Performance Analysis of Catch-Up Eye Movements in Visual Tracking

JENNA KANG, New York University, USA

BUDMONDE DUINKHARJAV, New York University, USA

NIALL L. WILLIAMS, New York University, USA

QI SUN, New York University, USA

(a) gaming scene with visual tracking
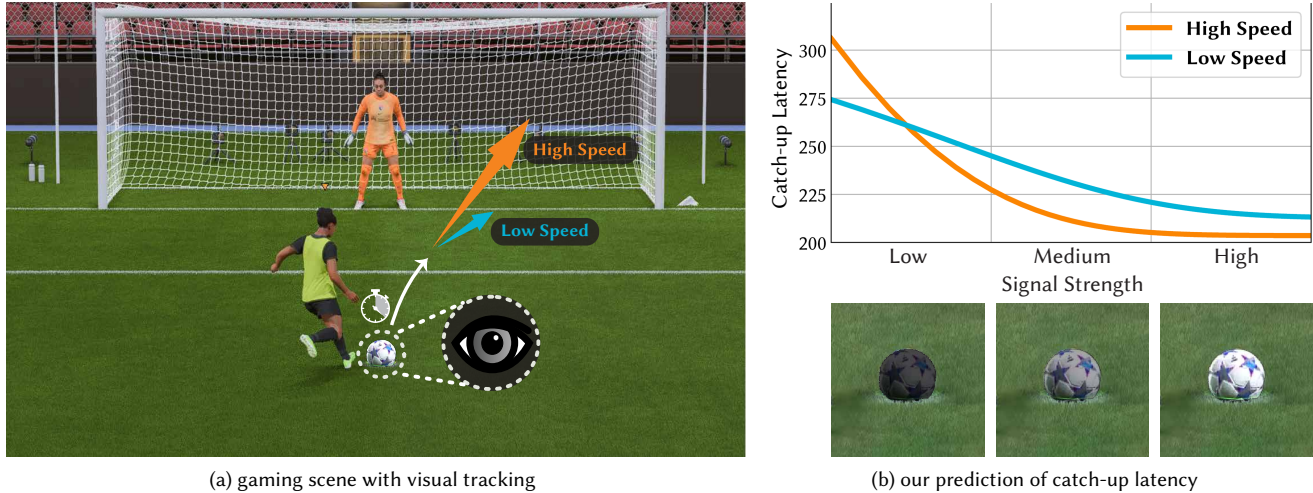

(b) our prediction of catch-up latency

Fig. 1. *Catch-up eye movement performance influenced by visual features.* (a) shows a soccer gaming scenario where the gatekeeper player must accurately follow the sudden motion of a kicked ball to take action. This process requires the player to visually estimate the change of the ball's motion pattern and adjust the eye movement accordingly to catch up. Depending on game designs and difficulties, the ball may appear differently on screen (varying visual signal strength) or move at different speeds. (b) depicts our prediction of user catch-up latency influenced by visual signal strengths and ball motion speeds.

In graphics applications featuring dynamically moving visual targets – such as film and gaming – we have to rotate our eyes to follow objects as they move across the screen. Because target motion is often unpredictable and ever-changing, we must rapidly respond to motion cues and adjust eye movements to maintain the target within the fovea, a process known as catch-up. This catch-up behavior reflects how efficiently the eyes react to and compensate for sudden changes in motion, making it a critical indicator for both task performance and the overall visual experience. In this work, we study and measure the eye catch-up performance during visual tracking. In particular, we present a behavioral analysis that predicts users' reaction latency to abrupt target motion based on target visibility. Our numerical analysis and human subject studies evidence the effectiveness and generalizability. We further show how the catch-up metric can be applied to evaluate video quality, adjust game difficulty, and optimize display configurations

for enhanced user performance. We envision this research to create a computational link between human perception and behavioral performance in dynamic graphics contexts.

CCS Concepts: • **Computing methodologies** → **Perception**; **Motion processing**.

Additional Key Words and Phrases: perceptual error, optical flow, visual perception, human performance, dynamic environment

Authors' Contact Information: Jenna Kang, jennakang@nyu.edu, New York University, Brooklyn, NY, USA; Budmonde Duinkharjav, budmonde@gmail.com, New York University, Brooklyn, NY, USA; Niall L. Williams, n.williams@nyu.edu, New York University, Brooklyn, NY, USA; Qi Sun, qisun@nyu.edu, New York University, Brooklyn, NY, USA.

1 Introduction

When playing a video game or watching a film, the ability for our eyes to follow a moving target on screen and maintain it within the clear foveal vision—known as *visual tracking* [39]—is a critical measure of both task performance and experiential quality [49]. Real-world applications are volatile; instead of uniform speed, targets usually move with unpredictable patterns, requiring observers to continuously adapt their eye movements to catch up. To do so, we rely on anticipatory estimates of the target location and velocity, followed by rapid eye movements to realign our gaze as early and accurately as possible [46, 47].

Studies in perceptual computer graphics commonly focus on measuring and modeling visual acuity [41], such as detecting artifacts due to altered rendering in resolution [58] and color [21, 35]. To further understand how visual content influences our *behaviors*, recent research has measured reactive visual performance, such as spatial landing errors [2] and temporal reaction latency [33]. Existing graphics research studies image-induced behaviors, such as the reaction latency to peripheral onset targets [20]. However, observers' catch-up performance to dynamic changes during continuous observation remains underexplored despite its ubiquity in downstream applications, primarily due to the difficulties of characterizing the complex, high-dimensional spatio-temporal visual factors.

Here, we measure and analyze users' catch-up latency to dynamically moving targets during continuous visual tracking. To this aim, we first conduct a series of psychophysical experiments to characterize the factors influencing the catch-up performance. We observed that the scales of visual signal strength (regardless of individual features) and speed changes (regardless of pre-conditioned eye movements) significantly impact the catch-up eye movement performance. Building on the data and insights, we regress an integrated and analytical prediction. Numerical analysis validates the prediction accuracy and generalizability. In addition, we leverage the research to practical scenarios, including evaluating animation design quality, managing game difficulty through character appearance and target motion adjustments, and optimizing video content for varying eye-display distances. We hope the research tightens the connection between visual perception and reactive behaviors for dynamic graphics applications.

Overview of scope and limitations. This work analyzes human visual tracking performance in diverse dynamic environments through the lens of catch-up eye movement latency, aiming to advance visual-performance-aware computer graphics literature beyond static settings. However, our data-driven model is not intended to represent the neurological mechanisms of sensorimotor control. Given the spatio-temporal complexities, we capture and predict relative effect trends to demonstrate their guidance potential rather than precise temporal values for direct algorithmic deployment.

2 Related Work

2.1 Visual Perception Guided Graphics

Measuring and modeling visual perception has been explored in both foundational and applied perception domains, such as design [25] and photography [30]. An understanding of peripheral vision enables foveated rendering, a technique that improves the efficiency of rendering for eye-tracked displays with a lower quality in the periphery [32, 59]. Many foveation techniques exist, such as rendering with lower-resolution frames [48], using image summary statistics [58], via neural networks trained with a perceptual loss [34], for neural radiance fields [18], with reduced color fidelity [21], with preserved motion cues [55], or attention-based foveation [37]. Perceptually-inspired image quality metrics predict the perceived quality of images based on how the visual system processes visual features [42, 43]. A common feature of existing perceptual models is that they usually focus on visual acuity. However, a significant gap

remains in understanding how we *react and perform* in real-world tasks.

2.2 Visual Behavior & Performance

Predicting users' gaze behavior when interacting with computer graphics systems is critical for understanding the overall experience and task performance. The combination of object and eye movements creates visual motion that serves as a cue for maintaining fixation and can alter the perceived location of objects in the visual field [60]. Souto and Kerzel [53] modeled a variable coupling of visual and motor selection as two separate processes that occur during tracking to coordinate catch-up saccades and pursuit. Arabadzhiyska et al. [1] developed a model to predict saccade dynamics in head-mounted displays. Duinkharjav et al. modeled saccade latency for static images based on image features [20, 22]. In neurophysiology domains beyond computer science, stochastic models have been established to simulate the underlying catch-up saccade mechanisms during smooth pursuit [46], which are found asymmetrically dependent on target velocity [7]. However, current behavioral models commonly consider static images instead of dynamic content with continued observation, or simplified stimuli that do not fully represent high-dimensional graphics features.

2.3 Eye Movement in Dynamic Content

Understanding observers' dynamic eye movements has implications for implementing and optimizing graphics applications. Researchers have developed learning-based methods for predicting gaze scanpaths in VR [5, 31, 44]. For instance, background motion *near* a target object can influence observers' ability to intercept motion [10], and therefore the following eye movements [61]. Meanwhile, increasing target contrast can increase observers' eye velocity gain during smooth pursuits [54]. Additionally, in head-mounted displays, visual artifacts such as phantom array and motion blur can impact saccadic error [27]. However, due to high-dimensional interactions between image features, we lack a comprehensive analyzes of the effects of appearance on eye movements for complex, naturalistic stimuli. Overall, while many metrics influence behavioral performance in visual tracking [38, 54], in practice, initial latency for catch-up eye movements is particularly relevant in computer graphics applications, as delays in foveating and tracking a visual target can cause observers to miss critical events.

