Title: Volume of β-bursts, but not their rate, predicts successful response inhibition

Abbreviated title: β-Burst volume predicts response inhibition

Authors: Nadja Enz[1], Kathy L. Ruddy[1], Laura M. Rueda-Delgado[1], Robert Whelan[1,2]

Affiliations: [1]School of Psychology and Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland, [2]Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland

Corresponding author email address: Robert Whelan; robert.whelan@tcd.ie

Number of pages: 40

Number of figures: 5
Number of tables: 2
Number of multimedia: 0
Number of 3D models: 0

Number of words for abstract: 250
Number of words for significance statement: 117
Number of words for introduction: 646
Number of words for discussion: 1499

Conflict of interest statement: The authors declare no competing financial interests.

Acknowledgements: We are grateful to Dr Jan Wessel for sharing his Stop Signal Task electroencephalography data, and for his advice in interpreting the data and analysis scripts. We would like to thank Dr Hanni Kiiski, Dr Laura O'Halloran, Rory Boyle, Laura Rai and Gabi Pragulbickaite for help with data collection and electroencephalography quality control. All machine learning calculations were performed on the Lonsdale cluster maintained by the Trinity Centre for High Performance Computing. This cluster was funded through grants from Science Foundation Ireland. Nadja Enz is supported by Irish Research Council postgraduate scholarship GOIPG/2018/537. Kathy L. Ruddy would like to acknowledge funding from the Irish Research Council GOIPD/2017/798 and Health Research Board, Ireland HRB-EIA-2019-003. Laura M. Rueda-Delgado is supported by the Science Foundation Ireland (SFI 18/IF/6272). Robert Whelan was supported by Science Foundation Ireland (16/ERCD/3797); European Foundation for Alcohol Research (ERAB); Brain & Behavior Research Foundation (23599); Health Research Board HRAPOR-2015-1075.

Abstract

In humans, impaired response inhibition is characteristic of a wide range of psychiatric diseases and of normal aging. It is hypothesised that the right inferior frontal cortex plays a key role by inhibiting the motor cortex via the basal ganglia. The electroencephalography-derived β-rhythm (15-29 Hz) is thought to reflect communication within this network, with increased right frontal β-power often observed prior to successful response inhibition. Recent literature suggests that averaging spectral power obscures the transient, burst-like nature of β-activity. There is evidence that the rate of β-bursts following a Stop signal is higher when a motor response is successfully inhibited. However, other characteristics of β-burst events, and their topographical properties, have not yet been examined. Here, we used a large human (male and female) electroencephalography Stop Signal Task dataset (n=218) to examine averaged normalised β-power, β-burst rate and β-burst 'volume' (which we defined as burst duration x frequency span x amplitude). We first sought to optimise the β-burst detection method. In order to find predictors across the whole scalp, and with high temporal precision, we then used machine learning to (1) classify successful vs. failed stopping and to (2) predict individual Stop Signal Reaction Time. β-Burst volume was significantly more predictive of successful and fast stopping than β-burst rate and normalised β-power. The classification model generalised to an external dataset (n=201). We suggest β-burst volume is a sensitive and reliable measure for investigation of human response inhibition.

Significance Statement

The electroencephalography-derived β-rhythm (15-29 Hz) is associated with the ability to inhibit ongoing actions. In this study, we sought to identify the specific characteristics of β-activity that contribute to successful and fast inhibition. In order to search for the most relevant features of β-activity – across the whole scalp and with high temporal precision – we employed machine learning on two large datasets. Spatial and temporal features of β-burst 'volume' (duration x frequency span x amplitude) predicted response inhibition outcomes in our data significantly better than β-burst rate and normalised β-power. These findings suggest that multidimensional measures of β-bursts, such as burst volume, can add to our understanding of human response inhibition.

Introduction

The ability to inhibit unwanted or inappropriate behaviours relies on effective response inhibition in the brain. The Stop Signal Task (SST) measures this cognitive process (Logan and Cowan, 1984) by requiring the participant to cancel an already initiated motor response following an infrequent Stop cue. If the response is withheld following the Stop cue then the trial is classified as a 'success'. If a response is made then the trial is classified as a 'failure'. The Stop Signal Reaction Time (SSRT) is an estimation of the covert latency of the action cancellation process (Verbruggen et al., 2019). The SSRT is ~200-250 ms in healthy adults, and slower SSRTs are characteristic of several psychiatric diseases (Lijffijt et al., 2005; Luijten et al., 2011) and normal aging (Hsieh and Lin, 2017).

Many previous studies have suggested that right inferior frontal cortex (rIFC) is involved in response inhibition through a rIFC-basal ganglia-motor cortex pathway (Aron et al., 2014; Wessel and Aron, 2017). Jana et al. (2020) recently supported the role of this pathway in action stopping by suggesting an exact temporal cascade: rIFC activation at 120 ms post Stop signal, global motor suppression at 140 ms, muscle inhibition at 160 ms and SSRT at 220 ms. Communication in this pathway may be facilitated through brain oscillations in the β-frequency band (15-29 Hz), and several studies have reported an increase in averaged β-power over the rIFC at the moment of stopping a movement (Swann et al., 2009; Wagner et al., 2017; Schaum et al., 2020).

Recent literature suggests that the cortical β-rhythm is by nature characterised by short-lasting, transient bursts and only appears to change in sustained amplitude if it is averaged over multiple trials (Feingold et al., 2015; Sherman et al., 2016). Various characteristics of β-bursts have been analysed in previous studies, such as rate, timing, probability, peak power, duration and interval time (Feingold et al., 2015; Sherman et al., 2016; Shin et al., 2017; Tinkhauser et al., 2017; Errington et al., 2020; Hannah et al., 2020; Jana et al., 2020; Seedat et al., 2020; Wessel, 2020). Three recent studies have investigated the role of β-bursts in human response inhibition using electroencephalography (EEG) and have

reported mixed findings. Wessel (2020) found a larger fronto-central β-burst rate for successful compared to failed action stopping whereas Jana et al. (2020) did not find a significant difference. Another study reported that earlier right frontal β-bursts were associated with faster *CancelTime*, a measure of stopping latency using electromyography (EMG) (Hannah et al., 2020).

Here, we aimed to further quantify the role and nature of β-bursts in human response inhibition in two large Stop Signal Task EEG datasets (n = 419) by using machine learning to search a wide range of spatial and temporal features. First, we sought to optimise the β-burst detection method. In a second step, we conducted two different single-trial analyses to (1) classify successful vs. failed Stop trials and (2) predict an individual's SSRT. We investigated two β-burst characteristics: rate and volume (a composite measure of burst duration, frequency span and amplitude) and also compared them to averaged normalised β-power. Given the high dimensional nature of the data, we employed a machine-learning approach, employing best-practice for quantifying the generalisability of our results (Poldrack et al., 2019). We trained the models on 60% of our data (internal validation set; n = 130) and subsequently applied the resulting models on the remaining, unseen 40% of our data (holdout validation set; n = 88). We then aimed to validate results on an external dataset (n = 201; Wessel (2020)). We hypothesised that β-burst features will be more predictive than averaged normalised β-power for classifying successful vs. failed Stop trials, and for predicting individual SSRTs. We also hypothesised that β-bursts over the right frontal scalp region would be most predictive.

Materials and Methods

Participants

The internal and holdout validation dataset consisted of 282 healthy adult human volunteers (age: 35.03 ± 14.72 years (mean ± SD); 175 female) who were pooled from four studies conducted in University College Dublin and Trinity College Dublin, Ireland. The studies were approved by the ethics committees of University College Dublin School of Psychology and Trinity College Dublin School of Psychology. Participants provided written informed consent. The raw data from this dataset formed the basis of another study that did not focus on spectral properties (Rueda-Delgado et al., 2019). The inclusion and exclusion criteria of each project are available in the Supplementary Material in Rueda-Delgado et al. (2019).

The external validation dataset consisted of 214 healthy adult human volunteers (age: 22.51 ± 6.67 years (mean ± SD); 121 female). This dataset was published as part of another study investigating β-bursts in response inhibition (Wessel, 2020) and is openly available on the Open Science Framework (https://osf.io/v3a78/).

Task

The task used for this study is described in detail in Rueda-Delgado et al. (2019). Briefly, participants underwent an adaptive visual SST. Each trial lasted 1000 ms and was preceded by a fixation cross (1000 ms duration). Participants were then presented with arrows pointing either to the right or left (Go stimulus) and they were instructed to respond with their right or left index finger, respectively, as fast as possible via an Xbox 360 game controller. In one of four Go trials, the Go stimulus was followed by an arrow pointing upwards (Stop signal) after a varying stop-signal delay (SSD). The participants were instructed to inhibit their button press on these Stop trials. The SSD was adjusted by a tracking algorithm, aiming to achieve a task difficulty resulting in 50% successful and 50% failed Stop trials. After a successful Stop trial, the SSD was increased, making the task harder and after a failed Stop trial, the SSD was decreased, making the task easier. The initial SSD was 250 ms and was subsequently

adjusted using a double-limit algorithm (see Richards et al., 1999). The SSD could vary between 50 ms and 450 ms. Following a Stop trial, the subsequent SSD value was chosen randomly between the current SSD and a pair of limits (higher or lower, as appropriate). These limits were designed to converge on the SSD that produced a 50% success rate and to be robust to fluctuations on individual trials. If a participant responded to the Go stimulus before Stop signal presentation, then the SSD was decreased for subsequent trials. The task consisted of 135 Go trials and 45 Stop trials and was presented in 3 blocks of 60 trials.

