Brain event-related potentials predict individual differences in inhibitory control

Rueda-Delgado[1], L.M., O'Halloran[1], L., Enz[1], N., Ruddy[1], K.L., Kiiski[1], H., Bennett[1], M., Farina[1], F., Jollans[1], L., Vahey[1], N., Whelan*[1,2], R.

[1]School of Psychology, Trinity College Dublin, Dublin 2, Ireland

[2]Global Brain Health Institute, Trinity College Dublin, Dublin 2, Ireland

*Corresponding author

• E-mail address: Robert.whelan@tcd.ie

• Full postal address: Rm 0.80 Lloyd building, Trinity College Dublin, Dublin 2, Ireland

https://doi.org/10.1016/j.ijpsycho.2019.03.013

Abstract

Stop-signal reaction time (SSRT), the time needed to cancel an already-initiated motor response, quantifies individual differences in inhibitory control. Electrophysiological correlates of SSRT have primarily focused on late event-related potential (ERP) components over midline scalp regions from successfully inhibited stop trials. SSRT is robustly associated with the P300, there is mixed evidence for N200 involvement, and there is little information on the role of early ERP components. Here, machine learning was first used to interrogate ERPs during both successful and failed stop trials from 64 scalp electrodes at 4 ms resolution (n=148). The most predictive model included data from both successful and failed stop trials, with a cross-validated Pearson's r of 0.32 between measured and predicted SSRT significantly higher than null models. From successful stop trials, spatiotemporal features overlapping the N200 in right frontal areas and the P300 in frontocentral areas predicted SSRT, as did early ERP activity (<200 ms). As a demonstration of the reproducibility of these findings, the application of this model to a separate dataset of 97 participants was also significant (r = 0.29). These results show that ERPs during failed stops are relevant to SSRT, and that both early and late ERP activity contribute to individual differences in SSRT. Notably, the right lateralized N200, which predicted SSRT here, is not often observed in neurotypical adults. Both the ascending slope and peak of the P300 component predicted SSRT. These results were replicable, both within the training sample and when applied to ERPs from a separate dataset.

Highlights

- A right lateralized N200 and the P300 during successful stop trials predicted individual differences in inhibition
- Data were internally and externally validated on out-of-sample data
- Application of machine learning methods to EEG data may improve reproducibility

1. Introduction

The ability to inhibit unwanted behaviors or to quickly cancel an already-initiated response relies on effective and rapid inhibitory control in the brain. Deficits in this system are characteristic of a wide range of psychological disorders, such as attention deficit hyperactivity disorder (ADHD), and a variety of addictive behaviors. The stop-signal task (SST) can assay inhibitory control by requiring participants to respond as quickly as possible to frequent 'Go' cues but to inhibit an already-initiated motor response following unexpected and infrequent 'Stop' cues (Verbruggen et al., 2008). On trials with a stop stimulus, the 'horse-race model' posits a race between two separate processes that are each triggered by the Go and Stop signal (Band et al., 2003). If the stop process is completed before the go process, subjects will successfully inhibit their responses and vice-versa (Verbruggen et al., 2008). The stop-signal reaction time (SSRT) indexes the time needed to successfully inhibit a response during the SST (Congdon et al., 2012), and is a reliable measure of deficits in inhibitory control. The SSRT is a measure of a covert mental process and can be calculated by subtracting the average stop signal delay from the participant's go reaction time. Shorter SSRTs indicate better inhibitory control. In neurologically healthy adults, SSRTs are approximately 200-270 ms (Dimoska et al., 2006; Hoptman et al., 2018; Huster et al., 2013; van Boxtel et al., 2001; Wessel and Aron, 2015; Wessel et al., 2016), but longer in adults with ADHD (Dawe et al., 2004; Lijffijt et al., 2005) and in individuals with addictions (Luijten et al., 2011).

The amplitude and latency of event-related potential (ERP) components, obtained using electroencephalography (EEG), are associated with performance during the SST (e.g., Kenemans, 2015). Two ERP components, the P300 and N200, are predominantly associated with response inhibition. A frontocentral N200, with a negative component peaking around 200-250 ms, is typically observed for failed relative to successful stops (Galdo-Alvarez et al., 2016; Greenhouse and Wessel, 2013). In ERPs time-locked to the erroneous response, an error-related negativity (ERN/Ne)/error positivity (Pe) complex for failed stops is often found. A right lateralized N200 has been observed during successful stops in adolescents with ADHD versus controls (Liotti et al., 2010), but is not usually observed in neurotypical adults (see Wessel & Aron, 2014 for a

discussion). Studies have consistently shown larger P300 amplitudes for successful vs. failed stop trials in healthy participants (Bekker et al., 2005; Greenhouse and Wessel, 2013; Kok et al., 2004; Lansbergen et al., 2007). This P300 has a frontocentral distribution (Wessel et al., 2014) and occurs approximately 150 ms after the N200. Although the N200 and P300 are both associated with inhibition during the SST, both components peak after, or just before, the SSRT. For this reason, some have argued that N200 and P300 do not directly reflect inhibition, but rather reflect conflict and evaluative processes, respectively (see Huster et al., 2013 for a detailed review). Earlier ERP components that are sometimes reported during stop trials include the N100 and P200 (Bekker et al., 2005; Dimoska et al., 2006; Hoptman et al., 2018; Kenemans, 2015; van Gaal et al., 2011).

In contrast to comparisons between successful and failed stop trials, findings with respect to ERP *correlates* of SSRT are more inconsistent. In visual SSTs, no significant correlations between a central N200 and SSRT have been reported (Anguera and Gazzaley et al., 2011; Hoptman et al., 2018), whereas van Gaal et al. (2011) found a positive correlation between these two variables. Using median splits to create two groups of fast and slow SSRTs, van Boxtel et al. (2001) reported higher N200 amplitude in groups with fast SSRT using a sample of 10 participants in a visual SST. P300 amplitude was negatively correlated with SSRT (Lansbergen et al., 2007; Smith et al., 2006), but van Gaal et al. (2011) found no significant correlation. Given that P300 peak latency is longer than the SSRT, Wessel and Aron (2015) correctly predicted that P300 *onset* latency would be a better correlate of individual differences in SSRT than P300 peak latency.

Mixed findings, such as those related to ERP correlates of SSRT, are common in cognitive neuroscience. One reason for this is low power. Acquiring brain data is expensive, as in functional magnetic resonance imaging (fMRI) and magnetoencephalography, and time-consuming, as in EEG. Therefore, many neuroscience studies have small sample sizes which, in combination with small effect sizes, increase the probability of false positive findings (Button et al., 2013; Eklund et al., 2016; Munafò et al., 2017). A second reason may be that data from EEG or fMRI are high

dimensional. In the case of EEG, there are often more than 64 channels, typically acquired at a sampling rate of over 256 Hz. In order to reduce the exposure to type I errors (i.e., false positives), data from high density EEG arrays are typically reduced in dimension, often by averaging over some time interval and/or selecting a subset of channels to define the ERP components. It is also possible that focusing on a subset of data, or correcting for multiple univariate tests, increases the probability of type II errors (i.e., false negatives). A multivariable approach, in which all data are used in a single model, will likely be more useful for predicting individual difference in inhibitory control (Jollans and Whelan, 2018). However, when the ratio of variables to participants increases, which is usually the case for EEG data, the model will inevitably fit to idiosyncrasies in the data (i.e., to noise) and will fail to generalize to unseen data (Whelan and Garavan, 2014): this is known as *overfitting*.

One approach that has promise for addressing both type I and type II errors, and for improving reproducibility by reducing overfitting, is machine learning. This approach is increasingly popular in cognitive neuroscience (e.g. Chung et al., 2018; Jollans and Whelan, 2018; Vu et al., 2018; Whelan et al., 2014; Woo et al., 2017) but is, relative to MRI, rarely used to interrogate ERPs (but cf. Kiiski et al., 2018; Stock et al., 2016; Vahid et al., 2018). In order to overcome the problem of fitting a model with many more variables than subjects, regression weights can be penalized (termed *regularization*), with the elastic net being one common regularization method (Zou and Hastie, 2005). Rather than relying on null-hypothesis significance testing, machine learning typically employs out-of-sample validation, in which models are tested against unseen data. Another benefit is the automatic selection of model hyperparameters by defining them from the data itself, which reduces the 'researcher degrees of freedom' (Baldwin, 2017; Simmons et al., 2011).

While machine learning has not been applied previously to SST data collected under EEG, a study using the Go/No Go task, which measures the blocking of response initiation rather than action cancellation, suggested that this approach holds promise for predicting the speed and

accuracy of inhibition. Stock et al. (2016) trained several machines to predict group membership (good accuracy/slow response vs. low accuracy/fast response). Dimensionality reduction was done in space (selecting only 16 channels) and time (selecting six time bins based on ERP components) to ensure that the number of features (input variables) did not exceed the number of participants. Classification accuracies on an external validation set showed that three of the four machines performed significantly better than chance, with areas under the curve of receiver operating characteristic of approximately 0.6 (0.5 represents classification at chance level). Contrary to expectations, ERP features from around 200 and 300 ms over lateral parietal and occipital electrodes, and not N200 and P300 components from midline electrodes, were most predictive of group membership.

Here, we applied machine learning to a large sample to predict SSRT from EEG data. Rather than performing dimension reduction on the data, we used the elastic net, which has benefits of performing feature selection (i.e., providing a sparse solution) and regularization in a single approach. Furthermore, we aimed to identify the *correlates* of SSRT performance, rather than a between-group classification (which reduces statistical power; Royston et al., 2006). Our purpose here was to identify processes associated with individual differences in response inhibition rather than successful or failed stopping *per se*. The ability to search for inhibitory control correlates over the entire ERP time-course should shed light on the relative importance of early versus later ERP components (e.g., N200, P300 onset and peak). Guided by previous literature, we first applied machine learning to successful stop trials. However, we also investigated the contribution of failed stop trials to inhibitory control, and finally we tested a model that included ERP activity from both successful and failed stop trials. To quantify reproducibility, we employed both internal and external validation, which is the gold standard for reproducible analysis.

2. Materials and Methods

2.1. Participants

Participants were pooled from four different studies conducted in University College Dublin and Trinity College Dublin for a total of 282 participants (mean = 35.03 years old, standard deviation = 14.72, range 16-69 years; 175 female). The exclusion criteria are listed per project in the Supplementary Material. The studies were approved by the University College Dublin School of Psychology ethics committee and the Trinity College Dublin School of Psychology ethics committee. Participants provided informed consent.

2.2. Task

Participants were seated in front of a cathode ray tube computer monitor with a screen resolution of 1024 x 768 pixels at a refresh rate of 75 Hz. The distance from the position of the chair to the monitor (screen size 32 x 24 cm) was standardized (screen to back of chair = 108 cm). Participants were asked to maintain their focus on the stimuli on the screen during the experiment.

Participants performed an adaptive event-related SST, which took approximately 9 min to complete. The task consisted of 135 Go trials interspersed with 45 Stop trials; with one randomized Stop trial appearing within four Go trials. The task was presented in 3 blocks of 60 trials. Each trial began with a central fixation cross for 1000 ms and the total duration of a trial was always 1000 ms. On every trial, participants were presented with arrows pointing either to the left or right, shown centrally on the screen for 750 ms. During Go trials, participants were required to make a single button-press response via an Xbox 360 game controller with their left or right index finger corresponding to the direction of the arrow as fast as possible. In stop trials, the Go stimulus was followed by an arrow pointing upwards (i.e., the Stop signal), which required participants to inhibit their motor responses. A tracking algorithm adjusted task difficulty by varying the stop-signal delay (SSD; the time interval between Go signal and Stop signal onsets). The aim was to produce 50% successful and 50% unsuccessful inhibition trials. The initial SSD

was 250 ms, but was adjusted according to a participant's performance, to between 50 ms and 450 ms. These limits were adjusted depending on task performance, making the SSD shorter (i.e., the task easier) after an unsuccessful stop trial, and the SSD longer (i.e., the task more difficult) after a successful stop trial. A moving average of go reaction times (RTs) began on the 10th trial. Participants were presented with a "Speed up!" prompt for 2 s if they failed to respond to 2 out of 5 Go trials, or if their last RT was longer than 1.5 times their average RT. If the participant responded to the Go stimulus before Stop stimulus presentation (i.e., responded during the SSD) then the SSD was adjusted downwards for subsequent trials.

The SSRT refers to the time taken to cancel a prepotent motor response after Stop stimulus presentation. According to the horse-race model (Logan and Cowan, 1984), the finish of the stop process can be estimated from a subject's distribution of RTs on Go trials. The left side of the distribution of the RTs on Go trials represents faster responses that increase the probability of failure to inhibit a response (i.e., the go process is more likely to win the race), whereas the right side represents slower responses that increase the probability of successful inhibition (i.e., the stop process is more likely to win the race; see Figure 1). The integration method of SSRT calculation was used (Logan and Cowan, 1984). That is, if a subject failed to inhibit on n% of Stop trials, the finishing time of the stop process was assumed to be approximately equal to the n^{th} percentile of the go RT distribution. The mean SSD was then subtracted from the n^{th} percentile of the go RT distribution, resulting in an estimate of SSRT. Participants with SSRT < 75 ms were excluded from the analysis.

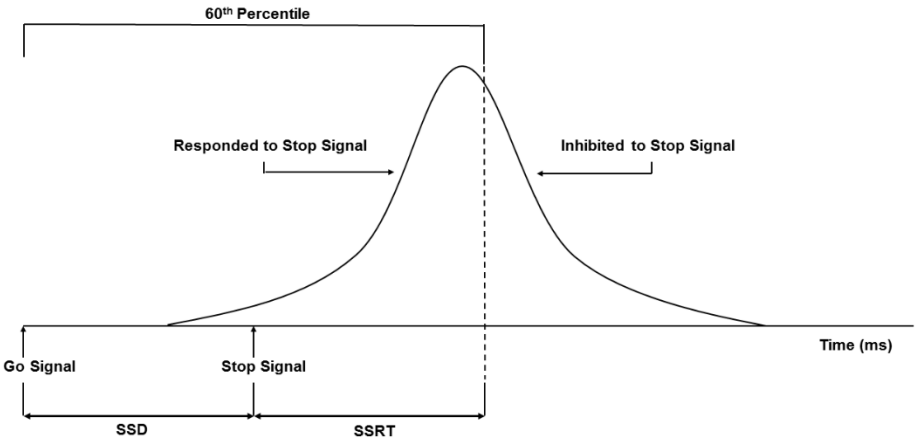

Figure 1. Illustration of the horse-race model of inhibitory control, where a participant's go RT distribution is superimposed onto a timeline for Stop trials. In this example, the participant has a successful stop rate of 40%, such that the upper 40% of the Go RT distribution corresponds to slower Go RTs that would produce successful stops, while the lower 60% of the Go RT distribution corresponds to faster Go RTs that would produce failed stops.