3 Measuring Catch-Up Eye Movement Latency

We determine the extent to which visual target visibility affects the latency of catch-up eye movement via eye-tracked psychophysical experiments. Specifically, we manipulate the visibility of targets by adjusting the luminance and color contrasts, as well as by introducing external noise. By analyzing the effects of changing target visibility across different perspectives of image manipulation, we identify common trends of human behavior, and ultimately regress a computational model for predicting human performance in tracking dynamic visual content.
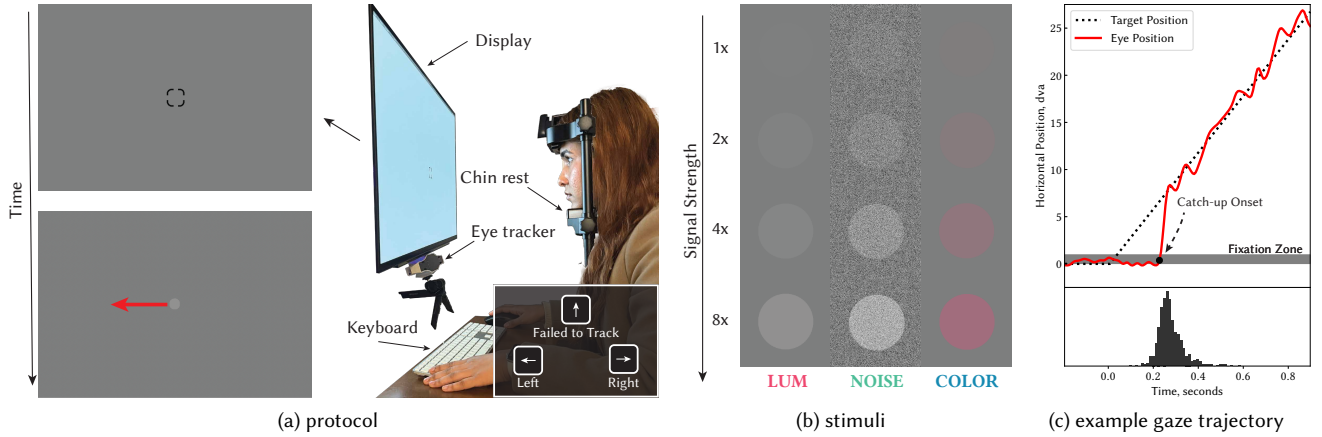
Fig. 2. *Experimental protocol, stimuli, and data.* (a) Participants visually tracked target stimuli that appeared on the display, and indicated via button press its direction of movement. Eye position traces were recorded using an eye tracker. (b) Target visibility was modulated by adjusting different visual features across the **LUM**, **NOISE**, and **COLOR** experiments. Stimuli with signal strength = 1× were calibrated to be at threshold of visibility, and scaled up by signal strength. (c) Catch-up onset distributions were determined by detecting the latest saccades that moved away from the fixation zone during each trial.

3.1 Experiment Design

Participants. Five participants (ages 22 - 28, 3 male) with normal or corrected-to-normal vision were recruited for a series of three psychophysical experiments. Each experiment was conducted during a separate session, and each session consisted of two blocks of 120 visual target tracking tasks each. Each experiment took a total of about 40 minutes to complete. All experimental protocols were approved by an institutional review board (IRB).

Setup. Experimental stimuli were displayed on a 27.5 inch OLED monitor (LG 27GR95QE) with a refresh rate of 240 Hz, a spatial resolution of 2560 × 1440 pixels, and subtended a horizontal field of view of 55°. As shown in Figure 2a, participants seated throughout the experiments, and their head positions were stabilized with a chin rest. Eye position signals were recorded using a 150 Hz eye tracker (GazePoint 3), and was calibrated before each block of trials.

Stimuli and design. Participants were instructed to visually track targets that appeared and moved across the screen. As illustrated in Figure 2b, targets were solid disks with a diameter of .5 dva (*degrees of visual angle*) placed on top of a neutral gray background (with measured mean luminance of 54 cd/m^2), and moved at a constant speed. The visibility and speed of the targets varied across trials and constituted the different conditions across which we compared the catch-up performance. We leverage the eye-tracking data collected as participants visually tracked the targets to quantify a measure of performance as further detailed in the *Analysis* paragraph.

One of the main factors that affect visual tracking performance is target speed [54]. Therefore, we varied the target speeds $v \in \{7.5, 15, 30\}$ dva/s (see supplementary video to compare conditions). Additionally, as discussed in Section 2.3, the visibility of visual targets significantly affects catch-up performance. Thus, across our three experiments, we modulated target visibility within three different features. Specifically, we studied how performance is affected by luminance contrast in low external noise (**LUM** experiment), and

high external noise (**NOISE** experiment), as well as by color contrast (**COLOR** experiment).

In the **LUM** and **NOISE** experiments, the luminance contrast of the target varied between conditions with zero background noise in **LUM**, and a uniform additive noise with RMS contrast of 23% added to the background in **NOISE**. In the **COLOR** experiment, the color of the disk varied along the $L - M$ axis in *DKL* coordinates [19], while the luminance was kept equal to the background. For applications of the *DKL* color-space in computer graphics, see [21, 43].

We parameterize target *visibility* in units of "just-detectable contrast thresholds", and refer to this parameter as the *signal strength*, s. That is, for a target stimulus with a contrast value of c, its corresponding signal strength equals $s = c/c_{th}$, where c_{th} is the just detectable threshold contrast (i.e., inverse of the sensitivity) of the target stimulus. In **LUM** and **NOISE**, contrast was quantified using Michelson contrast, whereas the *DKL* color contrast is used in the **COLOR**. Even though the specific contrast values of the targets vary significantly across experiments, the generalized *signal strength* parameterization allows us to validate our hypothesis that catch-up performance demonstrates similar first-order patterns, irrespective of the specific visual features being manipulated.

Since we require the contrast sensitivity thresholds calibrated for each individual in each experimental task (see further discussion in Section 6), prior to each experiment, the sensitivity in each task was determined via a 4-down-1-up adaptive staircase procedure. The calibration featured the same disk from its corresponding main experiment, moving either leftward or rightward at a speed of $v = 30$ dva/s. Participants progressed through the staircase by visually tracking the target and indicating the direction of the disk's motion via button press. Staircase steps were incremented in reciprocal contrast units with a step size = 5 for six reversals, and the average of the last three reversals were used to determine the sensitivity.

Based on the calibrated contrast threshold, the *signal strength* of the main experiments were determined. As illustrated in Figure 2b,

across the experiments, signal strength values of $s \in \{1, 2, 4, 8\}$ were used. Overall, the target's 3 speed conditions ×4 signal strength conditions were repeated 10 times per block for a total of 120 trials.

Procedure. Each trial began with a button press. As shown in Figure 2a, following a randomized stimulus onset delay of $0 − 1$ s, the target stimulus replaced the fixation crosshair and moved either leftward or rightward across the screen at a constant speed determined by the trial condition. Randomizing the stimulus onset delay and target movement direction minimized anticipatory eye movements. The target remained in motion for $1 − 1.5$ s before disappearing. At the end of each trial, participants indicated via button press whether the target object moved leftward, rightward, or if they failed to track it. Trials in which participants failed to track the target or responded incorrectly were repeated at the end of each block. Participants were instructed to take breaks at any time, and there was a mid-session break enforced for sessions exceeding 30 minutes to minimize fatigue and/or discomfort of the participants.

Analysis. As motivated in Section 2.3, the study focuses on the latency for catch-up eye movements, as delays in foveating and tracking a target can cause missed visual details and poor task performance. Catch-up movement is primarily driven by two concurrent mechanisms triggered at stimulus onset: pursuit eye movement control and a corrective saccadic eye movement [46]. Pursuit eye movement is tuned to synchronize the movements of the eye and the target so that a moving target remains foveated [51]. However, the onset of pursuit takes .1 s or longer to fully match the target's motion with an abrupt change [45]. During such delays, the positional error between the target and eye positions accumulates and necessitates a catch-up saccade that corrects this positional offset [46]. In our analysis, we leverage the onset of this corrective saccade as an indicative marker to quantify catch-up latency.

To detect the corrective saccade, we utilize the recorded eye position traces. Eye position traces were smoothed by a Butterworth filter with a 20 Hz cut-off prior to analysis. Trials where participants failed to maintain fixation at the beginning of the trial, and failed to track the target were excluded from analysis, and constituted 2.9% of all trials (see Supplement A for data processing details).

Figure 2c shows an example trace of target motion to the participant's gaze trajectory. Eye velocity was calculated using the central difference method and saccades were detected using a fixed velocity threshold [26]. Velocity cut-off criteria of 12.5, 25, and 50 dva/s were applied for stimuli moving at 7.5, 15, and 30 dva/s, respectively. The *main* corrective saccade was identified using a positional criterion that selected the latest saccade which shifted the eye position away from the initial fixation distribution by more than two standard deviations towards the target. The initial fixation distribution was established by aggregating the eye position data within the $−0.2 < t < 0.1$ s window relative to stimulus onset. Saccades detected during $t < 0.1$ s were excluded, as such movements could not have been programmed in response to the stimulus onset [4, 17]. Each saccade was manually inspected to ensure proper trial counting from the detection algorithm.

Catch-up latency was also examined across all trials to check for signs of fatigue or decline in performance. Participants exhibited
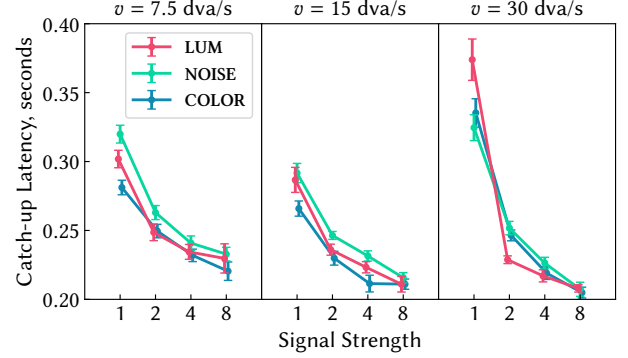

Fig. 3. *Catch-up performance of a representative participant.* Mean catch-up latencies are plotted against signal strength, and compared across targets at varying speeds. Each experiment is represented by a distinct color (see Figure 2b for the corresponding stimuli). Error bars denote ±1 Standard Error Mean (SEM). See Figure 4 for plots of the remaining participants.

consistent catch-up latency between early and late trials, an indicator of non-elevated fatigue [16]. Please see Supplement C for detailed plots on catch-up latency time throughout each participant's trials.

