Basic mechanisms of DBS for Parkinson’s disease: computational and experimental studies on neural dynamics
Çağnan, H.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 5

Automatic subthalamic nucleus detection from microelectrode recordings based on noise level and neuronal activity

Authors

Hayriye Cagnan, Kevin Dolan, Xuan He, Maria Fiorella Contarino, Richard Schuurman, Pepijn van den Munckhof, Wytse J. Wadman, Lo Bour, Hubert C.F. Martens

Under Review
Abstract

Micro electrode recording (MER) along surgical trajectories is commonly applied for refinement of the target location during Deep Brain Stimulation (DBS) surgery. In this study, we utilize automatically detected MER features in order to locate the subthalamic nucleus (STN) employing an unsupervised algorithm. The automated algorithm makes use of background noise level, compound firing rate and power spectral density along the trajectory and applies a threshold based method to detect the dorsal and the ventral borders of the STN. Depending on the combination of measures used for detection of the borders, the algorithm allocates confidence levels for the annotation made (i.e. high, medium and low). The algorithm has been applied to 258 trajectories obtained from 84 STN DBS implantations. MERs used in this study have not been pre-selected or pre-processed and include all the viable measurements made. Out of 258 trajectories, 239 trajectories were annotated by the surgical team as containing the STN versus 238 trajectories by the automated algorithm. The agreement level between the automatic annotations and the surgical annotations is 88%. Taking the surgical annotations as the golden standard, across all trajectories, the algorithm made true positive annotations in 231 trajectories, true negative annotations in 12 trajectories, false positive annotations in 7 trajectories and false negative annotations in 8 trajectories. We conclude that our algorithm is accurate and reliable in automatically identifying the STN and locating the dorsal and ventral borders of the nucleus; and in a near future could be implemented for on-line intra-operative use.
Chapter 5

5.1 Introduction

Deep brain stimulation (DBS) of the subthalamic nucleus (STN) is a widely used surgical technique in the management of late stage Parkinson’s disease (PD) motor symptoms. Amongst others, the efficacy of DBS is dependent on accurate localization of the target nucleus (Israel and Burchiel 2004; Molinuevo et al. 2003; Priori et al. 2003; Romanelli et al. 2004). Therefore, during DBS surgery, micro electrode recording (MER) of neuronal activity along the surgical trajectory is often performed to refine target location (Benazzouz et al. 2002; Hamel et al. 2003; Hutchinson et al. 1998; Israel and Burchiel 2004; Kim et al. 2006; Molinuevo et al. 2003; Priori et al. 2003; Temel et al. 2007; Zhuang and Li 2003). The main aim of MER mapping is accurate delineation of the functional boundaries of the STN and its surrounding structures (Benazzouz et al. 2002; Israel and Burchiel 2004; Kim et al. 2006; Molinuevo et al. 2003; Priori et al. 2003; Romanelli et al. 2004; Starr 2002). Typically, audio and visual conversions of the MERs are monitored and assessed by experts during surgery (Kim et al. 2006). Based on characteristic MER patterns, functional targets are identified. Dorsal to the STN, the thalamus is passed showing relatively slowly firing and bursting activity patterns (Benazzouz et al. 2002; Falkenberg et al. 2006; Israel and Burchiel 2004; Novak et al. 2007; Wong et al. 2009; Zaidel et al. 2009). In PD, hyperactivity of the STN neurons is reflected as increased background noise level, high firing rate, and irregular bursting activity patterns (Benazzouz et al. 2002; Falkenberg et al. 2006; Hutchinson et al. 1998; Israel and Burchiel 2004; Novak et al. 2007; Wong et al. 2009; Zaidel et al. 2009). Ventral to the STN, the substantia nigra pars reticulata (SNr) may be encountered, which is characterized by more regularly firing units (Benazzouz et al. 2002; Israel and Burchiel 2004). Correct interpretation of the MERs is important for targeting but requires time and expertise and can be challenging especially under surgical circumstances. As the number of centers performing DBS surgeries continues to grow there is a need for a system that can support surgical teams in real-time with reliable and objective identification of the target nucleus from the MERs. Several studies have addressed MER based automatic localization and visualization of the STN (Falkenberg et al. 2006; Moran et al. 2006; Novak et al. 2007; Wong et al. 2009;
Zaidel et al. 2009). Commonly utilized signal features are the background noise level and spike count (Falkenberg et al. 2006; Moran et al. 2006; Novak et al. 2007; Wong et al. 2009; Zaidel et al. 2009). Since synchronized oscillatory activity patterns in the theta and beta frequency bands seem to be characteristic in PD patients, power spectral density (PSD) based measures also have been used for STN detection and even for discrimination of the sub-territories of the STN (Brown and Williams 2005; Chen et al. 2007; Chen et al. 2006; Falkenberg et al. 2006; Moran et al. 2008; Pesenti et al. 2003; Priori et al. 2003; Trottenberg et al. 2006; Zaidel et al. 2009).