2.3. EEG recording and preprocessing

EEG data were recorded using the ActiveTwo Biosemi™ system in a soundproofed, darkened room from 70 electrodes (64 scalp electrodes), organized according to the 10–5 system (Oostenveld and Praamstra, 2001). Activity related to eye movement was recorded bilaterally from approximately 2 cm below the eye (vertical) and from the outer canthi (horizontal electrooculography). EEG data preprocessing was carried out using the EEGLAB toolbox (Delorme and Makeig, 2004; http://sccn.ucsd.edu/eeglab) in conjunction with the FASTER plug-in (Fully Automated Statistical Thresholding for EEG artefact Rejection; Nolan et al., 2010; http://sourceforge.net/projects/faster). The data were bandpass filtered between 0.1 and 95 Hz, notch filtered at 50 Hz and average referenced across all scalp electrodes. Data were subsequently epoched from 500 ms pre-stimulus to 2000 ms post-stimulus. FASTER identified and removed artefactual (i.e., non-neural) independent components, removed epochs containing

large artefacts (e.g., muscle twitches) and interpolated channels with poor signal quality. The remaining EEG data were then visually inspected to ensure good quality and that any remaining noisy data were removed. A total of 37 subjects across the whole data set were rejected from further analysis due to artefacts, missing data or SSRT <75 ms. Here, we report the results for the remaining 245 participants.

2.4. EEG ERP calculation

Three trial types were identified and epoched in the EEG: trials in which participants successfully responded after a Go cue (the epoch was defined -100 ms to 600 ms with respect to the Go cue), trials in which participants successfully inhibited after a Stop cue (Successful Stop) and trials in which participants failed to inhibit after a Stop cue (Failed Stop). In Stop trials, the epoch was defined with respect to the Stop cue, and averaged across epochs to obtain a Successful and Failed Stop ERP. In order to isolate activity associated specifically with stopping, the influence of the Go stimulus – which always appeared prior to each Stop cue – needed to be subtracted. According to the horse race model of inhibition in the SST, slower reaction times in go trials are related to successful inhibition (because the Stop process 'wins' the race), and faster reaction times are related to failed inhibition (because the Go process 'wins' the race; Logan and Cowan, 1984). For this reason and as it is standard in the SST literature (e.g., Kok et al., 2004), we defined "fast" and "slow" Go trials based on each participant's own median reaction time, and calculated a "fast" and "slow" Go ERP, which we then subtracted from the Failed and Successful Stop ERP, respectively (Palmwood et al., 2017; O'Halloran et al. 2019). As is typical for SST ERP analyses, baseline removal (-100 to 0 ms) was performed separately for Go and Stop ERPs prior to the subtraction of "fast" and "slow" Go trials from Stop trials (Kok et al., 2004; Palmwood et al. 2017). This procedure was applied per channel per participant.

2.5. Machine learning analysis

2.5.1. Data preparation

The dataset of 245 participants was divided into two groups: the internal validation set (n=148) and the external validation set (n=97). In order to balance the internal and external validation sets, we assigned individuals to each set randomly, but iterated this assignment until the female:male ratio (1.6:1, approximately) was similar across sets. 'Features' (the input to the machine learning model) consisted of ERPs averaged in 4-ms bins from 0 to 600 ms after the Stop cue at each of the 64 channels. We fitted models for the Successful and Failed Stop ERP separately, with 9,856 spatio-temporal features in each model. Additionally, we examined both Successful and Failed Stop trials (termed the 'Full model'), with 19,712 spatio-temporal features. We controlled for age and sex (Coxon et al., 2012; Hsieh and Lin, 2017, Rubia et al., 2013) by adding them as features. We also conducted a supplementary analysis in which age and sex were not included.

2.5.2. Analysis of the internal validation set

In order to choose the hyperparameters of the model (described below), we applied nested cross-validation on the internal validation set using 10 main folds (Figure 2). In each main fold, 90% of the data were used for training and 10% for testing. In each subfold (inner cross-validation), the data were z-scored and extreme values were replaced with a value of 3 (i.e., Winsorizing). The model was fitted with a range of hyperparameters using the elastic net (Friedman et al., 2010; Qian et al., 2013; Zou and Hastie, 2005), which is a regularization method for generalized linear models that includes l_1 regularization (i.e., lasso regularization - least absolute shrinkage and selection operator) and l_2 regularization (as in ridge regularization). Lasso regularization allows parameters to be 0, promoting parsimonious solutions; whereas ridge regularization allows parameters to be small but not to reach 0, avoiding overfitting (Murphy, 2012). In this model, the objective is to minimize the following equation:

$$\min\left[\frac{1}{2}\|Y - X\beta\|^2 + \lambda\alpha\|\beta\|_{l_1} + \frac{\lambda(1-\alpha)}{2}\|\beta\|_{l_2}^2\right]$$

where Y is the dependent variable (SSRT), X the input data with the ERPs and covariates, β the regression coefficients, λ the penalization for complexity and α is weighting parameter between ridge and lasso regression (Friedman et al., 2010). The complexity and weighting parameters (λ and α, respectively) are not known *a priori*. Therefore, a range of values were explored: 15 linearly-spaced values of both hyperparameters in the range of 0.01 to 10 and all their possible combinations (i.e., a search grid of 225 parameter-pair values). The prediction accuracy of each parameter combination was assessed using the mean squared error. The parameter combination that yielded the lowest error was selected per subfold. The mode of α and the median of λ across subfolds were selected as parameters per main fold. These optimal parameters from the nested cross validation were used to fit a model using the training set of the main fold (outer cross-validation). The prediction of the model on the test set of each main fold was saved and pooled across main folds.

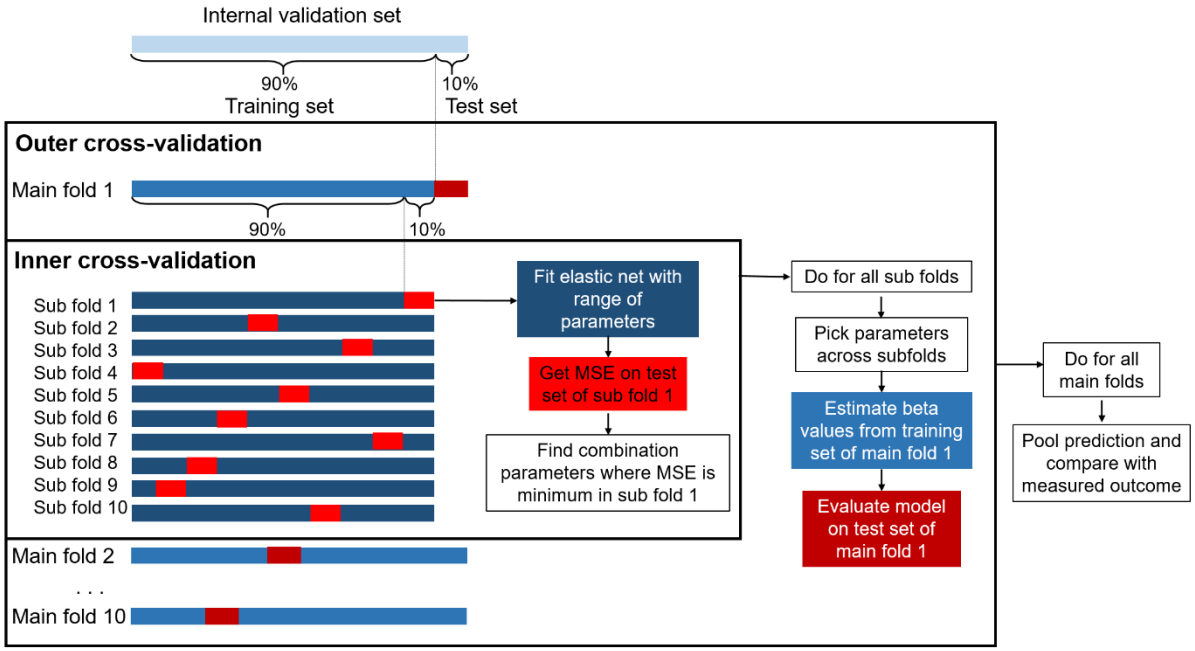

Figure 2. Procedure for the nested cross-validation. MSE=mean squared error.

The analysis on the internal validation set was performed 100 times where the training and test sets were randomly assigned. In order to quantify model performance further, we repeated

the entire procedure using random-label permutation (i.e., each subject was randomly assigned to an SSRT from a different subject). All null models included the EEG data and sex and age as covariates. The accuracy achieved using the null model was then compared to the accuracy of the model with real data (i.e., actual model) by performing a t-test. The actual model was deemed to be successful if the Pearson's *r* of the actual model was statistically significantly higher than the one of the null model ($p<0.05$). Cross-validated *r* is the most appropriate measure to use with linear regression conducted using machine learning (see for example Jollans et al., 2017). Results reported for the internal validation are mean values across all 100 iterations of the analysis. This procedure was applied to the Successful and Failed Stop ERPs separately, and to both Stop ERPs jointly.

2.5.3. Data visualization and interpretation

It can be challenging to adequately represent the results of machine learning analyses (Jollans & Whelan, 2016; Woo et al., 2017). Features are typically distributed in space and, in the case of EEG, in time (e.g., Kiiski et al., 2018; O'Halloran et al., 2019; Vahid et al., 2018). Furthermore, the final model includes features that share variance with each other. With elastic net regularization, some features can be assigned a weight of 0 (i.e., that feature is not selected in the model). To further interrogate the results, we first summed each feature's non-zero count in each main fold, and then averaged this value across the 100 iterations (the 'selection frequency'). We also calculated the selection frequency from the null models, and took the 95th percentile of the null distribution as a threshold. For comparison with previous literature, we highlighted spatio-temporal features in the surviving features that overlapped spatially and temporally with known SST-related components. As a further aid to interpretation, we created voxels (3-dimensional elements) whereby the first 2 dimensions corresponded to spatial locations from anterior to posterior and from left to right, with each electrode assigned a position in the spatial dimensions (i.e., in a 2D grid). Clusters with 3 or more voxels were kept for visualization and interpretation.

2.5.4. Analysis of the external validation set

The analysis of the internal validation set was applied to the Successful and Failed Stop ERPs separately, and then to a model with both combined. The best model based on the Pearson's correlation was then selected to test on the external validation set. The average beta values from 100 iterations of the internal validation set were used as the final model to test on the external validation set. The validation set (n=97) was scaled to the internal validation data (they were z-scored and Winsorized using the mean and standard deviation of the internal validation set). Prediction accuracy on the external validation set was quantified using the mean squared error and Pearson's correlation between the predicted and actual SSRT.

2.6. Mass-univariate approach

In order to compare the results of the machine learning approach with a standard mass-univariate test (i.e., testing each feature individually then correcting for multiple comparisons), we calculated the correlation between each single EEG spatio-temporal feature and the SSRT in the internal and external validation sets separately. We also calculated the correlation between EEG features and the shuffled SSRT 100 times (i.e., a random-label permutation). From each shuffle, we saved the maximum t value. Taking the 5th percentile from this maximum t-value distribution corrects for multiple comparisons at the 0.05 level, while allowing for correlations among features.

3. Results

3.1. Behavioral results

Table 1 displays the characteristics and statistical comparison of the internal and external validation sets. Within each set, the failed stop RTs were significantly faster than the mean Go RT ($t(147) = 25.25$, $p = 2.48*10^{-55}$, $t(96) = 23.27$, $p = 3*10^{-41}$ for internal and external validation datasets, respectively), satisfying an assumption of the race model. There was a significant difference in failed stop RT between internal and external validation datasets ($p = 0.03$, 16 ms).

	Internal validation set	External validation set	Statistical test (df)	p
Sex (Female:Male)	91:57	61:36	$X^2(1)=0.260$	0.61
Age (years)	35.27 (14.53)	33.96 (15.56)	t(243) = 0.670	0.50
SSRT (ms)	184 (53)	187 (44)	t(243) = -0.423	0.67
Mean Go RT (ms)	489 (63)	476 (69)	t(243)=0.816	0.42
Mean failed stop RT (ms)	424 (57)	408 (63)	t(243)=2.100	0.03
% successful stops	58.05 (9.62)	53.03 (9.44)	t(243)=1.442	0.15
No. of successful stop epochs	19.58 (4.79)	20.47 (3.83)	t(243)=-1.542	0.12
No. of failed stop epochs	16.02 (5.28)	16.62 (4.88)	t(243)=-0.893	0.37

Table 1. Characteristics and statistical comparison of the internal and external validation sets. SSRT: stop-signal reaction time; RT: reaction time; df: degrees of freedom. Means and standard deviations are reported, except for sex. For all analyses, t-tests were used, except for sex comparisons, which used chi-square.

3.2. Internal validation

3.2.1. Model with Successful Stop ERPs

The internal validation model with only Successful Stop data from 148 participants returned a mean cross-validated Pearson's r of 0.29 (SD = 0.04), which significantly out-performed the null model (mean r = -0.02) on over 97% of iterations, ($t(198) = 24.77$, $p = 1.48*10^{-62}$). 3.1% of features (304) from the actual model were above the selection frequency threshold (i.e., were selected in >16% of models; Figure 3). Age survived this threshold with a selection frequency of 96%, but sex did not. The spatiotemporal features that were selected in more than 50% of the iterations are listed in Table S3.

