Predicting Response Inhibition: A Deep Learning Approach Using Pre-Response Single-Trial EEG Data

Anna Grabowska (anna.belowska@doctoral.uj.edu.pl)
Doctoral School in the Social Sciences, Rynek Główny 34, 31-010 Krakow, Poland; Centre for Cognitive Science, Ingardena 3, 30-060 Krakow, Poland

Paweł Basoń (pawel.bason@doctoral.uj.edu.pl)
Doctoral School in the Social Sciences, Rynek Główny 34, 31-010 Krakow, Poland; Centre for Cognitive Science, Ingardena 3, 30-060 Krakow, Poland

Magdalena Senderecka (magdalena.senderecka@uj.edu.pl)
Centre for Cognitive Science, Ingardena 3, 30-060 Krakow, Poland

Abstract

Inhibitory control is a key component of executive functions. While it primarily depends on the extended motor network, early sensory processing of stimuli also plays a critical role in inhibition. This study examined whether deep neural networks could predict stop-signal task performance from early stop-related EEG signals in 225 participants, and whether including early go-related signals would enhance prediction accuracy. The best-performing model combined both go- and stop-related EEG data, revealing that successful inhibition was associated with reduced sensory processing of go stimuli and enhanced perception of stop signals. These results underscore the dynamic interplay between go and stop-signal processing and represent the first successful prediction of inhibition outcomes using non-motor EEG signals.

Introduction

Response inhibition, the ability to suppress automatic or habitual actions in favor of goal-directed behavior, is crucial for self-regulation and supports a wide range of cognitive and behavioral functions (Dillon & Pizzagalli, 2007; Hofmann, Schmeichel, & Baddeley, 2012). Impairments in this ability are associated with conditions such as ADHD or addictive behaviors (Chambers et al., 2009; Nigg et al., 2007).

The stop-signal task (SST) is widely used to study response inhibition. It requires rapid responses to frequent go stimuli and inhibition to infrequent stop signals, presented shortly after the go stimuli (Logan & Cowan, 1984). Using SST, electroencephalography (EEG) research has identified key event-related potential components associated with inhibitory control: the N2, associated with conflict monitoring and prominent in failed stops; the P3, linked to successful inhibition and adaptive control (Huster et al., 2013; Wessel & Aron, 2015); and the N1, enhanced in successful stops, highlighting the role of early perceptual processing (Bekker et al., 2005; Skippen et al., 2020). While brain–behavior correlations provide valuable insights, they cannot determine whether observed neural activity specifically reflects inhibitory control or general processes occurring in the same time window (Gholamipourbarogh et al., 2023). Further, given the often non-linear nature of brain–behavior relationships (Reuter et al., 2019), such studies may miss key dynamics. Recent machine learning (ML) approaches address these limitations by predicting stopping behavior directly from neural data. For example, Rueda-Delgado et al. (2021) found that features near the N2 and P3 time-windows predicted behavioral measures of inhibitory control, while Gholamipourbarogh et al. (2023) successfully used deep learning to distinguished response execution from inhibition in a go/no-go task.

Most EEG-based ML studies on inhibitory control target late post-stimulus activity, often within the P3 window. However, in speeded tasks like the SST, this activity may reflect post-response monitoring rather than inhibition itself (Gehring et al., 1993). Only Bode and Stahl (2014) have predicted errors before response execution, detecting them after motor initiation in a flanker task. Additionally, White et al. (2014) showed that faster go stimulus processing enhances stopping-related activity, suggesting a close interaction between response initiation and inhibition.

This study addresses both challenges by using deep neural networks to predict stopping behavior from EEG signals related to go, stop, or both types of stimuli, excluding post-response activity. Based on White et al. (2014), we hypothesize that the model using both go- and stop-related signals will perform best, as effective inhibition likely relies on the interaction between these processes.

