Low-Cost Real-Time Mental Load Adaptation for Augmented Reality Instructions - A Feasibility Study

Dennis Wolf* Tobias Wagner† Enrico Rukzio‡

Ulm University
Ulm, Germany

ABSTRACT

Since the introduction of augmented reality (AR) technology, in-situ instructions for manual tasks have been a central use case for a large body of previous work. However, most implementations provide identical sets of instructions to each user disregarding the user's current mental load. This is a major issue since previous work has shown the importance and potential of an adapted instruction fidelity for manual tasks such as playing an instrument. To implement a low-cost mental load adaptation for AR instructions, we evaluated a mobile off-the-shelf electroencephalographic (EEG) device for its suitability and feasibility to measure mental load while wearing a video see-through AR head-mounted display (HMD). In a first user experiment (n=12), data of EEG power band values and proprietary performance metrics of the manufacturer were collected and analysed regarding their validity to estimate the user's mental load. Our results indicate that our setup successfully induced different levels of mental effort. The proprietary performance metrics, however, only partially reflected the participants' current mental effort and require further analysis.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—Usability testing

1 INTRODUCTION

Dynamic difficulty adjustment systems can be found in many game environments where they are supposed to keep players in a state of flow and, thus, increase player enjoyment. Since it has been shown that the emotional and cognitive state of a user can impact performance in human-computer interaction in general [5, 8], this motivated us to explore the potential of adaptation for a non-gaming context such as in-situ instructions in AR.

While there has been research on input techniques and modalities for AR where users can switch between techniques or use them complementary depending on the context [10], research on context-sensitive output adaptation in real-time is sparse and mostly focuses on desktop applications rather then HMDs [11]. Previous research on assessing a user's cognitive or mental load relied mostly on expensive medical-grade sensors such as EEGs [2] and functional near-infrared spectroscopy (fNIRS) [1, 11], limiting not only the accessibility to the community and but also the potential application areas such as mobile scenarios. Furthermore, most approaches are limited to a post-hoc analysis of mental load [7].

To evaluate the feasibility of a low-cost framework that could measure the users' mental load in real-time and cover a wide range of scenarios, we performed a preliminary study with a low-cost EEG device and a video see-through AR HMD. Our aim was to evaluate the validity of our low-cost setup to measure mental load and to explore the potential of the manufacturer's performance metrics [4,9] to estimate the user's current mental state.

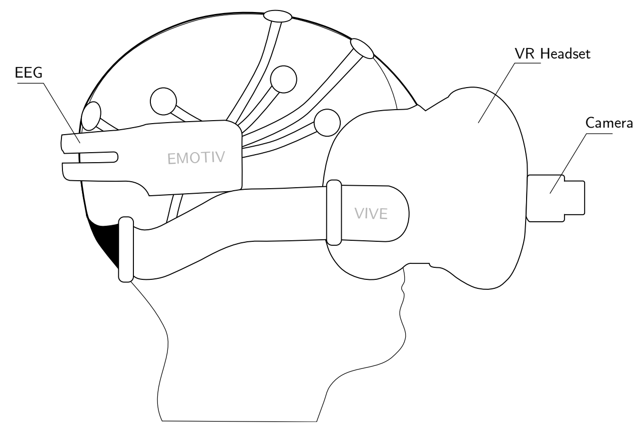

Figure 1: A ZED Mini stereo camera is mounted in front of the Vive HMD to provide high resolution video see-through AR. A custom head mount was 3D printed to accommodate the EEG beneath the HMD and reduce pressure on the electrodes.

2 PRELIMINARY STUDY

To collect EEG data on different levels of mental effort, n-back tasks were applied to stimulate the working memory of the participants [3]. The independent variable *Task* was defined with three levels of difficulty: *0-back*, *2-back*, and *3-back*. The measured variables were the self-reported mental effort on the Rating Scale Mental Effort (*RSME*) instrument that has a scale from 0 ('absolutely no effort') to 150 ('extreme effort') [12] and the performance metrics (*Performance*) that consist of the 7 categories *Interest*, *Stress*, *Relaxation*, *Excitement*, *Engagement*, *LongTermExcitement*, and *Focus* [4].

