2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3503-2838-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISMAR59233.2023.00124

Empirical Evaluation of the Effects of Visuo-Auditory Perceptual Information on Head Oriented Tracking of Dynamic Objects in VR

Mark Tolchinsky* Rohith Venkatakrishnan† Roshan Venkatakrishnan‡ Christopher C. Pagano§
Sabarish V. Babu¶
Clemson University
Clemson, SC, USA

Figure 1: The beginning of a feedback-enabled trial. Note the yellow ball indicating the direction of gaze.

ABSTRACT

As virtual reality (VR) technology sees more use in various fields, there is a greater need to understand how to effectively design dynamic virtual environments. As of now, there is still uncertainty in how well users of a VR system are capable of tracking moving targets in a virtual space. In this work, we examined the influence of sensory modality and visual feedback on the accuracy of head-gaze moving target tracking. To this end, a between subjects study was conducted wherein participants would receive targets that were visual, auditory, or audiovisual. Each participant performed two blocks of experimental trials, with a calibration block in between. Results indicate that audiovisual targets promoted greater improvement in tracking performance over single-modality targets, and that audio-only targets are more difficult to track than those of other modalities.

Keywords: perception-action, user studies, head related tracking, perceptuo-motor calibration

1 INTRODUCTION

As virtual reality (VR) technologies continue to develop, they will become increasingly powerful and accessible. Applications in entertainment [23], training [48], and therapy [21, 35] hints at the growing ubiquity of VR. Despite VR's growth, there is still uncertainty surrounding how humans perceive and interact with virtual environments. In some cases, there is a pervasive mismatch between the perception of and state of a virtual environment (VE). For example, distances are systematically underestimated in VR when compared to the real world [34]. One such gap in understanding involves how dynamic targets are tracked in a virtual space.

Tracking a moving target is a task so ubiquitous, the pattern of eye movements designed for the task has its own name – smooth pursuit. From keeping one's "eye on the ball" to tracking a vehicle on the road in order to avoid it, moving target tracking is a staple perceptual task in day to day life. While smooth pursuits focus on fixating on the moving target with the eyes, head movement plays a significant role in the process [26]. In fact, work by Mann, et al., suggests that elite cricket batters use their head to track a ball, and lead it with their eyes [32]. While such a strategy likely takes much practice to develop on a target as fast as a cricket ball, it may prove to be viable when tracking slower moving targets in virtual reality.

Head orientation, or "head gaze", is frequently utilized by VR systems. In particular, HMDs with no external input devices, such as Samsung Gear VR, use head gaze as the primary method of target selection by necessity [41]. Head gaze target tracking has been used in therapy as a guiding mechanism for neck stretches [35]. Other uses of head gaze as an interaction metaphor includes as direction of travel [11]. In fact, some studies suggest that head gaze may outperform eye gaze in tracking and selection tasks in XR environments [19, 22]. Head orientation has also been used as a low cost measure of visual attention in tasks such as traffic crossing, virtual humans and crowds in interactive scenarios in VR [2, 30, 47]. Furthermore, as head gaze does not require any equipment except the HMD itself, it's use would increase the accessibility of VR from an economic standpoint, as well as for those with motor impediments. As such, head gaze may be a good candidate as a metaphor for

*email: mtolchi@clemson.edu
†email: rohithv@clemson.edu
‡email: rvenkat@clemson.edu
§email: cpagano@clemson.edu
¶email: sbabu@clemson.edu

979-8-3503-2838-7/23/$31.00 ©2023 IEEE
DOI 10.1109/ISMAR59233.2023.00124

dynamic object tracking as well.

Visually guided tasks such as moving target tracking can be described in terms of control laws [49]. Control laws are a simple mapping between information from the environment and control parameters of an actor. In other words, control laws are a direct relationship between one's perception and their response to it [50]. As new information is perceived, a control law will transform it into an appropriate action. For example, a control law governing head-gaze tracking of a visible target may be to keep the target in the center of one's field of vision. As such, if the target is not perceived to be there, the control law stipulates the agent's action should be to move their head such that the target ends up closer to the center of the field of vision. If the visual stimulus afforded to an individual is unusual to them (for example, they are in VR), their control laws may not be valid for their environment. If this is the case, they must undergo perceptuo-motor recalibration to generate more effective control laws.

A common technique to facilitate perceptuo-motor calibration is closed-loop feedback [8]. In their work, Mohler et al. [36] define feedback as a stimulus which indicates to what extent an individual accomplished what they set out to do. In a closed-loop feedback system, this stimulus is provided in real time so that the user can adjust their behavior until their actions create the desired effect. Visual feedback has been shown to improve performance in various tasks, such as distance judgements in VR [1, 15, 36]. However, there has been no concerted effort to evaluate whether closed-loop feedback can help calibrate user tracking of a moving target in VR, and specifically how visuo-auditory feedback during calibration enhances head oriented tracking performance.

When presented with multisensory stimuli, humans generally prioritize visual input over those from other modalities [12]. However, research has shown that using multiple/different sources of sensory information can alter overall perception [24, 33]. In particular, multisensory stimulus can be integrated in order to more richly perceive its source [25]. On static target localization tasks, audiovisual stimuli were found to enable better performance when compared to stimuli that were audio or visual only. For example, work by Tannen et al. reveals superior localization performance for multisensory targets in a flight task [44]. Additionally, Hairston et al. have shown that under induced myopia, multisensory targets were localized more often than visual only targets [20]. As such, the use of additional sensory modalities can help increase the accessibility of VR.

In this work, we seek to investigate how smooth pursuit of a moving target is performed in virtual reality. By varying the target's sensory modality between subjects, we examine how visual dominance and audiovisual binding affect tracking performance. Furthermore, we evaluate if calibration via closed-loop feedback can be used to improve tracking performance.

2 Related Works

2.1 Multisensory Integration

Multisensory integration is a process by which input to multiple senses is combined into a single, coherent understanding. A robust body of work details the existence and mechanisms behind this phenomenon, particularly in audiovisual stimuli [33, 43]. However, there is also research to suggest that visual stimuli can dominate those of other senses [12], so whether or not binding of vision and audio influences perception-action coordination is still uncertain.

Hairston et al. examined the effect of stimulus modality on localization performance in the real world, for both visually-abled and myopic individuals [20]. Some participants were given myopia-inducing lenses to degrade their visual acuity. In all cases, participants were significantly worse at localizing targets which had no visual component. However, participants with induced myopia performed significantly better when localizing audiovisual targets than those who had vision-only. Accuracy for audiovisual targets was similar whether or not a participant was myopic, which was not the case for visual-only targets. Although the effects of audiovisual integration may not be immediately evident, the brain still makes use of it in situations where vision is degraded.

In VR, the effect of multisensory integration can be more easily studied as it is possible to arbitrarily manipulate the environment. Work by Yang et al. investigated how performance in a moving target selection task was affected by multisensory stimuli [51]. Participants were tasked with hitting a shuttlecock in a badminton scenario while provided indicators of varying input modalities to assist them. A visual indicator enabled better performance than other unimodal conditions, as expected. However, contrary to visual-dominance theory, adding audio or haptic information improved performance as compared to a vision-only condition.