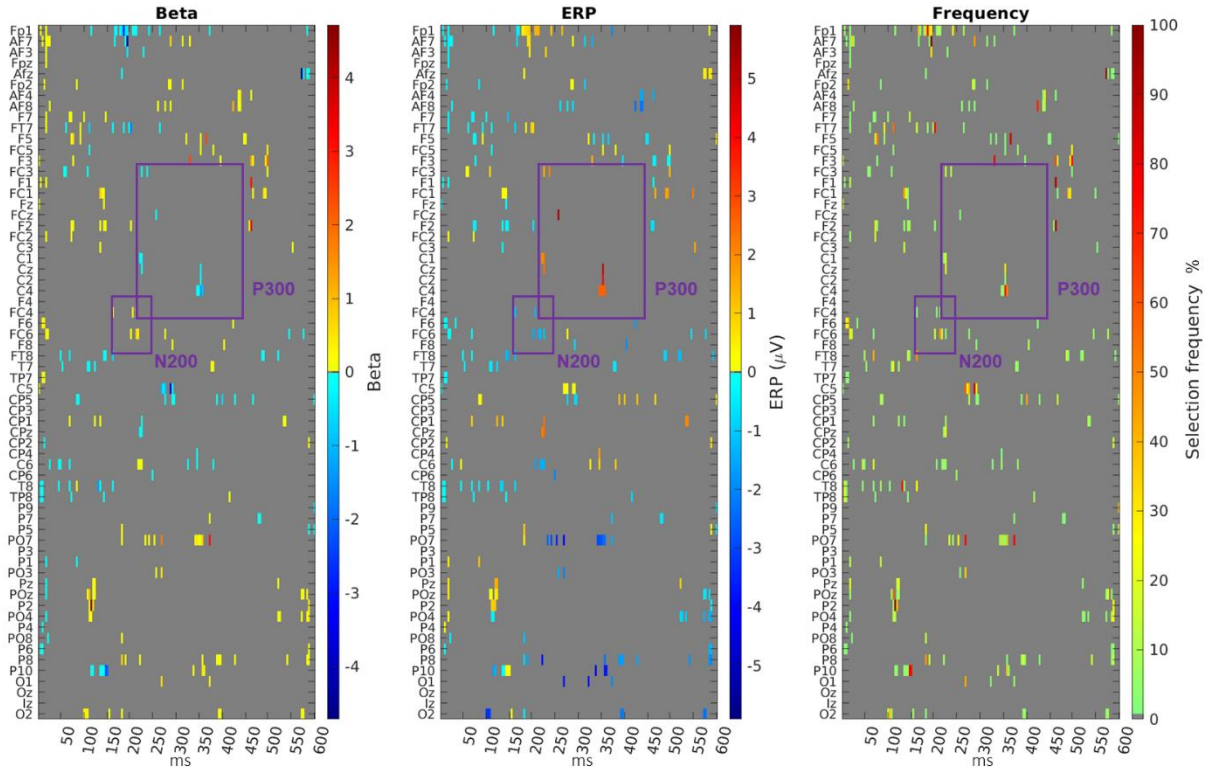

Figure 3. Maps of beta values, ERP voltage and selection frequency for the Successful Stop ERP-based features that survived the 95th percentile threshold from the null distribution. In the Beta map, warm colors indicate a positive ERP-SSRT relationship, whereas cool colors indicate a negative ERP-SSRT relationship. Features corresponding to the N200 and P300 components are highlighted in purple boxes.

Two clusters in the Successful Stop model overlapped with spatio-temporal characteristics of the P300 (see Tables 2, S4). Specifically, an early cluster was found between 222.7 and 226.5 ms, which might correspond to the ascending part of the P300. Additionally, a later cluster was found between 347.7 and 359.4 ms, which might correspond to the peak amplitude of the P300. The relationship between the amplitude of these clusters and SSRT was negative (i.e., lower amplitude associated with longer SSRT; see Beta map in Figure 3). Topographical plots are shown in Figure 4 and in Online Supplementary Videos 1-3. ERPs of central channels showing the P300 are shown in Figure S1. One cluster overlapped with spatio-

temporal characteristics of the N200 (i.e., 214.8-226.6 ms in right frontal channels) and had a positive relationship with SSRT (Figure 3, see Figures S2 and S6 for ERP plots). One cluster at approximately 100 ms with positive voltage was found over parietal-occipital areas and showed a positive relationship with SSRT. This cluster included the P2 electrode which was selected on 97% of the iterations (Table S3; see Figure S3 for ERP plots).

Time bin (ms)	Cluster size	Mean frequency (%)	Mean voltage (V)	Channels	Associated ERP component
109.4-125	10	25.9	0.85	POz, P2, CP1, Pz	P100
179.7-195.3	9	35.4	0.59	Fp1, Afz, Af7, Af3	
347.7-359.4	8	22.1	2.76	C6, C4, CP4, Cz, C2	Late P300
222.7-226.6	7	14.1	2.15	C1, CPz, FC3, CP1, Cz	Early P300
214.8-226.6	7	11.9	-1.24	FC4, FC6, C6	N200
101.6-117.2	5	17.8	-2.06	O2, PO4	
343.8-359.4	5	15.6	-2.65	PO7	
109.4-113.3	4	22.6	-0.38	T7, FT7, FC5	

Table 2. Clusters of more than three voxels identified between 100 and 400 ms in the Successful Stop model. For the full list of clusters, see Tables S4 and S9.

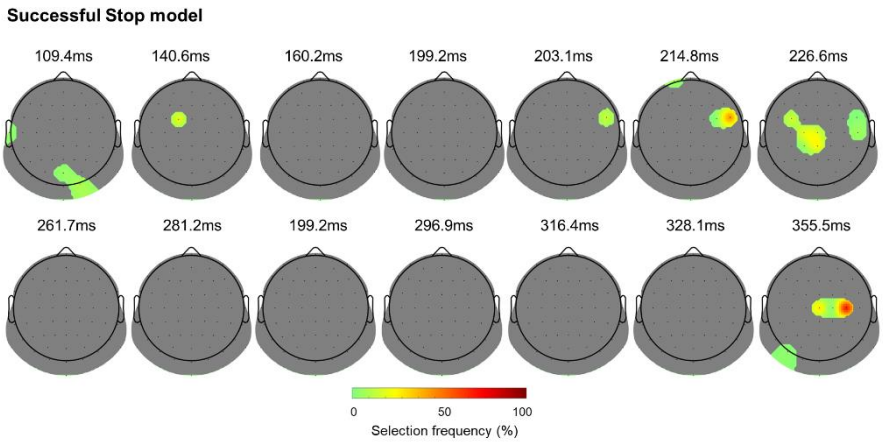

Figure 4. Topographic plots of the selection frequency at selected time bins between 100 and 400 ms for the Successful Stop model. The color represents the selection frequency of the features (channel-time point pair) across iterations that survived the 95th percentile threshold from the null distribution.

3.2.2. Model with Failed Stop ERPs

The internal validation of the Failed Stop model returned a mean cross-validated Pearson's r of 0.18 (SD = 0.04), which significantly out-performed the null model (mean r = -0.03) on over 92% of iterations ($t(198) = 17.01$, $p = 1.37*10^{-40}$). The 95th percentile of the null distribution corresponded to a 12% selection frequency: 3.8% of features from the Failed Stop ERPs were above this threshold (374 features, Figure 5), including age and sex (selection frequencies of 90% and 21%, respectively). Failed stop features with a selection frequency >50% are shown on Table S3.

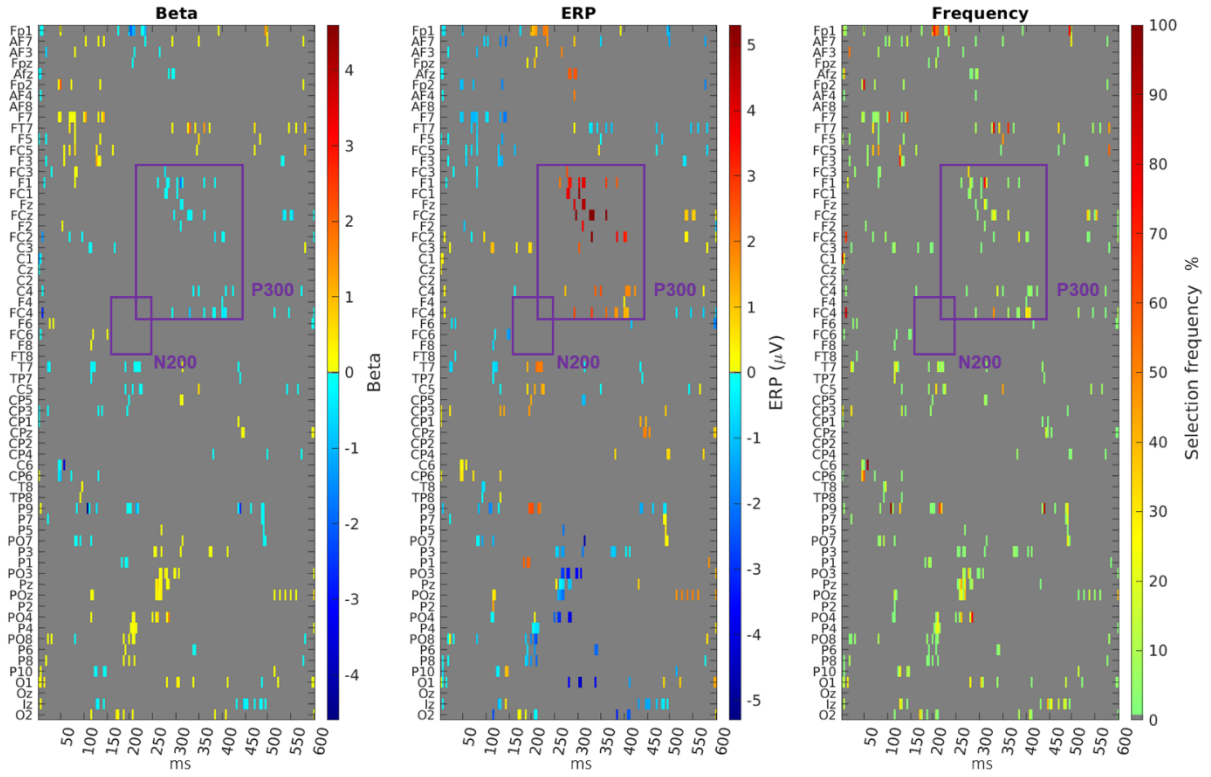

Figure 5. Maps of beta values, ERP voltage and selection frequency for the Failed Stop ERP-based features that survived the 95th percentile threshold from the null distribution. In the Beta map, warm colors indicate a positive ERP-SSRT relationship, whereas cool colors indicate negative

ERP-SSRT relationship. Features corresponding to the N200 and P300 components are highlighted in purple boxes.

Five clusters overlapped spatio-temporal characteristics of the P300 component (see Tables 3 and S5), in frontocentral areas in the time windows between 226.6 and 234.4 ms, 277.3 and 285.2 ms, 293 and 297 ms, 312.5 and 316.4 ms, and 328.1 and 335.9 ms. These clusters had a negative relationship with SSRT (see Beta map in Figure 5). Topographical plots are shown in Figure 6 and in Online Supplementary Videos 4-6. One cluster over parietal-occipital areas between 257.8 and 269.5 ms overlapped with the negative polarity of the P300 dipole (see Figures S5, S7).

Time bin (ms)	Cluster size	Mean frequency (%)	Mean voltage (V)	Channels	Associated ERP component
257.8-269.5	15	16.7	-1.64	Pz, POz, PO4, P3, PO3, P5	Dipole of P300
199.2-214.8	12	8.4	-1.07	PO8, P4, O2, P6, P8	
128.9-144.5	10	20.4	-1.27	FC5, F3, F7, Af7, C5, F5	
191.4-199.2	8	4.6	1.35	T7, C5, P1, C3, CP3	
277.3-285.2	7	25.5	-2.89	PO3, O1, Pz, PO4	
207-226.6	7	12.4	1.71	C5, T7	
277.3-285.2	7	9.3	3.17	FC3, FC1, F1, Afz	P300
199.2-210.9	6	39.7	1.10	Fp1, Af3, Fpz	
312.5-316.4	5	24.5	4.21	F1, Fz, F2	P300
328.1-335.9	5	20.1	5.54	FCz, FC2, FC4	P300
296.9-308.6	5	10.4	-3.98	PO3, O1	
312.5-316.4	5	4.1	-1.82	CP5, P3, T7, PO7	
293-297	5	4.0	3.42	Afz, Fz, Af4, FCz	P300
226.6-234.4	4	20.1	1.86	Fp1, Af7	P300
117.2-125	4	15.7	-0.61	T7, TP7	

Table 3. Clusters of more than three voxels identified between 100 and 400 ms in the Failed Stop model. For the full list of clusters, see Tables S5 and S10.

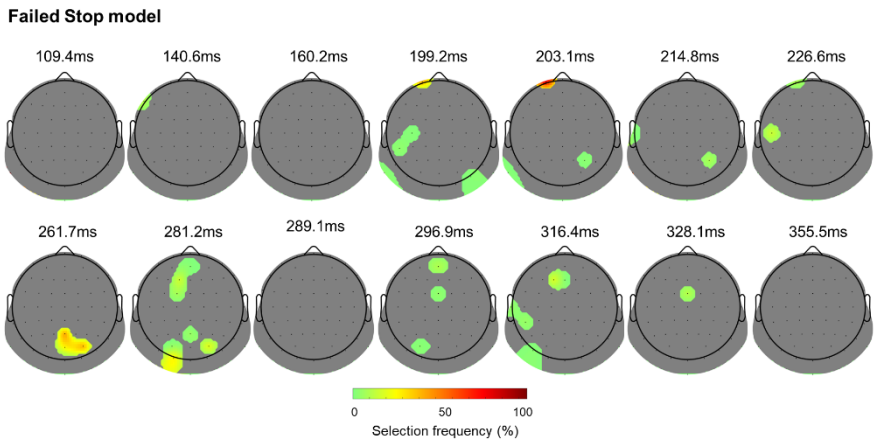

Figure 6. Topographic plots of the selection frequency at selected time bins between 100 and 400 ms for the Failed Stop model. The color represents the selection frequency of the features (channel-time point pair) across iterations that survived the 95th percentile threshold from the null distribution.

3.2.3. Full model

The mean cross-validated Pearson's r of the model including both the Successful and Failed Stop ERPs was 0.32 (SD = 0.04), which significantly out-performed the null model (mean r = -0.05) on over 99% of iterations ($t(198) = 30.18$, $p = 5.33*10^{-76}$). The 95th percentile of the null distribution corresponded to a selection frequency of 9% (Figures 7, 8, S4 and S5). 3.4% of features (666) from the actual model were above this threshold. Age survived this threshold with a selection frequency of 70%, but sex did not. The spatio-temporal features that were selected in more than 50% of the iterations are listed in Table S6. For the full list of features, see Table S11.

