Accelerating P300-BCI Based Communication in ALS via On-Device Large Language Model Integration

Bartu Atabek[1], Saqer Alshehri[1], Akshay Bharath[2], Joshua Miller[2],
Terry Heiman-Patterson[2], Hasan Ayaz[1]

[1]School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, PA

[2]Department of Neurology, Lewis Katz School of Medicine, Temple University, Philadelphia, PA

Synopsis

This study evaluates a new artificial intelligence (AI) enhanced brain-computer interface (BCI) for individuals with amyotrophic lateral sclerosis (ALS). By integrating mobile EEG with offline and on-device large-language-model, the system aims to move beyond traditional character-by-character spelling. We compared this hybrid approach against conventional P300 spellers across individuals with ALS, age-matched controls, and healthy young adults in real-world settings. Participants performed calibration, copy-spelling, and open-dialogue tasks to assess information transfer rates, cognitive workload, and usability. Results demonstrate the technical feasibility of on-device LLM integration, highlighting significant keystroke reduction and its potential to facilitate fluid, conversational-paced interaction in clinical populations.

Background

ALS is a neurodegenerative disease characterized by progressive muscular paralysis reflecting degeneration of motor neurons in the primary motor cortex, corticospinal tracts, brainstem, and spinal cord. For individuals with advanced ALS, BCIs represent a critical communication pathway when motor function is lost (Wolpaw et al., 2018). The P300 event-related potential (ERP) paradigm has emerged as particularly suitable because it relies on preserved attentional mechanisms rather than sustained visual fixation or learned motor strategies (Allison et al., 2020; Verbaarschot et al., 2021; Vansteensel et al., 2016). Despite demonstrated feasibility, traditional P300 spellers require character-by-character selection, transforming sentences into cognitively demanding exercises that undermine conversational flow (Rezeika et al., 2018).

Recent work has shown that integrating LLMs with P300 spellers can reduce required selections through predictive text (Hong et al., 2024; Lebedev et al., 2025; Wang et al., 2025). However, existing implementations rely on cloud-based inference, raising privacy, latency, and network dependency concerns, while validation has occurred primarily with young healthy participants. Furthermore, these approaches utilize LLMs primarily as autocompletion engines, failing to leverage contextual reasoning and memory capabilities.

This study addresses these gaps by evaluating an augmented P300 BCI powered by on-device LLM across three cohorts, integrating context-aware prediction and response generation that learns from user input patterns while maintaining privacy through local processing.

Methods

This ongoing study targets thirty participants across three cohorts of ten individuals each. To date, sixteen have completed the protocol: six with ALS (M = 63.6, SD = 2.92), five age-matched controls (M = 62, SD = 7.90), and five young controls (M = 27.4, SD = 4,33). EEG data were acquired using the Mentalab Explore Pro headset with 8 channels, sampled at 250 Hz with bandpass filtering between 0.5 and 30 Hz and 60 Hz notch filtering.

The system employed a checkerboard-based P300 paradigm (Townsend et al., 2010) using a stepwise LDA classifier where randomized pattern flashes elicited attention-modulated ERPs for character selection. Each session began with calibration on twenty unique characters. Participants

then completed two tasks in counterbalanced order using conventional character-by-character spelling and an LLM-augmented system running locally deployed GPT-OSS-20B with hardware acceleration.

The copy-spelling task required typing "PLEASE REPOSITION PILLOW", representing common positioning and comfort needs communicated by individuals with ALS through assistive devices (Ball et al., 2004). The conversational task presented randomly selected open-ended questions from categories including daily life, food and drinks, entertainment, or travel, prompting responses averaging nine characters. In the conventional system, all tasks required individual character selections. In the LLM system, copy-spelling employed context-aware word prediction while conversational tasks utilized full response generation based on accumulated session context.

Performance metrics included online accuracy, total duration, time per character, typing speed, and information transfer rate (ITR). Following each task, participants completed the NASA Task Load Index (TLX).

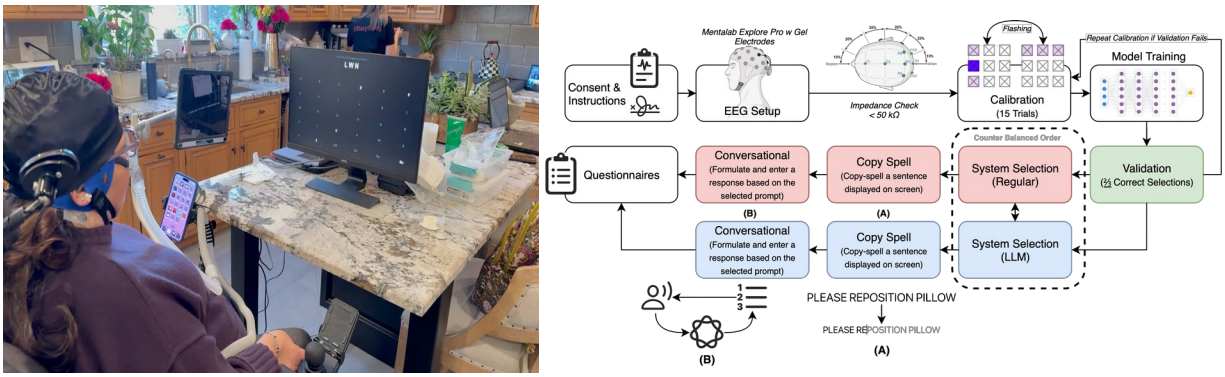

Figure 1. Left: At-home experimental setup. **Right**: Experimental Protocol

Results

During copy-spelling, the conventional system achieved 15.58 seconds per character and 20.50 bits per minute ITR. The LLM system recorded a longer per-character time of 22.77 seconds, attributable to the inference latency of the locally running language model. Despite this delay, the LLM system achieved a substantially higher ITR of 70.21 bits per minute. Selection accuracy remained comparable at 74.05% for the conventional and 82.41% for the LLM condition.

