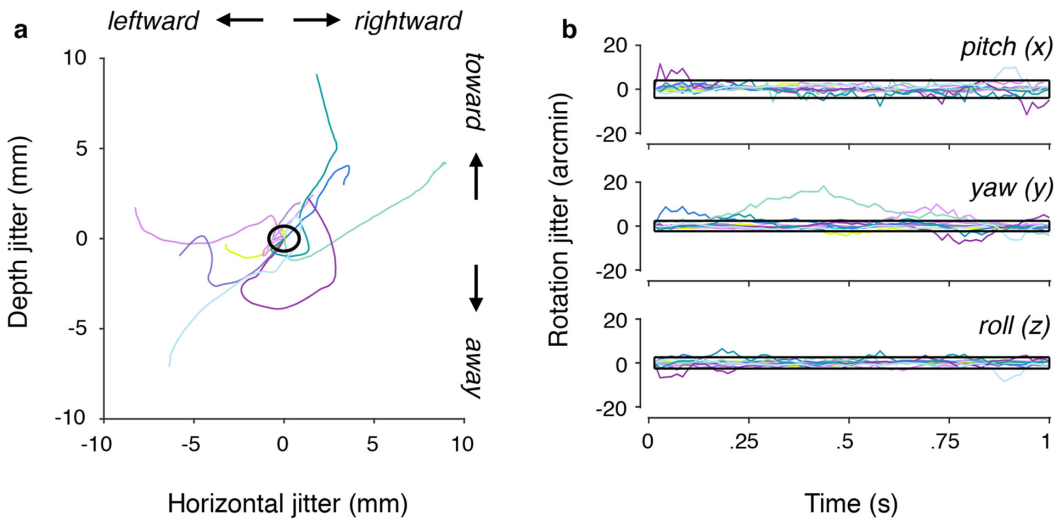

Figure 3. Characterization of head jitter. Observers make small involuntary head movements under naturalistic viewing conditions. (a) Translation jitter (mm) in the X (lateral) and Z (depth) directions at each time point sampled (at 75 Hz) during the target's 1 s presentation for 10 trials when head-tracking was on. For presentation purposes, the smaller translational head jitter in the Y (vertical) direction is not depicted. The black ellipse corresponds to the between-subjects 95th percentiles for the horizontal and vertical translations between any two of the ~13.33 ms sampled time points. (b) Rotation jitter about the three axes during the target 1 s presentation for the same 10 trials using the same color coding. The black boxes correspond to the between-subjects 95th percentiles for the rotations between any two of the ~13.33 ms sampled time points along each axis. The trajectories correspond to data for a representative observer (S22) in Experiment 1.

unrealistically large and likely reflected tracking errors (~3.3% of all time points). Path integration skipped those points. Three observers in each experiment were excluded from further analysis due to excessive tracking errors (~15% of their time points were excluded) compared to the remaining observer data. Translation jitter was typically positively skewed, so the median translation was computed for each observer and averaged across observers.

To quantify rotation, we computed the total angular distance that the head rotated in 3D space ("rotation jitter") during the target's 1s trajectory on each trial. The rotation components were extracted from the "cyclopean" view matrix to give M_R and decomposed to determine the amount of rotation about each axis in the following order: y (yaw), z (roll), and x (pitch). The observer's orientation at the start of the trial was represented by the vector (0, 0, 1), which corresponded to looking straight ahead. The direction vector at each subsequent time point was calculated by computing the dot product of M_R and the starting vector (0, 0, 1). Total rotation jitter was computed by summing the total head rotation between consecutive time points (i.e., the absolute angle between two successive vectors). Point-to-point estimates ≥ ~28.5 arcmin (corresponding to an angular velocity of ~36°/s) were excluded (~2.5% of all time points). When an erroneous tracking time point was identified, path integration skipped that point. Rotation jitter was typically positively skewed, so the median rotation was computed for each observer and averaged across observers.

Statistical analysis

We used analysis of variance to statistically test our results concerning head movement (translation and rotation jitter) and behavior (% target interceptions, d' and c). Analyses incorporated within-subject (repeated-measures) fixed effects, viewing condition (off, on, or lagged head-tracking) in Experiment 1 and lag in Experiment 2. Block order and participant were included as random effects. Bonferroni-corrected t-tests were used for multiple comparisons where indicated. False discovery rate (FDR) corrected t-tests were used for statistical analyses of binned measures (i.e., d' and c).

Results

Head jitter enhances motion-in-depth perception

To identify the contribution of head jitter in motion-in-depth perception, 24 observers performed a virtual-reality based 3D motion extrapolation task (3D Pong) in Experiment 1, while we manipulated the presented visual scene relative to an observer's head position (see Figure 1).

We first compared performance in terms of the percentage of 3D motion target interceptions. Our results revealed a significant increase in behavioral performance when the scene updated according to observers' head movements (head-tracking on) compared to when the scene did not update in response to observers' head movements (head-tracking off; $M_{on} = 45.1\%$, $M_{off} = 38.9\%$; $t(23) = -2.77$, $p < 0.005$, one-tailed). Inspection of Figure 2a suggests that the improvement in performance is asymmetric across presented target directions. Specifically, target interceptions were more likely when the targets directly approached or receded from the observer compared to when the target moved laterally. Moreover, the impact of head jitter was most pronounced for these directly approaching or receding trajectories.

We wished to characterize this impact more directly. Inspection of observers' raw settings when head-tracking was off (Figure 2b, left plot) reveals a prominent negative diagonal, which corresponds to misreports of the direction of the target's motion-in-depth (i.e., the z component). In fact, such misreports occurred on a considerable proportion of head-tracking off trials ($M = 21.4\%$, $SD = 14.2\%$). When the scene did update according to observers' head movements (head-tracking on), motion-in-depth misreports significantly dropped ($M = 15.3\%$, $SD = 13.2\%$; $t(23) = 3.27$, $p = 0.002$, one-tailed) illustrated by a reduction of the prominence of the negative diagonal (Figure 2b, right plot). This result suggests that head jitter cues are especially useful in resolving uncertainty associated with the motion-in-depth direction and is therefore consistent with our hypothesis that extraretinal cues support more accurate perception of approaching versus receding motion.

