Neural Correlates of Move Quality During Chess Games: a Low-Cost EEG Study

Matthew Russell, William Xia, Samuel Youkeles, Alexander Gu, Robert J.K. Jacob
Tufts University
Keywords: Chess, EEG, MUSE 2

ABSTRACT

Consumer-grade electroencephalography (EEG) presents promising opportunities for applied Brain-Computer Interfaces (BCI). We examine the relationship between neural activity and chess move quality using the MUSE 2. Linear mixed-effects modeling identifies significant positive correlations between move quality and power in beta and gamma bands. With machine learning, we investigate the potential application of these findings towards real-time implicit BCI. Using two-class classification with leave-one-out cross-validation across participants, we achieve 54% accuracy; average accuracy improves to 60% when considering the best-models per-participant. These findings indicate that consumer-grade EEG devices can detect meaningful cognitive variation during complex decision-making, and demonstrates potential for using this data for BCI.

INTRODUCTION

Electroencephalography (EEG) has emerged as a powerful tool in Brain-Computer Interface (BCI) research, with established capabilities in measuring cognitive workload [1] and differentiating task difficulties in ergonomics and human-computer interaction studies [2]. Recent advances in consumer-grade devices have expanded EEG's accessibility beyond clinical settings, enabling researchers across diverse fields to leverage this technology [3]. Yet, significant challenges remain in bridging the gap between clinical findings, consumer device validation, and practical applications [4], particularly in establishing reliable real-time signal processing methods for BCI applications [5]. Our research addresses these challenges by employing the MUSE 2, a wireless and portable EEG device previously validated for cognitive workload measurement [6,7], to explore its practical application in realistic BCI scenarios during chess games.

MATERIALS AND METHODS

We recruited 17 healthy participants (mean age 20.5 ± 2.03 years, 1 female) to play five 5-minute games against a computer opponent with adjustable difficulty (levels 1-8). Participants selected the difficulty level of the computer opponent based on their own skill level, and could adjust the difficulty by one after each game. Participants played on a local copy of the Lichess open source engine [8], using a 13-inch M2 MacBook Air equipped using a mouse. Data was collected from the Muse 2 device using the Mind Monitor [9] application, which applies a 60Hz notch filter and records delta (0.4-4hz), theta (4-7hz), alpha (8-12hz), beta (13-30hz), and gamma (30-80hz) waves at 10Hz [9, 10]. Connection to the Mind Monitor application was not perfect: whenever disconnection occurred all moves afterwards for that game were lost. On average, 3.6 games' worth of data were collected per participant (sd 1.15). Samples were filtered to exclude eye-blinks, jaw clenches, or invalid connection to the forehead (10% of samples) [9]. The final dataset contained EEG data from 1038 moves across 14 participants.

Move quality was evaluated using Stockfish 15 [11]. With the engine, we evaluated the game positions both before and after the move in centipawns. We converted these evaluations into win probabilities (W) using a logistic function: $W = \frac{1.0}{1.0 + 10^{\frac{-p}{400}}}$, where p is the centipawn position evaluation [12]. Move quality score was defined as the difference between the post-move and pre-move win probabilities, measuring deviation from theoretically optimal play. Move quality was normalized per-participant by transformation into a categorical variable based on median split, dividing the participants' moves into the Higher Quality and Lower Quality moves.

We focused our analysis on the MUSE device's prefrontal probes (AF7 and AF8) [9]. We calculated the mean and standard deviation within each frequency band per-probe, averaged these statistics across probes. We performed statistical modeling in R [13], fitting Linear Mixed-Effects models [14] using move quality as a fixed effect and with games nested within participants as random effects. Model checks were made for linearity, homogeneity of variance, outliers, and normality [15]. Effect sizes are reported as partial epsilon-squared (ε_p^2) [16]; Bonferroni correction is used across models. Machine Learning was performed in Python on the same data but with leave-one-out cross-validation at the participant level, and utilized Random Forest, Support Vector Machines, Quadratic Discriminant Analysis, Linear Discriminant Analysis, K-Nearest Neighbors, Multi-Layer Perceptron, and the Extreme Gradient Boosting classifier; ML analysis was done in the scikit-learn library [17].

RESULTS

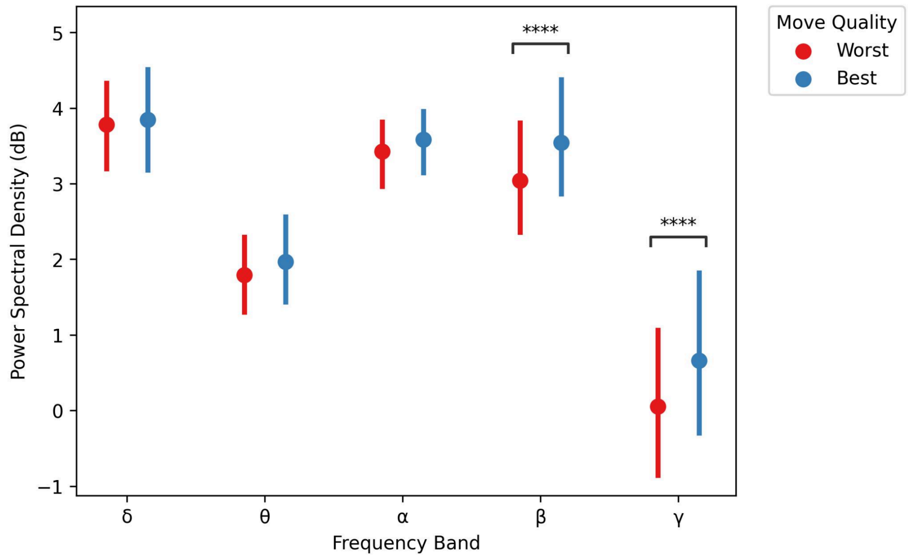

Figure 1: Power in EEG bands comparing lower and higher performing chess moves. Significant differences are observed in the beta and gamma bands.

