Prediction of pathological subthalamic nucleus beta burst occurrence in Parkinson's disease

Bahman Abdi-Sargezeh[1,2], Sepehr Shirani[3], Abhinav Sharma[1,2], Alexander Green[2], Harith Akram[4], Ludvic Zrinzo[4], Patricia Limousin[4], Tom Foltynie[4], Timothy Denison[1], Huiling Tan[1,2], Vladimir Litvak[5], Ashwini Oswal[1,2]

Author affiliations:

[1]MRC Brain Network Dynamics Unit, University of Oxford, Oxford, UK

[2]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK

[3]Department of Computer Science, Nottingham Trent University, Nottingham, UK

[4]Department of Clinical and Movement Neurosciences, University College London, London, UK

[5]The Welcome Centre for Human Neuroimaging, University College London, UK

Correspondence to: Bahman Abdi-Sargezeh or Ashwini Oswal

MRC Brain Network Dynamics Unit, Mansfield Road, Oxford, OX1 3TH

bahman.abdisargezeh@ndcn.ox.ac.uk or ashwini.oswal@ndcn.ox.ac.uk

Keywords: Beta burst prediction; DBS; deep learning; adaptive deep brain stimulation; local field potential.

Abstract

The cortico-basal ganglia network in Parkinson's disease (PD) is characterised by the emergence of transient episodes of exaggerated beta frequency oscillatory synchrony known as bursts. Although beta bursts of prolonged duration and amplitude are well recognised to have a detrimental effect on motor function in PD, the neurophysiological mechanisms leading to burst initiation remain poorly understood. Related to this is the question of whether there exist features of basal ganglia activity which can reliably predict the onset of beta bursts. Current state-of-the-art adaptive Deep Brain Stimulation (aDBS) algorithms for PD involve the reactive delivery of stimulation following burst detection and are unable to stimulate proactively so as to prevent burst onset. The discovery of a predictive biomarker would allow for such proactive stimulation, thereby offering further potential for improvements in both the efficacy and side effect profile of aDBS.

Here we use deep neural networks to address the hypothesis that beta bursts can be predicted from invasive subthalamic nucleus (STN) recordings in PD patients. We developed a neural network which was able to predict bursts 31.6ms prior to their onset, with a high sensitivity and a low false positive rate (mean performance metrics: sensitivity = 84.8%, precision = 91.5%, area under precision recall curve = 0.87 and false positive rate = 7.6 per minute). Furthermore, by considering data segments that our network labelled as being predictive, we show that a dip in the beta amplitude (a fall followed by a subsequent rise) is a predictive biomarker for subsequent burst occurrence.

Our findings demonstrate proof-of-principle for the feasibility of beta burst prediction and inform the development of a new type of intelligent DBS approach with the capability of stimulating proactively to prevent beta burst occurrence.

Introduction

Parkinson's disease (PD) is a common neurodegenerative condition which is characterised by nigrostriatal dopamine depletion and the emergence of stereotyped patterns of oscillatory synchrony within cortico-basal ganglia circuits[1,2]. Excessive synchronisation across the beta (13-30 Hz) frequency range characterises the parkinsonian dopamine depleted state and is believed to relate directly to motoric impairment[3,4]. Therapeutic approaches such as both STN

DBS and dopaminergic medication lead to a suppression of basal ganglia beta oscillatory synchrony, with the degree of suppression correlating positively with motor improvements[5–8]. Furthermore, a causal effect of beta oscillations on movement is suggested by the observation that entraining motor cortical beta rhythms results in movement slowing[9].

Recent observations highlight that beta activity is not continuous but occurs in short-lived packets known as **bursts** [10]. Although the mechanisms of beta burst generation remain poorly understood, it is increasingly believed that **bursts** of longer duration and amplitude may be particularly detrimental to motor function in PD[11]. This finding has led to beta activity being used as a feedback signal in amplitude-responsive closed loop DBS, where stimulation is delivered only when beta amplitude rises above a certain threshold[12–16]. Studies reveal that beta triggered adaptive DBS (aDBS) is more effective than conventional continuous DBS[15,16]. Additionally, by virtue of selectively targeting a pathophysiological signal of interest, aDBS may offer additional benefits including reduced battery consumption and a lower incidence of stimulation induced side effects such as dyskinesias, gait impairment and speech impairment[17].

One drawback of aDBS is that stimulation is initiated after some fixed delay following the actual occurrence of a **burst**[15]. This delay, which can be up to hundreds of milliseconds, will be the sum of the time taken for the burst to be detected and the system delay between burst detection and stimulation initiation[15]. This means that aDBS can only reactively suppress beta bursts after they have developed and propagated within the cortico-basal ganglia circuit. The discovery of a reliable biomarker that could allow for the **prediction** of bursts would facilitate the development of **proactive DBS** approaches with the capability of either preventing bursts or suppressing them earlier following their onset[18]. Proactive DBS approaches are likely to be more energy efficient [19] and could lead to improvements in both the efficacy and the side effect profile of DBS for PD.

In this work, we test the **hypothesis** that STN activity includes features which reliably **predict** beta **burst** onset. To address this question, we developed a deep neural network architecture - based on a combination of Convolutional Neural Networks (**CNNs**) and long-short term memory networks (**LSTMs**) – that modelled burst prediction as a binary **classification** problem (outputting either 1 or 0 depending on whether or not a burst was expected to occur). Our network was trained on STN local field potential (LFP) activity recorded from PD patients undergoing functional neurosurgery for the insertion of STN DBS electrodes. Importantly, we

train and test our neural network architecture on patient specific beta band filtered STN activity, which has emerged as a robust biomarker of motoric state for aDBS applications [15,20–22].

Our trained network was able to reliably **predict** burst onset in previously unseen test data. Importantly, by considering predictive segments within the test data, we reveal a pathophysiological **state transition** characterised by a reduction followed by a rise in beta amplitude that occurs prior to the onset of beta bursts. Our findings provide proof of principle for the feasibility of **burst prediction** for **proactive** DBS, as well as shedding light on the pathophysiological mechanisms of beta burst generation.

Methods

Patients and experimental details

We studied STN activity in 15 PD patients undergoing bilateral implantation of STN DBS electrodes at the National Hospital for Neurology and Neurosurgery (UCL). In all cases, a Medtronic model 3389 electrode with four platinum-iridium contacts was implanted. Recordings were performed 3-6 days after electrode implantation, before connection and insertion of the implantable pulse generator (see Error! Reference source not found. for further clinical details). Further details of the surgical procedure can be found in other reports [23,24]. PD diagnoses were made in accordance with the Queen Square Brain Bank Criteria[25]. All patients provided written informed consent and research protocols were approved by the local research ethics committee.

To maximise the probability of beta burst occurrence, recordings were performed following overnight withdrawal from dopaminergic medication (OFF medication). LFP activity was collected using a battery powered and mains optically isolated BrainAmp system (Brain Products) with a sampling frequency of 2400 Hz. Three bipolar channels (0-1, 1-2, 2-3) were recorded from each electrode and the data were subsequently high pass filtered at 1 Hz in the hardware to avoid amplifier saturation due to large DC offsets. Although recordings of cortical activity using magnetoencephalography (MEG) [5,18,26,27] were collected at the same time as LFP recordings, in this analysis we focus explicitly on beta burst prediction from the STN LFP alone.

A neurologist was present during the recordings and patients were requested to keep their eyes open and to remain still. Either one or two rest recording sessions were performed. The duration of each session varied between 188 and 253 seconds (with a mean and standard error of the mean (SEM) of 197.6±3.6, see Error! Reference source not found.).