To our best knowledge, the combination of a low-cost EEG device and a video see-trough HMD has not been evaluated for measuring mental load in previous work before. The delay of video see-through AR in combination with the physical pressure exerted by the EEG might impact the mental effort created by n-back tasks. Hence, the first hypothesis to evaluate was:

H1: There is a significant increase in mental effort (*RSME*) for increasing *Task* difficulty presented in our setup.

The calculations behind the proprietary performance metrics of the manufacturer are not known and therefore require a validation of their feasibility to reliably measure mental load in real-time. Thus, the following hypothesis was tested in this experiment:

*e-mail: dennis.wolf@uni-ulm.de
†e-mail: tobias.wagner@uni-ulm.de
‡e-mail: enrico.rukzio@uni-ulm.de

	Metric						
Part.	Interest	Stress	Relaxation	Excitement	Engagement	LT Excitement	Focus
1	$\chi^2(2) = 14.0^{**}$	$\chi^2(2) = 14.0^{**}$	$\chi^2(2) = 10.17^{*}$	$\chi^2(2) = 2.67$	$\chi^2(2) = 3.17$	$\chi^2(2) = 13.17^{*}$	$\chi^2(2) = 8.17^{*}$
2	$\chi^2(2) = 1.5$	$\chi^2(2) = 6.5^{*}$	$\chi^2(2) = 3.07$	$\chi^2(2) = 2.17$	$\chi^2(2) = 8.0^{*}$	$\chi^2(2) = 6.5^{*}$	$\chi^2(2) = 7.17^{*}$
3	$\chi^2(2) = 1.5$	$\chi^2(2) = 13.17^{*}$	$\chi^2(2) = 0.5$	$\chi^2(2) = 20.67^{**}$	$\chi^2(2) = 0$	$\chi^2(2) = 24.0^{**}$	$\chi^2(2) = 10.67^{*}$
4	$\chi^2(2) = 12.17^{*}$	$\chi^2(2) = 8.67^{*}$	$\chi^2(2) = 6.5^{*}$	$\chi^2(2) = 6.5^{*}$	$\chi^2(2) = 16.17^{**}$	$\chi^2(2) = 13.5^{*}$	$\chi^2(2) = 4.5$
5	$\chi^2(2) = 8.17^{*}$	$\chi^2(2) = 1.17$	$\chi^2(2) = 3.17$	$\chi^2(2) = 6.0^{*}$	$\chi^2(2) = 11.17^{*}$	$\chi^2(2) = 15.5^{**}$	$\chi^2(2) = 10.5^{*}$
6	$\chi^2(2) = 1.5$	$\chi^2(2) = 15.17^{*}$	$\chi^2(2) = 6.5^{*}$	$\chi^2(2) = 6.7$	$\chi^2(2) = 9.5^{*}$	$\chi^2(2) = 8.17^{*}$	$\chi^2(2) = 3.17$
7	$\chi^2(2) = 8.67^{*}$	$\chi^2(2) = 6.17^{**}$	$\chi^2(2) = 5.17$	$\chi^2(2) = 10.17^{*}$	$\chi^2(2) = 4.67$	$\chi^2(2) = 8.17^{**}$	$\chi^2(2) = 16.17^{**}$
8	$\chi^2(2) = 22.17^{**}$	$\chi^2(2) = 24.0^{**}$	$\chi^2(2) = 4.41$	$\chi^2(2) = 14.0^{*}$	$\chi^2(2) = 22.55^{**}$	$\chi^2(2) = 15.5^{**}$	$\chi^2(2) = 11.17^{*}$
9	$\chi^2(2) = 12.17^{*}$	$\chi^2(2) = 13.17^{*}$	$\chi^2(2) = 15.5^{**}$	$\chi^2(2) = 6.5^{*}$	$\chi^2(2) = 10.17^{*}$	$\chi^2(2) = 22.17^{**}$	$\chi^2(2) = 7.17^{*}$
10	$\chi^2(2) = 2.17$	$\chi^2(2) = .17$	$\chi^2(2) = 4.17$	$\chi^2(2) = 8.17^{*}$	$\chi^2(2) = .5$	$\chi^2(2) = 9.5^{*}$	$\chi^2(2) = 10.67^{*}$
11	$\chi^2(2) = 14.0^{*}$	$\chi^2(2) = 15.5^{**}$	$\chi^2(2) = 18.77^{**}$	$\chi^2(2) = 12.67^{*}$	$\chi^2(2) = 1.17$	$\chi^2(2) = 18.0^{**}$	$\chi^2(2) = 4.5$
12	$\chi^2(2) = 10.67^{*}$	$\chi^2(2) = 14.0^{*}$	$\chi^2(2) = 5.17$	$\chi^2(2) = 7.17^{*}$	$\chi^2(2) = 11.72^{*}$	$\chi^2(2) = 8.17^{*}$	$\chi^2(2) = 9.5^{*}$

