Image Features Influence Reaction Time: A Learned Probabilistic Perceptual Model for Saccade Latency

BUDMONDE DUINKHARJAV, New York University, USA
PRANEETH CHAKRAVARTHULA, Princeton University, USA
RACHEL BROWN, NVIDIA, USA
ANJUL PATNEY, NVIDIA, USA
QI SUN, New York University, USA

We aim to ask and answer an essential question “how quickly do we react after observing a displayed visual target?” To this end, we present psychophysical studies that characterize the remarkable disconnect between human saccadic behaviors and spatial visual acuity. Building on the results of our studies, we develop a perceptual model to predict temporal gaze behavior, particularly saccadic latency, as a function of the statistics of a displayed image. Specifically, we implement a neurologically-inspired probabilistic model that mimics the accumulation of confidence that leads to a perceptual decision. We validate our model with a series of objective measurements and user studies using an eye-tracked VR display. The results demonstrate that our model prediction is in statistical alignment with real-world human behavior. Further, we establish that many sub-threshold image modifications commonly introduced in graphics pipelines may significantly alter human reaction timing, even if the differences are visually undetectable. Finally, we show that our model can serve as a metric to predict and alter reaction latency of users in interactive computer graphics applications, thus may improve gaze-contingent rendering, design of virtual experiences, and player performance in e-sports. We illustrate this with two examples: estimating competition fairness in a video game with two different team colors, and tuning display viewing distance to minimize player reaction time.

Fig. 1. Illustration of our model and applications. Our model predicts the reaction latency for users to identify and saccade to a peripheral target, as shown in (a). Based on our psychophysical data collected for stimuli with varying visual characteristics, we model the likelihood distribution of the time users take to process, react, and saccade to a target. If replacing the black placeholder in (a) with the three target candidates shown in (b), the resulting retinal images exhibit identical perceptual similarity in terms of visual acuity, with all FoxVideoVDP scores > 9.5 per [Mantiuk et al. 2021]. However, they may trigger significantly faster (leftmost of (b)) or slower (rightmost of (b)) reaction latencies with up to about 35ms difference, significantly affecting task performance.

3D asset credits to Slavyer at Sketchfab Inc as well as Counter-Strike: Global Offensive © Valve Corporation.

Authors’ addresses: Budmonde Duinkharjav, budmonde@gmail.com, New York University, USA; Praneeth Chakravarthula, cpk@cs.unc.edu, Princeton University, USA; Rachel Brown, rachelabrown347@gmail.com, NVIDIA, USA; Anjul Patney, anjul.patney@gmail.com, NVIDIA, USA; Qi Sun, qisun@nyu.edu, New York University, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.
0730-0301/2022/7-ART144 $15.00
https://doi.org/10.1145/3528223.3530055

CCS Concepts: • Computing methodologies → Perception; Virtual reality.

Additional Key Words and Phrases: Virtual Reality, Augmented Reality, Visual Perception, Human Performance, Esports, Gaze-Contingent Rendering

ACM Reference Format:
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., Vol. 41, No. 4, Article 144. Publication date: July 2022.
1 INTRODUCTION

Measuring, modeling, and predicting how humans perceive and act on displayed visual content are important tasks in computer graphics, with applications in cinematic, real-time rendering, virtual/augmented reality (VR/AR), display optimization, esports, video compression/streaming, and visual design [Dunn et al. 2020; Mantiuk et al. 2004; Patney et al. 2016; Serrano et al. 2017; Sitzmann et al. 2018]. Perceptual image quality metrics predict the likelihood of visibility of image artifacts that result from creative and technical design, or are a side-effect of rendering, processing, or transmission. While many such metrics already exist, research is primarily focused on modeling the spatial/temporal acuity of the human visual system (HVS), not on how viewers “react” after seeing the stimuli. Although visibility may be closely related to behavior, there is evidence that perceptually identical stimuli frequently result in significantly different reactions for viewers [Mulckhuyse and Theeuwes 2010; Sperling and Carrasco 2015]. Since responses are critical in many interactive applications such as esports and user interfaces, metrics that predict user reactive performance are arguably in emerging and crucial demand.

Researchers have so far exhaustively studied the acuity of the human visual system and established a significant body of perceptual image-quality metrics [Hore and Ziou 2010], as well as perceptually-optimized computer graphics techniques [Krajancich et al. 2021; Patney et al. 2016]. Such methods have unlocked significant performance and memory optimizations, as well as quality improvements. A recent example is gaze-contingent rendering, which perceptually optimizes rendering complexity for wide field-of-view (FoV) AR/VR displays. The rendering methods achieve great savings in computation [Patney et al. 2016], bandwidth [Chen et al. 2022; Kaplanyan et al. 2019; Krajancich et al. 2021], or enhancing visual cues [Krajancich et al. 2020]. Researchers have also proposed perceptually-based metrics that predict the visibility of artifacts in a user’s peripheral vision [Mantiuk et al. 2021; Tursun et al. 2019]. However, to our best knowledge, there is still very limited characterization of changes to human reaction times when observing naturalistic visual stimuli that are otherwise perceptually identical.

This paper proposes an analytical model for a user’s reaction time as evidenced by their eye movements. Human eyes change visual fixation three to four times every second [Fabius et al. 2019] via rapid exploratory movements called saccades. Saccades allow for perceptual shifts of attention to better understand one’s surroundings and to localize objects of interest, e.g., potential dangers [Purves et al. 2008]. Saccadic reaction latencies, after the eye observes an stimulus, are closely tied to performance in a broad range of real-time applications. For instance, subtly (as low as 4ms [Kim et al. 2019]) altered saccade latency can significantly determine performance in competitive esports [Koposov et al. 2020]. Each saccade involves perceiving a stimulus, identifying the target [Lisi et al. 2019], sending oculomotor neural signals, and controlling the extraocular muscles to reorient the eyeballs. Due to these complex mechanisms, fully characterizing changes in saccade/fixation as a function of changes in visual stimuli remains an open problem in vision science and computer graphics.

Note that, unlike with visual quality metrics, both high and low visibility of a target could hypothetically induce a longer processing time for fine details or blurred content. That may lead to potential non-correlation between acuity and saccadic latency [Kalesnykas and Hallett 1994]. Therefore, we present a visual-oculomotor and probabilistic model of the correlation between gaze-contingent visual stimuli and the timing of humans’ cognitive decisions to perform saccades. Our model adapts to different viewing tasks that involve reacting to a stimulus in the visual field. Psychophysical experiments demonstrate our model’s statistical effectiveness and real-world applications such as customizing players’ action timing, measuring esports competition fairness, and predicting performance with various display environments.

We first conduct a preliminary psychophysical study. The results reveal that peripheral artifact visibility fails to fully explain changes in saccadic reaction latency. By leveraging the data, we develop a probabilistic and closed-form visual-oculomotor model predicting the likelihood of human saccadic latency after observing a given visual stimulus. The model is neurologically inspired to depict a decision-making process and is further established via machine learning with our studied data. We validate our model using various evaluation data partitioning and several realistic visual search tasks. Our results demonstrate that meaningful changes in behavior may occur even without visible artifacts, establishing the importance of predicting user gaze behaviors. Our main contributions are:

- a series of psychophysical experiments characterizing the disconnect between visual acuity and eye movement latencies toward a target;
- a neurologically-inspired and closed-form probabilistic model of human reaction time, built to effectively predict saccade onset latency by learning from real user data;
- studies with natural images and complex tasks showing how subtle target appearance changes may significantly alter reaction performance, following our model’s predictions;
- demonstrations of the model’s additional applications to measuring competitive game fairness and optimizing player performance in esports.

In addition, we provide the full source code and all raw user study data to the community at www.github.com/NYU-ICL/gaze-timing.

2 RELATED WORK

We propose a model for eye motion based on the characteristics of the human visual system, and our primary applications lie in the area of interactive computer graphics. Our model is an alternative to metrics for visibility of image artifacts, which serve for similar applications. Hence, we begin by discussing state-of-the-art in peripheral image perception and corresponding image quality metrics, as well as recent advances in the area of gaze-contingent rendering. We then discuss existing models for saccadic gaze motion and their applications in computer graphics and related areas.

