Analysis and Classification of Microelectrode Recordings in Deep Brain Stimulation Surgery

José Carlos F. V. Lima¹, MD. Manuel H. de Carvalho², Prof. José M. B. Dias¹
¹Instituto Superior Técnico – Universidade de Lisboa, ²Hospital Universitário Santa Maria - Lisboa

Abstract – In late years neurosurgery has had an increased acceptance in scientific community for the treatment of Parkinson’s Disease due to the good outcomes of Deep Brain Stimulation (DBS) clinical procedure. Of utmost importance for its success is the precise location of the very small target region in the brain. In the current study we research the most effective methods for the automated classification of the spontaneous activity of this target and develop a decision support software application in Matlab to accurately identify the optimum location for the implant of chronic stimulation electrode directly from spontaneous local neural activity. We show it is possible to estimate the human labeling of DBS microelectrode recordings, as belonging to the Subthalamic Nucleus region, with an average accuracy of at least 92.8%, using a Multinomial Logistic Classifier, trained with the first 25 principal components of the extracted features, from a dataset composed of 1000 signals from 4 patients only. This system has the potential to improve performance of the medical teams and reduce DBS surgery morbidity indicators.

Index Terms – Deep Brain Stimulation; DBS; targeting automation; Parkinson’s Disease; MER; biomedical signal processing; pattern recognition; classification.

I. INTRODUCTION

Deep Brain Stimulation (DBS) is a therapeutic technique particularly suitable to treat the symptoms of advanced drug refractory Parkinsonism, essential tremor and dystonia (Lyons, 2011). It consists of the stimulation of deep structures in the brain by means of implanted electrodes connected to an external electronic device, as to modulate its electric activity and improve the functions degraded by the pathological condition.

DBS exact mechanism is still not fully understood, nevertheless, on the last twenty years, scientists have shown evidence that its use can dramatically reduce motor disorder symptoms in Parkinson’s Disease (PD) (Ardouin C, 1999; Bejjani BP, 2000; Burchiel KJ, 1999; Houeto JL, 2000; Krack P P. P., 1998; Molinuevo, 2003; Koller W, 1997; Kumar K, 1999; Limousin P, 1999; Simuni, 2002) e.g.: bradykinesia; dyskinesia; tremor; gait disturbance and rigidity (Krack P B. A., 2003; Pahwa, 2009; Kleiner-Fisman G, 2003; Anderson VC, 2005; Krause M, 2001; Fahn, 2003).

Subthalaric Nucleus (STN) is today’s preferred implantation location (Levine, 2003; Pralong, 2004). This is functionally divided in 3 sub-regions according to their efferent projections: the Ventromedial Limbic Region; the Associative Region; and the Dorsolateral Somatosensory Region (Bertler À, 1959; Sano I, 1959). To achieve the best results in the treatment of PD, its motor disorder symptoms, and reduce the risk of cognitive or behavioral adverse effects, DBS electrodes should be implanted as close as possible to the sensorimotor sub-region (Godinho F, 2006; Herzog J, 2004). The stimulation of adjacent functional regions has been proven to cause adverse effects like involuntary muscle contraction (Internal Capsule), diploria (Oculomotor Nerve Root), ataxia (Superior Cerebellar Peduncle), and paresthesias (Medial Lemniscus and Spinothalamic Pathways) (Bériault, Xiao, Bailey, & Collins, 2012).

Estimation of the exact coordinates of the STN is usually done with both X-Ray Computerized Tomography (CT) and Magnetic Resonance Imaging (MRI), together with standard averaged stereotactic atlases. Completely personalized mapping approaches are being increasingly used as imaging technology improve (Nakano N, 2012; Slavin, 2005; Massey, 2012). The optimal trajectory is defined for both hemispheres so as to reduce the danger of damaging blood vessels and brain tissue. To assure maximum correspondence between planned coordinates and the actual procedure, it is used a stereotactic reference frame, and final re-referencing is done by X-ray imaging of the head of the patient with the frame already in place. However, owing to present CT low anatomical discrimination and MRI spatial resolution limitations, to errors in the MRI-CT image registration process and other caused by brain shifts, the information gathered in the planning phase is not enough to validate the location of target point and the boarding trajectory. To increase precision, neuronavigation is supported by the intraoperative Microelectrode Recording (MER) of spontaneous neural activity, which is today interpreted in raw by eye of a specialized neurologist. The difficulty, time needed, and inherent subjectivity of human judgment of these electric signals calls for the development of a specific computerized decision support system.