Table 1. Clinical characteristics of patients. LDE = levodopa dose equivalent. The total pre-operative UPDRS part III score is presented in the on and off medication states. S1 and S2 indicate the length of recording sessions in seconds.

Case	Age and gender	Disease duration (years)	Preoperative medication (mg)	UPDRS III pre-op off/on	Data length (s)	Number of bursts in training dataset	Number of bursts in validation dataset	Number of bursts in test dataset
1	54 F	10	LDE 1158	35/9	S1=192 S2=193	288	136	130
2	54 M	15	LDE 1150	53/19	S1=190 S2=193	214	137	140
3	58 F	3	LDE 390	41/9	S1=191	167	54	56
4	47 M	8	LDE 2264	46/4	S1=190 S2=188	279	114	116
5	57 M	8	LDE 1229	41/19	S1=238 S2=195	291	163	149
6	60 M	27	LDE 2048	63/8	S1=194 S2=189	279	140	130
7	57 M	17	LDE 1460	54/14	S1=190	209	44	44
8	52 M	13	LDE 1484	35/10	S1=191	188	44	55
9	58 M	11	LDE 1320	43/25	S1=201	220	60	45
10	72 M	9	LDE 1281	28/5	S1=192	209	41	50
11	51 M	8	LDE 650	49/21	S1=189	202	34	45
12	60 M	11	LDE 1012	28/5	S1=190 S2=201	267	165	171
13	43 M	9	LDE 1650	63/40	S1=253	245	43	46
14	41 M	6	LDE 1220	50/22	S1=188	217	38	41
15	58 M	12	LDE 1500	38/14	S1=202	194	46	29
Mean ± SEM	54.18	11.1	1321.1	44.5/14.9	197.6	231.3	83.9	83,1
	1.9	1.4	117.2	2.8/2.4	3.6	10.0	12.8	12.3

Determination of beta peak frequency and labelling of beta bursts

To determine the beta peak frequency, the power spectrum of the LFP from each bipolar contact was obtained using short-time Fourier transform (STFT). A Hamming window with a length of 1 second and an overlap of 50% was used for spectral estimation. The squared magnitude of the resulting complex spectrum was computed and averaged across all time windows for visualisation between frequencies of 1 and 100 Hz (with a resolution of 1 Hz; see Error! Reference source not found.**A** for an exemplar spectrum). For each participant the single bipolar LFP channel with the highest amplitude peak within the beta frequency range (13-30 Hz) - which we term the **beta channel** - was selected for further analysis.

LFP time series from the **beta channel** were bandpass filtered within a ± 3 Hz window centred on the beta peak frequency, using a causal 6th order Butterworth filter implemented in the SciPy library for Python (https://scipy.org/). The application of a causal filter served to prevent future samples from impacting current or past filter outputs which would be used for prediction. We limited our analysis to individual patient beta band filtered signals as these are typically used in aDBS implementations[15,20,21]. After filtering, data were downsampled to 600 Hz (to reduce computational cost) and rectified prior to the peak values being linearly interpolated to produce the beta amplitude envelope of the signal. Finally, beta burst timings were defined as time points where the beta amplitude envelope exceeded its 75th percentile [10,20] (see **Figure 1B**).

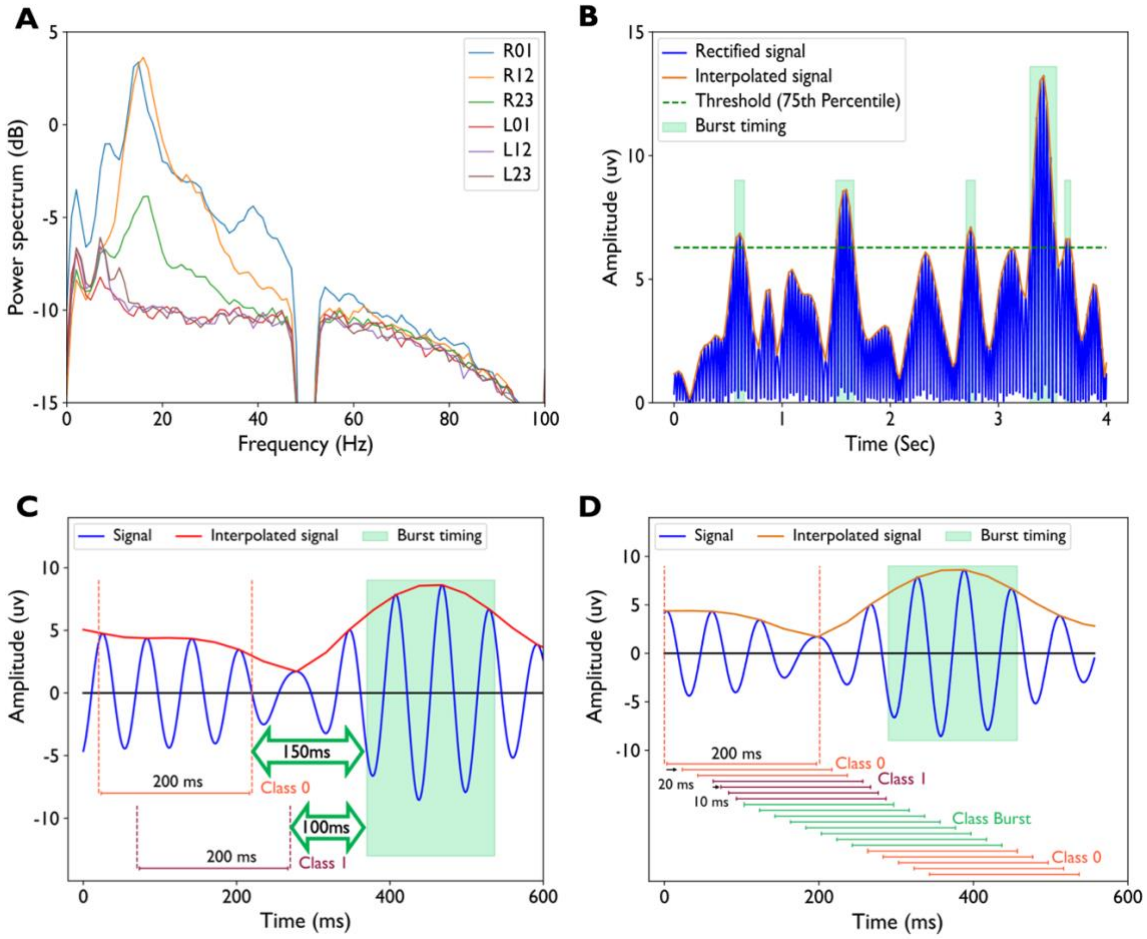

Figure 1. Determination of beta peak frequency, labelling of beta bursts and data segmentation for training of the neural network. (A) Exemplar power spectra of right and left hemisphere STN LFP signals (recorded from bipolar contact pairs 01, 12 and 23) from patient 12 are displayed. In this case, the right hemispheric channel 12 had the highest amplitude peak within the beta frequency range (peak frequency at 16 Hz) and was therefore selected as the *beta channel* for further analysis. **(B)** A 4s long segment of data from the right hemispheric channel 12 is displayed. Data were first filtered ($\pm$3 Hz) around the beta peak frequency before being rectified and interpolated. The 75th percentile of the interpolated signal amplitude distribution (dashed green line) was used as a threshold to define the onset and offset of beta bursts (burst timings are shown in the green rectangular boxes). **(C)** Illustration of the fixed window approach, where 200ms long data segments ending at fixed time intervals (0, 20, 40, 60, 80, or 100ms) prior to the onset of a burst are labelled as being predictive (Class 1) of subsequent burst onset. Non-predictive data segments (Class 0) terminated at least 150ms prior to the onset of a burst. **(D)** Demonstration of the sliding window approach. A 200ms long window with a stride length of 20ms was passed along the beta filtered time series. Four windowed segments with a shortened stride length of 10ms (see main text) were labelled as being predictive of subsequent burst occurrence (Class 1). Windows that ended during the occurrence of a burst (Class Burst) were excluded from subsequent analysis, whilst the remaining data segments were labelled as being non-predictive (Class 0).