2.1 Human Vision and Image Quality Metrics

The HVS is a complex bio-system that perceives the visual world using a series of optical components, retinal photoreceptors,
neural structures, including the optic nerve and visual cortex. Almost all of these are highly non-uniform across the visual field. As a consequence, our peripheral vision exhibits numerous peculiarities— in addition to having significantly lower acuity than foveal (central) vision, it is known to have nonuniform acuity for detection vs. resolution tasks [Thibos et al. 1987b,a], high sensitivity for moving stimuli [McKee and Nakayama 1984] including high critical flicker-fusion frequency [Hartmann et al. 1979], and reduced color perception [Cohen et al. 2020; Noorlander et al. 1983].

Several models describe aspects of peripheral visual perception and predict the visibility of stimuli at various visual eccentricities. Many build on contrast sensitivity measurements across the visual field [Barten 1999; Cajar et al. 2016; Daly 1992; Kelly 1979]. While most metrics only consider spatial image characteristics [Rimac-Drije et al. 2010; Rimac-Drije et al. 2011; Wang et al. 2001], some recent developments also account for spatio-temporal perception [Krajancich et al. 2021; Mantiuk et al. 2021]. However, all these metrics for peripheral image quality only account for visibility of artifacts and not for change of user behavior that might exist between perceptually identical stimuli.

2.2 Gaze-Contingent Computer Graphics

Gaze-contingent rendering utilizes high-speed eye-tracking to identify the user’s gaze locations and to modify the rendered images or display accordingly. Such techniques can often provide additional perceptual cues to improve immersion of displayed content, e.g., perceived parallax [Konrad et al. 2020], stereo depth perception [Krajancich et al. 2020], and depth-of-field [Duchowski et al. 2014; Hillaire et al. 2008; Mauderer et al. 2014]. Leveraging the perceptual differences between foveal and peripheral vision also improves interactive computer graphics methods, which forms the basis for gaze-contingent foveated rendering and displays. Foveated methods reduce the complexity of graphics or display in the peripheral visual field, which results in an overall improvement of performance or perceived quality [Franke et al. 2021; Guenter et al. 2012; Kaplanyan et al. 2019; Koskela et al. 2019, 2016; Meng et al. 2018; Patney et al. 2016; Polychronakis et al. 2021; Sun et al. 2017, 2020; Walton et al. 2021; Weier et al. 2016].

Existing gaze-contingent and foveated rendering algorithms rely on how gaze location causes differences in spatiotemporal visual perception. Typically, these methods are evaluated on the noticeability of any visual artifacts they may introduce. Instead, we attempt to model gaze behavior changes, e.g., saccadic latency, after observing a stimulus. The reactive changes occur due to both subthreshold and suprathreshold differences in visual stimuli, which can be leveraged to predict or improve user performance in immersive visual tasks.

2.3 Models for Saccadic Eye Motion

Saccadic eye motion exhibits several unique qualities. We find limited studies in the literature that model the latency before the onset of a saccade. Further, they do not characterize image features in a way that may build a general-purpose metric. We discuss research that explains the mechanism for latency that leads to a saccade and build on this mechanism for our model.

First, the large, rapid gaze changes that occur during a saccade are ballistic and exhibit a predictable trajectory [Bahill et al. 1975; Kowler 2011]. Saccade amplitude, velocity, and duration are nonlinearly related, and velocities of short saccades tend to have an asymmetric, bell-shaped velocity profile [Bahill et al. 1975]. This behavior lends to models that can characterize a saccade profile by only partially observing it, and with a sufficiently fast eye-tracker, can predict where an ongoing saccade will land via regression [Arabzhiyska et al. 2017] or neural network [Morales et al. 2021]. It can also help improve the perceived latency of gaze-contingent computer graphics, as studied by Albert et al. [2017].

Second, during a saccade as well as for a short period after, the HVS experiences temporary perceptual blindness, as known as saccadic suppression, and has been well-studied in literature [Burr et al. 1994; Diamond et al. 2000; Ibbotson and Cloherty 2009; Matin 1975]. Saccadic suppression naturally helps gaze-contingent graphics be tolerant to higher eye-tracking latencies [Albert et al. 2017], and has also been leveraged in VR redirected walking [Sun et al. 2018].

Third, saccadic eye movements have been shown to be inaccurate and often undershoot their target [Becker and Fuchs 1969; Deubel et al. 1982]. The magnitude of error in the landing position depends on the degree of uncertainty in the target location as well as sensory noise [van Beers 2007] and adaptation [Cotti et al. 2009]. Researchers have proposed to model this uncertainty based on visual characteristics of the target [Carpenter 2004; Lisi et al. 2019]. Such models could help predict as well as correct for saccade errors by reducing the frequency and size of missed and catch-up saccades. However, we are not aware of such practical application of these models in prior work.

Finally, another area of saccade behavior models is subtle gaze direction [Bailey et al. 2009], which seeks to stimulate saccades toward specific targets by modeling how the spatio-temporal characteristics of a stimulus may attract a user’s gaze.

While the above studies thoroughly explain and utilize various saccadic behaviors, we are not aware of computational models of pre-saccadic reaction time as a function of changes in naturalistic stimuli. Predicting this latency is potentially valuable for a variety of applications, such as esports [Koposov et al. 2020].

Various research endeavors to explain the cognitive decision that precedes and results in a saccadic eye movement. Several neuroeconomic approaches describe a generic decision-making process where the perceptual measure of visual evidence is accumulated stochastically. Once a threshold amount of evidence is collected, a decision is made [Mazurek et al. 2003]. The diffusion model [Palmer et al. 2005; Ratcliff 1978] suggests that the stochastic process may be modeled as a random walk akin to Brownian Motion. Carpenter [1995] assumes the rate of accumulation of visual evidence as a random variable which stays constant throughout individual saccades, as verified in [Reddi et al. 2003]. The above mechanisms suggest that sub-threshold changes in a stimulus could non-trivially influence the latency of saccade onset, as has been verified by [Mulckhuysen and Theeuwes 2010; Spering and Carrasco 2015]. Further, since stronger stimuli accumulate faster than weaker ones, the amount of time required must be correlated with various visual characteristics [Bell et al. 2006; Carpenter 2004; Mahadevan et al. 2018] and task modalities [Yamagishi and Furukawa 2020]. There is also evidence
We first conduct a psychophysical experiment with parameterized details are specified in Table 1. During the study, participants remained seated and perceived stimuli through the stereo display. Before each experiment, a five-point eye-tracking calibration was applied on each individual.

Participants. The psychophysical study was performed with $n = 5$ participants (ages 22 – 28, 3 female) with normal or corrected-to-normal vision. The participants were instructed to perform a series of two-alternative forced choice (2AFC) tasks for each trial. The experiment was conducted during a single session split into 10 blocks, with each block containing 225 trials, i.e., 11250 trials in total with all the participants. The procedure took around 2 hours. With competitive scenarios such as esports, we further demonstrate how our model can guide virtual asset design toward optimizing performance and competition fairness.

3 PILOT STUDY: MEASURING SACCADIC LATENCY

We first conduct a psychophysical experiment with parameterized stimuli to observe and measure the correlation between image characteristics and the time it takes to process them in order to trigger a saccade, and whether/how the correlation differs from that of visual acuity. We anticipate the data collected from the participants to serve as the inspirational and statistical basis toward a closed-form predictive model.

Setup. The study was performed with an eye-tracked HTC Vive Pro Eye head-mounted display as shown in Figure 2a. The hardware details are specified in Table 1. During the study, participants remained seated and perceived stimuli through the stereo display. Before each experiment, a five-point eye-tracking calibration was applied on each individual.

Table 1. Specifications of the HTC Vive Pro Eye display used in our studies.

| Feature                | Value          |
|------------------------|----------------|
| Display Resolution     | 1440 × 1600 pixels per eye |
| Refresh Rate           | 90 Hz          |
| Peak Luminance         | 143 cd/m²      |
| Field of View          | 110° diagonal  |
| Eye Tracker Frequency  | 120 Hz         |

that the latency is a function of the target eccentricity and demographic [Mazumdar et al. 2019] but exhibits a distinct trend from that of visual acuity [Kalesnykas and Hallett 1994]. Inspired by these discoveries, we develop an analytical and data-driven model for saccadic reaction latency, dependent on visual contrast, frequency, and target locations. With competitive scenarios such as esports, we further demonstrate how our model can guide virtual asset design toward optimizing performance and competition fairness.