2.2 Moving Target Tracking

Tracking a moving target is a highly ubiquitous activity in dynamic scenarios. For example, athletes are expected to have a keen ability to track very quick targets. Mallek et al. conducted a study wherein the performance of novice and experienced athletes in a target tracking task was compared [31]. Participants were instructed to track a moving target on a screen by using a stylus on a tablet. Tracking performance was generally superior for experienced athletes, suggesting that target tracking by users is an acquired skill.

Work on the efficacy of head-oriented tracking is significantly less common. Leung et al. examined how well participants could track moving multisensory targets in the real world using only head rotation [27]. They found that for both initial localization and tracking, audiovisual and visual only targets were not significantly different, but audio only targets facilitated much worse performance. Additionally, performance using audio-only targets degraded much more when the target moved faster as opposed to the other two conditions. However, they note that the audio and visual stimuli are generated using different systems, and so may not be sufficiently similar to enable audiovisual binding to improve performance.

2.3 Calibration and Closed-Loop Feedback

Perceptuo-motor calibration is the process of learning that is facilitated via a task in which participants' actions are scaled/calibrated by providing corrective feedback [13, 42, 45]. Studies show that calibration to perceptual information can occur relatively quickly when individuals have access to closed-loop interaction with the environment [1, 16]. Research on this front has shown that users' perceptual judgments and interaction performance can be improved after calibration or attunement [5, 6, 8, 40, 46]. Each individual calibrates to their surroundings, creating their own perception-action system. If either action potential or sensory inputs are manipulated, the prior calibration may be rendered ineffective, and the perception-action system must be recalibrated for the individual to effectively act upon what they perceive. Displacement prisms are a common method for perturbing visual stimulus [7]. In such experiments, participants perform tasks while looking through prisms that distort their view. While the visual information participants receive does not match up with their actions in a way they are used to, it is still self-consistent. As such, over time, participants recalibrate to use the new visual stimulus to accurately perform their tasks.

Closed-loop feedback is a common method for recalibration of perceptuo-motor tasks in both the real world and VR [14, 28]. This paradigm is often used to calibrate distance judgements in VR where they are inaccurate without training. For example, work by Mohler et al. used various forms of closed-loop feedback to improve distance estimation of walking in VR [36]. In particular, visual feedback provided before and after a blind walk tended to improve the accuracy of future blind walks. Ebrahimi et al. found a similar result when investigating depth judgements via physical reaches [15]. While open-loop calibration led to an overestimation of distances

when reaching, closed-loop calibration enabled participants to make much more accurate distance estimations. Other works have also investigated calibration of depth perception with respect to auditory information [28].

However, despite the broad body of work on training via calibration, there has been no research specifically focused on using the paradigm to train head-based tracking. Furthermore, the influence of sensory modality on moving target tracking performance has not been thoroughly investigated, especially in VEs. Finally, moving target tracking tends to be examined only as a facilitator of selection or localization tasks, with little work going towards evaluating the performance of tracking in and of itself. As such, this work will focus on the effects of sensory input modality, closed-loop feedback calibration, and target speed on the performance of head-based tracking in virtual reality.

3 Research Questions and Hypotheses

Our study was pre-approved by our University's Institutional Review Board. In this empirical evaluation, we compared and contrasted the effects of different perceptual channels of information (visuo-auditory, vision only or audio only) on participants' dynamic target tracking ability using head oriented tracking in VR. Additionally, we also compared and contrasted the effects of calibration or attunement through visual feedback of the participants' tracking location information relative to the dynamic target location on the participants' head oriented tracking performance in VR. Thus, our research questions were as follows:

RQ1: To what extent does the different perceptual modalities of information (conditions - visuo-auditory, vision only and audio only) affect participants' tracking performance?

RQ2: To what extent does the participants' tracking performance improve due to calibration or learning?

RQ3: To what extent does the velocity of the targets affect participants' tracking performance in the different perceptual conditions?

Our hypothesis were as follows:

H1: Participants' tracking performance is expected to be better in visuo-auditory condition, as compared to vision only, which in-turn is expected to be better than the audio only condition.

H2: Participants' tracking performance is expected to be enhanced or improved as a result of calibration or learning.

H3: Participants' tracking performance is expected to be better in lower speed target trials as compared to high speed target trials.

Regarding **H1**, audiovisual binding has a significant effect on how humans perceive their environment [10]. As such, we expect that participants' tracking performance will be superior in the condition with audiovisual information pertaining to the target than vision only or audio only. From a vision perspective, we are evaluating to what extent participants can track a moving target in their center of vision. Whereas from a auditory perspective, we are evaluating to what extent participants can utilize the binaural information to track a moving target with the center of their head orientation. Regarding **H2**, research has shown that visuo-motor calibration enhances perception-action (i.e. depth and size perception) [1, 17]. Thus, we hypothesize that overall calibration with visuo-auditory feedback will improve participants' head oriented tracking performance in all the conditions. Regarding **H3**, we intend to search for a relationship between target speed and tracking performance. We expect higher target speeds to produce inferior performance because faster targets are more difficult to track.

4 System and Experiment Design

4.1 Apparatus

This experiment utilized a VE in which participants performed tasks. The environment was displayed on an HTC Vive Pro Eye with a refresh rate of 90 Hz, connected to a desktop computer workstation with a dedicated NVIDIA 2060 graphics card. Participants were asked to stand for the duration of the experiment.

4.2 Participants

A total of 30 participants (18 female) were recruited for this study ranging from 18-24 years of age. All participants were recruited from Clemson Univerity's SONA pool. As inclusion criteria, all participants were required to have normal or corrected to normal vision, and not be hard of hearing. 27 of the participants reported having less than 5 hours of VR experience. We also ensured that participants did not have any motor impediments to head and body motion.

4.3 Simulation Design

Figure 2: The simulation environment as viewed from above.

The VE used for this experiment consists primarily of a road in a rural environment (see figure 2). During trials, the participant stands 6m away from the center (or 4m away from the edge) of a 4m wide main road perpendicular to their initial facing direction. A second road, parallel to the participant's initial facing direction, helps them keep their bearings in the scene. These two roads intersect directly in front of the participant. The main road is terminated at both ends by tunnels from which targets emerge from and disappear into. These tunnels are located 40m away from the origin, so the participant can see 80m of the main road in total. From one tunnel to the other spanned an angle of 160°, encompassing the tracking of a target from side to side. This angular range was chosen as it was determined to be less than the maximum comfortable range of side to side cervical spinal (neck) rotation of 170° (maximum rotation is 180°), as per human biomechanics of head rotation [29, 52].

The moving targets in this simulation took the form of motorcyclists riding at various speeds along the main road. The target would emerge from one of the two tunnels at random, and move directly forward at a constant velocity until withdrawing into the opposite tunnel. Participants were instructed to track the rider who sat upright on the center of a motorbike, as the target of tracking. Each target could travel at one of 3 speeds to ensure that participants do not get too accustomed to the tracking task, and to provide a reasonable variation in target speed. In conditions where the target is audible, it makes the noise of a motorcycle engine. The audio is fully spatialized via interaural time difference and SteamVR's default HRTF, attenuated by distance, and transformed to imitate the natural Doppler effect using SteamVR's physical engine [37]. In all cases where the target was both visible and audible, the sensory information was congruent.

