User Self-Motion Modulates the Perceptibility of Jitter for World-locked Objects in Augmented Reality

Hope Lutwak*
New York University
Meta

T. Scott Murdison†
Meta

Kevin W. Rio‡
Meta

ABSTRACT

A key feature of augmented reality (AR) is the ability to display virtual content that appears stationary as users move throughout the physical world ('world-locked rendering'). Imperfect world-locked rendering gives rise to perceptual artifacts that can negatively impact user experience. One example is random variation in the position of virtual objects that are intended to be stationary ('jitter'). The human visual system is highly attuned to detect moving objects, and moreover it can disambiguate between the retinal velocities that arise from object motion and self-motion, respectively. In this study, we investigated how the perceptibility of AR object jitter varies as a function of user self-motion. Using a commercially available AR HMD to display a 3D textured cube, we measured sensitivity to added jitter versus a no-jitter reference using a two-interval forced choice task. Three user motion conditions (stationary, head rotation, and walking) and three object placement conditions (floating in free space, on a desk, and against a wall) were tested in a full factorial design. We hypothesized that (1) as users move their head and eyes during self-motion, their sensitivity to jitter will decrease, due to added retinal velocity; and (2) rendering virtual objects near physical surfaces will increase sensitivity to jitter, by providing proximal veridical visual cues. Psychometric thresholds indicated that users were significantly less sensitive to jitter during self-motion than when they were stationary, consistent with hypothesis (1). Users were also more sensitive to jitter in one of the two object placement conditions, providing partial support for hypothesis (2). To generalize beyond distinct user motion and object placement conditions, we also analyzed eye tracking data. The amount of retinal slip (i.e. how much gaze drifted across the virtual object) predicted jitter thresholds better than recorded head movements alone, suggesting a retinally-driven decrease in jitter sensitivity during self-motion. These results can be used to inform requirements for AR world-locked rendering systems, as well as how these may be updated dynamically using online measurement of user head and eye movements.

Keywords: Augmented reality, jitter, visual perception, world-locked rendering.

1 INTRODUCTION

Virtual, mixed, and augmented reality (VR, MR, and AR, respectively) have the potential to become foundational computing platforms over the coming decades. AR and MR, in particular, promise to blend digital content and the physical ("real") world, embedding information within a user's natural environment. To enable this, an essential technology is the ability to estimate and represent the geometry of the physical world, and display virtual content that appears colocated with and stable relative to it. This process is known as world-locking (WL) or world-locked rendering. It is a key feature that distinguishes true AR systems from other see-through displays, such as heads-up displays (HUDs) and smartglasses.

Generating convincing WL content in an AR head-mounted display (HMD) requires a complex computational pipeline that begins with input from cameras, depth sensors, and IMUs; followed by several stages of processing, head pose estimation, and rendering; and ultimately culminates in the precisely-timed illumination of photons on the display. This is a technical challenge under any circumstances, but is made even more difficult by the razor-thin size, weight, power, and thermal budgets required to build compact AR glasses that can be worn for long durations. In order to meet these constraints, every component of the AR WL rendering system must be highly optimized.

We propose using the study of human vision to inform this optimization, by generating data about users' fundamental perceptual capabilities and limits, which can serve to bound the space of engineering solutions. In this study, we manipulated the magnitude of one WL rendering error, jitter (defined as high-frequency spatiotemporal 3D position errors). We systematically added jitter to the 3D position of WL objects in an AR HMD, and measured participants' ability to detect the added jitter relative to a no-jitter reference across different classes of user self-motion and configurations of virtual content relative to physical surfaces. We also propose an underlying visual mechanism (retinal slip) to explain some of the observed variation in sensitivity, as well as generalize beyond the distinct experimental conditions that were tested.

The results from this study provide foundational empirical data about human sensitivity to the precision of AR WL rendering, as well as how that sensitivity varies as a function of user self-motion and virtual object placement. To our knowledge, this is the first public report of perceptual data on AR WL jitter sensitivity collected from freely-moving observers wearing 6 degrees of freedom (6DoF) AR HMDs. We believe that these results can be used by AR systems architects and experience designers to ground engineering requirements for WL rendering, enabling compelling experiences in consumer devices.

2 BACKGROUND

2.1 Motion sensitivity in the human visual system

The human visual system is highly attuned to detect motion [10]. Psychophysical experiments have demonstrated that observers can detect motion from displacements as small as 10 arcseconds, equivalent to 1/360º of visual angle [48, 59, 81, 92] and smaller than the width of a single photoreceptor [68, 69]. This implies that motion discrimination is not fundamentally limited by the spatial resolution of the retina, placing it within a class of perceptual capabilities known as 'hyperacuities' along with Vernier acuity and stereoacuity [40, 53, 91]. Small differences in relative velocity are also readily perceived. In random dot stereograms, observers can discriminate between differences in speed as small as 5% [16] and differences in direction as small as 1º [88].

More than simply detecting motion, human vision is designed to guide effective action in dynamic environments [28,33,54]. Consider a hunter chasing down its prey [60], an outfielder running towards a fly ball [21, 52], or a pedestrian walking over uneven terrain [51].

*email: hlutwak@nyu.edu
†email: smurdison@meta.com
‡email: kevinwrio@meta.com

More directly, consider a user moving in an AR environment while aiming towards a target with a controller or their gaze [19, 80, 90].

Dynamic tasks like these produce a constantly-changing pattern of retinal stimulation known as optic flow [27, 41, 46]. A key challenge for the visual system is to extract information by disambiguating between optic flow caused by self-motion (e.g. radial flow due to locomotion [45, 86], translational and rotational flow due to eye movements [44, 85]) and the movements of external objects [84]. Dozens of mechanisms across all levels of visual cortex are dedicated to solving this problem [55], and solutions include 'parsing' optic flow into subcomponents [71, 84], extracting higher-order visual information [47], and visual-vestibular integration [50, 64], among others. A detailed summary of all of these mechanisms is outside the scope of this paper, but several comprehensive review articles have been written (e.g. [33, 43]).

Of particular relevance, the visual system is able to detect moving objects during self-motion [87, 95] and throughout eye movements [26]. Sensitivity to extrinsic motion is decreased during saccades at low spatial frequencies, but there is little to no loss at high spatial frequencies [9]. Human observers however do not perfectly detect object motion. Our visual system has varying sensitivity to different types of 3D motion [13, 95], as well as differences in sensitivity across environmental conditions [84] or eye and head orientations [57, 58]. Object motion can also influence scene perception [22].

In sum, the human visual system has evolved over millions of years to be highly attuned to detect both object motion and self-motion, and to disambiguate them in natural viewing contexts. This sensitivity sets a high bar for the goal of presenting virtual objects that appear stable relative to the physical world in AR/MR, a feature known as world-locking rendering.

2.2 World-locked rendering in AR/VR

World-locked rendering refers to the capability to present virtual content so that it appears colocated with and stable relative to the physical world, as a user moves throughout it. In order to correctly render WL objects, an AR or MR display system must build a map of the physical world and continuously estimate the user's 6DoF pose (3DoF orientation, x/y/z + 3DoF orientation, yaw/pitch/roll) within it using a global coordinate system. This is typically accomplished by fusing data from cameras, depth sensors, and inertial measurement units (IMUs) using a class of algorithms known as simultaneous localization and mapping (SLAM) [18, 34]. This is aided by forward prediction to estimate the user's future position at the time photons will be emitted from the display, taking into account delays due to rendering, frame buffering, display response time, and other sources of latency [79, 94].

Errors can arise at any stage of the WL rendering pipeline. These include tracking errors, approximations or heuristics to facilitate real-time rendering of complex scenes, overshooting in forward prediction, and so on. These errors result in perceptual artifacts for users, which break the illusion of WL and can reduce immersion, interfere with interactions, and cause visual discomfort [4, 49].

A canonical example of these errors is known as jitter. Jitter is a general term in signal processing that refers to random fluctuations to a signal over time [17], and by convention is limited to relatively high frequencies. ITU-T G.810 classifies jitter as variations at frequencies of 10 Hz and higher [35]. In the context of virtual and augmented reality displays, jitter is used to describe at least two different and, in principle, orthogonal artifacts. Positional or spatial jitter refers to fluctuations in the 2D (x/y) or 3D (x/y/z) position of user interface elements. Similarly, rotational jitter [5] refers to fluctuations in orientation (typically in 3D, yaw/pitch/roll). Temporal jitter refers to stochastic fluctuations in latency, the delay between movement (e.g. of an input device, like a mouse; or an HMD) and updated rendering.