To further validate data consistency across individuals, we leverage a generalized linear mixed model (GLMM) to measure individual participants' data variability [8]:

$$\text{catch-up latency} \sim \text{signal strength} \times \text{object speed} + (1 \mid \text{ID}),$$

where signal strength, object speed, and their interaction were modeled as fixed effects, and participant ID was included as a random intercept [23, 24]. Additionally, we also conducted a post-hoc power analysis to evaluate the sensitivity of our experimental design [9]

3.2 Results

We present our main findings for a representative participant's experimental data across the 12 conditions in Figure 3. Catch-up latencies improve by up to 0.166/0.117/0.130 s in **LUM/NOISE/COLOR** respectively, based on both the target speed and signal strength. Increasing signal strength provides diminishing benefits until performance nears its peak at a mean latency of .216/.218/.212. Notably, at high signal strength, faster targets induce faster catch-up, while at low signal strength, the opposite holds.

Two-way ANOVA tests revealed a significant main effect of target speed ($p < 0.01$) and signal strength ($p < 0.001$) on catch-up latency across all experiments, indicating that target speed and signal strength strongly influence catch-up latency (see Supplement B for detailed ANOVA results). Additionally, there was a significant interaction between target speed and signal strength across all experiments ($p < 0.001$), suggesting that the effect of object speed on latency depends on the level of signal strength.

The Spearman correlation of mean latencies found between **LUM** & **NOISE**, **LUM** & **COLOR**, and **NOISE** & **COLOR** experiments were $\rho = 0.96$ ($p < 0.001$), $\rho = 0.94$ ($p < 0.001$), and $\rho = 0.99$ ($p < 0.001$) respectively, indicating a very strong positive monotonic relationship. These results suggest that the catch-up latencies measured in each experiment are highly correlated to each other with a high degree of consistency for the individual.
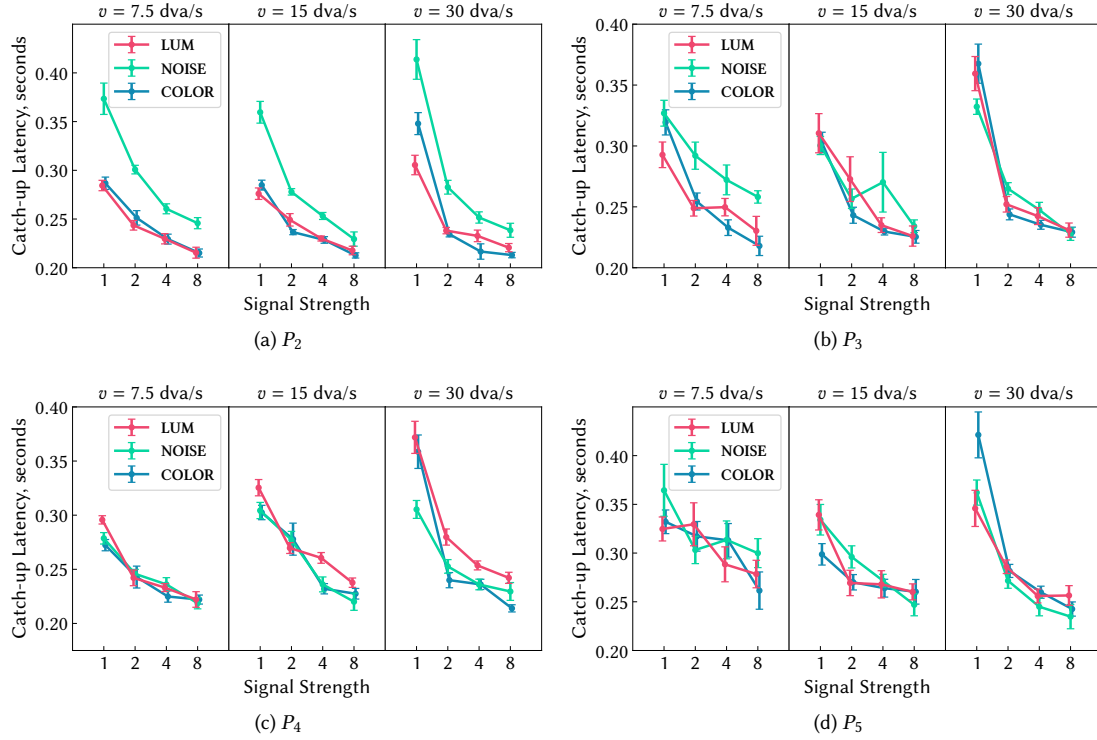
Fig. 4. *Catch-up performance of main study participants.* Mean catch-up latencies of main study participants are visualized in the same style as in Figure 3.

The remaining four participants demonstrate similar trends (Figure 4) and statistical significances across conditions and experiments, as detailed in Supplement B. All the 15 experimental datasets exhibit a significant signal strength effect on latency. Among them, 12 exhibit significant target speed effects on latency, and 12 exhibit a significant interaction.

The GLMM analysis estimated a small variance component for participants (random intercept of 0.0 for **LUM**, **NOISE**, and **COLOR**), indicating that individual differences contributed little to the overall variance. This suggests that with five participants and a large number of trials per participant ($\geq$ 210 per condition), our design provided precise within-participant estimates of catch-up latencies. For the power analysis, we assumed $\alpha = 0.05$, a sample size of $n = 5$ participants, and at least 210 repeated measurements per participant. Under these conditions, a sensitivity analysis showed that our study had 80% power ($1 - \beta = 0.80$) to detect small effect sizes ($\eta_p^2 \geq 0.0297$).

3.3 Discussion

The ANOVA results indicate that both target speed and signal strength significantly influence catch-up eye movement latencies. Thus, these factors should be considered independently, as evidenced by the significant interaction between them. As shown in Figure 3, increasing signal strength accelerates catch-up. Similar

relationships between the *visibility* of visual targets and their corresponding reaction times have been observed in different visual task contexts [20].

Crucially, apart from target visibility, its speed also affects catch-up performance. Faster speeds impair latencies at low signal strengths, suggesting that the visual tracking task becomes more challenging. We hypothesize that this decline in performance occurs because faster-moving low-visibility targets exit the fovea and enter peripheral vision more quickly, where visual acuity is reduced, and target localization becomes more challenging. However, the effect of this performance deterioration might not apply when the signal strength is high and the target can be localized easily even in the periphery.

The observed consistency across different visual features modulated by their signal strength suggests that the catch-up onset may depend on a unified signal encoding of the target visibility. While further physiological research is needed to validate this hypothesis, the presence of such first-order effects provides practical guidance to develop computational methods for downstream graphics applications. Motivated by these statistical insights, we then integrate all experimental data collected in our studies to construct an executable prediction for catch-up latency. Across all participants, the dataset comprised approximately 1100 trials, providing a substantial amount of data for estimating within-participant effects.

3.4 Regressing a Predictive Model

As discussed in Section 3.2, increasing *signal strength* yields diminishing performance benefits, in agreement with prior work indicating that catch-up latency plateaus and performance saturates beyond a certain ceiling [54]. To effectively model this saturation behavior, we define the catch-up onset rate,

$$R = 1/t, \tag{1}$$

where t represents the mean latency. This approach is consistent with established practices in models such as the Drift Diffusion Model (DDM) [50] and LATER models [14], both of which are commonly used to model response latencies in neuroscience and psychology (see [20] for a review relevant to perceptual graphics applications). The adaptation rate, R, is modeled as a sigmoid function:

$$R(s) = (R_{max} - R_{min})\frac{1 - e^{-\lambda s}}{1 + e^{-\lambda s}} + R_{min}, \tag{2}$$

where R_{max} is the peak rate, R_{min} is the minimum rate with $R(s = 0) = R_{min}$, λ determines the slope of the sigmoid, and s is signal strength.

To enable the model to predict the catch-up rate across different target speeds, v, we fit polynomial functions to R_{max}, R_{min}, and λ based on our experimental data. Polynomial fits using scipy `curve_fit` to all **LUM**/**NOISE**/**COLOR** data from Figure 3 yields the following coefficients:

$$R_{max}(\log(v)) = 2.40 + .506\log(v) - .075\log(v)^2$$
$$R_{min}(\log(v)) = -3.19 + 4.74\log(v) - 1.06\log(v)^2 \tag{3}$$
$$\lambda(v) = -.174 + .441\log(v).$$

The model predicts mean catch-up latencies, as in Figure 5, by first calculating the rate using Equations (2) and (3), and then deriving the mean latency using the relationship in Equation (1).

Our dataset suggests that the range of catch-up latency increases with velocity. As in 3.1 and [56], higher velocities likely cause the target to move faster into the periphery, leading to poorer reaction performance bounds. Therefore, the quadratic term in R is used to depict this effect, where the performance bounds decrease with faster velocities. Furthermore, a higher-order polynomial allows the model to capture nonlinear dynamics, and a quadratic term being the highest term helps to prevent overfitting to our data, as is common with very high-order functions.

The sigmoid formulation is motivated by the way performance gradually increases with signal strength, reflecting both diminishing gains at high intensities and the inherent variability in perceptual and visuomotor processes. This reflects the empirical observation that responses transition smoothly, from chance to asymptotic performance, due to uncertainty in sensory processing [36].