4.4 Procedure

After the informed consent process, each participant's hearing acuity was measured using the Widex online hearing test to ensure that they could hear well enough out of each ear to spatialize audio. The participant was then asked to fill out a demographics questionnaire which included questions about prior VR experience. Then, the participant's interpupillary distance was measured via the Dotty Eye Measure iOS app and used to adjust the HMD. At this stage, the experimenter briefed the participant on their task, informing them that they were to use the sensory information provided by the target to track it with their head to the best of their ability. At this stage, the participant donned the HMD and was allowed to familiarize themselves with the environment. To encourage presence, the participant's first task was to walk forward to a designated marker

on the ground of the VE before trials began. Once the participant was standing on the marker, they used their head gaze to input an ID on a diegetic numpad after which the trials began.

Each participant was randomly assigned one of 3 conditions (audiovisual, visual, audio). Participants in the audiovisual condition received targets that were both visible and audible. Participants in the visual condition received targets that were visible but silent, and those in the audio condition received targets that were audible but invisible. The experimenter informed each participant what kind of target they would be receiving. Although the target is not visible in the audio condition, the visual environment is still shown to the participant. The VE was presented even when the target cannot be seen moving through it in order to maintain consistency between different conditions and phases, as the environment must be presented when the target is visible. Furthermore, the environment contains information about the target's travel, such as the road upon which it travels and the tunnels on either end that may help participants initiate tracking. The lack of this visual information may degrade performance even when the target is not visible. Before the first phase, the experimenter informed each participant how they were expected to track the targets. The instruction was to "imagine a ray going forward from between your eyes, and attempt to always hit the target with the ray". They were also explicitly informed that they were not to track the target with their eyes.

The trials were split up into 3 phases of 30 trials each, with a break between the first and second phase. The first and third phases were identical, and will be referred to as "pre-test" and "post-test". The second phase, called "calibration", provided participants with visual feedback in the form of a yellow sphere indicating where their head gaze is pointing. During the calibration phase, the participant could use the sphere to adjust their head gaze so that it matches with the position of the target. In addition, participants in all conditions were provided with audiovisual targets during the calibration phase.

At the start of each phase, participants were given 3 practice trials in order to become accustomed to the new conditions. During each trial, the target would travel along the main road one time, moving either right to left or left to right (see figure 3). Before the appearance of the target, the participant would see an arrow pointing toward direction it would emerge from, and hear a sound localized at the target's origin. In addition, in each trial the target would travel at one of 3 constant speeds: 20, 30, or 40 mph in the pre- and post-tests, and 25, 35, or 45 mph during the calibration phase. These trials occurred in random order, but were counterbalanced so that there was an equal number of each type of trial throughout the phase.

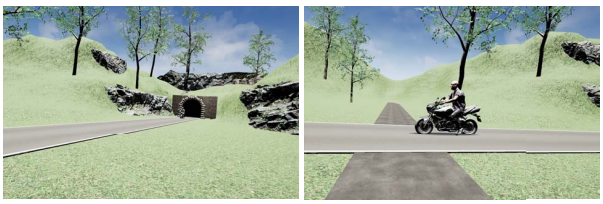


Figure 3: The progression of a normal trial.

4.5 Data preparation

During trials, the simulation recorded the position and orientation of both the vehicle and the participant's head on every frame. Prior work with similar tasks has shown that tracking tends to begin with an onset phase where the target is localized [27]. In order to isolate the periods of time when the subject had already acquired the target and was actively tracking it, we used a polyline splitting algorithm (see figure 4) [4]. The tracking data was compared to a straight line, and once the distance between them crossed a threshhold, the line was split into two lines to include that point. This process was repeated until the polyline approximately matched the tracking data, upon which two inflections could be identified, representing the initialization and conclusion of active head tracking. For regression analysis, the data was further transformed such that the target angle is negative when moving towards the subject, and positive when moving away. As such, the direction of target travel in each trial is made irrelevant in regression.

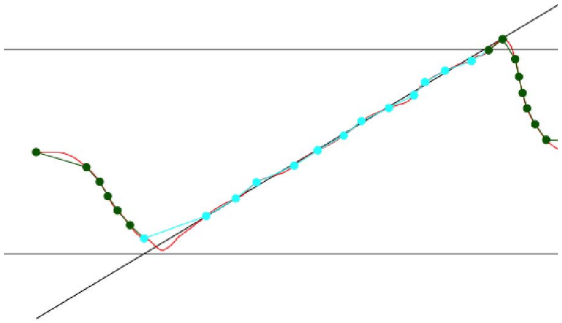


Figure 4: A time v position graph demonstrating polyline splitting applied to a single trial, where the red curve is the participant's gaze, and light blue represents tracking.

4.6 Measures

Once the data was filtered to only include periods of active tracking, several measures were taken and averaged by trial. Note that several of these measures use the concept of a gaze point in their calculation. The gaze point is defined as the intersection between the participant's gaze vector and a Y-Z (upright) plane passing through the origin. In other words, the gaze point is located where the participant's gaze vector passes over the center of the main road on which the target travels. PointY and PointZ refer to the point's position on this plane on the Y and Z axes, respectively. VehicleY refers to the target's position on the Y axis of this plane.

Tracking time is the difference in seconds between the time when tracking began and when tracking concluded, as determined by the polyline algorithm.

$$TT = t_c - t_b$$

Proportional tracking time is calculated by dividing the tracking time of each trial by the amount of time the target was perceptible during said trial. The audio was attenuated such that the target was visible and audible for the same duration of each trial.

$$PTT = \frac{TT}{t_p}$$

Tracking latency is the difference in seconds between the target's presence in a location and the gaze point's presence in the same location. During each frame of tracking, the target was present in a unique location. This same location must have been visited by the participant's gaze at some point during the trial; the difference in these times is the latency.

$$TL = avg(t_{targetatN} - t_{pointatN})$$

Unsigned polar error is the absolute value of the angular difference, in degrees, between the vector from the participant's head to the target, and the gaze vector. This value is calculated by treating these two vectors as the hypotenuses of right triangles, with a third point at the origin.

$$UPS = avg(|\frac{tan^{-1}(PointY/PlayerX)}{tan^{-1}(VehicleY/PlayerX)}|)$$

Path length is the distance (cm) that the participant's gaze point travelled during tracking. The differences in the position of the gaze point between each set of consecutive frames is added together.

$$PL = \sum_{n=0}^{N} \sqrt{(PointY_{n+1} - PointY_n)^2 + (PointZ_{n+1} - PointZ_n)^2}$$

Tracking velocity is the mean velocity, in centimeters/second, of the gaze point during tracking.

$$TV = PL/TT$$

Proportional tracking velocity is calculated by dividing the tracking velocity of each trial by the vehicle velocity during that trial.

$$PTV = TV/VV$$

5 Results

On all the quantitative objective data, parametric ANOVA analyses were conducted on the data after carefully verifying that the underlying assumptions were met – namely the data in the samples were normally distributed and error variance between samples were equivalent. We ensured that Box's test of equality of covariance matrix was not significant. Levene's test was conducted to verify homogeneity of variance, and Mauchly's test of sphericity was conducted to ensure that the error variance in groups of samples was equivalent. Pairwise post-hoc tests between levels of the between-subjects variables was conducted using Tukey's HSD analysis, whereas between levels of the within-subjects variables was conducted using the Bonferroni adjusted alpha method. Greenhouse-Geisser correction and adjustment to degrees of freedom were applied when Mauchly's test of sphericity was violated.