In practice, positional and temporal jitter are often correlated, because allowing for longer latencies can reduce both positional and temporal jitter [63, 94]. But in principle, these two artifacts are orthogonal and can be manipulated independently [63, 79]. In this paper, we will consider only positional jitter, and use it synonymously with "jitter." Nevertheless, temporal jitter is also an important artifact and warrants further investigation. Temporal jitter has been demonstrated to have deleterious effects, including causing interaction difficulties [32] and contributing to simulator sickness [78]. However, because these effects arise from different mechanisms (both perceptually and in the architecture of AR/VR systems) and can be manipulated independently, we focus here on positional jitter.

2.3 Previous research on positional jitter in HCI and AR/VR

The effects of positional jitter in 2D UI have been widely reported [37, 63]. In VR, positional jitter has been has been shown to decrease user performance in 3D targeting tasks [4, 6], increase error rate for gaze interaction modalities [19, 56], and decrease the sense of presence when interacting with virtual agents [49].

By contrast, relatively little empirical data on human sensitivity to positional jitter in WL AR is publicly available. This is due, at least in part, to the relatively recent arrival of commercially-available, widely-distributed 6DoF see-through AR systems. The Microsoft HoloLens and Magic Leap One were released in 2016 and 2018, respectively. Using the Microsoft HoloLens 2, Wilmott and colleagues [93] measured the perceptibility of jitter in WL AR objects, across different luminance background levels, contrast ratios, and object distances. They found that observers were highly sensitive to jitter and could reliably detect this motion artifact at magnitudes of $\sim$3 arcminutes in some conditions. They also showed that jitter was more perceptible at greater viewing distances, and at higher background (physical world) luminances. Guan and colleagues used a custom-built display apparatus [30] to measure sensitivity to WL rendering errors arising from incorrect interpupillary distance (IPD) and departures from nominal eye relief [31]. Their display system included a bite bar to stabilize head position, eye tracking, a wide field-of-view stereoscopic display, and a method for calibrating and then continuously estimating the position of the user's head and eyes in space, in order to render WL objects above physical references with very high precision (<3 mm). They reported thresholds in the range of 1 arcminute for disparity errors, and 3 arcminutes for visual direction change.

While these studies [31, 93] provided some of the first published data on sensitivity to WL rendering errors, they also constrained observer movement in order to maintain experimental control. In [93], observers wore a 6DoF AR HMD, but were instructed to remain stationary, thereby limiting their self-motion only to small fluctuations needed to maintain postural stability [36, 67]. The display used in [31] constrained user self-motion to 1DoF yaw rotation. We are not aware of published data that characterizes human sensitivity to jitter in WL AR while users are able to move freely in 6DoF, which we present here.

2.4 Motivation & Contributions

In this study, we investigate how jitter thresholds for objects in WL AR vary under naturalistic and representative user motion conditions. We measured sensitivity while users were stationary (similar to [93]), shaking their heads left-to-right in 1D yaw rotation (similar to [31]), and walking freely along a semicircular arc. These data allow for comparison to previous results for 0DoF and 1DoF user motion, as well as elucidating the more general case of 6DoF. In addition to user motion, we also manipulated the proximity of WL objects to different configurations of physical surfaces. We measured sensitivity while virtual objects were floating in space (far from physical surfaces), horizontally coincident with the surface of a physical desk, and vertically coincident with a physical wall.

We had two chief hypotheses. First, we hypothesized that sensitivity to jitter would decrease during user motion, due to the added retinal velocity that arises from imperfectly-stabilized fixation (known as retinal slip). Second, we hypothesized that sensitivity to jitter would increase when displaying virtual objects near physical surfaces, due to the presence of proximal and veridical visual information.

By quantifying human sensitivity to jitter for different user self-motion and object placement conditions, we aim to better understand the perceptual mechanisms that enable WL AR objects to appear stable within the physical world. In turn, this understanding can be used to derive requirements for and to optimize the performance of WL rendering in AR systems. This is part of a broader movement to use the study of visual perception to inform the design of AR/VR displays [12], which are fundamentally coupled to and dependent upon human vision.

3 METHODS

The current study quantified user perception of WL positional jitter artifacts for AR objects. We employed a psychophysical staircase procedure to measure the magnitude of added jitter needed to reliably discriminate jittering from (rendered-to-be) stationary virtual content. Participants viewed WL virtual objects on a commercially available AR HMD. On each trial, a stationary and a jittering cube were presented sequentially in random order and participants reported which object appeared to jitter. The pattern of responses across added jitter magnitudes was used to estimate each participant's detection threshold, our primary dependent variable. We manipulated two independent variables (user self-motion, WL object placement), with 3 levels per variable, in a full factorial design (for a total of 9 conditions).

3.1 Participants

21 participants (13 identified as male; 8 female) began the experiment, and 19 completed all conditions in the experiment. The remaining 2 participants did not complete all 9 conditions because they did not return for the second of two experimental sessions. In addition, one response data file from one participant was missing. Taken together, these represent a loss of 4.7% (9 out of 189 conditions) of potential data.

Mean age across participants was 36.1 years (min: 27, max: 51). All participants had prior experience with AR/VR devices, and reported normal or corrected-to-normal vision. Study design and protocols were approved by the WCG IRB (WIRB-Copernicus Group Institutional Review Board).

3.2 Apparatus

A Microsoft HoloLens 2 (Redmond, WA, USA) head-mounted display (HMD) was used to render and display WL AR content. The HoloLens 2 is a self-contained AR system, with onboard compute, power (battery), display, see-through optics, speakers, sensors, eye tracking, and head tracking.

The experiment took place in a large room (7 m length, 4 m width) with white walls. The background luminance of the physical world was 60 nits, measured using a photometer (SpectraScan PR-788, Photo Research, New Syracuse NY, USA).

3.3 Stimuli

Participants viewed a 3D virtual cube, generated by a custom Unity (San Francisco, CA, USA) application and displayed on the HoloLens 2. The cube's size (20 cm side length) was chosen such that it subtended $\sim 8^\circ$ of visual angle when viewed at a distance of 1.5 m. This enabled participants to make yaw head rotations while keeping the cube within the fixed field of view (FOV) of the HoloLens 2 ($\sim 52^\circ$ diagonal; 43º H x 29º V), which was validated during pilot testing. Each face of the cube was textured with a 2D grayscale 1/f Gaussian noise pattern (see Figure 1A) to approximate the spatial frequency statistics found in real world scenes [20, 70].

The 3D position of the virtual cube in world coordinates was computed and updated by the tracking and rendering systems of the HoloLens 2. This position served as the 'ground truth' location of the virtual cube. Errors and artifacts from the WL rendering system of the HoloLens 2 itself include some magnitude of jitter, which we term 'baseline jitter' (see Limitations, Section 7). To this, we introduced 'added jitter' by perturbing the 3D position of the virtual cube away from ground truth. At a rate of 15 Hz, the position of the midpoint of the cube was selected randomly from a uniform distribution in 3 dimensions about the ground truth position, forming a sphere of possible positions (see Figure 1C). The rate of 15 Hz was chosen as it is representative of typical SLAM camera capture rates (e.g. [73, 75], and is evenly divisible into the HoloLens 2 refresh rate (60 Hz), reducing the likelihood of introducing temporal jitter [2] or judder [14] motion artifacts. The cube's final rendered position was updated at a rate of 60 Hz using a spline interpolation between the randomly-chosen points, in order to match the render and refresh rate of the display.

The magnitude of added jitter that we manipulated throughout the experiment was defined by the radius of the sphere described above. The range of possible radii spanned [0, 0.3 cm], with the radius on a given trial specified according to an adaptive staircase procedure (see Section 3.7). The maximum magnitude of 0.3 cm was chosen based on pilot testing such that it would be unambiguously perceived as jittering by a typical observer.