Participants in the external dataset performed a slightly different version of the SST which is described in Wessel (2020). The main differences were a higher Stop trial probability (0.33), a different visual Stop signal (arrow turned from white to red), different response buttons (q and p buttons on a QWERTY keyboard), the tracking algorithm (+/- 50 ms) was implemented independently for left- and rightward Go stimuli, and a larger number of trials (6 blocks of 50 trials).

SSRT analysis

The SSRT was calculated using the integration method with replacement of Go omissions by the maximum RT (Verbruggen et al., 2019). All Go trials were included in the Go RT distribution, including Go trials with choice errors. Premature responses on failed Stop trials were included when calculating the probability of responding on a Stop trial (p(respond|signal)) and mean SSD. Participants with SSRT < 125 ms and > 98th percentile of total SSRT distribution (303.7 ms) were excluded from the analysis (n = 47). The same criteria were applied to the external dataset (SSRT < 125 ms and > 98th percentile SSRT distribution (385.7 ms)) and 11 participants were excluded.

EEG recording and preprocessing

64-channel EEG data in the 10-5 system were recorded during the SST in a soundproof, darkened room using the ActiveTwo Biosemi system. Four additional electrodes recorded the electrooculogram from ~2 cm below the eyes (vertical eye movements) and from

the outer canthi (horizontal eye movements). Two further electrodes recorded from bilateral mastoids.

EEG data were digitised with a sampling rate of 512 Hz. EEG data preprocessing was carried out using the EEGLAB toolbox (Delorme and Makeig, 2004; http://sccn.ucsd.edu/eeglab) in conjunction with the Fully Automated Statistical Thresholding for EEG artefact Rejection plug-in (FASTER; Nolan et al., 2010; http://sourceforge.net/projects/faster). The data were initially bandpass filtered between 0.1 Hz and 95 Hz, notch filtered at 50 Hz and average referenced across all scalp electrodes. Data were subsequently epoched from 500 ms prior to Go/Stop stimulus onset to 2000 ms after Go/Stop stimulus onset for Go trials and Stop trials, respectively. 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 by trained raters to ensure good quality and that any remaining noisy data were removed. Specifically, trained raters identified any remaining artefacts in independent components (e.g., eyeblinks), epochs containing idiosyncratic muscle/movement or transient electrode artifacts, high amount of alpha and interpolated any channels that were noisy throughout all epochs of a participant. Datasets with a large amount of artefacts (e.g., due to chewing) and/or with >40% of epochs removed were excluded from further analysis. The remaining datasets were further checked for noisy data by calculating each dataset's event-related potential (ERP) and calculating z-scores across all ERPs. Datasets with z-scores > 15 were excluded. After exclusion of these, new z-scores were calculated and datasets with z-scores > 3 were excluded from further analysis. Fifteen participants were rejected from further analysis due to these artefacts or noisy data (in accordance with the exclusion criteria in Rueda-Delgado et al., 2019). The remaining 220 participants were subjected to further analysis. Details for EEG recording and preprocessing steps of the external dataset can be found in the original publication (Wessel, 2020). The external data was recorded using a different 62-channel array. We therefore interpolated the

channel locations from the preprocessed external data to match our 64-channel locations using the EEGLAB function *pop_interp*.

After channel interpolation, both internal and external EEG data were analysed identically. The analysis was applied to individual successful and failed Stop trial epochs (-500 ms to +2000 ms with respect to the Stop signal) using custom scripts in MATLAB 2017b and 2018b (Mathworks, USA). After preprocessing, EEG data were transformed using the current source density method (CSD; Kayser and Tenke, 2006; https://psychophysiology.cpmc.columbia.edu/software/CSDtoolbox/index.html) which is a reference-free montage to attenuate the effect of volume conduction in scalp EEG.

Time-frequency transformation

For all epochs, 2-dimensional representations of each electrode's time-frequency were estimated using a complex Morlet wavelet (range of logarithmically spaced 4-10 cycles for 15 linearly spaced frequencies across 15-29 Hz). The squared magnitude of the convolved data was calculated to obtain power estimates. β-Bursts were extracted from time-frequency power estimates without baseline normalisation. Baseline normalisation was separately performed to compute averaged normalised β-power. Time-frequency power estimates were converted to decibel (dB) using a 100 ms baseline prior to the Stop signal. β-Power was estimated over whole epochs and was only later subdivided into smaller time bins. It therefore does not matter if time bin sizes are smaller than the β-cycle length.

B-burst detection

B-burst detection was performed using the same general method as described in Wessel (2020), originally reported in Shin et al. (2017). For each time-frequency power matrix, local maxima were detected using the MATLAB function *imregionalmax*, from -25 ms to +1000 ms with respect to the Stop signal. The *imregionalmax* function identifies data with higher values that are surrounded by data with lower values (starting at -25 ms is necessary to inspect lower values for maxima located close to 0 ms). β-Bursts were then defined as local maxima

that exceeded a defined threshold. While Wessel (2020) and Shin et al. (2017) employed a β-burst detection threshold of 6x median power calculated from -500 ms to +1000 ms with respect to the Stop signal (across all trials per subject), we used machine learning to search across six different thresholds ranging from 1x to 6x median power of each individual time-frequency power matrix. We also tested two different baselines to calculate median power from, either using the whole epoch including stopping (-500 ms to +1000 ms with respect to the Stop signal) or using 300 ms during the fixation cross period. The purpose of this was to establish an optimal burst detection method for detecting β-bursts in human EEG that yielded the highest predictive ability for the behaviour under investigation (the 6x median power threshold employed by Shin et al. (2017) was decided following a search procedure through a range of thresholds based on pre-stimulus β-power derived from murine local field potentials and human magnetoencephalography (MEG)).

β-features extraction

The time-frequency power matrices of each participant were epoched from -125 ms to +100 ms with respect to each participant's SSRT, calculated for each of the 64 scalp electrodes. For each individual 225 ms time-frequency power matrix, features were individually extracted for either 25 ms, 45 ms or 75 ms time bins (i.e., 9, 5 or 3 bins, respectively). Three types of beta-band activity features were extracted in each time bin (**Figure 1**): (1) *β-Burst rate*; the sum of the number of supra-threshold bursts (2) *β-Burst volume*; the area under the curve of supra-threshold datapoints, individually calculated for each frequency and subsequently summed up over all frequencies within each time bin to obtain volume (β-burst volume was calculated per time bin, and not per burst, so that each trial would have the same number of features for inclusion in the regression models) (3) *Normalised β-power*; the mean of the baseline normalised power estimates.

Data preparation for the machine learning analysis

Our internal dataset (n = 218) was divided into two groups: the *internal* validation set (n = 130) and the *holdout* validation set (n = 88). In order to balance the internal and holdout

validation sets with respect to gender, participants were assigned to each set randomly but the assignment was iterated until the female:male ratio fell within the range 1.545:1 to 1.655:1 in each set. Data from Wessel (2020) served as the external validation set (n = 201).

The goal of the analysis was to predict individual trial outcomes (successful or failed stopping) and, separately, individual SSRTs from EEG data. Thus, each trial was regarded as an independent observation. This resulted in 2546 successful Stop trials and 2087 failed Stop trials for the internal validation set, 1814 successful Stop trials and 1454 failed Stop trials for the holdout validation set and 7809 successful Stop trials and 6449 failed Stop trials for the external validation set. For each trial, we obtained a value for each electrode and time bin for each of the three β-features as described in *'B-features extraction'*.

Experimental design and statistical analysis

Behavioural analysis

In addition to the SSRT analysis described earlier, means and standard deviations were extracted for each subject for the following behavioural SST measures: Go trial reaction time (RT), failed Stop trial RT, SSD, number of successful Stop trials, number of failed Stop trials, probability of successful stopping, probability of Go omissions, probability of choice errors. These measures were compared between the internal and the holdout as well as the external validation sets using a two sample t-test. Go trial RT and failed Stop trial RT were compared within each validation set using a paired t-test and effect sizes were estimated using Cohen's d. Participants were excluded from analysis if their failed Stop trial RT was larger than Go trial RT (Bissett et al., 2019; Verbruggen et al., 2019), violating an assumption of the horse-race model.