4 Evaluating Prediction Effectiveness

4.1 Numerical Analysis

We first show that the model (1) generalizes across visual features – e.g., a model trained on luminance can predict behavior for color – and (2) accurately captures behavioral patterns in Section 3.2.
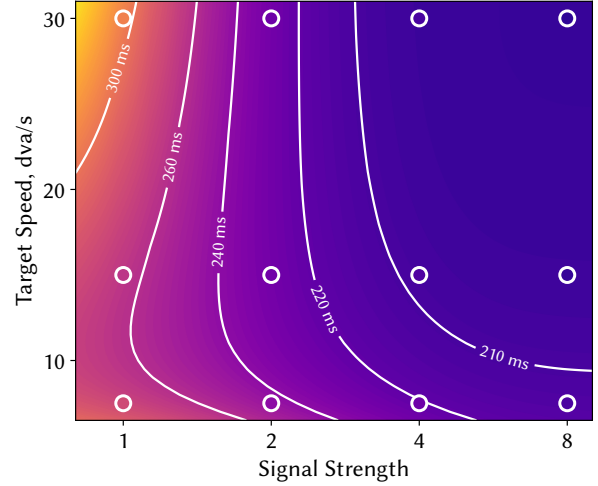
Fig. 5. *Model visualization.* The predicted mean catch-up latencies of our model fitted to all experiment data, compared to the mean latency data it was fit on. Latencies for each target speed, and signal strength conditions are depicted in different colors, along with overlaid level sets indicating equivalent performance. Measured latencies are shown as individual points using the same color mapping.

Generalizability. We fit our model's parameters described in Equations (2) and (3) to the **LUM**, **NOISE**, and **COLOR** experimental results, separately, and measure how well each model's predictions agree with the results of the other two experiments. The R^2 scores for the **LUM** model were .99/.90/.93 across the **LUM**/**NOISE**/**COLOR** datasets respectively. Similarly, **NOISE** and **COLOR** model scores were .90/.97/.79 and .95/.83/.99 respectively. As expected, R^2 scores are highest for models evaluated against their respective trained datasets. We further analyzed the data variability using signal-to-noise ratio (SNR) and observed low SNR (-4.3dB and -0.1dB) for the two participants with slightly lower R^2 (0.72) vs. the average 1.14dB for the other three. This suggests that these participants may exhibit generally higher individual noise, but their average trend still aligns with other participants.

Crucially, we observe that each model presents high fitting scores across all datasets ($R^2 > .79$), indicating that our signal strength-based model generalizes across different visual features [15].

Cross-user Generalizability. We use a leave-one-out cross validation to evaluate fitted models against unseen data. Specifically, model parameters in Equation (3) are fitted using data from four participants, and goodness-of-fit is assessed on the remaining. The R^2 scores for each participant (.80/.96/.92/.72/.72) indicate that our model effectively captures behavioral trends excluded from the fitting process.

4.2 Extending to Continuous Viewing

In addition to the singular trials in Section 3.1, we conducted another study to test our model with continuous viewing, where the target's movement spans a broader range of directions and dynamics.

(a) continuous viewing stimuli
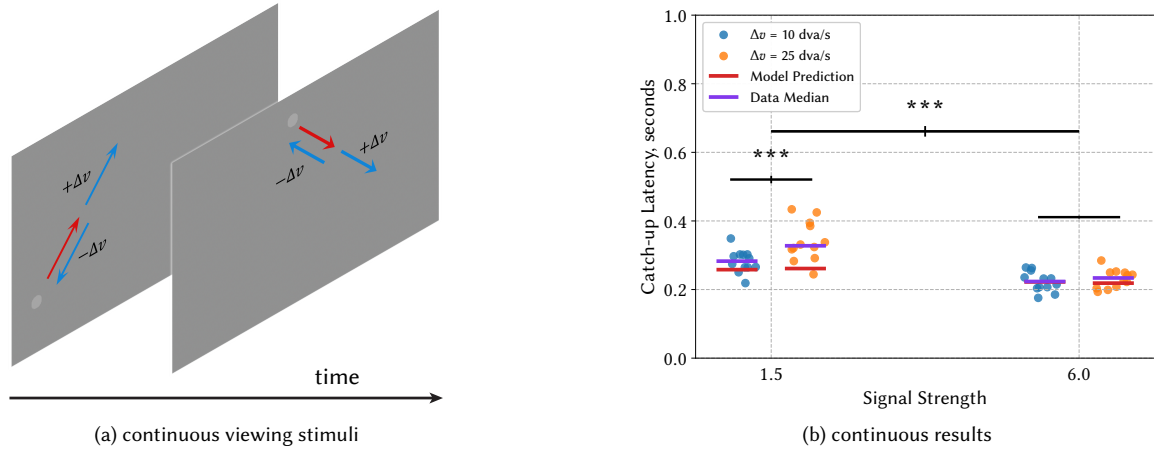

(b) continuous results

Fig. 6. *Stimuli and results for continuous viewing study.* The stimuli (a). The stimuli were altered along global luminance contrast similar to Figure 2, where the disk first moved at 5 dva/s along the red line for 1-1.5s, then changed by Δv. The study results are shown in (b). The individual points reflect mean responses for each participant, color coded by target velocity. Error bars above the scatter plots indicate ±1 SEM of the difference between target velocity within each signal strength, and between different signal strengths overall. ***p < 0.001, **p < 0.01, *p < 0.05

Participants and setup. Twelve participants (ages 21 - 28, 5 female) with normal or corrected-to-normal vision were recruited to perform this *continuous* visual tracking task, as detailed below.

Stimuli, conditions, and procedure. Participants were instructed to visually track a target that continuously moved around the screen for the entire duration. The target was the same solid disk as in Section 3.1 for the **LUM** condition. We also applied similar eye tracker and sensitivity threshold calibration procedure. The target's motion and visibility were modulated across conditions, with two speeds ($v \in \{10, 25\}$ dva/s), and three signal strengths ($s \in \{1.5, 6, 8\}$).

Notably, the target disk moved continuously in random directions instead of being limited to horizontal trajectories. At the start of each trial, the target was launched in a random direction (anywhere 360 deg) with an initial velocity of 5 dva/s. After 1-1.5s, the target's speed suddenly changed by Δv of either 10 or 25 dva/s, while maintaining the same trajectory. The random direction for each trial was algorithmically selected such that the target would not move beyond the edge of the screen.

Although each trial was analyzed independently, the study was conducted in a continuous fashion with no breaks between trials. That is, upon the completion of each trial, a new trial began immediately, again in a new random direction. Please refer to the supplementary video for an example.

Analysis, results, and discussion. Following a similar methodology in Section 3.1, the $s = 8$ condition was used as a normalization benchmark. Since target motion was now presented in a random direction spanning 360 deg, rather than restricted to horizontal motion, we first projected the x and y gaze coordinates to the direction of target motion. This transformation allowed us to analyze the data along the target trajectory as in Section 3.1.

A two-way ANOVA reveals a significant main effect of both signal strength ($F = 56.19, p < 0.0001$) and target speed ($F = 19.51, p <$

0.01) on catch-up latency, with a significant interaction between groups ($F = 13.19, p < 0.001$). Mean catch-up latencies were aggregated from each participant, and were compared to our model predictions in Figure 6b. A repeated measures ANOVA on the mean catch-up latencies reveals a significant effect of target speed on catch-up latency for $s = 1.5$ ($F = 29.40, p < 0.001$), other than $s = 6$ where the predicted latency difference between target speeds is also very low (3.52 ms, shorter than one eye tracking frame) due to the high visibility. The analysis shows the model's prediction trend extends to scenarios where the targets move omnidirectionally and observers perform visual tracking continuously.

5 Application Case Studies

The regressed performance model enables us to support various graphics applications. Rather than conducting ad-hoc user studies to analyze performance for each content design, our model provides approximate predictions, enabling rapid and cost-effective estimations that inform design decisions.

5.1 Analyzing Video Quality on Event Tracking

In video production, creators define how target elements appear, but ensuring fast-moving events remain trackable is an underexplored challenge critical for viewer comprehension and engagement [52]. While existing perceptual metrics assess visual quality, our model offers a complementary perspective by evaluating their impact on event tracking. Here, we demonstrate in the presence of complex stimuli and scene backgrounds as a user study.

Participants and setup. Fourteen participants[1] (ages 22 - 28, 6 female) with normal or corrected-to-normal vision were recruited for a similar psychophysical experiment as in Section 3.1, where they were asked to complete a series of visual target tracking tasks. The

---

[1]Two participants (1 female) were excluded for inability to perform the tasks (excessive blinking from dry eyes, which disrupted eye-tracking accuracy, and colorblindness).
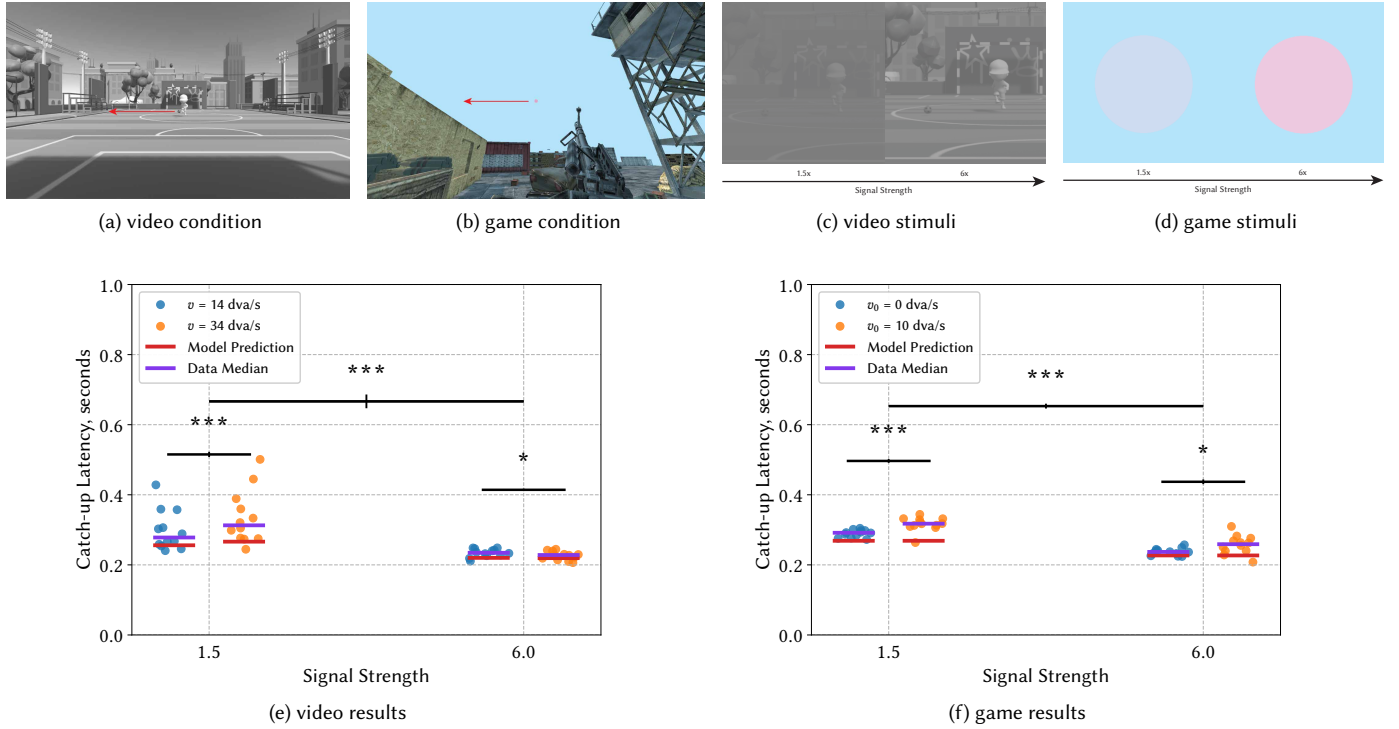
(a) video condition (b) game condition (c) video stimuli (d) game stimuli