After verifying that the assumptions were met, we subjected each quantitative objective measure to a 3 speed (low, medium, high) x condition (audiovisual, vision only, audio only) x phase (pre-test, calibration, post-test) mixed model ANOVA analysis. The within-subjects variables were phase and speed of the target, and the between-subjects variable was condition. Post-hoc pairwise comparisons for the between-subjects variables (condition) were conducted using Tukey's HSD test while those for the within-subjects variables (phase, speed) and their interactions were conducted using the Bonferroni method. These effects are annotated in each figure along with their significance levels.

5.1 Tracking Time (s):

The ANOVA analysis revealed a significant main effect of condition F(2, 25) = 5.78, $p = 0.009$, part. $\eta^2 = 0.32$, phase F(2, 100) = 36.76, $p < 0.001$, part. $\eta^2 = 0.59$, and speed F(1.27, 65.85) = 587.80, $p < 0.001$,part. $\eta^2 = 0.96$. The ANOVA analysis also revealed a significant speed x phase interaction effect F(2.63, 65.85) = 3.07, $p = 0.04$, part. $\eta^2 = 0.11$. See figures 5a and 5b.

5.2 Proportional tracking time:

The ANOVA analysis revealed significant main effects of condition F(2, 25) = 6.110, $p = 0.007$, part. $\eta^2 = 0.33$, phase F(2, 100) = 120.591, $p < 0.001$, part. $\eta^2 = 0.83$, and speed F(1.629, 85.384) = 4.578, $p = 0.022$, part. $\eta^2 = 0.16$. Additionally, the analysis revealed a significant phase x speed interaction effect F(3.415, 85.384) = 4.386, $p = 0.005$, part. $\eta^2 = 0.15$. See figures 5c and 6a.

5.3 Tracking latency (s):

The ANOVA analysis revealed significant main effects of phase F(2, 100) = 168.31, $p < .001$, part. $\eta^2 = 0.87$, and speed F(1.267, 100) = 78.685, $p < .001$, part. $\eta^2 = 0.76$. Additionally, the analysis revealed significant interaction effects in phase x condition F(4, 100) = 2.615, $p = 0.046$, part. $\eta^2 = .17$, phase x speed F(2.425, 100) = 2.709, $p < .001$, part. $\eta^2 = .55$, and phase x speed x condition F(4.851, 100) = 2.493, $p = .042$, part. $\eta^2 = .17$. To explore the three-way interaction effect, we employed a block analysis using the phase x condition and phase x speed interaction effects. See figures 6b and 6c.

5.4 Unsigned polar error (°):

The ANOVA analysis revealed significant main effects of phase F(2,100) = 99.520, $p < 0.001$, part. $\eta^2 = 0.80$, and condition F(2, 25) = 13.034, $p < 0.001$, part. $\eta^2 = 0.51$. Additionally, the analysis revealed significant interaction effects in phase x condition F(4, 100) = 6.338, $p < 0.001$, part. $\eta^2 = 0.34$ and phase x speed F(1.700, 100) = 3.472, $p = 0.047$, part. $\eta^2 = 0.12$. See figures 7b and 7a.

5.5 Unsigned error (cm):

The ANOVA analysis revealed a significant main effect of phase F(2, 100) = 146.259, $p < 0.001$, part. $\eta^2 = 0.85$, as well as significant interaction effects between phase x speed F(2.826, 70.638) = 3.695, $p = 0.017$, part. $\eta^2 = 0.12$ and phase x speed x condition F(5.651, 70.638) = 2.624, $p = 0.026$, part. $\eta^2 = 0.17$. To explore the three-way interaction effect, we used a block analysis using the phase x condition and phase x speed interaction effects. See fig 7c and 8a.

5.6 Path Length (cm):

The ANOVA analysis revealed a significant main effect of phase F(2, 100) = 257.50, $p < 0.001$, part. $\eta^2 = 0.91$. In addition, the analysis revealed significant interaction effects between phase x condition F(4, 100) = 2.85, $p = 0.033$, part. $\eta^2 = 0.19$ and phase x speed F(3.526, 100) = 3.084, $p = 0.025$, part. $\eta^2 = 0.11$. See figures 8b and 8c.

5.7 Proportional tracking velocity:

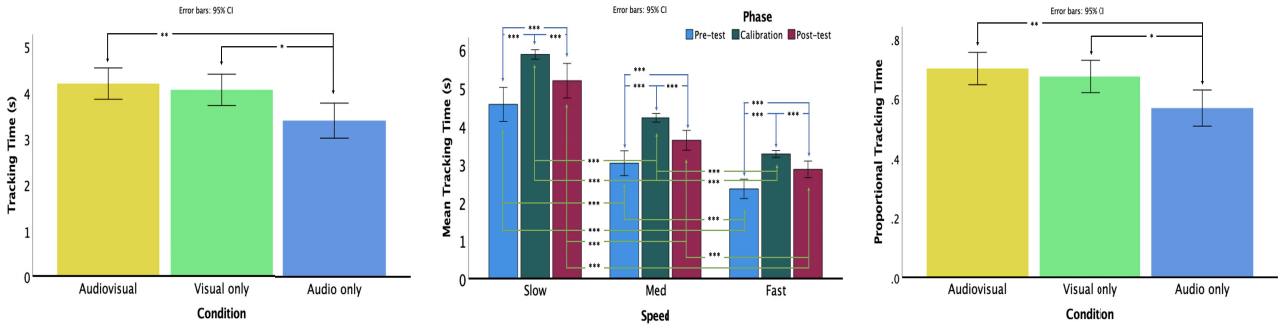
The ANOVA analysis revealed a significant main effect of phase F(1.534, 69.501) = 24.609, $p < 0.001$, part. $\eta^2 = 0.50$. Additionally, the analysis revealed a significant interaction effect in speed x condition F(4, 100) = 2.681, $p = 0.04$, part. $\eta^2 = 0.18$. Post-hoc pairwise comparisons on the speed x interaction effect were not significant.

5.8 Regression Analysis

Regression analysis was conducted to examine the relationship between actual target angle and mean signed polar error between the participant's head gaze and the target location. The mean signed polar error is the angular difference in degrees between the vector from the participant's head to the target and the gaze vector. The formula for the mean signed polar error is described below:

$$signedpolarerror = avg(\frac{tan^{-1}(PointY/PlayerX)}{tan^{-1}(VehicleY/PlayerX)})$$

Regression analysis has been preferred in classical perception and motor control research in protocols in which researchers need to predict a continuous dependent variable (mean signed polar error) from a continuous independent variable (target angle) between different conditions or sessions in an experiment, some example of which in research in the virtual world and real world include [3, 9, 18, 38, 39]. One of the contributions of the regression analyses is that researchers can use the regression equations to predict unseen users' signed polar error of head oriented tracking to a target object's angle in visuo-auditory, vision only and auditory only conditions, as well as in pre-calibration, calibration and post-calibration trials. Also, slopes and intercepts given by the regression equations are more useful than other descriptive statistics, as they describe the lawful function that predicts the participants' signed polar error from the target's angle in the different perceptual conditions, and pre-, during, and post- learning situations of the experiment. Cubic regression models seemed to fit the data the best, as compared to other linear and non-linear models.