To quantify this pattern, we computed sensitivity (d') when head-tracking was on compared to off. We found that sensitivity was greater with head-tracking on compared to head-tracking off for all binned trajectories (Figure 2c, left plot; $p < 0.05$ FDR-corrected), suggesting a general head jitter-based enhancement in motion-in-depth sensitivity.

Finally, further inspection of the left plot in Figure 2b, reveals a large density of dots in the upper half of the plot, indicating a bias to disproportionally report object motion as receding rather than approaching, a pattern of results reported in previous studies (Fulvio et al., 2015; Rokers, Fulvio, Pillow, & Cooper, 2018). This bias was significantly reduced when head-tracking was turned on. Significant declines in bias were most pronounced for objects moving along oblique trajectories (Figure 2c, right plot; $p < 0.05$ FDR-corrected for near-oblique trajectories, $p < 0.1$ FDR-corrected for all other trajectories).

These results suggest that head jitter is an important source of sensory information in resolving the direction of object motion-in-depth. In the context of the motion-in-depth estimation task used here, we found that the sensory information associated with head jitter improved estimates of object motion, increasing sensitivity and reducing biases. We next characterized the nature of observer head movement to better understand how head movement contributes to the behavioral effects. Specifically, we asked how the head movement observed in these experiments compared to the large voluntary movements that have been previously shown to be important in resolving position-in-depth through motion parallax signals.

We first analyzed observer head jitter when head tracking was on (see Figure 3 for example recorded head jitter at each 75 Hz-sampled time point, during the 1s target motion of 10 trials from a representative observer). The magnitude of head jitter during the 1s presentation of the target had an average median total 3D head translation of 11.7 mm across observers. Considered separately in the horizontal, vertical, and depth directions, median translations were 5.8, 3.1, and 6.2 mm respectively. In addition, average median total 3D head rotation during the 1s target presentation was

151.6 arcmin across observers. Considered separately for pitch, yaw, and roll, median rotations were 78.2, 95.9, and 74.5 arcmin, respectively. Thus head movements were small in comparison to movements typically considered in the context of motion parallax, but as will be described below, these head movements produce motion parallax signals that are within the limits of visual acuity.

We next assessed whether the head jitter reflected purposeful movements. If purposeful, we should expect a systematic relationship between the presented stimuli and the direction of head motion. If involuntary, we should expect the pattern of head-motion to exhibit a pink noise (1/f) pattern common to many physiological processes.

To test for a systematic relationship between the presented stimuli and the direction of head motion, we quantified the likelihood of the observer's head opposing the target's motion direction. Such a strategy is beneficial because it maximizes motion parallax signals. We found no evidence for a systematic relationship between target direction and observer head motion as head motion direction was independent of target motion direction and not more likely to oppose target motion ($t(20) = 1.164$, $p > 0.05$, two-tailed).

To test for 1/f noise patterns in the head jitter, we applied a discrete-Fourier transform to the continuous raw head movement trace along the three translation axes to obtain the spectral density of the head jitter. We found spectral densities for all three movement axes consistent with a pink (1/f) signal (see Supplementary Figure S2 for the results of this analysis for three observers including the representative observer whose data are depicted in Figure 3). This result supports the interpretation that head jitter during head-free viewing is likely due to physiological noise and not under voluntary control. This interpretation will be discussed further below.

The above results suggest that the benefits of head jitter are opportunistic—observers do not appear to adopt a particular head jitter "strategy" during the experiment. We next quantified head jitter for the same observers when the visual display did not update according to head movement (head-tracking off). Importantly, in this condition, head jitter did not inform the direction of the target's motion, and in fact would be a source of sensory cue conflict if observers did rely on head jitter. As a result, a beneficial strategy in this condition would be to suppress head jitter and reduce cue conflicts.

Indeed, magnitude of head jitter was reduced when head-tracking was off: average median total 3D head translation was 4.0 mm compared to 11.7 mm with head-tracking on ($t(7547) = 7.42$, $p < 0.001$). Similarly, average median total 3D head rotation with head-tracking off was 46.4 arcmin compared to 151.6 arcmin with head-tracking on ($t(7557) = 8.68$, $p < 0.001$). These results suggest that when head motion may produce cue conflicts, observers suppress head jitter.

Considering alternative explanations

The results so far support the interpretation that head jitter is a source of sensory information and improves perceptual estimates of motion-in-depth. Because we rendered the visual scene on the display's discrete pixel array, updating of the rendered scene over time according to small head movements effectively increases the spatial resolution of the display. That is, small head movements will cause scene content to slide across the pixel grid when head tracking is on and that by itself may be sufficient to enhance performance.

We performed two additional analyses, and an additional experiment to exclude such an explanation. First, according to this explanation head jitter should be especially beneficial when the target's motion-in-depth is small, so that it produces an excursion of just a few pixels on the visual display. In that case the combination of pixel grid sliding and anti-aliased rendering might afford a significant benefit that does not depend on a reliable relationship between head jitter and scene geometry.

To address this concern, we performed a median split of the trials in both head-tracking conditions (i.e., on and off) according to the target's z displacement. We subsequently computed the percentage of targets intercepted as a function of both head-tracking condition and target displacement. First, we found no significant interaction between target z-displacement and head-tracking condition (which should be the case if the effect was purely based on anti-aliasing in the small z-displacement case). Second, we found that for both small and large z-displacements, performance was significantly better with head-tracking on than off (small: $p = .044$; large: $p = .01$). Therefore, if anything, the effect in the large z-motion condition is more robust.