For the present study we researched most published scientific literature to date specifically related to the DBS targeting automation problem, using a web search engine to access available free online databases like Pubmed, PlosOne and ResearchGate. We found several studies using a multitude of methods and with different scopes of
analysis, the first dating back to 2006. Table 1 systematizes most important implementation aspects which correspond roughly to the following: frequency band, signal enhancing, feature extraction and selection, region discriminated, availability of an online implementation and performance.

The main objective of the present study was then to develop and apply the signal processing and pattern recognition techniques to display, analyze, and classify MER-signals acquired during DBS surgery, with the final aim of automating the process of targeting the STN region in the brain, and if possible the optimum sub-region for the chronic implantation of a stimulation electrode. By developing and testing first offline, this project is meant to be a first step towards creating a decision support system to be used online by the medical teams, potentially increasing targeting speed, accuracy, and reliability, reducing surgery time and postoperative morbidity.

The software tool herein described is able to perform, among other, the following operations: a) Load, display, and listen simultaneously MER-signals acquired during a DBS surgery; b) Process MER-signals offline aimed at feature extraction; c) Label MER-signals acquired along the microelectrode trajectory; d) Supervised classification of MER-signals. Furthermore, most relevant signal features present in literature were implemented and tested, and most used supervised classification models performance was assessed.

II. METHODS

With the final aim of automating the targeting of the optimum implantation location for the STN-DBS electrode, present study used the Machine Learning flow which is summarized in Figure 2.

Database comprised of a total of 118 patients with an advanced state of Parkinson’s Disease (PD) that underwent Subthalamic Nucleus (STN) DBS surgery in Santa Maria Hospital in Lisbon, Portugal. Records included both intraoperative microelectrode electrophysiology Recordings of spontaneous neuronal activity, as well as the Intraoperative Classification (human labels or classes), done by an experienced neurologist assisting surgery, with a qualitative rating of the signals. These had to be first translated to digital format and added to the database to serve as the ground truth for subsequent supervised classification estimation.

MER signals were first pre-processed for enhancing signal-to-noise ratio. They were segmented and tagged, after which most significant features were extracted, selected and transformed. To estimate the classification of MER-signals, various supervised classification methods using different sets of features were tested, namely: Linear Discriminant Classifier (LDC), K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Binomial and Multinomial Logistic Regression (MLR). Optimization and performance were done using k-fold schemes.

All stages of analysis were done offline with the aid of a software application developed in Matlab specifically for this purpose and presented ahead.

A. Electrophysiology Data Acquisition

MER signals used were acquired during bilateral STN-DBS surgery using a maximum of 5 independent probing tracks parallel to the approaching surgical route axis, 2mm apart from each other, with central, medial, lateral, anterior and posterior relative positions to the body axis, as shown in the Figure 3. MER signals were acquired simultaneously from all trajectories, in a stationary state at positions consecutively closer to the target reference, which was preoperatively determined through combined MRI and CT. Probes were advanced in steps of 1 or 0.5mm, starting from position -10mm and ending at +3.5mm relative to the target. Acquisition was made at 24 kHz for periods of 10s, corresponding to a total of 24 thousand samples per signal, and with a Digital to Analog Converter (DAC) resolution of 16 bit. In average there were 240 MER signals per patient.

Recordings were done using an intraoperative microelectrode Leadpoint™ System (Medtronic Inc.). The electric probes used were monopolar microelectrodes (22660, FHC, Inc. Bowdoin, ME) with a tungsten tip with a diameter that can go down to 1-3μm, and an isolated shank in stainless steel with Ø200μm, measuring impedances between 0.5 and 10 MΩ at a frequency of 1 kHz.