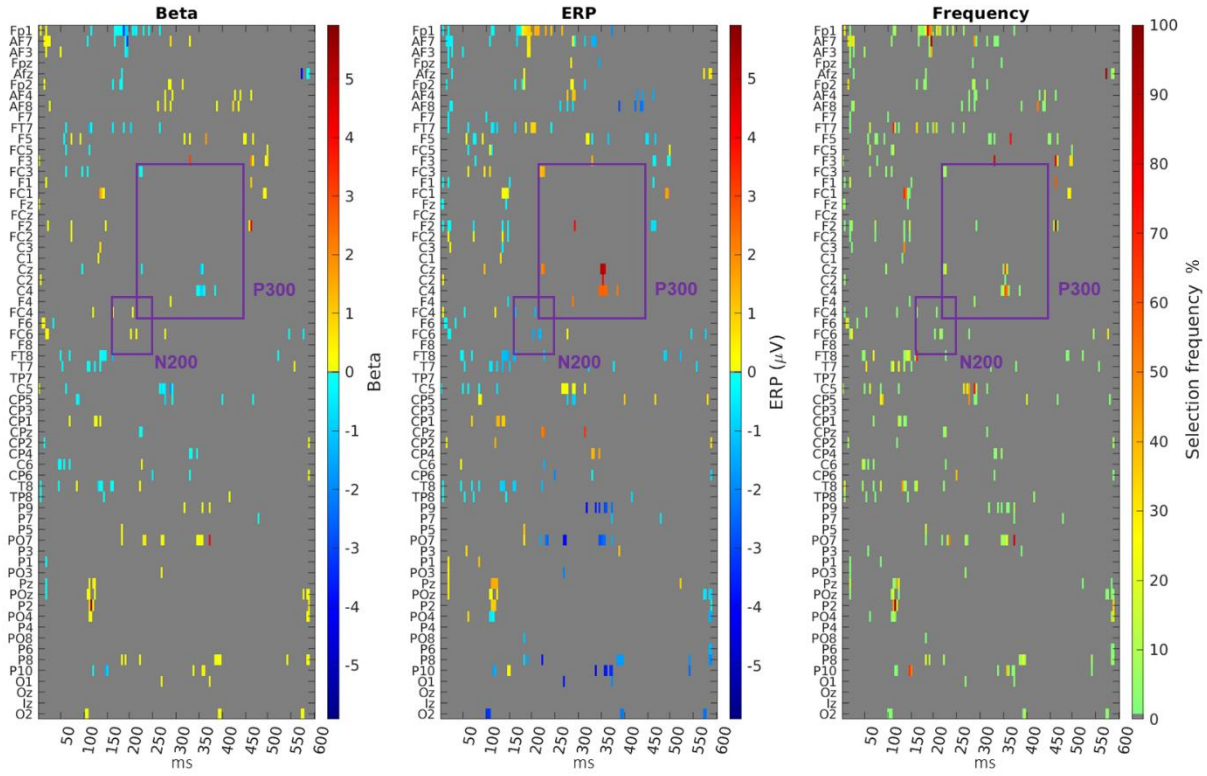

Figure 7. Maps of beta values, ERP voltage and selection frequency for the Successful Stop ERP-based features that survived the 95th percentile threshold from the null distribution in the model including both Successful and Failed Stop ERPs. In the Beta map, warm colors indicate a positive ERP-SSRT relationship, whereas cool colors indicate negative ERP-SSRT relationship. Features corresponding to the N200 and P300 components are highlighted in purple boxes.

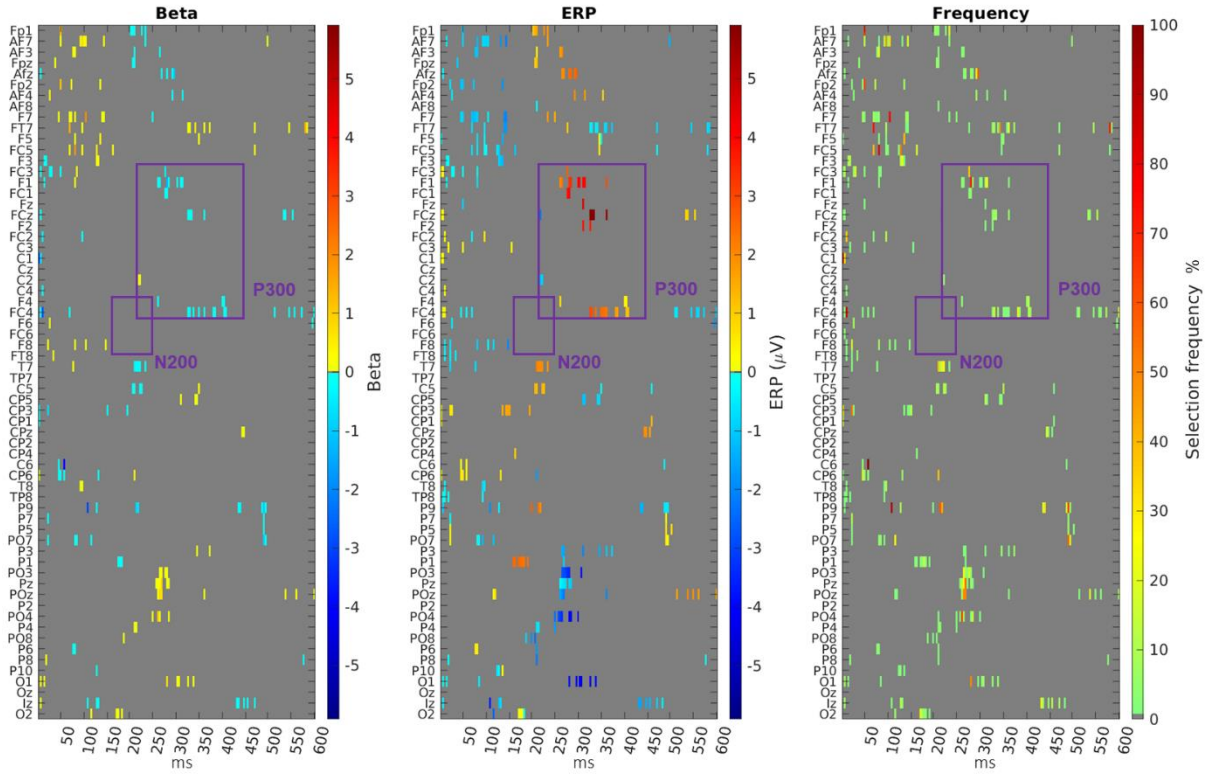

Figure 8. Maps of beta values, ERP voltage and selection frequency for the Failed Stop ERP-based features that survived the 95th percentile threshold from the null distribution in the model including both Successful and Failed Stop ERPs. In the Beta map, warm colors indicate a positive ERP-SSRT relationship, whereas cool colors indicate negative ERP-SSRT relationship. Features corresponding to the N200 and P300 components are highlighted in purple boxes.

Topographical plots with the cluster threshold are shown in Figure 9 and in Online Supplementary Videos 7-12. Six clusters in the Failed Stop ERP and two clusters in the Successful Stop ERP overlapped with spatio-temporal characteristics of the P300 (see Table 4, S7-S8). The relationship between the amplitude of these clusters and SSRT was negative (i.e., lower amplitude associated with longer SSRT; see Beta map in Figure 8.). One cluster in both Successful and Failed Stop ERPs was concomitant with the negative pole of the P300 in parietal-occipital areas (frontal and parietal-occipital channel activity were significantly negatively correlated, see Figure S7). Additionally, we found one cluster at approximately 200 ms over frontal areas with positive

voltage in both the Successful and Failed Stop ERPs. There was a positive cluster at approximately 100 ms over parietal areas, primarily in the Successful Stop ERP (see Table 4). The cluster includes the P2 electrode, which was selected on 99.8% of the iterations (Table S6). The relationship between P100 amplitude and SSRT was positive. In the Successful Stop ERP, there was also one cluster with negative voltage at approximately 100 ms over occipital areas and one cluster at approximately 120 ms with positive voltage over centro-parietal areas. There were no identified clusters of more than two voxels that overlapped with spatio-temporal characteristics of the N200.

Time bin (ms)	Cluster size	Mean frequency (%)	Mean voltage (V)	Channels	Associated ERP component
Failed Stop					
257.8-285.2	24	17.1	-2.07	Pz, PO4, POz, P3, PO3, P1, O1	Dipole of P300
125-144.5	11	13.1	-1.40	FC5, F3, Af7, F5, F7, FT7	
207-226.6	8	16.4	1.66	C5, T7	
277.3-285.2	7	17.9	3.17	FC3, FC1, F1, FC1, Afz	P300
207-214.8	7	7.3	-1.30	PO8, PO4, P4, CP6, P6, P8	
328.1-335.9	6	7.4	4.83	FCz, F2, FC4	P300
203.1-210.9	6	3.9	1.11	Fp1, AF3, Fpz	P200
261.7-269.5	6	3.2	1.99	AF3, F1, Afz	P300
312.5-316.4	5	7.9	3.62	F1, Fz, F2, Af4	P300
343.8-351.6	4	5.7	-0.75	CP5, P3, C5	
171.9-183.6	4	5.7	2.38	P1	
144.5-156.3	4	3.9	1.29	CP3, C3	
351.6-363.3	4	1.9	2.38	FC4	P300
Successful Stop					
109.4-148.4	20	20.9	0.86	POZ, PZ, P2, CP2, CP1, C1, FC1	P100
179.7-199.2	12	31.0	0.58	Fp1, Afz, Fp2, Af7, Af3	P200
347.7-363.3	10	21.9	3.38	C4, CP4, CZ, C2	P300
132.8-148.4	8	14.1	-0.99	T8, FT8, TP8	
285.2-293	7	5.1	1.26	Fpz, Fp2, Af4, F4, F2	P300
101.6-117.2	6	4.4	-1.80	O2, PO4	N100
265.6-277.3	5	24.7	0.12	FT7-C5	
109.4-117.2	5	23.3	-0.43	T7, FT7, FC5	
160.2-164.1	4	23.6	-1.09	FT8, T8, FT8	
386.7-398.4	4	12.1	-1.69	P8	
347.7-359.4	4	12.0	-2.59	PO7	
222.7-226.6	4	11.2	2.26	Cz, CPz	P200
269.5-273.4	4	6.0	-3.48	PO7, PO3, O1	Dipole of P300
168-175.8	4	5.1	-0.48	Af7, Fp1	
144.5-148.4	4	4.6	-0.32	Fz, F2, FC2, F4	
332-339.8	4	2.1	-1.10	F5, Af7	

Table 4. Clusters of more than two voxels identified between 100 and 400 ms in the Failed and Successful Stop ERPs from the Full model. For the full list of clusters, see Table S3, S4, S12.

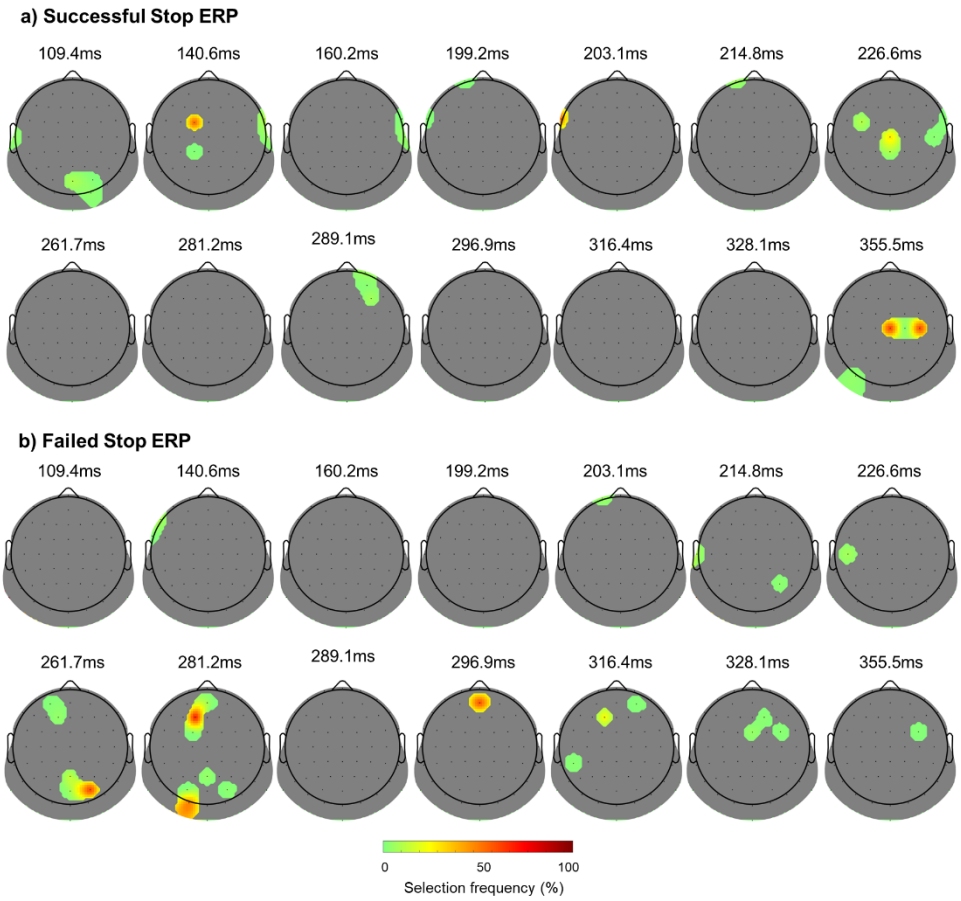

Figure 9. Topographic plots of the selection frequency at selected time bins between 100 and 400 ms for the model including both Successful and Failed Stop ERPs. The color represents the selection frequency of the features (channel-time point pair) across iterations that survived the 95th percentile threshold from the null distribution.

3.2.4. Models without covariates

To quantify the effect of including the age and sex covariates in the Successful Stop, Failed Stop and Full models, we fitted the same models without covariates and found similar results – see Supplementary Table S2, indicating that model accuracy did not depend on the inclusion of these covariates.

