Perception of Looming Motion in Virtual Reality Egocentric Interception Tasks

Robert A. Rolin, Jolande Fooken, Miriam Spering, and Dinesh K. Pai, *Member, IEEE*

Abstract—Motion in depth is commonly misperceived in Virtual Reality (VR), making it difficult to intercept moving objects, for example, in games. We investigate whether motion cues could be modified to improve these interactions in VR. We developed a time-to-contact estimation task, in which observers (n=18) had to indicate by button press when a looming virtual object would collide with their head. We show that users consistently underestimate speed. We construct a user-specific model of motion-in-depth perception, and use this model to propose a novel method to modify monocular depth cues tailored to the specific user, correcting individual response errors in speed estimation. A user study was conducted in a simulated baseball environment and observers were asked to hit a looming baseball back in the direction of the pitcher. The study was conducted with and without intervention and demonstrates the effectiveness of the method in reducing interception errors following cue modifications. The intervention was particularly effective at fast ball speeds where performance is most limited by the user's sensorimotor constraints. The proposed approach is easy to implement and could improve the user experience of interacting with dynamic virtual environments.

Index Terms—virtual reality, motion perception, time to contact, augmented reality, games

1 Introduction

Virtual Reality (VR) and related 3D display technologies have recently experienced tremendous growth in promise and popularity, but have significant limitations. Human vision, carefully tuned to integrating multiple cues from the real world such as disparity, vergence, and accommodation, can incorrectly perceive the virtual world in these displays (e.g., [1]). However, accurate motion perception is essential for tasks such as interception and collision avoidance that are central to many VR applications that simulate baseball and other ball sports, as well as in many video games. We describe a novel method for correcting the perception of moving objects, such as balls, that move towards the observer in VR. In a calibration stage, we first learn user-specific parameters based on a user's errors in speed judgments to characterize their perception of virtual motion. In the second stage, we modify motion-in-depth cues to facilitate more accurate judgments for that user. Using a baseball hitting simulation in VR, we show that our modified motion-in-depth cues resulted in more accurate and consistent interactions of moving objects in VR. The presented method has been tested with objects moving at constant speed straight towards the user. Our model and correction algorithms apply to this case only. However, the method could be adapted to apply to other forms of motion, such as motion in a fronto-parallel plane or curved trajectories. Our motion correction is analogous to the ubiquitous Gamma correction required to correct for non-linearity in light output of individual CRT displays independent of the viewer. Ours is a simple technique that can similarly improve motion perception in VR displays tailored to the individual user.

• *R. Rolin and D. K. Pai (Computer Science); J. Fooken and M. Spering (Ophthalmology & Visual Sciences); University of British Columbia, Vancouver, BC, Canada. Correspondence to pai@cs.ubc.ca.*

2 Related Work

2.1 Time to Contact

Behavioral studies in humans, using visual psychophysics, have provided a deep understanding of motion sensitivity – our ability to detect moving objects and to discriminate their motion direction or speed [2], [3]. It is an area of interest due to its importance in everyday life for tasks such as collision avoidance or interception, two areas of applied VR research. The most relevant aspect for our study is perception of time-to-contact (TTC), the ability to estimate the time when a moving object will reach a target location. In a typical TTC experiment, observers judge when an object will collide with another object when one or both objects are moving. For most TTC experiments, a combination of speed, distance, and viewing time are varied, i.e., the object will disappear at some point in its collision trajectory. The goal of such research is to determine what visual information is needed in order to estimate TTC accurately [4], [5]. For egocentric TTC tasks, in which the target moves towards the observer, it has been hypothesized that humans rely on information derived from the target's visual angle and its rate of expansion or change of size on the retina (looming), captured by the so-called variable Tau, a monocular predictor of TTC [6], [7], $\tau = \theta/(\partial\theta/\partial t)$. Here θ is the visual angle subtended by the object, and $\partial\theta/\partial t$ is the rate of expansion on the retina.

However, Tau might not provide sufficient information to account for human-level accuracy. Instead, the visual system might rely on binocular indicators of TTC and take into account target kinematics such as distance, speed, and acceleration [8], [9], [10], [11]. For example, Gray and Regan [12] proposed the following equation for TTC based only on binocular information for objects moving along a straight line through a point midway between the eyes at constant speed: $TTC \approx I/D(\partial\delta/\partial t)$. Here D is the distance from object to observer, $(\partial\delta/\partial t)$ is the rate of change of relative disparity and I is the interpupillary distance.

The accuracy and relative strength of these signals have been investigated in TTC tasks using perceptual time estimates or the manual interception of a moving target [12], [13], [14], [15], [16].

2.2 Perception of Distance and Motion in VR

In the real world, human observers can develop expertise in TTC tasks such as hitting or catching a ball with high accuracy and precision. Yet, in the laboratory, observers' performance in TTC tasks is highly variable and speed is frequently misjudged, with the sign of the error depending on factors such as viewing time. [17], [18]. These TTC estimation errors occur both in laboratory tasks involving regular displays as well as in VR using head-mounted displays [19]. The use of head-mounted displays may also result in an apparent compression of distance (space) in egocentric distance estimation tasks in virtual or augmented indoor as compared to real-world indoor settings [20], [21], [22], [23]. In these studies, an observer might first view a target on the ground, and then be asked to walk towards it with occluded vision (blind-walking task). Consistent underestimations of object size as well as distance in VR by up to 85% have been observed, and many contributing factors have been discussed in the literature. For example, head-mounted displays provide fewer depth cues [24] and a limited field of view [25]. However, field of view did not affect distance estimation when subjects were free to move their head [25], indicating that depth perception deficits must be due to other factors, such as lack of experience in interacting with VR [26], or restrictions in the ergonomics of acting in VR. The magnitude of estimation errors of size and distance can be alleviated by giving observers visual-motor feedback during locomotion tasks, presumably resulting in visual-motor recalibration of space [27].