Labelling of data for neural network classification

Our goal was to detect time domain data features that would predict subsequent beta burst occurrence with high sensitivity and specificity. The ideal predictive feature should be present before all bursting episodes and absent prior to non-bursting time periods. Consequently, we could frame burst prediction as a binary classification model, which would output either 0 or 1 depending on whether a burst was expected to occur. To train (and test) our classification model

we segmented beta band filtered LFP time series using either **sliding window** or **fixed window** approaches (see **Figure 1C and 1D**).

The fixed window approach allowed us to evaluate prediction performance at fixed time intervals prior to the onset of bursts. In this approach, 200ms long segments of the band filtered LFP that ended at fixed time intervals (either 0, 20, 40, 60, 80 or 100ms) prior to the onset of each burst were labelled as being predictive (**Class 1** in **Figure 1C**) for network training. For each **Class 1** data segment, a corresponding 200ms long data segment that was not predictive of subsequent burst occurrence (**Class 0**) was selected at random from data terminating at least 150ms before burst onset.

In the sliding window approach, we used a 200ms long time window with a stride length of 20ms (Figure 1D). If the end of a windowed segment overlapped with the occurrence of a burst, the segment was excluded from subsequent analysis (see **green** segments labelled **Class Burst** in Error! Reference source not found.**1D**). For training the network to learn burst predictive features, we used the last four 200ms long data windows terminating prior to burst onset (referred to as **Class 1** in **Figure 1D**). For these predictive data segments, a smaller stride length of 10ms was selected, such that the first predictive window started within 240ms of burst onset and the last predictive window ended within 10ms of burst onset. The use of a sliding window approach in this instance represents a form of data augmentation; a widely used technique in machine learning which can be used to significantly expand the size of the training dataset [28,29]. The remaining 200ms long data segments were associated with the absence of a subsequent burst and were labelled as being non-predictive (referred to as **Class 0** in **Figure 1D**).

Consecutive bursts separated by less than 40ms were treated as a single burst. The length of the sliding window and the stride length were selected empirically based on performance metrics for the validation datasets (see below section titled **Beta burst prediction network**). After testing three different window lengths (200, 250 and 300ms) and five stride lengths (10, 20, 30, 40 and 50ms), we observed that values of 200ms and 20ms respectively resulted in optimal performance.

Beta burst prediction network

Our burst prediction network (Error! Reference source not found.**)** is based a combination of CNNs and LSTMs, which were implemented in TensorFlow using the Keras API (https://www.tensorflow.org and https://keras.io/api/). 1D convolution (Conv1D with 128 filters, a kernel size of 5 and a stride length of 1) and 1D depthwise convolution (depthwise Conv1D with a kernel size of 5 and depth multiplier of 2) were applied to the beta filtered LFP prior to the use of a rectified linear unit (ReLU) activation function and max pooling (MaxPool), which served to half the number of temporal features. This process was repeated twice to capture deep temporal features of the signal with reasonable model complexity[30]. The resulting temporal features were inputted to two sequential LSTM layers (with a dropout rate of 0.5), prior to a sigmoid function being used to perform binary classification. The model was fit using the Adam optimiser to minimise the categorical cross-entropy loss function[31]. Similar combinations of CNN and LSTM architectures have been used previously for EEG signal classification[32,33]. CNNs, by virtue of using multiple filters, excel at capturing local data features, but unlike LSTMs are unable to learn long-term temporal dependencies which may facilitate the final prediction [34].

LFP data from each patient were split into training, validation, and test subsets. The number of beta bursts in each subset for each patient is shown in **Table 1**. Training data were used to fit multiple candidate models, whilst the validation dataset could be used to select the best model architecture and to optimise model hyperparameters (e.g., learning rate and the threshold of the final sigmoid output for binary classification). Final model performance was evaluated and reported using the test data.

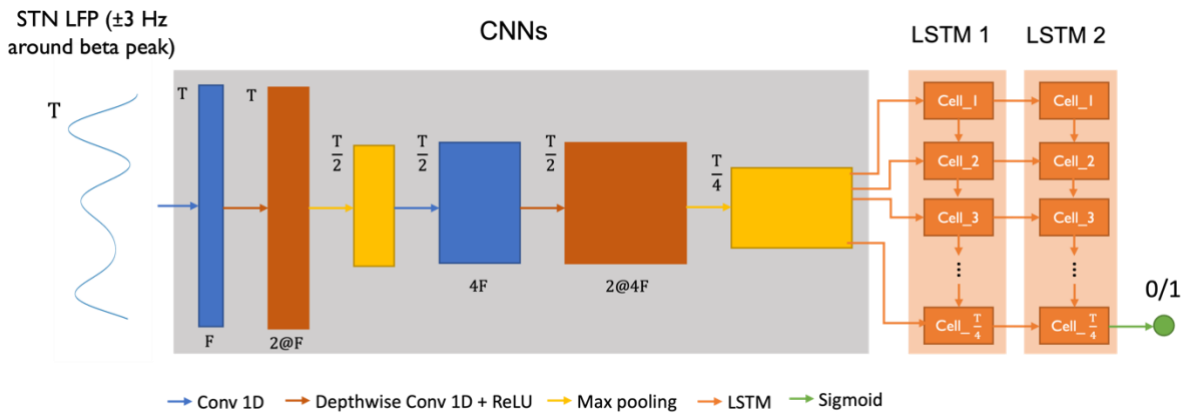

Figure 2. Architecture of the burst prediction network. The input beta filtered LFP (of dimension *T x 1*) is passed through a sequence of Conv1D, depthwise Conv1D, ReLU, and Max pooling (see main text). This process is repeated before the resulting outputs are passed to two sequential LSTM layers. The output of the final LSTM cell is used for binary classification using a sigmoid activation function. In this analysis, data were segmented into 200ms long epochs with a sampling rate of 600Hz, meaning that *T* = 120. *F* represents the number of filters in the convolutional layers and was selected to be 128. The number of depthwise convolution output channels for each input channel was selected to be 2 (denoted by 2@).

Prediction performance metrics

For analysis using the fixed window approach, test data were segmented as previously described (see section titled '**Labelling of data for neural network classification**'), using the same six different time windows (0, 20, 40, 60, 80 and 100ms prior to burst onset) for predictive data. Data were either labelled as being predictive (**Class 1**) or non-predictive (**Class 0**) such that there were an equal number of both data types. Similarly for the sliding window approach, test data were segmented into 200ms long windows with a stride length of 20ms. Segments overlapping with time periods of burst occurrence were excluded, whilst those terminating within 60ms of burst onset were labelled as being predictive (**Class 1**). The remaining data segments were labelled non-predictive (**Class 0**), meaning that the number of non-predictive segments was substantially greater than the number of predictive data segments. Segmented data were then passed to the trained network so that the resulting prediction could be compared to the input class. Based on the comparison of test data labels and the model prediction, we calculated the following classification metrics:

- **True positives** (TP) – this is the number of burst predictive data segments (Class 1) that were correctly identified by the model.
- **True negatives** (TN) – this is the number of non-predictive data segments (Class 0) that were correctly identified by the model.

