Perceptibility of Jitter in Augmented Reality Head-Mounted Displays

James P. Wilmott* Ian M. Erkelens† T. Scott Murdison‡ Kevin W. Rio§
Meta Meta Meta Meta

ABSTRACT

When using a see-through augmented reality head-mounted display system (AR HMD), a user's perception of virtual content may be degraded by a variety of perceptual artifacts resulting from the architecture of rendering and display pipelines. In particular, virtual content that is rendered to appear stationary in the real world (world-locked) can be susceptible to spatial and temporal 3D position errors. A subset of these errors, termed jitter, result from mismatches between the spatial localization, rendering, and display pipelines, and can manifest as perceived motion of intended-to-be stationary content. Here, we employ psychophysical methods to quantify the perceptibility of jitter artifacts in an AR HMD. For some viewing conditions, participants perceived jitter that was smaller than the pixel pitch of the testbed (i.e., subpixel jitter). In general, we found that jitter perceptibility increased as viewing distance increased and decreased as background luminance increased. We did not find that the contrast ratio of virtual content, age, or experience with AR/VR modulated jitter perceptibility. Taken together, this study quantifies the degree of jitter that a user can perceive in an AR HMD and demonstrates that it is critical to consider the capabilities and limits of the human visual system when designing the next generation of spatial computing platforms.

Index Terms: Human-centered computing—Human-computer interaction (HCI)—HCI design and evaluation methods—Laboratory experiments; Human-centered computing—Human-computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality

1 INTRODUCTION

The recent advent of head-worn, spatial computing platforms like virtual reality (VR) and AR necessitates the development of systems that produce convincing and comfortable extensions of a user's perceptual experience. In particular, building AR HMD systems requires solving many unique engineering challenges associated with implementing high performance computing in a wearable form factor [2, 19]. An important differentiator for AR compared to VR is the ability to successfully render virtual content to appear as if it is co-located with physical objects in the real world, referred to as world-locking (WL). WL virtual content in AR serves a central role in constructing a metaverse, a hybrid virtual/physical world that is persistent across time and shared among interconnected users.

A key component of a successful WL rendering pipeline is the use of algorithms that generate an estimate of the user's 6 Degree of Freedom (6 DoF) position and orientation (pose) within a global world coordinate system, which is then used to update the rendered location of virtual content to maintain a convincing experience of WL. While making accurate 6 DoF pose estimates is a nontrival problem, Simultaneous Localization and Mapping (SLAM) algorithms originally developed for robot navigation can successfully be used to (approximately) solve it [6, 9]. However, a number of estimation errors can occur based on the architecture of the system used to implement the WL algorithm. For example, the time required to render and display each frame requires online algorithms to make predictions about the user's 6 DoF head pose in real time based on the previously estimated pose. This latency can be relatively large - often exceeding tens of milliseconds - and is particularly a problem when position changes rapidly (e.g., during head movements). Though predictive algorithms attempt to compensate for this latency [6, 31, 42], user movements are noisy and nonlinear, resulting in some magnitude of practically unavoidable spatial rendering errors. Additionally, the hardware and software used in AR SLAM systems is susceptible to noise due to a variety of factors, for example differences in the timing of IMU and camera sampling rates, or erroneous estimates of the 3D geometry of the world (which the user's position is estimated relative to). These and other factors have the potential to induce a variety of spatiotemporal WL artifacts that may significantly detract from a user's experience if they are readily perceptible, affect image quality, and/or lead to discomfort.

In the current study, we measure the perceptibility of a specific artifact known to result from the use of 6DoF pose estimation algorithms consisting of high frequency 3D position errors, termed jitter. An observer experiencing jitter will see virtual content that appears to randomly 'jump' or 'vibrate' from one position to another around a center of mass at or near the ground truth rendering position, i.e., where the content should be rendered if WL position and orientation were estimated perfectly. Jitter artifacts have the capacity to significantly hinder a user's experience by increasing the difficulty of interaction with virtual content [3] and causing a user to experience discomfort or (in extreme cases, for blocked-light displays) motion sickness [38, 39]. Moreover, it is possible that certain negative effects of jitter may be exacerbated in AR HMDs (compared to non-AR displays) because the user can simultaneously see both jittering virtual content and the non-jittering (stable) physical world.

While some early studies have been conducted to characterize the magnitude of jitter resulting from the use of different prediction algorithms (e.g., [31]), there are currently no publicly available measurements of how jitter impacts a user's experience while using an AR HMD. These are important measurements that can help constrain the engineering of high-quality AR systems. However, before quantifying how jitter impacts the overall user experience (which is a function of a multitude of factors), it is crucial to understand the visual system's fundamental sensitivity to jitter. Put another way, one cannot begin to properly account for jitter artifacts without understanding the magnitude of jitter that users can perceive in the first place. Here, we employ psychophysical methods to quantify jitter perceptibility in AR HMDs. These measurements provide a generalizable standard that can be used to inform the design and engineering of WL rendering AR systems.

2 BACKGROUND

2.1 Impacts of jitter on user experience in AR and physical displays

While there are few studies related to jitter perceptibility in AR systems, recent work has produced evidence that jitter negatively impacts user experience in AR. Louis and colleagues [24] observed that AR content rendered with jitter magnitudes of >0.5 mm de-

---

*e-mail: jwilmott@fb.com
†e-mail: ian.erkelens@fb.com
‡e-mail: smurdison@fb.com
§e-mail: kevinwrio@fb.com

tracted from a user's feelings of presence (operationally defined by the experimenters as the perceived realness of content). However, the AR content in this study was displayed to observers via image projection onto a spherical physical screen coupled with shutter glasses, leaving it unclear whether these results generalize to WL content viewed through an HMD.

More generally, studies have been conducted to understand how latency (delay in time from input to update) and jitter impact a user's input capabilities in a physical 2D computer monitor setup. Increasing latency reduces movement accuracy and increases movement times for tasks requiring movement of a computer mouse to intercept a target [27]. Similarly, jitter degrades user performance when completing 3D pointing movements to touch a physical monitor [33, 41] and in VR [3]. Jitter is also a concern in the design and engineering of modern touch screen devices like smartphones, where the screen position of content is determined by the user's touch actions (e.g., scrolling and swiping). Touch action comfort and accuracy are significantly reduced when content jitters [1, 29].

Taken together, the existing literature indicates that jitter has the potential to significantly harm user experience in most existing computing platforms.

2.2 Factors that may modulate jitter perceptibility in AR HMDs

Visual perception of jitter is the result of a complex set of motion-processing mechanisms in the human visual system. Each feature of a moving object (e.g., direction and speed) elicits unique spatiotemporal patterns of neural activity that are decoded in higher levels of visual processing to yield an observer's perception of motion (for reviews, see [5, 30]). Given that vision-related neural activity is a function of the patterns of photoreceptor stimulation on the retina, any variables that modulate the retinal image have the capacity to subsequently impact the brain's ability to detect and estimate object motion. Such factors include locomotion (due to optic flow), eccentricity of objects in the observer's field of view (more central vs. peripherally located), and the spatial frequency content of objects (e.g., text has greater energy at high frequencies, smooth shapes have more energy at lower frequencies).