Associated with each surgery session there were also made records with a qualitative evaluation for each EDT of the signal suitability for the implantation of the stimulation electrode. These, provided in paper, were digitized to build the final dataset with the signal-label pairs. The neurologist’s evaluation was available in two different ways: one in form of boxes which estimated coarsely the position range where stimulation would potentially have better effect, and another in a 6 integer scale from -1 to 4, indicating the probability of the signal belonging to the target region, more strictly related with the physical properties of the signal. Figure 4 shows an example of such a classification record.

B. Signals Pre-processing

All signals were filtered at time of acquisition by the recording software through a frequency band of 500 to 5000 Hz with an unknown method. Following, to remove low frequency drifting and high frequency noise that were still visible in the signals, we used a zero-phase filtering, as implemented in filter Matlab function, with a frequency band between 300 and 5000 Hz. The signal processing application gives total freedom to the user to define the size and location of the signal segments that will constitute the training sets. For the performance tests we segmented the original 10s MER in portions of 1s. Episodic artifacts in the frequency band of interest, which are virtually impossible to remove without distorting the signals, as well as inhomogeneities during and across signals, from different trajectories, hemisphere or patient, were naturally washed out by averaging. Robustness to these factors is improved with increasing the number of signals and reducing the segments length.

We dedicated special attention to signal preprocessing needed for the detection of spikes, which is the base of multiple features found in literature to have good discriminatory power for the region of interest. To remove
the background noise and artifacts that degraded performance of spike detection we tested several strategies: common causal and non-causal Fourier filtering; wavelet packet filtering; pattern recognition over wavelet coefficients; and statistical analysis of a set of time and frequency domain features through machine learning approaches. Figure 4 exemplifies MER pre-processing using wavelets. In the top frame of left column it is shown a segment of 1000 samples of the original signal, chosen for its good Signal-to-Noise Ratio (SNR). In the middle we present an image of the correspondent wavelet decomposition coefficients for the different time scales. For this test analysis it was used a redundant wavelet decomposition with a simple Haar template and a non-dyadic scale with a small overlap. Having the coefficient values for each sample at different scales, we then applied a Principal Component Analysis (PCA) to find the biggest variance in the feature space, and reduce its dimension to improve clustering performance. At the bottom frame we show a bidimensional scatter of every signal sample according to their two first principal components, with a color labeling done by simple thresholding. In the right column of Figure 5 we exemplify the application of the Matlab software developed in the present study, for the segmentation of the signal, feature extraction, and analysis of time domain features of a 10s MER signal.

For action potentials or spikes detection we used the noise mode (Kim, 2007), which we find using the amplitude distribution of the maxima of the Hilbert Transform of the signal, as in (Dolan, 2009). This transformation returns an almost perfect Rayleigh distribution for pure band-limited white Gaussian noise, allowing for a clear separation between a lower (noise) and a higher (spikes) mode. Having found the noise level of the signal, all maxima above 4.5 times this value were labeled as spikes (Cagnan H., 2011).

C. Feature Extraction

Signal features used in these study were gathered through an exhaustive research of scientific literature published to date specifically related to the DBS targeting automation problem using multitunit activity (MUA) MER, and available in free access online databases like Pubmed, PlosOne and ResearchGate. Some of most relevant articles found were (Hudgins B1, 1993; Kiryu, 1994; Farry, 1996; Zecca, 2002; Oskoei, 2006; Phinyomark, 2009; Solnik, 2010; Nayak, 2011; Rechy-Ramirez, 2011; Hamed, 2012). Reported most used and performant features are presented here in four categories: time domain, frequency domain, time-frequency domain, and action potentials domain.