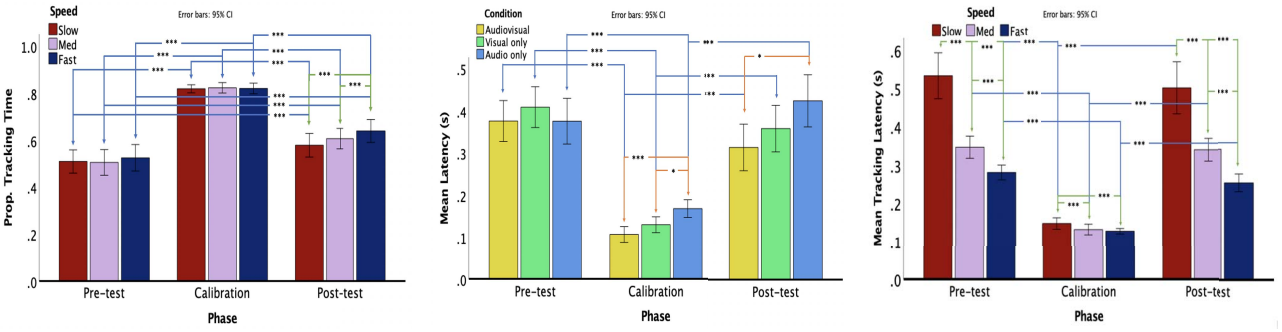


(a) Mean tracking time by condition

(b) Mean tracking time in Phase by Speed interaction

(c) Mean proportional tracking time by condition

Figure 5



(a) Mean proportional tracking time in speed by phase interaction

(b) Mean latency in Phase by Condition interaction

(c) Mean latency in Phase by Speed interaction

Figure 6

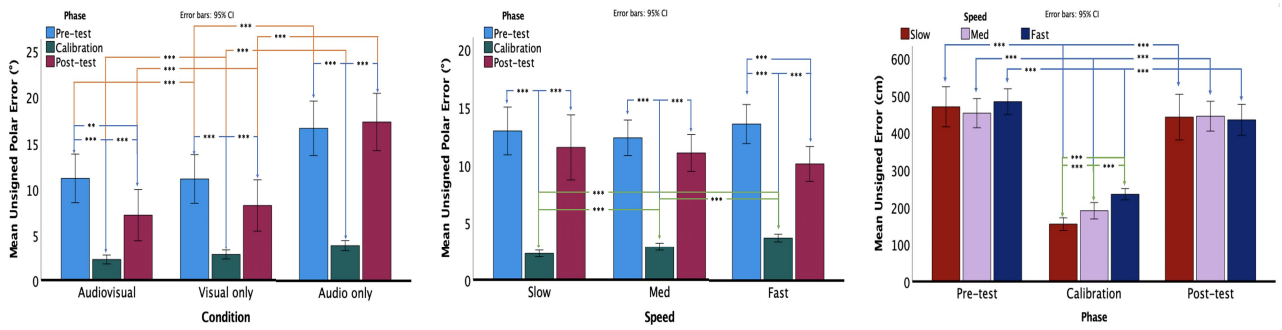
The cubic regression models by condition were as follows. In the visuo-auditory binded condition, the participants' $MeanSignedPolarError = -0.21 \times VehicleAngle - 0.00023 \times VehicleAngle^2 + 0.000028 \times VehicleAngle^3 - 3.0$ [$R^2 = 0.47$], in the vision only condition the participants' $MeanSignedPolarError = -0.22 \times VehicleAngle - 0.00064 \times VehicleAngle^2 + 0.000033 \times VehicleAngle^3 - 2.57$ [$R^2 = 0.47$], and in the audio only condition the participants' $MeanSignedPolarError = -0.40 \times VehicleAngle - 0.0012 \times VehicleAngle^2 + 0.00006 \times VehicleAngle^3 + 2.0$ [$R^2 = 0.54$] (See Figure 9).

The cubic regression models by session were as follows. In the pre-test session, the participants' $MeanSignedPolarError = -0.35 \times VehicleAngle - 0.00048 \times VehicleAngle^2 + 0.000042 \times VehicleAngle^3 - 3.3$ [$R^2 = 0.67$], in the calibration session the participants' $MeanSignedPolarError = -0.13 \times VehicleAngle - 0.0014 \times VehicleAngle^2 + 0.000027 \times VehicleAngle^3 + 3.64$ [$R^2 = 0.71$], and in the post-test session the participants' $MeanSignedPolarError = -0.32 \times VehicleAngle - 0.000056 \times VehicleAngle^2 + 0.000046 \times VehicleAngle^3 - 4.6$ [$R^2 = 0.54$] (See Figure 10).

6 Discussion

Our first hypothesis was that tracking performance would be best in the audiovisual condition, and better in the visual only condition than the audio only condition. This hypothesis was partially supported; while the audio only condition consistently saw the worst performance, audiovisual did not seem to offer significant improvement over vision only. Analysis of polar error revealed that across all phases of the experiment, tracking in the audio-only condition was significantly less accurate, than in the other conditions. In particular, the average polar error during the post-test phase was over twice as high in the audio only condition, as compared to the others. Furthermore, tracking time was shortest in the audio condition, indicating that auditory information may not have been enough to track the target at extreme angles. Finally, the audio condition performed worse than the audiovisual condition in latency and path lengths metrics during post-test trials. Even after attempts at recalibration, participants seemed unable to use only audio information to track. However, there was almost no difference in performance between the audiovisual and vision only conditions throughout all phases. Thus, there appeared to be no performance-enhancing effect of audiovisual binding in this study. These findings are consistent with similar studies conducted in the real world, wherein audio only performance was worst and audiovisual performance, which was not better than visual only [20, 27]. This suggests a similar minimal effect of binding in VR when compared to the real world. Hairston et al. found that the impact of audiovisual binding is greater when visual stimulus is degraded in the real world, which may later be thoroughly investigated in VR.