The main objective of this study is the development and validation of an unassisted algorithm that identifies the STN and the SNr from surgical MERs. To this end, we firstly assess the predictive value of the background noise level, firing rate and PSD for delineation of the STN borders. We then construct an algorithm to automatically identify the STN and the SNr using these measures. The algorithm detects the dorsal and ventral borders of the STN using different combinations of the measures and allocates qualitative confidence levels (i.e. high, medium, low) based on combinations of measures used. To test the algorithm and assess its reliability, we have applied it to a large number of consecutive recordings without any pre-selection of the data and compared its results to the surgical annotations and to independently made offline annotations by two MER experts.
5.2 Patients and Methods

5.2.1 MERs, surgical and expert annotations

Over a period of four years, 48 PD patients received DBS in the STN (84 electrode implantations) at the Academic Medical Center of Amsterdam. For each hemisphere, one to five MER trajectories were performed. MERs were collected by the Department of Neurology/Clinical Neurophysiology of the Academic Medical Center of Amsterdam and were acquired using the Leadpoint™ system (Medtronic Inc.) (Kim et al. 2006). Details about the surgical procedure and MER acquisition are provided as supplementary material (SM1). Anonymized MERs (258 MER trajectories, 6064 recording sites) were retrospectively analyzed for this study.

For the 84 STN implantations, annotations made during the DBS surgery were retrieved. These surgical annotations indicated dorsal and ventral boundaries of the STN as well as the entry to the SNr. In addition, for a set of 42 randomly chosen trajectories expert annotations were independently created off-line by two MER experts (LB and MFC), blinded to the surgical annotations. These 42 trajectories have been used as training data sets to optimize the algorithm.

5.2.2 Automatic STN Detection Algorithm

a) Signal Processing

Prior to automatic localization of the STN and the SNr, signal processing is used for delineation of noise, artifacts and spikes in the MER. First, the noise-level is estimated using the envelope of the MER (Figs. 1A and 1B) (Dolan et al. 2009). The estimated noise-level is used to detect high amplitude artifacts and frequency spectrum of the MER is used to detect low amplitude artifacts (Fig. 1A). Automatically detected artifacts are removed from the MER and excluded from further analysis. Sections of the MER exceeding an amplitude threshold, which is based on the estimated noise-level, are compared to a spike template and are marked either as spikes or as low amplitude artifacts (Fig. 1B). Sections of the MER, not meeting the spike criteria and subsequently marked as low amplitude artifacts are also excluded from further analysis. More details on noise-level estimation, artifact
Automatic STN detection

removal and spike detection can be found in supplementary materials (SM2).

![Figure 1: A: Example of 5 second MER epoch containing mechanical artifacts automatically detected by the software (in grey). Background signal noise is given in black. B: Example of 25 ms MER epoch containing automatically detected spiking activity (highlighted by grey boxes).]

b) Feature Set

For reliable STN detection, robust and distinctive MER features need to be extracted. We computed noise level, compound firing rate and measures based on PSD, low band index (3-12 Hz), beta band index (13-30 Hz) and gamma band index (31-100 Hz) (supplementary material, SM3), and determined which features showed the strongest correlation with the surgical annotations. In agreement with previous work, we found that (ordered with decreasing selectivity) noise level, compound firing rate, gamma and beta power were good indicators for the presence of the STN (Falkenberg et al. 2006; Trottenberg et al. 2006; Wong et al. 2009; Zaidel et al. 2009). Changes in power in the 3-12 Hz range were not specific to the STN. Therefore the low band index has not been included in the feature set used in automatic localization of the STN. The methods used for calculating noise-level,
compound firing rate and PSD are described in the supplementary material (SM3).

c) STN and SNr detection

A threshold based algorithm is constructed for detection of the STN and the SNr. The algorithm uses as inputs the estimated noise level, the compound firing rate, and beta band and gamma band indices, observed at each site along a trajectory. Details on noise level, firing rate, beta and gamma band thresholds can be found in the supplementary materials (SM4).