3.4 Input & Output

User input and responses were recorded using a numberpad (Microsoft, Redmond, WA, USA) that was wirelessly paired to the HoloLens 2 via Bluetooth. To initiate a trial, participants pressed the 'Enter' key on the numberpad. After each pair of stimulus intervals was presented in a trial, participants indicated their response by pressing the '1' or '2' key. Participants were able to consistently maintain gaze on the target cube throughout each trial. After a few trials, participants typically placed their thumbs on each of the two response keys, and no longer needed to look down to select which key to press. We validated post hoc that participants were fixating on the cube for >90% of each trial in the stationary condition, using eye tracking data.

Eye tracking data was collected using the eye tracking system of the HoloLens 2 [82]. Before each session, an experimenter instructed each participant to complete the eye tracking calibration procedure built into the Windows operating system. We used the gaze direction output from the HoloLens 2 (vector on the unit sphere, centered at the cyclopean headset origin) and calculated the angular velocity at each timestep throughout the trial. This provided an estimated gaze position in world coordinates at a rate of 30 Hz.

Responses and eye tracking data were logged as .csv files on the HoloLens 2, and later transferred to a PC for analysis in MATLAB (MathWorks, Natick, MA, USA).

3.5 Experimental design

To manipulate the independent variable of participant self-motion, participants were instructed to move in three specific ways: to remain stationary ($\sim$0DoF), shake their head "no" (1DoF yaw rotation), and walk along a semicircular path (6DoF). For both the stationary and shaking conditions, participants sat in an office chair, whose seat surface was $\sim$45 cm above the floor. For the stationary condition, participants were instructed to remain as still as possible. For the shaking conditions, participants were instructed to rotate their heads back and forth to the beat of a metronome that played at 1.33 Hz, resulting in a nominal head rotation speed of 30°/s. This speed was chosen to be representative of typical speeds observed during VOR in natural viewing. For the walking condition, participants were

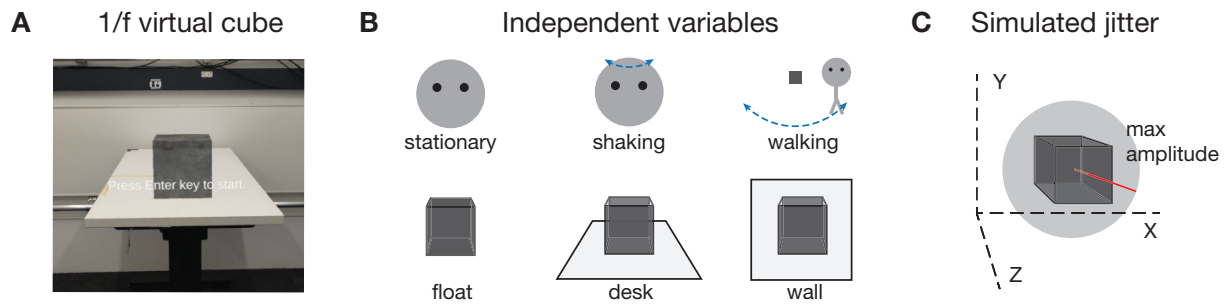

Figure 1: Stimulus and experimental design. (A) Example photograph of a participant's point-of-view photo during the experiment, showing both the physical world and the rendered virtual cube. The cube is textured with a 1/f 2D noise pattern on all sides (0.2 m side length). In this photograph, the cube is aligned horizontally with the surface of a physical desk, at a distance of 1 m from the wall of the physical room. (B) Array of stimulus conditions. There are 3 levels of the user motion factor: stationary, shaking one's head back and forth (yaw rotation), and walking along a semicircular path. There are 3 levels of the object placement factor: floating in space, aligned horizontally with the surface of a physical desk, and aligned horizontally with the surface of a physical wall. (C) Added jitter magnitude is defined by the radius of a 3D sphere, centered about the ground truth position of the cube, containing possible perturbed positions. A random point from a uniform distribution within the sphere is chosen at a rate of 15 Hz, and the cube's final rendered position is updated at a rate of 60 Hz using a spline interpolation between the randomly-chosen points.

instructed to walk along a marked path and coordinate each step to the beat of a metronome that played at 1 Hz, resulting in a nominal translational walking speed of 0.4 m/s. This speed was chosen to be slightly slower than natural walking speed, because it has been shown that performance on cognitive tasks decreases at faster walking speed [62]. In other words, we chose a relatively slow speed in order to potentially maximize participants' sensitivity to detecting jitter. The marked path that participants walked was semicircular with a 1.5 m radius, in order to keep the distance between the participant and the virtual cube (approximately) consistent.

To manipulate the independent variable of AR object placement, the virtual cube was placed into one of three different configurations: floating in free space, horizontally aligned on the surface of a desk, and vertically aligned against a wall. Participants were instructed to fixate on the virtual cube throughout each trial. Before each object placement block (see Section 3.5), an experimenter placed the virtual cube at the desired location in world coordinates using a custom Unity application which provided the ability to manually adjust the cube's position. In all conditions, the cube was placed 0.9 m above the floor and 1.5 m away from the participant (measured by distance along the floor). These values were chosen based on pilot testing to maximize participants' ability to fixate on the virtual cube throughout all 3 user motion conditions. In the float and desk conditions, the cube was placed a distance of 1 m away from the wall. For the wall condition, the cube was placed so that the farthest face was vertically aligned with the surface of the wall. For the desk condition, the cube was placed so that the bottom face was horizontally aligned with the surface of a physical desk. The physical desk was not present for the float or wall conditions.

The experiment was a full factorial design, with 2 factors (user motion and object placement), and 3 levels of each factor (for user motion: stationary, shaking, walking; for object placement: float, desk, wall), for a total of 9 conditions. Factors are illustrated in Figure 1B. The experiment was a blocked design with respect to user motion, with each participant completing all stationary conditions, followed by all shaking conditions, followed by all walking conditions. Within each user motion block, the 3 object placement conditions were randomized and counterbalanced. This blocked design also allowed participants to practice and standardize their behavior in each self-motion condition, reducing uncontrolled variability. However, this design does introduce the possibility of order effects across user motion conditions (see Section 5.1).

3.6 Protocol

Each session began with 5 practice trials at a very high jitter magnitude (1 cm) in order to demonstrate jitter artifacts to each participant. The intent of these practice trials was to ensure that participants were responding based on their perception of jitter, rather than other artifacts (such as chromatic aberrations, rainbows, etc.) that were not manipulated or controlled. Participants responded correctly on the majority (91%) of practice trials; incorrect responses can likely be attributed to common causes such as accidental keypresses, attentional lapses, and misapprehension or miscommunication of task instructions.

A two-interval forced choice (2IFC) psychophysical task was used to measure the magnitude of added jitter at which an observer could reliably distinguish between a virtual cube rendered to be stationary, and one with added positional jitter. During each trial of the experiment, participants were presented with two 1.0 s intervals, with a 0.5 s interstimulus interval. During one randomly selected interval, jitter was added to the cube's position. Participants indicated the interval in which they perceived added jitter by pressing a key on the Bluetooth keyboard.

3.7 Psychophysical parameters

The magnitude of added jitter on each trial varied using an adaptive staircase method. Staircase parameters were chosen based on best practice recommendations from Garcia-Perez and colleagues [24, 25], and refined by pilot testing. For each of the 9 conditions in the experiment, we used two interleaved staircases to estimate the threshold at which participants could reliably detect added jitter. One staircase started at a high jitter magnitude (0.3 cm), and one started at a lower jitter magnitude (0.1 cm). Each staircase began with a preliminary phase that used a 1-down, 1-up rule (step size = 0.1 cm for the high staircase, 0.01 cm for the low staircase), whereby one correct response would decrease jitter magnitude on the following trial, and one incorrect response would increase jitter magnitude. After the first trial in which participants made an incorrect response, a 3-down 1-up rule was used, whereby jitter magnitude was decreased after 3 successive correct responses, and was increased after 1 incorrect response. This rule converges toward an estimate of the 80% detection threshold [38]. While using the 3-down, 1-up rule, jitter magnitude was increased by 0.05 cm for the high staircase (0.0375 cm cm for the low), and decreased by 0.0375 cm for both staircases. Unequal step sizes have been shown to sample the psychometric function more efficiently than equal