- **False positives** (FP) – this refers to non-predictive data segments (Class 0) where the model incorrectly predicted subsequent burst occurrence.
- **False negatives** (FN) – this refers to predictive data segments (Class 1), where the model failed to predict subsequent burst occurrence.
- **False positives per minute** (FP/min) – this quantifies the rate of false positive occurrence.
- **Accuracy** (ACC) – this indicates the proportion of correct predictions and is defined as: ACC = (TP+TN)/(TP+TN+FP+FN).
- **Precision** (PRC) – tells us the proportion of correct positive predictions and is defined as: PRC = TP/(TP+FP).
- **Sensitivity** (SEN) – also known as recall, this illustrates how well the model predicts the occurrence of bursts and is defined as: SEN=TP/(TP+FN).
- **Area under the receiver operating characteristic curve** (AUC-ROC) – this is a performance metric that considers the trade-off between sensitivity and the false positive rate at various thresholds. It provides a measure of the classifiers ability to correctly distinguish between classes and was used for the fixed window approach where class membership was balanced.
- **Area under the precision-recall curve** (AUC-PR) – this is a performance metric that considers the trade-off between precision and sensitivity at various thresholds and is therefore not influenced by a disproportionately high occurrence of true negative predictions. The AUC-PR was used instead of the AUC-ROC for the sliding window approach, owing to its increased effectiveness for classification evaluation on imbalanced datasets [35,36].
- **Prediction time prior to burst occurrence** (PT-PBO) – this metric was computed for the sliding window approach as the mean time interval between the end of the first true positive prediction (within 60ms of burst onset) and the start of a burst.

Results

STN beta bursts can be predicted in advance of their onset

Prediction metrics for the sliding window approach applied to individual patient data are shown in **Table 2**. Our model achieved a high sensitivity and precision for burst prediction, with a low

false positive rate (mean sensitivity = 84.8%; mean precision = 91.5%; mean AUC-PR = 0.87; mean FP/min = 7.6). Importantly, across patients the mean prediction time prior to burst occurrence was 31.6ms.

Similarly, for the fixed window approach we were able to achieve high levels of sensitivity, precision and AUC-ROC when using predictive data segments that terminated up to 20ms before burst onset. **Supplementary Table 1** shows burst prediction performance metrics for each patient, for each of the six different predictive window termination timepoints relative to burst onset (0, 20, 60, 80 and 100ms). This information is summarised in **Figure 3A**, which shows mean estimates of accuracy, sensitivity, precision, and AUC-ROC across patients for the six different predictive window termination timepoints.

Table 2. Individual patient burst prediction performance metrics for the sliding window approach. The lowermost row shows the mean and standard error for each metric.

Case	TP	FN	FP	TN	SEN (%)	PRC (%)	AUC-PR	FP/min	PT-PBO (ms)
1	117	12	17	3747	90.7	87.3	0.89	10.5	34±1
2	140	10	9	2906	93.3	93.9	0.89	5.6	30±1
3	49	12	1	785	80.3	98.0	0.88	2.0	24±1
4	115	22	9	3486	83.9	92.7	0.88	5.7	36±1
5	123	39	10	2707	75.9	92.4	0.89	6.1	32±1
6	116	34	8	3344	77.3	93.5	0.89	5.0	34±1
7	46	3	1	793	93.8	97.8	0.90	2.1	35±2
8	50	6	10	793	89.2	83.3	0.86	20.9	34±2
9	39	8	0	1036	83.0	100	0.89	0.0	32±2
10	51	5	2	847	91.0	96.2	0.89	4.1	29±2
12	41	12	7	716	77.36	85.4	0.78	14.8	26±2
12	170	29	14	2418	85.4	92.3	0.92	8.3	34±1
13	40	7	5	1364	85.1	88.9	0.84	7.9	29±2
14	42	10	2	919	80.7	95.4	0.83	4.2	28±2
15	29	5	9	1155	85.2	76.2	0.88	17.8	37±2

Mean±SEM	77±11	14±3	6±1	1801±297	84.8±1.5	91.5±1.6	0.87±0.00	7.6±1.5	31.6±1.0

Although the sensitivity drops substantially as the predictive window is shifted to greater than 20ms prior to burst onset, the accuracy, precision and AUC-ROC of the model are less detrimentally affected. This reflects the fact that false negative predictions are more likely than false positive predictions as the predictive window is moved further away from burst onset.

Classification threshold controls the trade-off between sensitivity and false positive rate

An ideal burst prediction model should have a high sensitivity and a low false positive rate. The trade-off between these two-performance metrics is determined by the sigmoid classification threshold of the output layer (see **Figure 2**), which binarises burst predictions. **Figure 3B** shows the effect of varying the classification threshold on the sensitivity and false positive rate of each patient's burst prediction model. Lowering the threshold results in increased sensitivity at the cost of an increased false positive rate. For all patients, our pipeline achieved sensitivities in excess of 60% with a false positive rate of less than 3 per minute. We were also able to achieve sensitivities of greater than 80%, with false positive rates of less than 10 per minute in all but one patient (Patient 11).

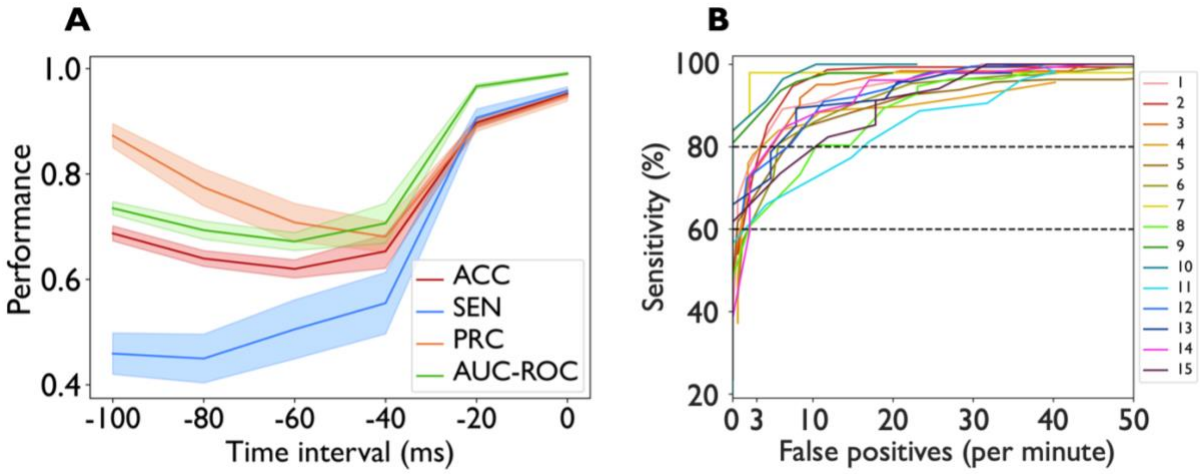

Figure 3. (A) Burst prediction performance metrics for the fixed window approach. Results for each of the six different predictive window termination timepoints relative to burst onset are averaged across patients (ACC = accuracy; SEN = sensitivity; PRC = precision; AUC-ROC = area under the ROC curve). Solid lines represent the mean, with the shaded regions indicating the standard error of the mean. The sensitivity for burst prediction drops dramatically when the 200ms long predictive window terminates more than 20ms prior to burst onset (see main Results for further discussion). **(B) Classification threshold and the trade-off between sensitivity and false positive rate.** Each coloured line represents a different patient. Lowering the classification threshold increases sensitivity, at the cost of an increased false positive rate. For all patients, our pipeline achieved sensitivities in excess of 60% with a false positive rate of less than 3 per minute. We were also able to achieve sensitivities of greater than 80%, with a false positive rate of less than 10 per minute in all but one patient (Patient 11).