3.1 Results

Using a statistical saccade detection method in Section 5.3, we identify the saccadic latency as the duration between appearance of the primary stimulus (Gabor patch) and the first frame of a participant’s saccade. We notice that the saccade latencies exhibit an asymmetrical distribution as shown in Figure 2c. As the various features of the stimulus are modulated, the overall shape of the distribution remained consistent while the mean saccade latency varied by as much as 25% or 100 ms as shown in Figure 3. Increasing the contrast of the stimuli decreases reaction latency, while increasing the frequency increases the latency. Further, increasing the eccentricity does not always reveal a monotonic effect, but instead a U-shaped effect with the lowest mean latency values (265 ms) plateauing at 10°. For breakdown visualizations, please refer to Figure 3/Appendix E for the effects of individual characteristics and participants.

We also analyzed FoVVideoDVP [Mantiuk et al. 2021] scores for all stimuli used in the experiment. Using the median condition’s image due to the patches not being detectable by all participants. The eccentricity range was chosen to cover common scenarios since the human gaze does not typically go outside 10° from the center [Hatada et al. 1980], and most natural saccade sizes are less than 15° [Bahill 1975]. Unless otherwise specified, we use Weber contrast in all our experiments and as input to our model.

At the beginning of each trial, the participants fixated at a cross shown in the center of the screen. Once they successfully fixated on the cross, it disappeared and a pair of Gaussian patches appeared at 10° eccentricities to the left and right of fixation. These patches served as the target locations to which the participants would saccade to indicate their decision about the stimulus. After a small delay—chosen randomly between 300 and 500 ms to avoid learning effects—the primary stimulus (Gabor patch) appeared either at the center of the screen (eccentricity=0°), or at some eccentricity in the inferior peripheral vision (eccentricity=10° or 20°). We instructed the participants to identify whether the Gabor stimulus was oriented at a rotation of 45° clockwise from vertical, as shown in Figure 2a, or 45° counter-clockwise from vertical. We further instructed them to saccade to the target patch corresponding to their determination, right for clockwise and left for counter-clockwise. During each trial we recorded the subjects gaze at a rate of 120 FPS using the display’s built-in eye tracker.

We varied the eccentricity, contrast, and frequency of Gabor patterns across trials such that each combination of variables was shown 5 times in each block for 10 blocks, yielding a total of 50 trials per condition. To ensure the participants were completing the task correctly, we discarded all trials where they do not complete the task correctly, and repeat all mistaken trials at the end of the block until all trials are completed. The order of these conditions was randomly shuffled within each block to eliminate any bias. Meanwhile, all features of the Gaussian target patches (only being used to cue the saccade direction) remained unchanged throughout the trials. For the practice session at the beginning of the experiment, each participant performed one block of the study with identical settings as in the actual study. Please refer to our supplementary video for an animated illustration.

ACM Trans. Graph., Vol. 41, No. 4, Article 144. Publication date: July 2022.
The above results and analysis reveal several remarkable discoveries (out of 10) value across all the stimuli being studied. Yet, they result in a significant difference in saccadic latency. This visibility of the stimuli is improved, and vice versa. Surprisingly, the reaction latency when the stimulus is at the fovea is higher as compared to when it is in mid-periphery.

\begin{align*}
\text{Frequency (\text{cpd})} & \quad \text{Contrast} \\
0.05 & \quad 0.22 & \quad 0.53 & \quad 1.00
\end{align*}

(c = 0.53, f = 2 cpd, \(e = 10^\circ\)) as reference, we observed 9.52 ± 0.03 (out of 10) value across all the stimuli being studied.

3.2 Discussion

The above results and analysis reveal several remarkable discoveries on the relationships between visual characteristics and saccadic latency. First, using a state-of-the-art peripheral image similarity metric [Mantiuk et al. 2021], we conclude that visual differences between our stimuli are all well below the perceivable thresholds. Yet, they result in a significant difference in saccadic latency. This evidence confirms that perception of visual differences cannot alone explain the changes in saccadic latencies. Second, the asymmetrical probability distribution agrees with the discoveries of prior work in measuring similar visual-oculomotor reactive latencies [Carpenter and Williams 1995; Lisi et al. 2019; Palmer et al. 2011]. Third, at a given eccentricity, as the visibility of the stimuli improves (either by increasing contrast or by modulating the frequency), the latency decreases. Meanwhile, the latency rises toward infinity whenever visibility reduces and approaches the Contrast Sensitivity Function (CSF) threshold. Lastly, we observe a surprising effect that the saccade latencies for a stimulus at the fovea are longer than in mid-periphery. We hypothesize that the more analytic purpose of the fovea causes feature extraction to take longer, similar to the results reported by Kalensnykas et al. [1994].

The collected data and the observations drive our development of a closed-form probabilistic model inspired by the computational process of decision making, as detailed in the next section.

4 PROBABILISTIC MODEL OF SACCADIC REACTION

Driven by our preliminary study discoveries, we aim to establish the computational relationship between saccadic reaction latencies and visual characteristics of stimuli. In Section 4.1, we model the saccadic behaviors as a random process of decision-making and incorporate perceptual, cognitive, and individual uncertainties. Section 4.2 shows how image characteristics such as contrast, frequency and eccentricity relate to our model of saccadic onset latency for a task. Sections 4.3 and 4.4 detail how we leverage our preliminary study data to learn the resulting parameters for our model.

4.1 Random Process Model for General Decision-Making

The Drift Diffusion Model (DDM) has been leveraged in neuroeconomics and psychology to model decision-making behaviors in perceptual tasks [Fudson et al. 2020]. Using this model, we may quantify the reaction latencies for performing various tasks such as...
to decide whether to stop a car upon seeing an approaching object, or to correctly identify a friend from foe in a video game. Throughout a decision process, DDM presents a measure of “evidence” which is used to quantify how much confidence an individual needs to reach a decision. “Evidence”, in this context, is accumulated over time, and once the amount of evidence reaches a minimum threshold, a decision/action is triggered, as visualized in Figure 4. As soon as a decision is made, the evidence is reset, and the process restarts, preparing for the next decision.

The process of evidence accumulation is modeled as a random stochastic process to reflect human reaction uncertainties and inconsistencies. As the name of the DDM suggests, the observed evidence is modeled as a diffusion process with non-zero drift, also known as Brownian motion.

**Integration-and-action.** We represent the measure of evidence accumulation after time \( t \) via a stochastic process, \( \{A(t; v)\}_{t \geq 0} \). The process is modeled as the trajectory following Brownian motion with mean drift rate \( v \). Formally, the process is expressed as

\[
\begin{align*}
A(0; v) &= 0, \\
A(t; v) &= vt + W(t) \\
W(t) &\sim \mathcal{G}(0, t),
\end{align*}
\]

where \( \mathcal{G} \) is the Gaussian distribution.

As illustrated in Figure 4, at the beginning of a decision-making process, no evidence has been accumulated to affect the confidence of reaching a decision. Therefore, the initial evidence amount, \( A(0; v) \), equals to zero. As time, \( t \), progresses, confidence builds up with a mean rate of \( v \), while also accumulating noise proportional to \( t \) due to various uncertainties in the process of evidence accumulation.

The distribution of evidence, \( A(t; v) \), at a given point in time, \( t \), can be simplified and expressed as a Gaussian distribution,

\[
A(t; v) \sim \mathcal{G}(vt, t).
\]

However, we aim to characterize the distribution of the latency when enough evidence is collected to trigger a decision. More concretely, we seek to find the distribution of the earliest time, \( T(\alpha; v) \), when the integrated evidence reaches a given threshold, \( \alpha \):

\[
T(\alpha; v) := \inf_t \{ A(t; v) = \alpha \}.
\]

This measure of earliest reach to the threshold, \( T(\alpha; v) \), corresponds to the reaction latency of the decision being made.