Machine learning analysis of the internal validation set

The detailed method for the machine learning used in this study is described in Rueda-Delgado et al. (2019) and is briefly explained here. We used logistic regression to classify successful vs. failed Stop trials ('*Stop trial classification*') and, separately, linear regression to

predict individual SSRT for each trial ('*SSRT prediction*'). For both analyses we used a form of penalised regression (specifically, the *Elastic Net*; Zou and Hastie, 2005) to attenuate overfitting (Jollans et al., 2019). The Elastic Net constrains the size of regression coefficients and can also set regression coefficients to zero (i.e., it is a form of feature selection). A modification for this study was that the range of the two hyperparameters (alpha and lambda) were expanded, and the search grid altered. For logistic regression, alpha ranged from 0.01 to 1 across 10 logarithmically-spaced values and lambda ranged between 0.2 to 1 across 10 linearly-spaced values (i.e., a search grid of 100 parameter-pair values). For linear regression, alpha ranged from 0.3 to 1 in 10 linearly-spaced values and lambda ranged from 3 to 7 in 10 linearly-spaced values. Model hyperparameters were determined using nested cross-validation. In contrast to Rueda-Delgado et al. (2019), here we used 5-fold rather than 10-fold cross-validation. The main- and subfolds were assigned across subjects. In each subfold, the data were z-scored and then Winsorized (i.e., values > |3| were replaced with a value of +/-3). The entire analysis was iterated 100 times for each model, with a new random assignment of training and test sets on each iteration.

We adjusted for age and sex by adding them as additional features to each model (Rubia et al., 2013; Hsieh and Lin, 2017). For the internal and holdout validation sets we additionally controlled for the four different data collection projects by adding 3 dummy-coded regressors.

To quantify the model performance, the entire procedure was repeated using random-label permutation (i.e., to generate a 'null' model) where trial outcomes were randomly shuffled (i.e., the successful or failed Stop trial label for the logistic regression, or SSRT for the linear regression). Note that covariates were not shuffled for null models, thereby quantifying their contribution to the actual model (Dinga et al., 2020). That is, any increase in performance between the actual and null models can be attributed to the contribution of the EEG data. The performance metrics for the logistic regression consisted of the area under the receiver operating characteristic curve (AROC; chance = 0.5) and Brier score (i.e., mean squared error

between predicted probabilities and expected values) (Dinga et al., 2019). The performance metrics for the linear regression consisted of mean absolute error (MAE) and coefficient of determination (R^2) (Poldrack et al., 2019). Results reported for the internal validation are mean values across all 100 iterations of the analysis. To assess the significance of the actual model predictions we calculated *P* which is equal to the fraction of iterations on which the performance metrics of the null model were more predictive than the performance metrics of the actual model (cf., Greene et al., 2018).

We sought to first determine the optimal β-burst detection method by utilizing the internal validation dataset. For this, we tested two parameters: (1) six different thresholds (1x, 2x, 3x, 4x, 5x or 6x median power) and (2) two different baselines to calculate median power of, either using the whole epoch including stopping (-500 ms to +1000 ms with respect to the Stop signal) or using 300 ms during the fixation cross period. In addition to these burst detection parameters, we also tested a third parameter: (3) three different time bin sizes to partition the 225 ms time-frequency power matrix (9x25 ms, 5x45 ms or 3x75 ms). We tested all 36 possible combinations of these three parameters using a “burst features” model that consisted of β-burst rate and β-burst volume. Depending on the time bin size, this model consisted of 1152, 640 or 384 spatio-temporal features (2 burst features x 64 channels x 9, 5 or 3 time bins). The analysis was iterated 100 times for each combination. We ran this analysis both for the Stop trial classification and for the SSRT prediction. We then determined the most predictive parameter combination across both analyses.

Following determination of the optimal β-burst detection method and time bin size, the internal validation analysis was tested on five final main models using the most predictive parameter combination (see Results): (1) *Full features* model (β-burst rate, β-burst volume, normalised β-power; 1728 spatio-temporal features), (2) *Burst features* model (β-burst rate, β-burst volume; 1152 spatio-temporal features) and three Single feature models (3) *B-burst rate* (576 spatio-temporal features), (4) *B-burst volume* (576 spatio-temporal features) and (5) *Normalised β-power* (576 spatio-temporal features). The performance metrics of the five main

models were compared using a one-way repeated measures ANOVA test (omnibus p threshold = 0.05) and a post hoc t-test with Bonferroni correction to determine if the performance metric means of the models were significantly different. Effect sizes were estimated using eta-squared (η^2).

Validation analysis of the holdout and external validation sets

The results of the five main models from the internal validation were tested on the holdout and external validation sets. This involved applying the optimum burst detection parameters and time bin size as well as regression weights from the internal set to the external and holdout sets (neither the optimum burst detection parameters and time bin size nor the regression weights could be derived from the holdout and external sets as this would violate the independence between training and test sets (Poldrack et al., 2019)). The holdout and external validation sets were scaled to the internal validation data by z-scoring and Winsorizing using the mean and standard deviation of the internal validation set. The mean SSRTs of the holdout and external validation sets were 5 ms and 51 ms longer, respectively. Therefore, for the SSRT prediction, the SSRTs in the holdout and external validation sets were scaled to the same mean and standard deviation of the internal validation set SSRT. For each model, the average regression coefficients from the 100 iterations of the internal validation were then applied to the holdout and external validation data. The same metrics to evaluate model performance were utilised for holdout and external validation sets. This procedure was again repeated using random-label permutation (null model).

Data interpretation

To further interrogate the data and identify the most predictive features in successful models, we calculated the 'selection frequency' of each individual feature. This was calculated by summing each feature's non-zero count in each main fold and subsequently averaging across the 100 iterations. Features were then ranked first by selection frequency and second by the absolute regression coefficient. *Solely to aid interpretation*, we applied spatial and temporal feature reduction to summarise the results. We created boxplots by grouping data

from 64 electrodes into 8 different regions (**Table 1**) and amalgamated the data into 3 time bins (-125 ms to -50 ms before SSRT, -50 ms to +25 ms around SSRT, +25 ms to +100 ms after SSRT).

Code accessibility

Custom written scripts can be downloaded on the Open Science Framework at https://osf.io/4tznd/.

Results

Behavioural results

The behavioural data and statistical comparisons of all three validation sets are displayed in **Table 2**. Failed Stop RT was larger than Go RT for 2 participants in the internal validation set, 0 participants in the holdout validation set and 2 participants in the external validation set. These data were therefore excluded from further analysis. Within each validation set, failed Stop RTs were significantly faster than Go RTs (internal: $t(129) = 28.51$, $p = 1.6*10^{-57}$, $d = 2.50$; holdout: $t(87) = 23.74$, $p = 9.0*10^{-40}$, $d = 2.53$; external: $t(200) = 33.54$, $p = 4.6*10^{-84}$, $d = 2.37$), in accordance with the horse-race model assumption. There were no significant differences in the behavioural measures between internal and holdout validation sets. There were significant differences between every behavioural SST measure except the SSD when comparing internal and external validation sets, which used slightly different versions of the SST.

Machine learning results

Optimal β-burst detection method and time bin size

The results of the analyses to determine the optimal β-burst detection method and optimal time bin size are displayed in **Figure 2**. The differences in performance metrics across the 36 models were small. Both the Stop trial classification and the SSRT prediction yielded most predictive results using 1x median power as the burst detection threshold. However, after visual inspection of the data it was apparent that thresholds below 2x median power occasionally mis-registered background β-oscillatory activity as bursts. For interpretability (i.e., to distinguish between averaged β-power and β-burst events), therefore, we used 2x median power as the burst detection threshold for both Stop trial classification and SSRT prediction. Further, there was no difference between using either baseline method to calculate median power. This was further supported when correlating the data coming from either baseline which showed that these data correlated (for β-burst rate: $r = 0.83$, for β-burst volume:

r = 0.99). Therefore, we chose -500 ms to +1000 ms as a baseline for compatibility with previous studies (i.e., Shin et al., 2017; Wessel, 2020). Finally, there was little difference among different time bin sizes. We therefore chose the 25-ms time bin size because it afforded greater temporal precision.

Stop trial classification (logistic regression results of five main models)

Figure 3A displays the results for the Stop trial classification for all five main models plus corresponding null models for internal, holdout and external validation sets.

We aimed to find the features that best classified successful vs. failed Stop trials. The mean AROCs of the five models were unequal according to a one-way repeated measures ANOVA ($F(4,396) = 1185$, $p = 8.1*10^{-219}$, $\eta^2 = 0.92$). Post hoc comparisons (Bonferroni corrected at 0.05/10 = 0.005) indicated that mean AROC of the single feature β-burst volume model (mean AROC = 0.57, mean Brier score = 0.244) was significantly larger than mean AROC of the full features model (mean AROC = 0.56, mean Brier score = 0.245) and of the burst features model (mean AROC = 0.56, mean Brier score = 0.246). The full features model and the burst features model were the only two models which were not significantly different from each other. The single feature β-burst volume model outperformed the null model on every iteration ($P_{AROC} = 0$, $P_{Brier} = 0$). The holdout (mean AROC = 0.58, mean Brier score = 0.243) and the external (mean AROC = 0.57, mean Brier score = 0.244) validations yielded similar results when applying the regression coefficients from the internal validation single feature β-burst volume model. As a next step, we compared the mean AROCs of the three single feature models and the post hoc test revealed that they were all significantly different from each other. The β-burst rate model showed the worst classification accuracy (mean AROC = 0.51, mean Brier score = 0.248) although generally outperformed the null model ($P_{AROC} = 0.24$, $P_{Brier} = 0.25$). The null model results of the Stop trial classification (AROCs around 0.5, i.e., chance level) indicate that the model accuracies did not depend on the inclusion of the covariates.