Materials and Methods

A total of 225 volunteers (113 F, 1 non-binary), aged 18–39 (M = 23.64, SD = 4.18), with normal or corrected-to-normal vision, were recruited from the general population. Participants received verbal and written information about the study's purpose and procedures. The protocol was approved by the local Research Ethics Committee, all participants provided written informed consent and were compensated monetarily for their time. During EEG recording, they performed a stop-signal task with stop-signal delays adaptively adjusted using a standard tracking procedure (Verbruggen et al., 2019). The stop-signal delay ranged from 100 to 400 ms in 50 ms intervals.

The EEG signal was continuously recorded at 256 Hz using 64 Ag/AgCl electrodes with preamplifiers (BioSemi Active-Two system) and referenced online to CMS-DRL ground. Off-line, the signal was re-referenced to the average of the two mastoid electrodes, band-pass filtered between 0.1 Hz and 40 Hz using a Butterworth filter, and notch filtered at 50 Hz. The

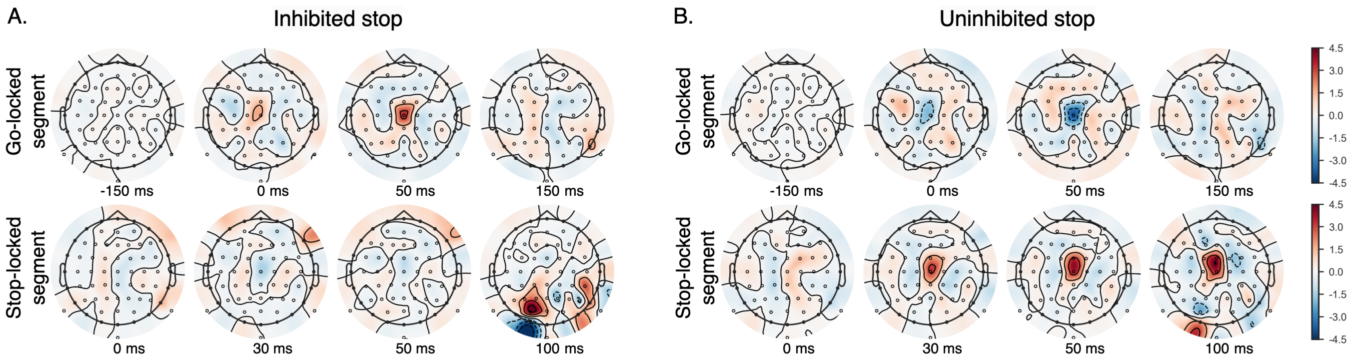

Figure 1: Saliency maps illustrating the relevance of spatio-temporal signals for classifying inhibited and uninhibited stop trials. The topographical plots display saliency maps at specific times during the segments.

data were segmented into 2-second epochs around go stimuli (-900 to 1100 ms) and aligned to the pre-stimulus baseline from -900 to -800 ms. Eye blinks were removed using the (Gratton, Coles, & Donchin, 1983) algorithm, and noisy epochs were rejected through an automatic procedure using AutoReject Python package (Jas et al., 2017).

Epochs were subsequently divided into go-locked and stop-locked segments. To ensure that the input to the model captured sufficient stimulus-related information while avoiding contamination from stop-related activity in go segments and response-related activity in stop segments, we excluded trials in which the stop-signal delay was shorter than 150 ms or the stop-response interval was shorter than 100 ms. Go-locked signals were then extracted from -200 to 150 ms around the go stimulus, and stop-locked signals from 0 to 100 ms around the stop. After selection, participants had in average 31.40 inhibited trials (SD = 14.80) and 28.87 uninhibited trials (SD = 7.97). We used the EEGNet architecture to classify inhibited and uninhibited trials in both channel and source space. EEGNet is a well-established deep learning model for EEG-based brain-state decoding (Lawhern et al., 2018). The following parameters were used: F1 = 16, D = 2, F2 = 32, and temporal kernel length = 64 (full specifications in Lawhern et al., 2018). We created three models: the go model, based on go signals; the stop model, based on stop signals, and the go-stop model, based on concatenated go and stop signals. Model performance was evaluated using subject-wise 5-fold cross-validation. In each fold, five participants from the training set were randomly selected for early stopping validation. The model with the lowest cross-entropy on the validation set was saved after 50 epochs. Performance metrics: accuracy, AUC (Junge & Dettori, 2018), recall, and specificity, were calculated on the testing sets. Final results were obtained by averaging the metrics across all five test folds.