Solving for \( T(\alpha; v) \) using Equations (1) and (3) (see Appendix A for derivation), we find that \( T(\alpha; v) \) follows the Inverse Gaussian (IG), or Wald, distribution [Folks and Chhikara 1978]:

\[
T(\alpha; v) \sim I\mathcal{G}(\alpha, v),
\]

which has a probability density function of

\[
h(t; \alpha, v) = \frac{\alpha}{\sqrt{2\pi t^3}} \exp \frac{-(\alpha - vt)^2}{2t}.
\]

Intuitively, random variables for evidence, \( A(t; v) \) and latency, \( T(\alpha; v) \), roughly describe inverse functions of the same stochastic process, and fittingly follow Gaussian, and Inverse Gaussian distributions respectively.

### 4.2 Saccade as a Visual-Oculomotor Decision-Making

Equation (4) depicts generic decision-making latencies, parameterized with the evidence threshold, \( \alpha \), and mean evidence integration rate, \( v \). A saccade is a representative example of visual-oculomotor decisions. A saccade latency results from processing visual stimuli to gain enough evidence of the target, and then deciding to make an eye movement. In this section, we determine how these two parameters in Equation (4) relate to the characteristics of visual content, such as contrast, frequency, and eccentricity, as well as the nature of the visual task of interest.

The evidence threshold, \( \alpha \), is primarily dependent only on the nature of the visual task. That is, what a visual task is asking an individual to execute determines how confident the individual needs to be in their belief to successfully make a saccadic decision. For instance, Palmer et al. [2011] show that \( \alpha \) is invariant to modulations in signal strength of the visual stimuli but varies across different tasks such as feature search, conjunction search, and spatial configuration search. In fact, Reddi et al. [2003] demonstrate that if an individual is instructed to observe the same visual stimulus while executing different tasks, the evidence integration rate, \( v \), stays constant while the evidence threshold, \( \alpha \), varies between the tasks.

We thus correlate \( \alpha \) with a visual task via a task description, \( D \in \mathcal{D} \), where \( \mathcal{D} \) is the set of all visual task descriptions which trigger viewers’ saccades toward a target. Examples include searching for characters of the opposing team in a given esports game, comparing and choosing a preferred target, or the task described in Section 3.

The evidence integration rate, \( v \), is shown to change depending on the difficulty to process the visual content [Palmer et al. 2011]. Results from our preliminary study (Section 3.2) reveal that visual characteristics have a complex, and non-monotonic relationship with processing difficulty of the stimulus. The effects are also, to some extent, naturally uncertain due to motor/neural noise and individual variances.
In the following sections, we determine how to obtain the analytical forms of $\alpha(D)$ and $v(c, f, e)$ respectively by leveraging our collected data. Further, note that, as intuitively visualized in Figure 4, $v$ already universally determines the relative proportion between the mean and variance of latencies. It depends on the stimuli characteristics, and together with task-dependent $\alpha$ that determines the absolute time.

4.3 Measuring Evidence Threshold $\alpha$ in eq. (6)

To fit our model to the pilot user study data from Section 3, we need to compute $\alpha$ for the task $D_{\text{pilot}}$, and also fit a function which maps $(c, f, e)$ values to corresponding $v$ values.

Since our model is built with a single task description, denoted as $D_{\text{pilot}}$, and is kept invariant throughout the entire study, we only evaluate a single value of $\alpha$. We leverage how the mean, $E$, and variance, $\nabla$, of the IG distribution relate to its $\alpha$ parameter to derive...
an estimation given a sample of saccade latency $T_{sac}(D_{pilot}, \cdot)$, 

$$\hat{a}(D_{pilot}) = \sqrt[3]{\frac{E[T_{sac}(D_{pilot}, \cdot)]}{V[T_{sac}(D_{pilot}, \cdot)]}}.$$  \hspace{1cm} (7) 

Note that $a$ only depends on the task but not the stimuli. Therefore, we can choose any $(c, f, e)$ in order to compute $a$. Extending the model from the pilot study to novel tasks (e.g., stimulus comparison) only requires a calibration via Equation (7) using a sample drawn from the novel task.

4.4 Learning Evidence Integration Rate $\nu$ in eq. (6)

With the task-calibrated $\alpha$, we optimize the evidence integration rate, $\nu$ via maximum likelihood estimation (MLE) given a combination of stimulus’ contrast, frequency, and eccentricity, $(c, f, e)$. Note that since the task description, $D$, does not correlate with $\nu$, the learned result is applicable to any arbitrary scenario and task. Mathematically, we formulate $\nu$ as a function of $(c, f, e)$. To ensure local smoothness of the desired function within this input domain, we model the function as a Radial Basis Function Neural Network (RBFNN),

$$\nu(c, f, e) = \sum_{i=1}^{N} \lambda_i \rho \left( \left\| \frac{c}{\sigma_i} - b_i \right\|, \sigma_i \right),$$ \hspace{1cm} (8) 

where $b_i$ indicates the individual radial basis centers, and $\rho$ is a Gaussian Basis function. In our experiments, we choose $N = 20$. Using our collected data from Section 3, we jointly train the RBFNN’s weights $\lambda$, centers $b_i$, and Gaussian deviations $\sigma$.

We need $\nu$ label values from our user study to obtain the RBFNN parameters. Similar to Section 4.3, we utilize the relationship between the evidence integration rate, $\nu$, and the mean, $E$, of the IG distribution:

$$\nu(c, f, e) \sim E[T_{sac}(\cdot, c, f, e)]^{-1}.$$ \hspace{1cm} (9) 

The proportionality constant of this relationship depends on the unit of time measurement and is set to 1 for simplicity, and scaled to appropriate units as necessary.

Detailed learning implementation can be referred to in Section 5.4. The final fitted model is visualized in Figure 5. Eccentricity’s U-shaped effects on saccadic latency can be observed.

5 IMPLEMENTATION

5.1 Data Normalization

While saccade latencies vary according to the trends observed in Section 3, the absolute values measured across different individuals can vary significantly. In order to aggregate the data efficiently without introducing large amounts of noise caused by individual variances, we normalize the data as a first step during analysis. Specifically, we pick a calibration condition from all the data collected from a single experiment block and set the mean normalized duration of this condition equal to 1. For the pilot experiment from Section 3, the calibration condition was $c, f, e = [1.0, 1.0, 0.0]$. For the dual task experiment from Section 6.3, the calibration condition was $c_f, c_p = [1.0, 1.0]$. For the natural task experiment from Section 6.2, the calibration condition was the control group. The choice of all calibration conditions are arbitrary and we validated that it does not affect the model’s predictive ability. See Appendix B for detailed description of normalization.

5.2 Hardware and Stimuli Generation

All user study systems were implemented in the Unity Game Engine and run on the HTC Vive Pro Eye with specifications in Table 1. The synthetic image datasets used in Section 6.2 are generated using 3D-assets purchased from the Unity Asset Store and rendered using the Cycles Rendering Engine included in Blender. When evaluating the FoVVideoDVP scores of the stimuli used in our work, we used the authors’ open-source implementations which provided configurations for the HTC Vive Pro Eye. Some visualizations used in the demo video were created using the Manim Mathematical Animation Framework [2022].

5.3 Saccadic Latency Detection

Our method of detecting reaction times for saccadic events is measured by the time of onset of the “primary” saccade that is used to move the gaze to the target location. We define the “primary” saccade as the saccade that is onset and offset within 3° of the intended gaze origin and target locations respectively. For saccade detection we use the method presented by Engbert and Mergenthaler [2006]. Note that this saccade detection algorithm is not just limited to microsaccade detection as the title might suggest, and has been used by Lisi et al. [2019] for normal saccade detection.

5.4 Integration Rate Learning

To train the rate parameter, $\nu$ in Section 4.4, we train a Radial Basis Function Neural Network (as formulated in Equation (8)) implemented using the auto-differentiation library, Pytorch. We use the Adam optimizer with a learning rate of 0.1. The training used 2000 epochs and took ~ 180 seconds on a single NVIDIA RTX 3090 GPU.

6 EVALUATION

We gauge the ability of our model to generalize and consider possible applications via a series of psychophysical and simulated experiments. In Section 6.1, we first measure robustness and generalizability using data collected in Section 3. Next, in Section 6.2 we provide three example tasks featuring complex stimuli (athletics, esports, and photographic scenes), and compare the model’s predictions with our user study data. In Section 6.3, we evaluate how our model extends to more sophisticated tasks containing multiple stimuli of interest. Finally, we present two possible applications of our model: evaluating esports competition fairness, and estimating how human-display configuration can alter in-game target searching performance in Section 7.

