Brain-Enhanced Applications: A/B Testing in Virtual Reality Education through EEG and Machine Learning

Vitor G. Inserra
Department of Computer Science
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina
vgasparetto-inserra@unc.edu

***Abstract*—This study investigates the use of EEG data to generate enhancements for user experience in educational Virtual Reality applications. By leveraging high-frequency data, Machine Learning models can identify cognitive states, offering targeted insights for application improvement or performance data for A/B testing. This approach suggests that cognitive states, including attention, load and memorization, can be classified with high accuracy, enabling the analysis of cognitive performance throughout usage of an application. Such data-driven advancements aim to optimize e-learning effectiveness and adaptability, with implications for addressing educational gaps and promoting enhancement strategies for emerging technologies.**

Keywords—EEG, Virtual Reality, E-learning, Education Software, Split Testing, A/B Testing.

I. INTRODUCTION

With the popularization of A/B or *split* testing and the successful results of this advent in the mass software industry, limitations to this practice have been found relative to the sample sizes and data availability required for testing [1]. While typical strategies for A/B testing may work for some applications, software that highly depend on cognitive processes–such as education software–lack informative data for testing, which limit insight into user experience. However, with the current status of *Electroencephalography* (EEG) devices, we may leverage an alternative for the low-resolution data through continuous readings with the ability to be classified through Machine Learning into cognitive states throughout the usage of software. Besides, there are currently many high-quality non-invasive dry EEG devices that do not require the user to sanitize after usage.

EEG combined with eye-tracking, head positioning, and other information, may generate particular insights relative to the user's perception of a software application. Furthermore, it creates a significantly more effective strategy for enhancing educational methods through objective evaluation. This project is particularly focused on conducting those tests in Virtual Reality (VR) for three reasons. First, VR provides an immersive learning experience with rich and complex environments that would require extremely informative data to be tested upon. Second, typical VR devices may provide more sensing strategies and be discreetly infused with electrode sensors, such as EEG, to achieve that goal. Third, the controlled nature of the user's visual environment reduces noise in the data and removes many of the external factors that could either ruin the quality of the data or skew the analysis algorithms due to the lack of context relative to the extraneous factor.

II. SYSTEM OVERVIEW

The current version of the system runs on the Meta Quest Pro VR Headset (show in Figure 1a) and a *Python-FastAPI* server in the backend. With that, a dry EEG device (show in Figure 1b) is placed on the user. Data is collected continuously throughout the completion of an assessment that requires spatial cognition and memorization. The user is then asked to complete the assessment and is assigned to one of two slightly different versions. Then, as the software runs on the VR headset, the user's eye-tracking and head movement data is sent over internet to the backend service in batches. Meanwhile, the server will request and process EEG data, matching the timestamps collected from eye-tracking. These readings, coming in at 274,440 floating points per minute, are saved to a *PostgreSQL* database to be analyzed for post-collection analysis.

Figure 1: a) The Meta Quest Pro Virtual Reality Headset. b) The OpenBCI "Mark IV" Ultracortex EEG Headset.

After the raw EEG signals are processed, the data can be analyzed by ML classifiers and decoded into cognitive states. In combination, eye-tracking and head movement data are injected into the classifier's analysis. The user's cognitive performance should then be evaluated within 3 categories:

1. attention level,
2. memory performance, and
3. cognitive effort.

Lastly, each part of the journey is labeled with this computed cognitive state by using timestamp, head-tracking, and eye-tracking data.