1) Time domain features

Extracted time domain features are synthetized in Table 2. These features were design to somehow account for amplitude and slope of the waveform (1 and 2), as well as some measure of frequency (3 and 4) and duration (5). Mean Absolute Value (1), Mean Absolute Value Slope (2), Zero Crossings (3) and the Slope Sign Changes (4) can be found in (Hudgins B1, 1993). In (Huang, 2000) we can find Root Mean Square (6) and Variance (7); from (Phinyomark, 2009) we have Waveform or Curve Length (8), Zero Crossings (9), Slope Sign Changes (10), Willison Amplitude (11), and Amplitude probability distribution (17-36); Simple Square Integral (12) has been proposed by (Ciecierski, 2012); from (Kaiser, 1990; Mukhopadhyay, 1998) we have feature Average Power (13); and from (Kim, 2007) we have Noise mode (14). Authors like (Zaidel, 2009; Biopac Systems, 2010; Cagnan H., 2011) also use the average power of the signal in different spectral bands.

Author (Subasi, 2010) uses the Average Power of the Wavelet coefficients. Kurtosis (15) and Skewness (16) amplitude distributions were introduced in present study.

2) Frequency domain features

Extracted frequency domain features are synthetized in Table 3. Frequency Median (37) and Frequency Mean (38) are used in (Oskoei, 2006); different relations of the bands of the Frequency Spectrum (39-58) are used in the articles (Zaidel, 2009; Cagnan H., 2011).

3) Action potentials domain features

Extracted action potential domain features are synthetized in Table 4. According to several authors (Pralong Etienne, 2002; Seifried, 2012), spike related features of the signal, like e.g. Neuronal Spiking Frequency and Burst Index, are sufficient, not only for the vertical localization of the STN, but also for targeting specific sub-regions of interest like the Dorsolateral Oscillatory Region (DLOR). As time or amplitude spike domain features we have: Spike Amplitude Distribution (72-81), Spike Maximum Amplitude Median (71), which is at the same time related to the proximity, density and synchrony of neuronal activity; and the Median Spike Amplitude Differential (60). As features measuring spike periodicity, or spike frequency domain features we have, from (Favre, 1999): Spike count (59), which gives an absolute value of neuronal activity; Inter-spike interval histogram (82-101); Inter-spike interval distribution entropy (61); Inter-spike interval mean (62); Inter-spike interval median (63); and Inter-spike interval standard deviation (64). From (Legendy, 1985) we have the Modified Burst Index (65), and from (Favre, 1999) the Pause Index (66) and the Pause Ratio (67).

Additionally, in this study, we added the following features related to individual spike shape: Spike depolarization time (68), Spike repolarization time (69), Spike refractory period (70) and the Spike mean shape itself (102-139).

D. Feature space reduction

The discriminatory power of the 140 different features was assessed through wrapping the classifiers over different subsets after z-score normalization, composed of sub-selections of the original features, or of increasing number of principal components. Tested subsets were chosen considering feature low collinearity and separability ranking for improved classification performance and generalization. Figure 6 shows the feature correlation matrix and the ranking according to the Class Separability Criterion of (Theodoridis, 1999).
E. Classification

Having implemented the feature extraction stage and inspected them thoroughly, we tested 4 different classifier models, namely: K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) and Logistic Regression, first in their binary versions, and after in the multinomial. For the binary classification human labels were binned, as the original dataset comprised of instances classified according to an ordinal scale with six consecutive scores from -1 to 4. We chose the value 2 as the dividing threshold, so all scores below it were set to zero, whether those equal or greater were set to 1. Non-classified signals were assumed to belong to class zero. The dataset used for training was composed of the 140 original features from a total of 2000 MER signals corresponding to all recorded positions of 9 different patients. Following a wrapper methodology, classifier estimation scores were used as a way to select the best features subset from the original group, and also to define any configuration parameters, which are detailed in the results and discussion section. In the case where original features were used, subsets were assembled by hand, making an intelligent guess combining both feature order in the CS ranking table and the values of the coefficients in the correlation matrix. In the case where Principal Component projection was used, since interdependence is assured, features were selected incrementally from most to less significant PC. For the performance tests, we used 10-fold cross-validation schemes, for which the data set was divided successively in train and validation subsets and accuracy averaged over all iterations. In the binary case, accuracy was calculated simply as a rate of correct and false estimations, together with the errors of type I and type II. In the multinomial case the accuracy was calculated as the mean error distance to the right value as a percentage of the maximum distance possible.