The lack of available depth cues (e.g., blur) in VR displays has been discussed extensively in the literature. Most modern VR headsets use fixed focus lenses, providing binocular stereo cues but not accommodation cues. This shortcoming, known as the "vergence-accommodation conflict" leads to visual fatigue and reduced ability to perceive 3D structure [1]. Because of these problems, many attempts have been made to create stereoscopic displays with multiple focal lengths, resulting in a clear benefit when this conflict is minimized [1], [28], [29], [30]. Another outcome of having a fixed focal distance is a lack of defocus blur, an important depth cue [31], [32], [33], [34]. Accurate models have been developed to predict how blur will interact with other relative depth cues in an image to produce an apparent scale of the image's contents [35]. The use of blur in gaze-contingent stereoscopic displays has been shown to enhance realism, improve quantitative depth perception, and reduce discomfort [36], [37], [38], [39]. Additionally, the use of blur has been credited with significant reduction in rendering costs [40], [41].

Similarly, displays that incorporate head tracking, stereo, and a large field-of-view enhance spatial judgments [42] and understanding, e.g., in VR exploration, navigation, and visualization tasks [43]. Despite these recent advances in VR technology, perceptual errors remain. Yet, VR is widely used as a tool for studying interception in naturalistic tasks such as ball sports [44], [45], [46]. There is also growing interest in using VR for training athletes, with both academic research in implementing realistic sports-simulations [47], [48], and commercial products available [49]. The current study cannot resolve the debate as to what the contributing factors to misperceptions in VR are; rather, it presents a recalibration method tailored at the individual user to correct such estimation errors, with the ultimate goal of contributing to more useful experiences in VR.

3 Methods

Our system is composed of two main parts. In the first part, *User Calibration*, we determine the parameters of a model which characterizes how a user perceives motion-in-depth in VR. In the second part, *Motion Correction*, we use a novel strategy for modifying the motion-in-depth cues of an arbitrary VR object using the parameters determined from the calibration. The modifications are meant to present a set of stimuli that will increase a user's ability to interpret the movement of the original object and thereby increase accuracy when interacting with virtual objects.

To be concrete, let $\boldsymbol{v} = (\dot{x}\ \dot{y}\ \dot{z})^T$ be the velocity of a small object, expressed in a world-fixed coordinate frame with origin located at the nominal location of the observer's head and oriented using the typical convention in graphics ($-z$ is the viewing direction, y pointing up). We conducted a preliminary study in which observers were presented with balls either moving in the fronto-parallel plane at $z = -30m$ along $+x$, or moving towards the viewer from $(0\ 0\ -30)^T$ along $+z$. The results suggested that horizontal motion could be perceived accurately across speeds and viewing times in VR, but looming motion was misperceived. Thus we focus on modeling how an input motion at speed $\dot{z}$ is actually perceived, and modifying other cues (e.g., the rate of change of size) so that the stimulus will be perceived as actually moving at speed $\dot{z}$.

In this paper we use a virtual baseball environment to test our methods. The task of hitting a baseball provides a common scenario in which speed estimation plays a crucial role.

3.1 User Calibration

The goal of the calibration procedure is to relate the physical object's speed, $\dot{z}$, to the user's perception of the speed, $\dot{z}_p$. In other words, we construct an empirical model $\dot{z}_p = f(\dot{z})$ by estimating the parameters of the model.

We derive an observer's perceived speed from their subjective estimates of TTC. We asked observers to indicate, by button press, when a virtual looming object will collide with their head. The object disappears after traveling three quarters of the distance to the user to prevent any feedback on the actual time of contact. The response times from a representative subject can be seen in Fig. 1a. To determine a perceived speed for a given object speed, we take the median response time over all trials for that speed. By dividing the distance traveled by the median response time, we arrive at an estimate of perceived speed. The estimates

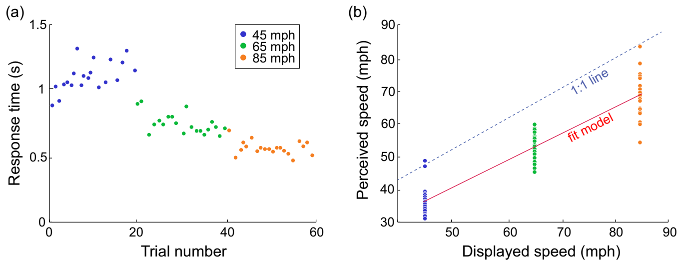

Fig. 1. Calibration data from one subject. (a) Response times for the calibration procedure, differentiated by display speed. The three speeds correspond to actual times-to-contact of 0.90, 0.62 and 0.47 seconds. (b) The model fit to the subject's perceived speeds.

of perceived speed for a subject are shown in Fig. 1b.

After determining $(\dot{z}, \dot{z}_p)$ pairs, we fit a model to this data to allow us to predict perceived speed for any model speed. The data seem to be well approximated as an affine function

$$\dot{z}_p = f(\dot{z}) = w_0 + w_1 \dot{z}, \tag{1}$$

where model parameters w_0 and w_1 are estimated using least squares regression.

Without much a priori knowledge on how VR affects the perception of motion, we cannot justify using very complicated models. An alternative may be to map actual TTC to user response time instead of mapping actual speed to perceived speed. Such a model would be more appropriate if the distance traveled by objects was not constant. Since our calibration and test environments had objects traveling the same distance, mapping based on time and speed are equivalent.

3.2 Motion Correction

Once we have a model that can predict how a user will perceive speed in VR, we can modify depth cues to change the user's perception of virtual motion. Motion through depth is indicated primarily via two depth cues. The first is the monocular motion-in-depth cue, τ, which corresponds to the rate of change of apparent size of an object as it moves through different depths. An object appears larger when it is close to a viewer and smaller when it is far away. The second is a biocular motion-in-depth cue, the rate of binocular disparity change, comes from the difference of the images created by an object on a viewer's left and right retinas.