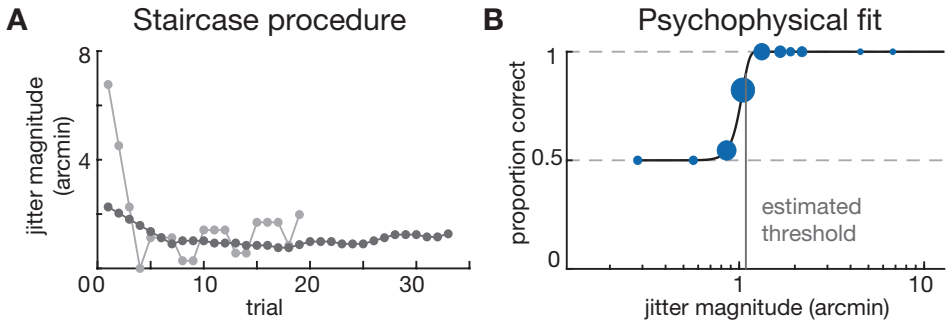

Figure 2: Estimating detection thresholds for added jitter. (A) Example psychophysical staircase from one participant during the shaking-desk condition. The high staircase (whose initial jitter magnitude is well above the expected threshold) is shown in light gray, and the low staircase (whose initial jitter magnitude is near or below the expected threshold) is shown in dark gray. (B) Example response data collected using the staircase procedure. Data is grouped by jitter magnitude. Larger circles indicate that more trials were completed for a given magnitude. The best-fit psychometric curve (for details, see Section 3.7) is shown in black. The gray vertical line indicates the estimated threshold for 80% correct performance.

ones [24]. Each staircase proceeded until one of two criteria were met: a maximum number of trials (60 per staircase), or a maximum of 8 reversals (a change from correct to incorrect responses, or vice versa), whichever came first. A sample staircase is shown in Figure 2A.

Advanced adaptive psychophysical methods, such as QUEST [89] and AEPsych [61], are able to more efficiently sample the stimulus space than the more conventional up-down rules used in this study. One difficulty in applying these methods is that they require additional real-time computation (to generate the stimulus magnitude to present on each trial, based on the pattern of past responses), which poses a challenge for the limited processing and thermal capacities of some current HMD technology. Future experiments could explore using these methods in order to generate more robust data, with fewer trials.

4 Analysis

Before beginning data analysis, jitter magnitudes were converted from metric units (meters) to angular units. Visual angle is the angle that an object subtends at an observer's eye, which also corresponds to its size on the retinal image, typically expressed in arcminutes (1/60°). It is expressed by $V = 2 \arctan(S/2D)$, where S is the object's size, and D is the distance between the object and the observer – in this case, between the virtual cube and the participant. We recorded this distance D continuously throughout the experiment, and computed the mean distance over each trial to serve as the denominator in the equation above when converting jitter amplitude from centimeters to arcminutes.

We fit a psychometric curve for each staircase on each trial, using psignifit [74], a publicly available MATLAB package. This software computes the best fit psychometric curve (Weibull function) given a set of psychophysical data using Bayesian inference to estimate the parameters in a beta-binomial model. A sample psychometric curve fit to a set of response data is shown in Figure 2B.

For a small number of staircases, this approach yielded a poor fit due to inadequate sampling of the stimulus space during one or both staircases. The staircase parameters (such as initial jitter magnitude, step size, up-down rule, etc.) were chosen based on best practices to efficiently target an 80% threshold (see Section 3.7 for details), but certain patterns of responses could result in inadequate sampling of the stimulus space. In particular, incorrect or highly variable responses at the beginning of a staircase could result in a poorly-fit psychometric function. For example, if a participant made an incorrect response on an early trial followed by mostly correct responses, the high and low staircases would descend or ascend too slowly, respectively, which did not allow them to converge towards the true threshold. We determined a psychometric curve to be poorly-fit if either or both of these criteria were met: (1) jitter magnitudes in the high staircase did not intersect with the low staircase (indicating that these staircases did not adequately sample jitter magnitudes near the participant's threshold), and (2) the estimated threshold was above most (>80%) of the tested jitter magnitudes (indicating that these staircases oversampled lower jitter magnitudes, which were far below the participant's threshold). Using these criteria, we removed a total of 20 sessions from across 12 participants (representing 11% of the 180 sessions that were analyzed). We also removed one participant's data entirely, because their estimated thresholds in 5 of the 9 conditions were more than 8 standard deviations above the mean, suggesting that they were unable to perform the task correctly.

In order to measure the effects of user motion and object placement on detection thresholds for added jitter, we used a linear mixed-effects regression model [65]. Mixed-effects regression is preferred for repeated measures experimental designs because it employs a model structure with parameters that control for variance within each participant's responses (random effects) while testing for the effects of independent variables across participants (fixed effects parameters). We included fixed effects parameters corresponding to planned contrasts comparing each level of our independent variables; namely, there were fixed effect parameters for the different user motion (stationary, shaking, walking) and object placement (float, desk, wall) conditions. For user motion, the two movement conditions (shake, walk) were compared to the stationary condition. For object placement, the desk and wall conditions were compared to the float condition. To control for individual differences in sensitivity across participants, we included a random effect variable. The complete model is specified as follows:

$$\begin{aligned}\text{Threshold} = \beta_{\text{intercept}} &+ \beta_{\text{shaking vs. stationary}} + \beta_{\text{desk vs. float}} \\ &+ \beta_{\text{wall vs. float}} + \beta_{\text{Participant (random effect)}}\end{aligned}$$

In a subsequent analysis, we analyzed eye tracking data to generalize beyond the discrete user motion conditions defined in the experiment. We calculated a proxy of retinal slip, a term used to describe motion on the retina caused by differences between eye velocity and target velocity during saccadic and smooth pursuit eye movements [15]. Given a known jitter magnitude (target velocity), higher or lower eye movement velocities should result in more or less retinal slip, respectively. We hypothesized that this would result in higher thresholds (lower sensitivity) at higher eye movement velocities. We used the change in gaze direction over time (measured in °/s) as an estimate

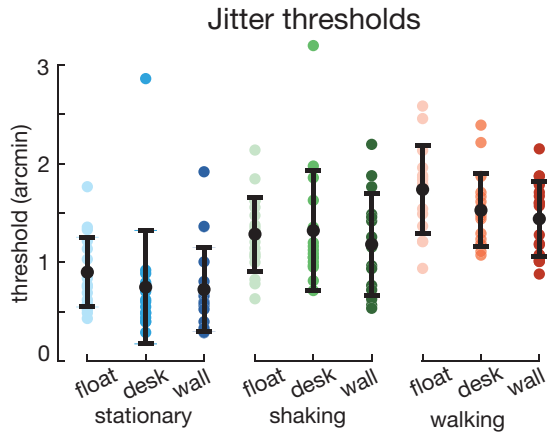

Figure 3: Summarizing the effects of user motion and object placement on jitter thresholds. Mean jitter thresholds across participants for each condition in the experiment; error bars represent standard deviation of the mean. Individual participants shown as dots (color-coded for user motion levels; shaded for object placement levels).

Parameter	Estimate (SE)	t value [DoF]	p value
intercept	0.89 (0.090)	9.87 [94]	<0.001*
shaking vs. stationary	0.47 (0.099)	4.74 [38]	<0.001*
walking vs. stationary	0.78 (0.098)	7.95 [38]	<0.001*
desk vs. float	-0.10 (0.068)	-1.49 [94]	0.14
wall vs. float	-0.20 (0.066)	-3.01 [94]	0.003*

Table 1: Results of linear-mixed effects regression analysis.

of eye movement velocity. For each trial in the experiment, we used the gaze vector output from the HMD to calculate mean gaze speed (°/s) over each trial. We took the mean of these time series data across all trials for each condition, and took the grand mean across each condition as a measure of retinal slip for that condition.

5 Results

5.1 Jitter thresholds are higher (sensitivity is lower) during user motion

For each condition in the experiment, we estimated each participant's threshold for jitter magnitude using the psychometric curve-fitting approach described in Section 4. The means across participants for each combination of user motion (stationary, shaking, walking), and object placement (float, desk, wall) factors are shown in Figure 3. Across all conditions in the experiment, mean detection thresholds for added jitter were in the range of 0.5 to 2.0 arcminutes.