F. Software application

To reach the goal of this study, a new software application was developed from scratch. Chosen programming language was Matlab due to its built-in functions for machine learning pattern recognition; its computation performance, which allows the analysis of large data sets, fast enough for real-time applications; its reliability and tradition as a scientific mathematics and statistics tool; and finally because it also allows the construction of Graphical User Interfaces (GUI), essential for the application to be used by a final non-programmer user. The software allows loading and navigation through a database of MER files in binary format, the visual display of each signal in its original form and to listen their audio representation. It allows as well the application of a few standard signal processing transformations as a preview, namely: low, high and band-pass Fourier filtering; Fast Fourier Transform (FFT); Hilbert Transform; thresholding; absolute value; extraction of the autocorrelogram, amplitude histogram, Root Mean Square (RMS) value; detection of action potential; and others. It also provided an interface to save to file and load or build new signal-label pair datasets, by choosing subsets of signals from the database, segment patches to desired size, and editing labels if necessary. From these datasets user can extract up to 140 features and check their collinearity through bidimensional scatter plots and correlation matrix; or analyze their relevance for the targeting by displaying their values together with human labeling across the probing route, individually or in group, and in 1 or 2 dimensions. The software allows the loading of a trained classifier from file, or to train one of the 6 available models (binary KNN, SVM, LDA or Logistic; or multinomial KNN and Logistic) using a newly or previously assembled training signal-label dataset. Having trained a classifier, user can then apply it to newly unseen MER data and estimate their labels, from one signal segment to the entire database at a time, in seconds. Figure 7 show the GUI of the software application.

III. RESULTS AND DISCUSSION

1) Feature extraction, analysis and selection results

The definition of the different features and their analytical expression give us a first grasp on the inter-feature correlation. The correlation matrix give us a fast way of seeing the correlation between a vast number at same time. Furthermore, the direct inspection of the pairwise scatter plots can make evident the structure of these correlations. From the different independent variables possible combinations, we can distinguish between different groups of features according to their interdependence (see Figure 8): a first group, corresponding to the first two rows, with features with a clear and strong linear interdependence; a second group, at row three, showing a strong correlation but according to an unknown nonlinear mapping; and finally at the bottom row, we can see a group of features with very little or no apparent correlation, distinguished by a higher and more randomly mixed dispersion of samples.

In terms of discriminatory power, see we that many of the extracted features belonging to time domain have very strong correlation with human classification labels. A bigger subset of features, which includes numbers 1 to 8 (INT, MAV, MMAV2, MMAV2, MED, RMS, VAR, WL) and 11 to 14 (WAMP, SSI, PWRA, NM) showed positive correlation; and a smaller subset, composed by features 9, 10, 15 and 16 (ZC, SSC, AKUR, ASKW) showed negative correlation. Their plots are shown in Figure 9. The trajectory chosen by the surgery team for the final implantation (shaded) is normally the one having the strongest correlation of all, with the clearest transition at the superior border when entering the target area, followed by a high plateau and a smooth descent into the exit in inferior border. In the Figure 10 we show the remaining 20 time domain features corresponding to the bins of the amplitude histogram between values 0 and 100 µV in modulus. It is visible a strong correlation with class labeling below, which in this case correspond to the chosen ones for DBS implantation. We can see that the correlation between amplitude distribution and EDT is negative for the first two amplitude bins (0-5 and 5-10 µV), and becomes positive.
from the third bin (10-15 µV) onward. This is in line with the separation observed in previous analysis of the MER signals between background noise and neural activity. In general noise levels lay below 10 µV. As we go higher in the amplitude bins, the HIST feature becomes more uncorrelated with the human labeling along the route, signaling strongly intermediate depths inside the STN with high amplitude neural activity. From the figures we can see a strong negative correlation between both frequency mean (FMN) and median (FMD) features, with the first varying more smoothly than the second. This is explained by the fact that background noise is mostly located in higher frequencies (>2500Hz), while neuronal activity composed of the transient AP events, which gets stronger in the target area, has a lower frequency. If we superimpose the mean of the bottom half of the frequency spectrum with the top half, this frequency shift becomes very clear (bottom row). An interesting aspect of these graphs is that the start of the target area, as labeled by the specialist, seems to correspond exactly with the point where the two lines representing the two halves of the spectrum cross each other. Neuronal activity dynamics is then correlated with STN location both in time and frequency domains.