Here, we focus on quantifying how three variables that are especially relevant in the context of displaying WL content in an AR HMD impact jitter perceptibility: viewing distance of content in depth, background luminance, and contrast ratio. Viewing distance is important to consider because WL rendering engines must be able to present objects at various locations and it is useful to understand how perceptibility differs at close versus far distances. We presented virtual content at three distances corresponding to close (1 m [1 D]), intermediate (2 m [0.5 D]) and far (5 m [0.2 D]) locations in depth. Background luminance is a function of the light sources surrounding a user and will thus vary dramatically across the variety of real world situations in which AR HMDs may be used. Here, we test three levels of background luminance that approximate dark (10 Cd/m^2), moderately bright (60 Cd/m^2) and bright (100 Cd/m^2) indoor use cases. Finally, contrast ratio is an important consideration for HMDs because display luminance is often constrained by form factor requirements, panel or projector limitations, power budgets, and/or thermal tolerances. The (additive) contrast ratio of content in see-through AR may be defined as ($L_{\text{virtual content}} + L_{\text{real world}}$) : $L_{\text{real world}}$ where L is luminance in Cd/m^2. We selected contrast ratios ranging from 1.9:1 to 9.5:1 as these are realizable in current commercially-available AR HMDs, including the testbed used in this study (the Microsoft HoloLens 2).

2.3 Psychophysical methods for measuring jitter perceptibility

Here, we operationally define jitter perceptibility as the ability of an observer to reliably discriminate jittering from non-jittering virtual content. How does one accurately measure perceptual discriminability? Behavioral methods that rely on self-report, such as survey responses, are inappropriate because they are susceptible to biases in the user's ability to accurately assess and communicate their perceptual experience. A class of methods has been developed over the last century to quantify perception in a manner that reduces the potential for such biases to contaminate perceptual measurements, termed psychophysics. Below, we develop the logic of the psychophysical approach using the experimental design we implement in this study as an example.

To obtain a quantitative assessment of the observer's ability to discriminate jittering from non-jittering virtual content, we seek a measurement of the minimum magnitude of jitter required for an observer to reliably report moving content as in motion. Consistent with previous literature, we use the term threshold to describe this value. Any amount of jitter below threshold is indiscriminable from non-jittering content, while any suprathreshold jitter is readily perceptible and may negatively impact user experience. Central to the psychophysical approach is the notion that perception may be described as a decision-making process, where an observer uses the visual information available in the retinal image to decide when a particular signal (e.g., jitter) is present [17, 26, 28]. We thus design an experiment that requires observers to view two examples of virtual content, one rendered to jitter and the other rendered as stationary, across two successive intervals and report which interval contained the object that jittered. Importantly, everything about the content presented in each interval is the same except for the presence or absence of jitter. In such 2 interval forced choice (2IFC) tasks, the observer makes discrete perceptual decisions based solely on the physical quantity that is under control of the experimenter (in our case jitter), thereby reducing the possibility of measurement contamination by non-perceptual factors. When the magnitude of jitter is large, the observer will correctly report the jittering interval with ease. If the magnitude of jitter is systematically decreased, at some point the observer will be unable to perceive a difference between the two intervals and will perform the task at chance level across repeated presentations of the same stimulus. It has been shown that the psychometric function relating the intensity of a physical variable to the probability of correct report ranges between 50%-100% correct and takes on a characteristic 'S' shape that is often modelled as a Weibull or cumulative Gaussian function [28]. The threshold determining when jittering content is reliably discriminable from stationary content is often approximated as the 80% point of the psychometric function [12, 13, 28]. In this study, we implement a standard psychophysical technique termed a staircase to iteratively adjust the magnitude of jitter to obtain measurements of observer's discriminability around this 80% threshold [12, 13, 20, 22]. We then use the pattern of responses to estimate thresholds separately for different combinations of the independent variables listed above. Note that threshold and perceptibility are inversely related: a larger threshold means jitter is less perceptible in that viewing condition, while a smaller threshold indicates jitter is more perceptible.

3 METHODS

The current study characterized observers' perception of 3D WL jitter artifacts. We employed a psychophysical staircase procedure to measure the magnitude of added jitter needed to reliably discriminate jittering from (rendered-to-be) stationary WL virtual content. Observers viewed WL virtual content on a commercially available AR HMD (Microsoft HoloLens 2). 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 jitter magnitudes was used to estimate the observer's threshold. We manipulated the following factors to determine how jitter perceptibility varies with viewing conditions: viewing distance of the virtual content, real world luminance, and contrast ratio.