Depending on combinations of measures used in delineation of the STN borders, the algorithm allocates different qualitative confidence levels: high, medium or low (Figs. 2 and 3).

i. Annotations with High Confidence

High confidence STN detection is achieved when both increased noise level and increased neuronal unit activity (high firing rate and raised beta or gamma PSD) in a particular MER is detected. From dorsal to ventral, the first site along the trajectory exceeding the noise level threshold, the firing rate threshold and either the beta band threshold or the gamma band threshold is marked as being within the target structure (site \(x\)) and the corresponding trajectory is flagged as containing STN. From site \(x\) the algorithm searches the dorsal and ventral boundaries by marking the sites \((y\) and \(yy\) respectively) where MER noise level drops and stays below the noise level threshold. The dorsal boundary is further refined by incrementally shifting the boundary in the dorsal direction (site \(w\)) to include consecutive sites exceeding the firing rate threshold and either the beta band threshold or the gamma band threshold.

ii. Annotations with Medium Confidence

In case the firing rate threshold is not exceeded within the region which has exceeded the noise level threshold, the algorithm searches for consecutive sites of MERs exceeding the noise level threshold which are subsequently marked as STN.
Automatic STN detection

Figure 2: Schematic representation of the automatic STN detection algorithm: At each recording site, first, noise level is estimated, artifacts are automatically removed and spikes are detected. PSD is computed, after artifacts have been removed. Then, noise level, firing rate, beta (13-30 Hz) and gamma (31-100 Hz) band indices are computed at each recording site of a trajectory. Per trajectory, noise level, firing rate, beta and gamma band thresholds are calculated. Depending on the combination of features which exceed the threshold, annotations are made per trajectory using three different confidence levels: high, medium and low.

If the sites exceeding the firing rate threshold also exceed either the beta band threshold or the gamma band threshold, precede and are adjacent to the consecutive sites exceeding the noise level threshold, the dorsal border of the STN is shifted to include the sites exceeding the firing rate threshold.
Figure 3: A: annotation made based on noise level, firing rate and PSD changes. Site x is at -2.5 mm; where all measures exceed their respective thresholds. Dorsal border is at -2.5 mm (site y) and ventral border is at 2 mm (site yy). For this trajectory, the automatic annotation is in full agreement with the surgical annotations. B: Region which exceeds noise level threshold (at -1.5 to -1 mm) does not coincide with the region which exceeds firing rate threshold (-2 mm). The algorithm has indicated that STN spans from -2 mm to -1 mm. The surgical team has annotated the region from -3 mm to -0.5 mm.
Automatic STN detection

mm as STN. C: Annotation has been made based on firing rate and PSD. The algorithm has annotated the region from 1 mm to 2.5 mm as STN. In this example, the automated annotation fully agrees with the surgical annotation. Beta band (13-30 Hz) and gamma band (31-100 Hz) indices are equivalent to $10 \log \left( \frac{\text{average power in 13-30 Hz}}{\text{average power}} \right)$ and $10 \log \left( \frac{\text{average power in 31-100 Hz}}{\text{average power}} \right)$, respectively.

iii. Annotations with Low Confidence

In case the noise level threshold is not exceeded in a given trajectory, the algorithm searches for consecutive clusters of MERs exceeding the firing rate threshold and either the beta band threshold or the gamma band threshold. These sites are subsequently marked as STN.

From the ventral border of the STN (site yy), the algorithm searches for a supra-threshold increase in the noise level and the firing rate, and annotates these sites as SNr. The algorithm requires a gap of one site recording between STN and SNr for annotation. For the cases that in the trajectory, there is no gap between STN and SNr, the algorithm is not able to annotate SNr.