6.1 Model Performance and Generalizability

We present an analytical model whose parameters are learned based on user-collected data. Thus, we need to evaluate our model’s performance regarding both prediction accuracy and generalizability beyond the specific trials and subjects included in the training data. To this end, we perform a comparative analysis of alternate training and testing partitions of the dataset.
We also visualize the baseline Q-Q plot for the model’s performance on \( D \) way as in Sections 4.3 and 4.4, and then measure the accuracy of the K.S. Test, and partition, across both partition protocols. The K.S. test fails to reject the null hypothesis supports the accuracy and generalizability of sampled data is drawn from the same distribution; failing to reject rejection on the K.S. test indicates a goodness-of-fit test between the reserved test data and our predic-
tion \([\text{Massey Jr 1951}]\), and show the Quantile-Quantile (Q-Q) plot. In Figure 6 indicate an overestimation of saccade latencies and vice versa. Data below the figure to contextualize the volume of data being shown. As defined in Section 6.1, (a) shows the results with random partition (1), (b) shows the results with subject_01’s data partition (2). We observe a distribution agreement between our model prediction and the unseen testing dataset; the closer the Q-Q curves are to the diagonal line, the more accurate the predictions are. The corresponding K.S. tests evidence the observation.

**Protocol.** For each analysis, we reserve a different partition of the dataset from Section 3 for testing, re-train the model in the same way as in Sections 4.3 and 4.4, and then measure the accuracy of the re-trained model prediction on the reserved test data. We perform two types of partitioning protocols for reserving the test set:

1. Random: a random selection drawn from all data points (20%),
2. Subject: all data from each individual subject (20%).

**Metrics and results.** We perform the Kolmogorov–Smirnov (K.S.) goodness-of-fit test between the reserved test data and our prediction \([\text{Massey Jr 1951}]\), and show the Quantile-Quantile (Q-Q) plot \([\text{Gnanadesikan and Wilk 1968}]\) in Figure 6. A significant difference on the K.S. test indicates a rejection of the null hypothesis that the sampled data is drawn from the same distribution; failing to reject the null hypothesis supports the accuracy and generalizability of our model. The Q-Q plot visualizes the correspondence of two probability distributions at each quantile. Data below the \( y = x \) line in Figure 6 indicate an overestimation of saccade latencies and vice versa for data above the line.

Figure 6 shows the Q-Q plot for the training and testing sets across both partition protocols. The K.S. test fails to reject the null hypothesis that the observed user saccade latency distribution is drawn from our model-predicted distribution for (1) the random partition, \( D = .2, p = .99 \), and (2) the individual subject partitions:

| Subject ID | S1    | S2    | S3    | S4    | S5    |
|------------|-------|-------|-------|-------|-------|
| K.S. analysis | \( D = .3 \) | \( D = .2 \) | \( D = .2 \) | \( D = .2 \) | \( D = .1 \) |
|             | \( p = .79 \) | \( p = .99 \) | \( p = .99 \) | \( p = .99 \) | \( p = 1.0 \) |

where \( D \) is the distance metric between two CDFs according to the K.S. Test, and \( p \) is the \( p \)-value corresponding to the distance metric.

**Discussion.** The above analysis demonstrates that our model does not predict statistically different distributions compared to unseen observations across various partitioning protocols. The results of the randomly partitioned study (1) demonstrate the generalizability of our model across trials without observed overfitting. Analysis of the subject-partitioned study (2) verifies our model’s applicability to unseen users, and thus general human saccadic behaviors.

We further performed an ablation study with individual visual characteristic conditions \((c, f, e)\) provided in Appendix F. It compares each condition’s contribution to our model by training an ablated version of the model with various conditions of the dataset missing, and quantifying the size of the regression in the model’s predictive ability. The results show that across ablation conditions, the Mean Squared Error (MSE) between ground truth and our model’s prediction regresses by 50% on average.

### 6.2 Predicting Saccadic Behaviors with Altered Target Appearances in Natural Tasks

In Section 3, we observed that unnoticeably subtle visual changes may induce significantly varied reactive latencies, as was formulated and predicted by our model in Section 4. In this experiment, we evaluate our model’s application in several realistic target search scenarios such as esports, and real-world photographs.

Via a series of psychophysical studies, we seek to determine: (1) whether our model can extend to predicting saccadic reaction latencies with natural task/stimuli; (2) whether we can imperceptibly alter the appearance of objects while still introducing enough reactive latency to materially influence real-world task performance. We answer these queries in our experiment and compare our findings to the model predictions.

**Participants and setup.** We recruited 14 participants (ages 22 – 33, 3 female) for this series of 2AFC experiments. Two participants were excluded for inability to perform the tasks (self-reported difficulty perceiving peripheral stimuli and target identification accuracy greater than one standard deviation below the mean), resulting in 12 final participants. Two of the 12 participated in the preliminary study in Section 3. The study was conducted during a 10-minute sessions consisting of 153 trials per scene for each participant. The hardware and setup remain the same as in Section 3.

**Scenes and stimuli.** To simulate a broad range of applications, our user study stimuli consisted of three groups of images: (1) a synthetic soccer scene, (2) a synthetic first-person view as an analog for esports, and (3) digital photographs of an indoor shelf. Each group contained 51 different images; each has the target stimuli appearing at different locations (to avoid learning effects) on the visual field, and serve as a trial. The background and targets from each evaluation group are shown in Figure 7. Although shown in color in the paper for visual clarity, all images were rendered with grayscale on display to avoid bias from color cues.

**Tasks.** Participants were instructed to complete a similar 2AFC decision task across all trial. At the beginning of each trial, they were shown a background image containing several task-irrelevant objects. After a randomized 1 – 1.5 second delay, an additional task-relevant stimulus, either a target or non-target, appeared on the
Subjects. Across each image set, we tested three variations of the target stimulus in order to measure how changes in image features affect saccade latencies. In one variation the target stimulus had increased contrast and/or decreased frequency (Accelerated), in another variation the target had decreased contrast and/or increased frequency (Deferred), and a third unfiltered variation was used as a control group (Control). Each participant performed 51 images × 3 conditions × 3 scenes, resulting in 459 trials total, i.e., 5508 trials across the experiment. Measuring the precise frequencies affecting saccade latency is a complex task requiring pooling from multiband. Investigating a comprehensive pooling strategy is beyond the scope of this work. Therefore, we approximate the representative frequency as the Laplacian pyramid layer with the highest corresponding contrast, for those images without a uniform frequency pattern. Contrast and eccentricity computations were trivial to compute without requiring pooling operations.

Results. We present the results of our experiments in Figure 7. We again use the K.S. statistical test to evaluate alignment between predicted and measured histograms across the different scenes for each condition. We report the results of these tests below:

Fig. 7. User data and our model prediction of saccade latencies among varied target appearances. Saccade latency modulation correlates with the contrast of stimuli as shown in the three distinct scenes (and target candidates) shown in (a)/(b)/(c). Each scene presents distinct visual characteristics including low polygon 3D scenes, dense geometries, or natural scenes. (d) illustrates the study procedure over time. With the Control condition as reference, all others show FovVideoVDP scores above 9.5, indicating identical perceptual appearance per [Mantiuk et al. 2021]. Using the shooter scene as example, (e) shows the user latency data in histograms, and our model predicted latency in curves. A significant agreement can be observed. Please refer to our supplementary videos for an animated visualization. (f) shows the mean relative durations (with Control as 0% pedestal) of Deferred/Accelerated. The error bars indicate standard error. Full statistical analysis on all scenes can be seen in Section 6.2. Each individual’s raw probabilistic distributions are provided in Appendix G. 3D asset credits to haykel-shaba (a), and Slavyer (b) at Sketchfab Inc.
Please refer to Appendix G for the collected saccadic latency distributions of individual participants and scenes.

Using the Control images as reference, we additionally calculate the FovVideoVDP values for all images in our dataset. We find the mean values to be above 9.5 for all Accelerated/Deferred images, which indicates observers would be approximately at chance for detecting differences between them.