Binocular disparity is a function of the actual distance of the object and vergence; thus manipulating binocular disparity can only be done by manipulating the actual position of the object in the two images. Since we are attempting to facilitate more accurate interactions, adjusting the position of an object may be challenging. One would want to ensure that an object is in the position a user expects it to be at the time when they want to interact with it.

Adjusting the size of objects can be easily accomplished in VR software. Thus to facilitate more accurate perceptions our technique is to present the position of the object according to the actual model speed while giving the object the monocular depth cues of an object traveling at a speed that will be perceived as the actual model speed (see Fig. 2).

Specifically, given an desired velocity to display, $\boldsymbol{v} = (\dot{x}\ \dot{y}\ \dot{z})^T$, we compute a modified velocity

$$\boldsymbol{v}' = (\dot{x}\ \dot{y}\ f^{-1}(\dot{z}))^T, \tag{2}$$

where f is the model described in Eq. 1. We then determine the visual angle subtended by the object moving with velocity $\boldsymbol{v}'$ and update the size of the original object to subtend the same visual angle. The visual angle of a spherical object is given by

$$\theta = 2\arctan(\frac{r}{d}) \tag{3}$$

where r is the radius of the object and d is the object's distance to the viewer. It can be seen that the size of the original object should be scaled by d/d' where d is the distance between the original object and the viewer and d' is the distance between the object moving with velocity $\boldsymbol{v}'$ and the viewer.

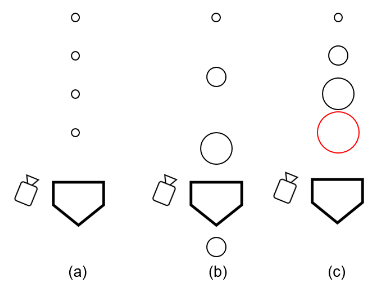

Fig. 2. (a) The position of a ball at four points in a pitch moving with velocity $\boldsymbol{v}$ with actual size shown. (b) The position of a ball moving with velocity $\boldsymbol{v}'$ with perceived size shown. (c) The output of the modifications. The position of the ball with velocity $\boldsymbol{v}$ and perceived size of ball with velocity $\boldsymbol{v}'$. After a certain distance, perceived sizes are extrapolated from previous values.

For looming objects, the procedure above will only produce desirable results during the time-span in which both actual and simulated objects are in front of the viewer (Fig. 2). For example, consider a user who underestimates speed. When the original object is halfway to the user the simulated object might be three quarters of the way to the user. At this point the simulated object will subtend a larger visual angle so the original object will be enlarged. However, once the simulated object passes the viewer it will start to subtend a smaller visual angle. We do not want our original object to start getting smaller because that would not help in making the object seem like it is moving faster. To work around this, we monitor the size of the object and detect when it has stopped increasing (or decreasing). Once that is detected there are number of sensible things to do. We could keep the size constant from that point on. Alternatively, we could maintain the rate of expansion the object had before. We found that keeping size constant caused an artifact in the motion of the object which makes it seem like it is suddenly changing speed or direction. We also found that maintaining the rate of expansion, which is geometric, could cause objects to become unnaturally large. What we found worked best was continuing a linear rate of expansion using the two most recent sizes to determine the rate.

There are other possible modifications that could be used to affect how an object's motion is perceived. In this paper we keep the position of the display speed and the apparent size of the modified display speed. One other technique we tried was maintaining the apparent size of the display speed and the position of the modified display speed. A third technique we tried was restricting all moving objects to be within d_{max} meters. Instead of the baseball starting 18m away, it started at some $d_{max} < 18$m away but maintained the apparent size of the display speed and would arrive at home plate at the same time as the display speed. In preliminary experiments comparing these modifications, we found modifying apparent size, the technique used in this paper, to be the most effective.

It is important to note that observers perceived the ball as rigid despite the size manipulation. In pilot experiments, authors were blinded to block order and were unable to identify whether they participated in a baseline or intervention block. Further, even though participants would sometimes report that one block seemed easier than another, they were unable to identify the difference between baseline and intervention.

4 User Study

We conducted a user study to evaluate the depth cue modification. Human subjects completed the calibration procedure and then played a baseball game with modified motion-in-depth cues, altering perception of the ball's movement. We chose to implement the user study using a realistic baseball scenario to investigate transfer and applicability of our simple calibration model, derived from an abstract task, to a more naturalistic environment. The baseball task relies on the same TTC estimation, yet employs a naturalistic and engaging game-like interactive component. We acknowledge that such a task might introduce additional complexity and noise based on subjects' individual baseline skill level. However, our study excluded subjects with extensive prior experience in interceptive sports (see below) thus minimizing such influence. Subjects were randomly assigned to one of two groups: one group played the baseball game without depth cue modification in block 1 (baseline), then completed the same task with depth cue modification in block 2 (treatment). The other group completed the tasks in reverse order, first treatment then baseline, to control for any effects of training from block 1 to block 2. Subjects were encouraged to take breaks between blocks; including breaks the study took between 45 and 60 minutes to complete.

4.1 Participants

We recruited 21 subjects (mean age: 26 yrs, std = 6.3 yrs; 10 of them female), undergraduate or graduate students at the University of British Columbia. We excluded three participants with experience playing interceptive ball sports (e.g., baseball, softball, tennis) at a competitive level. The remaining 18 participants were randomly assigned to one of two groups ($n = 9$ per group, 5 females in each group). All subjects were unaware of the experimental hypothesis, and provided written informed consent prior to participating. Study protocols were approved by the UBC Behavioural Research Ethics Board.