5.2.3 Statistical analyses used to determine the feature set

Correlation of features (i.e. noise-level, compound firing rate and low band, beta band and gamma band indices) with STN anatomy was determined by comparing feature values inside the surgically annotated STN with the ones outside the STN. Only trajectories containing surgically annotated STNs of length greater than 2 depths have been included in the analysis. We applied a paired $t$-test to detect changes in noise level, compound firing rate and low band, beta band and gamma band indices observed on average inside the first half and the second half of surgically annotated STN to average values outside the STN ($p=0.01$ indicated statistical significance).

Using surgical annotations to define dorsal and ventral borders of the STN, we computed if a measure was significantly higher ($p=0.05$) at the dorsal border of the STN than the average value of a measure observed at the preceding sites and if measures observed at the site following
the ventral border of the STN were significantly less \( (p=0.05) \) than the average value of the measure observed inside the STN.

5.2.4 Validation of the algorithm

Validation is performed by applying the automatic algorithm to the MERs and comparing its output to the surgical (258 trajectories, 6064 sites) or expert annotations (42 trajectories, 1011 sites). The agreement percentage between different annotation sets is computed by calculating the ratio between the number of sites that the two annotations agree on and the total number of sites. Inter-rater reliability is calculated based on Cohen’s Kappa statistic (Cohen 1960). Taking surgical annotations as the golden standard, false negatives were defined as the number of trajectories annotated by the surgical team and not by the algorithm. False positives were defined as the number of trajectories annotated by the algorithm and not by the surgical team. Moreover we tested the accuracy of the algorithm in detecting the dorsal and ventral border of the STN, by calculating the difference between the surgically annotated ventral and dorsal STN borders and the automatically detected borders.
5.3 Results

5.3.1 Predictive value of Noise Level, Firing Rate and Power Spectral Density

Based on 239 out of 258 trajectories, where surgical annotations indicated the presence of the STN, we assessed 1) if measures such as noise level, firing rate and PSD are significantly different inside the STN from the values observed outside the STN, and 2) in what percentage of the trajectories these measures were significantly different from the values preceding the surgically annotated dorsal boundary and ventral boundary of the STN.

Noise level, compound firing rate and beta band and gamma band indices all showed higher values throughout the STN ($p=0.01$). The low band index showed no differentiation between outside and inside the STN ($p=0.01$).

<table>
<thead>
<tr>
<th></th>
<th>Noise Level</th>
<th>Firing Rate</th>
<th>Beta Band</th>
<th>Gamma Band</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dorsal</td>
<td>70 %</td>
<td>47 %</td>
<td>18 %</td>
<td>17 %</td>
</tr>
<tr>
<td>Ventral</td>
<td>36 %</td>
<td>26 %</td>
<td>19 %</td>
<td>28 %</td>
</tr>
</tbody>
</table>

Table 1: Percentage of trajectories ($n = 239$) where a measure was significantly higher ($p = 0.05$) at the dorsal border of the STN than the average value of a measure observed at the preceding sites and where a measure observed at the site following the ventral border of the STN was significantly less ($p = 0.05$) than the average value of the measure observed inside the STN.

Additionally, using surgical annotations as indicators for the dorsal and the ventral borders of the STN, we computed if a measure was significantly higher ($p=0.05$) at the dorsal border of the STN than the average value of the measure observed at the preceding sites and if measures observed at the site following the ventral border of the STN were significantly less ($p=0.05$) than the average value of the measure observed inside the STN (Table 1). Results indicate that noise level and firing rate increase are good indicators for the dorsal border of the STN. Moreover, decrease in noise level and firing rate between inside the STN and the site following the ventral border (exiting the STN) are not as pronounced as the increase in noise level and firing rate.
between outside the STN and the site marking the dorsal border (entering the STN).