3.3. External validation

The model including both Successful and Failed Stop ERPs showed the best performance in the internal validation dataset (see Table S2). This model, with 19,715 beta values, including the intercept, was tested on the validation set (n = 97). The Pearson's r between the measured SSRT and the predicted SSRT derived from ERPs was 0.29 (p = 0.005; Figure S5). The data point with SSRT of 370.9 ms may be considered an outlier: after exclusion of this data point the correlation remained significant (r = 0.21, p = 0.044). A robust regression with all data also showed that this relationship was significant (slope = 0.17, p = 0.004).

3.4. Mass-univariate approach

In order to demonstrate the effect of type I errors (false positives), we first explored the data using an uncorrected significance threshold of alpha equal to 0.05 separately in the internal and external validation sets and quantified the overlap between both analyses. Of the significant features (voltage at each electrode at each time point) that survived the threshold in the internal validation set, only 32% were also significant in the external validation set. Similarly, of the significant features that survived the threshold in the external validation set, 31% were significant in the internal validation set. Thus, approximately two-thirds of significant features would fail to replicate when using an uncorrected threshold. In order to demonstrate the effect of type II errors (false negatives), we applied a correction for multiple comparisons based on the maximum statistic approach, which yielded a corrected alpha value for the internal and external validation sets of $7.37*10^{-7}$ and $2.01*10^{-6}$, respectively. No features in either set survived this correction.

4. Discussion

Here, we related individual differences between a behavioral measure of inhibitory control, SSRT, and electrophysiological activity recorded across the entire scalp and the ERP time-course from Stop trials. The use of machine learning facilitated the inclusion of more than 9,000 features per participant in order to assess the relationship of Successful and Failed Stop ERPs to SSRT. Notably,

the optimal features for predicting SSRT were widespread spatially, included early ERP activity, and included both lateral and midline electrodes, rather than being confined to the N200-P300 complex over midline electrodes. A model that included both Successful and Failed Stop ERPs predicted SSRT more accurately than either condition separately, and these results were externally validated in a separate sample. ERPs from Successful Stop trials were better predictors of SSRT than ERPs from Failed Stop trials.

4.1. Successful Stop ERPs predict SSRT

ERP activity from Successful Stop trials reliably predicted individual differences in SSRT. We identified clusters with spatio-temporal features from 222.7 to 226.6 ms and from 347.7 to 359.4 ms over frontocentral areas, which correspond with early and later P300 activity, respectively. The amplitude of the ERP in these clusters was negatively associated with SSRT (i.e., a higher amplitude was associated with a faster SSRT). These results are concordant with previous studies, which reported a negative association between the P300 amplitude and SSRT, by using either direct correlation (Hoptman et al., 2018) or a median split (Lansbergen et al., 2017); however non-significant correlations have also been reported (Anguera & Gazzaley, 2011). Anguera and Gazzaley (2011) reported a positive relation between the P300 onset latency and SSRT. This was confirmed by Wessel and colleagues who demonstrated that P300 onset latency is a better predictor of SSRT than peak latency or amplitude (Dutra et al., 2018; Wessel and Aron, 2015; Wessel et al., 2016). Although the timing of our first P300 cluster (starting at 222.7 ms) was similar to Wessel and Aron's (225.3 ms), our SSRT was faster (184 ms vs. 227.5 ms), perhaps due to different response modalities between the studies. Nonetheless, the early P300 cluster in our results likely reflects amplitude changes in the ascending slope of the P300 component, with the later cluster related to the peak amplitude (see Figure S4 and S5).

We identified features consistent with a right-frontal N200 component and we suggest this right lateralized N200 reflects individual differences in inhibitory control. The association between these features and SSRT was positive (i.e., more negative N200 was associated with faster SSRT) and these predictive features were only observed in the stop success condition. It is

worth noting that participants in our study were all adults, and a right lateralized N200 is not typically observed in adults (see Wessel & Aron, p. 478). We did not find a frontocentral N200 that related to SSRT. Unlike the consistent relationship between the P300 and SSRT, SSRT association with a frontocentral N200 is less well defined in the literature. Higher N200 amplitude (more negativity) in those with fast SSRT has been reported (van Boxtel et al., 2001; van Gaal et al., 2011). However, no association between the N200 and the SSRT has also been reported (Anguera and Gazzaley et al., 2011; Hoptman et al., 2018). These mixed findings may stem from the different choice in electrodes (typically, Fz and Cz, although van Gaal et al. included fronto-lateral electrodes), in time windows (typically 180-250 ms, but van Gaal et al. defined a later window from 281-366 ms) and smaller sample sizes ($n<42$). Importantly, van Gaal et al. (2011) reported post-hoc correlations between the N200 amplitude across the scalp and the SSRT, and found higher values in right lateralized areas, which agrees with our findings. In summary, a data-driven approach in combination with a large sample facilitated the inclusion of the entire ERP time-course and scalp, and the subsequent disentangling of the right lateralized and frontocentral N200.

Although we found evidence of right lateralized N200 and P300 association with SSRT, the best prediction of SSRT was achieved by taking into account ERPs across the entire scalp before and after the N200-P300 components. Other machine learning approaches have reported similar findings. In a Go/No Go task, Stock et al. (2016; also Vahid et al., 2018) used ERP-derived features in a machine learning approach to predict if a person was accurate/slow vs. less accurate/fast. The right-frontal N200 and frontal P300 components did not discriminate between groups. Rather, features that discriminated between groups occurred at 145-210 ms over occipital areas for No-Go responses, at 350-440 ms over parietal-occipital areas for No-Go responses, and 250-310 ms for Go responses. Similarly, we found that ERP features typically associated with SSRT, overlapping with the N200 and P300 components, were not the most important predictors, but rather earlier and later ERP features in parietal and occipital areas predicted SSRT (some of these predictors were the dipoles of the P300).

4.2. Failed Stop ERPs predict SSRT

Most previous studies focused on Successful Stop trials to investigate inhibitory control, with relatively little attention paid to the Failed Stop ERP. We found that the Failed Stop ERP is predictive of SSRT, although not to the same extent as the Successful Stop model. We identified clusters that spanned the P300 temporal window, with higher P300 amplitude associated with better SSRT. Contrary to the Successful Stop model, we did not find any relationship between the N200 and SSRT in the Failed Stop trials. This suggests that the N200 and P300 components may reflect different processes following a failed stop: the N200, performance monitoring, and the P300, both performance monitoring and motor inhibition. It is worth noting that Dimoska et al. (2006) found significant correlations between the *stimulus*-locked N200/P300 and the *response*-locked error-related negativity (ERN) and error positivity (Pe) on stop fail trials. In our adaptive version of the SST, which aims to produce 50% failed stop trials, failed responses will tend to occur shortly after the stop signal. The N200/P300 complex in the Failed Stop trials is therefore likely to incorporate the ERN/Pe complex and to reflect error processing. Future work may seek to better disentangle the N200/P300 from the ERN/Pe.

4.3. Predicting SSRT with both Successful and Failed Stop ERPs

The separate Successful and Failed Stop models were individually able to predict SSRT. However, the best prediction was achieved when using both the Successful and Failed Stop ERPs. After thresholding based on the null distribution, the surviving features of the Full model were similar to those from the separate models. Our results suggest that there is some unique variance in each Stop type that contributes to SSRT.

The C4 electrode was predictive of SSRT in at 340 ms in Successful Stop ERPs in the Full model, a time period after the mean SSRT. It is likely that the activity in the C4 electrode reflects activity from the right motor cortex (Kuo et al., 2014; Pereira et al., 2017). Involvement of the motor cortex in inhibitory control has also been reported in a study using machine learning in a Go/No Go task (Vahid et al., 2018). Vahid et al. (2018) reported that the ERP activity from the C3

electrode (over the left motor cortex) at approximately 322 ms after No-Go stimulus presentation could predict with 68% accuracy if a subject was a good performer (i.e., with good accuracy and fast response) or a bad performer (i.e., with bad accuracy and slow response). Vahid et al. proposed that, since the C3 feature was before the mean Go RT, it reflected an inhibitory command in the motor cortex directly controlling the response (subjects needed to respond with their right hand/left motor cortex). However, the mean RT may be too late in the inhibitory process (mean: 489 ms in our data). In the SST, the behavioral proxy for inhibitory control is the SSRT (mean: 184 ms in our data). In the literature, it is a matter of debate if the P300 peak, which occurs between 300 and 400 ms after the Stop cue – and is shorter than the mean RT, reflects inhibition and not evaluative processes because it occurs after the SSRT (Huster et al., 2013). As such, we interpret our finding in the C4 electrode (the non-dominant hemisphere of most subjects) as a process of inter-hemispheric inhibitory maintenance, rather than a starting inhibitory command. The task we used – an adaptive stop signal task – challenges the inhibitory capacity of the participants, resulting in an objectively more difficult task (i.e., shorter SSDs) for participants with faster SSRTs than for those with slower SSRTs (D'Alberto, 2018). Activity in the non-dominant hemisphere and inter-hemispheric connectivity is increased with task difficulty (Rueda-Delgado et al., 2016). Therefore, we suggest that the predictive power of C4 is related to an increased involvement of the non-dominant hemisphere to maintain the inhibition of the dominant motor cortex, which becomes more crucial when the task is more challenging.

4.4. Challenges of machine learning approaches

We used penalized regression, which allows for more direct interpretation of the model estimates (unlike other approaches such as Random Forest). However, the best predictors of the SSRT in all three models were spatially and temporally sparse. Features outside the N200/P300 complex, which have been seldom investigated, were included in the machine learning method and were predictive of SSRT. Interpretation of such findings is a current challenge of machine learning applications in medical sciences (Woo et al., 2017).

4.5. Contribution to reproducibility debate

There is growing concern about the utility of null-hypothesis significance testing (Trafimow et al., 2018), reproducibility in science in general (Munafò et al., 2017) and cognitive neuroscience in particular. We demonstrated that a mass-univariate approach created false positives when multiple comparisons were uncorrected (Type I error), whereas correcting for multiple comparisons across the whole spatio-temporal space resulted in no features surviving the threshold (Type II error). When compared with a mass-univariate approach, machine learning identified features that correlated with SSRT and reproduced this correlation on previously unseen data. The advantages of a machine learning approach, like the elastic net used here, is that key features can be identified from a large search space, and correlated variables can be accommodated. Typically, the whole spatio-temporal space is not examined under EEG (with more than 32 channels), with analyses guided by previous research. In contrast, in our Full model we used almost 20,000 features per participant, many of which were highly correlated. Machine learning may be particularly useful for testing new paradigms because it does not restrict the analysis to specific time intervals or electrodes.

Inhibitory deficits, such as those in addiction and ADHD or those observed during development and normal aging, are detectable with SSRT (Dawe et al., 2004; Lijffijt et al., 2005; Luijten et al., 2011; Winstanley et al., 2006). Therefore, the SST may be a valuable clinical measurement tool. However, in this study, we focused on healthy adults, thus, the resulting model may be suitable to that population exclusively. The true generalizability of the resulting model will need to be tested on populations with inhibitory deficits. Future research may also explore other approaches with higher or lower levels of interpretability, such as other generalized linear models, random forests, support vector machines or neural networks.

In this paper, individual differences in inhibitory control were predicted using machine learning on a large dataset of ERPs from 64 scalp electrodes at resolution of 4 ms. We restricted experimenter degrees of freedom by retaining all the spatio-temporal information and using internal nested cross-validation to choose the final model parameters. Model performance and

feature relevance was elucidated by comparing the original model with a model created from random-label permuted data. Additionally, we tested the model on an external dataset, with similar performance. In addition to the well-established N200 and P300 components, early ERP features were also useful for predicting SSRT, and supports the idea that machine learning can yield new insights into the electrophysiology of inhibitory control.

Acknowledgements

Robert Whelan is supported by Science Foundation Ireland (16/ERCD/3797); European Foundation for Alcohol Research (ERAB); Brain & Behavior Research Foundation (23599); Health Research Board HRA-POR- 2015-1075. Kathy Ruddy is supported by Irish Research Council fellowship GOIPD/2017/798. Nadja Enz is supported by Irish Research Council postgraduate scholarship GOIPG/2018/537. Laura O'Halloran is supported by Irish Research Council postgraduate scholarship GOIPG/2016/1635.

References

Anguera, J. A., Gazzaley, A., 2012. Dissociation of motor and sensory inhibition processes in normal aging. Clinical Neurophysiology 123(4), 730-740.

Baldwin, S. A., 2017. Improving the rigor of psychophysiology research. International Journal of Psychophysiology 111, 5-16.

Band, G.P.H., van der Molen, M.W., Logan, G.D., 2003. Horse-race model simulations of the stop-signal procedure. Acta Psychologica 112(2), 105-142.

Bekker, E.M., Kenemans, J.L., Hoeksma, M.R., Talsma, D., Verbaten, M.N., 2005. The pure electrophysiology of stopping. International Journal of Psychophysiology 55(2), 191-198.

Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., Munafo, M.R., 2013. Power failure: Why small sample size undermines the reliability of neuroscience. Nature Reviews -Neuroscience 14(5), 365-376.

Chung, Y., Addington, J., Bearden, C. E., et al., 2018. Use of machine learning to determine deviance in neuroanatomical maturity associated with future psychosis in youths at clinically high risk. JAMA Psychiatry 75(9), 960-968.

Congdon, E., Mumford, J., Cohen, J., Galvan, A., Canli, T., Poldrack, R., 2012. Measurement and Reliability of Response Inhibition. Frontiers in Psychology 3(37).

D'Alberto, N., Chaarani, B., Orr, C., Spechler, P., Albaugh, M., Allgaier, N. et al., 2018. Individual differences in stop-related activity are inflated by the adaptive algorithm in the stop signal task. Human Brain Mapping 39(8), 3263-3276.

Dawe, S., Gullo, M.J., Loxton, N.J., 2004. Reward drive and rash impulsiveness as dimensions of impulsivity: Implications for substance misuse. Addictive Behaviors 29(7), 1389-1405.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 134(1), 9-21.

Dimoska, A., Johnstone, S.J., Barry, R.J., 2006. The auditory-evoked N2 and P3 components in the stop-signal task: Indices of inhibition, response-conflict or error-detection? Brain and Cognition 62(2), 98-112.

Dutra, I., Waller, D. A., Wessel, J. R., 2018. Perceptual surprise improves action stopping by non-selectively suppressing motor activity via a neural mechanism for motor inhibition. The Journal of Neuroscience, 3091-3017.