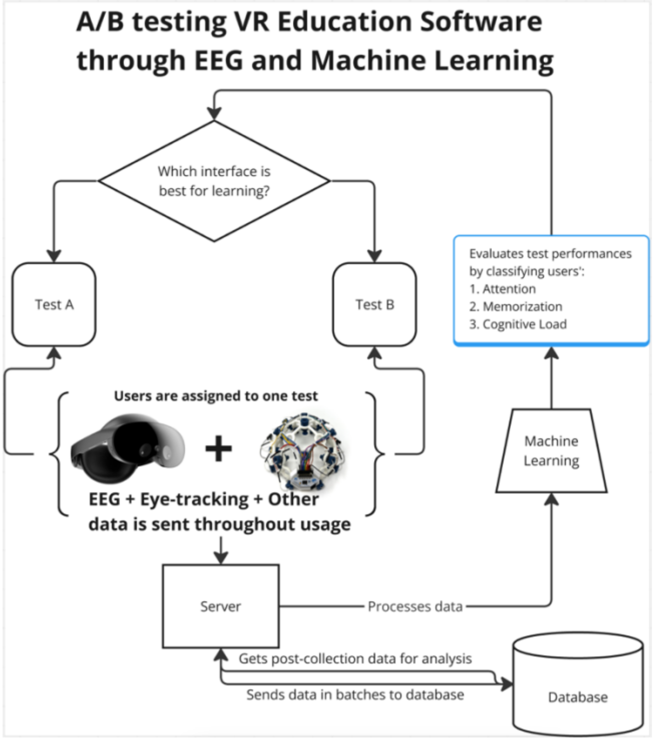

Figure 2: Diagram of A/B testing methodology using EEG and VR over education software.

This method enables complex interface experimentation, providing comprehensive conclusions about the user journey to the experimenter. Naturally, *split* testing is expected to generate enhancements by elimination for the educational software being tested. The main goal of the *split* test is to capture which assessment is able to maintain cognitive engagement and performance successfully, showing which version is better optimized for users' learning processes.

III. Supporting Work

Previous success in decoding cognitive states through classifiers and in using EEG to analyze user experience enables EEG-based A/B testing as a form to enhance software interfaces. Thus, this project combines strategies from User Experience, EEG, and A/B testing research.

A. *Decoding Cognitive States*

A 1999 study first found EEG can measure cognitive and memory performance, because they were reflected by high amplitudes in alpha and theta bandwidths [3]. After a few years of development in the EEG realm, a 2022 study collected data throughout Tibetan monastic debates to classify cognitive states of attention and distraction. They achieved high predictability, with accuracy reaching 95.86% for attention and distraction states by alpha, delta and theta waves in the *fronto-central* regions [4].

B. *User Experience through EEG*

Usage of EEG has started to gain some attention in UX research. Previous experiments have highlighted this method as a reliable tool for evaluating software interfaces as it can objectively assess UX by tracking cognitive and emotional responses to different interface designs. For example, a 2023 study compared participants' EEG responses to various adaptive graphical menus. The study analyzed UX factors like cognitive load, engagement, attraction, and memory. Findings revealed that different menu structures and fonts created distinct brain responses, whereas color-based designs elicited more uniform responses across users [2].