To quantify the effects of user motion and object placement conditions on jitter thresholds, we fit a linear mixed-effects model as described in Section 3. Parameter estimates and p-values for the full model are reported in Table 1. Jitter thresholds were significantly higher (p<0.001) for the shaking (1.27 $\pm$ 0.11 arcmin; M $\pm$ SE) and walking (1.56 $\pm$ 0.09 arcmin) motion conditions compared to the stationary condition (0.79 $\pm$ 0.10 arcmin). Relative to the floating object placement condition (1.29 $\pm$ 0.11 arcmin), jitter thresholds were significantly lower (p<0.01) for the wall placement condition (1.11 $\pm$ 0.12 arcmin). They were also lower for the desk placement condition, but this difference was not significant (p>0.05).

5.2 Eye movements predict threshold increases

To generalize beyond the discrete user motion conditions defined in the experiment, we analyzed eye tracking data (as described in Section 4). Figure 4 shows the mean gaze speed as a function of the estimated threshold for each session. Across the three user motion conditions, there was an overall increase in gaze speeds (stationary > shaking > walking). There was also more variability in mean gaze speed for the walking condition (which ranged from ~6-15°/s) compared to the shaking and float conditions (which ranged from ~0-2 and 2-6°/s, respectively).

We quantified the relationship between eye movements and jitter detection thresholds by applying a linear model and computing Spearmans' correlation coefficient (ρ). There was a significant correlation (ρ = 0.6, p = 1.35e-15), indicating that jitter thresholds scale with gaze speed. The correlation coefficients for similar analyses applied to head rotation speed (ρ = .44, p=6.8e-09) and translational speed (ρ = .53, p=6.8e-13) were also significant, but lower than for gaze speed. These results suggest that the amount of retinal slip (as estimated using gaze speed) can predict jitter thresholds better than measured head movements or translational movements alone, suggesting a gaze-driven decrease in object motion sensitivity during self-motion.

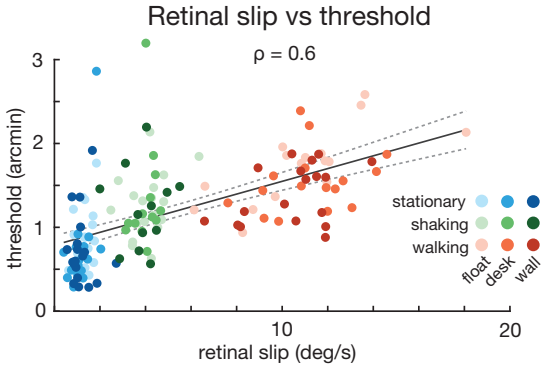

Figure 4: Retinal slip predicts jitter detection thresholds. Each data point represents a single session from one participant. Session type is indicated by color (for user motion levels) and shading (for object placement levels). The black line indicates the regression fit and the dotted black lines show the 95% confidence interval. There is a significant correlation between retinal slip (as measured by mean gaze speed in °/s) and jitter detection thresholds (ρ = 0.6, p<0.05).

6 Discussion

6.1 Jitter as a function of user self-motion and WL object placement

This study measured participants' ability to detect spatiotemporal noise (positional jitter) that was added to the 3D position of WL objects in AR, across experimental conditions specifying categories of user self-motion (stationary, shaking, walking) and the arrangement of virtual objects relative to surfaces in the physical world (floating, horizontally aligned with the surface of a desk, and vertically aligned against a wall).

We observed that jitter detection thresholds were roughly twice as high during the two user motion conditions (shaking and walking) compared to when participants were stationary. We hypothesize that the decrease in sensitivity when users are in motion may be due to imperfect retinal stabilization (retinal slip) of the target object, which we discuss in detail in the following section. Alternatively, we cannot rule out the possibility that the baseline jitter of the HoloLens 2 also varied across user motion conditions (and therefore contributed to decreased sensitivity), which we discuss along with other limitations in Section 7.

It is worth noting that the differences we observed between the user motion and stationary conditions runs in the opposite direction

than what would be expected based on perceptual learning due to order effects. Each participant saw all stationary, followed by all shaking, followed by all walking conditions. If participants became more sensitive over the course of the experiment, sensitivity – the opposite of what we observed. Future experiments could test for order effects directly using a fully counterbalanced design, as opposed a blocked design.

For the three object placement conditions (float, desk, wall), we hypothesized that viewing virtual objects in close proximity to stable references in the physical world (e.g. on a desk or against a wall) would provide more proximal veridical 3D cues that the visual system could use to compare with the motion of a jittering object. We found mixed results that provide partial support for this hypothesis. Jitter thresholds were lower (sensitivity was higher) for the wall placement condition compared to floating objects. The difference between the desk and float conditions trended in the same direction, but it was not statistically significant.

Some of the failure to observe a significant effect in the desk placement condition may be attributable to our choice of stimulus parameters, which in turn affected the presence and salience of 3D cues to object location. We did not render UI elements such as cast shadows (of the virtual cube on the physical surface of the desk), which may have enhanced the degree to which the physical world surfaces provided visual information that could facilitate detecting the (in)stability of virtual content.

Another explanation for the failure to observe a consistent effect for the object placement manipulation arises from variation in viewing angles across participants and object placement levels. Certain viewing angles are especially informative for judging the arrangement of 3D objects – so-called canonical views [7]. For example, to judge whether an object is coincident with the surface of a desk or slightly above it, the optimal viewing angle would arise when the point of observation is located at the same height, and parallel to, the surface of the desk. This reduces the task of judging the distance between the bottom of the object and the surface of the desk to one dimension (height). Other viewing angles, such as looking down on the object and desk from above, require a more complex visual transformation and resulting perceptual judgment [76]. For this reason, we expected thresholds to be lower in the stationary-desk condition compared to the stationary-wall condition. In our experiment, however, the virtual cube and desk were both placed 0.9 m above the ground, which was always below eye level for participants sitting in the chair used in those conditions. Participants also varied in height, resulting in differences in their viewing angle across object placement conditions. For example, due to their stature, taller participants had a steeper viewing angle in the desk condition, which may have made them less sensitive to positional errors coincident with the surface of the desk. These changes in the visual information that is available to observers in the different object placement conditions are difficult to fully eliminate in 6DoF motion, an example of the design tradeoffs between experimental control and naturalistic or ecological freedom of movement.

6.2 Jitter as a function of eye movements

We also found that jitter thresholds were significantly correlated with a measure of eye movement velocity (mean gaze speed); higher mean gaze speeds were associated with higher thresholds (and equivalently, lower sensitivity). These results indicate that participants were less sensitive to the motion caused by added jitter when they were moving than when they were stationary. We hypothesize that this is due to the additional retinal motion that arises during self-motion, and in particular retinal slip (retinal motion caused by differences between eye velocity and target velocity), which makes it more difficult for the visual system to isolate and detect the motion that arises from jitter in the AR object's 3D position.

Interestingly, this decrease in sensitivity to jitter during self-motion occurs despite mechanisms within the human visual system that are designed to stabilize the retinal image during self-motion [39,51].These mechanisms are especially potent during self-motion similar to the shaking condition in our experiment, where the vestibular-ocular reflex (VOR) helps stabilize the eyes to maintain fixation on a target object as the head rotates, which induces radial optic flow on the retina [42, 44]. During pure head yaw rotation, the human visual system is able to accurately compensate for head rotation by counterrotating the eyes (rotational VOR), and can maintain stable fixation quite well [1, 11]. Because of the nature of the task and the AR HMD used to present stimuli in our study, we observed additional sources of eye movement velocities that might be attributable to conditions of the experiment, suggesting that participants were not performing a pure rotational VOR to fixate a single fixed point throughout the entire trial. In the shaking condition, participants were asked to fixate on the virtual cube while moving their heads back and forth to the beat of a metronome. Due to the limited FOV of the AR HMD, the edge of the virtual cube would approach the boundary of the display FOV after $\sim$20° of head rotation in either direction. Participants were instructed to change direction when this occurred to keep the virtual cube visible (within the display FOV of the HMD) throughout each trial. Examining eye tracking trajectories, we find that participants shifted their gaze across the cube at specific points during the head rotational movement, to fixate on the edge of the cube nearest to the boundary of the FOV (when rotating the head left, on the left edge of the cube, and vice versa). Due to this constant scanning across the cube, we observed higher gaze speeds for the shaking condition compared to stationary. These gaze speeds were also higher than what we would expect if participants had been able to perfectly fixate a single point on the cube's surface (which should yield mean gaze speeds near zero). This indicates that, in addition to the stabilization of the retinal image afforded by VOR, participants made additional saccadic eye movements due to task and hardware constraints unique to our experiment. Future experiments could isolate VOR more narrowly by instructing participants to fixate at a single point, providing a fixation target on the surface of the object, using a display with a wider FOV and/or using a smaller virtual object (so that it remains within the FOV across larger head rotations). These experimental changes would allow for a more precise investigation into whether retinal slip can explain changes in jitter perceptibility, because retinal slip should be low during a well-conducted rotational VOR, without the additional retinal motion due to the saccadic eye movements that we observed across the surface of the cube.

