VR Reading Aid

Nancy Liddle Columbia University
New York, NY, USA
nll2128@columbia.edu

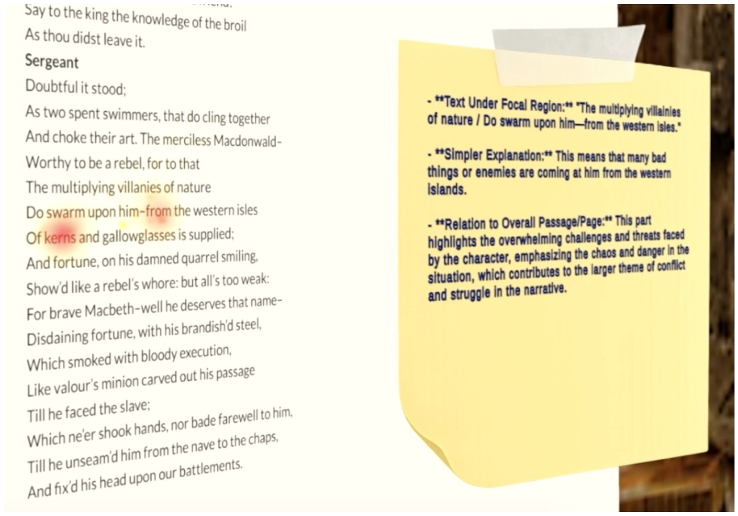

Figure 1: Example Usage of the Reading Aid

Abstract

Encountering confusion is an inevitability while learning: if a student comes across text that alludes to or discusses an unfamiliar phrase, algorithm, or event, they often disengage long enough to search for an explanation. This breaks the learning flow and increases cognitive overhead. In order for the brain to retain information, it has to be in a relaxed but attentive state. Prolonged high cognitive load can lead to being overwhelmed and is not conducive to learning [7] [8]. Immersive environments facilitated by a head-mounted display offer the opportunity to enhance the learning experience and scaffold user cognition. This project presents a Virtual Reality reading assistant that estimates moments of elevated mental effort using the HP Reverb G2 Omnicept Edition's multi-modal physiological sensing. When this signal exceeds a threshold, this acts as a trigger for the system to identify the region of confusion by aggregating the gaze clusters of the previous 30 seconds. A large language model is than queried and the context-aware explanation is overlaid beside the user's point of fixation. In this abstract, we discuss the implementation of the immersive environment.

Index Terms: Virtual reality, cognitive load, eye tracking, pupillometry, PPG, and large language models

1 Introduction & Related Work

Virtual Reality has risen in popularity as a potential learning modality due to its ability to facilitate a controlled environment. Prior work shows AR can improve EFL learners' reading comprehension [1]. Additionally, VR has been shown to modulate brain rhythms associated with neuroplasticity, learning, and memory [6]. Identifying when and where confusion occurred is the first step in enhancing the learning process. Cognitive load is used to describe how much mental effort is required to complete a task. Cognitive Load can infer memory retention, stress levels, confusion, attention, and learning performance. There is extensive literature identifying the predictive powers certain features have for cognitive load, including facial features, pulse, blood pressure, eye movements, pupillometry, EEG signals, and voice volume [3]. Head-mounted displays that capture multimodal data often combine eye tracking to infer user cognitive state [3][9]. Recent work demonstrates real-time cognitive load inference in VR using multimodal physiological markers [9], and has explored attention-related signals for adaptive XR interaction [5]. Wearable sensing in AR has been used to distinguish internal vs. external attention [4]. Researchers have explored the cognitive load classification accuracy of classical machine learning models trained on these features, including: k-nearest neighbor (kNN), naive Bayes (NB), logistic regression, linear discriminant analysis (LDA), support vector machines(SVM), ensemble methods (random forest, XGBoost), and neural networks [9]. For this study, we implemented a VR Reading Aid that builds upon these studies.
This paper explores the effectiveness of:

- Scaffolding the learning experience in VR with real-time cognitive load detection.
- Identifying points of confusion by implementing a gaze heatmap over the provided text.
- Gleaning insight from an LLM after querying for a contextually aware explanation of the identified point of confusion.

2 Methodology

2.1 Classifying High Cognitive Load

The HP Reverb G2 Omnicept headset has built-in capabilities to estimate cognitive-load levels. The headset receives multimodal streams of data, including pupillometry/eye tracking and pulse, to classify cognitive load as a continuous value between $[0,1]$ with an uncertainty estimate. The model was developed in previous work [9]. Rather than re-training or modifying this model, we consume the headset-provided estimate at runtime and use it only to trigger an intervention when mental effort appears elevated for a sustained period. Cognitive load does not necessarily indicate confusion but can reflect focused effort, novelty, or engagement. We use it as a heuristic to decide when to localize the point of confusion in the text. Should the user's cognitive load exceed a threshold, we identify the most likely source of difficulty on the page using gaze clustering.

2.2 Estimating Region of Confusion with DBSCAN

We follow the gaze-density visualization approach used by Zhuang et al. [10] as an inspiration for highlighting where attention concentrates during reading. However, Zhuang et al. combine eye tracking with EEG to classify when and what type of confusion may have occurred, whereas our system uses eye tracking–based localization only and does not attempt to classify confusion types. Our use of the gaze signal is limited to selecting candidate text regions that triggered high cognitive load. We overlay a gaze heatmap, an estimate for regions of concentration, onto the reader's view. Points of gaze intensity, rendered in red, are found utilizing a strategy developed by Zhuang et al. [10]. The gaze heatmap is calculated by the following steps:

2.2.1 Identifying Gaze Location

Let $P = \{p_i\}_{i=1}^{N}$ be all the gaze points from the previous 30 seconds, and $p_i = (x_i, y_i)$ denote the (x, y) coordinates of gaze point i. Unlike the 2D text plane that is utilized by Zhuang et al, our text is rendered in a 3D VR scene. Therefore, to identify (x_i, y_i), we cast a ray from the camera into the scene along the headset's gaze direction. If the ray intersects the text plane, we take the 3D intersection point and project it into the camera view to obtain a normalized 2D gaze location on the screen.