Beta amplitude modulations predict subsequent burst occurrence

Our results reveal that beta burst occurrence may be predicted up to ~30ms prior to burst onset. But is there a consistent data feature which facilitates this prediction, that could be used in aDBS applications? To address this question, we examined mean beta amplitude envelopes for test data segments leading to TP, FP, and FN predictions. This procedure was performed separately for the sliding window and fixed window approaches. In the case of TP and FN predictions for the sliding window approach, we selected the data segment corresponding to the latest positive or negative prediction that occurred within a 20ms window prior to burst onset. Similarly for the fixed window approach, we selected predictive windows that terminated 20ms prior to burst initiation, since predictive windows that terminated earlier than this led to poor prediction performance (see **Figure 3A**).

Figure 4 and **Figure 5** show individual patient results for the sliding and fixed window approaches. The group trends are seen more clearly in **Figure 6**, which displays normalised beta amplitudes averaged across all patients, separately for the fixed and sliding window approaches. Normalisation to a value between 0 and 1 was performed for each patient prior to averaging, to account for interindividual differences in beta amplitudes. Importantly, for TP predictions there is a consistent dip (fall followed by a subsequent rise) in the beta amplitude across patients, which occurs approximately 80ms prior to beta burst onset (Figure 6B; note that in this figure, the predictive window ends 20ms prior to burst onset). As expected, the beta amplitude profile of FP predictions closely matched that of TP predictions. FN predictions in contrast exhibited less pronounced beta amplitude dips. Taken together, these results reveal that dips in the beta amplitude occurs consistently approximately 80ms prior to burst onset and can be a reliable predictive biomarker of burst occurrence (see **Discussion** for further comment regarding this phenomenon).

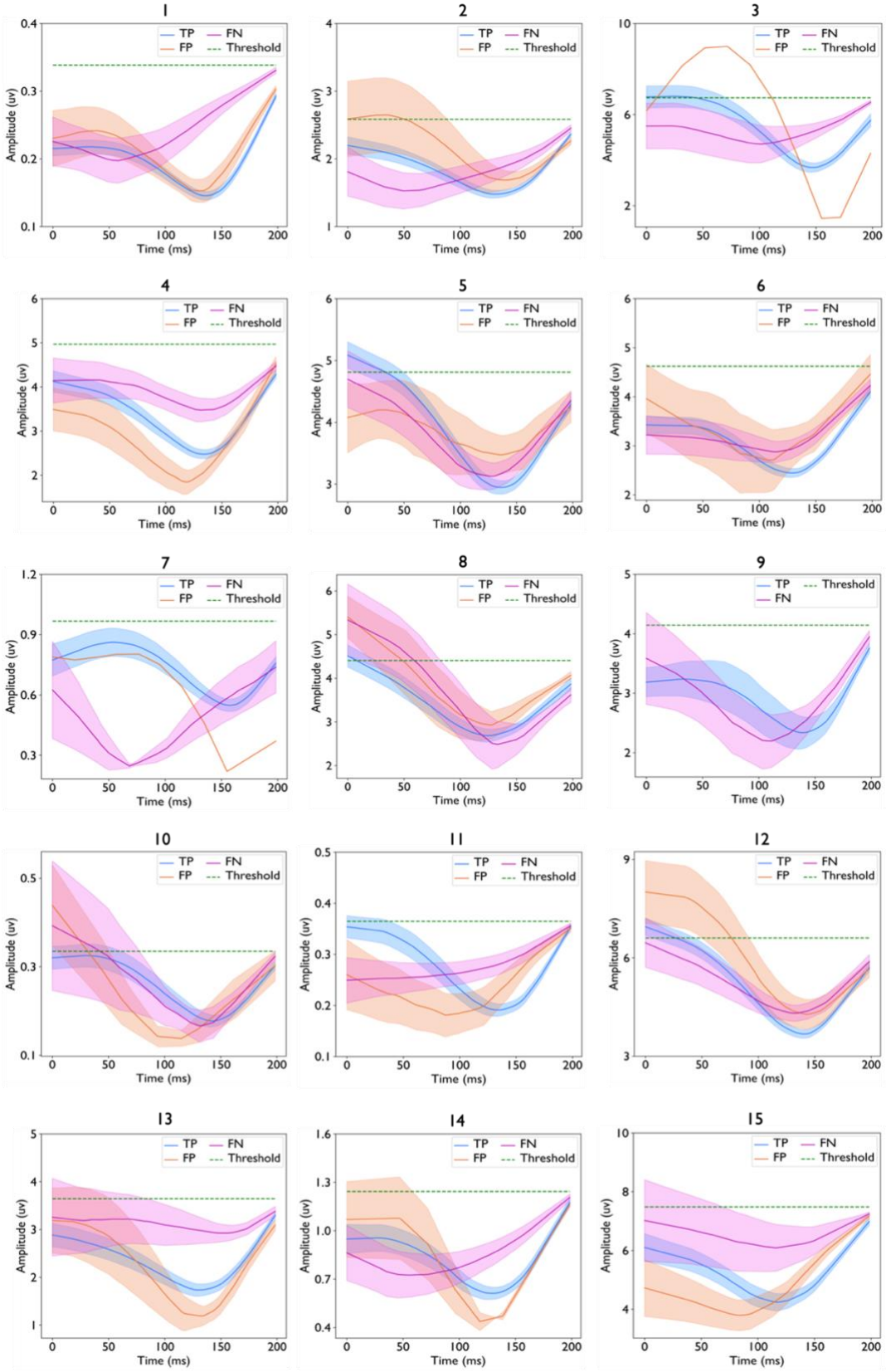

Figure 4. Individual patient beta envelope amplitudes for (200ms long) data segments leading to TP, TN, and FP predictions with the sliding window approach. Solid lines indicate the mean, whilst shaded areas represent the standard error of the mean. For each patient the amplitude threshold (75th percentile) for defining burst occurrence is also indicated in the green dashed line. For patients 3 and 7 there was only 1 FP prediction, whilst for patient 9 there were no FP predictions. See main text for discussion of key findings.