(e) video results (f) game results

Fig. 7. *Scene, stimuli, and results of the application case study for videos and games.* The scenes evaluated were a soccer field (a) and an FPS game map (b). The stimuli were altered along global luminance contrast in (c) and color contrast in (d). The study results are shown in (e) and (f). The individual points reflect mean responses for each participant, color coded by target velocity. Error bars above the scatter plots indicate ±1 SEM of the difference between target velocity within each signal strength, and between different signal strengths overall. ***p < 0.001, **p < 0.01, *p < 0.05

overall experimental setup, protocols, and IRB remain the same, but feature major changes in the visual stimuli.

Stimuli and design. During each trial, as shown in Figure 7a, participants viewed a scene featuring a soccer goal-keeping video and were instructed to visually track a target ball (Figure 7c) moving either leftward or rightward. The target's motion and visibility were modulated across conditions, with two target speeds ($v \in \{14, 34\}$ dva/s), and three target visibility levels ($s \in \{1.5, 6, 8\}$). The target speed was controlled by applying a force to the ball to simulate more realistic motion rather than maintaining a constant velocity. Target speed was calculated by averaging the ball's velocity over the first .2 s of motion. Each condition was repeated 20 times for a total of 120 trials and was completed in about 25 min during a single session. To avoid learning effects from repeated exposure to the same clips, each instance of a condition included slight variations in motion trajectories and camera angles. Before the experiment, participants completed an adaptive staircase to calibrate the global contrast of the video to determine the just-detectable threshold of successfully tracking the target.

Analysis and results. Catch-up latency measurement analysis was unchanged from Section 3.1, albeit with more relaxed gaze tracking error criteria (see Supplement B). Similar to Section 4.2, the $s = 8$ condition was used as a benchmark. A two-way ANOVA reveals a

significant main effect of both signal strength ($F = 24.12, p < 0.001$) and target speed ($F = 13.53, p < 0.01$) on catch-up latency, with a significant interaction between groups ($F = 20.54, p < 0.001$). Mean catch-up latencies were aggregated from each participant, and were compared to our model predictions in Figure 7e. A repeated measures ANOVA on the mean catch-up latencies reveals a significant effect of target speed on catch-up latency for both $s = 1.5$ ($F = 24.96, p < 0.0001$) and $s = 6$ ($F = 6.71, p < 0.05$) conditions.

Discussion. The result suggests that our model can effectively predict periods when observers are unable to visually track moving visual events, even when the target and scene backgrounds are more complex. The results reflect the significant effects of signal strength and target speed on catch-up latency as discussed in Section 3.3. This contrasts with prior studies on static images, such as [20], which do not account for target speed as a factor. Therefore, this study underscores the model's potential application in providing appearance and motion design guidance to ensure critical dynamic events are visible to observers.

5.2 Designing Difficulty Levels in Gaming

In gaming, how quickly a player can track and hit moving targets is a core indicator of game difficulties [13, 57]. However, designing different game difficulties commonly requires iterative player tests; automated difficulty control remains an open challenge [6]. Here,

experiment with the model's capability in guiding game difficulty design via players' catch-up delay to targets as a metric. The study was completed by the same participants as Section 5.1, following nearly identical procedures, except for environment and target stimulus appearance differences. Notably, we extend the conditions to *non-zero initial* velocity before transitioning to another. This is to simulate shooter game scenarios.

Stimuli and conditions. At the start of each trial, as illustrated in Figure 7b, a solid disk target of the same design as in the **COLOR** experiment appeared to replace the fixation crosshair, and moved at an initial speed of $v_0 \in \{0, 10\}$ dva/s. After $1-1.5$ s, the target's speed suddenly changed by $\Delta v = 20$ dva/s in either horizontal direction. Throughout every trial, the camera moved in a random forward direction. Target visibility conditions and number of repetitions remained unchanged as in Section 5.1.

Analysis and results. The same analysis procedure from Section 4.2 was used to detect catch-up saccades, with the s = 8 condition was used as a benchmark. A two-way ANOVA reveals a significant main effect of both signal strength ($F = 210.99, p < 0.001$) and target speed ($F = 20.84, p < 0.001$) on catch-up latency, but with no significant interaction effect. Mean catch-up latencies were aggregated as in Section 5.1, and compared to model predictions in Figure 7f. A repeated measures ANOVA on the mean catch-up latencies reveals a significant effect of target speed on catch-up latency for both $s = 1.5$ ($F = 24.96, p < 0.0001$) and $s = 6$ ($F = 6.71, p < 0.05$) conditions.

Discussion. We observe that, despite the added complexity of the initial velocities, the behavioral trends remain consistent with our model predictions. Notably, a faster initial velocity significantly increases catch-up latency, as in Figure 7f. This motivates interesting future work on understanding object tracking while the eyes are already engaged in pursuit. Overall, our model successfully captured performance trends, even under extremely challenging conditions. Observers faced increased target movement complexity, global optic flow from camera motion, and an off-white background adaptation—conditions not tested in our prior experiments. The consistency of the behavioral trends in these scenarios suggests that our findings can reliably extend to more complex and interactive contexts, serving as a potential numerical guideline for game difficulty design that reduces player testing cycles.

5.3 Viewing Conditions Altering Visual Performance

In Figure 8, we show examples where lighting and eye-display distances may induce visual tracking performance variances.

Lighting conditions. While driving at night, drivers' target tracking performance may be significantly compromised. Although existing video quality metrics quantify how different lighting changes appear to observers [42], the effect scales may differ from task performance. In Figure 8a, we took a rendered animation and simulated varied lighting levels. The temporally averaged SSIM-based visual similarity and our predicted catch-up latency display differing trends, showing the necessity of evaluating performance as a complementary perspective. We envision studying their analytical correlations as exciting future exploration.

Eye-display distance. Eye-display distance often changes due to ergonomics and display environments (e.g., VR/AR headsets), affecting perceived size, and speed of observed content. Users' catch-up performance is affected not only by target speed changes, but also its size; Visual stimulus size modulates users' contrast sensitivity to the stimulus [3], and by extension, modulates signal strength as well. By jointly accounting for the change in contrast sensitivity, and target speed given eye-display viewing conditions, we are able to bootstrap our behavioral model.

As a proof-of-concept visualization, in Figure 8b we show how the catch-up latency of adapting a $d = 1$ inch wide visual target with a spatial frequency of 1 cycles-per-degree moving at $v = 5$ inch/s across a 27 inch display changes as a function of both its luminance contrast, and eye-display distance. We compute the underlying signal strength (visualized as a colormap) for each contrast and eye-display distance condition by applying Barten [3]'s contrast sensitivity function on the described stimulus, which outputs signal strength of the stimulus as a direct input for our model, and overlay the resultant catch-up latency prediction according to our model.

6 Limitations and Future Work

Individual vs. unified visibility model. One of our core aims is to establish is how target visibility and motion jointly influence visual catch-up efficiency. Since establishing a fine-grained and comprehensive visibility function for signal strength itself is not our focus, we performed a calibration procedure for participants. We envision that recent advancements in cross-population and unified visibility models [11, 41] may shed light on a statistical model to bypass individual calibrations.

Multi-target and anisotropic optical flows. In Section 5.2, we validated the non-effect from camera motion-induced retinal optical flow during first-person shooter gameplay. In the experiments, participants were instructed to track a single moving target. However, in real-world scenarios, multiple peripheral targets may appear and move anisotropically, potentially influencing localization performance [40]. Additionally, the size of the target may extend outside of the fovea, which has been shown to impact the number of catch-up saccades for the same target motion [29]. A promising future research direction could be exploring the eye movement performance in the visual optical flow space to establish a robust and generalizable model for complex interactive applications.

Unpredictable target trajectories. Our experiments focused on simple, predictable target trajectories with constant velocity, yet real-world motion is far more varied and often unpredictable. Prior work highlights the influence of contextual cues on tracking predictability [28] and shows that acceleration affects pursuit dynamics [12]. To better approximate naturalistic viewing, future studies should incorporate more complex trajectories, including acceleration and other nonuniform motions, where anticipation is limited. Additionally, targets in applied settings frequently appear in the periphery, where saccadic dynamics differ and may confound latency. Incorporating peripheral presentations will be important for testing model robustness and revealing other latency patterns.