Results

From the three created models, only the stop and go-stop models yielded mean cross-validated accuracy and ROC values that exceeded chance level. The stop model achieved moderate performance, with AUC values at 55.56%. However, its higher recall compared to specificity suggests a slight bias toward detecting inhibited trials. In contrast, the go-stop model demonstrated more balanced performance, with both recall and specificity closely aligned, indicating better calibration. It also outperformed the other models in overall accuracy and AUC (57.72%). Detailed results for all models are presented in Table 1. We used a saliency map approach (Ancona et al., 2018) to identify key time points and channels contributing most to the model's predictions, thereby highlighting neural activity relevant to inhibitory control. The average saliency maps from the test folds are shown in Figure 1.

Discussion

The go-stop model, which integrates both go and stop signals, outperformed the individual go and stop models. These results confirm the interaction between go process initiation and inhibition in the stop-signal task, supporting the hypothesis proposed by White et al. (2014). As expected, go-locked and stop-locked activity within the first 100 ms time window were the most discriminative between conditions, with reduced go and enhanced stop activity in inhibited trials. Thus, our results indicate that successful inhibition is associated with blunted sensory processing of go stimuli and enhanced sensory processing of stop signals. This study demonstrates the successful prediction of inhibition outcomes from EEG data that are not directly related to motor processes. Using a deep learning approach, we show that early perceptual processing of both go and stop stimuli is predictive of successful inhibition.

Table 1: Detailed results of the inhibited and uninhibited trials classification task for the go, stop, and go-stop models.

Model	Accuracy		AUC		Recall		Specificity	
	M	SD	M	SD	M	SD	M	SD
go	50.85	1.34	50.75	1.29	52.46	4.97	49.04	4.23
stop	55.65	1.18	55.56	1.13	58.01	4.17	53.12	3.12
go-stop	57.76	0.86	57.72	1.05	57.47	4.76	57.96	6.42

Acknowledgments

The study was supported by Sonata Bis grant 2020/38/E/HS6/00490 from the National Science Centre of Poland.

References

Ancona, M., Ceolini, E., Öztireli, C., & Gross, M. (2018, March). *Towards better understanding of gradient-based attribution methods for Deep Neural Networks.* arXiv. (arXiv:1711.06104 [cs]) doi: 10.48550/arXiv.1711.06104

Bekker, E. M., Kenemans, J. L., Hoeksma, M. R., Talsma, D., & Verbaten, M. N. (2005, February). The pure electrophysiology of stopping. *International Journal of Psychophysiology*, *55*(2), 191–198. doi: 10.1016/j.ijpsycho.2004.07.005

Bode, S., & Stahl, J. (2014, December). Predicting errors from patterns of event-related potentials preceding an overt response. *Biological Psychology*, *103*, 357–369. doi: 10.1016/j.biopsycho.2014.10.002

Chambers, C. D., Garavan, H., & Bellgrove, M. A. (2009, May). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neuroscience and Biobehavioral Reviews*, *33*(5), 631–646. doi: 10.1016/j.neubiorev.2008.08.016

Dillon, D. G., & Pizzagalli, D. A. (2007, December). Inhibition of Action, Thought, and Emotion: A Selective Neurobiological Review. *Applied & Preventive Psychology: Journal of the American Association of Applied and Preventive Psychology*, *12*(3), 99–114. doi: 10.1016/j.appsy.2007.09.004