From the sequence of plots in the top row of Figure 12 we can see that the median of the Spikes Peak Amplitude (MAX) is increased in depths corresponding to the higher class scores, which corroborates the existence of a higher neural activity in the target region. Something interesting is that the Median Maximum Spike Amplitude Differential (MSA), which is a feature that measures the irregularity in firing patterns, closely follows the MAX feature. This could mean that as neuronal activity in STN increases, it also gets proportionally more erratic or, which is more plausible, that more signal sources are being captured by the microelectrode. Because we know that action potentials have standardized firing amplitudes, and that the voltage potential gradient recorded is greater when more active neuronal tissue is close to the probe, aligned and in phase, this dynamics should be explained by an increased density of active cell groups, aligned and in synchrony. The figures in the middles row with the Median Action Potential Shape (SHAPE) and also the bottom row figures showing their binned Amplitude Distribution (SAD) also confirmed this increased activity correlated with the labeling.

2) Classification results

Binomial Classifiers Results

For the binary K-Nearest Neighbors Classifier it was achieved a maximum correct rate of 85.5% with a feature subset comprising of numbers 19, 4 and 57, which are respectively: the 3rd Amplitude Histogram bin (HIST), corresponding to peaks with 10 to 15µV; the Modified Mean Absolute Value 2 (MMAV2); and the 19th Frequency Spectrum band, which covers the spectrum from 4500 to 4750 Hz (FTB.19). These results are shown in Figure 13 together with the accuracy and false positives evolution with increased number of Principal Components used.

Using a Support Vector Machine classifier model, it was achieved a maximum accuracy of 90.59% with the first 10 PCs, a Linear Kernel function and an L1 soft-margin (Figure 14). Accuracy rises very quickly from first to second PC, reaching a stable plateau shortly after at 5, and a local maximum of 87.82% at the 7th principal component. The linear kernel function is the one with better performance, which can mean that the present problem is, for its most part, linearly separable. It was noticed though that SVM classifier takes much longer to training the KNN, and even non-convergence occurs at times, when using the Sequential Minimal Optimization (SMO) algorithm (Fan, 2005), and especially for bigger number of PCs.

With the Linear Discriminant Analysis Classifier, in a 10-fold cross-validation population of 1000, we were able to achieve a mean maximum accuracy of 90.9%, at the expense of using all first 50 PCs, and with the Binary Logistic Classifier, 90.9% with the first 9 principal components only. For the last, with a 100-fold cross-validation, the mean accuracy drops only to 90.7%, with 3.3% false positives and 6% false negative. When first 70 PCs or more are used, Maximum Likelihood Estimation doesn’t converge under a 100 iterations limit. Because LDA and Logistic Classifiers output the probabilities of the different classes, they can be tuned, for instance, to reduce false negatives, which is a great advantage in real life clinical applications.

Multinomial Classifiers Results

Multinomial K-nearest Neighbors classifier was first tested with only 5 neighbors and the 3 most significant PCs of the original feature set. Average accuracy using a different number of folds in the cross-validation averaged at 90.5% (Figure 16). Standard deviation increased with the number of folds of the test, which is expected since the test subset correspondingly decreases. The model was also tested with increasing number of PCs and increasing number of neighbors, with the results illustrated in Figure 16. From these results, it is clear that performance plains when we approximate to the 20 most significant PCs, with an absolute maximum average accuracy of 92.40±1.48% reached at 27 PCs. Increasing the number of neighbors causes a very strong increase in performance until 5 where it reaches a local maximum of 91.46% and then stabilizes to start decreasing again only at 23.