We also debriefed each participant after the experiment on their thoughts regarding the tasks, and most participants reported no self-awareness of reaction time difference.

Discussion. Our results demonstrate agreement between the predictions made by our model and the observed saccadic latency distributions across 12 participants. We find significant differences in saccadic latency across conditions, despite identical perceptual appearance evidenced by the FovVideoVDP metrics.

Our prediction of the photographic scene results show correct trends and distribution ratios, albeit for a scaled absolute time (in ms) relative to the measured data. We attribute this scale variance to the fact that natural images contain wide frequency bands and our single-frequency pooling in the Laplacian Pyramid may discard significant frequency information. This motivates interesting future work on multi-frequency pooling models tailored for reaction time, see Section 8.

6.3 Extending to Foveal-Peripheral Dual Tasks

In various real-world scenarios, humans perform tasks by jointly analyzing both foveal and peripheral content, such as with reading, film watching, and architectural design. In this experiment, we extend and evaluate our model to such applications considering dual tasks.

Modeling. Our visual system processes foveal and peripheral stimuli independently and in parallel for a variety of tasks [Ludwig et al. 2014]. That is, the foveal and peripheral pathways gather information concurrently, and the decision to trigger a saccade waits until both processes have finished. We hypothesize that these independent foveal and peripheral stimulus processing units operate using the integration-and-action process as described in Section 4.1.

In this model, processing times for both the fovea, $T_f$, and the peripheral, $T_p$, follow Equation (4), and can be adapted to specific task descriptions and stimulus characteristics as shown in Equation (6):

$$T_f \sim \mathcal{IG}(\alpha_f, \nu_f)$$
$$T_p \sim \mathcal{IG}(\alpha_p, \nu_p),$$

where we create some shorthands for convenience:

$$\alpha_f = \alpha(D_f), \nu_f = \nu(c_f, f_f, \alpha_f = 0^\circ)$$
$$\nu_p = \nu(c_p, f_p, \alpha_p = 10^\circ).$$

We hypothesize that the observed user saccadic latencies are drawn from our $T_{\text{dual}}$-predicted distribution ($D = 0.1$ and $p = 1.0$).

Models which consider only the peripheral contrast (shown in Figure 8b), or only the foveal contrast fail to accurately predict the saccade latencies. We run the K.S. test for both of these conditions and present the Q-Q plot comparing the data to our model predictions in Figure 8d. The K.S. statistical test again fails to reject the null hypothesis that the observed user saccadic latencies are drawn from our $T_{\text{dual}}$-predicted distribution ($D = 0.1$ and $p = 1.0$).

Setup. To evaluate our hypothetical model for dual tasks, we conducted a user study to measure how saccade latency changes as we modulate foveal and peripheral stimuli independently. Unfortunately, it is not possible to explicitly compute the $\alpha_f$ and $\alpha_p$ values as in Equation (7), because a user study for the dual task can only sample the total saccade latency from Equation (12). That is, the individual distributions, $T_f$ and $T_p$ are not measured directly. Since finding these threshold values directly is not possible, we infer them via maximum-likelihood estimation (MLE) of the overall distribution of $T_{\text{dual}}$, given a dataset of size $n$:

$$\alpha_f, \alpha_p = \arg \max \sum_i \log L(\alpha_f, \alpha_p; i^{(i)}, v_f, v_p).$$

Please refer to Appendix C for the derivation of the likelihood function for $T_{\text{dual}}$. The hardware setup in this experiment is the same as described in Table 1.

Participants. We recruited $n = 12$ participants (ages 22-33, 3 female) with normal or corrected to normal vision for a series of 2AFC experiments. The study was conducted during a single 10 minute session, including a total of 240 trials for each participant.

Stimuli and Tasks. At the beginning of each trial participants are shown three Gabor patches as illustrated in Figure 8a: one at the fovea, and two in the left and right peripheries at equal eccentricities of $10^\circ$. The foveal Gabor is tilted either $45^\circ$ or $-45^\circ$ from the vertical axis; with chance probability, one of the peripheral Gabors is selected to have the same tilt as the foveal Gabor, while the other has the opposite tilt. The task is to identify and saccade to the peripheral Gabor of the same orientation as the foveal Gabor. For each trial, the central and peripheral Gabor contrast values are sampled from $[0.05, 0.22, 0.53, 1.0]$, drawn independently. That is, taking all combinations of central-peripheral Gabor contrast possibilities yields a total of 16 conditions. The frequency of all Gabors was fixed to 2.0 cpd (cycles-per-degree). Each participant also performed 15 randomly ordered practice trials before the start of the experiment.

Results. In Figure 8c, we show the relationship of both foveal and peripheral contrasts with saccade latency, as well as the ground truth data collected from our user study overlaid on top of the surface plot. The MLE regression produces threshold values of $\alpha_f = 3.21$ and $\alpha_p = 3.56$. Hence, the threshold ratio between the foveal and peripheral components is $1 : 1.04$. Similar to Section 6.1, we present the Q-Q plot comparing the data to our model predictions in Figure 8d. The K.S. statistical test again fails to reject the null hypothesis that the observed user saccadic latencies are drawn from our $T_{\text{dual}}$-predicted distribution ($D = 0.1$ and $p = 1.0$).
A major application of our model is to measure and optimize human performance in competitive, real-time, or time-sensitive tasks such as defense, piloting, and esports. In this evaluation, we use esports as an example. In real-world professional gameplay, we deploy our model to: 1) measure game fairness in terms of character skin design that triggers varied gaze motion performance between two teams; 2) measure and optimize the human target search performance under various screen resolutions, eye-display distances, and compare the performance with traditional and immersive displays.

Data. We collected professional replay videos from a popular esports game, Counter-Strike: Global Offense via YouTube. The data contains a half hour long video footage where we uniformly sampled 95 frames from beginning to the end. For each frame, we exploit the virtual human tracking model YoLO [2016] that predicts the team ID (Counter-Terrorist, CT and Terrorist TR), and bounding boxes. We assume the observers gaze lies in the middle of the screen, and apply our model to predict the time when the viewer reorients their gaze to each target. We measure the visual characteristics with a common display setting: a Samsung 32inch CH32H711 monitor, 2K 16:9 resolution, 70cm width, 300cd/m² brightness, and \(1.33 \text{D} \) (dioptr = \( \text{m}^{-1} \)) eye-display distance (50° FoV).

7 APPLICATION CASE STUDY: ESPORTS FAIRNESS

7.1 Competition Fairness in Target Searching

The game has two opposing teams of characters. Regardless of game task design and differences in tools, game fairness is an important concern in esports [Chen et al. 2014]. Using our model and the detected targets, we measure the average saccade latency of individual groups.

Figure 9b shows the results. We observed a significant difference between CT and TR groups: the average normalized latencies are 0.92 ± 0.02 for searching CTs and 0.95 ± 0.04 for searching TRs, indicating a 3.3% difference. Given previous literature indicating the mean saccade latency for CS:GO professional players to be about 282ms [Velichkovsky et al. 2019], 3.3% results in a 9.3ms reaction variance. One-way repeated measures ANOVA showed the group’s significant main effect on the saccadic latency, \(F_{1,93} = 11.4, p = .001\).

The results demonstrate a statistically significant difference between the two groups, in terms of them perceiving, processing, and reacting to appeared targets. That is, a TR/CT saccading to the other group is significantly faster/slower with no less than 2/1 frames on a 60/120FPS displays. The speed difference is remarkably higher than the minimum latency, as low as 4ms, that leads to allo-

7.2 Optimizing Player Performance

A natural and extensively asked question is the role of eye-display distance (e.g., regular monitors vs. VR displays) and screen resolution in professional competitions. Using our model, we measure the

predictions: \(D = 0.9\) and \(p = 0.002\) for the foveal-only model, and \(D = 0.8\) and \(p = 0.002\) for the periphery-only model.

Discussion. When humans perform tasks involving both foveal and peripheral analysis, we observe that a model considering only one eccentricity fails to predict saccade latencies, as illustrated in Figure 8d and demonstrated by the K.S. tests. By comparison, the joint model we propose in Equation (12), inspired by prior discoveries on visual mechanisms, successfully predicts the latency distribution.