SSRT prediction (linear regression results of five main models)

Figure 3B displays the results for the SSRT prediction for all five main models plus corresponding null models for internal, holdout and external validation sets.

The mean MAEs of the five models were unequal according to a one-way repeated measures ANOVA ($F(4,396) = 36.23$, $p = 8.3*10^{-26}$, $\eta^2 = 0.27$). Post hoc comparisons (Bonferroni corrected at 0.05/10 = 0.005) indicated that mean MAEs of the full features model (mean MAE = 28.77, mean $R^2 = 0.02$), the burst features model (mean MAE = 28.78, mean $R^2 = 0.02$) and the single feature β-burst volume model (mean MAE = 28.82, mean $R^2 = 0.01$) were not significantly different, however, they were significantly more predictive than the single feature models β-burst rate and normalised β-power. There was a modest performance improvement in comparison to the null model: full features model ($P_{MAE} = 0.27$, $P_{R2} = 0.28$), burst features model ($P_{MAE} = 0.29$, $P_{R2} = 0.23$) and single feature β-burst volume model ($P_{MAE} = 0.34$, $P_{R2} = 0.36$). The respective holdout and the external validations yielded similar results. The mean MAEs of the two remaining single features models, β-burst rate (mean MAE = 29.18, mean $R^2 = 0.00$) and normalised β-power (mean MAE = 29.17, mean $R^2 = 0.01$), were not significantly different. Neither model outperformed the null models (β-burst rate: $P_{MAE} = 0.52$, $P_{R2} = 0.54$); normalised β-power: $P_{MAE} = 0.49$, $P_{R2} = 0.49$). The coefficients of determination (R^2) of the null models of the SSRT prediction are slightly above zero, showing an additional, but quite small, effect of the covariates on the SSRT prediction.

Spatial and temporal dynamics of β-burst volume

The β-burst volume model could predict Stop trial classification. We therefore present the spatial and temporal dynamics for this model in more detail. The models did not robustly predict SSRT and therefore we do not interpret their detailed spatial and temporal dynamics here.

Stop trial classification

Figure 4 shows the spatial and temporal dynamics of β-burst volume for classifying successful vs. failed Stop trials. Larger β-burst volume in right frontal, fronto-central and

bilateral sensorimotor was predictive of successful stopping -125 ms to -50 ms before SSRT. In centroparietal and occipital sites smaller β-burst volume was predictive of successful stopping (from the other perspective; larger β-burst volume was predictive of failed stopping). A similar pattern was observed from -50 ms to +25 ms around the SSRT where smaller β-burst volume in centroparietal and occipital sites was predictive of successful stopping. After the SSRT (+25 ms to +100 ms), larger β-burst volume in bilateral sensorimotor areas was predictive of successful stopping.

Spatial and temporal dynamics of response inhibition across validation sets

Figure 5 presents topoplots of successful minus failed Stop trials. The three main rows show the data for each β-feature (i.e., β-rate, burst volume and normalised β-power). Topoplots are shown for each validation set (i.e., internal, holdout and external) separately in sub-rows. The pattern of activation across time for β-burst volume and normalised β-power is broadly similar across all three validation sets. B-burst rate is less consistent across validation sets due to the low frequency of burst detection.

Discussion

Here, we investigated the role of β-bursts for successful response inhibition and their impact upon the speed of stopping (SSRT). We used machine learning to search across a wide span of temporal and spatial features on a large dataset, and employed best practice for validation, including external validation. Results partially supported our hypothesis. B-burst volume, but not rate, was the superior predictor for Stop trial classification. However, normalised β-power was a better predictor than β-burst rate. Stop trial classification (success vs. fail) generalised to an external dataset, but prediction of individual SSRT did not.

To the best of our knowledge, this study is the first to introduce β-burst volume as a key behaviourally relevant measure of human action stopping. Burst volume is a composite measure, capturing in single metric features previously studied in isolation: burst peak amplitude (Feingold et al., 2015; Sherman et al., 2016; Shin et al., 2017; Tinkhauser et al., 2017; Little et al., 2019), burst duration (Feingold et al., 2015; Sherman et al., 2016; Shin et al., 2017; Tinkhauser et al., 2017; Little et al., 2019), and burst frequency span (Shin et al., 2017). Burst amplitude increases in proportion with burst duration (Tinkhauser et al., 2017), suggesting a single generating mechanism and therefore suitable to be combined. The inclusion of a third dimension (frequency span) allows a compact assessment across the entire set of beta-range frequencies (Zich et al., 2020). The β-rhythm likely inhibits neural processing (Sherman et al., 2016) and the brevity of β-bursts may be critical for intact messaging in the brain (Feingold et al., 2015). In patients with Parkinson's disease shorter β-burst duration was related to improved motor function (Tinkhauser et al., 2017). Taken together, these findings suggest burst volume better captures the multifaceted nature of the β-bursts rather than merely their presence or rate.

Machine learning facilitates a data-driven approach, searching a large area of scalp and wide temporal window. Increased right frontal β-burst volume from -100 ms to -50 ms before SSRT was predictive of successful stopping. This is spatially and temporally consistent with β-rhythm findings using electrocorticography (ECoG; Swann et al., 2009, 2012), EEG

(Wagner et al., 2017; Hannah et al., 2020; Jana et al., 2020) and MEG (Schaum et al., 2020). The rIFC is thought to act as a brake for motor output that can be triggered by a unexpected event (Aron et al., 2014; Schaum et al., 2020). We used scalp EEG, which cannot identify the source generator. However, studies using ECoG (Swann et al., 2009, 2012), MEG/functional magnetic resonance imaging (fMRI; Schaum et al., 2020) and fMRI-guided repetitive transcranial magnetic stimulation (Sundby et al., 2021) have empirically demonstrated the link between right frontal scalp activity and rIFC. The larger β-burst volume recorded from right frontal scalp therefore likely indicates rIFC-related motor braking triggered by the Stop signal. Additionally, we found that early increased bilateral sensorimotor β-burst volume from -125 ms to the time of SSRT predicted successful stopping. This is consistent with β-burst studies (Jana et al., 2020; Wessel, 2020) that related this finding to global motor suppression during non-selective movement stopping (Badry et al., 2009; Wessel and Aron, 2013). An unexpected result occurred in most time bins: a relative decrease in β-burst volume in centroparietal and occipital areas for successful Stop trials, potentially indicating relatively more focused activity in successful stopping. Finally, +75 to +100 ms after the SSRT, there was a clear sensorimotor signature where decreasing β-burst volume predicted failed stopping, consistent with the well-known β-desynchronisation over motor cortex during motor preparation, which occurs during failed Stop trials where a button press takes place (Zhang et al., 2008; Swann et al., 2009; Fonken et al., 2016).

In contrast to the findings relating to burst volume, burst rate was not a robust predictor of stopping behaviour. Similarly, Jana et al. (2020) found no difference in right frontal β-burst rate between successful vs. failed Stop trials, but Wessel (2020) (i.e., our external dataset) did at a single fronto-central electrode (FCz). The external dataset SST had a higher Stop trial probability than the internal dataset (0.33 vs. 0.25). Increased Stop trial probability may recruit a more proactive response inhibition (Castro-Meneses et al., 2015). Some studies suggested that pre-SMA is more strongly activated during proactive response inhibition (Sharp et al.,

2010; Hu et al., 2015), potentially explaining a more fronto-central signature in the external dataset (also cf., Leunissen et al., 2020).

Several methods for β-burst detection amidst background brain activity have been described (cf., Shin et al., 2017; Tinkhauser et al., 2017; Little et al., 2019; Seedat et al., 2020). There is no consensus on the methodology for optimal β-burst detection and therefore we tested several thresholds. A threshold of 2x median β-power was most sensitive for investigation of the brain's stopping process, suggesting β-burst information might be lost by using a higher detection threshold. Threshold selection is also relevant to the relationship between burst rate and volume. At higher thresholds, only the larger amplitude bursts will be identified and burst rate will be somewhat confounded with burst volume. Additionally, we tested two different baselines to calculate median power: there was no difference in results and therefore it appears either method can be used for burst threshold calculation.

Stopping behaviour can be described by formal computational models (cf., Bissett et al., 2019). For example, the Interactive Race (Boucher et al., 2007) and the Blocked Input (Logan et al., 2015) models both characterise Stop and Go processes as stochastic accumulators. For both models, the Stop unit should be active on every successfully inhibited trial (Logan et al., 2015; Errington et al., 2020) to either inhibit or block the Go unit. Despite evidence for β-involvement in stopping, studies to date have not shown that β-bursts are causally linked as a mechanism for the Stop unit. We consider a number of potential, not mutually exclusive, explanations (see also Errington et al., 2020; Wessel, 2020). First, infrequency of β-burst detection is due to the low signal-to-noise ratio (SNR) in EEG (e.g., Jana et al., 2020). However, Errington and colleagues demonstrated similar frequency of β-burst detection using intracranial electrodes in macaques, a method with better signal fidelity than EEG. It therefore seems unlikely that poor SNR is the sole reason for the low detection rate. Second, β-activity is produced by multiple brain sources that reflect different processes: our results clearly support this interpretation. B-activity topography changed over time, and different scalp regions were predictive of stopping at different times. Third, (as suggested by

Errington et al., 2020), some informative β-bursts may be subthreshold. Here, we utilised a training dataset to optimise burst detection threshold, which was lower than previous studies. There was an improvement in prediction accuracy using this lower threshold, but it did not approach the one-to-one mapping necessary to characterise the Stop unit. We suggest that brain activity needs to be recorded with improved SNR (perhaps recording many more trials for EEG). In addition, more sophisticated methods (e.g., Deep Learning (Abrol et al., 2021)) may improve detection of informative bursts and could better integrate information from multiple regions to adequately capture the complexity of β-band activity in stopping behaviour.