Second, head translation and rotation will have different effects on scene geometry. Although both will cause geometry to slide across the pixel grid, head translation will provide more substantial motion parallax information than head rotation. Thus, if head jitter is a source of sensory information, the improvement in target interceptions should largely be explained by translational head jitter, rather than rotational head jitter.

We ran a step-wise generalized linear regression model (glm) to determine the best model (i.e., with the lowest Akaike information criterion) to account for observers' interception performance on each trial in the head-tracking on condition. The fitting tested translational and rotational head jitter as predictors, and the best model included translational head jitter only ($\beta_{translationaljitter} = 0.081$, $p < 0.001$; $\chi^2 = 13.7$, $p < 0.001$ against the constant model). Given that translational and rotational head jitter did not affect the display in the head-tracking off condition, we expected that neither head movement type would predict performance. Indeed a base (constant or intercept) model was the best model in a second step-wise glm analysis for the head-tracking off condition ($p > 0.1$ for other models against the constant model). Taken together, the above results provide further support for a role of head jitter in 3D motion direction discrimination.

Head-motion contingent updating needs to be consistent and fast

To further rule out that our findings were simply due to an effective increase in spatial resolution or some other unanticipated consequence of incorporating jitter in stimulus rendering, Experiment 1 observers ($n = 24$) also took part in a lagged head-tracking condition in which the head-motion contingent updating was delayed on each trial with a random, unpredictable, lag between 0 image frames (the inherent latency of the headset) and 38 images frames (beyond the latency inherent to the device). If head jitter improved performance by way of an effective increase in the spatial resolution of the display, observer performance should be comparable in the head tracking on and lagged conditions. If, on the other hand, head jitter improves performance by providing additional sensory cues, performance should be comparable in the lagged and off head-tracking conditions.

Behavioral performance was consistent with the prediction that head jitter improved performance by providing additional sensory cues and not due to an effective increase in the spatial resolution of the display. Target interceptions were significantly less frequent when head-tracking was lagged rather than on ($M_{on} = 45.1\%$, $M_{lagged} = 39.0\%$; $t(23) = -2.34$, $p = 0.014$, one-tailed; Figure 4a). In fact, target interceptions with lagged head-tracking were not statistically different from target interceptions with head-tracking turned off ($M_{off} = 38.9\%$, $M_{lagged} = 39.0\%$; $t(23) = 0.04$, $p = 0.96$, two-tailed).

We found no differences in terms of sensitivity when head-tracking was lagged: d' was at a level intermediate between the sensitivities observed when head-tracking was on and when it was off, and was not significantly different from either for any trajectories (Figure 4b, left). However, motion-in-depth reports were significantly more biased when head-tracking was lagged compared to when head-tracking was on ($t(23) = 1.92$, $p = 0.03$, one-tailed) and not significantly different from head-tracking was off on average across trials ($t(23) = -0.32$, $p = 0.75$, two-tailed) with differences primarily occurring for head-on and near-oblique trajectories ($p < 0.01$, FDR-corrected, all others $p > 0.1$ FDR-corrected; Figure 4b, right).

In an earlier section, we provided evidence that observers modulate head jitter depending on whether head jitter can or cannot provide useful information. Given that lagged head-tracking also introduces cue conflicts, we expected a similar reduction in head jitter. Indeed, the magnitude of translational head jitter was significantly smaller when head tracking was lagged rather than on ($t(7547) = 2.48$, $p = 0.01$). Furthermore, translational head jitter with lagged head-tracking was not different from head jitter when head-tracking was off (Figure 4c). Rotation jitter followed a somewhat similar pattern. Rotation jitter was significantly smaller when head tracking was lagged rather than on ($t(7557) = 2.97$, $p < 0.01$; Figure 4c). However, rotational head jitter was significantly larger compared to when head-tracking was turned off ($t(7557) = 3.47$, $p < 0.001$).

The behavioral performance supports the interpretation that estimates of object motion-in-depth are improved when head jitter contributes *reliable* sensory cues. However, the significantly smaller magnitudes of head jitter when head-tracking was lagged are also consistent with an alternative explanation. Performance may degrade when the magnitude of head jitter is too small to provide *adequate* signals. To rule this out, we split the lagged head-tracking trials according to the observed magnitude of translation jitter. We selected the trials with translations on par with those when head-tracking was on. These "large jitter trials" had translation jitter $\geq$ the lower bound of the 95% confidence interval on the mean translation jitter in the head tracking on condition, (yielding ~27% of trials on average across observers). The remaining lagged head-tracking trials were grouped as "small jitter trials." For both the small and large jitter trials, we computed the percentage of targets intercepted.

We did not find evidence for a difference in performance between the large and small jitter trials ($t(20) = 0.17$, $p = .87$, two-tailed). We further note that the mean lag between the small and larger jitter trials was not significantly different ($t(20) = 0.36$, $p = .72$, two-tailed). Critically however, even in the lagged but large jitter trials, target interceptions were significantly less frequent when compared to interceptions in the head tracking on condition ($t(20) = 2.42$, $p = .03$, two-tailed). Taken together, these results suggest that performance was degraded in the lagged head-tracking condition not because head jitter became too small to provide adequate retinal signals, but rather because