(a)

(b)

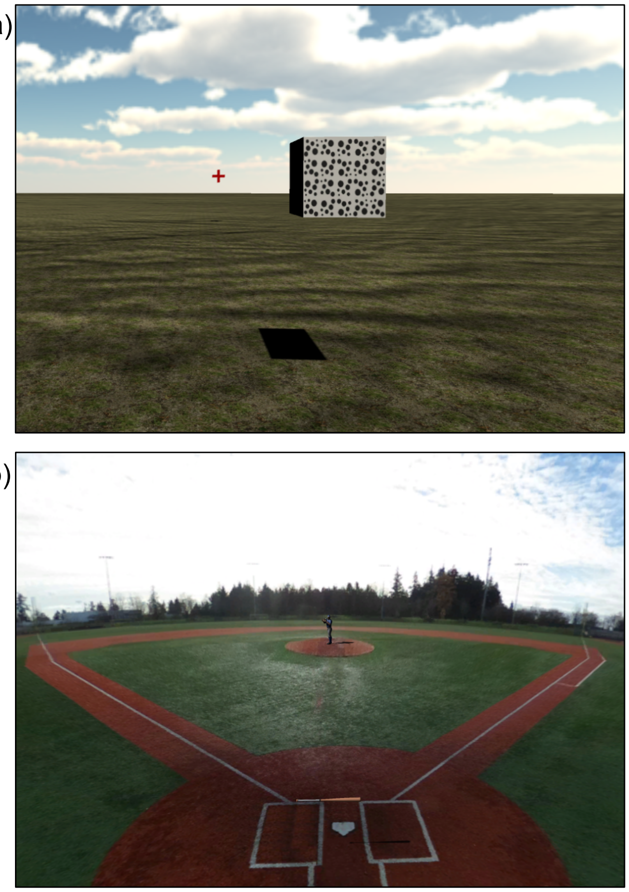

Fig. 3. Virtual environments from the subject's perspective. (a) Calibration task: indicate time of collision with button press. (b) Baseball task: hit ball back to pitcher.

4.2 Virtual Environment

Visual stimuli were presented in an Oculus Rift CV1 headset (Oculus VR, Menlo Park, CA) connected to a computer with a 2.3 GHz processor, 448 GB RAM, and a GTX Titan X graphics card. The virtual visual environment was a custom built application developed in Unity. For the calibration task, the virtual environment was a large open field, consisting of a ground plane, tiled with an image of grass (Fig. 3a). For the baseball task, the virtual environment was a realistic baseball field (Fig. 3b). This environment was captured as a 360-deg photo with a Ricoh Theta S camera. The image was then re-projected onto a hemisphere to provide a ground plane at the appropriate depth. A 3D model of a pitcher, purchased from the Unity Asset Store, was placed over the pitcher's mound.

4.3 Visual Stimuli, Task and Procedure

4.3.1 Calibration Task

A red fixation cross was placed 18 m in front of the subject (Fig. 3a) and presented for 0.5 to 1 sec. Upon disappearance of the fixation cross, a textured cube (side length 30 cm) appeared and moved towards the subject at a speed of either 45, 65, or 85 mph (corresponding to TTCs of 0.90, 0.62 and 0.47 sec). The cube never reached the subject but disappeared at a random distance 4 to 5 m away from the subject. The task was to press a button on the Oculus Touch controller at the estimated TTC with the subject's head. Subjects completed 20 consecutive trials at each speed for a total of 60 trials.

4.3.2 Baseball Task

Subjects were placed standing in the virtual environment in the batter's location and a natural stance that depended on the subject's handedness (e.g., right-handed subjects' left foot forward, and vice versa for lefthanders). Subjects held an Oculus Touch controller that corresponded to a baseball bat in the virtual environment (Fig. 3b). They were asked to hit a looming, pitched ball as directly as possible back towards the pitcher by physically making a swinging motion. The animation of the pitcher was identical across speeds to prevent the subject from using postural cues to infer the ball's speed. A model of the natural flight of a baseball was developed based on parameters (Table 1) derived from the literature [50], [51]. We used an enhanced Euler update with the following equation for forces (F) and acceleration ($\boldsymbol{a}$):

$$\boldsymbol{a} = \frac{1}{m}(F_{gravity} + F_{lift} + F_{drag}),$$

where $F_{gravity} = -g\ m(0\ 1\ 0)^T$, $F_{lift} = \frac{1}{2}\rho C_l \sin(\theta) A\|\boldsymbol{v}\|\boldsymbol{v}$, and $F_{drag} = -\frac{1}{2}\rho\ C_d A\|\boldsymbol{v}\|\boldsymbol{v}$.

TABLE 1
Parameters for Baseball Simulation.

Variable	Description	Value
m	Mass	$0.145kg$
r	Radius	$0.037m$
g	Acceleration of gravity	$9.81m/s$
v	Translational velocity	varies throughout pitch
ω	Rotational velocity	$\{40, 0, 0\}Hz$
ρ	Density of air	$1.204kg/m^3$
θ	Angle between v and ω	varies throughout pitch
A	Cross-sectional area	$0.004m^2$
s	Spin factor	$r * \frac{\|\omega\|}{\|v\|}$
C_l	Lift coefficient	$\begin{cases} 1.5 * s & s \leq 0.1 \\ 0.09 + 0.6 * s & otherwise \end{cases}$
C_d	Drag coefficient	0.3

The ball moved at 40, 60 or 80 mph, directly over home plate, and through the strike zone. The order of pitch velocities was randomized from trial to trial. Subjects completed this task at a self-directed pace. Each trial started with pressing a button on the Oculus Touch controller, initializing the pitching animation.

We instructed subjects to start each swing with the bat pointing behind them, and finish with the bat pointing in front. In a normal baseball scenario, it is possible for the batter to have perfect timing but swing either above or below the ball, thus missing it. Because our focus was on the timing of interception, rather than its spatial accuracy, we extended the collision bounds of the bat to be a large vertical plane (Fig. 4).

Upon successfully hitting the ball, a subject would feel haptic feedback from the controller and see the ball bounce off the bat in the direction determined by the bat's yaw-angle (Fig. 4). The vertical component of the direction was sampled uniformly at random between zero and the z (depth) component of the velocity. This had the effect of producing hits that ranged from low (grounders) to high (pop flies). The total force applied to the ball was made proportional to the speed of the subject's hand during the swing.