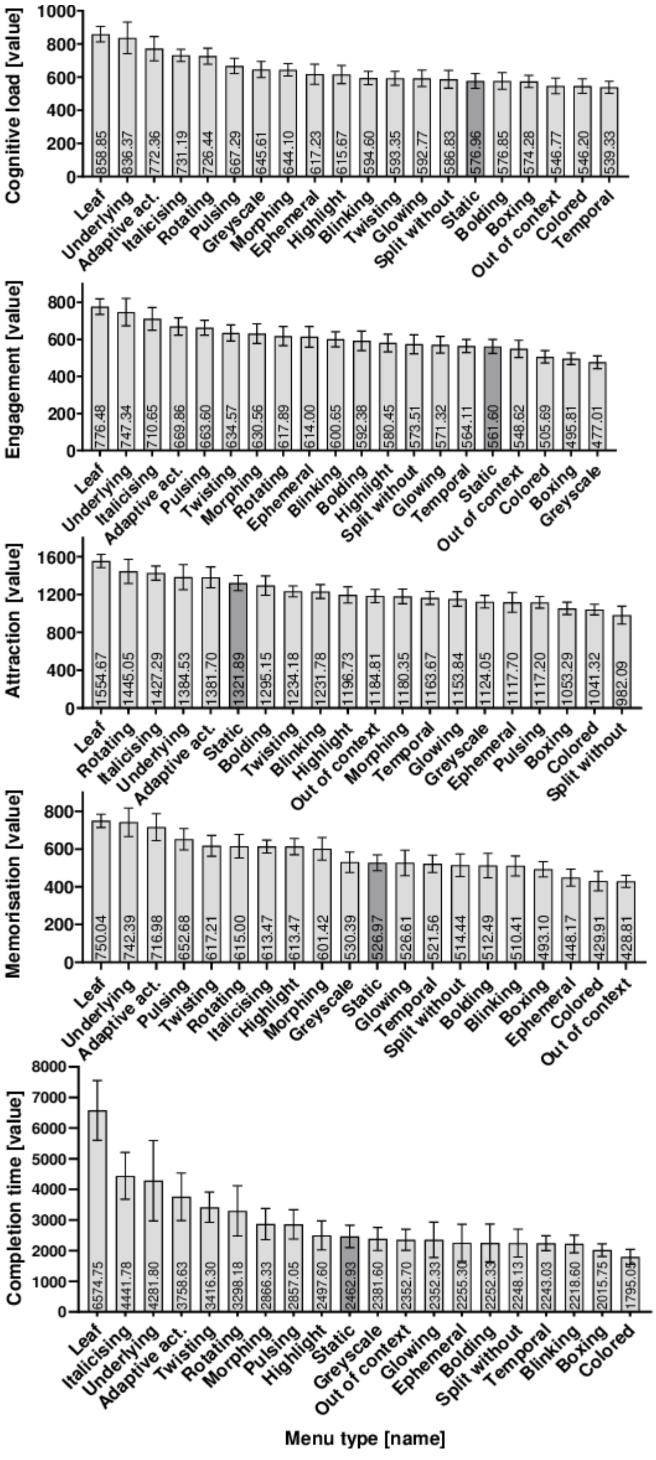

Figure 3: Results collected for each type of adaptive menu in "Measuring User Experience of Adaptive User Interfaces using EEG: A Replication Study" [2].

C. *A/B Testing*

In *"Trustworthy Online Controlled Experiments"*, Ronny Kohavi highlights a successful *split* test at Microsoft Bing aimed at improving user engagement. The test compared the original layout with one that included related searches. Data such as click-through rates, dwell time, and task completion rates were tracked. Results showed that the "related searches" feature increased user engagement, with higher click-through rates and longer dwell times. Ultimately, this experiment was shown to have had relevant impact in revenue [1].

This is one of many examples that show the success of A/B testing in enhancing user engagement [5]. Even with sporadic and relatively uninformative data, when compared to EEG sensing, *split* testing is capable of generating significant enhancements. These examples, however, required a massive user base and many repeated interactions from those users. In contrast, the use of richer data for *split* tests would allow statistically significant results with much smaller sample sizes.

IV. Conclusion

EEG-based studies have been successful in analyzing a plethora of attributes related to UX that could not have been possible through traditional studies. At best, traditional UX studies attempt to indirectly predict what EEG provides overtly. With that said, this project attempts to generate data-driven enhancements to a piece of software through *split* style testing. Moreover, the underlying concept behind this project can be explored in order to use other forms of testing. EEG can then be leveraged to evaluate A/B/n or multivariate tests, besides many other strategies. Additionally, EEG can be leveraged to provide self-enhancing apps that automatically rearrange layout, control the journey's pace, change colors etc. as it gathers continuous real-time information of the user's cognitive state throughout their journey. Exploring these complex methods of data-driven enhancements will first require success in correctly setting up A/B tests in the EEG-VR format.

Education software becomes a viable target for this form of testing as it requires rich and continuous data that is highly user-specific—given users vary dramatically in cognitive attributes and learning abilities. Therefore, this project aims at structuring efficient collection for that data and training ML algorithms to classify cognitive performance that serves as criteria for A/B tests over the usage of a spatial cognition and memorization assessment.

V. References

[1] Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press.

[2] Gaspar-Figueiredo, D., Abrahão, S., Insfrán, E., & Vanderdonckt, J. (2023). Measuring User Experience of Adaptive User Interfaces using EEG: A Replication Study. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering, EASE 2023* (pp. 52-61). ACM.

[3] Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29, 169-195.

[4] Kaushik, P., Moye, A., van Vugt, M., & Roy, P. P. (2022). Decoding the cognitive states of attention and distraction in a real-life setting using EEG. *Scientific Reports*, *12*(1), 20649.

[5] Dmitriev, P., Gupta, S., Kohavi, R., Deng, A., Raff, P., & Vermeer, L. (2017). *A/B Testing at Scale*. Tutorial presented at SIGIR 2017 and KDD 2017. ExP Platform.