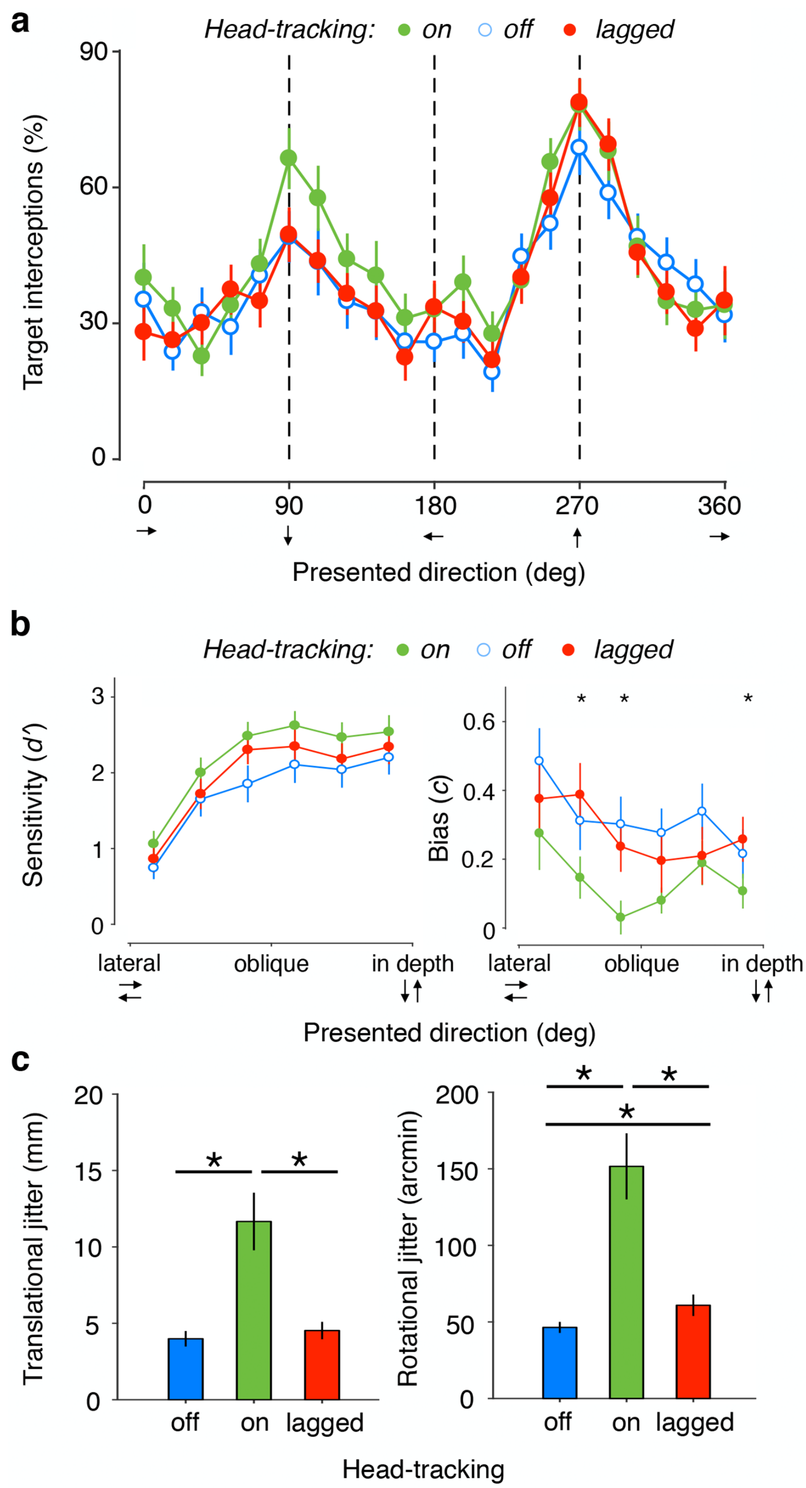

Figure 4. Reliability of head-motion contingent updating impacts both visual performance and head jitter. (a) Visual performance (percentage of targets intercepted) as a function of presented motion direction. The presented stimuli when head-tracking was lagged (red symbols), were identical to those presented when head-tracking was on (green symbols) and when head-tracking was off (blue symbols). However, with lagged head-tracking, the visual scene was updated according to head movement with a temporal lag on each trial, randomly chosen from the 0-38 image frame interval. Performance with lagged head-tracking was worse than with head-tracking on, but not statistically different from performance with head-tracking off. These results suggest that performance is not enhanced when the visual signals provided by head jitter are unreliable. (b.) Left: Sensitivity to approaching motion quantified by d' as a function of binned presented target direction when head-tracking was on (green), off (blue), and lagged (red). Performance in the lagged condition was generally intermediate relative to performance when head-tracking was on and off. Right: Bias in

→

←
motion-in-depth direction report quantified by c as a function of binned presented target direction when head-tracking was on (green), off (blue), and lagged (red). Data points in (a) and (b) reflect between-subject averages in 15° bins, and error bars correspond to ± 1 SEM. In the right panel of (b) * correspond to FDR-corrected p values < 0.1 for t-tests comparing performance with head-tracking on and lagged. No other comparisons involving the lagged head-tracking performance were statistically significant. (c) Head jitter as a function of viewing condition. Left: 3D translation head jitter (mm) for the three viewing conditions. Right: 3D rotation jitter (arcmin) for the three viewing conditions. Head jitter was significantly smaller under conditions in which head-motion contingent information was unavailable (head-tracking off) or unreliable (head-tracking lagged) compared to when it was reliable (head-tracking on). Asterisks (*) correspond to a significant difference between two viewing conditions at the Bonferroni-corrected alpha level of .0167 for three comparisons. Error bars correspond to ± 1 SEM.

lagged head-tracking provided an unreliable sensory cue.

In summary, the results for lagged head-tracking indicate that an increase in effective spatial resolution of the display cannot fully account for the improvement observed with head-tracking on, indicating head jitter is exploited as a cue. These results further support the role of head jitter in motion-in-depth perception. These results also suggest that when head jitter is uninformative (head-tracking off) or is unreliable (lagged head-tracking), observers suppress head motion, a strategy that is beneficial because it reduces sensory cue conflicts.