The internal validation for the SSRT prediction revealed a weak model that did not generalise to the holdout and external validation data. It is possible that β-band activity is unimportant for speed of stopping. Another potential explanation is that we relied on the stopping speed derived from a button press to measure SSRT. While this is a widely used measure in response inhibition research, it has its drawbacks (e.g., Skippen et al., 2020). Most importantly for our study, the SSRT is only a summary measure for each participant and does not provide single trial information. It is likely that using single trial EMG data to calculate *CancelTime* (Jana et al., 2020; also cf., Thunberg et al., 2020) as the stopping outcome measure would have yielded a more predictive model.

In conclusion, β-burst volume was superior to both β-burst rate and averaged normalised β-power for classifying successful response inhibition. To date, several studies have shown a statistical association between β-band activity and stopping (e.g., Wagner et al., 2017; Jana et al., 2020; Schaum et al., 2020; Wessel, 2020). Here, a machine learning approach reliably showed that β-bursts are implicated in human action stopping. The models significantly outperformed a random model suggesting a real effect of β-activity and the classification models were validated on unseen and independent external data. This first validation of neural EEG SST data on a holdout and external dataset therefore fulfils the requirement for establishing prediction (Poldrack et al., 2019). These results support the

emerging view that transient β-bursts are a more accurate representation of oscillatory β-activity in the brain, and suggest an important role for β-bursts in human response inhibition.

References

Abrol A, Fu Z, Salman M, Silva R, Du Y, Plis S, Calhoun V (2021) Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. Nat Commun 12:353.

Aron AR, Robbins TW, Poldrack RA (2014) Inhibition and the right inferior frontal cortex: One decade on. Trends Cogn Sci 18:177–185.

Badry R, Mima T, Aso T, Nakatsuka M, Abe M, Fathi D, Foly N, Nagiub H, Nagamine T, Fukuyama H (2009) Suppression of human cortico-motoneuronal excitability during the Stop-signal task. Clin Neurophysiol 120:1717–1723.

Bissett PG, Jones HM, Poldrack RA, Logan GD (2019) Severe violations of independence in response inhibition tasks. PsyArXiv.

Boucher L, Palmeri TJ, Logan GD, Schall JD (2007) Inhibitory control in mind and brain: An interactive race model of countermanding saccades. Psychol Rev 114:376–397.

Castro-Meneses LJ, Johnson BW, Sowman PF (2015) The effects of impulsivity and proactive inhibition on reactive inhibition and the go process: Insights from vocal and manual stop signal tasks. Front Hum Neurosci 9:1–12.

Delorme A, Makeig S (2004) EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21.

Dinga R, Penninx BWJH, Veltman DJ, Schmaal L, Marquand AF (2019) Beyond accuracy: Measures for assessing machine learning models, pitfalls and guidelines. bioRxiv:743138.

Dinga R, Schmaal L, Penninx BWJH, Veltman DJ, Marquand AF (2020) Controlling for effects of confounding variables on machine learning predictions. bioRxiv:255034.

Errington SP, Woodman GF, Schall JD (2020) Dissociation of Medial Frontal β-Bursts and Executive Control. J Neurosci 40:9272–9282.

Feingold J, Gibson DJ, Depasquale B, Graybiel AM (2015) Bursts of β-oscillation

differentiate postperformance activity in the striatum and motor cortex of monkeys performing movement tasks. Proc Natl Acad Sci U S A 112:13687–13692.

Fonken YM, Rieger JW, Tzvi E, Crone NE, Chang E, Parvizi J, Knight RT, Krämer UM (2016) Frontal and motor cortex contributions to response inhibition: evidence from electrocorticography. J Neurophysiol 115:2224–2236.

Greene AS, Gao S, Scheinost D, Constable RT (2018) Task-induced brain state manipulation improves prediction of individual traits. Nat Commun 9:2807.

Hannah R, Muralidharan V, Sundby KK, Aron AR (2020) Temporally-precise disruption of prefrontal cortex informed by the timing of β-bursts impairs human action-stopping. Neuroimage 222:117222.

Hsieh S, Lin YC (2017) Stopping ability in younger and older adults: Behavioral and event-related potential. Cogn Affect Behav Neurosci 17:348–363.

Hu S, Ide JS, Zhang S, Li C-SR (2015) Anticipating conflict: Neural correlates of a Bayesian belief and its motor consequence. Neuroimage 119:286–295.

Jana S, Hannah R, Muralidharan V, Aron AR (2020) Temporal cascade of frontal, motor and muscle processes underlying human action-stopping. Elife 9:e50371.

Jollans L, Boyle R, Artiges E, Banaschewski T, Desrivières S, Grigis A, Martinot JL, Paus T, Smolka MN, Walter H, Schumann G, Garavan H, Whelan R (2019) Quantifying performance of machine learning methods for neuroimaging data. Neuroimage 199:351–365.

Kayser J, Tenke CE (2006) Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. Clin Neurophysiol 117:348–368.

Leunissen I, van Steenkiste M, Heise KF, Monteiro TS, Dunovan K, Mantini D, Coxon JP, Swinnen SP (2020) Effects of beta- and gamma-band rhythmic stimulation on motor inhibition. bioRxiv:422006.

Lijffijt M, Kenemans JL, Verbaten MN, Van Engeland H (2005) A meta-analytic review of stopping performance in attention-deficit/ hyperactivity disorder: Deficient inhibitory

608 motor control? J Abnorm Psychol 114:216–222.
609 Little S, Bonaiuto J, Barnes G, Bestmann S (2019) Human motor cortical β-bursts relate to
610 movement planning and response errors. PLoS Biol 17:e3000479.
611 Logan GD, Cowan WB (1984) On the ability to inhibit thought and action: A theory of an act
612 of control. Psychol Rev 91:295–327.
613 Logan GD, Yamaguchi M, Schall JD, Palmeri TJ (2015) Inhibitory control in mind and brain
614 2.0: Blocked-input models of saccadic countermanding. Psychol Rev 122:115–147.
615 Luijten M, Littel M, Franken IHA (2011) Deficits in inhibitory control in smokers during a
616 Go/Nogo task: An investigation using event-related brain potentials. PLoS One
617 6:e18898.
618 Nolan H, Whelan R, Reilly RB (2010) FASTER: Fully Automated Statistical Thresholding for
619 EEG artifact Rejection. J Neurosci Methods 192:152–162.
620 Poldrack RA, Huckins G, Varoquaux G (2019) Establishment of Best Practices for Evidence
621 for Prediction: A Review. JAMA Psychiatry 77:534–540.
622 Richards JB, Zhang L, Mitchell SH, de Wit H (1999) Delay or probability discounting in a
623 model of impulsive behavior: effect of alcohol. J Exp Anal Behav 71:121–143.
624 Rubia K, Lim L, Ecker C, Halari R, Giampietro V, Simmons A, Brammer M, Smith A (2013)
625 Effects of age and gender on neural networks of motor response inhibition: From
626 adolescence to mid-adulthood. Neuroimage 83:690–703.
627 Rueda-Delgado LM, Halloran LO, Enz N, Ruddy KL, Kiiski H, Bennett M, Farina F, Jollans L,
628 Vahey N, Whelan R (2019) Brain event-related potentials predict individual differences
629 in inhibitory control. Int J Psychophysiol In Press.
630 Schaum M, Pinzuti E, Sebastian A, Lieb K, Fries P, Mobascher A, Jung P, Wibral M,
631 Tüscher O (2020) Cortical network mechanisms of response inhibition. bioRxiv:940841.
632 Seedat ZA, Quinn AJ, Vidaurre D, Liuzzi L, Gascoyne LE, Hunt BAE, O'Neill GC, Pakenham
633 DO, Mullinger KJ, Morris PG, Woolrich MW, Brookes MJ (2020) The role of transient
634 spectral 'bursts' in functional connectivity: A magnetoencephalography study.
635 Neuroimage 209:116537.

Sharp DJ, Bonnelle V, De Boissezon X, Beckmann CF, James SG, Patel MC, Mehta MA (2010) Distinct frontal systems for response inhibition, attentional capture, and error processing. Proc Natl Acad Sci U S A 107:6106–6111.

Sherman MA, Lee S, Law R, Haegens S, Thorn CA, Hämäläinen MS, Moore CI, Jones SR (2016) Neural mechanisms of transient neocortical β-rhythms: Converging evidence from humans, computational modeling, monkeys, and mice. Proc Natl Acad Sci U S A 113:E4885–E4894.