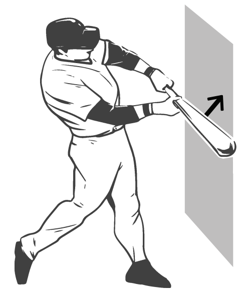

Fig. 4. The collision bounds of the bat were extended vertically. The bounce angle of the ball was determined by the bat's rotation around the vertical axis plus a random vertical component. Image source: shutterstock.com

Subjects completed 50 trials at each speed in randomized order for a total of 150 trials per block; half of the subjects were assigned to a block order in which they received baseline first and then treatment, the other half completed treatment first and then baseline.

4.4 Data Analysis

The azimuth angle of the ball's bounce off the bat was the main variable of interest in our study. We examined the "hitting error," which is defined as the absolute value of the hitting angle. A value of 0 is ideal and would indicate a hit that went directly back towards the pitcher. We expected subjects to perform better in the treatment as compared to the baseline block across all speeds. We also expected subjects to hit more accurately over time, as they got more practice. We thus also analyzed learning rate, i.e. how hit angle changed over the course of the experiment. Linear mixed models were used to analyze hit angle data. These models represent a response variable as a linear combination of predictors, which include both fixed effects (non-random quantities) and random effects, and are well suited for data with multiple correlated measurements per subject.

The overall time course of hit angles was modeled with a line (intercept and slope), a fixed effect of treatment, ball speed, and a fixed interaction between treatment and speed. Following a Growth Curve Analysis procedure [52], these factors were added progressively only when they led to a model that explained the data significantly better (stepwise regression with forward selection). The same results were obtained regardless of the order in which factors were added. The model also included participant random effects on both the slope and intercept terms corresponding to inter-subject variation in starting ability and learning rates on the task. Significance values were obtained from t-tests using the Satterthwaite approximation to degrees of freedom; significance levels were consistent with the Kenward-Roger approximation [53]. All statistical analyses were conducted in R.

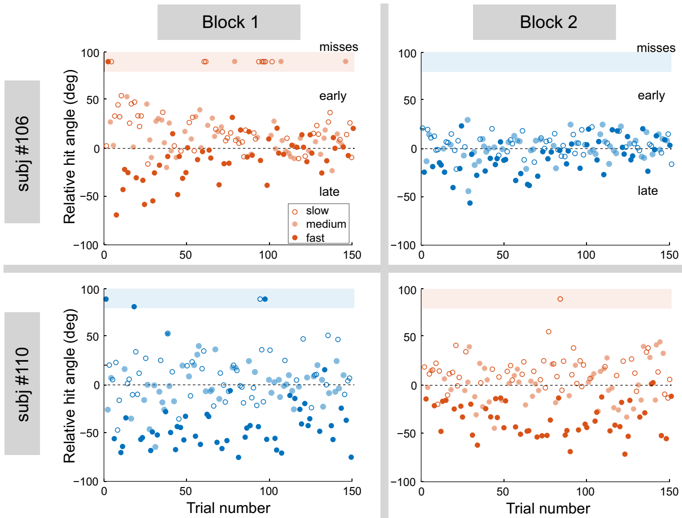

Fig. 5. Relative hit angles (in degrees) as a function of trial number for two representative subjects. Subject 106 completed block 1 with intervention (orange data points), and block 2 without (blue data points). Subject 110 completed blocks in the reverse order, first the baseline condition, then the intervention. Ball speed denoted by symbol type. Negative hit angles indicate late hits, positive hit angles indicate early hits. Misses are coded at 90° hit angle by default, indicated by shaded box at top of each panel.

5 Results

Across all conditions, subjects performed well in the task. On average, they responded with a mean hit angle of M = -3.7°, (std = 33.5°), indicating that the ball was hit slightly late on average. Subjects missed the ball in 8.7% (std = 6.1°) of all trials on average. Hitting performance depended on speed: subjects tended to overestimate balls at slow speed (hit early), and underestimate balls at fast speed (hit late). Subjects performed best at the medium speed (hit angle M = -1.8°, std = 30.1°), as compared to the slow speed (M = 13.5°, std = 26.9°), or the fast speed (M = -23.3°, std = 32.8°).

Variability between subjects was high both in terms of accuracy and precision. Fig. 5 shows two example profiles of hit angles as a function of trial number, for two representative subjects with different block order; each data point is the hit angle in a given trial. These two subjects performed at different overall levels. Subject 106 rapidly improved in hitting accuracy in the first 50 trials, then saturated at a relatively high performance level (M = 6.1° across all speeds) and maintained this level in block 2 (M = -2.4°). Subject 110 performed at a lower level (block 1: M = -14.7°, block 2: M = -10.1°). Results from these two representative subjects also reveal differences in response precision over time, with smaller improvements in subject 110 vs. 106, indicating potential differences in learning rate.

Our statistical models address the main question to what extent performance differences are due to the intervention. To this end, we first consider data from block 1 only, and disregard block 2, where performance improved due to training and repetition across all conditions. The models revealed a significant effect of treatment (i.e., a significant decrease in the intercept term) for the fastest two speeds (60 mph: estimate = -9.2°, SE = 1.81° , $p < 0.001$; 80 mph: estimate = -10.1°, SE = 1.85°, $p < 0.001$; estimates are given