The results of Experiment 1 demonstrate that observers are sensitive to micro-parallax signals produced by small involuntary head movements. Furthermore, the decline in performance with lagged head-tracking demonstrates that observers are sensitive to the temporal lag between the vestibular and visual signals. The average lag used in Experiment 1 was about 253 ms (about 19 image frames) beyond the inherent latency of the display. Large temporal lags are associated with increased motion sickness (Biocca, 1992; Hettinger & Riccio, 1992; Pausch, Crea, & Conway, 1992). Consequently, VR development has been focused on reducing lag, and consensus is that lag should be 20 ms or smaller (Abrash, 2012; LaValle, Yershova, Katsev, & Antonov, 2014; Yao, Heath, Davies, Forsyth, Mitchell, & Hoberman, 2017). However, is 20 ms a reasonable benchmark? Addressing this requires better understanding of the sensitivity to the visual information accompanying head jitter and more specifically characterizing the tolerance for temporal lag.

To assess the sensitivity to temporal lag in visual information accompanying head jitter, we leveraged the high frame rate of the device (75 Hz). In a new set of observers, we measured behavioral performance on the 3D motion extrapolation task and head jitter. Observers completed three experimental blocks with small levels of added lag: 0, 1, and 2 image frames beyond the inherent latency of the system. They also completed a fourth block in which all three added lags were randomly interleaved among trials (“mixed”).

As expected, we found a significant decrease in the observers’ ability to intercept the target with increase in added lag ($F(3,76) = 2.93$, $p = 0.04$). In particular, performance was significantly worse in the largest added lag condition of two image frames compared to the smallest added lag condition of 0 image frames ($t(76) = -2.69$, $p < 0.01$; Figure 5a). Although the impact of added lag appeared to impact response bias, neither bias nor sensitivity was significantly affected by the small added lags used in this experiment ($p > 0.1$, FDR-corrected for all trajectories and both measures except for a trending increase in bias for near-oblique trajectories with 1 frame of lag added; see Figure 5b).

Even the small added display lags in Experiment 2 were sufficient to suppress head jitter. Increase in added lag led to a significant reduction in translation head jitter ($F(3,6114) = 2.74$, $p = 0.042$). Specifically, translation jitter was significantly smaller with 2 image frames of added lag compared to both the 0 images frames added lag ($t(6114) = -2.14$, $p = 0.03$), and mixed conditions ($t(6114) = -2.16$, $p = 0.03$; Figure 5c). Although similar trends were evident in rotation jitter, these effects were not significant ($F(3,6116) = 1.48$, $p = 0.22$; Figure 5c). It is noteworthy that head jitter was not suppressed in the mixed lag condition of Experiment 2, considering that head jitter was significantly suppressed in the variable lag condition of Experiment 1. We return to this point in the Discussion. Previous efforts have found mixed results concerning the role of head jitter on perceptual accuracy (Louw et al., 2007; de la Malla et al., 2016). Our results reveal the tight relationship between head jitter and display lag. Added lag as little as ~27 ms (two image frames) is sufficient to impact both perceptual performance and head jitter magnitude.

Finally, the magnitude of head jitter associated with enhanced perceptual performance across both of our experiments was small. When head tracking was on, the average horizontal translation was 5.8 mm in Experiment 1 and 2.7 mm in Experiment 2. Thus one might be concerned about the magnitude of the retinal signals associated with these small head movements. Specifically, how do the retinal signals produced by head jitter relate to the limits of visual acuity?