2.2.2 DBSCAN Algorithm Application

We apply the DBSCAN algorithm to P to identify regions where gaze samples are densely concentrated, while treating isolated samples as noise [2]. In DBSCAN, a location is considered a cluster if m neighbors lie within an ε radius.

For each $p_i \in P$, sum the number of points within a ε radius. If the count meets or exceeds m, mark the point as a core point. For each core point that is not already assigned to a cluster, a new cluster is created. This is performed recursively until all density-connected points are found. Two points a and b are density-connected if there exists a chain of points where each point is within the neighborhood of the next and at least one point in the chain is a core point. This chaining process ensures that all points in a cluster are connected through a series of dense regions. After running the algorithm, if any point remains outside a cluster, it is considered noise.

2.2.3 Identifying Spurious Gaze Clusters

Not all gaze clusters reflect meaningful reading behavior. For each cluster C_k, we compute the total dwell time and estimate how many times the gaze returns to that region. Should both metrics surpass our threshold, the cluster is retained.

2.2.4 What is included in the LLM Query

We find the location of confusion by determining the median point within the cluster that had the longest summed duration.

2.3 Explanation overlay

When the HP Reverb G2 Omnicept identifies that cognitive load has surpassed a threshold, we take a screenshot of the user's view (including the overlaid gaze heatmap). We then prompt Open Router with the screenshot from the reader's view. The overlaid gaze heatmap allows the prompted LLM to identify where in the text confusion occurred by the intensity of the red. The prompt asks the LLM to first identify the phrase of confusion, explain that phrase, and then explain the phrase within the larger context. The returned explanation is rendered as a small annotation on a yellow sticky note: positioned adjacent to the identified region of concentration. To reduce distraction, the note appears only after the LLM response is available. To handle repeated triggers, we enforce a cooldown period and display at most one note at a time. If a new trigger occurs, the system updates the content of the existing note, and relocates it to the new region.

3 Conclusion & Future Study

This work presents a sensor-driven VR reading assistant that detects elevated mental effort, localizes the likely source of confusion using gaze clustering, and overlays context-aware explanations. This project aims to reduce cognitive load to enhance the learning process. Cognitive load is not equivalent to confusion; it can reflect focused effort, novelty, or task engagement. Over-triggering risks annoyance and may distract. LLM outputs can be incomplete or incorrect. However, this project proves the potential of an immersive environment to enhance the learning process by providing aid in real time. The system aims to support learners and multilingual readers. Future research would address the concerns expressed above and evaluate learning and productivity gains yielded by the Reading-Aid.

References

[1] S. Ebadi and F. Ashrafabadi. An exploration into the impact of Augmented Reality on EFL learners' reading comprehension. *Education and Information Technologies*, 27:9745–9765, Apr. 2022. doi: 10.1007/s10639-022-11021-8 1

[2] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of KDD-96 (Knowledge Discovery and Data Mining)*, pp. 226–231. AAAI Press, 1996. 2

[3] M. Kolnes, A. Uusberg, and S. Nieuwenhuis. Broadening of attention dilates the pupil. *Attention, Perception & Psychophysics*, 86(1):146–158, Jan. 2024. doi: 10.3758/s13414-023-02793-3 1

[4] N. Kosmyna and E. Hauptmann. Are you still watching?: assessing internal and external attention in AR using wearable brain sensing glasses. In N. Argaman, H. Hua, and D. K. Nikolov, eds., *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR) V*, vol. 12913 of *Proceedings of SPIE*, p. 1291310. SPIE, 2024. doi: 10.1117/12.3023316 1

[5] N. Kosmyna, C.-Y. Hu, Y. Wang, Q. Wu, C. Scheirer, and P. Maes. A pilot study using covert visuospatial attention as an EEG-based brain computer interface to enhance AR interaction. In *Proceedings of the 2021 International Symposium on Wearable Computers (ISWC '21)*, pp. 43–47. Association for Computing Machinery, 2021. doi: 10.1145/3460421.3480420 1

[6] C. Seydel. Virtual reality boosts brain rhythms crucial for neuroplasticity, learning and memory. UCLA Health, June 2021. Accessed: 2025-12-19. 1

[7] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285, 1988. doi: 10.1207/s15516709cog1202_4 1

[8] J. Sweller, J. J. G. van Merriënboer, and F. G. W. C. Paas. Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251–296, 1998. doi: 10.1023/A:1022193728205 1

[9] J. Wei, E. Siegel, P. Sundaramoorthy, A. Gomes, S. Zhang, M. Vankipuram, K. Smathers, S. Ghosh, H. Horii, J. Bailenson, and R. T. Ballagas. Cognitive load inference using physiological markers

in virtual reality. In *2025 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2025. doi: 10.1109/VR59515.2025.00098 1, 2

[10] H. Zhuang, D. Baradari, N. Kosmyna, A. Balyan, C. Albrecht, S. Chen, and P. Maes. Detecting reading-induced confusion using EEG and eye tracking. arXiv, Aug. 2025. doi: 10.48550/arXiv.2508.14442 2