(a) Lighting conditions
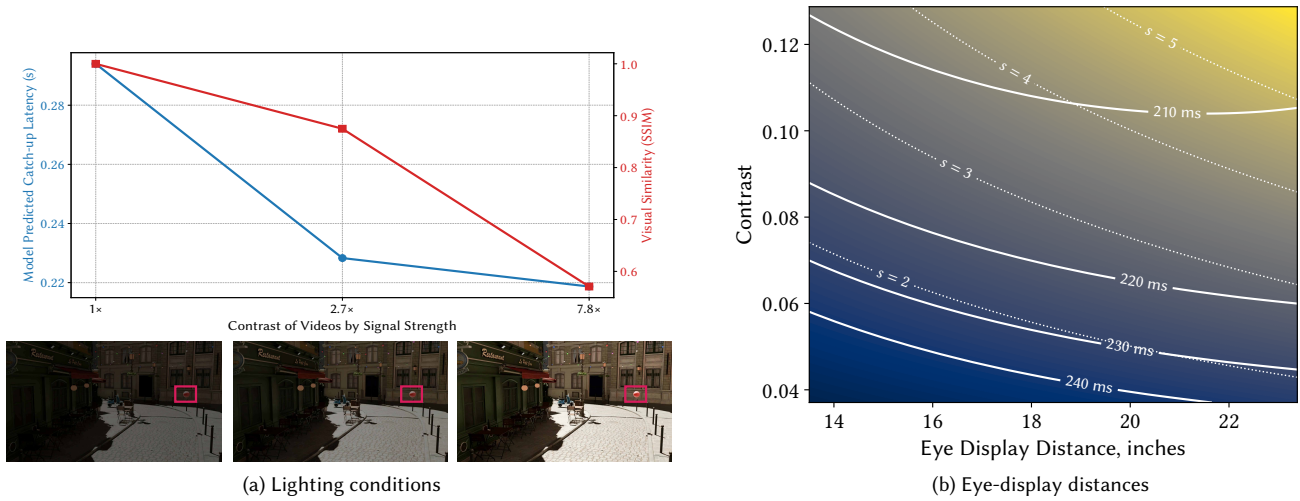
(b) Eye-display distances

Fig. 8. *Viewing conditions influence visual tracking.* (a) Comparing visual similarity with catch-up latency. With three rendered lighting conditions of an animation (bottom), the red and blue curves show SSIM-based visual similarity approximation and our catch-up latency prediction, respectively. (b) Suggesting content appearance with eye-display distance. Predicted catch-up latencies are visualized as solid contour plots overlaid on a colormap of the signal strength which was estimated using Barton's contrast sensitivity function for the stimulus in Section 5.3. Signal strength contours are visualized via dotted lines.

Effect trends vs. numerical precision. Across diverse settings and applications, our model demonstrates generalizability in predicting trends in how context affects catch-up eye movement performance. However, we emphasize that precise numerical accuracy in-the-wild is beyond its current capability due to factors such as spatio-temporal complexity, variable viewing conditions (e.g., omnidirectional), and eye-tracking precision. We envision that augmenting this approach with large-scale gaze behavioral datasets and probabilistic modeling could enable more accurate predictions to inform future rendering algorithms and display systems.

7 Conclusion

In this work, we measure and analyze how spatio-temporal visual features influence human catch-up eye movement latency in dynamic visual target tracking. With validations in various application scenarios, we demonstrate our model's applicability in predicting and guiding interactive visual content designs, considering observers' task performance and viewing experience. We hope the research could foster new research on collaborative ground on modeling the relationship between visual perception and its influence on behavioral performance for practical and interactive graphics.

Acknowledgments

References

[1] Elena Arabadzhiyska, Cara Tursun, Hans-Peter Seidel, and Piotr Didyk. 2023. Practical saccade prediction for head-mounted displays: Towards a comprehensive model. *ACM Transactions on Applied Perceptions* 20, 1 (2023), 1–23.

[2] Elena Arabadzhiyska, Okan Tarhan Tursun, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. 2017. Saccade landing position prediction for gaze-contingent rendering. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1–12.

[3] Peter GJ Barten. 1999. *Contrast sensitivity of the human eye and its effects on image quality.* SPIE press.

[4] Wolfang Becker and Reinhart Jürgens. 1979. An analysis of the saccadic system by means of double step stimuli. *Vision research* 19, 9 (1979), 967–983.

[5] Edurne Bernal-Berdun, Daniel Martin, Sandra Malpica, Pedro J Perez, Diego Gutierrez, Belen Masia, and Ana Serrano. 2023. D-SAV360: A Dataset of Gaze Scanpaths on 360° Ambisonic Videos. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[6] Glen Berseth, M Brandon Haworth, Mubbasir Kapadia, and Petros Faloutsos. 2014. Characterizing and optimizing game level difficulty. In *Proceedings of the 7th International Conference on Motion in Games*. 153–160.

[7] Hans-Joachim Bieg, Lewis L Chuang, Heinrich H Bülthoff, and Jean-Pierre Bresciani. 2015. Asymmetric saccade reaction times to smooth pursuit. *Experimental brain research* 233 (2015), 2527–2538.

[8] Benjamin M Bolker, Mollie E Brooks, Connie J Clark, Shane W Geange, John R Poulsen, M Henry H Stevens, and Jada-Simone S White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution* 24, 3 (2009), 127–135.

[9] Roser Bono, Rafael Alarcón, and María J. Blanca. 2021. Report Quality of Generalized Linear Mixed Models in Psychology: A Systematic Review. *Frontiers in Psychology* Volume 12 - 2021 (2021). https://doi.org/10.3389/fpsyg.2021.666182

[10] Eli Brenner and Jeroen BJ Smeets. 2015. How moving backgrounds influence interception. *PLoS One* 10, 3 (2015), e0119903.

[11] Yancheng Cai, Ali Bozorgian, Maliha Ashraf, Robert Wanat, and K Rafał Mantiuk. 2024. elaTCSF: A Temporal Contrast Sensitivity Function for Flicker Detection and Modeling Variable Refresh Rate Flicker. In *SIGGRAPH Asia 2024 Conference Papers*. 1–11.

[12] Vanessa Carneiro Morita, David Souto, Guillaume S. Masson, and Anna Montagnini. 2025. Anticipatory smooth pursuit eye movements scale with the probability of visual motion: The role of target speed and acceleration. *Journal of Vision* 25, 1 (01 2025), 2–2. https://doi.org/10.1167/jov.25.1.2 arXiv:https://arvojournals.org/arvo/content_public/journal/jov/938700/i1534-7362-25-1-2_1735884269.5659.pdf

[13] Loïc Caroux and Axelle Mouginé. 2022. Influence of visual background complexity and task difficulty on action video game players' performance. *Entertainment Computing* 41 (2022), 100471.

[14] Roger HS Carpenter and MLL Williams. 1995. Neural computation of log likelihood in control of saccadic eye movements. *Nature* 377, 6544 (1995), 59–62.

[15] Samprit Chatterjee and Ali S Hadi. 2015. *Regression analysis by example.* John Wiley & Sons.

[16] Jui-Tai Chen, Ying-Chun Kuo, Tzu-Yu Hsu, and Chin-An Wang. 2022. Fatigue and Arousal Modulations Revealed by Saccade and Pupil Dynamics. *International Journal of Environmental Research and Public Health* 19, 15 (2022). https://doi.org/10.3390/ijerph19159234

[17] Sophie De Brouwer, Demet Yuksel, Gunnar Blohm, Marcus Missal, and Philippe Lefèvre. 2002. What triggers catch-up saccades during visual tracking? *Journal of neurophysiology* 87, 3 (2002), 1646–1650.

[18] Nianchen Deng, Zhenyi He, Jiannan Ye, Budmonde Duinkharjav, Praneeth Chakravarthula, Xubo Yang, and Qi Sun. 2022. Fov-nerf: Foveated neural radiance fields for virtual reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3854–3864.

[19] Andrew M Derrington, John Krauskopf, and Peter Lennie. 1984. Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of physiology* 357, 1 (1984), 241–265.

[20] Budmonde Duinkharjav, Praneeth Chakravarthula, Rachel Brown, Anjul Patney, and Qi Sun. 2022. Image Features Influence Reaction Time: A Learned Probabilistic Perceptual Model for Saccade Latency. *ACM Trans. Graph. (Proc. SIGGRAPH)* 41, 4 (2022), 144:1–144:15.

[21] Budmonde Duinkharjav, Kenneth Chen, Abhishek Tyagi, Jiayi He, Yuhao Zhu, and Qi Sun. 2022. Color-perception-guided display power reduction for virtual reality. *ACM Transactions on Graphics (TOG)* 41, 6 (2022), 1–16.

[22] Budmonde Duinkharjav, Benjamin Liang, Anjul Patney, Rachel Brown, and Qi Sun. 2023. The Shortest Route Is Not Always the Fastest: Probability-Modeled Stereoscopic Eye Movement Completion Time in VR. *ACM Transactions on Graphics (TOG)* 42, 6 (2023), 1–14.

[23] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41, 4 (01 Nov 2009), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

[24] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39, 2 (01 May 2007), 175–191. https://doi.org/10.3758/BF03193146

[25] Camilo Fosco, Vincent Casser, Amish Kumar Bedi, Peter O'Donovan, Aaron Hertzmann, and Zoya Bylinskii. 2020. Predicting visual importance across graphic design types. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology.* 249–260.

[26] Agostino Gibaldi and Silvio P Sabatini. 2021. The saccade main sequence revised: A fast and repeatable tool for oculomotor analysis. *Behavior Research Methods* 53 (2021), 167–187.