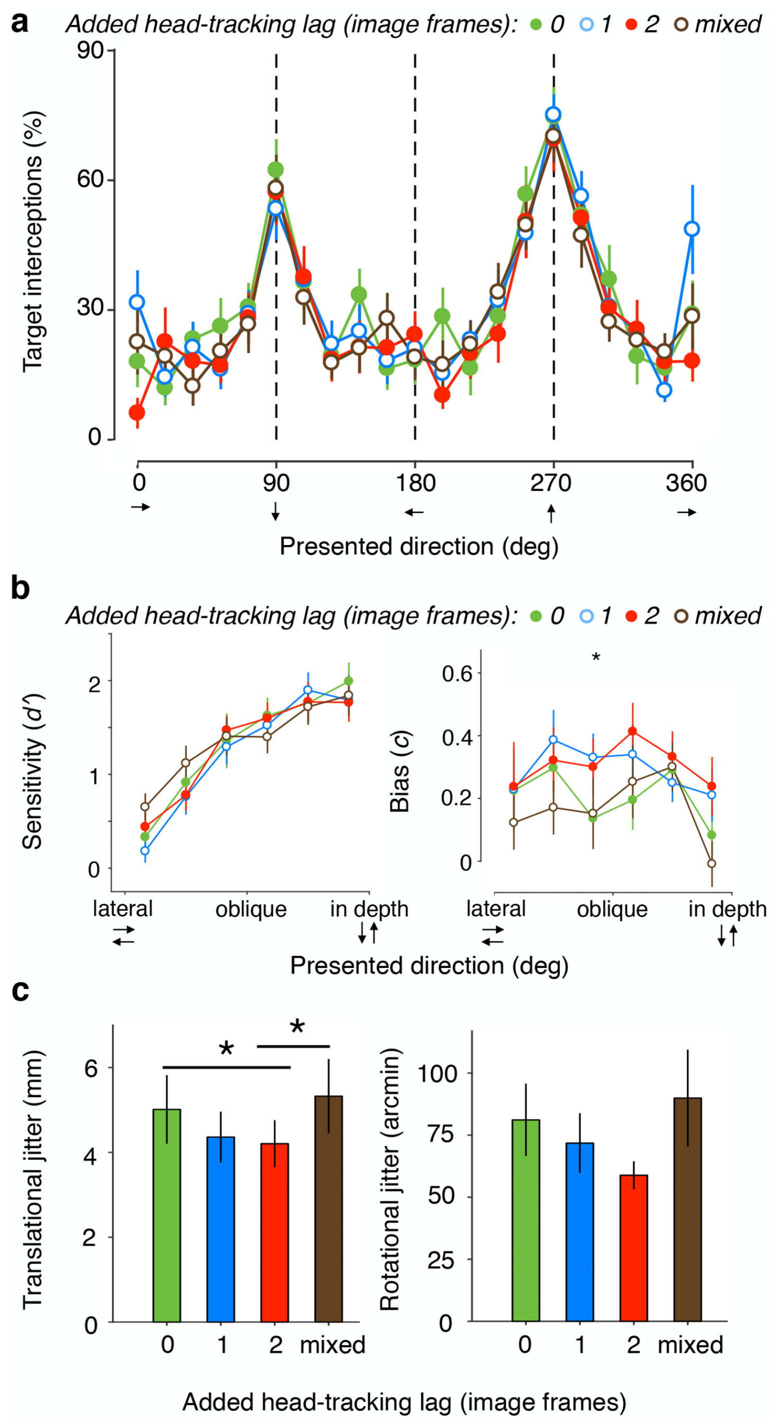

Figure 5. The impact of display lag on both visual performance and head jitter in a 3D motion perception task. The visual display was updated according to head movement at various delays (0, 1, 2 image frames or "mixed" in which all three lags were randomly assigned across trials). (a) Target interceptions and head jitter as a function of lag. Target interceptions (left): A 2 image frame lag in addition to the lag inherent to the system was sufficient to significantly degrade visual performance. (b) Sensitivity (right) and bias (left) as a function of lag. Lag did not significantly impact either measure. Error bars correspond to ±1 SEM. * symbols correspond to FDR-corrected p values < 0.1 for t-tests comparing performance with 0 and 1 frame of lag added. (c) Translation head jitter (mm) as a function of lag (left). Rotation head jitter (arcmin) as a function of lag (right). A significant reduction in head translation occurred with two image frames of lag in addition to the lag inherent to the system. A similar trend was observed in head rotation; however, this effect failed to reach significance. These results further demonstrate that head jitter enhances visual performance when head-motion contingent updating is both fast and reliable. Error bars correspond to ±1 SEM. Asterisk (*)corresponds to significance at the Bonferroni-corrected alpha level.

We first computed the magnitude of the retinal signal. The average horizontal translation of 5.8 mm due to head motion in Experiment 1 corresponds to an angular translation of 44 arcmin at the 45 cm viewing distance. In Experiment 2, a horizontal translation of 2.7 mm corresponds to an angular translation of 7.73 arcmin at the 120 cm viewing distance. Given that hyperacuity limits are in the range of 20 arcsec or less (Westheimer, 1979; Westheimer & McKee, 1977; Westheimer & McKee, 1980), the retinal motions measured here are an order of magnitude larger than the limits of hyperacuity.

In addition to signals of retinal position, head motion also produces retinal velocity signals. Head motion produces a retinal velocity signal of 44 arcmin/s at the 45 cm viewing distance, and a velocity of 7.73 arcmin/s at the 120 cm viewing distance. These values are well above the detection threshold reported by Ujike and Ono (2001) of around 0.26 arcmin/s provided the head moves slowly. Thus both the retinal location and retinal velocity signals measured for our participants generally fall within the limits of visual sensitivity. An open question for future work is which of these signals is the predominant contributor to perceptual enhancement based on head jitter.

Discussion

We assessed the role of small, involuntary head movements (“head jitter”) on motion-in-depth perception. Head jitter significantly enhanced perceptual performance, increased perceptual sensitivity and reduced response bias. Head jitter especially improved the perception of motion trajectories directly approaching the observer, and reduced the tendency to report approaching targets as receding that is characteristic of behavior in this task under head-fixed conditions (Fulvio et al., 2015; Rokers, Fulvio, Pillow, & Cooper, 2018).

Prior work has suggested that motion parallax information provided by head-free viewing produces retinal velocity signals above detection thresholds (Ujike & Ono, 2001; Aytekin & Rucci, 2012), and enhances depth perception in static scenes (Louw et al., 2007; de la Malla et al., 2016). Here we show that such head motions enhance the perception of motion-in-depth as well. Although enhanced performance in the previous work can be explained on the basis of observers having access to multiple redundant views of the same static visual scene, our results using moving stimuli suggest more sophisticated neural mechanisms.

Why has the contribution of head jitter to visual perception historically been underexplored? First, there are technical reasons. Until very recently low-latency high-accuracy head-tracking has been expensive and difficult to implement. The advent of commercially available virtual reality headsets has made this technology substantially more accessible. Second, the desire to carefully control experimental viewing conditions has led to the widespread adoption of head restraints, such as a chin-rests or bite bars. Unfortunately, such restraints limit the study of visual perception to fixed viewpoint conditions. Third, there is the view that eye movements compensate for head movements such that head jitter has relatively little impact on the content of the retinal image. However, compensatory eye movements are not perfect, producing retinal slip (Ferman, Collewijn, Jansen, & van den Berg, 1987). Furthermore, retinal image velocities are greater under head-free viewing conditions, than under head-fixed conditions (Aytekin, Victor, Rucci, 2014). Although we were not able to measure eye movements within the VR headset, characterizing the interaction between eye and head movements will be important in further characterizing how observers use sensory information produced by head jitter, especially under naturalistic 3D content viewing conditions. Nevertheless, our results show that in addition to the monocular and binocular cues to motion-in-depth, small involuntary head movements play an important perceptual role.

The observed magnitude of head movement in our experiments was small, on the order of millimeters and arcminutes when head-tracking was on. When head-tracking was off or lagged, head jitter was further reduced. We believe this effect is due to an involuntary reduction in head movement when the visual cues associated with head jitter are absent or unreliable. Head jitter was much less reduced in Experiment 2 where lag was substantially smaller and less variable.