Lastly, class estimation with the Multinomial Logistic Classifier it was achieved an average accuracy of 92.8% using the first 25 principal components of the features as predictors (Figure 19). From the Figure 19 we can see that, at least when using only the 1st PC as predictor, classes 4 and 0 dichotomize very well but not the other. This indicate that it might be hard for the classifier to discriminate the remaining intermediate classes. It means that, for the dataset and predictors used, it is probably unnecessary to use multinomial analysis, and that a simpler two class nominal classification can be enough or more adequate.
IV. CONCLUSIONS

With the application developed in the present study we are able to estimate virtually instantly the clinical evaluation of the quality of each MER acquired during DBS-STN surgery, as corresponding to the optimum position in the target area for the implantation of a stimulation electrode. Offline performance reached an average maximum accuracy of 92.8% of the ground truth (human labeling done intraoperatively by a specialist neurologist) with a Multinomial Logistic Classifier model using the first 25 principal components of the 140 extracted features, and trained with 1000 signal from only 4 patients (Figure 1).

Table 1 – Classification performance summary.

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Data set</th>
<th>Feature Set</th>
<th>Avg. Max. Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Nearest Neighbors</td>
<td>4 patients (1000 signals)</td>
<td>25 PCs</td>
<td>88.53±3.66</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>9 patients (2020 signals)</td>
<td>10 PCs</td>
<td>90.59±4.98</td>
</tr>
<tr>
<td>Linear Discriminant</td>
<td>4 patients (1000 signals)</td>
<td>50 PCs</td>
<td>90.9</td>
</tr>
<tr>
<td>Logistic</td>
<td>4 patients (1000 signals)</td>
<td>9 PCs</td>
<td>90.92±2.22</td>
</tr>
<tr>
<td>K-Nearest Neighbors</td>
<td>9 patients (2020 signals)</td>
<td>27 PCs</td>
<td>92.41±9.92</td>
</tr>
<tr>
<td>Logistic</td>
<td>4 patients (1000 signals)</td>
<td>25 PCs</td>
<td>92.8</td>
</tr>
</tbody>
</table>

Figure 1 – Classification stage performance summary.

Matlab application

As a product of this study it was developed a fully functional Matlab application with a user friendly graphical interface, which in fulfillment of the original objectives allows a user to: a) Load, display, and listen to MER-signals acquired during a DBS surgery; b) Perform offline several signal preprocessing techniques to enhance its SNR; c) Extract up to 140 different features of the signals specific to the identification of STN brain region neural activity, and perform selection and normalization; c) Manually edit the classification of existing data and build new data sets; d) Train different classifier algorithms and estimate the matching score of MER-signals at each depth to the target region for chronic implantation of a DBS-STN stimulation electrode based on provided information from presently resident neurologist intraoperative evaluation of previous surgeries.

Signal pre-processing

Preprocessing of the signals positively contribute to subsequent stages of analysis and are of special relevance for efficient detection of action potentials which is a fundamental step to discriminate different neural activity patterns and locate the neuroanatomical region of interest in the present study - the STN. We show that non-causal Fourier based filters can increase SNR and preserve at same time most of the AP dynamics, while causal versions don’t, for which they should be avoided. We also show that wavelet transforms are superior both in preserving AP dynamics as well as identifying segments of the signal with different physical characteristics when used together with machine learning, allowing a very fine tailoring of the signals for subsequent stages of analysis, the efficient detection of APs and removal of odd artifacts.