The conversational task revealed more pronounced differences. Conventional spelling required 15.83 seconds per character with 10 total selections on average. The LLM system reduced selections to 2 while achieving 10.53 seconds per character through full response generation.

TLX scores varied across groups. Individuals with ALS reported low workload for both systems (Regular: 15, LLM: 27), with the LLM introducing marginally higher perceived effort due to inference latency. Young controls showed the highest workload overall but a significant reduction with the LLM (53.7 vs. 39.7), while age-matched controls showed the opposite pattern (41.8 vs. 50.3).

Discussion

This study demonstrates the technical feasibility of integrating offline large language models with portable wireless P300 brain-computer interfaces while providing critical validation in individuals with ALS. The results reveal an important tradeoff inherent to language model augmentation: inference latency penalties are offset by substantial reductions in required character selections, yielding net improvements in communication efficiency. Although the LLM system introduced per-character delays due to on-device inference, this cost was recouped at the message level, where fewer selections were needed overall and word-level prediction simultaneously reduced spelling errors through autocorrection. This finding challenges the assumption that raw information transfer

rate alone determines practical BCI utility, suggesting that keystroke economy and cognitive sustainability warrant equal consideration in system design.

The comparable performance observed across participant cohorts, including individuals with ALS, age-matched controls, and young healthy adults, addresses a significant gap in existing literature where validation has occurred primarily in laboratory settings with young participants (Hong et al., 2024; Lebedev et al., 2025). That individuals with ALS achieved similar outcomes to control groups supports the viability of this approach for the target clinical population. The on-device processing architecture addresses privacy concerns inherent to cloud-based language model implementations while enabling operation independent of network connectivity, both critical factors for assistive technology adoption (Tuskey, 2025).

Subjective usability assessments revealed modest differentiation between systems, likely attributable to limited prior BCI experience among participants. This novelty effect suggests that as brain-computer interfaces become more prevalent, users may develop more discriminating preferences. Future work should examine longitudinal adoption patterns, optimize language model architectures for reduced inference latency, and evaluate voice synthesis personalization to fully realize naturalistic communication goals.

References

Allison, B.Z., Kübler, A., and Jin, J. (2020). 30+ years of P300 brain–computer interfaces. Psychophysiology 57(7), e13569.

Ball, L.J., Beukelman, D.R., and Pattee, G.L. (2004). Acceptance of augmentative and alternative communication technology by persons with amyotrophic lateral sclerosis. Augmentative and Alternative Communication 20(2), 113–122.

Hong, J., Wang, W., and Najafizadeh, L. (2024). Chatbci: A p300 speller bci leveraging large language models for improved sentence composition in realistic scenarios. arXiv preprint arXiv:2411.15395.

Lebedev, M.A., Makarova, A.V., Kleeva, D.F., and Maysuradze, A.I. (2025). Connecting a P300 speller to a large language model. bioRxiv, 2025.2011.2006.686984.

Rezeika, A., Benda, M., Stawicki, P., Gembler, F., Saboor, A., and Volosyak, I. (2018). Brain–computer interface spellers: A review. Brain sciences 8(4), 57.

Townsend, G., LaPallo, B.K., Boulay, C.B., Krusienski, D.J., Frye, G., Hauser, C., et al. (2010). A novel P300-based brain–computer interface stimulus presentation paradigm: moving beyond rows and columns. Clinical neurophysiology 121(7), 1109–1120.

Tuskey, J. (2025). The Psychological Risks of BCI-LLM Cognitive "Enhancement". Neuroethics 18(3), 50.

Vansteensel, M.J., Pels, E.G., Bleichner, M.G., Branco, M.P., Denison, T., Freudenburg, Z.V., et al. (2016). Fully implanted brain–computer interface in a locked-in patient with ALS. New England Journal of Medicine 375(21), 2060–2066.

Verbaarschot, C., Tump, D., Lutu, A., Borhanazad, M., Thielen, J., van den Broek, P., et al. (2021). A visual brain-computer interface as communication aid for patients with amyotrophic lateral sclerosis. Clinical Neurophysiology 132(10), 2404–2415.

Wang, J., Zhong, Y., Huang, C., and Yao, L. (2025). MindChat: Enhancing BCI Spelling with Large Language Models in Realistic Scenarios. arXiv preprint arXiv:2507.21435.

Wolpaw, J.R., Bedlack, R.S., Reda, D.J., Ringer, R.J., Banks, P.G., Vaughan, T.M., et al. (2018). Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis. Neurology 91(3), e258–e267.