Our second hypothesis was that tracking performance would be improved as a result of a closed-loop calibration phase. This hypothesis was partially supported. Especially in the audiovisual condition, tracking performance was significantly better in the post-test as compared to the pre-test. On average, polar error was reduced by approximately 4° after calibration in the audiovisual condition. While not quite statistically significant, a similar trend was observed in polar error in the visual condition. Across all conditions, tracking time was longer in the post-test compared to the pre-test, suggesting that training may have improved performance at more extreme angles. However, while metrics indicating accuracy generally improved, there were few significant effects of training on efficiency

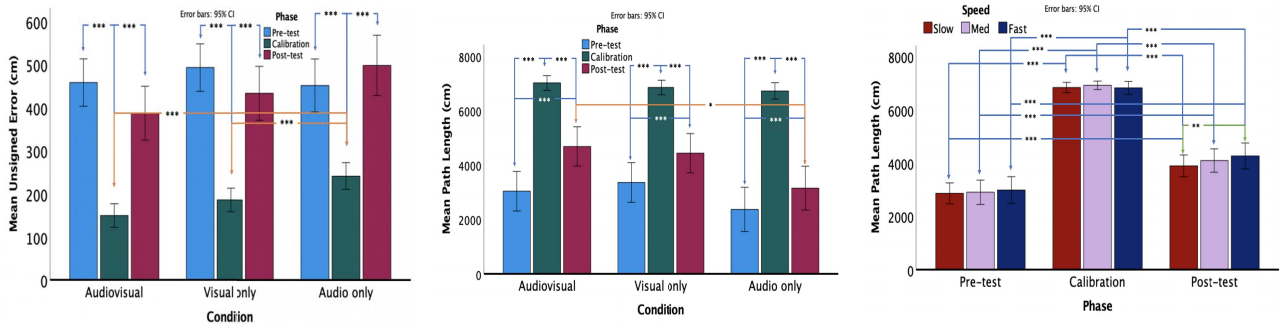


(a) Unsigned polar error in Phase by Condition interaction

(b) Unsigned polar error in Phase by Speed interaction

(c) Unsigned error in Speed by Phase interaction

Figure 7



(a) Unsigned error in Phase by Condition interaction

(b) Path length in Phase by Condition interaction

(c) Path length in Phase by Speed interaction

Figure 8

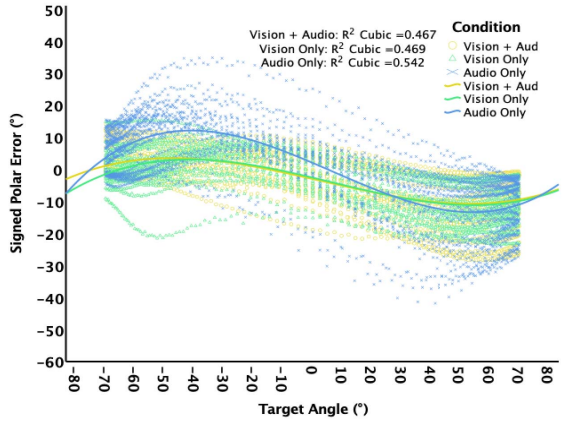


Figure 9: Cubic regression graph of signed polar error by target angle by condition.

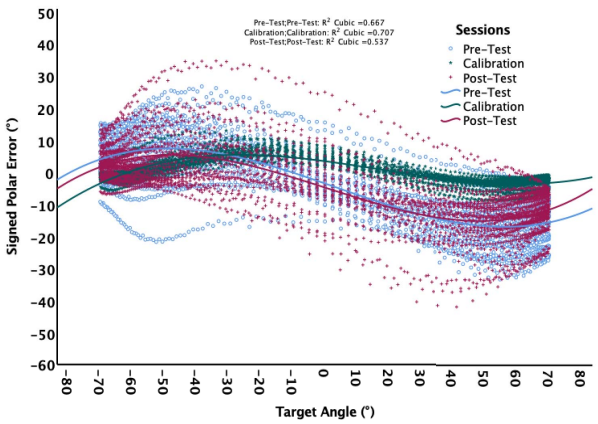


Figure 10: Cubic regression graph of signed polar error by target angle by session.

metrics such as latency and velocity. There is no significant body of work examining how training affects head gaze tracking, so further study may be necessary to determine to what extent this task can be calibrated.

Our third hypothesis was that tracking performance would be adversely affected by an increase in target speed. This hypothesis was not supported. Most notably, polar error was generally not affected by target speed, except during the calibration phase. When there was visual feedback, participants may have been tracking by attempting to align the feedback sphere with the target; a faster target means it is harder to align the visual elements. However, as this effect did not carry over to trials with no feedback, it may be possible that participants tracked with a different technique, such as using the target's velocity to predict its future position rather than attempting to align with its present position. Furthermore, speed had little influence on proportional tracking time or proportional tracking velocity during trials with no feedback. In fact, the proportional tracking time during the post-test was closer to 1 for fast trials than

slow ones, indicating the participants tracked the targets for a longer portion of the trial. These findings conflict with Leung et al., which found that for a similar task in the real world, an increase in speed did degrade performance [27]. However, our study used a narrow range of speeds, so it is possible our work did not contain trials difficult enough to degrade performance.

The regression analysis provided some interesting insights into the relationship between target angle and the signed polar error between perceptual conditions and phases. In all cases, the local maximum occurs when the target angle is negative, and the local minimum occurs when the target angle is positive. Across conditions and phases, participants would tend to lead the target (positive error) as it was traveling towards them, and trail the target (negative error) as it moved away from them. Furthermore, the intercepts of most of the models were negative, indicating that participants would generally begin to trail the target before it traveled past them. Across conditions, the models support the finding that tracking performance was worst in the audio only condition. The local maxima of the audiovisual and visual models have signed polar errors of 3.4 and 3.1 respectively, whereas the audio model's local maximum has a y value of 12.2. Similarly, the minima of the audiovisual and visual models have signed polar errors of -10.6 and -11.1, whereas the audio model's has a y value of -13.7. These greater errors at the extremes of the model are indicative of worse performance from the audio only participants, while the audiovisual and visual models are relatively similar. While the audio participants performed worse at the extremes, the similarity of intercepts between all the models suggests that audio only may be sufficient for tracking targets in a narrow band of angles near the origin. Across phases, the models seem to indicate that performance during calibration was superior, and post-test performance may be slightly improved over pre-test performance. The calibration model's local maximum and minimum signed polar errors of 5.6 and -3.4 are much lower magnitude than the other models'. The calibration's local minimum is notably close to zero compared to the other models', indicating that the recalibration process may have particularly helped participants avoid trailing the target as a trial progressed. The post-test model's extrema (5.5 & -15.0) are lower in magnitude than those of the pre-test's (7.7 & -17.0), suggesting a potential improvement in performance overall.

This work has provided us with some interesting insights about users' abilities to head-track a moving target in immersive VR. With respect to the target's sensory information, as expected, users seem to be less accurate in tracking such moving targets when tracking is to be performed solely based on spatial auditory information. Interestingly, additionally providing audio information does not seem to significantly improve how accurately users are able to track a target that they already have visual information of, once the target is captured. In other words, audio-visual binding may not be a crucial requisite for accurately tracking a moving target when users already have visual information pertaining to the target. We also find that users can perceptually improve their performance in tracking a moving target when there is visual information of the target and closed-loop visual feedback is provided, enabling perceptual calibration under such circumstances. We further learned that the speed of the target does not seem to drastically influence users' performance in tracking a moving target. Our results suggest that the sensory information pertaining to a moving target has a more significant impact on target-tracking than the speed with which the target moves. Overall, these findings are relevant to VR designers and developers, informing them about aspects to consider when designing virtual experiences that support the smooth pursuit of a moving target. Apropos of this, our findings highlight the effects of the target's sensory information and its speed, further demonstrating the existence of a perceptual calibration aftereffect that shows how closed-loop feedback can be utilized to improve smooth pursuit.