[27] Alexander Goettker, Kevin MacKenzie, and T. Scott Murdison. 2020. Differences between oculomotor and perceptual artifacts for temporally limited head mounted displays. *Journal of the Society for Information Display* 28 (06 2020), 509–519. https://doi.org/10.1002/jsid.912

[28] Alexander Goettker, Hemanth Pidaparthy, Doris I. Braun, James H. Elder, and Karl R. Gegenfurtner. 2021. Ice hockey spectators use contextual cues to guide predictive eye movements. *Current Biology* 31, 16 (23 Aug 2021), R991–R992. https://doi.org/10.1016/j.cub.2021.06.087

[29] {Stephen J.} Heinen, Elena Potapchuk, and {Scott N.J.} Watamaniuk. 2016. A foveal target increases catch-up saccade frequency during smooth pursuit. *Journal of Neurophysiology* 115, 3 (1 March 2016), 1220–1227. https://doi.org/10.1152/jn.00774.2015 Publisher Copyright: © 2016 the American Physiological Society..

[30] Aaron Hertzmann. 2024. Toward a theory of perspective perception in pictures. *Journal of Vision* 24, 4 (2024), 23–23.

[31] Zhiming Hu, Sheng Li, Congyi Zhang, Kangrui Yi, Guoping Wang, and Dinesh Manocha. 2020. DGaze: CNN-Based Gaze Prediction in Dynamic Scenes. *IEEE Trans Vis Comput Graph* 26, 5 (Feb. 2020), 1902–1911.

[32] Susmija Jabbireddy, Xuetong Sun, Xiaoxu Meng, and Amitabh Varshney. 2022. Foveated rendering: Motivation, taxonomy, and research directions. *arXiv preprint arXiv:2205.04529* (2022).

[33] Daniel Jiménez Navarro, Xi Peng, Yunxiang Zhang, Karol Myszkowski, Hans-Peter Seidel, Qi Sun, and Ana Serrano. 2024. Accelerating Saccadic Response through Spatial and Temporal Cross-Modal Misalignments. In *ACM SIGGRAPH 2024 Conference Papers.* 1–12.

[34] Anton S Kaplanyan, Anton Sochenov, Thomas Leimkühler, Mikhail Okunev, Todd Goodall, and Gizem Rufo. 2019. DeepFovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.

[35] Min H Kim, Tim Weyrich, and Jan Kautz. 2009. Modeling human color perception under extended luminance levels. In *ACM SIGGRAPH 2009 papers.* 1–9.

[36] Frederick A.A. Kingdom and Nicolaas Prins. 2016. Chapter 4 - Psychometric Functions. In *Psychophysics (Second Edition)* (second edition ed.), Frederick A.A. Kingdom and Nicolaas Prins (Eds.). Academic Press, San Diego, 55–117. https://doi.org/10.1016/B978-0-12-407156-8.00004-9

[37] Brooke Krajancich, Petr Kellnhofer, and Gordon Wetzstein. 2023. Towards attention–aware foveated rendering. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–10.

[38] Li Li and Raine Chen. 2016. Action video game play increases the connection of pursuit eye movements and dynamic visual processing with visuomotor control. *Journal of Vision* 16, 12 (2016), 1357–1357.

[39] Stephen G Lisberger, Edward Joseph Morris, and Lawrence Tychsen. 1987. Visual motion processing and sensory-motor integration for smooth pursuit eye movements. *Annual review of neuroscience* 10 (1987), 97–129.

[40] Casimir JH Ludwig, J Rhys Davies, and Miguel P Eckstein. 2014. Foveal analysis and peripheral selection during active visual sampling. *Proceedings of the National Academy of Sciences* 111, 2 (2014), E291–E299.

[41] Rafał K Mantiuk, Maliha Ashraf, and Alexandre Chapiro. 2022. stelaCSF: A unified model of contrast sensitivity as the function of spatio-temporal frequency, eccentricity, luminance and area. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–16.

[42] Rafał K Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. Fovvideovdp: A visible difference predictor for wide field-of-view video. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–19.

[43] Rafal K Mantiuk, Param Hanji, Maliha Ashraf, Yuta Asano, and Alexandre Chapiro. 2024. ColorVideoVDP: A visual difference predictor for image, video and display distortions. *arXiv preprint arXiv:2401.11485* (2024).

[44] Daniel Martin, Ana Serrano, Alexander W Bergman, Gordon Wetzstein, and Belen Masia. 2022. Scangan360: A generative model of realistic scanpaths for 360 images. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2003–2013.

[45] Marcus Missal and Stephen J Heinen. 2017. Stopping smooth pursuit. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372, 1718 (2017), 20160200.

[46] Omri Nachmani, Jonathan Coutinho, Aarlenne Z Khan, Philippe Lefèvre, and Gunnar Blohm. 2020. Predicted position error triggers catch-up saccades during sustained smooth pursuit. *eneuro* 7, 1 (2020).

[47] Chloé Pasturel, Anna Montagnini, and Laurent Udo Perrinet. 2020. Humans adapt their anticipatory eye movements to the volatility of visual motion properties. *PLoS computational biology* 16, 4 (2020), e1007438.

[48] Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–12.

[49] Andriy Pavlovych and Wolfgang Stuerzlinger. 2011. Target following performance in the presence of latency, jitter, and signal dropouts. In *Proceedings of Graphics Interface 2011.* 33–40.

[50] Roger Ratcliff. 1978. A theory of memory retrieval. *Psychological review* 85, 2 (1978), 59.

[51] Do A Robinson. 1965. The mechanics of human smooth pursuit eye movement. *The Journal of Physiology* 180, 3 (1965), 569.

[52] Tim J Smith. 2013. Watching you watch movies: Using eye tracking to inform film theory. (2013).

[53] David Souto and Dirk Kerzel. 2021. Visual selective attention and the control of tracking eye movements: A critical review. *Journal of Neurophysiology* 125 (03 2021). https://doi.org/10.1152/jn.00145.2019

[54] Miriam Spering, Dirk Kerzel, Doris I Braun, Michael J Hawken, and Karl R Gegenfurtner. 2005. Effects of contrast on smooth pursuit eye movements. *Journal of vision* 5, 5 (2005), 6–6.

[55] Taimoor Tariq and Piotr Didyk. 2024. Towards Motion Metamers for Foveated Rendering. *ACM Transactions on Graphics (TOG)* 43, 4 (2024), 1–10.

[56] Paul D. Tynan and Robert Sekuler. 1982. Motion processing in peripheral vision: Reaction time and perceived velocity. *Vision Research* 22, 1 (1982), 61–68. https://doi.org/10.1016/0042-6989(82)90167-5

[57] Boris B Velichkovsky, Nikita Khromov, Alexander Korotin, Evgeny Burnaev, and Andrey Somov. 2019. Visual fixations duration as an indicator of skill level in esports. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part I 17.* Springer, 397–405.

[58] David R Walton, Rafael Kuffner Dos Anjos, Sebastian Friston, David Swapp, Kaan Akşit, Anthony Steed, and Tobias Ritschel. 2021. Beyond blur: Real-time ventral metamers for foveated rendering. *ACM Transactions on Graphics* 40, 4 (2021), 1–14.

[59] Lili Wang, Xuehuai Shi, and Yi Liu. 2023. Foveated rendering: A state-of-the-art survey. *Computational Visual Media* 9, 2 (2023), 195–228.

[60] David Whitney. 2002. The influence of visual motion on perceived position. *Trends in cognitive sciences* 6, 5 (2002), 211–216.

[61] Xiuyun Wu and Miriam Spering. 2022. Tracking and perceiving diverse motion signals: Directional biases in human smooth pursuit and perception. *Plos one* 17, 9 (2022), e0275324.

A Eye-Tracking Data Pre-processing

We implement an objective method for determining whether the eye position recordings are usable in further analysis. First, we segment the eye position traces of each trial, relative to stimulus onset time, t, into an initial fixation phase ($-0.2 \leq t \leq 0.1$ s) [4, 17], a catch-up onset phase ($0.1 \leq t \leq 0.6$ s) [17], and a steady-state tracking phase ($0.6 \leq t \leq 0.9$ s), [17, 54].

Using the eye position recordings from each segment, we require that the median fixation position be within 2 dva of the central fixation positions, and the spread of eye positions, measured in standard deviations of the gaze distribution, to be less than $\sigma < 1$ dva. In the application studies of Section 5, the fixation phase tolerances were doubled due to larger tracker errors observed due to the more complex visual stimuli presented throughout the trials. During the steady-state tracking phase, the median offset of the eye position recording relative to the target stimulus was required to be within 4 dva.

To ensure accurate temporal alignment, the study was configured to record the screen position of the stimulus at each frame, synchronized with the eye tracker's position data at the corresponding frame. Data points in which the eye tracker detected invalid gaze samples or blinks were identified and discarded prior to analysis.

B Main Results

We present the detailed ANOVA test results presented in Section 3.2.

Table 1. *Experiment significance.* ANOVA results for the three experiments of Section 3 show significance of both target speed, v, and signal strength, s, as well as a strong interaction effect between the variables.

