Predicting BCI Performance with the Detectability Index

B. O. Mainsah[1], L. M. Collins[1*], C. S. Throckmorton[1]

[1]Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA

*Duke University, 130 Hudson Hall, Durham, NC, USA. E-mail: leslie.collins@duke.edu

Introduction: Predicting a user's performance with a BCI is important for several reasons. For example, the predictions can be used to select parameters for the current stimulus paradigm (e.g. the amount of data to collect), to estimate the impact of switching to a new stimulus paradigm, or to avoid the frustrating scenario of a user attempting to use the system despite the low likelihood of success. However, prediction can be challenging since previous performance may not always be indicative of current performance. A previous method to predict P300 speller performance was limited to the row-column paradigm (RCP) and static data collection [1]. With the increased interest in other stimulus paradigms and dynamic stopping algorithms, a generalized method is desirable. We have developed a Bayesian-based method for predicting BCI performance that is independent of stimulus paradigm and accounts for dynamic data collection. The prediction method relies on data that is already available from system calibration and accounts for possible changes in user performance after calibration.

Material, Methods and Results: The focus of this work is on P300 spellers; however, this technique could be applied to any probabilistic decision-based BCI in which 1-of-M choices is selected. A user's performance level with the P300 speller depends on how well the system's classifier can distinguish between target and non-target electroencephalography (EEG) responses. Under a Gaussian assumption, this performance level can be quantified by the parameters of the two class conditional classifier likelihoods via a distance measure called the *detectability index,* d [2]. The classifier likelihoods, and therefore d, can be estimated from the BCI calibration data to initially pre-asses BCI performance, and changes in user performance level accounted for by different d values. Our new method to predict performance for a Bayesian dynamic stopping (DS) algorithm is independent of stimulus paradigm and relies on analytical calculations or Monte Carlo (MC) simulations, parameterized by d.

Given a stimulus paradigm, detectability index and data collection limit, performance estimates can be derived based on the cumulative likelihood ratio for each character. For stimulus paradigms with (i) a two-stage character selection process, (ii) equally-sized and pairwise disjoint flash groups at each stage, and (iii) fixed character-to-flash group assignments during a selection process (e.g. RCP), we derive tractable analytical solutions for approximating accuracy, and the lower bound of the expected stopping time. For other paradigms that satisfy only the last property, we propose an alternate analytic solution to approximate accuracy. Alternatively, performance estimates for any paradigm can be obtained from MC simulations of P300 spelling runs. The proposed methods were initially verified with MC simulations using synthetic and EEG data, and validated with results from several online studies. Fig. 1 shows that the user performances from an online ALS participant study [3] follow the trends predicted by offline simulations, according to their respective detectability indices.

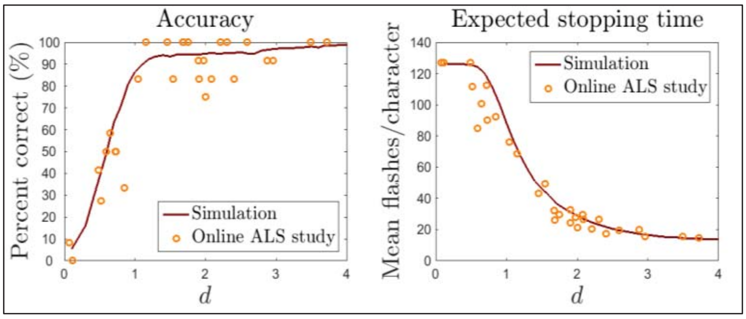

Figure 1. *Online vs. predicted P300 speller performances with Bayesian dynamic stopping using the checkerboard paradigm, based on a user's detectability index,* d.

Discussion: We have shown that the detectability index metric allows us to predict the performance of the Bayesian DS algorithm for a given stimulus paradigm. This provides a convenient way to potentially compare the performance of stimulus paradigms across a range of performance levels prior to online testing. Future work includes using objective performance functions based on the detectability index to develop custom-designed stimulus paradigms that facilitate the distinction between characters via their respective flash patterns, and take into account physiological limitations (e.g. refractory effects) to improve P300 speller performance.

Significance: The proposed method provides a useful tool to pre-asses BCI performance with the Bayesian DS algorithm with a given stimulus paradigm without extensive online testing. It can also be used to determine a suitable data collection limit to achieve a certain accuracy level given a user's performance level.

Acknowledgements: This research was funded by the NIH under grant number R33 DC010470.

References

[1] K. Colwell *et al.* Projected accuracy metric for the P300 Speller. *IEEE Trans Neural Syst Rehabil Eng,* 22(5): 921-925, 2014.
[2] A. J. Simpson and M. J. Fitter. What is the best index of detectability? *Psychological Bulletin,* 80(6): 481-488, 1973.
[3] B. O. Mainsah *et al.* Increasing BCI communication rates with dynamic stopping towards more practical use: an ALS study. *J Neural Eng,* 12(1): 016013, 2015.