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993, November). A Neural System for Error Detection and Compensation. *Psychological Science*, *4*(6), 385–390. doi: 10.1111/j.1467-9280.1993.tb00586.x

Gholamipourbarogh, N., Vahid, A., Mückschel, M., & Beste, C. (2023). Deep learning on independent spatial EEG activity patterns delineates time windows relevant for response inhibition. *Psychophysiology*, *60*(10), e14328. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/psyp.14328) doi: 10.1111/psyp.14328

Gratton, G., Coles, M. G. H., & Donchin, E. (1983, April). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. doi: 10.1016/0013-4694(83)90135-9

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences*, *16*(3), 174–180. (Place: Netherlands Publisher: Elsevier Science) doi: 10.1016/j.tics.2012.01.006

Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M., & Herrmann, C. S. (2013, March). Electroencephalography of response inhibition tasks: Functional networks and cognitive contributions. *International Journal of Psychophysiology*, *87*(3), 217–233. doi: 10.1016/j.ijpsycho.2012.08.001

Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F., & Gramfort, A. (2017, October). Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, *159*, 417–429. doi: 10.1016/j.neuroimage.2017.06.030

Junge, M. R. J., & Dettori, J. R. (2018, June). ROC Solid: Receiver Operator Characteristic (ROC) Curves as a Foundation for Better Diagnostic Tests. *Global Spine Journal*, *8*(4), 424–429. doi: 10.1177/2192568218778294

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018, July). EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *15*(5), 056013. (Publisher: IOP Publishing) doi: 10.1088/1741-2552/aace8c

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*(3), 295–327. (Place: US Publisher: American Psychological Association) doi: 10.1037/0033-295X.91.3.295

Nigg, J. T., Carr, L., Martel, M., & Henderson, J. M. (2007). Concepts of inhibition and developmental psychopathology. In *Inhibition in cognition* (pp. 259–277). Washington, DC, US: American Psychological Association. doi: 10.1037/11587-013

Reuter, E.-M., Vieluf, S., Koutsandreou, F., Hübner, L., Budde, H., Godde, B., & Voelcker-Rehage, C. (2019, January). A Non-linear Relationship Between Selective Attention and Associated ERP Markers Across the Lifespan. *Frontiers in Psychology*, *10*. (Publisher: Frontiers) doi: 10.3389/fpsyg.2019.00030

Rueda-Delgado, L. M., O'Halloran, L., Enz, N., Ruddy, K. L., Kiiski, H., Bennett, M., ... Whelan, R. (2021, May). Brain event-related potentials predict individual differences in inhibitory control. *International Journal of Psychophysiology*, *163*, 22–34. doi: 10.1016/j.ijpsycho.2019.03.013

Skippen, P., Fulham, W. R., Michie, P. T., Matzke, D., Heathcote, A., & Karayanidis, F. (2020, October). Reconsidering electrophysiological markers of response inhibition in light of trigger failures in the stop-signal task. *Psychophysiology*, *57*(10), e13619. doi: 10.1111/psyp.13619

Verbruggen, F., Aron, A. R., Band, G. P., Beste, C., Bissett, P. G., Brockett, A. T., ... Boehler, C. N. (2019, April). A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *eLife*, *8*, e46323. doi: 10.7554/eLife.46323

Wessel, J. R., & Aron, A. R. (2015, April). It's not too late: the onset of the frontocentral P3 indexes successful response inhibition in the stop-signal paradigm. *Psychophysiology*, *52*(4), 472–480. doi: 10.1111/psyp.12374

White, C. N., Congdon, E., Mumford, J. A., Karlsgodt, K. H., Sabb, F. W., Freimer, N. B., ... Poldrack, R. A. (2014, August). Decomposing Decision Components in the Stop-signal Task: A Model-based Approach to Individual Differences in Inhibitory Control. *Journal of Cognitive Neuroscience*, *26*(8), 1601–1614. doi: 10.1162/jocn_a_00567