Participant	Experiment	Factor	F-Statistic	p-value	η_p^2	Participant	Experiment	Factor	F-Statistic	p-value	η_p^2
	LUM	v	6.27	0.0025	0.0530		LUM	v	3.48	0.032	0.0299
		s	133.54	< 0.001	0.6414			s	114.51	< 0.001	0.6032
		$v \times s$	13.80	< 0.001	0.2698			$v \times s$	2.23	0.041	0.0560
	NOISE	v	10.94	< 0.0001	0.0886		NOISE	v	11.71	< 0.0001	0.0932
P_1		s	168.47	< 0.0001	0.6920	P_4		s	69.99	< 0.0001	0.4794
		$v \times s$	3.14	0.0056	0.0774			$v \times s$	1.26	0.028	0.0322
	COLOR	v	10.94	< 0.0001	0.1364		COLOR	v	2.39	0.09	0.0206
		s	168.47	< 0.0001	0.6374			s	68.57	< 0.0001	0.4743
		$v \times s$	3.14	0.0056	0.2412			$v \times s$	5.54	< 0.0001	0.1274
	LUM	v	1.52	0.025	0.0133		LUM	v	2.71	0.069	0.0245
		s	90.94	< 0.0001	0.5469			s	16.15	< 0.0001	0.1832
		$v \times s$	2.37	0.003	0.0594			$v \times s$	1.59	0.15	0.0423
	NOISE	v	3.64	0.028	0.0316		NOISE	v	8.63	< 0.001	0.0802
P_2		s	140.68	< 0.0001	0.6543	P_5		s	20.16	< 0.0001	0.2340
		$v \times s$	2.57	0.020	0.0647			$v \times s$	1.32	0.25	0.0386
	COLOR	v	4.33	0.014	0.0374		COLOR	v	6.93	0.0012	0.0598
		s	156.27	< 0.0001	0.6777			s	26.75	< 0.0001	0.2691
		$v \times s$	12.80	< 0.0001	0.2563			$v \times s$	7.94	< 0.0001	0.1794
	LUM	v	2.49	0.085	0.0224						
		s	43.85	< 0.001	0.3764						
		$v \times s$	3.45	0.0028	0.0869						
	NOISE	v	2.49	0.0053	0.0452						
P_3		s	43.85	< 0.0001	0.2957						
		$v \times s$	3.45	0.15	0.0402						
	COLOR	v	5.76	0.0036	0.0493						
		s	121.87	< 0.0001	0.6222						
		$v \times s$	4.83	< 0.001	0.1157						

C Participant Fatigue

We present the heat maps of catch-up latency based on trial progress per condition type from Section 3.
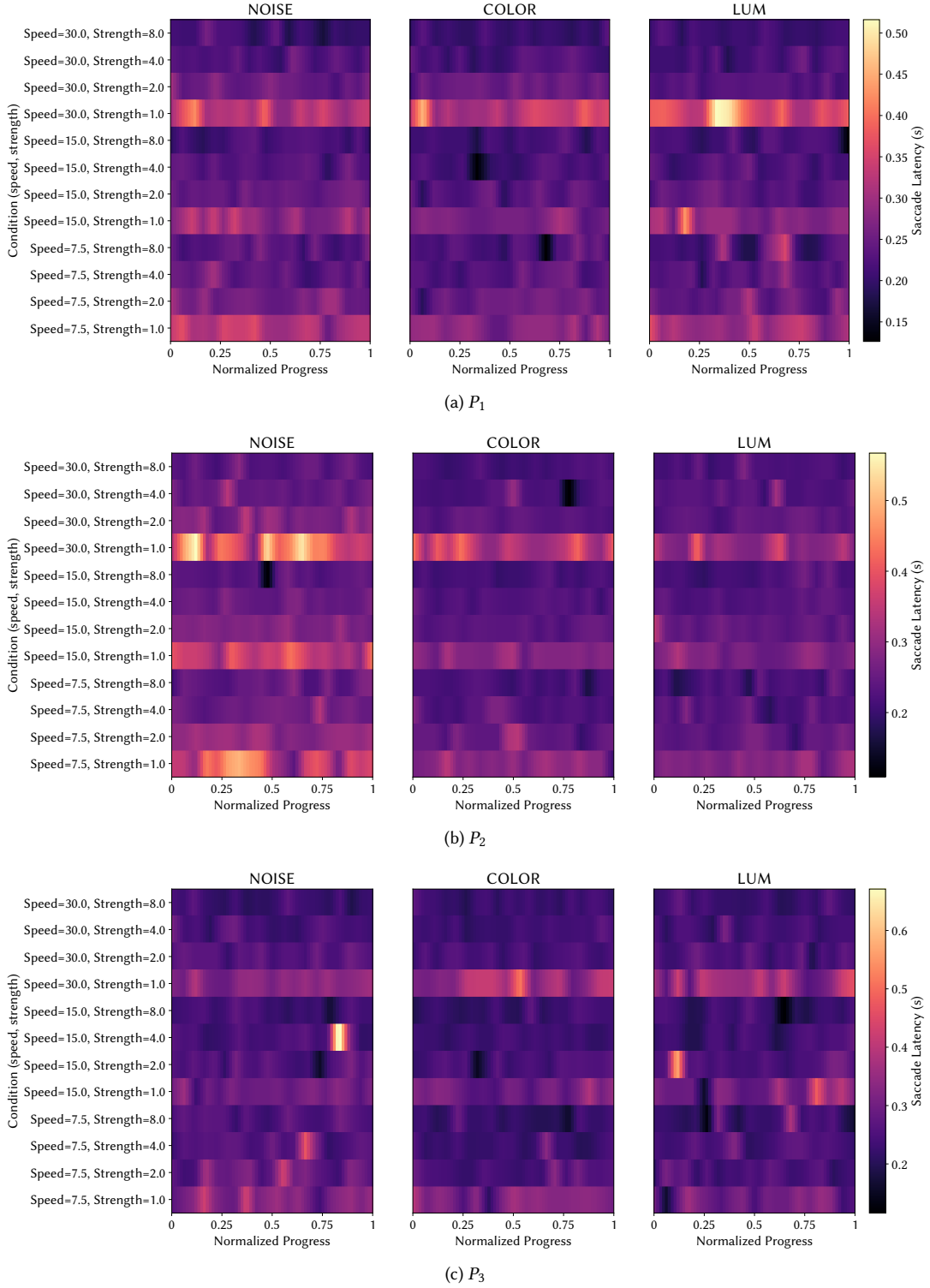
(a) P_1

(b) P_2

(c) P_3

Fig. 9. *Catch-up performance of main study participants.* Each panel shows a heatmap of normalized experiment progress (x-axis) versus object speed and signal strength (y-axis). The three subplots per participant correspond to LUM, NOISE, and COLOR conditions, using a shared color scale for mean saccade latencies (s). Rows denote speed–strength combinations, columns show normalized trial progress (0–1), and lighter colors indicate longer latencies.
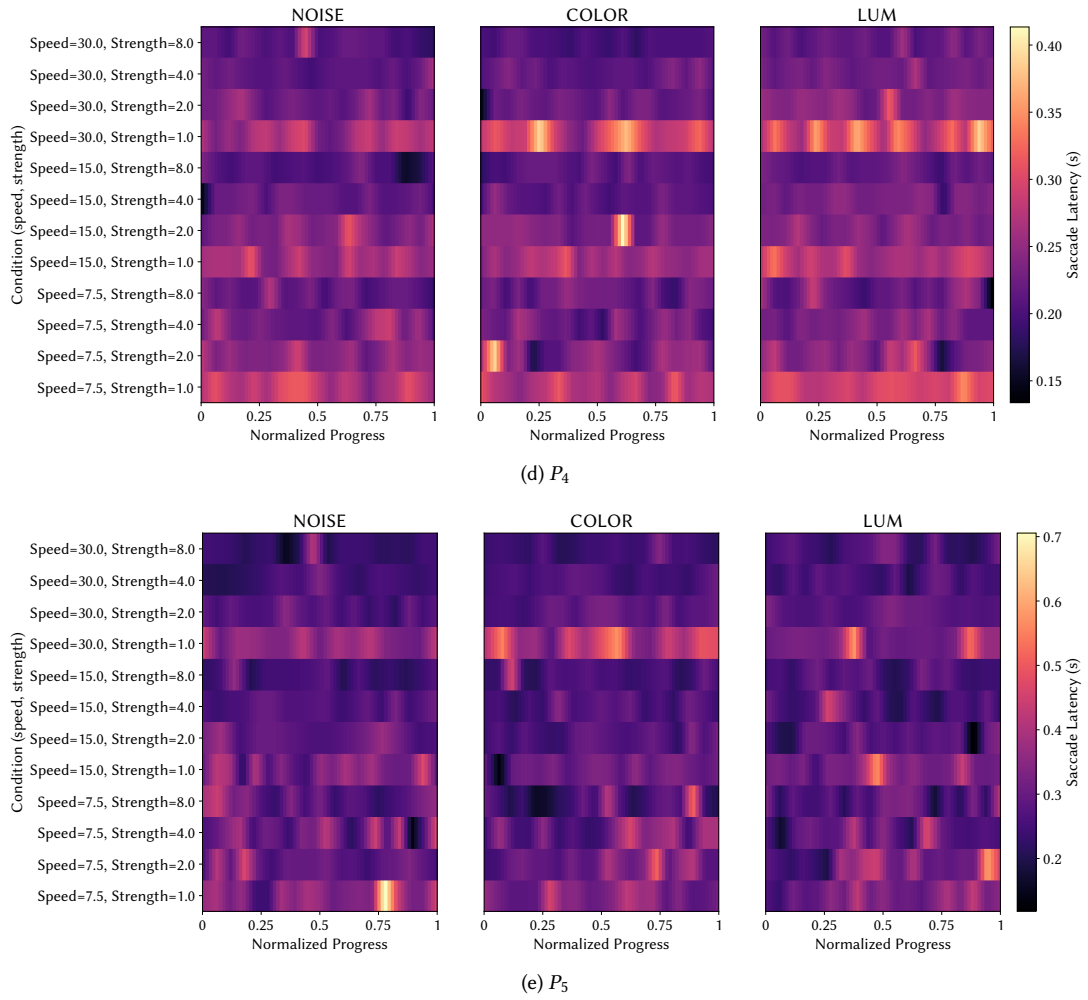
(d) P_4

(e) P_5

Fig. 9. *Catch-up latency time based on progress per condition type.* Each panel shows a heatmap of normalized experiment progress (x-axis) versus object speed and signal strength (y-axis). The three subplots per participant correspond to LUM, NOISE, and COLOR conditions, using a shared color scale for mean saccade latencies (s). Rows denote speed–strength combinations, columns show normalized trial progress (0–1), and lighter colors indicate longer latencies. The consistent latencies over time demonstrate low fatigue of subjects.