Interestingly reductions in head movement do not only seem to occur when the cues associated with head jitter are absent or unreliable, but also when they are thought to be so. In the absence of task feedback observers inexperienced with VR suppress head jitter even when it provides reliable sensory cues (Fulvio & Rokers, 2017). We interpreted this finding as reflecting the observers’ extensive prior real-world experience with traditional displays (e.g., TV screens, computer displays, cell phones), where head movement is an unreliable cue to depth and motion-in-depth. In that work, we found that observers require explicit visual and auditory feedback in order to incorporate available cues to motion-in-depth when making motion-in-depth direction judgments. In the current study, we focused on identifying the cues that participants pick up on when feedback is provided. Our results suggest that the benefits largely derive from head jitter-related motion parallax signals, which improve sensitivity and reduce bias in motion in depth perception.

Importantly, we found that the spectral density of head movements was consistent with a 1/f (pink) noise pattern, suggesting that head jitter was due to random physiological drift. Observers were not aware that the magnitude of their head jitter depended on the presence or absence of head-tracking contingent updating of the visual display. Furthermore, we have found no evidence that observers develop a head movement strategy that maximizes motion parallax information on a trial-by-trial basis.

Finally, why is it that observers dynamically suppress head movements, rather than simply treating head jitter as a source of sensory noise when it provides unreliable or conflicting information? We hypothesize that the suppression of head jitter reduces visuovestibular cue conflicts and the development of motion sickness (Golding, Markey, & Stott, 1995; Golding, Bles, Bos, Haynes, & Gresty , 2003). Although observers in our study did not report discomfort beyond that of more traditional stereoscopic setups, we suspect that this is in part because observers self-regulate. The onset of discomfort causes the observer to reflexively reduce head motion.

Conclusion

In conclusion, the current study provides insight into the sensory processes underlying accurate motion-in-depth perception in 3D environments. In particular, our results reveal the role of underappreciated sensory cues provided by head jitter. Observers take advantage of the signals accompanying such small head movements but are highly-sensitive to their reliability.

Keywords: 3D motion perception, naturalistic perception, head movement, virtual reality, head-tracking

Acknowledgments

J.M.F. and B.R. were supported by funding provided by Facebook Reality and Google Daydream.

Commercial relationships: none.
Corresponding author: Jacqueline M. Fulvio.
Email: jacqueline.fulvio@wisc.edu.
Address: Department of Psychology, University of Wisconsin – Madison, WI 53706, USA.

References

Abrash, M. (2012). "Latency: the *sine qua non* of AR and VR," Dec. 2012. [Online]. Available: http://blogs.valvesoftware.com/abrash/latency-the-sine-qua-non-of-ar-and-vr/.

Aytekin, M., & Rucci, M. (2012). Motion parallax from microscopic head movements during visual fixation. *Vision Research, 70*, 7–17, https://doi.org/10.1016/j.visres.2012.07.017.

Aytekin, M., Victor, J. D., & Rucci, M. (2014). The visual input to the retina during natural head-free fixation. *Journal of Neuroscience, 34*, 12701–12715, https://doi.org/10.1523/JNEUROSCI.0229-14.2014.

Ball, K., & Sekuler, R. (1980). Human vision favors centrifugal motion. *Perception, 9*(3), 317–325, https://doi.org/10.1068/p090317.

Biocca, F. (1992). Will simulation sickness slow down the diffusion of virtual environment technology? *Presence: Teleoperators and Virtual Environments, 1*, 334–343, https://doi.org/10.1162/pres.1992.1.3.334.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436, https://doi.org/10.1163/156856897X00357.

Chang, C. M., Hsu, C. H., Hsu, C. F., & Chen, K. T. (2016). Performance measurements of virtual reality systems: Quantifying the timing and positioning accuracy. In *Proceedings of the 24th ACM International Conference on Multimedia* (pp. 655–659), https://doi.org/10.1145/2964284.2967303.

Cooper, E. A., van Ginkel, M., & Rokers, B. (2016). Sensitivity and bias in the discrimination of two-dimensional and three-dimensional motion direction. *Journal of vision, 16*(10):5, 1–11, https://doi.org/10.1167/16.10.5.

de la Malla, C., Buiteman, S., Otters, W., Smeets, J.B., & Brenner, E. (2016). How various aspects of motion parallax influence distance judgments, even when we think we are standing still. *Journal of Vision, 16*, 1–14, https://doi.org/10.1167/16.9.8.

Edwards, M., & Badcock, D. R. (1993). Asymmetries in the sensitivity to motion in depth: A centripetal bias. *Perception, 22*(9), 1013–1023, https://doi.org/10.1068/p221013.

Ferman, L., Collewijn, H., Jansen, T. C., & van den Berg, A. V. (1987). Human gaze stability in the horizontal, vertical and torsional direction during voluntary head movements, evaluated with a three-dimensional scleral induction coil technique. *Vision Research, 27*, 811–828, https://doi.org/10.1016/0042-6989(87)90078-2.

Fulvio, J.M., & Rokers, B. (2017). Use of cues in virtual reality depends on visual feedback. *Scientific Reports, 7*, 1–13, https://doi.org/10.1038/s41598-017-16161-3.

Fulvio, J. M., Rosen, M. L., & Rokers, B. (2015). Sensory uncertainty leads to systematic misperception of the direction of motion in depth. *Attention, Perception, and Psychophysics, 77*, 1685–1696, https://doi.org/10.3758/s13414-015-0881-x.

Golding, J. F., Bles, W., Bos, J. E., Haynes, T., & Gresty, M. A. (2003). Motion sickness and tilts of the inertial force environment: active suspension systems vs. active passengers. *Aviation, Space, and Environmental Medicine, 74*, 220–227.

