2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR)

2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR) | 979-8-3315-2157-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/AIXVR63409.2025.00046

Toward Computationally Simulating the Effect of Gaze Guidance on Interactive Event Segmentation within Immersive Virtual Environments

Monthir Ali
School of Computing
University of Utah
Salt Lake City, USA
monthir.ali@utah.edu

Rogelio E. Cardona-Rivera
Division of Games
University of Utah
Salt Lake City, USA
r.cardona.rivera@utah.edu

***Abstract*—In this paper, we present an approach to computationally simulate the effect on a person's event segmentation that could hypothetically be induced via gaze guidance techniques within an immersive virtual environment. We propose a methodology to model gaze guidance via rational modifications to the Fluid Events Model of segmentation. We aim to provide a robust basis for discerning the relationship between immersive environment gaze guidance and its ultimate effect on a user's experience.**

***Index Terms*—immersive environment, psychology, attention, event cognition, computer-based experiment.**

I. INTRODUCTION

We are interested in discovering how guiding a person's attention through the use of *cues* within immersive environments impacts their event cognition in a computationally-rendered virtual world. There are *many* meaningfully different kinds of cues that have been employed within immersive virtual worlds [1]–[4]. This wide variety of cue types remains under-explored in terms of their impact on event cognition.

We sought to fill this gap by exploring *how possibly* [5] to account for the effect of gaze guidance on event cognition within interactive immersive environments.

This paper presents our method of constructing a simulation designed to support the following computerized thought experiment [6]: how might we adapt the Fluid Events Model (FEM) [7] – a validated behavioral model of interactive event segmentation – to accommodate a hypothetical influence on segmentation due to *spatial cues* [8]?

II. RELATED WORK AND BACKGROUND

This paper derives a method to modify the FEM through a principled extension that covers the hypothetical cue-based impact on *event segmentation*—the automatic component of our perceptual processing that breaks down daily occurrences and actions into a discrete set of meaningful events [9]. An *event* is a segment of time at a given location that is conceived by an individual to have a start and an end [10].

The FEM is illustrated in Figure 1. This model predicts the probability of people perceiving new events and changing their ongoing actions (activities carried out by individuals to complete a task) due to two types of factors: *event-structure factors* and *experience-based factors*. Event-structure factors (denoted with a hammer icon) include recent changes in event structure, suitability of the current action to the event, and time spent on task. Experience-based factors (denoted with a person icon) involve recent action shifts, frequency of action shifts, performance dips, and a person's propensity to switch actions within the current task (flexibility).

Our work lays the groundwork for a modified FEM that we expect will capture a new event-based factor: one that reflects our hypotheses for how spatial cues influence the baseline FEM's prediction of event segmentation likelihood. In the subsections below, we discuss the concepts and formalisms that are needed to understand the groundwork we lay out.

A. The CueInfluence Factor

The FEM calculates the probability perceiving a new event. We propose modifying the FEM's underlying formula to include a new event-structure factor: *CueInfluence*. This factor ought to reflect how cues operate within our field of view, as described next.

B. Cues within the Field of View

Covert vs. overt in the context of spatial attention cueing refers to how attention is drawn to a point in space. Researchers have looked into the differences between foveal vision and peripheral vision—the two main regions that form the humans eye's field of vision. The transition between these regions does not occur at hard boundaries but is rather seamless. This has led researchers to have different definitions of where these regions lie within the vision field, thereby introducing vagueness in terminology.

For the purpose of defining the central visual field in our work, we adopted the definition by Strasburger et al. [11]. They refer to the central visual field region as that of the fovea and parafovea ($\varphi \approx 8$ degree of arc diameter). This area represents the region where overt cues operate (illustrated in Fig. 2). Anything outside of that is peripheral vision.