Eklund, A., Nichols, T.E., Knutsson, H., 2016. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proc. Natl. Acad. Sci. U. S. A. 113(28), 7900-7905.

Fotiou, F., Fountoulakis, K.N., Iacovides, A., Kaprinis, G., 2003. Pattern-reversed visual evoked potentials in subtypes of major depression. Psychiatry Research 118(3), 259-271.

Foxe, J.J., Doniger, G.M., Javitt, D.C., 2001. Early visual processing deficits in schizophrenia: impaired P1 generation revealed by high-density electrical mapping. Neuroreport 12(17), 3815-3820.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of statistical software 33(1), 1-22.

Galdo-Alvarez, S., Bonilla, F.M., González-Villar, A.J., Carrillo-de-la-Peña, M.T., 2016. Functional Equivalence of Imagined vs. Real Performance of an Inhibitory Task: An EEG/ERP Study. Frontiers in Human Neuroscience 10, 467.

Greenhouse, I., Wessel, J.R., 2013. EEG signatures associated with stopping are sensitive to preparation. Psychophysiology 50(9), 900-908.

Hoptman, M.J., Parker, E.M., Nair-Collins, S., Dias, E.C., Ross, M.E., DiCostanzo, J.N., Sehatpour, P., Javitt, D.C., 2018. Sensory and cross-network contributions to response inhibition in patients with schizophrenia. NeuroImage: Clinical 18, 31-39.

Huster, R. J., Eichele, T., Enriquez-Geppert, S., Wollbrink, A., Kugel, H., Konrad, C., Pantev, C., 2011. Multimodal imaging of functional networks and event-related potentials in performance monitoring. Neuroimage 56(3), 1588-1597.

Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M., Herrmann, C. S., 2013. Electroencephalography of response inhibition tasks: functional networks and cognitive contributions. International Journal of Psychophysiology 87(3), 217-233.

Ioannidis, J.P.A., 2005. Why most published research findings are false. PLoS Med 2(8), e124.

Jollans, L., Whelan, R., 2016. The clinical added value of imaging: a perspective from outcome prediction. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 1(5), 423-432.

Jollans, L., Whelan, R., 2018. Neuromarkers for Mental Disorders: Harnessing Population Neuroscience. Frontiers in Psychiatry 9, 242.

Jollans, L., Whelan, R., Venables, L., Turnbull, O.H., Cella, M., Dymond, S., 2017. Computational EEG modelling of decision making under ambiguity reveals spatio-temporal dynamics of outcome evaluation. Behavioural Brain Research 321, 28-35.

Kenemans, J.L., 2015. Specific proactive and generic reactive inhibition. Neuroscience & Biobehavioral Reviews 56, 115-126.

Kiiski, H., Jollans, L., Donnchadha, S.Ó., Nolan, H., Lonergan, R., Kelly, S., O'Brien, M.C., Kinsella, K., Bramham, J., Burke, T., Hutchinson, M., Tubridy, N., Reilly, R.B., Whelan, R., 2018. Machine learning eeg to predict cognitive functioning and processing speed over a 2-year period in multiple sclerosis patients and controls. Brain Topography. 31(3), 346-363.

Kok, A., Ramautar, J.R., De Ruiter, M.B., Band, G.P.H., Ridderinkhof, K.R., 2004. ERP components associated with successful and unsuccessful stopping in a stop-signal task. Psychophysiology 41(1), 9-20.

Kuo, C.C., Luu, P., Morgan. K.K., Dow, M., Davey, C., et al. 2014. Localizing movement-related primary sensorimotor cortices with multi-band EEG frequency changes and functional MRI. PLOS ONE 9(11), e112103.

Lansbergen, M.M., Böcker, K.B.E., Bekker, E.M., Kenemans, J.L., 2007. Neural correlates of stopping and self-reported impulsivity. Clinical Neurophysiology 118(9), 2089-2103.

Lijffijt, M., Kenemans, J.L., Verbaten, M.N., van Engeland, H., 2005. A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: deficient inhibitory motor control? Journal of Abnormal Psychology 114(2), 216-222.

Liotti, M., Pliszka, S. R., Higgins, K., Perez III, R., Semrud-Clikeman, M., 2010. Evidence for specificity of ERP abnormalities during response inhibition in ADHD children: A comparison with reading disorder children without ADHD. Brain and cognition 72(2), 228-237.

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.

Luck, S.J., 1995. Multiple mechanisms of visual-spatial attention: recent evidence from human electrophysiology. Behavioural Brain Research 71(1), 113-123.

Luijten, M., Littel, M., Franken, I.H.A., 2011. Deficits in Inhibitory Control in Smokers During a Go/NoGo Task: An Investigation Using Event-Related Brain Potentials. PLOS ONE 6(4), e18898.

Maurage, P., Philippot, P., Verbanck, P., Noel, X., Kornreich, C., Hanak, C., Campanella, S., 2007. Is the P300 deficit in alcoholism associated with early visual impairments (P100, N170)? An oddball paradigm. Clinical Neurophysiology 118(3), 633-644.

Miller, G.A., & Chapman, J.P. 2001. Misunderstanding analysis of covariance. Journal of Abnormal Psychology 110, 40-48.

Munafò, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J.J., Ioannidis, J.P.A., 2017. A manifesto for reproducible science. Nature Human Behaviour 1, 0021.

Murphy, K.P., 2012. Machine learning: A probabilistic perspective. MIT Press, United States of America, pp. 225-232.

Nichols, T.E., Das, S., Eickhoff, S.B., Evans, A.C., Glatard, T., Hanke, M., Kriegeskorte, N., Milham, M.P., Poldrack, R.A., Poline, J.-B., Proal, E., Thirion, B., Van Essen, D.C., White, T., Yeo, B.T.T., 2017. Best practices in data analysis and sharing in neuroimaging using MRI. Nature Neuroscience 20, 299.

Nolan, H., Whelan, R., Reilly, R.B., 2010. FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. Journal of Neuroscience Methods 192(1), 152-162.

Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. Clinical Neurophysiology 112(4), 713-719.

Palmwood, E.N., Krompinger, J.W., Simons, R.F., 2017. Electrophysiological indicators of inhibitory control deficits in depression. Biological Psychology 130, 1-10.

Pereira, J., Ofner, P., Schwarz, A., Sburlea, A. I., & Müller-Putz, G. R. 2017. EEG neural correlates of goal-directed movement intention. NeuroImage 149, 129-140.

Qian, J., Hastie, T., Friedman, J., Tibshirani, R., Simon, N., 2013. Glmnet for Matlab. http://www.stanford.edu/~hastie/glmnet_matlab/. 2018).

Royston, P., Altman, D.G., Sauerbrei, W., 2006. Dichotomizing continuous predictors in multiple regression: a bad idea. Statistics in Medicine 25(1), 127-141.

Rubia, K., Lim, L., Ecker, C., Halari, R., Giampietro, V., Simmons, A., Brammer, M., Smith, A., 2013. Effects of age and gender on neural networks of motor response inhibition: From adolescence to mid-adulthood. NeuroImage 83, 690-703.

Rueda-Delgado, L. M., Solesio-Jofre, E., Mantini, D., Dupont, P., Daffertshofer, A., Swinnen, S. P., 2017. Coordinative task difficulty and behavioural errors are associated with increased long-range beta band synchronization. NeuroImage 146(Supplement C), 883-893.

Simmons, J.P., Nelson, L.D., Simonsohn, U., 2011. False-positive psychology:undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological Science 22(11), 1359-1366.

Slagter, H.A., Prinssen, S., Reteig, L.C., Mazaheri, A., 2016. Facilitation and inhibition in attention: Functional dissociation of pre-stimulus alpha activity, P1, and N1 components. NeuroImage 125, 25-35.

Smith, J.L., Johnstone, S.J., Barry, R.J., 2006. Effects of pre-stimulus processing on subsequent events in a warned Go/NoGo paradigm: Response preparation, execution and inhibition. International Journal of Psychophysiology 61(2), 121-133.

Stock, A.-K., Popescu, F., Neuhaus, A.H., Beste, C., 2016. Single-subject prediction of response inhibition behavior by event-related potentials. Journal of Neurophysiology 115(3), 1252-1262.

Trafimow, D., Amrhein, V., Areshenkoff, C. N., Barrera-Causil, C. J., Beh, E. J., Bilgiç, Y. K., . . . Marmolejo-Ramos, F. 2018. Manipulating the alpha level cannot cure significance testing. Frontiers in Psychology 9(699).

Slagter, H.A., Prinssen, S., Reteig, L.C. and Mazaheri, A., 2016. Facilitation and inhibition in attention: functional dissociation of pre-stimulus alpha activity, P1, and N1 components. Neuroimage 125, pp.25-35.

Vahid, A., Mückschel, M., Neuhaus, A., Stock, A. K., Beste, C., 2018. Machine learning provides novel neurophysiological features that predict performance to inhibit automated responses. Scientific reports 8(1), 16235.

van Boxtel, G.J.M., van der Molen, M.W., Jennings, J.R., Brunia, C.H.M., 2001. A psychophysiological analysis of inhibitory motor control in the stop-signal paradigm. Biological Psychology 58(3), 229-262.

van Gaal, S., Lamme, V.A.F., Fahrenfort, J.J., Ridderinkhof, K.R., 2011. Dissociable brain mechanisms underlying the conscious and unconscious control of behavior. Journal of Cognitive Neuroscience 23(1), 91-105.

Verbruggen, F., Logan, G.D., Stevens, M.A., 2008. STOP-IT: Windows executable software for the stop-signal paradigm. Behavior Research Methods 40(2), 479-483.

Verona, E., & Miller, G.A. 2015. Analysis of covariance. In R.L. Cautin & S.O. Lilienfeld (Eds.), The encyclopedia of clinical psychology (p 136-142). New York: Wiley.

Vu, M.-A.T., Adalı, T., Ba, D., Buzsáki, G., Carlson, D., Heller, K., Liston, C., Rudin, C., Sohal, V.S., Widge, A.S., Mayberg, H.S., Sapiro, G., Dzirasa, K., 2018. A shared vision for machine learning in neuroscience. The Journal of Neuroscience 38(7), 1601-1607.

Vul, E., Harris, C., Winkielman, P., Pashler, H., 2009. Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. Perspectives on Psychological Science 4(3), 274-290.

Wessel, J.R., Aron, A.R., 2015. 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.

Wessel, J. R., Jenkinson, N., Brittain, J.-S., Voets, S. H. E. M., Aziz, T. Z., Aron, A. R., 2016. Surprise disrupts cognition via a fronto-basal ganglia suppressive mechanism. Nature Communications 7, 11195.

Whelan, R., Conrod, P. J., Poline, J.-B., Lourdusamy, A., Banaschewski, T., Barker, G. J., . . . Garavan, H., 2012. Adolescent impulsivity phenotypes characterized by distinct brain networks. Nature Neuroscience 15(6), 920-925.

Whelan, R., Garavan, H., 2014. When optimism hurts: inflated predictions in psychiatric Neuroimaging. Biological Psychiatry 75(9), 746-748.

Whelan, R., Watts, R., Orr, C. A., Althoff, R. R., Artiges, E., Banaschewski, T., . . . the, IMAGEN Consortium, 2014. Neuropsychosocial profiles of current and future adolescent alcohol misusers. Nature 512, 185.

Winstanley, C.A., Eagle, D.M., Robbins, T.W., 2006. Behavioral models of impulsivity in relation to ADHD: Translation between clinical and preclinical studies. Clinical Psychology Review 26(4), 379-395.

Woo, C.-W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: brain models in translational neuroimaging. Nature Neuroscience 20(3), 365-377.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2), 301-320.

Supplementary Material

S2.1. Inclusion/exclusion criteria

Data from four projects was pooled for this study. The distribution of the data per project is shown on Table S1. The exclusion criteria for each project is listed below.

Project No.	N	mean age	range	Number females
1	88	32	16-61	57
2	40	21.13	18-29	16
3	108	46.61	20-69	72
4	46	25.72	19-60	30
Total	**282**	**35.03**	**16-69**	**175**

Table S1. Distribution of participants per project.

Project 1: The focus of this project was to recruit first degree relatives of individuals with attention deficit/hyperactivity disorder and controls. The exclusion criteria included history of traumatic brain injury (e.g. concussion); medical conditions such as epilepsy; severe migraine; hearing and/or severe motor impairment; stroke and diabetes; psychiatric conditions such as psychosis, bipolar disorder, eating disorder, schizophrenia and personality disorder; learning and/or intellectual disability; current medication use of antidepressants (e.g. selective serotonin reuptake inhibitors (SSRIs)), benzodiazepines, antipsychotics or anticonvulsants.

Projects 2 and 4: The focus of this project was to recruit healthy participants. The exclusion criteria included under 18 years old; accident causing blunt force trauma to the head; stroke patients; Receptive Language difficulties; consumption of cannabis regularly; alcohol or drug problems; learning or intellectual disability; mental illness; physical disability that might negatively affect the performance in the study.

Project 3: The focus of this project was to recruit participants who smoke regularly. So the exclusion criteria included smokers of less than 5-10 cigarettes per day for the past 12 months;

smokers who had attempted to quit smoking during the past week; under 18 years old and over 70; English proficiency of at least B1 standard of English proficiency on the CEFR framework; personality disorder or other serious psychiatric diagnosis; currently taking any psychoactive drugs (other than tobacco), cannabis, alcohol, eltroxin (levothyroxine) or SSRI anti-depressants; anybody who had trouble controlling their alcohol/cannabis/drug use, has any learning difficulties and has suffered an injury that caused 30 minutes or more of unconsciousness.