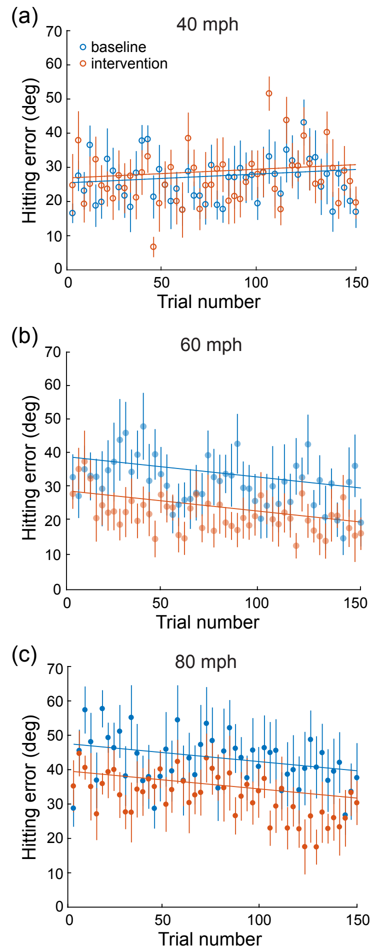

Fig. 6. Hitting error (in degrees) across all observers (n=9 per group) for three speeds. Lines are best fit model fits. Each data point is the average hitting error per trial across observers; each panel shows 50 trials per speed (total of 150 trials per block). Error bars denote standard errors.

relative to the baseline). Increases in speed were found to correspond to higher intercept terms; negative slopes of model fits for the two faster speeds (Fig. 6b,c) indicate improved hitting performance over time. The slowest speed neither showed a significant decrease in hit angle over time, nor any effect of treatment (Fig. 6a). The significant effects of the intervention on TTC performance can be directly inferred by comparing model fits (orange and blue lines) in Figs. 6b,c).

Taken together, these results show that subjects' performance improved after depth cues were modified, but this finding depended on speed. Whereas treatment had no effect at the slowest speed, performance improvements were significant at the two faster speeds.

We next considered all data, and included block and group (treatment first or control first) as fixed-effects in our model. Results show that performance improved significantly from block 1 to block 2 (fixed effect of group, estimate = -4.5°, SE = 0.51°, $p < .001$). Importantly, treatment effects were similar to those obtained with a model including only block 1 data, indicating that treatment effects hold even when considering results obtained after training / experience (60 mph: estimate = -1.1°, SE = 1.23°, $p < .001$;

80 mph: estimate = -4.1°, SE = 1.25°, $p < .002$; estimates are given relative to the baseline).

6 Discussion

Moving and interacting with our dynamic visual environment require a scale for calibration in space and time. In virtual or augmented reality settings, object location and motion paths are generally misperceived. Here we introduced a novel intervention for presenting moving stimuli in VR, relying on an individualized recalibration. Our user study demonstrates that a correction of depth cues significantly affected users' interception performance, resulting in greater accuracy and faster learning at higher stimulus speeds (Fig. 6). Our method compensates underestimation of speed in VR as compared to performance in a baseline condition with non-adjusted cues. Critically, this method is easy to implement in any VR display, and only requires a brief user calibration.

Our model focuses on the size of expansion on the retina as the critical TTC cue. However, additional factors should be taken into consideration when estimating TTC in virtual or real settings. For example, our model used constant values of wind resistance (aerodynamic drag) and gravity, which are known to affect perception and motor performance. When intercepting a fast-moving object, observers rely on an internal model of gravity when predicting where to intercept [54]. Interestingly, observers also assume accelerated motion when intercepting balls approaching at constant velocity in VR [55], a finding with direct implications for our study, which showed balls moving at constant speed.

Our user intervention could incorporate variability in any of these measures, likely resulting in increased task difficulty. Future models could also take into account size and disparity, tailored to different object sizes [19]. Moreover, future methods could be adaptive, taking into account individual variations to determine how many trials would be needed for an accurate estimate of a given user's errors. Our study excluded observers with prior experience playing competitive ball sports, thereby minimizing the effect of expertise in baseline performance. Even though all observers improved from block 1 to block 2 in our study, individual learning rates varied (see Fig. 5 for two examples). Different numbers of trials may be necessary for observers to saturate at a high performance level.

Our perceptual results are similar to those obtained in studies using egocentric distance estimation tasks, reporting underestimation of distance [20], [21], [22], [23], [24], [25] as well as recovery from error when visual-motor feedback is introduced [26], [27]. However, our task differs from standard egocentric distance estimation tasks in several important ways. First, TTC estimation in general requires a combination of different cues that have to be decoded continuously over time as the ball approaches and before it is blanked. Because the visual target is shown for longer periods of time, depth cue adjustments are possible. Egocentric distance estimation usually involves only one view of the target before the observer is blindfolded; no further visual information is available as the observer starts approaching the target. Interventions thus have to rely on feedback provided to the observer. Distance judgments in VR can be corrected by training observers in an environment in which visual-motor feedback during locomotion is manipulated. For example, changing the rate of optic flow (faster or slower than walking speed) during training sessions led to improvements in subsequent distance estimation during blind-walking [26], [27], reflecting humans' ability to rapidly calibrate locomotion. Second, most distance estimation tasks in VR have been conducted in simulated indoor settings. By contrast, our task employed a simulated outdoor setting with rich feature cues providing distance and depth information (e.g., fence, trees). Third, underestimation of speed at long presentation duration in TTC tasks extends to regular displays [18], indicating that perceptual errors cannot be due to restrictions in field of view, depth cue availability or other VR-specific factors. Our intervention does not necessarily require VR displays. Without VR, baseline performance is likely lower due to loss of stereo cues and head tracking, but we expect the intervention could be effective in such settings as well.

When considering applications of our methods in computer graphics the following limitations have to be considered. The object in our calibration task included a sizeable shadow, which could have served as an additional depth or motion cue. For example, previous studies have demonstrated that the perception of object motion (trajectory estimation) was more veridical when the target object cast a shadow vs. when it did not [56]. In addition to manipulating additional TTC cues, future studies could investigate the effect of shadows as depth cues in VR.