6.1 Limitations

As one of few works in this area of research, our study was somewhat limited in scope. First of all, we used a relatively narrow band of speeds for our moving targets (20-40 mph). There are some contexts, such as in aviation or athletic training, where moving targets are considerably faster than those used in this study. Our findings on very fast targets are inconclusive, and as such may not be able to inform the design of such applications. The range of participant ages in this study was also narrow. Each participant used was of college age, and so it is uncertain whether these findings apply to the behavior of older individuals. Furthermore, we only used a head gaze-based tracking metaphor to evaluate performance. While head gaze is used for tracking moving targets, other techniques such as eye gaze and pointing with the hand are also common and may perform differently in VR than expected. Finally, the use of VR may lead to a loss of the fidelity of the target, particularly in the case of audio. In particular, the use of a default HRTF rather than a personalized one and the lack of any audio reverberation in the virtual environment may have caused a subtle loss of audio fidelity which could have affected participants' tracking.

7 Conclusion and Future Work

In this work, we investigated how moving target tracking performance in VR is influenced by sensory modality, calibration via feedback, and target speed. We discovered that using closed-loop visual feedback is effective for visuo-motor recalibration of target tracking. Training simulations involving moving targets, such as those for athletes or engineers, would benefit from using closed-loop feedback to improve users' tracking performance in the short-term. Furthermore, we found that spatial audio on its own may not be sufficiently informative for tracking a moving target, whereas adding audio to high-fidelity visual information may not lead to an improvement of tracking performance. It seems that users are capable of using only visual information to track, but audio-only targets should be avoided. Cubic regression models were developed using participant data from the study. These models can be used to predict hypothetical performance in future scenarios across various target angles and sensory modalities. Speed was found to have no effect on tracking performance, suggesting that the mechanism by which humans track moving targets is robust across many target speeds. While this finding may not be applicable to high-speed targets, the particular speed of targets such as cars in a city in a driving simulation may not degrade tracking performance.

In the future, we wish to explore more factors and tracking techniques. Low-fidelity visual information was found to be insufficient for accurate tracking in the real world. Given that the fidelity of information can be manipulated to a great degree in VR, we wish to investigate how degraded visual information must become for audiovisual binding to influence tracking performance. In this work, all the sensory information provided to participants was congruent; controlled conflicts involving mismatched visual and aural information may be investigated in later work. Other forms of sensory stimuli, such as haptic feedback, also offer potential conditions. Additionally, a future study with a wider range of speeds may be more generally applicable. We also wish to investigate how sensory modality (i.e. eye gaze and manual tracking) and training affect tracking performance using different techniques.

Acknowledgments

The authors wish to thank those who participated in this study. This work was supported in part by the US National Science Foundation (CISE IIS HCC) under Grant No. 2007435.

References

[1] B. M. Altenhoff, P. E. Napieralski, L. O. Long, J. W. Bertrand, C. C. Pagano, S. V. Babu, and T. A. Davis. Effects of calibration to visual

and haptic feedback on near-field depth perception in an immersive virtual environment. In *Proceedings of the ACM symposium on applied perception*, pp. 71–78, 2012. 2, 3

[2] S. V. Babu, T. Y. Grechkin, B. Chihak, C. Ziemer, J. K. Kearney, J. F. Cremer, and J. M. Plumert. An immersive virtual peer for studying social influences on child cyclists' road-crossing behavior. *IEEE transactions on visualization and computer graphics*, 17(1):14–25, 2010. 1

[3] S. V. Babu, H.-C. Huang, R. J. Teather, and J.-H. Chuang. Comparing the fidelity of contemporary pointing with controller interactions on performance of personal space target selection. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 404–413. IEEE, 2022. 5

[4] Y.-B. Bai, J.-H. Yong, C.-Y. Liu, X.-M. Liu, and Y. Meng. Polyline approach for approximating hausdorff distance between planar free-form curves. *Computer-Aided Design*, 43(6):687–698, 2011. 4

[5] A. Bhargava, R. Venkatakrishnan, R. Venkatakrishnan, K. Lucaites, H. Solini, A. C. Robb, C. C. Pagano, and S. V. Babu. Can i squeeze through? effects of self-avatars and calibration in a person-plus-virtual-object system on perceived lateral passability in vr. *IEEE Transactions on Visualization and Computer Graphics*, 2023. 2

[6] A. Bhargava, R. Venkatakrishnan, R. Venkatakrishnan, H. Solini, K. Lucaites, A. C. Robb, C. C. Pagano, and S. V. Babu. Did i hit the door? effects of self-avatars and calibration in a person-plus-virtual-object system on perceived frontal passability in vr. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4198–4210, 2021. 2

[7] G. Bingham and J. L. Romack. The rate of adaptation to displacement prisms remains constant despite acquisition of rapid calibration. *Journal of Experimental Psychology: Human Perception and Performance*, 25(5):1331, 1999. 2

[8] G. P. Bingham and C. C. Pagano. The necessity of a perception–action approach to definite distance perception: Monocular distance perception to guide reaching. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1):145, 1998. 2

[9] G. P. Bingham and C. C. Pagano. The necessity of a perception–action approach to definite distance perception: Monocular distance perception to guide reaching. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1):145, 1998. 5

[10] M. Bischoff, B. Walter, C. Blecker, K. Morgen, D. Vaitl, and G. Sammer. Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia*, 45(3):578–586, 2007. 3

[11] D. Bowman, D. Koller, and L. Hodges. Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques. In *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*, pp. 45–52, 1997. doi: 10.1109/VRAIS.1997.583043 1

[12] F. B. Colavita. Human sensory dominance. *Perception & Psychophysics*, 16(2):409–412, 1974. 2

[13] B. Day, E. Ebrahimi, L. S. Hartman, C. C. Pagano, A. C. Robb, and S. V. Babu. Examining the effects of altered avatars on perception-action in virtual reality. *Journal of Experimental Psychology: Applied*, 25(1):1, 2019. 2

[14] E. Ebrahimi. *Investigating embodied interaction in near-field perception-action re-calibration on performance in immersive virtual environments*. PhD thesis, Clemson University, 2017. 2

[15] E. Ebrahimi, B. Altenhoff, L. Hartman, J. A. Jones, S. V. Babu, C. C. Pagano, and T. A. Davis. Effects of visual and proprioceptive information in visuo-motor calibration during a closed-loop physical reach task in immersive virtual environments. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '14, p. 103–110. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2628257.2628268 2

[16] E. Ebrahimi, B. Altenhoff, L. Hartman, J. A. Jones, S. V. Babu, C. C. Pagano, and T. A. Davis. Effects of visual and proprioceptive information in visuo-motor calibration during a closed-loop physical reach task in immersive virtual environments. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 103–110, 2014. 2

[17] E. Ebrahimi, B. M. Altenhoff, C. C. Pagano, and S. V. Babu. Carryover effects of calibration to visual and proprioceptive information on near field distance judgments in 3d user interaction. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 97–104. IEEE, 2015. 3

[18] E. Ebrahimi, A. Robb, L. S. Hartman, C. C. Pagano, and S. V. Babu. Effects of anthropomorphic fidelity of self-avatars on reach boundary estimation in immersive virtual environments. In *Proceedings of the 15th ACM Symposium on Applied Perception*, pp. 1–8, 2018. 5

[19] A. Esteves, D. Verweij, L. Suraiya, R. Islam, Y. Lee, and I. Oakley. Smoothmoves: Smooth pursuits head movements for augmented reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, UIST '17, p. 167–178. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3126594.3126616 1