2771-7453/25/$31.00 ©2025 IEEE
DOI 10.1109/AIxVR63409.2025.00046

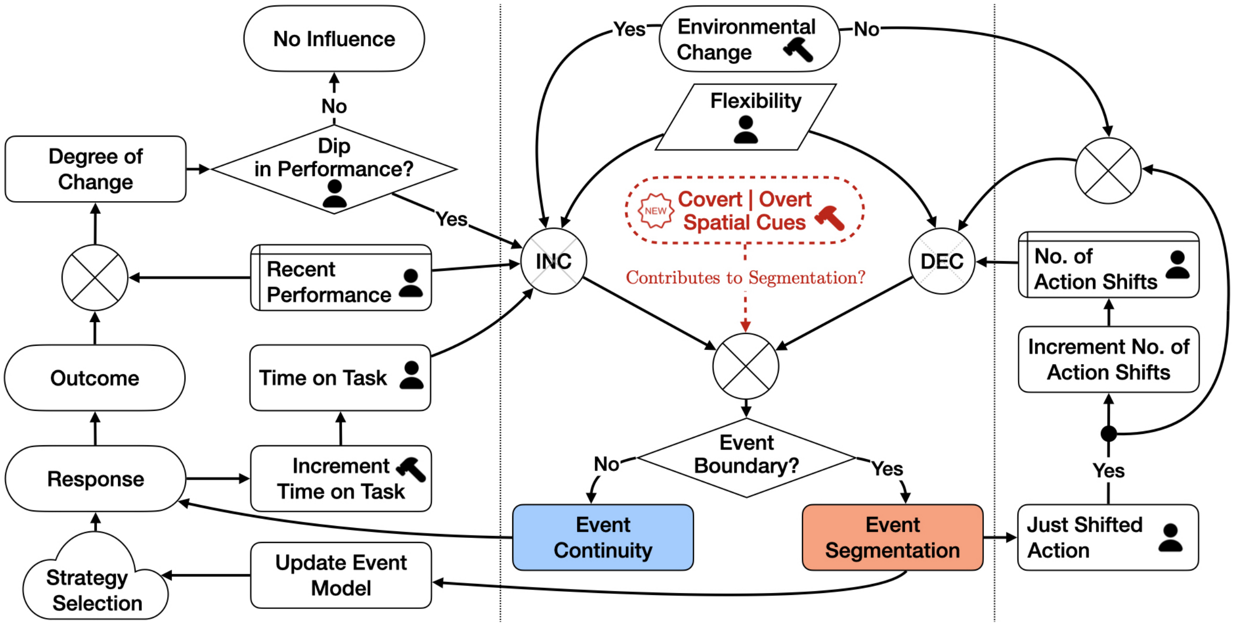

Fig. 1. A recreation of the Fluid Events Model Flowchart (ISO 5807:1985) with the addition of our proposed Spatial Cues component (highlighted in red). The two event-structure factors involved in the FEM are: 1) Environmental Change; alterations to the current task's environment or the event structure, and 2) Time on Task; time spent on current task. The five experience-based factors in the FEM are: 1) Just Shifted Action; whether a recent action switch has occurred, 2) Number of Action Shifts; frequency of action switching, 3) Performance Dip; recently experienced decline in performance, 4) Bad Shift; whether a performance dip occurred due to a recent action switch, and 5) Flexibility (Sensitivity); individual's inclination to try new actions. Our work focuses on event segmentation which is relevant to the event-structure factors. We posit this component factors directly into the probability of perceiving an event boundary.

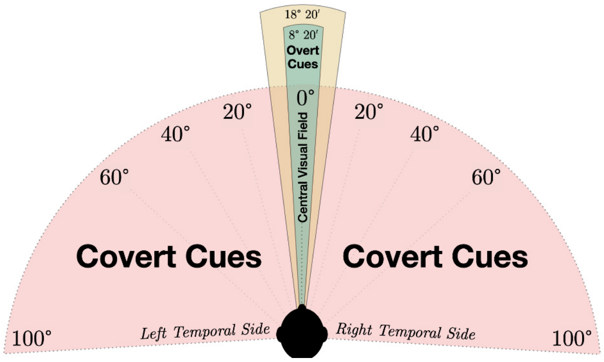

Fig. 2. The visual field has two principal sub-regions: (1) the central visual field, which we identify as the region that subtends $18°20'$ of visual angle; and (2) the peripheral visual field, which extends approximately $80°$ in both Temporal Left and Temporal Right directions.

Our eventual work will simulate the two aforementioned types of cues. For overt cues, we propose simulating a cue that conspicuously draws attention so that individuals are aware of the stimulus. Such techniques include highlighting and clear overlay elements (e.g., arrows pointing at targets) as used in previous studies [12], [13]. For covert cues, we propose simulating Bailey et al.'s Subtle Gaze Direction (SGD) technique [14]. SGD uses luminance (lightness) or color modulation to increase the salience of a particular region and *subtly* guide the gaze to it. This covert cue stimulus is terminated immediately whenever the angle between the center of the user's gaze and the cued region is within $10°$, preventing the gaze from fixating on the stimulus (illustrated in Fig. 3).