5.3.2 Comparison to Expert Annotations

<table>
<thead>
<tr>
<th></th>
<th>Automatic algorithm</th>
<th>Surgical annotation (AAP-IRR)</th>
<th>Expert 1 (AAP-IRR)</th>
<th>Expert 2 (AAP-IRR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatic algorithm</td>
<td></td>
<td>89% - 0.75</td>
<td>88% - 0.73</td>
<td>87% - 0.71</td>
</tr>
<tr>
<td>Surgical annotation</td>
<td></td>
<td>90% - 0.77</td>
<td>90% - 0.78</td>
<td></td>
</tr>
<tr>
<td>Expert 1</td>
<td></td>
<td></td>
<td></td>
<td>93% - 0.85</td>
</tr>
<tr>
<td>Expert 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Average agreement percentages (AAP) and inter-rater reliability (IRR) calculated over 42 randomly chosen trajectories (n=1011): Agreement percentage is the ratio between the total number of sites that the two sets of annotations agree on and the total number of sites recorded. Inter-rater reliability is based on Cohen’s kappa statistic.

The automated algorithm had 88% agreement with surgical annotations made on 84 patient data sets (258 trajectories, 6064 sites) and 87-88% agreement with the two experts’ independent annotations made on 42 randomly chosen trajectories (1011 sites). Additionally, we have compared the agreement level between the clinical experts and each expert with the surgical annotations made on the 42 trajectories in order to establish the inter-rater reliability (Table 2).

Out of 258 trajectories, the surgical team has indicated the presence of the STN in 239 trajectories and the absence of the STN in 19 trajectories. Taking surgical annotations made on 258 trajectories as the golden standard, our automatic algorithm has made true positive annotations in 231 trajectories and true negative annotations in 12 trajectories. The algorithm has performed false positive annotation of the STN in 7 trajectories and false negative annotation in 8 trajectories (Table 3). From any given data set, STN annotations have been made on one or more trajectories and false negatives do not occur at any of the channels which have been later on chosen as the final electrode position by the surgical team.
Automatic STN detection

<table>
<thead>
<tr>
<th>Trajectory annotation</th>
<th>Agreement with surgical annotations</th>
<th>TP</th>
<th>FP</th>
<th>TN</th>
<th>FN</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>88% (n=5192)</td>
<td>216</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>82% (n=141)</td>
<td>6</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>84% (n=262)</td>
<td>9</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No STN annotated</td>
<td>92% (n=469)</td>
<td>12</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>88% (n=6064)</td>
<td>231</td>
<td>7</td>
<td>12</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 3: Average percentages of agreement with the surgical annotations indicate that annotations made with high confidence have the highest agreement with the surgical annotations. Value of true positive (TP) annotations indicates the total number of trajectories containing STN according to both the surgical team and the automated algorithm. Value of false positive (FP) annotations indicates the total number of trajectories containing STN according to the automated algorithm but not the surgical team. Value of true negative (TN) annotations indicates the total number of trajectories not containing STN according to both the surgical team and the automated algorithm. Value of false negative (FN) annotations indicates the total number of trajectories not containing STN according to the automated algorithm but not the surgical team.

<table>
<thead>
<tr>
<th>Algorithm Confidence</th>
<th>Detection of the dorsal border (mean ± SD)</th>
<th>Detection of the ventral border (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High (5107 sites)</td>
<td>0 ± 2 depths</td>
<td>0 ± 2 depths</td>
</tr>
<tr>
<td>Medium (141 sites)</td>
<td>-1 ± 2 depths</td>
<td>2 ± 3 depths</td>
</tr>
<tr>
<td>Low (193 sites)</td>
<td>-2 ± 2 depths</td>
<td>0 ± 2 depths</td>
</tr>
<tr>
<td>Overall (5441 sites)</td>
<td>0 ± 2 depths</td>
<td>0 ± 2 depths</td>
</tr>
</tbody>
</table>

Table 4: Algorithm’s accuracy in detecting the dorsal and the ventral borders of the STN are computed by taking the average difference and the standard deviation of the difference between the annotations made by the surgical team and the automatic algorithm (231 trajectories). Negative values indicate that on average annotations made by the algorithm are more dorsal than the surgical annotations. Positive values indicate that on average annotations made by the algorithm are more ventral than the surgical annotations.
Over 231 trajectories identified to contain the STN both by the surgical team and the automatic algorithm, the accuracy in detecting the dorsal and ventral border is $0 \pm 2$ sites (MER recordings made at 0.5mm intervals) (Table 4).
5.4 Discussion

In this chapter, we have developed and validated an algorithm for automated detection of the STN from MERs. We tested our method by comparing it to surgical annotations and found a high agreement level (88%) and good inter-rater reliability (0.71-0.73), comparable to the inter-rater reliability between different experienced raters.