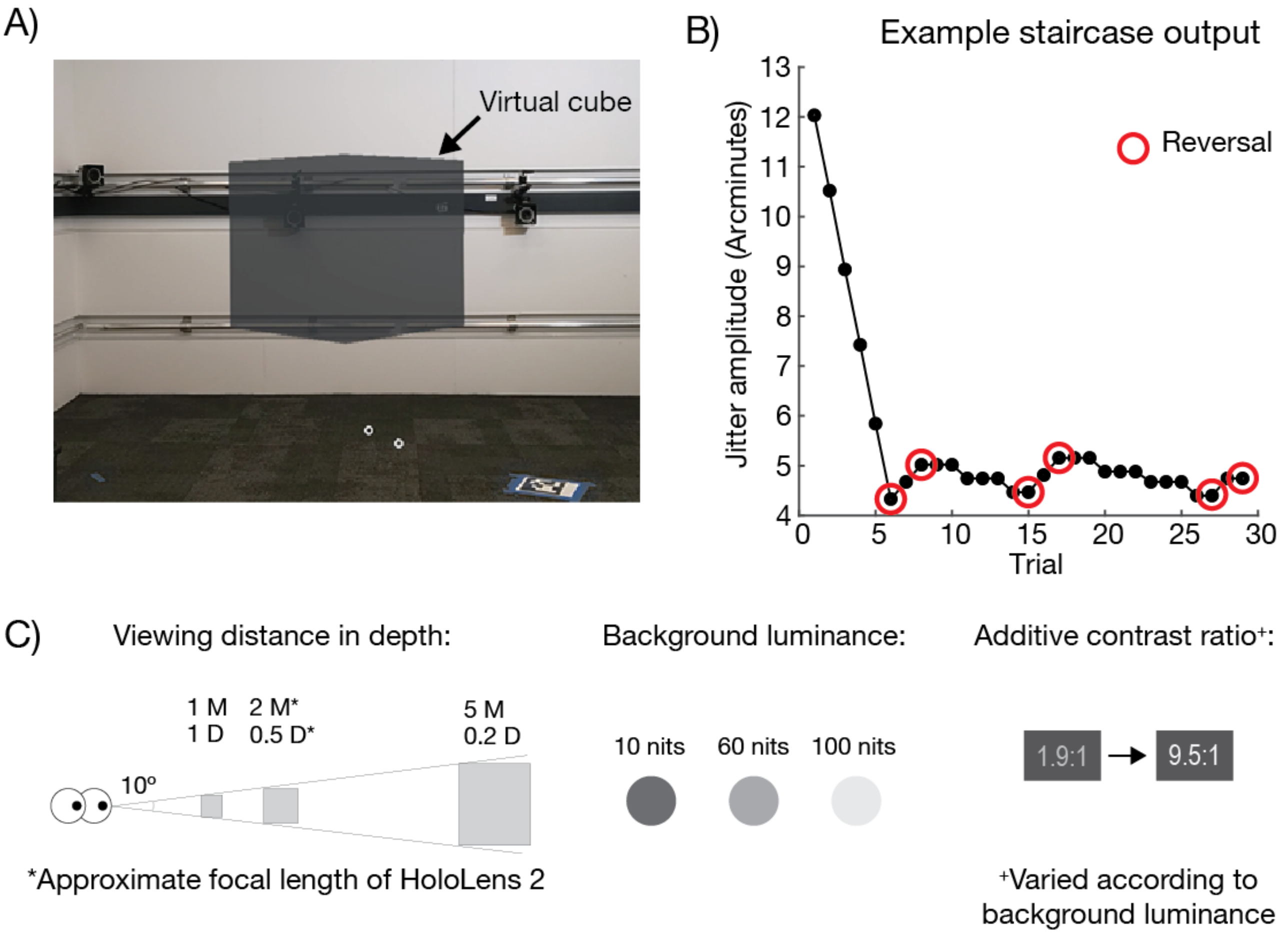

Figure 1: Experimental method. A) Example rendering of the experimental stimulus. B) Example output of a psychophysical staircase used in the experiment. The abscissae depict trial number while jitter magnitude in arcminutes is plotted on the ordinate. Reversals (trial where response correctness differs from the previous trial) are highlighted with red circles. C) Independent variables that were manipulated in the experiment.

3.1 Participants

Ten participants completed the study. Nine participants completed a post-study survey collecting demographic information. Of these respondents, 7 identified as male while 2 identified as female; the average age was 34.9 years (min: 26, max: 53); and the average number of years using AR/VR was 4.4 (min: 0.33, max: 12).

3.2 Apparatus and stimuli

Participants were seated in a large open room facing a white wall mounted with horizontal grey rails located 5.5 m away. Virtual content was displayed on a Microsoft HoloLens 2 (Redmond, WA, USA) running a custom application built in Unity (San Francisco, CA, USA). User input and responses were recorded using a wireless keyboard that was paired to the HoloLens via Bluetooth.

Stimuli were grayscale 3D cubes presented at the center of field of view in the left-right and up-down axes. The diagonal of each side of the cube subtended 10° of visual angle (note that while retinal size remained fixed at 10°, metric size varied with viewing distance). Each cube was rotated 45° about the plane perpendicular to the line of sight so that one edge pointed toward the participant. A point of view depiction of an example cube is displayed in Figure 1A.

On each trial, we presented two cubes which varied according to the magnitude of added 3D jitter. One cube (the reference) was always rendered with no added 3D jitter so that the only potential detectable jitter was that which resulted from the HoloLens 2 tracking system (see Section 6: Limitations below for discussion). The other cube (the target) was rendered to rhythmically vary in its 3D position using a spatiotemporal model of jitter. The specifications of the jitter model are as follows: 1) 3D position was varied according to a 10 Hz frequency oscillation, 2) the [X,Y,Z] jitter directions were simultaneously manipulated, resulting in a 3D ‘sphere’ of potential jitter directions, 3) a unique angular direction was chosen for each period of the oscillation from a uniform distribution of potential spherical angles, and 4) each period’s amplitude was randomly chosen from a uniform distribution in the interval [0, maximum trial amplitude].The maximum trial amplitude was determined by the staircase (see Section 3.5: Psychophysical staircase below for details). Across all participants and all trials, the range of maximum trial amplitudes used was [0, 60] arcminutes.

3.3 Experimental protocol

Participants completed 3 experimental sessions which varied according to the background luminance of the wall. Participants completed a standard 2IFC task. To begin a trial, participants pressed the spacebar, after which the reference and target were presented sequentially in the following sequence: the first stimulus was presented for 2000 ms, a blank screen was presented for 1000 ms, then the second stimulus was presented for 2000 ms. The order of presentation (i.e., whether the reference or target appeared first) was randomized on each trial. After the final stimulus, participants reported whether the first or second cube appeared to jitter more.

We manipulated three independent variables to determine how jitter perception varies as a function of viewing conditions (Figure 1C): viewing distance of the rendered object in depth, background luminance, and the contrast ratio of virtual content. Viewing distance was rendered to be 1 m (1D), 2 m (0.5 D; the approximate focal length of the HoloLens 2; [8]), or 5 m (0.2 D). Background luminance corresponded to the luminance at the observer's eye (i.e., after any dimming optics in the HoloLens 2) and varied between 10 Cd/m^2, 60 Cd/m^2, and 100 Cd/m^2 (approximating dark, moderately bright, and bright indoor luminance levels). Contrast ratio was calculated as ($L_{\text{virtual content}} + L_{\text{real world}}$):$L_{\text{real world}}$, where L is luminance in Cd/m^2. The minimum and maximum contrast ratios that can be presented on the HoloLens 2 depend on the device's display capabilities relative to the background luminance (see Section 3.4: HoloLens 2 display luminance measurement for details). Given these constraints, we were not able to present equated contrast ratios at each background luminance. We presented the following contrast ratios at each background luminance: at 10 Cd/m^2 background luminance, 3.1:1, 6.2:1, and 9.5:1; at 60 Cd/m^2 background luminance, 1.9:1, 3.1:1, and 4.1:1; at 100 Cd/m^2 background luminance, 1.9:1, 2.3:1, and 2.9:1. Background luminance was varied between sessions and was validated using a photometer (SpectraScan PR-788, Photo Research, New Syracuse NY, USA) at the beginning of each session. Viewing distance and contrast ratio were varied within sessions and the ordering of presentation for each combination of variables was randomized; background luminance was varied between sessions, which participants completed on different days. Participants completed one staircase for each combination of independent variables, resulting in 9 staircases per session (3 viewing distances x 3 contrast ratios) and a total of 27 staircases (9 x 3 background luminances) for each participant that completed all 3 sessions. The number of trials in each staircase was determined by the staircase procedure (see Section 3.5: Psychophysical staircase); namely, each staircase was terminated when participants reached a criterion for the maximum number of reversals (6) OR the maximum number of trials (60), whichever came first (see Section 3.5: Psychophysical staircase). Thus, the maximum number of trials 540 (9 staircases x 60 trials) per session and 1620 (540 x 3 sessions) per experiment, but most participants completed fewer trials (because they reached the criterion for maximum reversals before maximum trials on at least some staircases). After completion of the final session, each participant was asked to complete a pencil-and-paper survey demographic information (e.g., age and years of experience with AR/VR).