Statistical analysis demonstrated that both beta ($F_{(1,1004.9)} = 39.36$, $p<0.001$, $\varepsilon_p^2=0.04$) and gamma ($F_{(1,1006.9)} = 38.32$, $p<0.001$, $\varepsilon_p^2=0.05$) brain wave activity increased significantly during higher quality chess moves. Across ML models, Random Forest models achieved only an average accuracy of 54% in distinguishing between move qualities across all participants; however, averages across best-models per-participant yielded 60%. High variability across participants and models indicates potential for meta-classification techniques, or potentially model-based cross-validation on data from a given participant to determine optimal classifier or meta-classifier with which to work.

CONCLUSION:

This study demonstrates that the Muse 2 device can detect meaningful variations in neural activity during a complex cognitive task, and showed consistent increases in beta and gamma band activity during higher quality moves. Although variable machine learning results highlight current limitations in translating these neural patterns into reliable real-time predictions, high variance across participants and models indicates potential for new classification techniques to be applied for future interfaces. These findings contribute to our understanding of the potential and constraints of consumer EEG devices in BCI applications.

REFERENCES

[1] Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomics Science, 4(1-2), 113-131. https://doi.org/10.1080/14639220210159717

[2] Zarjam, P., Epps, J., & Lovell, N. H. (2015). Beyond subjective self-rating: EEG signal classification of cognitive workload. IEEE Transactions on Autonomous Mental Development, 7(4), 301-310. https://doi.org/10.1109/TAMD.2015.2441960

[3] Sabio, J., Williams, N. S., McArthur, G. M., & Badcock, N. A. (2024). A scoping review on the use of consumer-grade EEG devices for research. PLOS ONE, 19(3), e0291186. https://doi.org/10.1371/journal.pone.0291186

[4] Värbu, K., Muhammad, N., & Muhammad, Y. (2022). Past, present, and future of EEG-based BCI applications. Sensors, 22(9), 3331. https://doi.org/10.3390/s22093331

[5] McFarland, D. J., & Wolpaw, J. R. (2017). EEG-based brain–computer interfaces. Current Opinion in Biomedical Engineering, 4, 194-200. https://doi.org/10.1016/j.cobme.2017.11.004

[6] Peining, P., Tan, G., & Wai, A. P. (2017, August). Evaluation of consumer-grade EEG headsets for BCI drone control. In Proceedings of the IRC Conference on Science, Engineering, and Technology.

[7] Bird, J. J., Manso, L. J., Ribeiro, E. P., Ekárt, A., & Faria, D. R. (2018). A study on mental state classification using EEG-based brain-machine interface. In 2018 International Conference on Intelligent Systems (IS) (pp. 795-800). https://doi.org/10.1109/IS.2018.8710576

[8] Lichess. (n.d.). Lichess Open Source Project. Retrieved January 18, 2024, from https://github.com/lichess-org/lila

[9] Mind Monitor. (n.d.). Mind Monitor. Retrieved January 18, 2024, from https://mind-monitor.com/

[10] Kawala-Sterniuk A, Podpora M, Pelc M, Blaszczyszyn M, Gorzelanczyk EJ, Martinek R, Ozana S. Comparison of Smoothing Filters in Analysis of EEG Data for the Medical Diagnostics Purposes. Sensors (Basel). 2020 Feb 2;20(3):807. doi: 10.3390/s20030807. PMID: 32024267; PMCID: PMC7038754.

[11] The Stockfish developers. (2024). Stockfish [Computer software]. https://stockfishchess.org

[12] Biswas, T., & Regan, K. (2015, December). Measuring level-k reasoning, satisficing, and human error in game-play data. In 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA) (pp. 941-947). IEEE.

[13] R Core Team (2024). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

[14] Kuznetsova A, Brockhoff PB, Christensen RHB (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." Journal of Statistical Software, *82*(13), 1-26.

<https://doi.org/10.18637/jss.v082.i13>.

[15] Lüdecke et al., (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. Journal of Open Source Software, 6(60), 3139. https://doi.org/10.21105/joss.03139

[16] Mordkoff, J. T. (2019). A simple method for removing bias from a popular measure of standardized effect size: Adjusted partial eta squared. Advances in Methods and Practices in Psychological Science, 2(3), 228-232. https://doi.org/10.1177/2515245919855053

[17] Ben-Shachar M, Lüdecke D, Makowski D (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. Journal of Open Source Software, 5(56), 2815. doi: 10.21105/joss.02815

[18] Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. In ECML PKDD Workshop: Languages for Data Mining and Machine Learning (pp. 108-122).