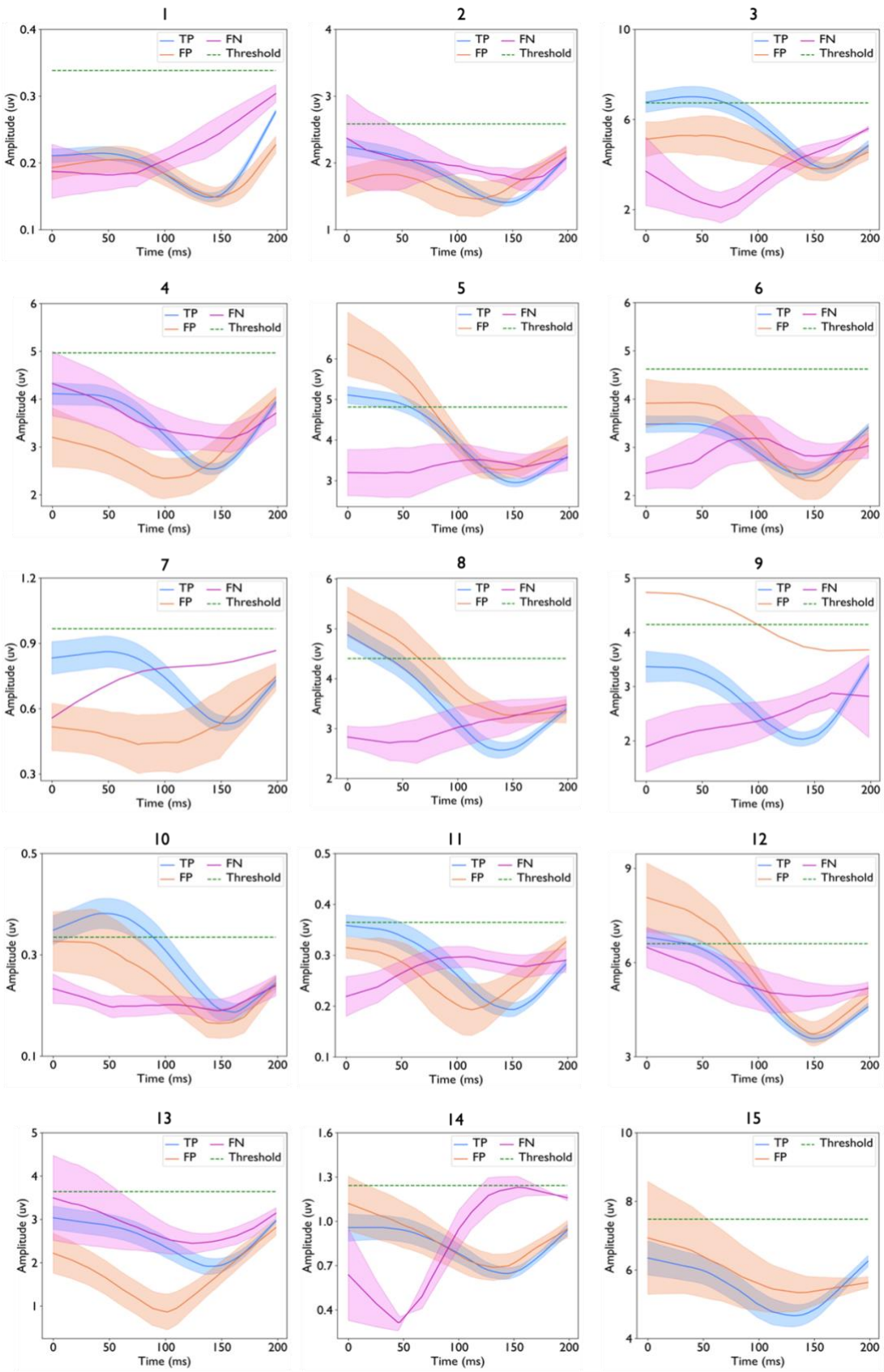

Figure 5. Individual patient beta envelope amplitudes for (200ms long) data segments leading to TP, TN, and FP predictions with the fixed window approach in which the predictive windows terminated 20ms prior to burst onset. Solid lines indicate the mean, whilst shaded areas represent the standard error of the mean. For each patient the amplitude threshold (75th percentile) for defining burst occurrence is also indicated in the green dashed line. For patients 7 there was 1 FN prediction, whilst for patient 15 there were no FN predictions. See main text for discussion of key findings.

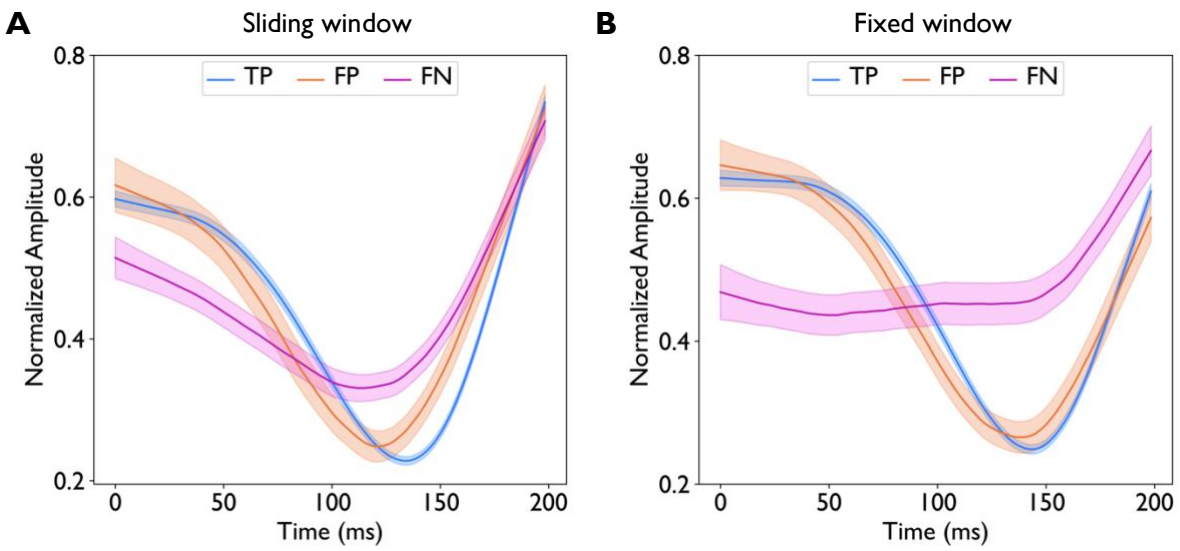

Figure 6. Normalised mean beta envelope amplitudes for (200ms long) data segments leading to TP, TN, and FP predictions separately for the sliding window (A) and fixed window approaches (B). For the fixed window approach, we selected predictive windows that terminated 20ms prior to burst onset. Solid lines indicate the mean across patients, whilst shaded areas represent the standard error of the mean. For data segments leading to TP predictions, there is a dip in the beta amplitude occurring ~80ms prior to beta burst onset. Data segments leading to FN predictions did not exhibit such a pronounced dip.

Validation of findings using surrogate data

We next sought to validate our findings, by generating surrogate data that preserved beta burst characteristics of the original data whilst destroying pathophysiological signal properties occurring within non-bursting time periods. We expected to see that this manipulation would lead to significantly impaired burst prediction performance metrics when training and testing our burst prediction network on surrogate data.

In order to generate patient specific surrogates, original signals were bandpass filtered (using a zero-phase 6th order Butterworth filter with pass band of ± 3 Hz around the beta peak frequency) to define periods of beta bursts. For the generation of surrogates, data from periods of bursting activity were preserved, whilst data from non-bursting time periods were randomly shuffled. Surrogate data were beta band pass filtered using the same causal filter as the original data. **Supplementary Table 2** shows the performance of the prediction model using surrogate data. As anticipated, the model achieved very poor performance compared to the original data, indicating that the network relies on the physiological signatures present in the original data to predict bursts, and the narrow-band filter has no effect on training the network.

Discussion

We have demonstrated that the onset of pathological beta bursts within the STN can be predicted in patients with PD. By training patient specific deep learning models, we show that bursts can be predicted with high sensitivity, accuracy, and precision up to 31.6ms prior to their onset. Importantly, our approach requires only modest amounts of training data - averaging less than 3 minutes for each patient. These results provide proof-of-principle for the feasibility of burst prediction and also shed light on the neurophysiological mechanisms leading to burst initiation. Interestingly for all patients, we observed that a predictive marker of burst occurrence was a dip in the beta amplitude that occurred approximately 80ms prior to burst onset.

Study limitations

Before considering the translational and neurophysiological significance of our findings, it is prudent to highlight potential limitations of our study. Firstly, STN recordings were performed a few days after electrode implantation, whilst patients were at rest and off dopaminergic medication. This meant that clinical constraints, including patient fatigue, allowed for only relatively short duration recordings. Future work may therefore focus on predicting bursts across different behavioural and medication states in patients with sensing enabled DBS devices that allow for prolonged recordings [13]. Secondly, although in the present study we focused on the prediction of bursts, it is becoming increasingly evident that bursts of different durations may have different pathophysiological roles [10,11,20]. Augmenting our model to predict burst durations (short vs. long) will therefore be an import area for future study. Thirdly, we have predicted STN beta bursts from STN activity and have therefore not included nodes within the cortico-basal ganglia circuit – such as primary motor cortex (M1) - whose activities may facilitate the prediction of STN bursts [18,37]. Although electrocorticographic recordings from M1 are currently performed only for research purposes during DBS procedures, there is growing evidence to suggest that these may provide complementary information about clinical states (e.g., dyskinesia), particularly when the fidelity of subcortical recordings is compromised by the presence of stimulation artefacts[38].