Despite these limitations, our method has the potential to improve user experience in dynamic environments requiring short reaction times, such as sports simulations or action video games. Manipulating user experience in VR can have rehabilitative benefits for visual impairments [57], [58]. Virtual environments have gained popularity for training in ball sports [59]. High perceptual fidelity is of fundamental importance in this setting, because users may rely on different 3D motion cues in a virtual environment than in a real scene. Given this challenge, it is not surprising that many VR sports simulations are suboptimal. Our methods could be used to improve such applications.

References

[1] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8, no. 3, pp. 33–33, 2008.

[2] K. Nakayama, "Biological image motion processing: a review," *Vision Research*, vol. 25, no. 5, pp. 625–660, 1985.

[3] T. D. Albright and G. R. Stoner, "Visual motion perception." *Proceedings of the National Academy of Sciences*, vol. 92, no. 7, pp. 2433–2440, 1995.

[4] D. Regan and R. Gray, "Visually guided collision avoidance and collision achievement," *Trends in Cognitive Sciences*, vol. 4, no. 3, pp. 99–107, 2000.

[5] H. Hecht and G. J. Savelsbergh, "Theories of time-to-contact judgment," *Advances in Psychology*, vol. 135, pp. 1–11, 2004.

[6] R. J. Bootsma and R. R. Oudejans, "Visual information about time-to-collision between two objects." *Journal of experimental psychology: human perception and performance*, vol. 19, no. 5, p. 1041, 1993.

[7] G. Savelsbergh, H. Whiting, and R. J. Bootsma, "Grasping tau." *Journal of experimental psychology: human perception and performance*, vol. 17, no. 2, p. 315, 1991.

[8] J. R. Tresilian, "Visually timed action: time-out for tau?" *Trends in cognitive sciences*, vol. 3, no. 8, pp. 301–310, 1999.

[9] J. P. Wann, "Anticipating arrival: Is the tau margin a specious theory?" *Journal of Experimental Psychology: Human Perception and Performance*, vol. 22, no. 4, p. 1031, 1996.

[10] J. B. Smeets, E. Brenner, S. Trébuchet, and D. R. Mestre, "Is judging time-to-contact based on tau?" *Perception*, vol. 25, no. 5, pp. 583–590, 1996.

[11] E. Brenner and J. B. Smeets, "How people achieve their amazing temporal precision in interception," *Journal of Vision*, vol. 15, no. 3, pp. 8–8, 2015.

[12] R. Gray and D. Regan, "Accuracy of estimating time to collision using binocular and monocular information," *Vision Research*, vol. 38, no. 4, pp. 499–512, 1998.

[13] D. Regan and K. I. Beverley, "Binocular and monocular stimuli for motion in depth: changing-disparity and changing-size feed the same motion-in-depth stage," *Vision Research*, vol. 19, no. 12, pp. 1331–1342, 1979.

[14] M. Lages, "Bayesian models of binocular 3-d motion perception," *Journal of Vision*, vol. 6, no. 4, pp. 14–14, 2006.

[15] J. López-Moliner, H. Supèr, and M. S. Keil, "The time course of estimating time-to-contact: Switching between sources of information," *Vision Research*, vol. 92, pp. 53–58, 2013.

[16] M. Zago, J. McIntyre, P. Senot, and F. Lacquaniti, "Visuo-motor coordination and internal models for object interception," *Experimental Brain Research*, vol. 192, no. 4, pp. 571–604, 2009.

[17] M. Manser and P. Hancock, "Influence of approach angle on estimates of time-to-contact," *Ecological Psychology*, vol. 8, no. 1, pp. 71–99, 1996.

[18] A. J. Lugtigheid and A. E. Welchman, "Evaluating methods to measure time-to-contact," *Vision Research*, vol. 51, no. 20, pp. 2234–2241, 2011.

[19] S. K. Rushton and J. P. Wann, "Weighted combination of size and disparity: a computational model for timing a ball catch," *Nature Neuroscience*, vol. 2, no. 2, pp. 186–190, 1999.

[20] J. M. Loomis, J. M. Knapp *et al.*, "Visual perception of egocentric distance in real and virtual environments," *Virtual and adaptive environments*, vol. 11, pp. 21–46, 2003.

[21] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson, "The effects of head-mounted display mechanics on distance judgments in virtual environments," in *Proceedings of the 1st Symposium on Applied perception in graphics and visualization*. ACM, 2004, pp. 35–38.

[22] V. Interrante, B. Ries, and L. Anderson, "Distance perception in immersive virtual environments, revisited," in *Virtual Reality Conference, 2006*. IEEE, 2006, pp. 3–10.

[23] J. A. Jones, J. E. Swan II, G. Singh, E. Kolstad, and S. R. Ellis, "The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception," in *Proceedings of the 5th symposium on Applied perception in graphics and visualization*. ACM, 2008, pp. 9–14.

[24] H. H. Hu, A. A. Gooch, W. B. Thompson, B. E. Smits, J. J. Rieser, and P. Shirley, "Visual cues for imminent object contact in realistic virtual environment," in *Proceedings of the conference on Visualization'00*. IEEE Computer Society Press, 2000, pp. 179–185.

[25] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, and W. B. Thompson, "The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments," *Perception*, vol. 34, no. 2, pp. 191–204, 2005.

[26] A. R. Richardson and D. Waller, "Interaction with an immersive virtual environment corrects users' distance estimates," *Human Factors*, vol. 49, no. 3, pp. 507–517, 2007.

[27] B. R. Kunz, S. H. Creem-Regehr, and W. B. Thompson, "Testing the mechanisms underlying improved distance judgments in virtual environments," 2015.

[28] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," in *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3. ACM, 2004, pp. 804–813.

[29] G. D. Love, D. M. Hoffman, P. J. Hands, J. Gao, A. K. Kirby, and M. S. Banks, "High-speed switchable lens enables the development of a volumetric stereoscopic display," *Optics Express*, vol. 17, no. 18, pp. 15 716–15 725, 2009.

[30] F.-C. Huang, D. Luebke, and G. Wetzstein, "The light field stereoscope," *ACM SIGGRAPH Emerging Technologies*, vol. 24, 2015.