Table 1: Results of the Friedman ANOVA for each participant. Significant differences ($^*p < .05$, $^{**}p < .001$) in the *Perfomance* metric values for the three *Task* conditions are highlighted.

H2: There is a significant difference between values of *Performance* for different levels of *Task*.

2.1 Apparatus

An Emotiv Epoc+ 14-channel EEG was used to measure real-time EEG data with a rate of 8 Hz for band power and 0.1 Hz for performance metrics. The data was steamed via Bluetooth and the manufacturer's *Cortex* SDK to a VR Notebook running a custom Unity 3D application. A Vive HMD was connected to the Notebook and attached on top of the EEG with a custom 3D-printed strap that leaves sufficient space around the ears to fit the EEG electrodes (see Fig. 1). Since the default camera of the Vive HMD (480p, 200ms delay) is not fit for low-latency and high-resolution video see-through AR, a ZED Mini stereo camera (720p, 60ms delay) was mounted on the HMD. The trigger of one Vive controller was used for selections during the n-back tasks and the touch pad of the second Vive controller to select values on the RSME scale that was presented as an interface attached to the controller. A marker was positioned at 50 cm distance in front of the participant and was used to position the virtual interface for the n-back task.

2.2 Participants and Procedure

12 participants (7 male, 5 female) with an average age of 23.17 (SD=1.46) were recruited from our institution. 58.33% reported to have already experienced AR at some point in their life. All participants had normal or corrected-to-normal vision.

After an introduction to the procedure and privacy policy, participants filled out a demographic questionnaire and had the opportunity to practise all three n-back task levels on a computer screen. Each n-back task consisted of 40 characters of which 12 were targets. Characters were displayed for 500 ms with a 2500 ms inter-stimulus interval. Only consonants were chosen to prevent chunking which might reduce mental effort [6]. After the training, the EEG and the Vive HMD were mounted on the participants' head and controllers were placed in their hands. Participants were seeing their environment via video see-through AR and a virtual interface displaying the n-back task in-front of them. In a comfortable sitting position participants went through two sessions with a break of 5 minutes. Each session started with a rest task by focusing on a centred cross for 2 minutes. A label above the interface indicated the level of difficulty of the next trial. Each level was presented twice per session and lasted for 2 minutes. Trials were counter-balanced with a Latin square. Afterwards, participants rated their mental effort on the RSME scale that was displayed as an interface attached to the second controller. A trigger press confirmed the selection and started the next trial.

2.3 Results

2.3.1 RSME

A Friedman ANOVA revealed significant differences in the *RSME* scores for the three *Task* conditions, $\chi^2(2) = 24.0, p < .001$. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < .0167$. Median (IQR) *RSME* levels for the *Task* levels *0-back*, *2-back*, and *3-back* were 19.125 (9.63 to 23.69), 55.25 (42.25 to 62.56) and 85.0 (71.88 to 91.81), respectively (see Fig. 2). There was a significant difference between the *0-back* and the *2-back* trials ($Z = -3.509, p = .002$), the *0-back* and the *3-back* trials ($Z = -3.509, p = .002$) and the *2-back* and the *3-back* trials ($Z = -3.061, p = .002$).