III. METHOD

As noted, the FEM predicts the likelihood of perceiving new events due to combined event-structure and experience-based factors. Our proposal is to modify the FEM by introducing a *CueInfluence* (event-structure) factor, which computes the expected influence of cues on the FEM's prediction score.

Depending on which type of cue is used (overt or covert), the formula contributes a different value to modify the baseline prediction score. We expect the cue influence equation to be comprised of the subtraction of a power function – usually used to describe decay and growth phenomena in real world [15] – from the value 1; this would reflect the growth of influence as the gaze proximity increases. In turn, this would produce a normalized value between 0.0 and 1.0, which fits the FEM's required prediction score criterion range. One candidate equation that satisfies these constraints is:

$$\text{CueInfluence} = 1 - \alpha^{\text{GazeProximity}} \tag{1}$$

Where:

- α is a weighting factor that modifies the influence of the cueing technique on event segmentation likelihood, and
- GazeProximity reflects the proximity of the gaze to the cue stimulus as illustrated in Fig. 4.

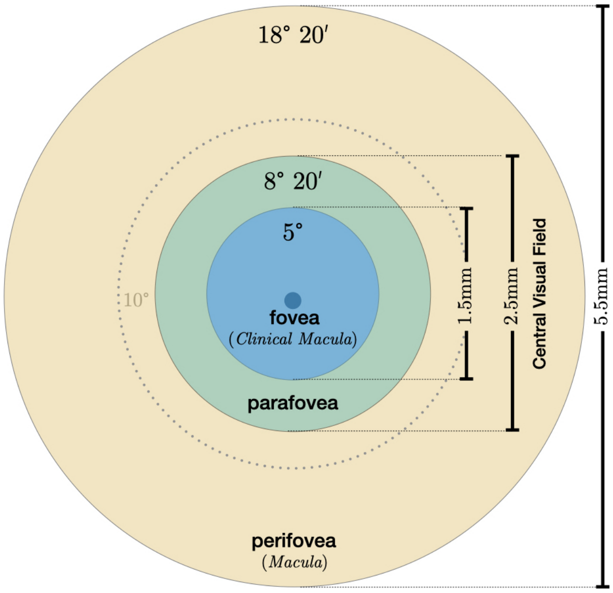

Fig. 3. Schematic diagram of the macula. The central visual field extends to the parafovea ($\varphi \approx 8$ visual angle). If the the center of gaze comes within 10° of a source of a covert cue stimulus, the stimulus is terminated immediately.

We posit that the cue influence equation should account for an expected dominance relationship between cue types as suggested by the literature [2]–[4], [13]. This ensures a dynamic adjustment of the influence based on the interplay between the cue type and the gaze proximity. While one might think that gaze proximity and cue type are dependent on each other (high proximity means overt and low proximity means covert) there is potentially a different way to interpret their relationship. We envision two possible configurations that could explain how these might be related differently.

The first configuration assumes no cue dominance and that both types of cues affect the gaze proximity equally, which is a reasonable baseline assumption. The second configuration assumes the presence of a *dominant* cue, which overrides the gaze proximity factor. The justification behind this second configuration lies in prior literature: several studies suggest that the efficacy of guiding gaze using cues is different depending on the type of cue that is being used [3], [4], [13]. This means that for a certain type of cue, *where* a person's gaze is fixated is of less importance relative to the cue's presence within your field of view; this is because that cue can be assumed to be reliably perceived and thereby succeed in drawing your gaze to a region of interest.

In order to conduct our intended simulation study, we will have to synthesize the data we would expect to see under different expectations (i.e., configurations) of how the different cues independently impact the likelihood of drawing your attention, and transitively, how this elicits segmentation. This synthetic data ought to be generated as a function of the cue type, because – as explained – we expect gaze behavior may be different when using overt cues versus covert cues.