7 APPLICATION CASE STUDY: ESPORTS FAIRNESS

METRIC AND PERFORMANCE OPTIMIZATION

A major application of our model is to measure and optimize human performance in competitive, real-time, or time-sensitive tasks such as defense, piloting, and esports. In this evaluation, we use esports as an example. In real-world professional gameplay, we deploy our model to: 1) measure game fairness in terms of character skin design that triggers varied gaze motion performance between two teams; 2) measure and optimize the human target search performance under various screen resolutions, eye-display distances, and compare the performance with traditional and immersive displays.

Data. We collected professional replay videos from a popular esports game, Counter-Strike: Global Offense via YouTube. The data contains a half hour long video footage where we uniformly sampled 95 frames from beginning to the end. For each frame, we exploit the virtual human tracking model YoLO [2016] that predicts the team ID (Counter-Terrorist, CT and Terrorist TR), and bounding boxes. We assume the observers gaze lies in the middle of the screen, and apply our model to predict the time when the viewer reorients their gaze to each target. We measure the visual characteristics with a common display setting: a Samsung 32inch CH32H711 monitor, 2K 16:9 resolution, 70cm width, 300cd/m² brightness, and \(1.33 \text{D} \) (dioptr = \( \text{m}^{-1} \)) eye-display distance (50° FoV).

7.1 Competition Fairness in Target Searching

The game has two opposing teams of characters. Regardless of game task design and differences in tools, game fairness is an important concern in esports [Chen et al. 2014]. Using our model and the detected targets, we measure the average saccade latency of individual groups.

Figure 9b shows the results. We observed a significant difference between CT and TR groups: the average normalized latencies are 0.92 ± 0.02 for searching CTs and 0.95 ± 0.04 for searching TRs, indicating a 3.3% difference. Given previous literature indicating the mean saccade latency for CS:GO professional players to be about 282ms [Velichkovsky et al. 2019], 3.3% results in a 9.3ms reaction variance. One-way repeated measures ANOVA showed the group’s significant main effect on the saccadic latency, \(F_{1,93} = 11.4, p = .001\).

The results demonstrate a statistically significant difference between the two groups, in terms of them perceiving, processing, and reacting to appeared targets. That is, a TR/CT saccading to the other group is significantly faster/slower with no less than 2/1 frames on a 60/120FPS displays. The speed difference is remarkably higher than the minimum latency, as low as 4ms, that leads to allo-

7.2 Optimizing Player Performance

A natural and extensively asked question is the role of eye-display distance (e.g., regular monitors vs. VR displays) and screen resolution in professional competitions. Using our model, we measure the

predictions: \(D = 0.9\) and \(p = 0.002\) for the foveal-only model, and \(D = 0.8\) and \(p = 0.002\) for the periphery-only model.

Discussion. When humans perform tasks involving both foveal and peripheral analysis, we observe that a model considering only one eccentricity fails to predict saccade latencies, as illustrated in Figure 8d and demonstrated by the K.S. tests. By comparison, the joint model we propose in Equation (12), inspired by prior discoveries on visual mechanisms, successfully predicts the latency distribution.

7 APPLICATION CASE STUDY: ESPORTS FAIRNESS

METRIC AND PERFORMANCE OPTIMIZATION

A major application of our model is to measure and optimize human performance in competitive, real-time, or time-sensitive tasks such as defense, piloting, and esports. In this evaluation, we use esports as an example. In real-world professional gameplay, we deploy our model to: 1) measure game fairness in terms of character skin design that triggers varied gaze motion performance between two teams; 2) measure and optimize the human target search performance under various screen resolutions, eye-display distances, and compare the performance with traditional and immersive displays.

Data. We collected professional replay videos from a popular esports game, Counter-Strike: Global Offense via YouTube. The data contains a half hour long video footage where we uniformly sampled 95 frames from beginning to the end. For each frame, we exploit the virtual human tracking model YoLO [2016] that predicts the team ID (Counter-Terrorist, CT and Terrorist TR), and bounding boxes. We assume the observers gaze lies in the middle of the screen, and apply our model to predict the time when the viewer reorients their gaze to each target. We measure the visual characteristics with a common display setting: a Samsung 32inch CH32H711 monitor, 2K 16:9 resolution, 70cm width, 300cd/m² brightness, and \(1.33 \text{D} \) (dioptr = \( \text{m}^{-1} \)) eye-display distance (50° FoV).

7.1 Competition Fairness in Target Searching

The game has two opposing teams of characters. Regardless of game task design and differences in tools, game fairness is an important concern in esports [Chen et al. 2014]. Using our model and the detected targets, we measure the average saccade latency of individual groups.

Figure 9b shows the results. We observed a significant difference between CT and TR groups: the average normalized latencies are 0.92 ± 0.02 for searching CTs and 0.95 ± 0.04 for searching TRs, indicating a 3.3% difference. Given previous literature indicating the mean saccade latency for CS:GO professional players to be about 282ms [Velichkovsky et al. 2019], 3.3% results in a 9.3ms reaction variance. One-way repeated measures ANOVA showed the group’s significant main effect on the saccadic latency, \(F_{1,93} = 11.4, p = .001\).

The results demonstrate a statistically significant difference between the two groups, in terms of them perceiving, processing, and reacting to appeared targets. That is, a TR/CT saccading to the other group is significantly faster/slower with no less than 2/1 frames on a 60/120FPS displays. The speed difference is remarkably higher than the minimum latency, as low as 4ms, that leads to altered esport performance among top-level competitors [Kim et al. 2019]. While this may have been one of the factors that contributed to the imbalanced competitive game performance (higher winning rate of TR on the map we analyzed) between these two groups \(^1\), in practice the asymmetric weapon and task designs might also have played a role.

7.2 Optimizing Player Performance

A natural and extensively asked question is the role of eye-display distance (e.g., regular monitors vs. VR displays) and screen resolution in professional competitions. Using our model, we measure the

1https://www.hltv.org/stats/teams/map/31/5995/g2
statistical saccade latency as a function of displays with the same dataset as Section 7.

Figure 9c visualizes our results by observing the altered reaction performance. As before, we use the mean saccade latency for CS:GO professional players to be about 282ms [Velichkovsky et al. 2019]. First, both teams are not at their best performance with the initial 1.33D eye-display distance, with the faster reaction of TRs searching CTs. However, the teams reveal different trends by changing the eye’s distance (thus FoV), which jointly alters target eccentricity (e) and frequency (f) (cf. Appendix D). Particularly, the minimal saccade latency towards CT targets is 273.5ms at 34.6° FoV (0.9D eye-display distance). In comparison, the minimal latency towards TR targets is 266.1ms at 61.5° FoV (1.6D eye-display distance). The two curves intersect at 1.69D with an identical latency of 266.3ms. We further simulate real-world use cases with different displays (for instance, gaming with mobile devices or training with VR displays). In this experiment, we use the measures from the iPhone 13 (5.78 × 2.53 inches) with the commonly suggested 30cm (or 3.3D) eye-display distance, leading to a 25.7° FoV. Under this circumstances, the saccade time to TR becomes higher than CT (324.3ms vs 259.4ms) Similarly, the measurement with our virtual reality HMD (90° overlapped FoV), the relative trend is swapped: saccading to a TR becomes shorter than to a CT, with a 21.2ms difference (276.0ms vs 288.0ms).

The above analysis indicates the sensitivity of eye display correlation in determining performance and fairness in time-sensitive and competitive scenarios. Surprisingly, the statistical performance bias may swap with different eye-display relationships. For instance, with a mobile/VR setting, the visual stimuli may bias with TR/CT players in terms of reaction performance. In addition to the commonly referred measurement of visual similarity and task/map fairness, our model presents a novel perspective in competitive and highly dynamic scenario design, such as athletics, esports and defense.