Eye movement velocity changes the most during the walking condition, when the eyes have to compensate for not only rotational VOR but also due to translation of the head and eyes through space (translational VOR). In our experiment, as participants walked along a semicircular path, their eyes needed to make both more and larger saccades in order to maintain fixation on the cube to perform the detection task, while simultaneously collecting the visual information needed to plan their walking path and guide movement [51]. Thus, we expected mean gaze speeds to be highest and most variable in the walking condition, and by extension, for there to be the most retinal slip in this condition. This was for at least two reasons. First, because translational VOR is generally less accurate than rotational VOR [29, 66], so we expected more retinal slip and therefore larger mean gaze speeds. Second, because participants needed to make more, and larger, saccadic eye movements while walking (both to maintain fixation on the virtual object, and to plan and guide locomotor behavior during walking), this would further increase mean gaze speeds. As opposed to the shaking condition in our experiment, where much of the increase in gaze speed we observed can be attributed to idiosyncrasies of the experimental stimuli and apparatus rather than imprecision during rotational VOR, the increased gaze speed we observed in the walking condition in our experiment is

likely to generalize to all translational locomotor behavior, regardless of the specific stimuli or display. Generally speaking, we would expect more retinal slip (and therefore gaze speed to be higher) during user self-motion that is primarily translational (like walking) as opposed to primarily rotational (like head-shaking).

6.3 Comparison to other results

The absolute measurement of the jitter detection thresholds we observed were lower than those found in a previous study [93] that measured jitter perceptibility using similar stimuli and the HoloLens 2 HMD, and which examined the effects of different viewing distances and background luminances. This suggests that properties of the stimuli used in the current experiment may have made it somewhat easier for participants to detect jitter. Specifically, we suspect that differences in retinal extent (using a larger object, which resulted in a larger visual angle) and surface texture (using a 1/f noise pattern, rather than a uniform color) made it easier for the visual system to detect motion in the WL AR object used here. These could have led participants to make fewer and/or smaller eye movements (resulting in less retinal slip, which we observed leads to lower thresholds), as well as providing motion cues at higher spatial frequencies, which have been shown to improve motion detection [8].

7 Limitations

There are several factors that limit the generalizability of our results, and which could be improved in future work. Many of these arise because of limitations in the capabilities and precision of current 6DoF AR HMDs, which we suspect at least partially explains the dearth of perceptual data on sensitivity to AR WL rendering errors.

A key limitation arises because we defined jitter based on noise added to the estimated 3D position of the virtual cube, generated by the HMD and exposed to the 3D rendering engine, rather than perturbing the object's true 3D position in world coordinates. We manipulated added jitter as an independent variable, but what users see and are sensitive to is the combined positional jitter that is the sum of baseline jitter in the HMD (which affects its estimate of the object's position in world coordinates) plus the added jitter that is applied in software. While there have been investigations to quantify the spatiotemporal accuracy of the tracking systems of AR HMDs [83], it has also been observed that tracking accuracy in smartphone-based AR varies depending on user motion [72]. This suggests that baseline jitter for 6DoF HMDs may be different across different types and magnitudes of user motion. For example, which movement pattern results in higher baseline jitter: faster but regular head rotations (shaking condition) or slower translational movements (walking condition)? Careful measurement of baseline jitter would allow us to better distinguish between whether changes in observed threshold are due to changes in baseline jitter, or to the independent variable of added jitter. Further work that manipulates eye movements within motion conditions would help to further disambiguate the effects, but require even more precise baseline measurements (not only characterizing the 6DoF position of the HMD, but also of the user's eye relative to the display). Future work should seek to measure baseline jitter to the extent possible, using methods such as those proposed by [77].

A related limitation owing to the current state-of-the-art in AR HMDs involves the eye tracking data used in our analysis of retinal slip. The frequency (30 Hz) of estimated gaze position is too low to detect saccades (which may be as high as 700°/sec [3, 23]). This limitation essentially acts as a low-pass filter when computing gaze speed. Thus, mean gaze speeds may have systematically underestimated eye movement velocities for conditions where we expect more or larger saccades (i.e. shaking, walking), thereby causing a bias towards lower mean gaze speeds. Future work could address this limitation and refine the data collected here by using an independent eye tracker, and in particular one that may not be suitable for a consumer product (e.g. because it is heavy, physiologically invasive, or requires extensive calibration). Nevertheless, this low-pass filtered eye tracking data we analyzed here is indicative of the kind of output that will likely be generated in consumer devices, so this measure is still useful as a way of experimenting with dynamic variations in WL requirements that could potentially be implemented in AR devices.

A third limitation arises because we simulated jitter artifacts by manipulating the 3D position of the cube, without directly manipulating its 3D orientation. WL AR systems use 6DoF tracking, which specifies the pose of the HMD using 3DoF position (x/y/z) and 3DoF rotation (yaw/pitch/roll). Errors can arise both in the estimates of position (typically generated by computations on images captured by SLAM cameras, at low frequency) and orientation (typically provided by IMUs, at high frequency). Here we manipulated positional jitter, and did not manipulate rotational jitter, which has been shown to have additional negative consequences for AR interactions [5, 6].

Finally, there are valid concerns about whether our results will generalize beyond the specific stimulus parameters that were tested. For example, while we designed our stimuli to reflect the most common sources of error in 6DoF tracking, and chose parameter values (e.g. in the frequency of added jitter) that represent current tracking architectures and performance, one could ask whether and how the results would differ for other stimulus parameters (e.g. at a different frequency). There is a danger of a combinatorial explosion in the number of experimental conditions that would be required to fully investigate such a multidimensional space. There are two potential solutions to this problem. One approach would be to use adaptive psychophysical methods, such as QUEST [89] and AEPsych [61], to more efficiently sample multidimensional spaces of stimulus parameters. A second approach is to develop models that generalize beyond specific experimental conditions by positing an underlying mechanism, as we attempted to do here using retinal slip. Both are recommended for future work.

8 Conclusion

World-locked rendering is a core capability of augmented reality displays. But it is a challenging technical problem, requiring an array of sensors and demanding computations to solve it in real-time. These solutions must be implemented within the narrow budgets of size, weight, and power needed to build practical consumer devices in compact form factors. To that end, understanding human sensitivity to WL rendering errors, such as jitter, can help AR system architects design and optimize WL rendering pipelines.

In this study, we have presented quantitative data on human perceptual sensitivity to jitter in a 6DoF AR HMD. Importantly, we provide data collected during naturalistic locomotion (walking), as well as during more constrained self-motion (stationary, ∼0DoF; yaw rotation, 1DoF), which allows for comparison to previously published results. We found that participants were approximately twice as sensitive when they were moving (shaking their heads or walking) compared to when they were stationary. We also varied the placement of virtual objects relative to physical surfaces, and found that participants were significantly more sensitive when virtual objects were vertically aligned against the surface of a wall compared to when they were floating. Finally, we proposed a visual mechanism (retinal slip) that correlates reasonably well with sensitivity, generalizes beyond distinct user motion conditions, and could potentially be estimated in real-time using eye tracking.

In addition to direct applications of the data presented here, we hope to demonstrate the utility of applying the methods of experimental perception science to inform the design of AR/VR systems. By quantifying fundamental human capacities and limits, we can build devices that are tightly coupled to and highly optimized for the visual systems of end users.

Acknowledgments

The authors would like to thank John Pella, Joseph Zhang, Sara Kenley, and Helen Ayele for their support with software development and data collection, as well as Anqi Xu, Alex Chen, James Wilmott, Takahiro Doi, and Ian Erkelens for their invaluable feedback.