3.4 HoloLens 2 display luminance measurement

To precisely determine the display luminance required to present the intended contrast ratios, we measured HoloLens 2 gamma curves for each background luminance level used in the study. Measurements took place inside an enclosed cabinet space with blackout curtains to ensure no light contamination. An adjustable overhead lighting source was used to approximate each background luminance level. A white background target (approximately 15 cm x 15 cm) was used to model the background of the experimental apparatus. Measurements were conducted using a ProMetric I29 colorimeter (Radiant Vision Systems, Redmond, WA, USA) with a custom folded lens. A custom Unity application presented a grayscale 3D cube and systematically varied luminance ranging from the minimum to the maximum capabilities of the display. The resulting gamma curves were used to determine the brightness settings used in the subsequent experiment. The specific HoloLens 2 device that was used for measurement was also used in the subsequent study to ensure the settings used in the experiment were accurate.

3.5 Psychophysical staircase

We adjusted the magnitude of jitter added to the target using a psychophysical staircase procedure to obtain estimates of threshold (see Section 2.3 Psychophysical methods for measuring jitter perceptibility for overview of threshold; [12, 13, 20, 22]). The participant's pattern of responses about which object appeared to jitter is used to adaptively adjust the jitter magnitude on future trials. The logic of a staircase is as follows. When the difference in jitter magnitude between the target and reference is large, the participant will be consistently correct at discriminating which cube was rendered to jitter. In these instances, we should reduce the difference in jitter magnitude by decreasing the magnitude of jitter added to the target to present values that are close to an observer's threshold. When the difference in jitter magnitude between target and reference is sufficiently small, the participant will be at a chance level of performance, and we should increase the magnitude of jitter added to the target. Across trials, the jitter magnitude returned from the staircase dynamically de/increases. An example of a series of jitter values returned from a staircase used in our experiment is depicted in Figure 1B. We define a reversal as a trial in which the correctness of a participant's response differed from the previous trial (i.e., correct on trial *n* and incorrect on trial *n*-1 or vice versa); these are depicted with red circles in Figure 1B. A staircase proceeds until either a certain number of reversals has been achieved or a maximum number of trials has passed. We used independent staircases for each participant and each combination of viewing distance, background luminance, and contrast ratio (27 staircases total; see Section 3.3: Experimental protocol).

To ensure efficient selection of jitter amplitudes around a participant's threshold, we adapted recommended staircase parameters that are known to target the 80% correct threshold [12, 13] and validated their efficiency using mathematical simulations and pilot testing. The parameter values used were as follows: 1) initial jitter amplitude was 0.0175 m (1° of visual angle at 1 m, 0.5° at 2 m, and 0.2° at 5 m); 2) minimum amplitude was 0 m; 3) before the first reversal, the number of correct (incorrect) trials before decrementing (incrementing) amplitude was 1 (1), i.e. a 1-down, 1-up rule; 4) after the first reversal, the number of correct (incorrect) trials before decrementing (incrementing) amplitude was 3 (1), i.e. a 3-down, 1-up rule; 4) decrement (increment) amplitude value was 0.00225 m (0.001 m) before the first reversal and 0.000375 m (0.0005 m) afterwards, consistent with the optimal ratio for a 3-down, 1-up rule provided in [13]; 5) maximum number of reversals before termination was 6; 6) maximum number of trials was 60. Each staircase terminated when a participant reached the criterion for maximum number of reversals OR maximum number of trials, whichever came first.

3.6 Data analysis

To obtain an estimate of the 80% correct threshold, we first transformed each trial's jitter amplitude from metric units into retinal angular units (arcminutes) for easier interpretability. For each staircase we calculated the average of the jitter magnitudes at each reversal point, consistent with previous approaches [12, 13]. The result was an estimated threshold for each participant and each combination of viewing distance, background luminance, and contrast ratio.

One participant completed only 2 out of 3 experimental sessions. We treated this session as missing data and excluded it from analysis. We also excluded outlier data points using the following criteria. A staircase typically yields efficient and accurate threshold estimates, but may result in inaccurate estimates if 1) only a small number of reversals occur or 2) the staircase only returns jitter magnitudes that are well above threshold level throughout the experiment, e.g., when the staircase terminates too early. In the first instance, the threshold estimate will be unreliable due to the small number of data points, while in the second instance threshold estimates will be very large relative to the estimates obtained from other staircases. We thus excluded all threshold estimates resulting from staircases of less than 4 reversals or those which were an order of magnitude larger than the participants' average of the estimates from the other staircases. After exclusion based on these criteria, we further excluded data points that were three standard deviations above a participants' mean across all remaining conditions (conducted separately for each participant). These criteria resulted in exclusion of 12.6% of threshold estimates (33/261).

Statistical modelling was conducted using custom scripts in MATLAB (MathWorks, Natick, MA, USA) and R [40]. To assess statistical significance of our independent variables, we implemented a linear mixed-effects regression model using the fitlme package in MATLAB. 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 (e.g., by controlling for random differences in performance between participants, termed random effects) while also allowing one to test the effect of independent variables across participants (e.g., by specifying fixed effects parameters). To control for gross differences in threshold between participants, we selected a random effects structure that allowed us to fit unique intercepts for each participant (1 parameter per participant). We included fixed effects parameters corresponding to planned contrasts comparing each level of our independent variables. For contrast ratio, we compared the lowest level (1.9:1) against each higher level (2.3:1, 2.9:1, 3.1:1, 4.1:1, 6.2:1, and 9.5:1). For viewing distance, we compared the distance that approximates the focal length of the HoloLens 2 (2 m) against the distances closer (1 m) and farther (5 m) away. For background luminance, we compared the dimmest level (10 Cd/m^2) against either brighter level (60 Cd/m^2 and 100 Cd/m^2). We also added two fixed effect covariate parameters to control for demographic factors: age and years of experience with AR/VR (only 9 participants reported this demographic information).

The final model is specified as:

$$\begin{aligned}\text{Threshold} = {} & \beta_{\text{intercept}} + \beta_{\text{2.3:1 vs. 1.9:1}} + \beta_{\text{2.9:1 vs. 1.9:1}} + \\ & \beta_{\text{3.1:1 vs. 1.9:1}} + \beta_{\text{4.1:1 vs. 1.9:1}} + \beta_{\text{6.2:1 vs. 1.9:1}} + \\ & \beta_{\text{9.5:1 vs. 1.9:1}} + \beta_{\text{60 Cd/m}^2\text{ vs. 10 Cd/m}^2} + \\ & \beta_{\text{100 Cd/m}^2\text{ vs. 10 Cd/m}^2} + \beta_{\text{1 m vs. 2 m}} + \\ & \beta_{\text{5 m vs. 2 m}} + \beta_{\text{age}} + \beta_{\text{AR/VR exp.}} + \\ & \beta_{\text{Participant (random effect)}}\end{aligned}$$

To validate the normality assumption of linear regression [35], we inspected a Quantile-Quantile (Q-Q) plot of the model residuals for (approximate) linearity [25, 44], rather than using a formal statistical test (e.g. Shapiro-Wilk [36]), which "almost always yield significant deviations from normality at large sample sizes" [14, 21]. To test whether a fixed effect parameter in the model is significantly different from 0, we must know the null distribution of the parameter estimates and their associated test statistics. It is generally not possible to know the exact form of these distributions when employing mixed-effects modeling. In order to assess statistical significance, we approximated the relevant null distributions using Satterthwaite's method [34]. Statistical significance was determined by comparing the p-value to a criterion of 0.05.