Shin H, Law R, Tsutsui S, Moore CI, Jones SR (2017) The rate of transient β-frequency events predicts behavior across tasks and species. Elife 6:e29086.

Skippen P, Fulham WR, Michie PT, Matzke D, Heathcote A, Karayanidis F (2020) Reconsidering electrophysiological markers of response inhibition in light of trigger failures in the stop-signal task. Psychophysiology 57:e13619.

Sundby KK, Jana S, Aron AR (2021) Double-blind disruption of right inferior frontal cortex with TMS reduces right frontal β-power for action stopping. J Neurophysiol 125:140–153.

Swann N, Tandon N, Canolty R, Ellmore TM, McEvoy LK, Dreyer S, DiSano M, Aron AR (2009) Intracranial EEG Reveals a Time- and Frequency-Specific Role for the Right Inferior Frontal Gyrus and Primary Motor Cortex in Stopping Initiated Responses. J Neurosci 29:12675–12685.

Swann NC, Cai W, Conner CR, Pieters TA, Claffey MP, George JS, Aron AR, Tandon N (2012) Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: Electrophysiological responses and functional and structural connectivity. Neuroimage 59:2860–2870.

Thunberg C, Messel MS, Raud L, Huster RJ (2020) tDCS over the inferior frontal gyri and visual cortices did not improve response inhibition. Sci Rep 10:7749.

Tinkhauser G, Pogosyan A, Little S, Beudel M, Herz DM, Tan H, Brown P (2017) The modulatory effect of adaptive deep brain stimulation on β-bursts in Parkinson's disease. Brain 140:1053–1067.

664 Verbruggen F et al. (2019) Capturing the ability to inhibit actions and impulsive behaviors: A
665 consensus guide to the stop-signal task. Elife 8:e46323.
666 Wagner J, Wessel JR, Ghahremani A, Aron AR (2017) Establishing a Right Frontal B-
667 Signature for Stopping Action in Scalp EEG: Implications for Testing Inhibitory Control
668 in Other Task Contexts. J Cogn Neurosci 30:107–118.
669 Wessel JR (2020) β-Bursts Reveal the Trial-to-Trial Dynamics of Movement Initiation and
670 Cancellation. J Neurosci 40:411–423.
671 Wessel JR, Aron AR (2013) Unexpected events induce motor slowing via a brain
672 mechanism for action-stopping with global suppressive effects. J Neurosci 33:18481–
673 18491.
674 Wessel JR, Aron AR (2017) On the Globality of Motor Suppression: Unexpected Events and
675 Their Influence on Behavior and Cognition. Neuron 93:259–280.
676 Zhang Y, Chen Y, Bressler SL, Ding M (2008) Response preparation and inhibition: The role
677 of the cortical sensorimotor β-rhythm. Neuroscience 156:238–246.
678 Zich C, Quinn AJ, Mardell LC, Ward NS, Bestmann S (2020) Dissecting Transient Burst
679 Events. Trends Cogn Sci 24:784–788.
680 Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Stat Soc
681 Ser B Stat Methodol 67:301–320.

682 **Figures**

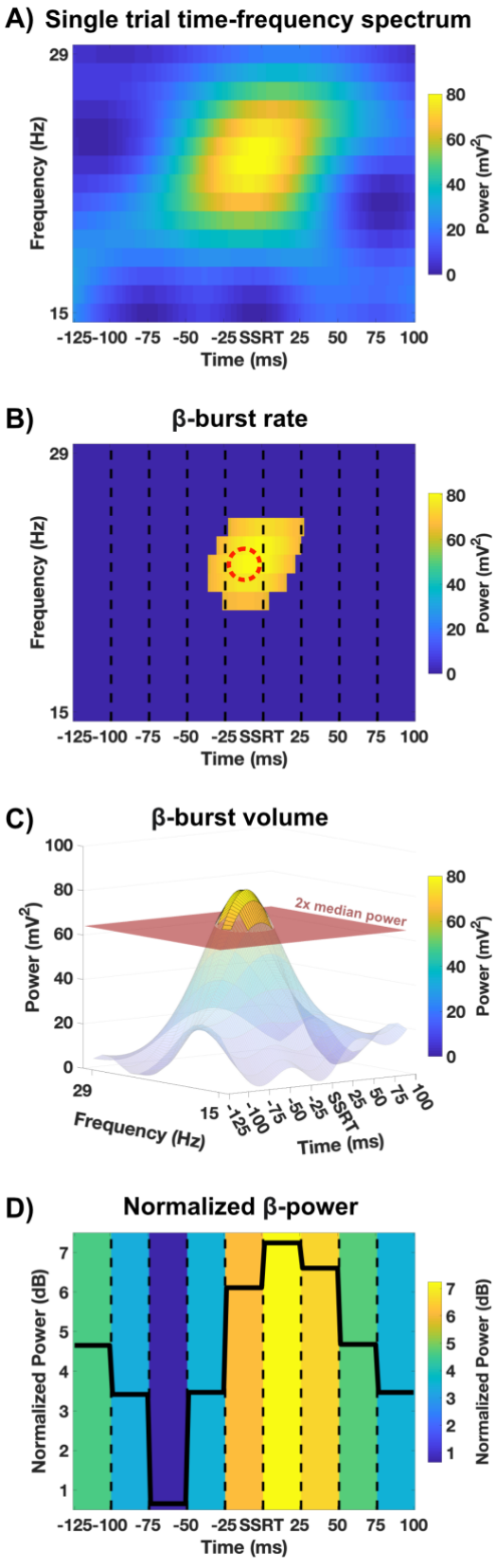
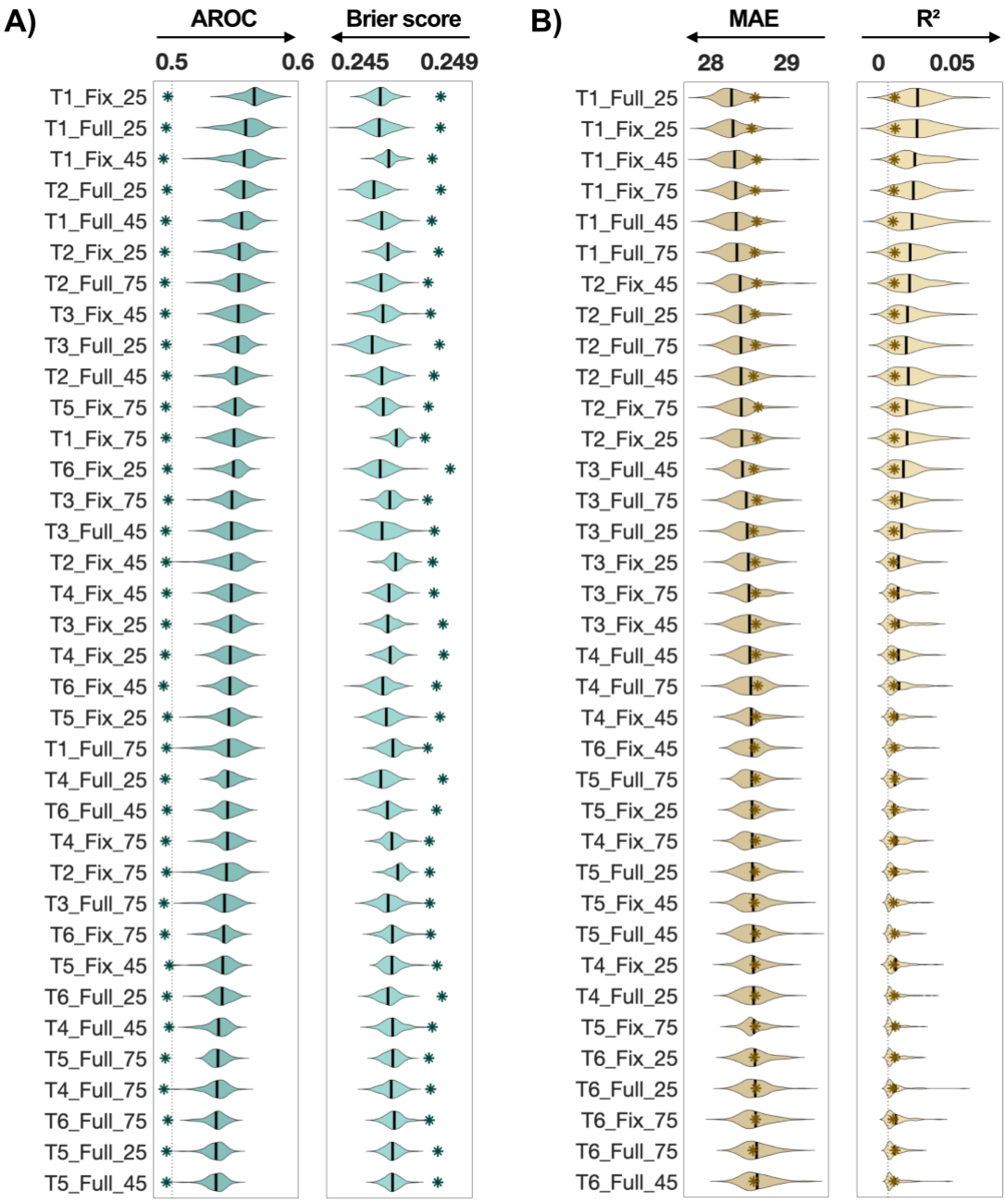