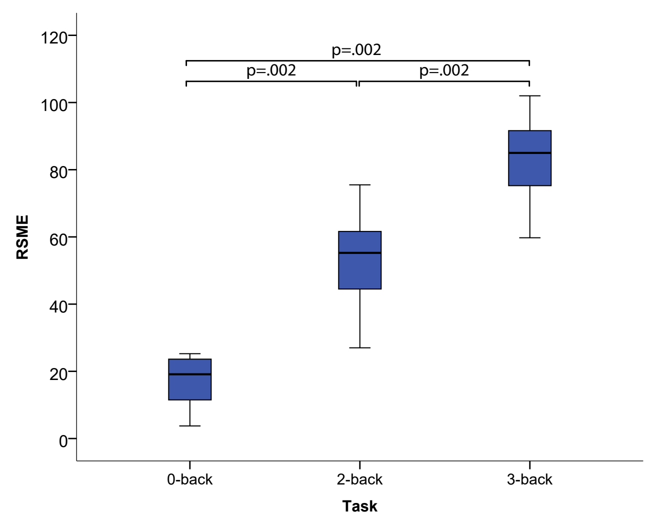

Figure 2: A box plot of subjective mental effort for each *Task* level measured with the RSME instrument on a scale from 0 to 150.

2.3.2 Performance

All values of the seven *Performance* metrics that were recorded during the second session were explored for each participant individually. Differences in the three *Task* conditions were analysed with a Friedman ANOVA. The results can be found in Table 1. Of seven *Performance* metrics *LongTermExcitement* showed significant differences for each participant.

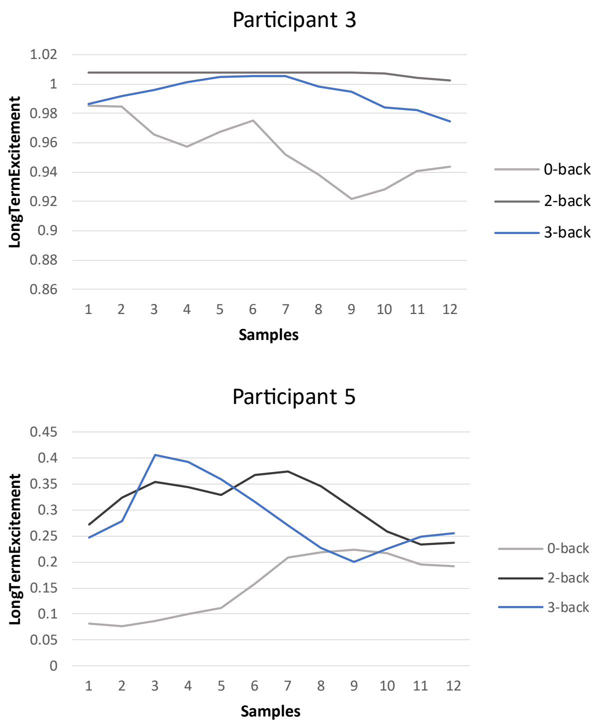

Figure 3: *LongTermExcitement* values of two exemplary participants during two minute trials (samples at 0.1 Hz) at each level of *Task*. Please note the different scaling of the y-axes.

3 Discussion

Our results indicate that mental effort was significantly higher for increasing levels of *Task*. We therefore accept *H1* and assume that our setup is valid to induce different levels of mental effort. As can be seen in Table 1, only *LongTermExcitement* expressed significant differences for all participants while *Relaxation* yielded the lowest number of significant results. We can therefore only partially accept **H2**. Although the results in Table 1 suggest that *LongTermExcitement* could be a reliable predictor of mental effort for all participants, the definition of a meaningful threshold for real-time adaptation is not trivial. First, the manufacturer defines *LongTermExcitement* as a metric that measures the overall mood rather than acute changes. Second, as can be see in Table 1, participants 3 and 5 displayed a highly significant difference in *LongTermExcitement* ($p < .001$). A close inspection of the metric values for both participants, however, reveals unique trends for each *Task* level giving tribute to the heterogeneity of EEG data (see Fig. 3). The results should therefore be interpreted with caution and require more research to define unique thresholds for each participant in the next user experiment. It should further be noted that the group differences were calculated for the complete duration of a trial which lasted two minutes whilst real-time adaptation should be able to react within a significantly smaller time window to fit the user's current mental effort.