4 RESULTS

4.1 Jitter thresholds indicate subpixel perceptibility

We first determined how the magnitude of jitter thresholds compared to the pixel pitch of the HoloLens 2 that was used to conduct the study. There is no definitive, publicly available value for the HoloLens 2's pixel pitch. Microsoft's device specifications [8] report the "holographic density" of the HoloLens 2 to be >2500 radiants (light points per radan), which converts to 1.4 arcminutes per pixel, assuming a 1:1 mapping from light points to pixels. This estimate is conservative compared to an independent report of display capability measurements [18], which estimates the pixel pitch at 3 arcminutes per pixel. We consider these two estimates as an approximate range of potential pixel pitch values with which to compare the measurements taken in our study.

The estimated thresholds for each combination of background luminance, viewing distance, and contrast ratio are depicted in Figure 2. Each plot depicts the data separately for each background luminance level (left: 10 Cd/m^2, middle: 60 Cd/m^2, right: 100 Cd/m^2). Within each plot, the abscissae correspond to viewing distance while threshold is plotted on the ordinate. Contrast ratio is color coded according to the legend located on the right. Each circle marker is the average threshold across all participants and error bars represent 95% confidence intervals. Individual participant thresholds are depicted with gray triangles. The range of potential pixel pitches are depicted with the dashed red lines. Surveying the data, it is clear that the average threshold in most conditions is at or below the upper bound pixel pitch estimate of 3 arcminutes (compare the relative positioning of each marker to the upper red line). This observation is confirmed quantitatively: 82% (187/228) of individual participant thresholds and 92.6% (25/27) of average thresholds are less than or equal to 3. Using the more conservative criterion, 27.6% (63/228) of individual participant thresholds and 3.7% (1/27) of average thresholds are less than or equal to 1.4 arcminutes. Taken together, this data indicates some participants perceived subpixel jitter in a subset of the viewing conditions tested in this study.

4.2 Effect of independent variables

We next determined whether our independent variables modulated jitter perceptibility by assessing the statistical significance of the corresponding parameters in our regression model. Parameter estimates and estimated p-values are reported in Table 1. For viewing distance, thresholds decreased with increasing distance: thresholds were significantly larger at 1 m compared to 2 m (β = 0.66, S.E. = 0.198, $t_{194.2}$ = 3.35, $p < 0.01$) and were larger at 2 m than at 5 m (β = -0.51, S.E. = 0.192, $t_{194.1}$ = -2.65, $p < 0.01$). For background luminance, thresholds were smaller at 10 Cd/m^2 than at 60 Cd/m^2 (β = 0.75, S.E. = 0.330, $t_{194.1}$ = 2.26, $p < 0.05$), while thresholds were marginally smaller at 10 Cd/m^2 than at 100 Cd/m^2 (β = 0.90, S.E. = 0.482, $t_{194.06}$ = 1.86, p = 0.06). None of the parameters testing differences between levels of contrast ratio were significant (all p's > 0.1).

To compute measures of effect size, we performed the likelihood ratio test [23] using a nested model approach, i.e. building towards the full model by specifying increasingly complex models that are special cases of one another. This approach quantifies whether the increased model complexity that results from adding parameters leads to a greater improvement in performance (goodness of fit) than would have been obtained by chance alone, by computing the ratio (LR) of log likehoods between the simpler and more complex model. Greater LR's indicate increasing improvement in goodness of fit for the more complex model [15]. Adding a main effect of viewing distance resulted in a significantly better fit (LR = 35.01, $p < .001$) over the null model (where the dependent variable is predicted by its overall mean), which was not the case for main effects of contrast (LR = 2.89, $p > .05$) or background luminance (LR = 1.23, $p > .05$). The full model, which included interactions, performed significantly better (LR = 7.09, $p < .05$) than a model with only main effects.

4.3 Effects of age and AR/VR experience

Our regression model included as covariates participant ages and years of experiences using AR/VR to ensure the parameter estimates of our independent variables were not contaminated by confounding variables. Jitter thresholds did not vary significantly with either age ($p > 0.9$) or AR/VR experience ($p > 0.3$).

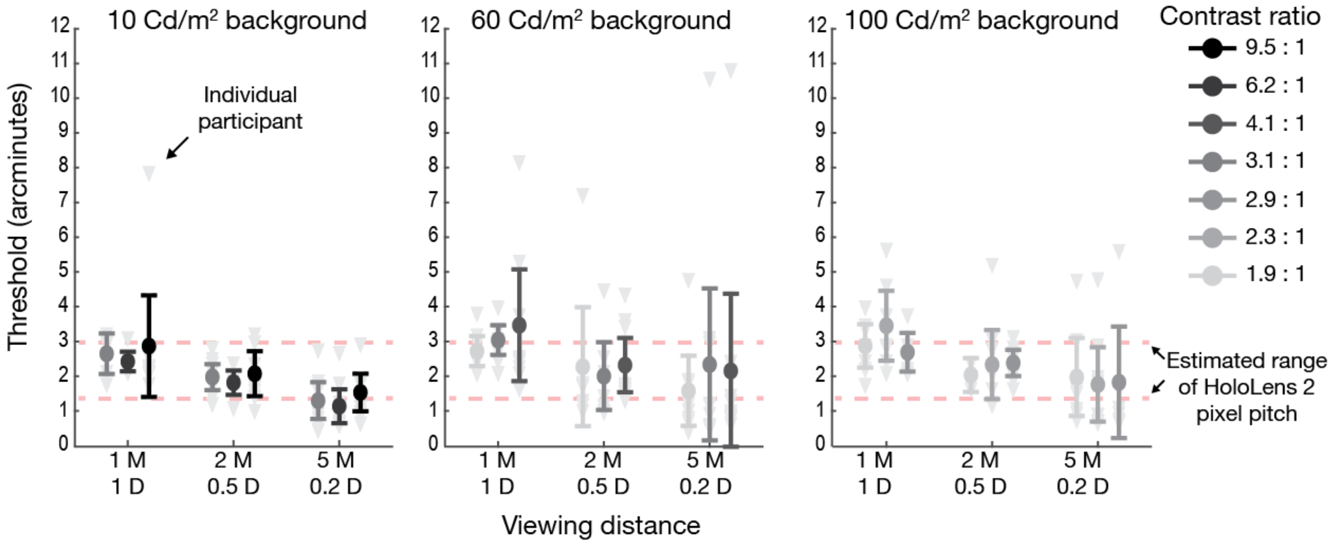

Figure 2: Results of the experiment. Data are plotted separately for each background luminance level (left: 10 Cd/m^2, middle: 60 Cd/m^2, right: 100 Cd/m^2). Within each plot, the abscissae depict viewing distance in depth and threshold is plotted on the ordinate. Circle markers represent the average threshold across all participants. Error bars represent 95% confidence intervals. Contrast ratio is color coded according to the legend located on the right. Individual participant thresholds are depicted with gray triangles. The range of estimated pixel pitches are depicted with dashed red lines.