A) Single trial time-frequency spectrum
Frequency (Hz)
29
15
-125 -100 -75 -50 -25 SSRT 25 50 75 100
Time (ms)
80
60
40
20
0
Power (mV²)
B) β-burst rate
Frequency (Hz)
Time (ms)
Power (mV²)
C) β-burst volume
Power (mV²)
2x median power
Frequency (Hz)
Time (ms)
D) Normalized β-power
Normalized Power (dB)
7
6
5
4
3
2
1
Time (ms)

Figure 1. β-Feature extraction from a representative single trial. A) Each trial consisted of a 15 (frequencies) x 117 (time points; 225 ms, 512 Hz) time-frequency power matrix. The 117 time points were divided into nine 25-ms time bins, from -125 ms before SSRT to +100 ms after SSRT. β-bursts were defined as local maxima which exceeded a predefined threshold. **B)** β-Burst rate: Shown are all datapoints exceeding the threshold (2x median power). Each local maximum (indicated with red dashed circle) exceeding the threshold was counted as one β-burst for the time bin where it occured. **C)** β-Burst volume: All timepoints exceeding the threshold (2x median power) were included for the volume calculation. The volume was spread across the different time bins and frequencies. **D)** Normalised β-power: The normalised power values over all frequencies were averaged for each time bin (means per time bin are indicated with black line). Abbreviation: SSRT: Stop Signal Reaction Time.



Legend for model names:

Burst detection threshold:	**'T1'** = 1x median power threshold (etc. for 1-6x median power)
Baseline to calculate median power:	**'Full'** = -500 ms to +1000 ms with respect to Stop signal
	'Fix' = 300 ms during fixation cross
Time bin size:	**'25'** = 25 ms; **'45'** = 45 ms; **'75'** = 75 ms

Figure 2. Machine learning results for 'Optimal β-burst detection method and time bin size' (36 parameter combination models: burst detection threshold x baseline for median power calculation x time bin size). A) Logistic regression results over 100 iterations for the Stop trial classification are shown. Two performance metrics are reported; AROC: area under the receiver operating characteristics curve and Brier score. The models are sorted by AROC from most predictive to least predictive. Respective null models of each model are indicated with an asterisk symbol (*). **B)** Linear regression results over 100 iterations for SSRT prediction are shown. Two performance metrics are reported; MAE: mean absolute error and R^2: coefficient of determination. The models are sorted by MAE from most predictive to least predictive. Respective null models of each model are indicated with an asterisk symbol (*). Arrows below each performance metric indicate the direction for higher prediction.

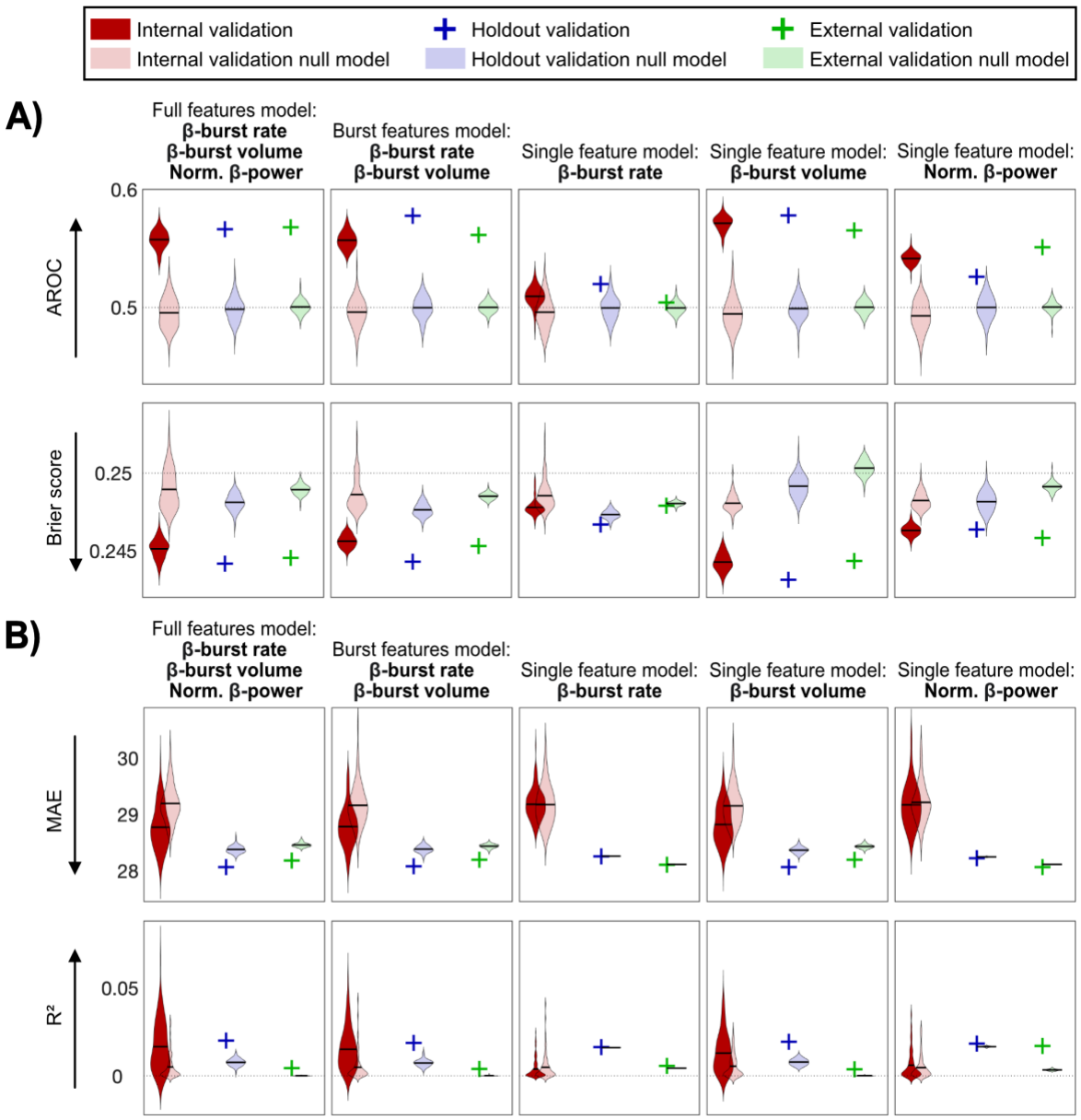

Internal validation
Internal validation null model
Holdout validation
Holdout validation null model
External validation
External validation null model
A)
Full features model:
β-burst rate
β-burst volume
Norm. β-power
Burst features model:
β-burst rate
β-burst volume
Single feature model:
β-burst rate
Single feature model:
β-burst volume
Single feature model:
Norm. β-power
AROC
0.6
0.5
Brier score
0.25
0.245
B)
MAE
30
29
28
R²
0.05
0

Figure 3. Machine learning results for internal, holdout and external validation and the corresponding null models of the 5 main models. A) Logistic regression results over 100 iterations for the Stop trial classification are shown. Two performance metrics are reported; AROC: area under the receiver operating characteristics curve and Brier score. **B)** Linear regression results over 100 iterations for the SSRT prediction are shown. Two performance metrics are reported; MAE: mean absolute error and R^2: coefficient of determination. Dotted lines represent chance level for the respective metric. Arrows beside each performance metric indicate the direction for higher prediction. The holdout and external validations are indicated with a single cross (instead of a distribution over 100 iterations) because mean regression coefficients from the internal validation were applied to the holdout and external data, which resulted in a single performance metric. Abbreviation: Norm.: Normalised.