Our novel automated algorithm outlined in this study can be implemented to run in real-time to assist clinicians in STN localization during DBS surgery. This would increase the reliability of MER interpretation and reduce surgical time. To test the algorithm’s robustness as needed for intra-operative usage, all levels of MER processing have been applied to a large number of consecutive trajectories without data pre-selection, (manual) data pre-processing or data cleaning. We thus aimed to avoid any bias toward clean “easier-to-interpret” recordings and to give an estimate of the reliability of our algorithm in “real life” situations. Statistical analyses demonstrated significant changes in noise-level, firing rate and PSD based measures inside the STN. However, individually, these measures are sub-optimal indicators of the STN boundaries (Table 3). Therefore our automatic algorithm makes use of combinations of features to detect the STN. The algorithm assigns a qualitative confidence level (high, medium, low) based on the combination of features used in detection of the STN (Figs. 2 and 3).

In 92% of the cases (220 out of 238 automatic annotations) the algorithm detected the STN based on typical STN hallmark MER features: increased background noise, increased firing rate, and oscillatory activity; and automatic annotations have been made with high confidence level (Falkenberg et al. 2006; Hutchinson et al. 1998; Israel and Burchiel 2004; Novak et al. 2007; Wong et al. 2009; Zaidel et al. 2009). In these cases, the agreement with surgical annotations (88%) is excellent and comparable to agreement between surgical annotations and off-line expert annotations. In the remaining cases, the three MER measures were not detected to be simultaneously increased in any individual recording and hence STN detection was performed with lower level of confidence (medium in 6/238 and low in 12/238). Combining measures for STN detection increases the accuracy of the algorithm and makes it more robust. Substantial changes in
the thresholds used in the algorithm have limited effect on the accuracy of the algorithm (Supplementary materials).

The algorithm has detected the STN in 7 trajectories which have not been annotated by the surgical team (i.e. false positives). Three of these annotations were made with low confidence. Visual inspection of the remaining four cases, where STN has been detected erroneously, revealed that the noise level estimate has been corrupted by low amplitude artifacts embedded in the noise, giving rise to artificially elevated noise-levels. Moreover, 8 trajectories which have been annotated by the surgical team have been missed by the algorithm (i.e. false negatives). This is due to noise level changes remaining sub-threshold and not having consecutive sites exhibiting high firing rate. Decreasing the noise-level threshold to reduce the number of false negatives would hamper the accuracy of dorsal and ventral border detection more than increasing the accuracy of detection.

In the present algorithm, noise level detection is important both for STN boundary detection and for spike-detection and artifact rejection. Hence an accurate and reliable estimate of the noise level is required. Several methods have been proposed for noise level estimation either using the root mean square (RMS) of the signal, the median of the distribution of the absolute value of the signal, or the envelope of the signal (Dolan et al. 2009; Pancrazio et al. 2003; Quiroga et al. 2004; Watkins et al. 2004). Previously, it has been demonstrated that presence of high frequency firing and signal artifacts could result in over estimation of the noise level for the RMS and median based methods, markedly leading to sub-optimal spike detection (Dolan et al. 2009; Quiroga et al. 2004). Noise level estimation based on the envelope of the signal is used in this study since it is more robust in the presence of high-frequency firing and artifacts (Dolan et al. 2009). Moreover, the envelope of the signal also can be used to determine which segments of the recording are artifacts since in this case the presence of high-amplitude artifacts hardly influences estimation of the noise level. The use of the envelope of the signal removes an important previous constraint in automatic MER signal processing which is that an artifact cannot be reliably detected without first estimating the noise level and that the noise level cannot be estimated accurately in the presence of artifacts when
methods based on the standard deviation or the median of the signal are being used.