Table 1: Results of mixed-effects regression model analysis. Star (*) denotes significance using a criterion of 0.05; dagger (†) denotes marginal significance.

Parameter	Estimate (SE)	*t* value [DoF]	*p* value
Intercept	1.92 (1.600)	1.20 [6.9]	0.27
2.3:1 vs. 1.9:1	0.12 (0.343)	0.34 [194.1]	0.73
2.9:1 vs. 1.9:1	-0.11 (0.353)	-0.31 [194.1]	0.76
3.1:1 vs. 1.9:1	0.43 (0.332)	1.28 [194.2]	0.20
4.1:1 vs. 1.9:1	0.40 (0.329)	1.21 [194.1]	0.23
6.2:1 vs. 1.9:1	0.37 (0.472)	0.78 [194.1]	0.43
9.5:1 vs. 1.9:1	0.73 (0.742)	1.55 [194.1]	0.12
60 Cd/m^2 vs. 10 Cd/m^2	0.75 (0.330)	2.26 [194.1]	0.02*
100 Cd/m^2 vs. 10 Cd/m^2	0.90 (0.482)	1.86 [194.1]	0.06 †
1 m vs. 2 m	0.66 (0.198)	3.35 [194.2]	<0.001*
5 m vs. 2 m	-0.51 (0.192)	-2.65 [194.1]	<0.01*
Age	-0.004 (0.041)	-0.10 [6]	0.93
AR/VR experience	-0.07 (0.073)	-1.0 [6]	0.37

5 DISCUSSION

Here, we provide the first psychophysical measurements of jitter perceptibility for WL virtual content in an AR HMD. Threshold and perceptibility are inversely related: larger thresholds indicate jitter was less perceptible in that viewing condition, while smaller thresholds indicate jitter was more perceptible. We observed that thresholds decreased with increasing viewing distance from the observer). One plausible explanation for the increased thresholds at 1 m compared to 2 m is that there were differences in the magnitude of vergence-accommodation conflict (VAC) experienced by the observers when viewing the virtual cubes at either distance. The approximate focal length of the HoloLens 2 is 2 m [32], indicating there should be little if any VAC for the cubes presented at this distance. However, for the 1 m condition the observer must accommodate at the 2 m (0.5 D) distance while converging at 1 m (1 D), resulting in a VAC of roughly 0.5 D. Previous studies have reported that virtual content presented beyond 0.5 D on either side of a fixed focus AR display (like the one employed in the HoloLens 2) appear blurred [10], while other studies employing VR displays have found that viewing stimuli more than 0.5 D closer or farther from the display focal plane results in discomfort [37]. The 0.5 D conflict experienced during the 1 m condition in our study is right at the boundary for noticeable blur/discomfort, suggesting these symptoms may have contributed to increased thresholds compared to the 2 m (0 D conflict) distance. However, VAC does not appear to be a plausible explanation for the reduced thresholds at 5 m (0.2 D) compared to 2 m; while the estimated conflict of -0.3 D is within the range in which we would expect little to no blurring/discomfort, we would not expect better performance than at the 2 m distance. It is possible that during the 5 m condition, observers may have accommodated and converged at the back wall (which was 5.5 m away) while still attending to the cube's motion. If so, the 5 m condition would represent a 'real world fixation' condition, which may have less blurring and/or discomfort than the 2 m condition where the observer is fixating at the focal plane of the AR display. Alternatively, it is possible that the observers did fixate the virtual cube, but any decrease in performance caused by VAC was mitigated due to increased perceived motion parallax resulting from the proximity of the cube to the back wall. We advocate for future work testing a larger range of VACs while controlling for physical object proximity to determine the cause of the distance effect reported here.

We also observed smaller thresholds (higher perceptibility) for the dimmest background luminance (10 Cd/m^2) compared to the intermediate (60 Cd/m^2) and brightest (100 Cd/m^2, albeit a marginally significant difference) levels. At first glance, it may appear that these differences are explained by unbalanced levels of contrast ratio at each background luminance, where a subset of more perceptible contrast ratios may have been presented at the dimmest background luminance. In fact, the highest contrast ratios were presented at the dimmest background luminance due to the display constraints of the HoloLens 2 (i.e., it was not possible to present higher contrast ratios at larger background luminances). However, this possibility is unlikely because we controlled for contrast ratio in our regression model. Moreover, at the single contrast ratio (3.1:1) that was used in both the 10 Cd/m^2 and 60 Cd/m^2 conditions, the average threshold was smaller for the dimmer condition (1.97 arcminutes at 10 Cd/m^2 vs. 2.49 arcminutes at 60 Cd/m^2), indicating that thresholds differed between background luminance levels even when contrast ratio was matched. If the perceptibility difference is in fact due to the background luminance itself, it is possible that observers were more sensitive to motion at the dimmer luminance because of neurobiological factors. At low light levels between roughly 0.01 Cd/m^2 and 5 Cd/m^2, visual acuity functions in the mesopic range, where a combination of motion-sensitive photoreceptor rods and cones are both active [43]. In contrast, at light levels above 5 Cd/m^2 vision is primarily driven by cones. It is possible that relatively more rods were active when our observers viewed the 10 Cd/m^2 background luminance condition compared to the more illuminant conditions, yielding higher perceptibility (and smaller thresholds). Alternatively, it is possible that head-tracking and WL rendering performance were degraded in the low luminance condition, which could affect performance by raising the level of baseline jitter in the reference stimulus (see Section 6: Limitations for discussion). Future work should test background luminance levels in the mesopic range while employing an experimental testbed with known head-tracking and WL rendering performance levels to determine how these factors modulate jitter perceptibility.

We did not observe a significant difference in threshold between levels of contrast ratio. While it is tempting to conclude that jitter perceptibility does not vary with contrast ratio, one must use caution when interpreting a null result. It is possible that jitter perception is modulated by contrast ratio but that we did not test an expansive enough range to observe it. For example, if contrast ratio were sufficiently small to be below an observer's threshold for contrast perception (e.g., close to 1:1), perceptibility would necessarily be worse than when contrast ratio is well above threshold. We intentionally chose a range of suprathreshold contrasts in this study to understand how jitter perceptibility varies at levels that are expected in AR HMDs, potentially limiting our ability to observe such an effect. At minimum, we may conclude that there is more evidence than not that jitter perceptibility does not vary with contrast ratio in the suprathreshold range tested in this study. A similar logic applies to the null results of the age and AR/VR experience covariates we included in our model. In particular, older observers are known to perform worse on some visual tasks due to age-related visual processing impairments, indicating we might expect larger jitter thresholds (reduced perceptibility) for older individuals. The average age of participants in our study was 34.9 and the oldest participant was 53, which means our sample may have skewed too young to observe any age differences. Future work should look to include a wider range of demographics.