S3.2.1 Event-related potential plots

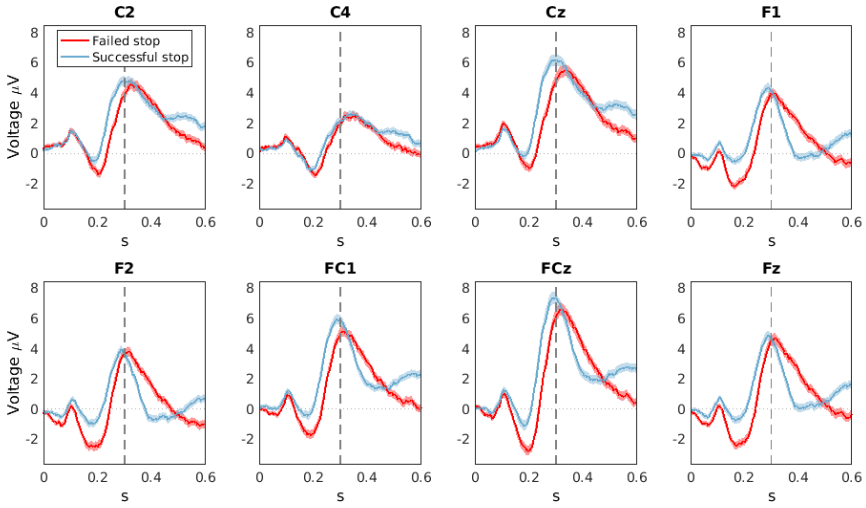

Figure S1. Grand average ERPs for the Failed and Successful Stop trials in selected channels over the frontal area. A positivity at around 300 ms (P300) is observed. Vertical dotted line indicates the 300 ms time stamp.

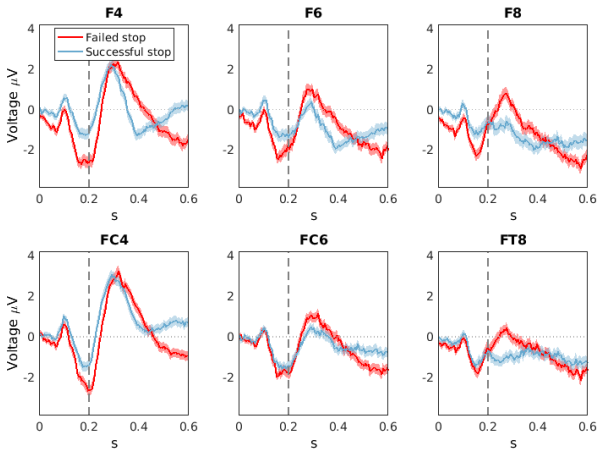

Figure S2. Grand average ERPs for the Failed and Successful Stop trials in selected channels over the right frontal area. A negativity at around 200 ms (N200) is observed. Vertical dotted line indicates the 200 ms time stamp.

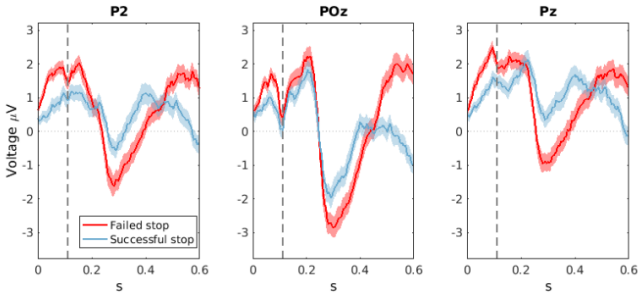

Figure S3. Grand average ERPs for the Failed and Successful Stop trials in selected channels over the parietal area. A positivity at around 100 ms is observed. Vertical dotted line indicates the 113 ms time stamp corresponding to the time window for channel P2 that was selected 99.8% of the iterations.

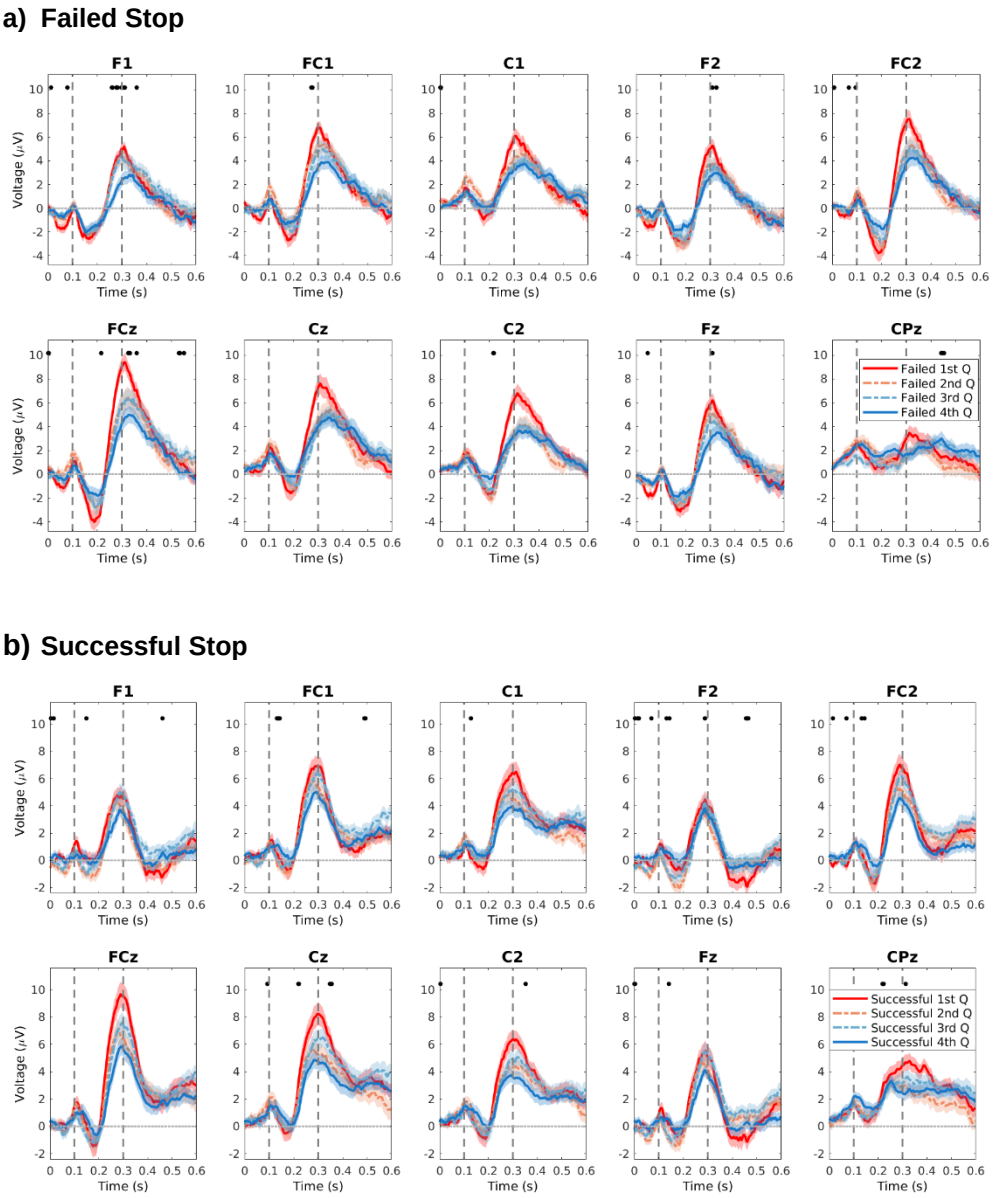

Figure S4. Grand average ERPs for the a) Failed and b) Successful Stop trials in selected channels over the frontal area per quartile (Q) of SSRT. Vertical dotted lines indicate the 100 and 300 ms time stamp. Black dots indicate spatio-temporal features in the Full model that survived the 95th percentile threshold of the null distribution.

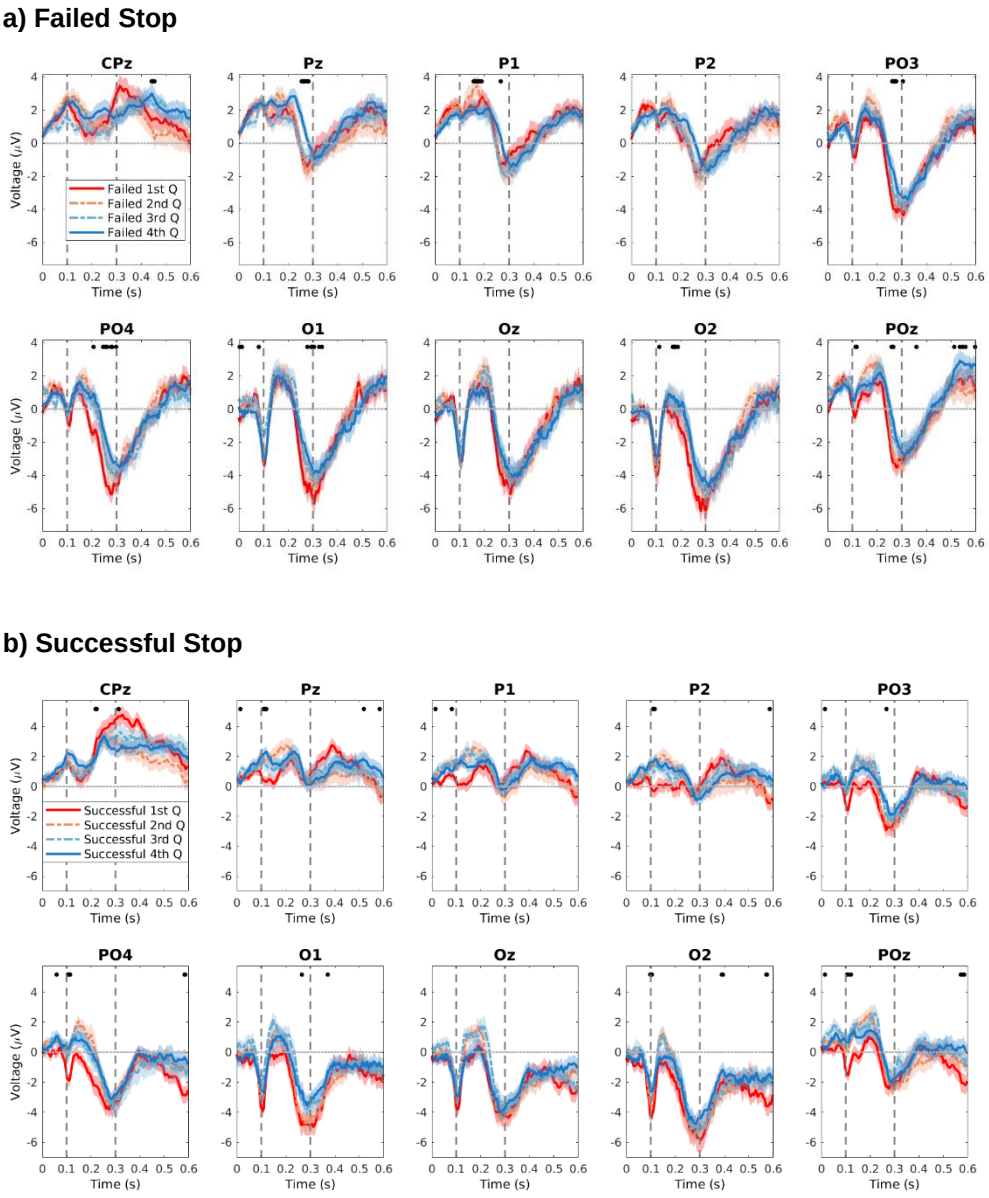

Figure S5. Grand average ERPs for the a) Failed and b) Successful Stop trials in selected channels over the parietal-occipital area per quartile (Q) of SSRT. Vertical dotted lines indicate the 100 and 300 ms time stamp. Black dots indicate spatio-temporal features in the Full model that survived the 95th percentile threshold of the null distribution.

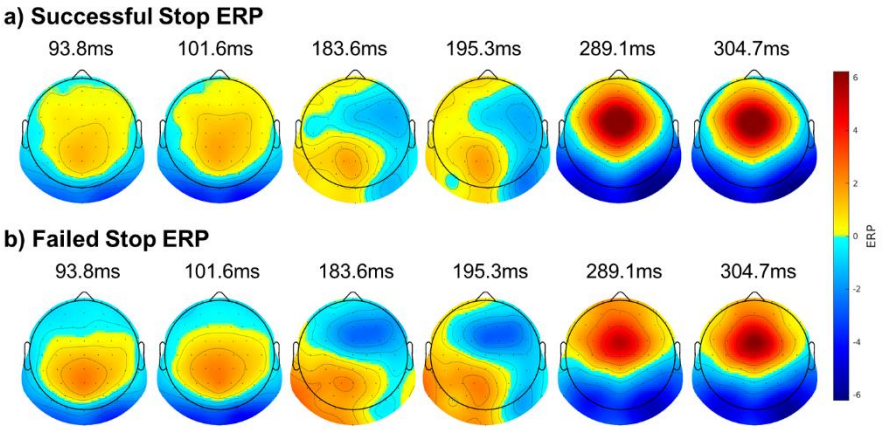

Figure S6. Grand average ERP in time bins close to 100, 200 and 300 ms for a) Successful Stops and b) Failed Stops.

S3.2 Models results

	EEG + covariates			EEG only		
	Successful Stop	**Failed Stop**	**Full model**	**Successful Stop**	**Failed Stop**	**Full model**
Original						
Average MSE	2577.5	2813	2536.2	2592.9	2843.4	2522.8
SD MSE	99.5	99.2	109.3	99.5	97.7	108.4
Average r	0.29	0.18	0.32	0.29	0.17	0.33
SD r	0.041	0.044	0.042	0.042	0.039	0.041
Null						
Average MSE	3005.2	2996.7	3073.8	2982.96	3005.8	3070.9
SD MSE	175.4	163.8	184.5	134.4	154.4	176.2
Average r	-0.02	-0.03	-0.05	-0.03	-0.029	-0.041
SD r	0.119	0.118	0.117	0.114	0.112	0.123

Table S2. Mean square error (MSE) and Pearson's correlation (r) of the original and null models using the Successful Stop data, Failed Stop data and joint Successful and Failed Stop data (Full model). These models were fitted including EEG data and covariates or including EEG data only. SD: standard deviation.