Previous studies have outlined methods for automatic detection and visualization of the STN based on objective and quantitative MER features (Falkenberg et al. 2006; Moran et al. 2006; Novak et al. 2007; Wong et al. 2009; Zaidel et al. 2009). In agreement with Wong et al. (2009), who described a STN visualization algorithm, we have observed that background noise and firing rate most consistently differentiated the STN from its surroundings. Falkenberg et al. (2006) demonstrated that appropriate visualization of power spectrum based measures overlapped well with surgical annotations. Zaidel et al. (2009) combined PSD and noise level to locate the STN and its sub-territories. Sub-territories were determined based on changes in the 13-30 Hz frequency band and these changes indicated oscillatory segments of the STN (Zaidel et al. 2009). In this study, we have investigated changes in the 3-12 Hz (i.e. low band index), 13-30 Hz (i.e. beta band index) and 31-100 Hz (i.e. gamma band index) frequency bands. We have observed that units inside the STN show significantly higher activities in the beta and gamma bands than the units outside of the STN ($p=0.01$). In agreement with Zaidel et al. (2009) we observed that these parameters per trajectory are valuable for localizing oscillatory regions within the STN. On the other hand, compared to noise level and compound firing rate, these parameters are less reliable for delineating dorsal and ventral borders of the STN (Table 1). Measures based on PSD, though, can be used for STN localization, provided they are combined with other measures such as noise level and firing rate.

In summary, the algorithm presented in this study reliably detects trajectories containing the STN and identifies the dorsal and the ventral borders of the STN with good accuracy. The automatic results show good consistency with the ratings made by experienced raters. The use of a simple threshold based method opens the door for this method to be implemented real-time during surgery leading to an on-line, easy and objective tool to support STN localization during DBS surgery.
5.5 Supplementary Materials

SM1 Patients, surgical procedure and MERs

The procedure for DBS was a one stage bilateral stereotactic approach, using frame-based 3D T1 and T2 weighted MRI reconstructions for target localization and path-planning, in conjunction with MER and macro test stimulation. The standard STN coordinates used were 12 mm lateral, 2 mm posterior and 4 mm below the mid-commissural point. When necessary, adjustments were made based on MRI anatomy. Surgical paths for the central channel was defined using the following criteria: anterior angulation to the inter-commissural line of 15-20°, lateral angulation from midline 20-30°, entry on top of a gyrus, and avoiding sulci, cortical surface veins, and the lateral ventricles. The implanted leads were model 3389 (Medtronic, Minneapolis, MN, USA).

All patients were awake during the surgical procedure and none of the patients were under sedatives. Surgery and MER were performed following overnight withdrawal of anti-parkinsonian medication. Extra-cellular single/multi-unit MER were performed using polyamide-coated tungsten microelectrodes (Medtronic microelectrode 291; impedance ~1.1MΩ measured at 220Hz, before starting recordings) with 10 µm exposure, mounted on a sliding cannula. The electrodes were placed in an array with central, lateral, medial, anterior and posterior positions placed 2 mm apart (i.e. Ben’s gun). Depending on the pre-operative MRI and clinical conditions of the patient, it was decided in some cases to record with less than five needles. Signals were recorded using amplifiers (10,000 times amplification) of the Leadpoint system (Medtronic), via a bootstrapping principle and were analog band-pass filtered between 500 and 5000 Hz (12 dB / oct). Signals were sampled at 12 kHz, using a 16-bit A/D converter and afterwards up-sampled to 24 kHz off-line. Starting from a remote position (i.e. approximately 6 mm to 8 mm dorsal to the MRI-based target (0 mm)), needles were advanced in steps of 0.5 mm towards the target, using a manual micro-drive. Recordings were stopped when electrical activity typical of substantia nigra cells was recognizable in at least one of the electrodes or when a significant decrease of electrical activity was present in all the recordings. Subsequently, intra-operative clinical testing was performed along selected trajectories at several sites. The definitive
electrode for chronic stimulation was implanted at the site with the best therapeutic window, i.e. best benefit on motor symptoms and highest threshold for side-effects.
SM2 Automatic MER Signal Processing

a) Noise level estimation

Noise level is estimated using the envelope of the MER signal (Dolan et al. 2009). The envelope of the signal is estimated as follows. First, the Hilbert Transform is taken of the original signal $x(t)$.