Finally, our finding that some participants perceived subpixel jitter for a subset of viewing conditions is not necessarily intuitive as it conflicts with the widely held notion that the pixel is the smallest unit of renderable precision. There are at least three reasons why this presumption is not true, and they are not mutually exclusive. First, any raster display has the capacity to render at subpixel resolution by distributing the intensity values intended for a given pixel across neighboring pixels (i.e., anti-aliasing). The virtual cubes used in this study were indeed anti-aliased in our Unity programs. Second, displays with a bit depth >1 (e.g., color raster displays like that used in the HoloLens 2) can render edges in virtual texture space before projecting them onto the display pixels, which in conjunction with anti-aliasing yield effective rendering resolution that is significantly higher than the pixel pitch of the display. Third, consider that an observer's perception of a rendered object is the result of a sophisticated estimation process that relies on the encoding and decoding of visual information via populations of neural activity. The brain's estimate of an objects' position is not based on pixels or rendering engines but rather a probabilistic spatiotemporal representation of the visual properties of the object. Thus, even in an extreme hypothetical case where a jittering cube is rendered on a display with purely discrete pixels (either 'on' or 'off,' with no modulation of intensity) and no anti-aliasing, the evoked pattern of neural activity may yield perception of a smoothly oscillating movement with subpixel magnitude. Taken together, it is entirely possible for the observers in this study to have perceived subpixel jitter, and our threshold measurements at or below the range of estimated HoloLens 2 pixel pitches indicate that some observers did.

More generally, this study adds to recent advancements in product and user experience research that advocate for increased inclusion of perceptual science into the design and engineering of AR and VR systems [4, 7, 10, 11, 16]. Recent advancements in spatial computing platforms are converging on solutions designed to extend a user's experience of reality by aiding and enhancing their perception. Psychophysical methods like the ones employed in this study offer valuable tools with which to quantify and understand perceptual experiences, an ability that is lacking in traditional user research methods predicated on survey responses and focus groups. Additionally, psychophysical methods have been developed over decades and offer the added benefit of widespread usage in modern basic and applied research, allowing researchers to employ both well established and novel methods to answer fundamental perceptual questions. In sum, we advocate for the following principle: one cannot build convincing AR/VR devices without understanding the impact of each component on users' perceptual systems.

6 LIMITATIONS

We used a commercially available AR HMD (the Microsoft HoloLens 2) to present our experimental stimuli. While the HoloLens 2 offers a platform for presenting AR virtual content, displayed content is susceptible to a variety of spatiotemporal artifacts resulting from the world-locking, rendering and display pipeline, including jitter. Thus, all objects rendered in our experiment possessed some amount of baseline jitter, including the rendered-to-be-stationary object that our participants used as a jiiterless reference when making judgements in our experimental task. Measurements of the spatiotemporal characteristics of this 'baseline jitter' are not publicly available, preventing us from knowing the magnitude of jitter that was present in the stationary reference. Moreover, we cannot know the spatiotemporal envelope of this jitter, or whether its magnitude varied as a function of our independent variables (e.g., if baseline jitter was larger at certain viewing distances or contrast ratios) or between participants (e.g., due to inter-participant differences in the frequency and magnitude of head movements during stimulus presentation). As such, our measurements do not provide measures of absolute jitter perceptibility (i.e., how much jitter is needed to discriminate moving from truly static content). Instead, our results should be interpreted as measurements of relative jitter perceptibility of the amount of added jitter necessary to reliably discriminate the moving target from the baseline jitter in the HoloLens 2 (that was present in the reference). More generally, the uncertainty associated with the specifics of the HoloLens 2 rendering pipeline and display capabilities make it unclear whether our results are generalizable to other AR HMDs, which may possess very different architectures and specifications. Future work should employ an experimental apparatus with well-characterized measurements of baseline spatiotemporal artifacts, and the ability to directly manipulate overall (i.e., baseline plus added) jitter, to obtain estimates of absolute perceptibility for this artifact.

7 CONCLUSIONS AND FUTURE WORK

Here, we found that WL virtual content jitter in AR HMDs 1) is detectable at subpixel resolution for some viewing conditions; 2) increases in perceptibility at farther viewing distances; and 3) is more detectable at dim compared to brighter background luminance levels. When designing and engineering an AR HMD system, we make the following recommendations based on the perceptual results presented. First, jitter tolerances (maximum magnitudes allowed) required to render convincing WL content may be smaller than the pixel pitch of the HMD, so pixel pitch should not automatically be assumed as a lower bound. Second, it may be possible to relax jitter tolerances slightly at closer viewing distances and/or in brighter ambient lighting conditions, so these along with other content and environmental variables should be considered when optimizing efficient WL rendering systems.

Future work should seek to determine whether the conclusions drawn from this study extend to virtual content other than 3D grayscale cubes; for example, text-based stimuli. Moreover, while this study focused on jitter perceptibility, a full characterization of the potential effects that jitter may have on a user's experience in AR should also include measurements of perceived quality, visual discomfort, and usability (e.g., how does jitter modulate a user's ability to interact with virtual content?). Finally, AR is often used in a wider variety of scenarios than the static viewing that observers conducted in this study. We advocate for an extension of our study to scenarios where an observer is standing, walking or performing simultaneous tasks (e.g., multi-tasking).

ACKNOWLEDGMENTS

The authors wish to thank John Pella, Justin Wang, Sara Kenley, and Helen Ayele for support in experiment implementation and data collection; Dana Qi and Hunter Herrin for collecting HoloLens 2 display measurements; and Zahid Hossain, Anqi Xu, Romain Bachy, and MJ Kim for valuable feedback and discussion.

REFERENCES

[1] 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. ACM, Virtual Event USA, Oct. 2020. doi: 10.1145/3379337.3415833

[2] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, Aug. 1997. doi: 10.1162/pres.1997.6.4.355

[3] 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 K. Arai, S. Kapoor, and R. Bhatia, eds., *Proceedings of the Future Technologies Conference (FTC) 2020, Volume 2*, Advances in Intelligent Systems and Computing, pp. 792–808. Springer International Publishing, Cham, 2021. doi: 10.1007/978-3-030-63089-8_52

[4] D. R. Blanc-Goldhammer and K. J. MacKenzie. The effects of natural scene statistics on text readability in additive displays. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1):1281–1285, Sept. 2018. Publisher: SAGE Publications Inc. doi: 10.1177/1541931218621294

[5] D. Burr and P. Thompson. Motion psychophysics: 1985-2010. *Vision Research*, 51(13):1431–1456, July 2011. doi: 10.1016/j.visres.2011.02.008

[6] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, 32(6):1309–1332, Dec. 2016. Conference Name: IEEE Transactions on Robotics. doi: 10.1109/TRO.2016.2624754

[7] S. A. Cholewiak, Z. Başgöze, O. Cakmakci, D. M. Hoffman, and E. A. Cooper. A perceptual eyebox for near-eye displays. *Optics Express*, 28(25):38008–38028, Dec. 2020. Publisher: Optica Publishing Group. doi: 10.1364/OE.408404

[8] S. Cooley, E. Miller, B. Hanson, J. Mathew, JoyJaz, J. Joubran, and S. Paniagua. HoloLens 2 Hardware, Nov. 2021.