Successful Stop				**Failed Stop**			
Channel	**Time bin (ms)**	**Mean selection frequency (%)**	**Beta value**	**Channel**	**Time bin (ms)**	**Mean selection frequency (%)**	**Beta value**
intercept		100	183.96	intercept		100	183.97
AF7	[191.4,195.3]	99.2	-4.4918	C6	[54.7,58.6]	97	-3.7096
P2	[113.3,117.2]	97	5.8951	P9	[105.5,109.4]	96.2	-5.7453
age		96.2	3.5007	P9	[437.5,441.4]	90.9	-2.7604
AFz	[570.3,574.2]	94.8	-5.3510	age		90	3.5177
F2	[460.9,464.8]	94	3.5016	FC4	[7.8,11.7]	86.4	-2.6323
F1	[460.9,464.8]	91.4	3.0217	Fp2	[46.9,50.8]	82.8	2.0983
C5	[285.2,289.1]	86.6	-3.7836	FT7	[328.1,332.0]	81.1	1.9004
F5	[363.3,367.2]	85.4	2.3534	Fp1	[492.2,496.1]	76.8	1.4600
Fp1	[183.6,187.5]	84	-2.8045	Fp1	[199.2,203.1]	76.3	-2.2002
P10	[144.5,148.4]	81.6	-2.3957	PO4	[281.2,285.2]	76.2	1.6592
F3	[328.1,332.0]	80.2	2.4100	F1	[308.6,312.5]	75.9	-1.7881
PO7	[371.1,375.0]	78.6	3.1033	C1	[0.0,3.9]	74.2	-1.4337
T8	[128.9,132.8]	75.6	-1.2840	F3	[125.0,128.9]	68.2	1.2096
F3	[496.1,500.0]	75.4	1.9547	FT7	[359.4,363.3]	67	1.3598
FT7	[199.2,203.1]	75.2	-1.6154	FC2	[7.8,11.7]	64.6	-1.3832
PO7	[265.6,269.5]	75	1.7559	P9	[210.9,214.8]	63.9	-1.1298
C4	[351.6,355.5]	73	-1.4352	F7	[101.6,105.5]	60.6	1.2712
P10	[148.4,152.3]	72.8	-1.9162	F7	[140.6,144.5]	60.5	1.1948

AF8	[421.9,425.8]	68.8	1.1477	CP6	[46.9,50.8]	58.5	-0.8742
F3	[460.9,464.8]	61	1.1792	Fp1	[230.5,234.4]	57.7	-1.1036
POz	[109.4,113.3]	57.2	1.0765	FC5	[578.1,582.0]	56.1	0.8508
C5	[265.6,269.5]	57	-1.0399	Fp1	[207.0,210.9]	55.5	-1.0656
FT7	[109.4,113.3]	54.8	-0.7523	AF3	[15.6,19.5]	53.5	-0.7243
C4	[355.5,359.4]	50.8	-0.9401	Fp1	[203.1,207.0]	53.4	-1.1347

Table S3. Features with a selection frequency >50% across iterations in the Successful and Failed Stop only models.

Time bin (ms)	Cluster size	Mean frequency (%)	Mean voltage (V)	Channels
<100ms				
15.6-27.3	8	11.8	-0.31	Fp2, Af7, Fp1, Af3, Fpz
11.7-27.3	5	20.1	-0.11	F6, FC6, C6
70.3-78.1	5	15.6	-0.26	F7, F5, F3
70.3-11.7	5	13.2	-0.25	FC6, C6, FT8, T8, TP8, P6
7.8-11.7	5	11.2	-0.17	T8, TP8, P6
19.5	4	11.0	0.52	P1, Pz, POz, PO4
>400ms				
574.2-578.1	5	9.4	-1.42	POz, PO4, O2
492.2-500	4	21.6	1.12	FC1, FC5, FC3
585.9-589.8	4	16.8	-0.95	POz, PO4, P2
437.5-441.4	4	13.3	-1.85	Af4, Af8
585.9-589.8	4	6.6	-1.59	P6, P8

Table S4. Clusters of more than three voxels identified before 100 ms and after 400 ms in the Successful Stop model.

Time bin (ms)	Cluster size	Average frequency (%)	Average voltage (V)	Channels
<100 ms				
70.3-82	12	12.8	-0.81	F7, FT7, FC5, F5, FC3, Af3, F1
3.9-7.8	5	25.3	0.41	CP3, C1, CP1, Cz
3.9-11.7	5	11.9	-0.36	Afz, Fp1
7.8-11.7	4	39.2	0.20	FC6, FC2, FC4, C4
46.9-50.8	4	26.8	0.40	C6, CP6
15.6-19.5	4	20.8	-0.30	F1, F5, F3, Af3
> 400 ms				
402.3-414	9	11.9	1.78	FC2, F4, FC4, C4
488.3-496.1	6	9.7	0.44	P7, P5, CP3, PO7
597.7-601.6	5	10.2	1.30	O1, PO3, Poz
535.2-539	4	13.3	0.65	FCz, FC2

597.7-601.7	4	4.3	-1.44	F6, F2, FC4

Table S5. Clusters of more than three voxels identified before 100 ms and after 400 ms in the Failed Stop model.

3.2.3 Full model

Different α and λ parameters were selected for each main fold within each of the 100 set assignments. We averaged the parameters across main folds and calculated the mean and standard error of the mean across set assignments. From this, the mean and SEM of α were 0.7244 and 0.0032, respectively, and the mean and SEM of the λ parameter were 5.559 and 0.046, respectively in the Full model.

Trial type	Channel	Time bin ms	Selection frequency (%)	Beta_value
intercept			100	183.8421
SubtractedSuccess	P2	[113.3,117.2]	99.8	7.3970
SubtractedFail	C6	[54.7,58.6]	99.3	-3.9078
SubtractedFail	C1	[0.0,3.9]	95.2	-3.2858
SubtractedSuccess	F2	[460.9,464.8]	94.9	4.3714
SubtractedSuccess	AFz	[570.3,574.2]	93.8	-4.2131
SubtractedSuccess	AF7	[191.4,195.3]	93.3	-3.0087
SubtractedFail	P9	[105.5,109.4]	91.4	-3.0003
SubtractedSuccess	*Fp1*	*[183.6,187.5]*	*90.5*	*-3.3313*
SubtractedSuccess	PO7	[371.1,375.0]	87.5	4.2395
SubtractedFail	FC4	[7.8,11.7]	86.7	-2.8128
SubtractedSuccess	F3	[328.1,332.0]	84.7	3.1098
SubtractedSuccess	F3	[460.9,464.8]	82.6	2.1106
SubtractedFail	FC5	[78.1,82.0]	81.9	1.7759
SubtractedFail	F7	[101.6,105.5]	78.7	1.9720
SubtractedSuccess	F5	[363.3,367.2]	72.8	1.7352
SubtractedFail	FT7	[66.4,70.3]	72.3	1.4353
SubtractedFail	FT7	[578.1,582.0]	72	1.3612
SubtractedSuccess	C5	[285.2,289.1]	71.2	-1.8167
age			70.4	1.4762
SubtractedSuccess	FT8	[160.2,164.1]	70.2	-1.2530
SubtractedFail	**F1**	**[277.3,281.2]**	**70**	**-1.2915**
SubtractedSuccess	FT7	[109.4,113.3]	69.5	-1.2231
SubtractedSuccess	FC1	[132.8,136.7]	67.8	1.6783
SubtractedSuccess	P10	[144.5,148.4]	67.7	-1.4137
SubtractedFail	Fp1	[46.9,50.8]	65.5	1.1790
SubtractedFail	P9	[488.3,492.2]	63.2	-1.0383
SubtractedSuccess	**Cz**	**[351.6,355.5]**	**62.9**	**-1.1838**
SubtractedFail	Fp2	[46.9,50.8]	62.8	1.1738
SubtractedFail	PO4	[257.8,261.7]	62.3	1.0844
SubtractedSuccess	**C4**	**[351.6,355.5]**	**61.5**	**-0.9149**
SubtractedFail	**AFz**	**[293.0,296.9]**	**59.6**	**-1.0800**
SubtractedSuccess	F1	[460.9,464.8]	59.1	1.0247
SubtractedFail	P9	[210.9,214.8]	58.7	-0.8713
SubtractedFail	Pz	[265.6,269.5]	56.5	0.8760
SubtractedFail	PO7	[488.3,492.2]	56.2	-0.8892
SubtractedSuccess	FC1	[136.7,140.6]	55.2	0.8971
SubtractedFail	CP3	[19.5,23.4]	54.2	-0.7684
SubtractedSuccess	AF8	[421.9,425.8]	52.2	0.7527
SubtractedSuccess	**C4**	**[355.5,359.4]**	**51.5**	**-0.8225**
SubtractedSuccess	*Fp1*	*[187.5,191.4]*	*51.4*	*-0.8813*
SubtractedFail	FC5	[125.0,128.9]	51.3	0.6595
SubtractedSuccess	FT7	[199.2,203.1]	51	-0.6938
SubtractedSuccess	PO7	[230.5,234.4]	50.8	0.6713
SubtractedFail	POz	[265.6,269.5]	50.3	0.6690
SubtractedSuccess	P10	[148.4,152.3]	50.1	-0.8411
SubtractedSuccess	**C5**	**[265.6,269.5]**	**50**	**-0.7488**

Table S6. Features with a selection frequency >50% across iterations in the Full model. Bold: Features that overlap with P300.

Time bin (ms)	Cluster size	Average frequency (%)	Average voltage (V)	Channels
<100ms				
3.9-11.7	13	24.2	0.24	FC3, CP3, C1, CP1, FCz, FC4, FC6, FC2, C4
70.3-82	13	19.0	-0.86	F7, FT7, FC5, AF3, F5, FC5, AF7, FC3, F1
93.8-105.5	8	13.5	-0.60	AF7, FT7, F5, FC5
19.5-27.3	5	12.0	0.36	C3, P5, CP3, PO7
7.8-11.7	5	1.4	-0.26	TP8, F8, FT8, T8
15.6-19.5	4	5.8	-0.30	AF7, F3, F1
>400ms				
492.2-496.1	4	21.2	0.38	P7, P5, PO7
402.3-410.2	4	8.6	0.78	F4, FC4

Table S7. Clusters of more than three voxels identified before 100 ms and after 400 ms in the Failed Stop ERP from the Full model.

Time bin (ms)	Cluster size	Average frequency (%)	Average voltage (V)	Channels
<100ms				
15.6-27.3	11	14.2	-0.30	AF7, FP2, FP1, AF3, F1, FPZ, AF7, AFZ
11.7-23.4	4	19.1	-0.13	F6, FC6
70.3-74.2	4	7.0	-0.26	FT8, C6, TP8, T8
19.5-23.4	4	5.6	-0.25	F2, AF8
46.9-50.8	4	5.6	-0.19	FT8, T8, C6
19.5	4	5.2	0.55	P1, PO3, Pz, POz
3.9-7.8	4	4.4	-0.10	Fz, F1, F2
>400ms				
585.9-589.8	5	15.5	-0.92	POz, PO4, P2, P6, P8, CP6
464.8-468.8	4	41.4	-0.43	F1, F3, F5,
460.9-468.8	4	27.6	-0.85	F2, AF4
574.2-578.1	4	3.4	-1.45	POz, O2

Table S8. Clusters of more than three voxels identified before 100 ms and after 400 ms in the Successful Stop ERP from the Full model.

We evaluated the relationship between the frontal and parietal-occipital channels during the Failed Stop ERP with Pearson's correlations. Ten frontal and seven parietal-occipital channels were selected over the window from 256 to 286 ms. This resulted in 70 pairs tested in 8 time bins. We calculated the percentage of pairs that showed a correlation that survived a significance

threshold of 0.05/8 (Bonferroni correction for 8 time points). More than 87% of the pairs survived this threshold (Figure S7, left).

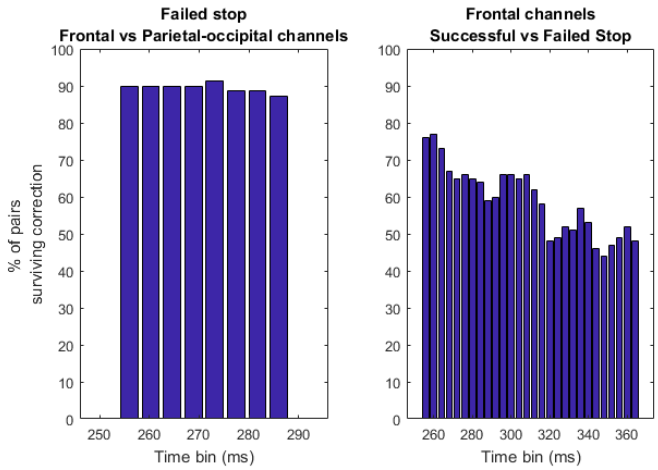

Figure S7. Percentage of pairs that survived a Bonferroni correction. Left: Percentage for the correlation between frontal and parietal-occipital channels during the Failed Stop ERP. Right: Percentage for the correlation between Successful and Failed Stop ERPs in frontal channels.

We also calculated the correlation between Successful and Failed Stop ERPs in the frontal channels. We selected 10 frontal channels in the interval from 256 to 364 ms. This resulted in 100 pairs from 28 time bins. After correcting for the time bins ($p<0.05/28$), up to 300 ms more than 59% of channels have a significant correlation. After 300 ms, more than 44% of channels show this correlation (Figure S7, right).

S3.3. External validation results

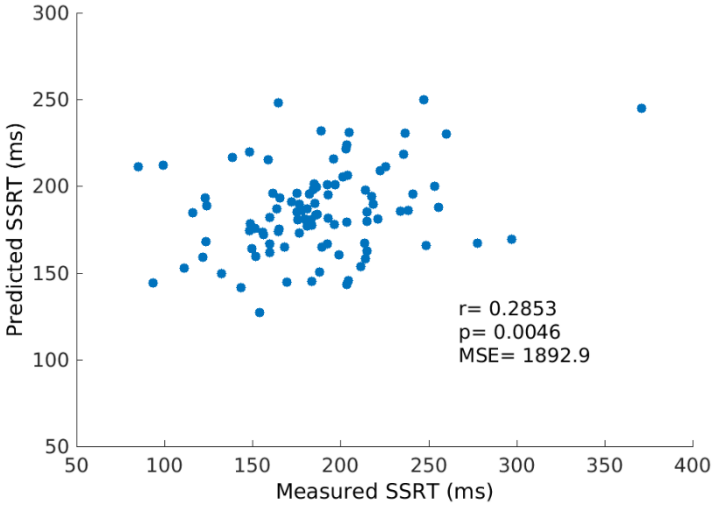

Figure S8. Predicted SSRT using the Full model from the cross-validation on the external validation set.

Additional supplementary files

Tables with full list of clusters with more than 2 voxels are shown in a separate Excel file for the three models (Table S9, S10, S12).

A table with the full list of features, sorted by the mean selection frequency, from the Full model is in a separate Excel file (Table S11.)

There are Supplementary Videos with the topographical plots across time of the beta map, the ERP and the selection frequency for the three models (Videos 1-12).