Golding, J. F., Markey, H. M., & Stott, J. R. R. (1995). The effects of motion direction, body axis, and posture on motion sickness induced by low frequency linear oscillation. *Aviation, Space, and Environmental Medicine, 66*, 1046–1051.

Hettinger, L. J., & Riccio, G. E. (1992). Visually induced motion sickness in virtual environments. *Presence: Teleoperators and Virtual Environments, 1*, 306–310, https://doi.org/10.1162/pres.1992.1.3.306.

Kijima, R., & Miyajima, K. (2016). Measurement of Head Mounted Display's latency in rotation and side effect caused by lag compensation by simultaneous observation—An example result using Oculus Rift DK2. In *2016 IEEE virtual reality (VR)* (pp. 203–204), https://doi.org/10.1109/VR.2016.7504724.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*, 1.

Kotaru, M., & Katti, S. (2017). Position tracking for virtual reality using commodity WiFi. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 68–78), https://doi.org/10.1109/CVPR.2017.286.

LaValle, S.M., Yershova, A., Katsev, M., & Antonov, M. (2014). Head tracking for the Oculus Rift. *IEEE International Conference on Robotics and Automation,* 187–194, https://doi.org/10.1109/ICRA.2014.6906608.

Louw, S., Smeets, J. B., & Brenner, E. (2007). Judging surface slant for placing objects: a role for motion parallax. *Experimental Brain Research, 183*, 149–158, https://doi.org/10.1007/s00221-007-1043-8.

Nadler, J. W., Angelaki, D. E., & DeAngelis, G. C. (2008). A neural representation of depth from motion parallax in macaque visual cortex. *Nature, 452*, 642–645, https://doi.org/10.1038/nature06814.

Naji, J. J., & Freeman, T. C. (2004). Perceiving depth order during pursuit eye movement. *Vision Research, 44*, 3025–3034, https://doi.org/10.1016/j.visres.2004.07.007.

Nawrot, M. (2003a). Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. *Vision Research, 43*, 1553–1562, https://doi.org/10.1016/S0042-6989(03)00144-5.

Nawrot, M. (2003b). Depth from motion parallax scales with eye movement gain. *Journal of Vision, 3*, 841–851, https://doi.org/10.1167/3.11.17.

Nawrot, M., & Joyce, L. (2006). The pursuit theory of motion parallax. *Vision Research, 46*, 4709–4725, https://doi.org/10.1016/j.visres.2006.07.006.

Nawrot, M., & Stroyan, K. (2009). The motion/pursuit law for visual depth perception from motion parallax. *Vision Research, 49*, 1969–1978, https://doi.org/10.1016/j.visres.2009.05.008.

Pausch, R., Crea, T., & Conway, M. (1992). A literature survey for virtual environments: Military flight simulator visual systems and simulator sickness. *Presence: Teleoperators and Virtual Environments, 1*, 344–363, https://doi.org/10.1162/pres.1992.1.3.344.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442, https://doi.org/10.1163/156856897X00366.

Raaen, K., & Kjellmo, I. (2015). Measuring latency in virtual reality systems. In *International Conference on Entertainment Computing* (pp. 457–462), https://doi.org/10.1007/978-3-319-24589-8_40.

Raymond, J. E (1994). Directional anisotropy of motion sensitivity across the visual field. *Vision Research, 34*(8), 1029–1037, https://doi.org/10.1016/0042-6989(94)90007-8.

Richards, W. (1985). Structure from stereo and motion. *JOSA A, 2*, 343–349, https://doi.org/10.1364/JOSAA.2.000343.

Rogers, B. J. (1993). Motion parallax and other dynamic cues for depth in humans. *Reviews of Oculomotor Research, 5*, 119–137.

Rogers, B. (2009). Motion parallax as an independent cue for depth perception: A retrospective. *Perception, 38*, 907.

Rogers, S., & Rogers, B. J. (1992). Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Perception and Psychophysics, 52*, 446–452, https://doi.org/10.3758/BF03206704.

Rokers, B., Fulvio, J.M., Pillow, J.W., & Cooper, E.A. (2018). Systematic misperceptions of 3-D motion explained by Bayesian inference. *Journal of Vision, 18*, 1–23, https://doi.org/10.1167/18.3.23.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers, 31*(1), 137–149, https://doi.org/10.3758/BF03207704.

Ujike, H., & Ono, H. (2001). Depth thresholds of motion parallax as a function of head movement velocity. *Vision Research, 41*, 2835–2843, https://doi.org/10.1016/S00426989(01)00164-X.

van Damme, W. J., & van de Grind, W. A. (1996). Non-visual information in structure-from-motion. *Vision Research, 36*, 3119–3127, https://doi.org/10.1016/0042-6989(96)00060-0.

Westheimer, G. (1979). Scaling of visual acuity measurements. *Archives of Ophthalmology, 97*(2), 327–330, https://doi.org/10.1001/archopht.1979.01020010173020.

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Research, 17*(8), 941–947, https://doi.org/10.1016/0042-6989(77)90069-4.

Westheimer, G., & McKee, S. P. (1980). Stereoscopic acuity with defocused and spatially filtered retinal images. *JOSA, 70*(7), 772–778, https://doi.org/10.1364/JOSA.70.000772.

Yao, R., Heath, T., Davies, A., Forsyth, T., Mitchell, N., & Hoberman, P. (2017). Oculus VR Best Practices Guide v 310-30000-02, https://static.oculus.com/documentation/pdfs/intro-vr/latest/bp.pdf.

Zhao, J., Allison, R. S., Vinnikov, M., & Jennings, S. (2017, March). Estimating the motion-to-photon latency in head mounted displays. In *2017 IEEE Virtual Reality (VR)* (pp. 313–314), https://doi.org/10.1109/VR.2017.7892302.