$$H(x(t)) = \int_{-\infty}^{\infty} \frac{x(t-\tau)}{\pi \tau} d\tau$$

Then, an analytical signal is constructed whose real part is $x(t)$ and imaginary part is $H(x(t))$. The instantaneous amplitude of this analytical signal is an estimate of the signal envelope (Dolan et al. 2009). For Gaussian distributed noise, the distribution of the instantaneous amplitudes of the signal is equivalent to the Rayleigh distribution, where the distribution’s mode corresponds to the standard deviation of the background noise. For realistic signals encountered during DBS surgery (i.e. containing high frequency firing and artifacts) the mode proves a robust statistic and is used as our noise level estimate (Dolan et al. 2009).

b) Artifact detection

Mechanical artifacts are common in MER and may lead to misinterpretation of the data. Therefore, prior to spike detection, artifacts are removed automatically from the signal in order to prevent spike dependent statistics (i.e. compound firing rate) and spectral measures from being biased. High amplitude artifacts are detected automatically by using a combined amplitude and frequency criterion relative to the estimated noise level; any artifact contaminated data segment is suppressed from further analysis (Fig. 1A). Amplitude criteria (i.e. 7 times the estimated noise level) and frequency criteria were determined empirically by comparing the automatic artifact rejection outcome to expert assessment of the same data (data assessment performed by HC, KD and LB). For the frequency analysis, the MER is divided into 50 ms windows and Fourier analysis is applied to each window. If the maximum frequency component of the 50 ms window is 2.5 times higher than the median maximum frequency component, then the window is marked as containing
artifacts. Smaller amplitude artifacts that may remain in the data are subsequently rejected during spike detection.

c) Spike detection

All signal segments exceeding 4 times the estimated noise level are flagged as events. Consecutive events are combined into biphasic spike candidates. By comparing spike candidates to a spike template, a spike candidate is either accepted as a spike or is flagged as false positive (Fig. 1B). Our spike template imposes a peak to peak spike width < 1 ms and total spike duration < 3 ms. No further spike sorting is performed for the present work.
SM3 Extracted MER features

a) Noise Level

Noise level is determined by re-applying the envelope based method to the MER data after initial automatic artifact removal (Dolan et al. 2009).

b) Compound Firing Rate

Compound firing rate at a specific site is the ratio between the total number of detected spikes and the total recording period in seconds.

c) Power Spectral Density

Power spectral density (PSD) of the MER is obtained by computing the power spectrum of the rectified MER after the artifacts have been automatically removed and the mean is subtracted (Zaidel et al. 2009). The power spectrum of the rectified signal is estimated by using the Welch’s method with a window length of 1 second and 50% overlapping windows. In order to quantify the oscillatory patterns, we defined the low band index (3-12 Hz), beta band index (13-30 Hz) and gamma band index (31-100 Hz). All indices are computed by taking the ratio between the average power in the pre-defined frequency band and the average power in the spectrum.
Threshold levels along each trajectory are made dependent on the MERs as follows. Noise level threshold is set to a multiple of the noise level observed from the first site of the trajectory up to 3 mm above the estimated target location (site 0 mm) \( \text{Noise level threshold} = \varphi \times \text{median [noise level up to -3mm]} \). Firing rate threshold is set to a multiple of the standard deviation of the firing rate with respect to the mean firing rate observed along the trajectory \( \text{Firing rate threshold} = \text{mean [firing rate]} + \chi \times \text{standard deviation [firing rate]} \). Beta band threshold is equivalent to the average beta power observed along the trajectory and gamma band threshold is set to the average gamma power observed along the trajectory.

**Figure SM4:** Relationship between noise level threshold, firing rate threshold, and agreement percentage with two expert annotations in the training set. A: The highest agreement percentage with the expert annotations is achieved when the noise level threshold is set at 1.3 times the median of the estimated baseline noise level (i.e. \( \varphi = 1.3 \)). Noise level estimation relies on the envelope of the signal. B: Agreement percentage with expert annotations is less sensitive to
changes in the firing rate threshold, since annotations predominantly depend on changes in the noise level. Therefore, the firing rate threshold is set at the mean firing rate observed along the trajectory (i.e. \( \chi = 0 \)).

The parameters (\( \varphi \) and \( \chi \)) determining the noise level and firing rate threshold levels were tuned by optimizing the agreement percentage between the automated annotation and the two expert annotations provided for the 42 trajectories (Fig. SM4).
References


Automatic STN detection