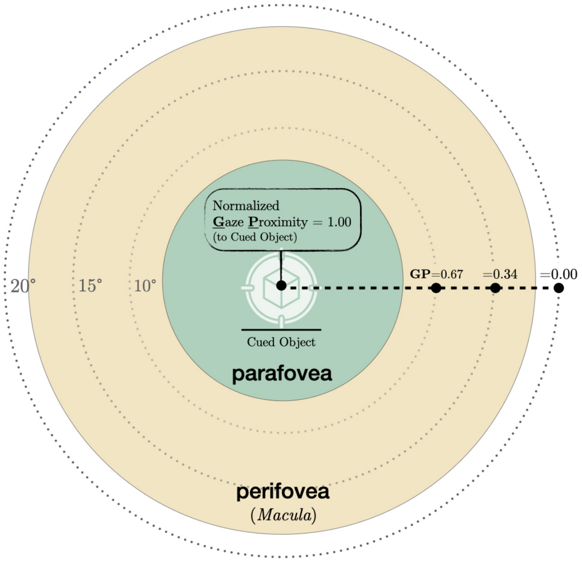

Fig. 4. The proximity of a simulated participant's gaze point to the cue stimulus. It has a maximum value of 1.0 when fixated on the cue and decreases to 0 as the angle between the center of gaze and the cue stimulus extends beyond 20°.

A. *Toward Synthesizing Overt Cue Data*

Prior literature has demonstrated the effectiveness of using overt cues to help users within immersive environments detect objects and changes in the scene [13], [16]–[18]. Hence, we anticipate that a majority of eye fixations will be reliably directed towards the overt cue stimulus. Thus, our synthetic data should reflect a distribution of eye gaze behaviors with the highest density centered on the cue stimulus. We expect this distribution to be left long-tailed.

B. *Toward Synthesizing Covert Cue Data*

Covert cues utilize the SGD technique and ensure that the stimulus is terminated before the angle between the center of gaze and the cue stimulus falls within 10° as previously described. Therefore, gaze fixations can either halt in that vicinity of where the stimulus is terminated (because it is no longer being guided), or proceed to fixate on the cued region that we intend the gaze to fixate on. As a result, our synthetic data should reflect a *bimodal* distribution with two comparable peaks at the following values: (1) GazeProximity $= 0.67$, indicating that the gaze halted in position after the termination of the covert cue (GazeProximity is equal to 0.67 when the angle between center of gaze and cue stimulus is exactly at $10°$ as shown in Fig. 4), and (2) GazeProximity $= 1.0$, indicating that gaze persisted towards and fixated on the cued region of interest.

IV. SUMMARY AND PROPOSED FUTURE WORK

In this paper, we proposed a theory to explore how adding cue-based information can affect the FEM segmentation-prediction performance.

In pursuit of that goal we had to reach two intermediate milestones. First, we needed to discover how to integrate cue-based information into the FEM model, which has corresponding computational implementation implications for a variety of *other* task-based and person-based factors as mediators of segmentation activity. Second, we posited various kinds of computational renditions of cue-based information in an attempt to make our theory independent of an ultimate computational model, but still consistent with the computational architecture that is expected by the FEM.

We plan on conducting a future simulation-based study to verify if our proposed method of incorporating cue-based information can result in a meaningful shift in the FEM baseline prediction score. To this end, we will modify the original FEM prediction score simulation code to include parts that account for the proposed *CueInfluence* factor. For that study, we expect to test the following hypotheses:

1) Both types of cues affect the GazeProximity factor equally and can influence event perception.
2) Overt cues dominate GazeProximity; they more effectively draw attention and influence event perception.
3) Covert cues dominate GazeProximity; they more effectively draw attention and influence event perception.

While our third hypothesis can seem contradictory to the difference between overt and covert cues discussed earlier, there exist several studies which suggest that people respond faster to subtle cues and that they are often harder to ignore [8], [19], [20]. As we detailed in our Method section in this work, we will assess the validity of our proposed configurations regarding the influence of spatial cues in eliciting event segmentation, as predicted by the FEM. We plan to reach out to the authors of the original FEM paper to obtain datasets used to validate their model. We will augment the datasets with synthesized data using the methodology discussed earlier in the paper, and evaluate the hypotheses above.

We will be successful in our aim if – after modifying the FEM – we are able to observe a meaningful shift in the prediction of event segmentation based solely on the introduction of the *CueInfluence* factor. This would suggest that – even in the presence of other factors which demonstrably impact event segmentation – we are able to "induce" segmentation as a function of *CueInfluence* alone. In turn, that would suggest that it is in principle possible to construct an experimental evaluation where we empirically observe the simulated effect (which is our eventual end goal).

Our work adopts the definitions of *central visual field* and *peripheral vision* as established by Strasburger et al. [11]. However, we recognize that there is no consensus on how these two regions are defined in the literature, meaning that our eventual simulation will be contextualized by this choice.

We expect this work will contribute to a deeper theoretical understanding of how spatial cues can affect event segmentation processes in immersive virtual environments. Further, the interdisciplinary nature of this work, bridging immersive technologies, cognitive psychology, and computational modeling, makes it relevant to the cognitive systems community.