References

[1] D. E. Angelaki, A. G. Shaikh, A. M. Green, and J. D. Dickman. Neurons compute internal models of the physical laws of motion. *Nature*, 430(6999):560–564, 2004. 7

[2] A. Antoine, M. Nancel, E. Ge, J. Zheng, N. Zolghadr, and G. Casiez. Modeling and reducing spatial jitter caused by asynchronous input and output rates. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 869–881, 2020. 3

[3] R. W. Baloh, A. W. Sills, W. E. Kumley, and V. Honrubia. Quantitative measurement of saccade amplitude, duration, and velocity. *Neurology*, 25(11):1065–1065, 1975. 8

[4] A. U. Batmaz, M. R. Seraji, J. Kneifel, and W. Stuerzlinger. No jitter please: Effects of rotational and positional jitter on 3D mid-air interaction. In *Proceedings of the Future Technologies Conference*, vol. 2, pp. 792–808, 2020. 2

[5] A. U. Batmaz and W. Stuerzlinger. The effect of rotational jitter on 3D pointing tasks. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–6, 2019. 2, 8

[6] A. U. Batmaz and W. Stuerzlinger. Rotational and positional jitter in virtual reality interaction in everyday VR. In *Everyday Virtual and Augmented Reality*, pp. 89–118. 2023. 2, 8

[7] V. Blanz, M. J. Tarr, and H. H. Bülthoff. What object attributes determine canonical views? *Perception*, 28(5):575–599, 1999. 7

[8] J. C. Boulton and C. L. Baker. Motion detection is dependent on spatial frequency not size. *Vision Research*, 31(1):77–87, 1991. 8

[9] D. Burr, J. Holt, J. Johnstone, and J. Ross. Selective depression of motion sensitivity during saccades. *The Journal of Physiology*, 333(1):1–15, 1982. 2

[10] D. Burr and P. Thompson. Motion psychophysics: 1985-2010. *Vision Research*, 51(13):1431–1456, July 2011. 1

[11] H. Collewijn and J. B. Smeets. Early components of the human vestibulo-ocular response to head rotation: latency and gain. *Journal of Neurophysiology*, 84(1):376–389, 2000. 7

[12] E. A. Cooper. The perceptual science of augmented reality. *Annual Review of Vision Science*, 9, 2023. 3

[13] E. A. Cooper, M. van Ginkel, and B. Rokers. Sensitivity and bias in the discrimination of two-dimensional and three-dimensional motion direction. *Journal of Vision*, 16(10):5–5, 2016. 2

[14] S. Daly, N. Xu, J. Crenshaw, and V. J. Zunjarrao. A psychophysical study exploring judder using fundamental signals and complex imagery. *SMPTE Motion Imaging Journal*, 124(7):62–70, 2015. 3

[15] S. de Brouwer, M. Missal, and P. Lefèvre. Role of retinal slip in the prediction of target motion during smooth and saccadic pursuit. *Journal of Neurophysiology*, 86(2):550–558, 2001. 5

[16] B. De Bruyn and G. A. Orban. Human velocity and direction discrimination measured with random dot patterns. *Vision Research*, 28(12):1323–1335, 1988. 1

[17] J. Dunn. Jitter theory. *Audio Precision TECNOTE*, 23:1–23, 2000. 2

[18] H. Durrant-Whyte and T. Bailey. Simultaneous localization and mapping: Part I. *IEEE Robotics Automation Magazine*, 13(2):99–110, June 2006. Conference Name: IEEE Robotics Automation Magazine. 2

[19] A. S. Fernandes, T. S. Murdison, and M. J. Proulx. Leveling the playing field: A comparative reevaluation of unmodified eye tracking as an input and interaction modality for VR. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2269–2279, 2023. 2

[20] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4:2379–2394, 1987. 3

[21] P. W. Fink, P. S. Foo, and W. H. Warren. Catching fly balls in virtual reality: A critical test of the outfielder problem. *Journal of Vision*, 9(13):14–14, 2009. 1

[22] R. L. French and G. C. DeAngelis. Scene-relative object motion biases depth percepts. *Scientific Reports*, 12(1):18480, 2022. 2

[23] A. Fuchs. Saccadic and smooth pursuit eye movements in the monkey. *The Journal of Physiology*, 191(3):609, 1967. 8

[24] M. A. Garcia-Perez. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, 38(12):1861–1881, 1998. 4, 5

[25] M. A. Garcia-Perez. Optimal setups for forced-choice staircases with fixed step sizes. *Spatial Vision*, 13(4):431–448, 2000. 4

[26] K. R. Gegenfurtner. The interaction between vision and eye movements. *Perception*, 45(12):1333–1357, 2016. 2

[27] J. J. Gibson. *The Perception of the Visual World*. Houghton Mifflin, 1950. 2

[28] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin, 1979. 1

[29] G. E. Grossman, R. J. Leigh, L. A. Abel, D. J. Lanska, and S. E. Thurston. Frequency and velocity of rotational head perturbations during locomotion. *Experimental Brain Research*, 70(3):470–476, 1988. 7

[30] P. Guan, O. Mercier, M. Shvartsman, and D. Lanman. Perceptual requirements for eye-tracked distortion correction in VR. In *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–8, 2022. 2

[31] P. Guan, E. Penner, J. Hegland, B. Letham, and D. Lanman. Perceptual requirements for world-locked rendering in AR and VR. *arXiv preprint arXiv:2303.15666*. 2

[32] A. Ham, J. Lim, and S. Kim. Do we need a faster mouse? Empirical evaluation of asynchronicity-induced jitter. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 743–753, 2021. 2

[33] M. M. Hayhoe. Vision and action. *Annual Review of Vision Science*, 3:389–413, 2017. 1, 2

[34] G. Huang. Visual-inertial navigation: A concise review. In *2019 International Conference on Robotics and Automation*, pp. 9572–9582, 2019. 2

[35] International Telecommunications Union. Recommendation G.810: Definitions and terminology for synchronization networks. Technical report, Telecommunication Standardization Sector, 1996. 2

[36] Y. Ivanenko and V. S. Gurfinkel. Human postural control. *Frontiers in Neuroscience*, 12:171, 2018. 2

[37] R. J. Jagacinski and D. L. Monk. Fitts' law in two dimensions with hand and head movements. *Journal of Motor Behavior*, 17(1):77–95, 1985. 2

[38] C. Kaernbach. Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49(3):227–229, 1991. 4

[39] W. M. King and N. Shanidze. Anticipatory eye movements stabilize gaze during self-generated head movements. *Annals of the New York Academy of Sciences*, 1233(1):219–225, 2011. 7

[40] S. A. Klein and D. M. Levi. Hyperacuity thresholds of 1 sec: Theoretical predictions and empirical validation. *Journal of the Optical Society of America A*, 2(7):1170–1190, 1985. 1

[41] J. J. Koenderink. Optic flow. *Vision Research*, 26(1):161–179, 1986. 2

[42] H. G. Krapp and R. Hengstenberg. Estimation of self-motion by optic flow processing in single visual interneurons. *Nature*, 384(6608):463–466, 1996. 7

[43] M. F. Land. Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, 25(3):296–324, 2006. 2

[44] M. Lappe and K. P. Hoffmann. Optic flow and eye movements. *International Review of Neurobiology*, 44:29–47, 2000. 2, 7

[45] D. N. Lee. A theory of visual control of braking based on information about time-to-collision. *Perception*, 5(4):437–459, 1976. 2

[46] D. N. Lee. The optic flow field: The foundation of vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 290(1038):169–179, 1980. 2

[47] D. N. Lee and P. E. Reddish. Plummeting gannets: A paradigm of ecological optics. *Nature*, 293(5830):293–294, 1981. 2

[48] G. E. Legge and F. W. Campbell. Displacement detection in human vision. *Vision Research*, 21:205–213, 1981. 1

[49] T. Louis, J. Troccaz, A. Rochet-Capellan, and F. Bérard. Is it real? Measuring the effect of resolution, latency, frame rate and jitter on the presence of virtual entities. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, pp. 5–16, 2019. 2

[50] P. R. MacNeilage, Z. Zhang, G. C. DeAngelis, and D. E. Angelaki. Vestibular Facilitation of Optic Flow Parsing. *PLOS ONE*, 7(7):e40264, 2012. 2