Translational potential of burst prediction

Beta frequency oscillatory activity within the cortico-basal ganglia circuit has proven to be a robust biomarker of motoric impairments – particularly bradykinesia and rigidity - in PD[4,7,39,40]. Moreover, aDBS approaches which involve the delivery of STN stimulation only after the detection of beta bursts are more effective in terms of achieving motor benefit and limiting stimulation related side effects, than conventional continuous DBS[15–17,41]. Although aDBS been shown to be effective, a key drawback of beta amplitude triggered approaches is that they cannot prevent the onset of pathological beta bursts, but rather respond with some fixed delay following burst onset. Consequently, aDBS may be acting too late to optimally prevent the initiation and propagation of pathological oscillatory activity within the cortico-basal ganglia circuit (akin to closing the stable door after the horse has bolted). This drawback speaks to the utility of patient specific burst prediction approaches which allow for the earlier detection of bursts, and consequently the earlier delivery of stimulation. Although our findings provide the first demonstration of proof-of-principle, further real-time clinical testing of our approach in externalized patients is warranted before moving towards the implementation of prediction-based stimulation algorithms in DBS devices.

Interestingly, our approach also draws important parallels with prior work exploring seizure forecasting from intracranial recordings. Deep learning-based approaches including CNNs, and LSTMs have proven to be successful in this regard and may allow for the more rapid delivery of seizure abortive treatments such as medication or therapeutic stimulation [42–44]. Theoretically, our approach may also be leveraged to predict other oscillatory features within and across frequency bands, which have been shown to relate to distinct cognitive and motor symptoms of PD [38,45–47].

Insights into network mechanisms leading to burst generation

A key finding of our work is that the beta amplitude displays a stereotyped pattern of modulation prior to the onset of bursts. **Figure 6** reveals the precise nature of this modulation, with the beta amplitude initially falling, reaching a nadir at approximately 80ms prior to burst onset, and then subsequently rising again towards burst initiation. But what are the neural mechanisms which might explain this phenomenon? Computational models relying upon stochastic (noisy) inputs to neuronal populations as a primary driver of bursting cannot easily

recapitulate the pre-burst amplitude dip [18,48,49]. Usually in such models, noisy inputs determine transitions of the system between stable (non-oscillatory) and unstable (oscillatory) states via a bifurcation, meaning that oscillatory amplitudes can only change in one direction (either increase or decrease). A more plausible explanation for the beta amplitude dip might relate to the communication through coherence hypothesis, which proposes that maximal oscillatory entrainment within a network requires optimal phase alignment of interacting oscillatory populations [50–52]. Switching to an optimal phase alignment for burst initiation requires a coordinated phase reset of oscillatory populations within the cortico-basal ganglia circuit [52], which could in theory account for the dip in the beta amplitude observed in our work.

Summary

We have harnessed deep neural networks to show that the timing of pathological beta bursts within the STN can be predicted in patients with PD. An important biomarker for this prediction is a dip in the beta amplitude, which occurs consistently ~80ms prior to burst onset and may be indicative of a phase reset of oscillatory cortico-basal ganglia neuronal populations. Crucially, our findings motivate the development of a new generation of predictive DBS algorithms which are capable of either preventing pathological bursts or terminating them earlier following their onset.

Funding

This work was supported by an MRC Clinician Scientist Fellowship (MR/W024810/1) held by AO. BA and AO acknowledge funding support from the Oxford University Hospitals Charity. The Wellcome Centre for Human Neuroimaging is supported by core funding from Wellcome (203147/Z/16/Z). H.T. is supported by the Medical Research Council UK (MC_UU_00003/2, MR/V00655X/1, MR/P012272/1), the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC) and the Rosetrees Trust, UK.

Acknowledgements

We thank Rafal Bogacz (University of Oxford) for helpful discussions regarding this work. Anonymized datasets used for the current study and computer code are available from the corresponding authors on reasonable request and will be deposited on the MRC Brain Network Dynamics Unit data-sharing platform (https://data.mrc.ox.ac.uk/data-set/).

References

1. Brown P, Oliviero A, Mazzone P, Insola A, Tonali P, Di Lazzaro V. Dopamine dependency of oscillations between subthalamic nucleus and pallidum in Parkinson's disease. *Journal of Neuroscience*. 2001;21(3):1033-1038.

2. Eusebio A, Thevathasan W, Doyle Gaynor L, et al. Deep brain stimulation can suppress pathological synchronisation in parkinsonian patients. *J Neurol Neurosurg Psychiatry*. 2011;82(5):569-573.

3. Oswal A, Brown P, Litvak V. Synchronized neural oscillations and the pathophysiology of Parkinson's disease. *Curr Opin Neurol*. 2013;26(6):662-670.

4. Neumann WJ, Degen K, Schneider GH, et al. Subthalamic synchronized oscillatory activity correlates with motor impairment in patients with Parkinson's disease. *Movement Disorders*. 2016;31(11):1748-1751.

5. Oswal A, Beudel M, Zrinzo L, et al. Deep brain stimulation modulates synchrony within spatially and spectrally distinct resting state networks in Parkinson's disease. *Brain*. 2016;139(5):1482-1496.

6. Kühn AA, Kupsch A, Schneider GH, Brown P. Reduction in subthalamic 8-35 Hz oscillatory activity correlates with clinical improvement in Parkinson's disease. *Eur J Neurosci*. 2006;23(7):1956-1960.

7. Kühn AA, Tsui A, Aziz T, et al. Pathological synchronisation in the subthalamic nucleus of patients with Parkinson's disease relates to both bradykinesia and rigidity. *Exp Neurol*. 2009;215(2):380-387.

8. Ray NJ, Jenkinson N, Wang S, et al. Local field potential beta activity in the subthalamic nucleus of patients with Parkinson's disease is associated with improvements in bradykinesia after dopamine and deep brain stimulation. *Exp Neurol*. 2008;213(1):108-113.

9. Pogosyan A, Gaynor LD, Eusebio A, Brown P. Boosting cortical activity at Beta-band frequencies slows movement in humans. *Current biology*. 2009;19(19):1637-1641.

10. Tinkhauser G, Pogosyan A, Tan H, Herz DM, Kühn AA, Brown P. Beta burst dynamics in Parkinson's disease off and on dopaminergic medication. *Brain*. 2017;140(11):2968-2981.

11. Torrecillos F, Tinkhauser G, Fischer P, et al. Modulation of Beta Bursts in the Subthalamic Nucleus Predicts Motor Performance. *Journal of Neuroscience*. 2018;38(41):8905-8917.

12. Velisar A, Syrkin-Nikolau J, Blumenfeld Z, et al. Dual threshold neural closed loop deep brain stimulation in Parkinson disease patients. *Brain Stimul*. 2019;12(4):868-876.

13. Gilron R, Little S, Perrone R, et al. Long-term wireless streaming of neural recordings for circuit discovery and adaptive stimulation in patients with Parkinson's disease HHS Public Access Author manuscript. *Nat Biotechnol*. 2021;39(9):1078-1085.

14. Little S, Beudel M, Zrinzo L, et al. Bilateral adaptive deep brain stimulation is effective in Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2016;87(7):717.