[9] 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. doi: 10.1109/MRA.2006.1638022

[10] I. M. Erkelens and K. J. MacKenzie. Vergence-accommodation conflicts in augmented reality: Impacts on perceived image quality. *SID Symposium Digest of Technical Papers*, 51(1):265–268, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sdtp.13855. doi: 10.1002/sdtp.13855

[11] J. M. Fulvio, M. Ji, and B. Rokers. Variations in visual sensitivity predict motion sickness in virtual reality. *Entertainment Computing*, 38:100423, May 2021. doi: 10.1016/j.entcom.2021.100423

[12] M. A. Garcia-Perez. Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties. *Vision Research*, 38(12):1861–1881, June 1998. doi: 10.1016/S0042-6989(97)00340-4

[13] M. A. Garcia-Perez. Optimal setups for forced-choice staircases with fixed step sizes. *Spatial Vision*, 13(4):431–448, 2000. doi: 10.1163/156856800741306

[14] A. Ghasemi and S. Zahediasl. Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2):486–489, 2012. doi: 10.5812/ijem.3505

[15] S. Glover and P. Dixon. Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review*, 11(5):791–806, Oct. 2004. doi: 10.3758/BF03196706

[16] A. Goettker, K. J. MacKenzie, and T. S. Murdison. Differences between oculomotor and perceptual artifacts for temporally limited head mounted displays. *Journal of the Society for Information Display*, 28(6):509–519, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jsid.912. doi: 10.1002/jsid.912

[17] D. M. Green and T. G. Birdsall. Detection and recognition. *Psychological Review*, 85(3):192–206, 1978. Place: US Publisher: American Psychological Association. doi: 10.1037/0033-295X.85.3.192

[18] K. Guttag. Hololens 2 Display Evaluation (Part 2: Comparison to Hololens 1), July 2020.

[19] Y. Itoh, T. Langlotz, J. Sutton, and A. Plopski. Towards indistinguishable augmented reality: A survey on optical see-through head-mounted displays. *ACM Computing Surveys*, 54(6):120:1–120:36, July 2021. doi: 10.1145/3453157

[20] C. Kaernbach. Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49(3):227–229, 1991. Place: US Publisher: Psychonomic Society. doi: 10.3758/BF03214307

[21] U. Knief and W. Forstmeier. Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, 53(6):2576–2590, Dec. 2021. doi: 10.3758/s13428-021-01587-5

[22] C. F. Lam, J. H. Mills, and J. R. Dubno. Placement of observations for the efficient estimation of a psychometric function. *The Journal of the Acoustical Society of America*, 99(6):3689–3693, June 1996. Publisher: Acoustical Society of America. doi: 10.1121/1.414966

[23] F. Lewis, A. Butler, and L. Gilbert. A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*, 2(2):155–162, 2011. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2041-210X.2010.00063.x. doi: 10.1111/j.2041-210X.2010.00063.x

[24] 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*, ISS '19, pp. 5–16. Association for Computing Machinery, New York, NY, USA, Nov. 2019. doi: 10.1145/3343055.3359710

[25] A. Loy, L. Follett, and H. Hofmann. Variations of Q–Q plots: The power of our eyes! *The American Statistician*, 70(2):202–214, Apr. 2016. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00031305.2015.1077728. doi: 10.1080/00031305.2015.1077728

[26] R. D. Luce. A threshold theory for simple detection experiments. *Psychological Review*, 70(1):61–79, 1963. Place: US Publisher: American Psychological Association. doi: 10.1037/h0039723

[27] I. S. MacKenzie and C. Ware. Lag as a determinant of human performance in interactive systems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pp. 488–493. Association for Computing Machinery, New

York, NY, USA, May 1993. doi: 10.1145/169059.169431

[28] N. A. Macmillan and C. D. Creelman. *Detection Theory: A User's Guide*. Psychology Press, New York, 2 ed., Aug. 2004. doi: 10.4324/9781410611147

[29] M. Nancel, D. Vogel, B. De Araujo, R. Jota, and G. Casiez. Next-point prediction metrics for perceived spatial errors. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, UIST '16, pp. 271–285. Association for Computing Machinery, New York, NY, USA, Oct. 2016. doi: 10.1145/2984511.2984590

[30] S. Nishida, T. Kawabe, M. Sawayama, and T. Fukiage. Motion perception: From detection to interpretation. *Annual Review of Vision Science*, 4(1):501–523, 2018. _eprint: https://doi.org/10.1146/annurev-vision-091517-034328. doi: 10.1146/annurev-vision-091517-034328

[31] M. Ouhyoung and J.-R. Wu. On latency compensation and its effects on head-motion trajectories in virtual environments. *The Visual Computer*, 16(2):79–90, Mar. 2000. doi: 10.1007/s003710050198

[32] E. J. Paul, S. Formicola, V. Tieto, and D. Coulter. Comfort - Mixed Reality, Oct. 2021.

[33] 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*, EICS '09, pp. 187–196. Association for Computing Machinery, New York, NY, USA, July 2009. doi: 10.1145/1570433.1570469

[34] F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6):110–114, 1946. Publisher: [International Biometric Society, Wiley]. doi: 10.2307/3002019

[35] A. F. Schmidt and C. Finan. Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, 98:146–151, June 2018. doi: 10.1016/j.jclinepi.2017.12.006

[36] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965. Publisher: [Oxford University Press, Biometrika Trust]. doi: 10.2307/2333709

[37] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8):11, 2011. doi: 10.1167/11.8.11

[38] 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, Mar. 2018. doi: 10.1109/VR.2018.8446195

[39] J.-P. Stauffert, F. Niebling, and M. E. Latoschik. Latency and cybersickness: Impact, causes, and measures. A review. *Frontiers in Virtual Reality*, 1, 2020.

[40] R. C. Team. R: A Language and Environment for Statistical Computing, 2020.

[41] 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, Mar. 2009. doi: 10.1109/3DUI.2009.4811204

[42] J. M. P. van Waveren. The asynchronous time warp for virtual reality on consumer hardware. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, VRST '16, pp. 37–46. Association for Computing Machinery, New York, NY, USA, Nov. 2016. doi: 10.1145/2993369.2993375

[43] A. J. Zele and D. Cao. Vision under mesopic and scotopic illumination. *Frontiers in Psychology*, 5, 2015.

[44] A. F. Zuur, E. N. Ieno, and C. S. Elphick. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1):3–14, 2010. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2041-210X.2009.00001.x. doi: 10.1111/j.2041-210X.2009.00001.x