DOI: 10.3217/978-3-99161-050-2-152

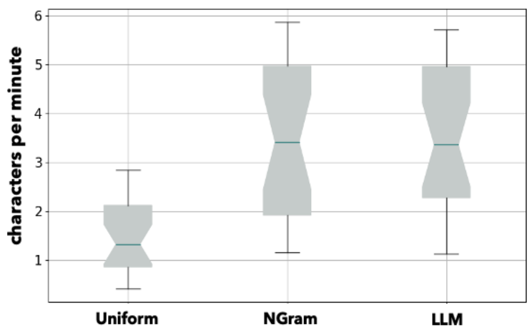
Simulated online typing performance in a cBCI using different language models

T. Memmott[1*], D. Gaines[2], M. Lawhead[1], D. Klee[1], B. Oken[1] & K. Vertanen[2] on behalf of the Consortium for Accessible Multimodal Brain-Body Interfaces (CAMBI)

[1]Oregon Health & Science University, Portland, OR, USA; [2]Michigan Technological University, Houghton, MI, USA
*707 SW Gaines St #1290, Portland, OR, 97239. E-mail: memmott@ohsu.edu

***Introduction*:** Communication Brain-Computer Interfaces (cBCIs) represent a crucial technological advancement for individuals with severe motor disabilities as they offer a direct pathway to express their thoughts and needs without physical movement. These systems commonly leverage the P300 ERP, a distinct neural response approximately 300–500ms after a novel stimulus. Language modeling presents a promising approach to enhancing the performance and usability of cBCIs. However, integrating language models with cBCI systems presents unique challenges, including balancing model complexity with real-time processing requirements and optimizing system performance parameters. This study utilizes simulations of online cBCI data to investigate the impact of different language models on typing rate and accuracy.

Figure 1: Simulated typing rate by language model.



***Methods, Materials, and Results*:** Twenty-four participants (22-49 years of age) were recruited for a cBCI study on mental effort conducted at Oregon Health & Science University (IRB #27415). The experiment consisted of participants writing using two interfaces using BciPy **[1]**: Rapid Serial Visual Presentation (RSVP) and Matrix presentation. Each presentation included a calibration phase followed by three tasks where participants spelled words. EEG data were collected at 300 Hz using the DSI-VR-300 (Wearable Sensing). Date were filtered using the default settings in BciPy: second order, 1–20Hz bandpass filter, 60Hz notch filter, and downsampling of two. The default signal model and settings (PcaRdaKde) were also used. Three conditions were compared: (I) UNIFORM: a baseline language model that assumes an equal probability for all characters; (II) NGRAM: a 12-gram statistical language model trained on conversational sentences; and (III) LLM: a 350M parameter version of the OPT large language model (LLM) that was fine-tuned on conversational sentences **[2]**. The effectiveness of each language model was evaluated using twelve phrases of varying complexity (six easy and six hard) from a previous study **[3]**. The average phrase length was 13 ± 4.7 characters. The copy phrase task implemented several thresholds for character selection: a maximum selection number of two times the phrase length, a maximum of eight inquiries per letter selection, and a decision threshold of 0.80. The backspace action was always presented and we used a starting backspace probability of 0.03. For each phrase and language model combination, the simulator performed 25 independent iterations using randomly sampled trial data from the earlier collected copy spelling tasks. The results were averaged across phrases and simulation runs for each participant using the three tested language models. Simulated typing rate was measured in characters per minute. The typing accuracy was calculated using the character distance between the final typed sentence and the reference phrase. The means between language models were compared using two-tailed t-tests with 10,000 permutations with statistical significance set at $p < 0.05$. As demonstrated in Figure 1, the NGRAM (M=3.5, SD=1.5) and LLM (M=3.5, SD=1.5) conditions significantly improved typing rate ($t(23) = -5.7$, $p < 0.0001$) when compared to UNIFORM (M=1.5, SD=0.8). However, there were no differences in simulated typing accuracy when compared to UNIFORM ($p > 0.5$) with the average accuracy being 83% for all conditions. There were no significant differences between the NGRAM and LLM conditions.

***Conclusion*:** These results show that integrating language modeling into a cBCI may benefit end-users by significantly increasing communication rate without compromising accuracy. The uniform condition, which had comparable accuracy to the more complex language models, did so by acquiring evidence more gradually. Further analyses of typed text across simulations and difficulty of phrases may yield more insights into language model performance in diverse contexts. These simulations demonstrated that language model integration could lead to a more efficient communication system for individuals dependent on this technology.

Acknowledgments and Disclosures: This research was funded by the National Institutes of Health (DC009834) and the National Science Foundation (IIS-1750193). The authors thank CAMBI and all associated researchers. The authors report no conflicts of interest.

***References*:**

[1] Memmott, T., Koçanaoğulları, A., Lawhead, M., Klee, D., Dudy, S., Fried-Oken, M., & Oken, B. (2021). BciPy: brain-computer interface software in Python. Brain- Computer Interfaces, 1-18.
[2] Gaines, D. and Vertanen, K. 2025. Adapting large language models for character-based augmentative and alternative communication. arXiv preprint. *arXiv:2501.10582*
[3] Gaines, D., Kristensson, P.O., and Vertanen, K. 2021. Enhancing the Composition Task in Text Entry Studies: Eliciting Difficult Text and Improving Error Rate Calculation. In Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '21). doi.org/10.1145/3411764.3445199