[51] J. S. Matthis, K. S. Muller, K. L. Bonnen, and M. M. Hayhoe. Retinal optic flow during natural locomotion. *PLOS Computational Biology*, 18(2):e1009575, 2022. 1, 7

[52] M. K. McBeath, D. M. Shaffer, and M. K. Kaiser. How baseball outfielders determine where to run to catch fly balls. *Science*, 268(5210):569–573, 1995. 1

[53] S. P. McKee, L. Welch, D. G. Taylor, and S. F. Bowne. Finding the common bond: Stereoacuity and the other hyperacuities. *Vision Research*, 30(6):879–891, 1990. 1

[54] D. Milner and M. Goodale. *The Visual Brain in Action, 2nd Edition*. Oxford University Press, 2006. 1

[55] M. C. Morrone, M. Tosetti, D. Montanaro, A. Fiorentini, G. Cioni, and D. C. Burr. A cortical area that responds specifically to optic flow, revealed by fMRI. *Nature Neuroscience*, 3(12):1322–1328, 2000. 2

[56] M. H. Mughrabi, A. K. Mutasim, W. Stuerzlinger, and A. U. Batmaz. My eyes hurt: Effects of jitter in 3D gaze tracking. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 310–315, 2022. 2

[57] T. S. Murdison, G. Blohm, and F. Bremmer. Saccade-induced changes in ocular torsion reveal predictive orientation perception. *Journal of Vision*, 19(11):10–10, 2019. 2

[58] T. S. Murdison, D. I. Standage, P. Lefèvre, and G. Blohm. Effector-dependent stochastic reference frame transformations alter decision-making. *Journal of Vision*, 22(8):1–1, 2022. 2

[59] K. Nakayama and C. W. Tyler. Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Research*, 21:427–433, 1981. 1

[60] V. Ngo, J. C. Gorman, M. F. De la Fuente, A. Souto, N. Schiel, and C. T. Miller. Active vision during prey capture in wild marmoset monkeys. *Current Biology*, 32(15):3423–3428, 2022. 1

[61] L. Owen, J. Browder, B. Letham, G. Stocek, C. Tymms, and M. Shvartsman. Adaptive nonparametric psychophysics. *arXiv preprint arXiv:2104.09549*. 5, 8

[62] P. Patel, M. Lamar, and T. Bhatt. Effect of type of cognitive task and walking speed on cognitive-motor interference during dual-task walking. *Neuroscience*, 260:140–148, 2014. 4

[63] A. Pavlovych and W. Stuerzlinger. The tradeoff between spatial jitter and latency in pointing tasks. In *Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, pp. 187–196, 2009. 2

[64] N. E. Peltier, D. E. Angelaki, and G. C. DeAngelis. Optic flow parsing in the macaque monkey. *Journal of Vision*, 20(10):8, 2020. 2

[65] J. C. Pinheiro and D. M. Bates. Unconstrained parametrizations for variance-covariance matrices. *Statistics and Computing*, 6(3):289–296, 1996. 5

[66] T. Pozzo, A. Berthoz, and L. Lefort. Head stabilization during various locomotor tasks in humans. *Experimental Brain Research*, 82(1):97–106, Aug. 1990. 7

[67] M. Riley, R. Balasubramaniam, and M. Turvey. Recurrence quantification analysis of postural fluctuations. *Gait & Posture*, 9(1):65–78, 1999. 2

[68] A. Roorda and D. R. Williams. The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520–522, 1999. 1

[69] E. A. Rossi and A. Roorda. The relationship between visual resolution and cone spacing in the human fovea. *Nature Neuroscience*, 13(2):156–157, 2010. 1

[70] D. L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6):814–817, 1994. 3

[71] S. K. Rushton, M. F. Bradshaw, and P. A. Warren. The pop out of scene-relative object movement against retinal motion due to self-movement. *Cognition*, 105(1):237–245, 2007. 2

[72] T. Scargill, J. Chen, and M. Gorlatova. Here to stay: Measuring hologram stability in markerless smartphone augmented reality. *arXiv preprint arXiv:2109.14757*, 2021. 8

[73] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers. The TUM VI benchmark for evaluating visual-inertial odometry. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1680–1687, 2018. 3

[74] H. Schütt, S. Harmeling, J. Macke, and F. Wichmann. Psignifit 4: Pain-free Bayesian inference for psychometric functions. *Journal of Vision*, 15(12):474–474, 2015. 5

[75] D. Sharafutdinov, M. Griguletskii, P. Kopanev, M. Kurenkov, G. Ferrer, A. Burkov, A. Gonnochenko, and D. Tsetserukou. Comparison of modern open-source visual SLAM approaches. *Journal of Intelligent & Robotic Systems*, 107(3):43, 2023. 3

[76] R. N. Shepard and J. Metzler. Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703, 1971. 7

[77] J. Simonen, T. Björk, T. Nikula, and K. Ryynänen. Measuring world-locking accuracy in AR/MR head-mounted displays. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) II*, vol. 11765, pp. 201–206, 2021. 8

[78] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Effects of latency jitter on simulator sickness in a search task. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 121–127, 2018. 2

[79] R. J. Teather, A. Pavlovych, W. Stuerzlinger, and I. S. MacKenzie. Effects of tracking technology, latency, and spatial jitter on object movement. In *2009 IEEE Symposium on 3D User Interfaces*, pp. 43–50, 2009. 2

[80] R. J. Teather and W. Stuerzlinger. Pointing at 3D targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 87–94, 2011. 2

[81] C. W. Tyler and J. Torres. Frequency response characteristics for sinusoidal movement in the fovea and periphery. *Perception & Psychophysics*, 12(2):232–236, 1972. 1

[82] D. Ungureanu, F. Bogo, S. Galliani, P. Sama, X. Duan, C. Meekhof, J. Stühmer, T. J. Cashman, B. Tekin, J. L. Schönberger, P. Olszta, and M. Pollefeys. HoloLens 2 Research Mode as a Tool for Computer Vision Research. *arXiv preprint arXiv.2008.11239*, 2020. 3

[83] R. Vassallo, A. Rankin, E. C. Chen, and T. M. Peters. Hologram stability evaluation for Microsoft Hololens. In *Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment*, vol. 10136, pp. 295–300, 2017. 8

[84] P. A. Warren and S. K. Rushton. Optic flow processing for the assessment of object movement during ego movement. *Current Biology*, 19(18):1555–1560, 2009. 2

[85] W. H. Warren and D. J. Hannon. Eye movements and optical flow. *Journal of the Optical Society of America A*, 7(1):160–169, 1990. 2

[86] W. H. Warren, B. A. Kay, W. D. Zosh, A. P. Duchon, and S. Sahuc. Optic flow is used to control human walking. *Nature Neuroscience*, 4(2):213–216, 2001. 2

[87] W. H. Warren Jr and J. A. Saunders. Perceiving heading in the presence of moving objects. *Perception*, 24(3):315–331, 1995. 2

[88] S. N. J. Watamaniuk and R. Sekuler. Temporal and spatial integration in dynamic random-dot stimuli. *Vision Research*, 32(12):2341–2347, 1992. 1

[89] A. B. Watson and D. G. Pelli. Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2):113–120, 1983. 5, 8

[90] Y. Wei, R. Shi, D. Yu, Y. Wang, Y. Li, L. Yu, and H.-N. Liang. Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2023. 2

[91] G. Westheimer. Editorial: Visual acuity and hyperacuity. *Investigative Ophthalmology*, 14:570–572, 1975. 1

[92] G. Westheimer. The spatial grain of the perifoveal visual field. *Vision Research*, 22:157–162, 1982. 1

[93] J. P. Wilmott, I. M. Erkelens, T. S. Murdison, and K. W. Rio. Perceptibility of jitter in augmented reality head-mounted displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality*, pp. 470–478, 2022. 2, 8

[94] J.-R. Wu and M. Ouhyoung. On latency compensation and its effects on head-motion trajectories in virtual environments. *The Visual Computer*, 16:79–90, 2000. 2

[95] X. Xing and J. A. Saunders. Perception of object motion during self-motion: Correlated biases in judgments of heading direction and object motion. *Journal of Vision*, 22(11):8–8, 2022. 2