8 LIMITATIONS AND FUTURE WORK
In our model, we consider a variety of factors which affect visual stimuli – contrast, frequency, eccentricity – in the context of several display environments (from mobile devices to VR displays). However, a multitude of other factors may affect reaction time to observed natural stimuli in the complex natural world. For instance, making a cognitive decision among multiple potential targets may depend on higher-level visual salience [Jarvenpa 1990] and object sizes [Lisi et al. 2019] whereas our model assumes pre-knowledge of the intended saccade target. While our work showed promising results for one foveal and one peripheral task, we cannot easily extrapolate our results for significantly more complex tasks. In addition, we rely on a single max-pooling of image frequencies using the highest contrast layer of the Laplacian pyramid (Section 6.2) for model application in natural images. This may introduce an approximation error, especially with complex and noisy natural images (Figure 7c). Using an approach similar to Mantiuk et al. [2021] and fitting a pooling model to user data collected from natural image stimuli would be a useful direction of future work. To avoid any bias introduced by color, we convert stimuli to gray-scale for the psychophysical experiments. We leave the task of creating a multi-band and color-aware model tailored for measuring reactive latency for future work. Our model also does not consider scenarios where motion [Jindal et al. 2021] and refresh rates [Krajancich et al. 2021] may play a critical role, especially during saccades [Schweitzer and Rolfs 2021]. We aim to extend our model to consider spatio-temporal effects and complex dynamic scenarios in the future.

9 CONCLUSION
We demonstrated a significant difference between human visual acuity (observation of a stimulus) and reaction latency (processing after the observation). We formulated this behavior using a neurologically-inspired probabilistic model that is motivated and evaluated by a series of psychophysical studies. The surprising gap we observe between observation and reaction raised new, previously unasked questions such as, “Are competitive digital activities such as esports fair among teams, given the design of their appearances?”, “How do human-display-content relationships alter our performance in virtual environments?”, and “What settings optimize reaction time without compromising visual content?”. We hope that our model’s answers to these questions inspire researchers to explore new avenues in interactive computer graphics and immersive virtual environments.

ACKNOWLEDGMENTS
The authors would like to thank Chris Wyman for his valuable suggestions. The research is partially supported by the NVIDIA Applied Research Accelerator Program and the DARPA PTC program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA.

REFERENCES
Rachel Albert, Anjul Patney, David Luebke, and Joohwan Kim. 2017. Latency Requirements for Foveated Rendering in Virtual Reality. ACM Transactions on Applied Perception 14, 4, Article 25 (Sep 2017), 13 pages. https://doi.org/10.1145/3127589
Elena Arabadzhiiyska, Okan Tarhan Tursun, Karol Myszkowski, Hans-Peter Seidel, and Piotr Didyk. 2017. Saccade Landing Position Prediction for Gaze-Contingent Rendering. ACM Trans. Graph. 36, 4, Article 50 (July 2017), 12 pages. https://doi.org/10.1145/3072959.3073642
A.Terry Bahill, Michael R. Clark, and Lawrence Stark. 1975. The main sequence, a tool for studying human eye movements. Mathematical Biosciences 24, 3 (1975), 191–204. https://doi.org/10.1016/0025-5564(75)90075-9
A Terry Bahill. 1975. Most naturally occurring human saccades have magnitudes of 15 deg or less. Invest. Ophthal&mol 14 (1975), 468–469
Reynold Bailey, Ann McNamara, Nisha Sudarananam, and Cindy Grimm. 2009. Subtle Gaze Direction. ACM Trans. Graph. 28, 4, Article 100 (Sept. 2009), 14 pages. https://doi.org/10.1145/1559755.1559757
Peter CJ Barton. 1999. Contrast sensitivity of the human eye and its effects on image quality. SPIE press.
W. Becker and A.F. Fuchs. 1969. Further properties of the human saccadic system: Eye movements and correction saccades with and without visual fixation points. Vision Research 9, 10 (1969), 1247–1258. https://doi.org/10.1016/0042-6989(69)90112-6
AH Bell, MA Meredith, AJ Van Opstal, and DougP Munoz. 2006. Stimulus intensity modifies saccadic reaction time and visual response latency in the superior colliculus. Experimental Brain Research 174, 1 (2006), 53–59.
DC Burr, MC Morrone, and J Ross. 1994. Selective suppression of the magnocellular visual pathway during saccadic eye movements. Nature 371, 6497 (1994), 511–513. https://doi.org/10.1038/37151A0
Anke Cajar, Ralf Engbert, and Jochen Laubrock. 2016. Spatial frequency processing in the central and peripheral visual field during scene viewing. Vision Research 127 (2016), 186–197.
RHS Carpenter. 2004. Contrast, probability, and saccadic latency: evidence for independence of detection and decision. Current Biology 14, 17 (2004), 1576–1580.
Fig. 9. Results of esports video dataset analysis (a) illustrates our simulated eye-display spatial relationship and our CS-GO gameplay dataset (including the automated labeling of the teams). Note that changing the eye-display distances results in varied FoVs, thus changing the perceived visual characteristics (eccentricity and frequency). (b) shows our model’s approximation of the team-wise target searching performance. The X-axis indicates the team splits. The Y-axis shows the predicted mean latencies with the semi-transparent error bar as standard error. The point where the two group’s mean latencies intersect is marked by the green circle. The lowest latencies of saccading for each team and the simulated FoVs of non-desktop display environments are dash-labeled.

Linus Franke, Laura Fink, Jana Martschinké, Kai Selgrad, and Marc Stamminger. 2021. Time-Warped Foveated Rendering for Virtual Reality Headsets. Computer Graphics Forum 40, 1 (2021), 110–123. https://doi.org/10.1111/cgf.14176
Drew Fudenberg, Whitney Newey, Philipp Strack, and Tomasz Strzelecki. 2020. Testing the drift-diffusion model. Proceedings of the National Academy of Sciences 117, 52 (2020), 33141–33148
Ramanathan Ganeshan and Martin B Willk. 1968. Probability plotting methods for the analysis of data. Biometrika 55, 1 (1968), 1–17.
Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D Graphics. ACM Transactions on Graphics 31, 6, Article 164 (nov 2012), 50 pages. https://doi.org/10.1145/2366145.2366183
E. Hartmann, B. Lachenmayr, and H. Brettel. 1979. The peripheral critical flicker frequency. Vision Research 19, 9 (1979), 1019–1023. https://doi.org/10.1016/0042-6989(79)90227-X
Toyohiko Hatada, Haruo Sakata, and Hideo Kunaka. 1980. Psychophysical analysis of the "sensation of reality" induced by a visual wide-field display. Smpte Journal 89, 8 (1980), 560–569
Sebastien Hillaire, Anatole Lecuyer, Remi Cozot, and Gery Casiez. 2008. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In 2008 IEEE Virtual Reality Conference. 47–50. https://doi.org/10.1109/VR.2008.480749
Alain Hore and Djemel Ziou. 2010. Image quality metrics: PSNR vs. SSIM. In 2010 20th international conference on pattern recognition. IEEE, 2366–2369.
Michael R. Ribtston and Shaun L. Cloherty. 2009. Visual Perception: Saccadic Omission—Suppression or Temporal Masking? Current Biology 19, 12 (2009), R493–R496. https://doi.org/10.1016/j.cub.2009.05.010
Sirrika J. Jarvenpaa. 1990. Graphic displays in decision making—the visual salience effect. Journal of Behavioral Decision Making 3, 4 (1990), 247–262.
Akhay Jindal, Krzysztof Wolski, Karol Myszkowski, and Rafal K. Mantiuk. 2021. Perceptual model for adaptive local shading and rendering rate. ACM Transactions on Graphics (TOG) 40, 6 (2021), 1–18.
RP Kalesnykas and PE Hallett. 1994. Retinal eccentricity and the latency of eye saccades. Vision research 34, 4 (1994), 517–531.
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.
Donald H. Kelley. 1979. Motion and vision. II. Stabilized spatio-temporal threshold surface. Jour 69, 10 (1979), 1349–1349.
Joowhan Kim, Josef Spjut, Morgan McGuire, Alexander Majercik, Ben Boudaoud, Rachel Albert, and David Luebke. 2019. Esports arm race: latency and refresh rate for competitive gaming tasks. Journal of Vision 19, 10 (2019), 218e–218e.
Robert Konrad, Anastasios Angelopoulos, and Gordon Wetzstein. 2020. Gaze-Contingent Ocular Parallax Rendering for Virtual Reality. ACM Trans. Graph. 39 (2020). Issue 2.