ACKNOWLEDGMENTS

This material is based upon work supported in part by the U.S. National Science Foundation (Grant #2046294).

REFERENCES

[1] K. E. Steiner and L. Voruganti, "A comparison of guidance cues in desktop virtual environments," *Virtual Reality*, vol. 7, pp. 140–147, 2004.

[2] M. Ward, A. Barde, P. N. Russell, M. Billinghurst, and W. S. Helton, "Visual Cues to Reorient Attention from Head Mounted Displays," in *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*, vol. 60, no. 1. Los Angeles, CA, USA: SAGE Publications, 2016, pp. 1574–1578.

[3] D. Lange, T. C. Stratmann, U. Gruenefeld, and S. Boll, "Hivefive: Immersion preserving Attention Guidance in Virtual Reality," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13.

[4] J. O. Wallgrün, M. M. Bagher, P. Sajjadi, and A. Klippel, "A Comparison of Visual attention Guiding Approaches for 360 Image-based VR tours," in *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*. IEEE, 2020, pp. 83–91.

[5] A. Bokulich, "How the Tiger Bush Got Its Stripes:'How Possibly' vs. 'How Actually' Model Explanations," *The Monist*, vol. 97, no. 3, pp. 321–338, 2014.

[6] I. van Rooij, O. Guest, F. G. Adolfi, R. de Haan, A. Kolokolova, and P. Rich, "Reclaiming AI as a Theoretical Tool for Cognitive Science," *PsyArXiv*, 2023. [Online]. Available: https://doi.org/10.31234/osf.io/4cbuv

[7] G. A. Radvansky, S. D'Mello, R. G. Abbott, B. Morgan, K. Fike, and A. K. Tamplin, "The Fluid Events Model: Predicting Continuous Task Action Change," *Quarterly Journal of Experimental Psychology*, vol. 68, no. 10, pp. 2051–2072, 2015.

[8] M. I. Posner, "Orienting of Attention," *Quarterly Journal of Experimental Psychology*, vol. 32, no. 1, pp. 3–25, 1980.

[9] G. A. Radvansky, "Across the Event Horizon," *Current Directions in Psychological Science*, vol. 21, no. 4, pp. 269–272, 2012.

[10] G. A. Radvansky and J. M. Zacks, "Event Boundaries in Memory and Cognition," *Current opinion in Behavioral Sciences*, vol. 17, pp. 133–140, 2017.

[11] H. Strasburger, I. Rentschler, and M. Jüttner, "Peripheral Vision and Pattern Recognition: A Review," *Journal of Vision*, vol. 11, no. 5, pp. 1–82, 2011.

[12] A. Khan, J. Matejka, G. Fitzmaurice, and G. Kurtenbach, "Spotlight: Directing Users' Attention on Large Displays," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 2005, pp. 791–798.

[13] Y.-C. Lin, Y.-J. Chang, H.-N. Hu, H.-T. Cheng, C.-W. Huang, and M. Sun, "Tell me Where to Look: Investigating Ways for Assisting Focus in 360° Video," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, 2017, pp. 2535–2545.

[14] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm, "Subtle gaze direction," *ACM Transactions on Graphics (TOG)*, vol. 28, no. 4, pp. 1–14, 2009.

[15] C. M. Pinto, A. M. Lopes, and J. T. Machado, "A review of power laws in real life phenomena," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 9, pp. 3558–3578, 2012.

[16] C. K. Friesen, J. Ristic, and A. Kingstone, "Attentional effects of counterpredictive gaze and arrow cues." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 30, no. 2, p. 319, 2004.

[17] B. Hommel, J. Pratt, L. Colzato, and R. Godijn, "Symbolic control of visual attention," *Psychological science*, vol. 12, no. 5, pp. 360–365, 2001.

[18] J. Tipples, "Eye gaze is not unique: Automatic orienting in response to uninformative arrows," *Psychonomic bulletin & review*, vol. 9, no. 2, pp. 314–318, 2002.

[19] M. I. Posner, C. R. Snyder, and B. J. Davidson, "Attention and the detection of signals." *Journal of Experimental Psychology: General*, vol. 109, no. 2, p. 160, 1980.

[20] L. Riggio and K. Kirsner, "The relationship between central cues and peripheral cues in covert visual orientation," *Perception & Psychophysics*, vol. 59, pp. 885–899, 1997.