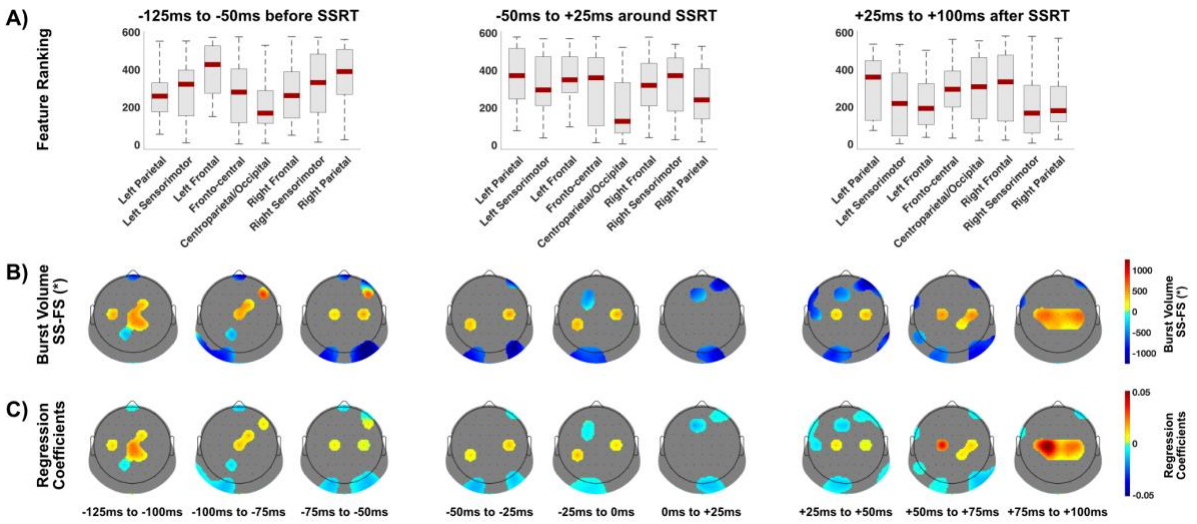
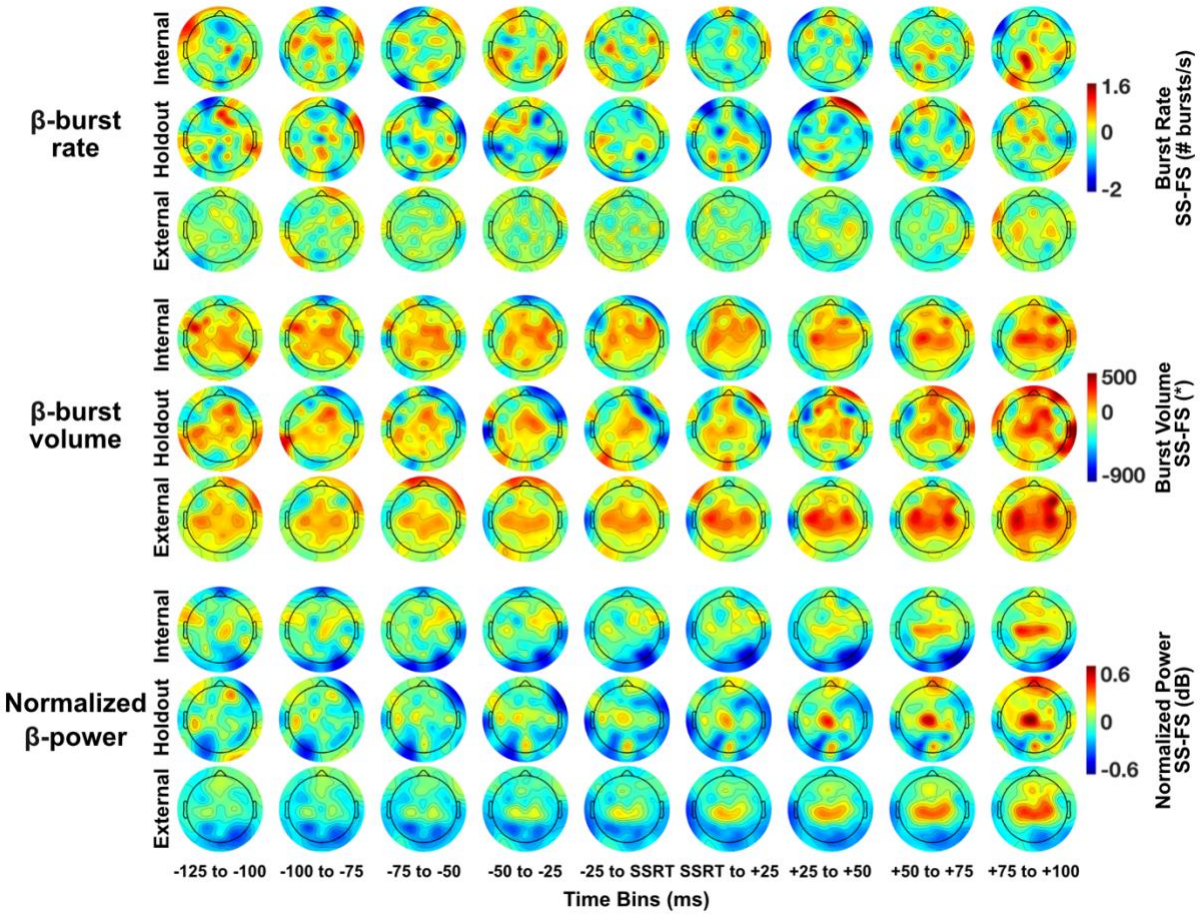


Figure 4. Spatial and temporal β-burst volume dynamics for the Stop trial classification. **A)** Boxplots display the machine learning feature ranking (ranked after selection frequency and absolute regression coefficient). Low ranking means more predictive in the machine learning analysis. These were averaged over three time bins (75 ms each) and for each time bin averaged over eight different brain regions. **B)** β-Burst volume data (successful minus failed Stop trials, a.u.) from the top 75 features (over all time bins) are shown for each 25 ms time bin with the rest of the electrodes masked. **C)** Regression coefficients from the machine learning logistic regression analysis from the top 75 features (over all time bins) are shown for each 25 ms bin with the rest of the electrodes masked. Abbreviations: SS: Successful Stop trial; FS: Failed Stop trial; SSRT: Stop Signal Reaction Time. *β-burst volume is burst duration x frequency span x amplitude.



716 **Figure 5. Topoplots of each β-feature (β-burst rate, β-burst volume, normalised β-**
717 **power) for the internal, holdout and external validation sets.** Topoplots are shown for
718 each 25 ms bin. β-Band frequency range was 15-29 Hz. B-bursts were calculated using 2x
719 median power as burst detection threshold. Means are shown for β-burst rate and normalised
720 β-power, medians are shown for β-burst volume. Data are shown for successful minus failed
721 Stop trials for each β-feature. Abbreviations: SS: Successful Stop trial; FS: Failed Stop trial;
722 SSRT: Stop Signal Reaction Time. *β-burst volume is burst duration x frequency span x
723 amplitude.

Tables

Table 1. Electrode labels (64 channels, 10-5 system) grouped into eight different brain regions. This step reduced spatial features to aid data interpretation.

Brain area	Electrode labels
Left parietal	CP5, TP7, P3, P5, P7, P9, PO3, PO7
Left sensorimotor	FC3, T7, C5, C3, C1, CZ, CP1, CP3
Left frontal	FP1, AF7, AF3, F7, F5, F3, FT7, FC5
Fronto-central	FPZ, AFZ, FZ, F1, F2, FC1, FC2, FCZ
Centroparietal/occipital	PZ, P1, P2, POZ, OZ, O1, O2, IZ
Right frontal	FP2, AF8, AF4, F8, F6, F4, FT8, FC6
Right sensorimotor	FC4, T7, C6, C4, C2, CPZ, CP2, CP4
Right parietal	CP6, TP8, P4, P6, P8, P10, PO4, PO8

Table 2. Characteristics and statistical comparison of the internal, holdout and external validation sets. For all analyses, t-tests were used, except for sex comparisons, for which we used a Chi-square test. Means and standard deviations are reported, except for sex. Abbreviations: df: degrees of freedom; SSRT: stop signal reaction time; SSD: stop-signal delay; RT: reaction time.

				Internal – Holdout		Internal – External	
	Internal validation set	Holdout validation set	External validation set	Statistical test (df)	p-value	Statistical test (df)	p-value
Sex (female:male)	78:52	55:33	120:81	$X^2(1) = 0.14$	0.71	$X^2(1) = 0.01$	0.96
Age (years)	35.39 (14.41)	33.78 (15.35)	22.70 (6.83)	t(216) = 0.79	0.43	t(329) = 10.76	$2.44*10^{-23}$
SSRT (ms)	190 (36)	195 (34)	241 (49)	t(216) = -1.13	0.26	t(329) = -10.31	$8.55*10^{-22}$
SSD (ms)	278 (61)	266 (66)	284 (123)	t(216) = 1.35	0.18	t(329) = -0.57	0.57
Mean Go RT (ms)	486 (62)	479 (71)	535 (105)	t(216) = 0.85	0.40	t(329) = -4.79	$2.55*10^{-6}$
Mean failed Stop RT (ms)	421 (58)	409 (66)	460 (92)	t(216) = 1.42	0.16	t(329) = -4.25	$2.79*10^{-5}$
No. of successful Stop epochs	22.68 (3.49)	22.78 (3.69)	38.85 (4.46)	t(216) = -0.22	0.83	t(329) = -35.00	$5.98*10^{-113}$
No. of failed Stop epochs	19.63 (3.70)	20.16 (3.95)	32.09 (5.11)	t(216) = -1.01	0.32	t(329) = -24.02	$2.30*10^{-74}$
Probability of successful Stop	0.54 (0.07)	0.53 (0.09)	0.52 (0.03)	t(216) = 0.51	0.61	t(329) = 3.17	0.002
Probability of Go omission	2.49 (3.53)	2.27 (4.16)	0.03 (0.04)	t(216) = 0.43	0.67	t(329) = 9.93	$1.65*10^{-20}$
Probability of choice errors	2.33 (2.00)	2.57 (3.37)	0.01 (0.02)	t(216) = -0.67	0.50	t(329) = 16.46	$8.20*10^{-45}$