4 Conclusion

In this work, we presented a preliminary study that aimed to evaluate the feasibility of a low-cost EEG device to measure a user's mental effort while wearing a video see-through AR HMD. We successfully induced three significantly different levels of mental effort via n-back tasks and found significant differences in some of the performance metric values provided by the manufacturer. Further research is necessary to interpret the results into meaningful thresholds that could distinguish between a low and a high mental effort and be applied to create a real-time adaptation system for AR. Furthermore, an analysis of the recorded power bands could yield additional features to derive a threshold.

Acknowledgments

This work was supported by the Emmy Noether research group "Mobile Interaction with Pervasive User Interfaces" funded by German Research Foundation (DFG) and by the Landesgraduiertenfrderung (LGFG) Scholarship for PhD students.

References

[1] D. Afergan, E. M. Peck, E. T. Solovey, A. Jenkins, S. W. Hincks, E. T. Brown, R. Chang, and R. J. Jacob. Dynamic difficulty using brain metrics of workload. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pp. 3797–3806. ACM, New York, NY, USA, 2014. doi: 10.1145/2556288.2557230

[2] A. O. G. Barbosa, D. R. Achanccaray, and M. A. Meggiolaro. Activation of a mobile robot through a brain computer interface. In *2010 IEEE International Conference on Robotics and Automation*, pp. 4815–4821, May 2010. doi: 10.1109/ROBOT.2010.5509150

[3] T. S. Braver, J. D. Cohen, L. E. Nystrom, J. Jonides, E. E. Smith, and D. C. Noll. A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, 5(1):49 – 62, 1997. doi: 10.1006/nimg.1996.0247

[4] Emotiv. Performance metrics. `https://emotiv.gitbook.io/emotivpro/data_streams/performance-metrics`.

[5] D. Giakoumis, A. Vogiannou, I. Kosunen, K. Moustakas, D. Tzovaras, and G. Hassapis. Identifying psychophysiological correlates of boredom and negative mood induced during hci. In *B-Interface*, pp. 3–12, 2010.

[6] M. A. Hogervorst, A.-M. Brouwer, and J. B. Van Erp. Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience*, 8:322, 2014.

[7] T. Kosch, M. Funk, A. Schmidt, and L. L. Chuang. Identifying cognitive assistance with mobile electroencephalography: A case study with in-situ projections for manual assembly. *Proc. ACM Hum.-Comput. Interact.*, 2(EICS):11:1–11:20, June 2018. doi: 10.1145/3229093

[8] R. Pekrun, T. Goetz, L. M. Daniels, R. H. Stupnisky, and R. P. Perry. Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, 102(3):531, 2010.

[9] M. Strmiska and Z. Koudelkov. Analysis of performance metrics using emotiv epoc +. *MATEC Web of Conferences*, 210:04046, 01 2018. doi: 10.1051/matecconf/201821004046

[10] D. Wolf, J. J. Dudley, and P. O. Kristensson. Performance envelopes of in-air direct and smartwatch indirect control for head-mounted augmented reality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 347–354. IEEE, 2018.

[11] B. F. Yuksel, K. B. Oleson, L. Harrison, E. M. Peck, D. Afergan, R. Chang, and R. J. Jacob. Learn piano with bach: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pp. 5372–5384. ACM, New York, NY, USA, 2016. doi: 10.1145/2858036.2858388

[12] F. R. H. Zijlstra. Efficiency in work behaviour: A design approach for modern tools. 1995.