[31] G. Mather, "Image blur as a pictorial depth cue," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 263, no. 1367, pp. 169–172, 1996.

[32] R. P. O'Shea, D. G. Govan, and R. Sekuler, "Blur and contrast as pictorial depth cues," *Perception*, vol. 26, no. 5, pp. 599–612, 1997.

[33] S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks, "Focus cues affect perceived depth," *Journal of Vision*, vol. 5, no. 10, pp. 7–7, 2005.

[34] D. Vishwanath and E. Blaser, "Retinal blur and the perception of egocentric distance," *Journal of Vision*, vol. 10, no. 10, pp. 26–26, 2010.

[35] R. T. Held, E. A. Cooper, J. F. Obrien, and M. S. Banks, "Using blur to affect perceived distance and size," *ACM Transactions on Graphics*, vol. 29, no. 2, 2010.

[36] A. T. Duchowski, D. H. House, J. Gestring, R. I. Wang, K. Krejtz, I. Krejtz, R. Mantiuk, and B. Bazyluk, "Reducing visual discomfort of 3d stereoscopic displays with gaze-contingent depth-of-field," in *Proceedings of the ACM Symposium on Applied Perception*. ACM, 2014, pp. 39–46.

[37] M. Mauderer, S. Conte, M. A. Nacenta, and D. Vishwanath, "Depth perception with gaze-contingent depth of field," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2014, pp. 217–226.

[38] G.-A. Koulieris, B. Bui, M. S. Banks, and G. Drettakis, "Accommodation and comfort in head-mounted displays," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 87, 2017.

[39] N. Padmanaban, R. Konrad, T. Stramer, E. A. Cooper, and G. Wetzstein, "Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays," *Proceedings of the National Academy of Sciences*, vol. 114, no. 9, pp. 2183–2188, 2017.

[40] B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder, "Foveated 3d graphics," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, p. 164, 2012.

[41] A. Patney, J. Kim, M. Salvi, A. Kaplanyan, C. Wyman, N. Benty, A. Lefohn, and D. Luebke, "Perceptually-based foveated virtual reality," in *ACM SIGGRAPH 2016 Emerging Technologies*. ACM, 2016, p. 17.

[42] E. D. Ragan, R. Kopper, P. Schuchardt, and D. A. Bowman, "Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 5, pp. 886–896, 2013.

[43] C. Ware and P. Mitchell, "Reevaluating stereo and motion cues for visualizing graphs in three dimensions," in *Proceedings of the 2nd symposium on Applied perception in graphics and visualization*. ACM, 2005, pp. 51–58.

[44] R. Gray, "Behavior of college baseball players in a virtual batting task." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 28, no. 5, p. 1131, 2002.

[45] 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*, vol. 9, no. 13, pp. 14–14, 2009.

[46] F. T. Zaal and R. J. Bootsma, "Virtual reality as a tool for the study of perception-action: The case of running to catch fly balls," *Presence: Teleoperators and Virtual Environments*, vol. 20, no. 1, pp. 93–103, 2011.

[47] T. Komura, A. Kuroda, and Y. Shinagawa, "Nicemeetvr: facing professional baseball pitchers in the virtual batting cage," in *Proceedings of the 2002 ACM symposium on Applied computing*. ACM, 2002, pp. 1060–1065.

[48] S. Xu, P. Song, C. L. Chin, G. G. Chua, Z. Huang, and S. Rahardja, "Tennis space: an interactive and immersive environment for tennis simulation," in *Image and Graphics, 2009. ICIG'09. Fifth International Conference on*. IEEE, 2009, pp. 652–657.

[49] L. G. Appelbaum and G. Erickson, "Sports vision training: A review of the state-of-the-art in digital training techniques," *International Review of Sport and Exercise Psychology*, pp. 1–30, 2016.

[50] R. Adair, "The physics of baseball," 2002.

[51] G. S. Sawicki, M. Hubbard, and W. J. Stronge, "How to hit home runs: Optimum baseball bat swing parameters for maximum range trajectories," *American Journal of Physics*, vol. 71, no. 11, pp. 1152–1162, 2003.

[52] D. Mirman, *Growth curve analysis and visualization using R*. CRC Press, 2016.

[53] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "Package lmertest," *R package version*, vol. 2, no. 0, 2015.

[54] F. Lacquaniti, G. Bosco, S. Gravano, I. Indovina, B. La Scaleia, V. Maffei, and M. Zago, "Gravity in the brain as a reference for space and time perception," *Multisensory research*, vol. 28, no. 5-6, pp. 397–426, 2015.

[55] M. Russo, B. Cesqui, B. La Scaleia, F. Ceccarelli, A. Maselli, A. Moscatelli, M. Zago, F. Lacquaniti, and A. dAvella, “Intercepting virtual balls approaching under different gravity conditions: evidence for spatial prediction,” *Journal of Neurophysiology*, vol. 118, no. 4, pp. 2421–2434, 2017.

[56] P. Mamassian, D. C. Knill, and D. Kersten, “The perception of cast shadows,” *Trends in cognitive sciences*, vol. 2, no. 8, pp. 288–295, 1998.

[57] P.-Y. Laffont, T. Martin, M. Gross, W. D. Tan, C. Lim, A. Au, and R. Wong, “Rectifeye: a vision-correcting system for virtual reality,” in *SIGGRAPH ASIA 2016 Posters*. ACM, 2016, p. 8.

[58] F.-C. Huang, G. Wetzstein, B. A. Barsky, and R. Raskar, “Eyeglasses-free display: towards correcting visual aberrations with computational light field displays,” *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, p. 59, 2014.

[59] H. C. Miles, S. R. Pop, S. J. Watt, G. P. Lawrence, and N. W. John, “A review of virtual environments for training in ball sports,” *Computers & Graphics*, vol. 36, no. 6, pp. 714–726, 2012.