[20] W. D. Hairston, P. J. Laurienti, G. Mishra, J. H. Burdette, and M. T. Wallace. Multisensory enhancement of localization under conditions of induced myopia. *Experimental brain research*, 152:404–408, 2003. 2, 6

[21] S. R. Harris, R. L. Kemmerling, and M. M. North. Brief virtual reality therapy for public speaking anxiety. *CyberPsychology & Behavior*, 5(6):543–550, 2002. PMID: 12556117. doi: 10.1089/109493102321018187 1

[22] K. A. M. Heydn, M. P. Dietrich, M. Barkowsky, G. Winterfeldt, S. von Mammen, and A. Nüchter. The golden bullet: A comparative study for target acquisition, pointing and shooting. In *2019 11th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1–8, 2019. doi: 10.1109/VS-Games.2019.8864589 1

[23] P. Hock, S. Benedikter, J. Gugenheimer, and E. Rukzio. Carvr: Enabling in-car virtual reality entertainment. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 4034–4044. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3025665 1

[24] Y.-H. Huang, R. Venkatakrishnan, R. Venkatakrishnan, S. V. Babu, and W.-C. Lin. Using audio reverberation to compensate distance compression in virtual reality. In *ACM Symposium on Applied Perception 2021*, pp. 1–10, 2021. 2

[25] C. Kayser and L. Shams. Multisensory causal inference in the brain. *PLoS biology*, 13(2):e1002075, 2015. 2

[26] J. Lanman, E. Bizzi, and J. Allum. The coordination of eye and head movement during smooth pursuit. *Brain Research*, 153(1):39–53, 1978. doi: 10.1016/0006-8993(78)91127-7 1

[27] J. Leung, V. Wei, M. Burgess, and S. Carlile. Head tracking of auditory, visual, and audio-visual targets. *Frontiers in neuroscience*, 9:493, 2016. 2, 4, 6, 8

[28] W.-Y. Lin, Y.-C. Wang, D.-R. Wu, R. Venkatakrishnan, R. Venkatakrishnan, E. Ebrahimi, C. Pagano, S. V. Babu, and W.-C. Lin. Empirical evaluation of calibration and long-term carryover effects of reverberation on egocentric auditory depth perception in vr. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 232–240. IEEE, 2022. 2, 3

[29] B. Lind, H. Sihlbom, A. Nordwall, and H. Malchau. Normal range of motion of the cervical spine. *Archives of physical medicine and rehabilitation*, 70(9):692–695, 1989. 3

[30] K.-Y. Liu, S.-K. Wong, M. Volonte, E. Ebrahimi, and S. V. Babu. Investigating the effects of leading and following behaviors of virtual humans in collaborative fine motor tasks in virtual reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 330–339. IEEE, 2022. 1

[31] M. Mallek, N. Benguigui, M. Dicks, and R. Thouvarecq. Sport expertise in perception–action coupling revealed in a visuomotor tracking task. *European journal of sport science*, 17(10):1270–1278, 2017. 2

[32] D. L. Mann, W. Spratford, and B. Abernethy. The head tracks and gaze predicts: how the world's best batters hit a ball. *PloS one*, 8(3):e58289, 2013. 1

[33] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264(5588):746–748, 1976. 2

[34] R. Messing and F. H. Durgin. Distance perception and the visual horizon in head-mounted displays. *ACM Trans. Appl. Percept.*, 2(3):234–250, jul 2005. doi: 10.1145/1077399.1077403 1

[35] Z. Mihajlovic, S. Popovic, K. Brkic, and K. Cosic. A system for head-neck rehabilitation exercises based on serious gaming and virtual reality. *Multimedia Tools and Applications*, 77:19113–19137, 2018. 1

[36] B. J. Mohler, S. H. Creem-Regehr, and W. B. Thompson. The influence of feedback on egocentric distance judgments in real and virtual envi-

ronments. APGV '06, p. 9–14. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1140491.1140493 2
[37] J. W. Murray. *Building virtual reality with Unity and Steam VR*. CRC Press, 2017. 3
[38] P. E. Napieralski, B. M. Altenhoff, J. W. Bertrand, L. O. Long, S. V. Babu, C. C. Pagano, J. Kern, and T. A. Davis. Near-field distance perception in real and virtual environments using both verbal and action responses. *ACM Transactions on Applied Perception*, 8(3), 8 2011. doi: 10.1145/2010325.2010328 5
[39] C. C. Pagano, R. P. Grutzmacher, and J. C. Jenkins. Comparing verbal and reaching responses to visually perceived egocentric distances. *Ecological Psychology*, 13(3):197–226, 2001. 5
[40] K. Ponto, M. Gleicher, R. G. Radwin, and H. J. Shin. Perceptual calibration for immersive display environments. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):691–700, 2013. 2
[41] Y. Y. Qian and R. J. Teather. The eyes don't have it: an empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pp. 91–98, 2017. 1
[42] J. J. Rieser, H. L. Pick, D. H. Ashmead, and A. E. Garing. Calibration of human locomotion and models of perceptual-motor organization. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3):480, 1995. 2
[43] C. Spence. Audiovisual multisensory integration. *Acoustical science and technology*, 28(2):61–70, 2007. 2
[44] R. S. Tannen, W. T. Nelson, R. S. Bolia, J. S. Warm, and W. N. Dember. Evaluating adaptive multisensory displays for target localization in a flight task. *The International journal of aviation psychology*, 14(3):297–312, 2004. 2
[45] S. van Andel, M. H. Cole, and G.-J. Pepping. A systematic review on perceptual-motor calibration to changes in action capabilities. *Human movement science*, 51:59–71, 2017. 2
[46] R. Venkatakrishnan, R. Venkatakrishnan, B. Raveendranath, C. C. Pagano, A. C. Robb, W.-C. Lin, and S. V. Babu. How virtual hand representations affect the perceptions of dynamic affordances in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2258–2268, 2023. 2
[47] M. Volonte, Y.-C. Hsu, K.-Y. Liu, J. P. Mazer, S.-K. Wong, and S. V. Babu. Effects of interacting with a crowd of emotional virtual humans on users' affective and non-verbal behaviors. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 293–302. IEEE, 2020. 1
[48] P. Wang, P. Wu, J. Wang, H.-L. Chi, and X. Wang. A critical review of the use of virtual reality in construction engineering education and training. *International Journal of Environmental Research and Public Health*, 15(6), 2018. doi: 10.3390/ijerph15061204 1
[49] W. H. Warren. Action modes and laws of control for the visual guidance of action. In *Advances in psychology*, vol. 50, pp. 339–379. Elsevier, 1988. 2
[50] W. H. Warren. The dynamics of perception and action. *Psychological review*, 113(2):358, 2006. 2
[51] L. Yang, W. Dong, J. Huang, T. Feng, W. Hong'an, and D. Guozhong. Influence of multi-modality on moving target selection in virtual reality. *Virtual Reality & Intelligent Hardware*, 1(3):303–315, 2019. 2
[52] J. W. Youdas, T. R. Garrett, V. J. Suman, C. L. Bogard, H. O. Hallman, and J. R. Carey. Normal range of motion of the cervical spine: an initial goniometric study. *Physical therapy*, 72(11):770–780, 1992. 3