15. Little S, Pogosyan A, Neal S, et al. Adaptive deep brain stimulation in advanced Parkinson disease. *Ann Neurol*. 2013;74(3):449-457.

16. Bronte-Stewart H, Martijn B, Ostrem J, et al. Chronic Adaptive DBS Provides Similar "On" Time with Trend of Improvement Compared to Continuous DBS in Parkinson's Disease and 98% of Participants Chose to Remain on aDBS (S2.008). *Neurology*. 2024;102(17_supplement_1).

17. Little S, Tripoliti E, Beudel M, et al. Adaptive deep brain stimulation for Parkinson's disease demonstrates reduced speech side effects compared to conventional stimulation in the acute setting. *J Neurol Neurosurg Psychiatry*. 2016;87(12):1388-1389.

18. Oswal A, Cao C, Yeh CH, et al. Neural signatures of hyperdirect pathway activity in Parkinson's disease. *Nature Communications 2021 12:1*. 2021;12(1):1-14.

19. Bocci T, Prenassi M, Arlotti M, et al. Eight-hours conventional versus adaptive deep brain stimulation of the subthalamic nucleus in Parkinson's disease. *NPJ Parkinsons Dis*. 2021;7(1).

20. Tinkhauser G, Pogosyan A, Little S, et al. The modulatory effect of adaptive deep brain stimulation on beta bursts in Parkinson's disease. *Brain*. 2017;140(4):1053-1067.

21. He S, Baig F, Merla A, et al. Beta-triggered adaptive deep brain stimulation during reaching movement in Parkinson's disease. *Brain*. 2023;146(12):5015-5030.

22. Neumann WJ, Gilron R, Little S, Tinkhauser G. Adaptive Deep Brain Stimulation: From Experimental Evidence Toward Practical Implementation. *Movement Disorders*. 2023;38(6):937-948.

23. Foltynie T, Zrinzo L, Martinez-Torres I, et al. MRI-guided STN DBS in Parkinson's disease without microelectrode recording: efficacy and safety. *J Neurol Neurosurg Psychiatry*. 2011;82(4):358-363.

24. Foltynie T, Hariz MI. Surgical management of Parkinson's disease. *Expert Rev Neurother*. 2010;10(6):903-914.

25. Gibb WR, Lees AJ. A comparison of clinical and pathological features of young- and old-onset Parkinson's disease. *Neurology*. 1988;38(9):1402-1406.

26. Oswal A, Abdi-Sargezeh B, Sharma A, et al. Spatiotemporal signal space separation for regions of interest: Application for extracting neuromagnetic responses evoked by deep brain stimulation. *Hum Brain Mapp*. 2024;45(2):e26602.

27. Oswal A, Brown P, Litvak V. Movement related dynamics of subthalmo-cortical alpha connectivity in Parkinson's disease. *Neuroimage*. 2013;70:132-142.

28. Iglesias G, Talavera E, González-Prieto Á, Mozo A, Gómez-Canaval S. Data Augmentation techniques in time series domain: a survey and taxonomy. *Neural Comput Appl*. 2023;35(14):10123-10145.

29. Iwana BK, Uchida S. An empirical survey of data augmentation for time series classification with neural networks. *PLoS One*. 2021;16(7):e0254841.

30. Lawhern VJ, Solon AJ, Waytowich NR, Gordon SM, Hung CP, Lance BJ. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. *J Neural Eng*. 2018;15(5).

31. Kingma DP, Ba JL. Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. Published online December 22, 2014.

32. Xu G, Ren T, Chen Y, Che W. A One-Dimensional CNN-LSTM Model for Epileptic Seizure Recognition Using EEG Signal Analysis. *Front Neurosci*. 2020;14:578126.

33. Wang X, Wang Y, Liu D, Wang Y, Wang Z. Automated recognition of epilepsy from EEG signals using a combining space–time algorithm of CNN-LSTM. *Scientific Reports 2023 13:1*. 2023;13(1):1-12. Accessed February 6, 2024. https://www.nature.com/articles/s41598-023-41537-z

34. Sainath TN, Vinyals O, Senior A, Sak H. Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. 2015;2015-August:4580-4584.

35. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS One*. 2015;10(3):e0118432.

36. Cook J, Ramadas V. When to consult precision-recall curves. *Stata Journal*. 2020;20(1):131-148.

37. Litvak V, Jha A, Eusebio A, et al. Resting oscillatory cortico-subthalamic connectivity in patients with Parkinson's disease. *Brain*. 2011;134(Pt 2):359-374.

38. Olaru M, Cernera S, Hahn A, et al. Motor network gamma oscillations in chronic home recordings predict dyskinesia in Parkinson's disease. *Brain*. Published online January 9, 2024. doi:10.1093/BRAIN/AWAE004

39. Neumann WJ, Staub-Bartelt F, Horn A, et al. Long term correlation of subthalamic beta band activity with motor impairment in patients with Parkinson's disease. *Clinical Neurophysiology*. 2017;128(11):2286-2291.

40. Williams D, Kühn A, Kupsch A, et al. The relationship between oscillatory activity and motor reaction time in the parkinsonian subthalamic nucleus. *Eur J Neurosci*. 2005;21(1):249-258.

41. Oehrn CR, Cernera S, Hammer LH, et al. Personalized chronic adaptive deep brain stimulation outperforms conventional stimulation in Parkinson's disease. *medRxiv*. Published online August 8, 2023. doi:10.1101/2023.08.03.23293450

42. Ilyas A, Alamoudi OA, Riley KO, Pati S. Pro-Ictal State in Human Temporal Lobe Epilepsy. *NEJM Evidence*. 2023;2(3).

43. Arcot Desai S, Afzal MF, Barry W, et al. Expert and deep learning model identification of iEEG seizures and seizure onset times. *Front Neurosci*. 2023;17:1156838.

44. Abdelhameed A, Bayoumi M. A Deep Learning Approach for Automatic Seizure Detection in Children With Epilepsy. *Front Comput Neurosci*. 2021;15:650050.

45. Ricciardi L, Fischer P, Mostofi A, et al. Neurophysiological Correlates of Trait Impulsivity in Parkinson's Disease. *Mov Disord*. 2021;36(9):2126-2135.

46. Yeh CH, Zhang C, Shi W, Lo MT, Tinkhauser G, Oswal A. Cross-Frequency Coupling and Intelligent Neuromodulation. *Cyborg and Bionic Systems*. 2023;4.

47. Yeh CH, Xu Y, Shi W, et al. Auditory cues modulate the short timescale dynamics of STN activity during stepping in Parkinson's disease. *Brain Stimul*. Published online April 16, 2024. doi:10.1016/J.BRS.2024.04.006

48. West TO, Duchet B, Farmer SF, Friston KJ, Cagnan H. When do bursts matter in the primary motor cortex? Investigating changes in the intermittencies of beta rhythms associated with movement states. *Prog Neurobiol*. 2023;221:102397.

49. Duchet B, Ghezzi F, Weerasinghe G, et al. Average beta burst duration profiles provide a signature of dynamical changes between the ON and OFF medication states in Parkinson's disease. *PLoS Comput Biol*. 2021;17(7):e1009116.

50. Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci*. 2005;9(10):474-480.

51. Bastos AM, Vezoli J, Fries P. Communication through coherence with inter-areal delays. *Curr Opin Neurobiol*. 2015;31:173-180.

52. Cagnan H, Mallet N, Moll CKE, et al. Temporal evolution of beta bursts in the parkinsonian cortical and basal ganglia network. *Proc Natl Acad Sci U S A*. 2019;116(32):16095-16104.