Feature extraction

Following an exhaustive literature research on best possible signal features to discriminate neuronal activity of the STN from its neighboring regions, its borders and sub-regions, we reached a final set with a total of 140. From these, time domain feature showed a very high collinearity and could be divided according to the correlation to the outcome classes in two subsets: one consisting of features 1 to 8 (INT, MAV, MMAV2, MMAV2, MED, RMS, VAR, WL) and from 11 to 14 (WAMP, SSI, PWRA, NM), which have positive correlation with outcomes, and another corresponding to features 9, 10, 15 and 16 (ZC, SSC, AKUR, ASKW), with have negative correlation. Frequency domain features also showed a strong correlation with the dependent variable and demonstrate that MER signals average frequency spectrum power consistently shifts from higher to lower frequency bands when STN anterior border is crossed. Features in the Action Potential domain further contribute to the discriminative power of the system by giving a direct measure of the neural activity. We could clearly identify an increase in spike count and average amplitude when the region with the best classification scores is reached, as well as a relatively consistent pattern in intra-spike features (DEP, REP and REF), which also showed a strong change when crossing the anterior border.

Feature selection

Selection of features which better correlate with the outcomes and at same time show less collinearity strongly contribute to improvement of the final performance of the trained classifier. Classifiers trained with the Principal Components have demonstrated better average maximum accuracies than those trained directly with most combinations of subsets taken from the 140 originally extracted features, even when Class Separability and Correlation Coefficients between features is considered. As shown, accuracies of 90.9% can be achieved with few as 9 principal components only. Between original extracted features, the three best ranking for CS are features 19, 4 and 57, which are respectively the 3rd Amplitude Histogram bin (HIST.3) from which covers peaks with 10 to 15μV, the Modified Mean Absolute Value (MMAV2) and the 19th Frequency Spectrum band which cover the spectrum from 4500 to 4750 Hz (FTB.19).

Class estimation performance

Average maximum accuracy values of more than 90% were achieved with every one of the classifiers tested using a training dataset composed of only 4 to 8 patients of a total of 118 records available, corresponding to 3.4-6.8% of the entire database. This results leave us optimistic about presently developed system, and to suppose that, once more records are added to the training phase, it will reach performance levels adequate for live trials and hopefully the use during Parkinson’s Disease DBS-STN surgeries.

Acknowledgements

First I would like to thank my brother who encouraged me to take this masters course while working. Thank
professor Raúl Carneiro Martins who, aware of my interest for the brain sciences invited me in first place to develop this thesis. Thank professor José Manuel Bioucas for the guidance, recognize doctor Manuel Herculano de Carvalho for the seminal vision, and all DBS-Parkinson surgery team of Santa Maria Hospital for their amazing job.

V. BIBLIOGRAPHY


Levine, C. B. (2003). Diagnosis and Treatment of Parkinson’s Disease: A Systematic Review of the Literature. AHRQ.


Pralong, E. (2004). Quality index for the quantification of the information recorded along standard microelectrode tracks to the subthalamic nucleus in parkinsonian patients.


VI. SUPPLEMENTARY MATERIAL

Figure 2 – Flow chart of the implemented machine learning system for the analysis of the DBS MER signals, from acquisition until class estimation.

Figure 3 – Relative position of the 5 parallel routes used for intraoperative microelectrode recording.

Figure 4 – Position-related qualitative evaluation record of intraoperative MER signals done by the electrophysiologist. Source: Santa Maria Hospital, Lisbon-Portugal.
Figure 5 – Left: Exemplification of signal segmentation (top) through manual clustering on principal components space (bottom) of the four first wavelet decomposition scales (middle). Right: Time feature vector (below) for of a MER signal (above) with a segmentation of 10 samples.

Figure 6 – Ranking of the 140 features extracted from MER signals according to the Class Separability Criterion of (Theodoridis, 1999) – Matlab’s rankfeatures – (left), and Correlation Coefficient Matrix (right) corresponding to the best 30.
Figure 7 – Graphical User Interface (GUI).
Figure 8 – Scatter plots for paired time domain features with color coded classes. First and second rows, with a strong linear correlation, features number: 12-13, 1-2, 1-4, 1-6, 1-8, 1-7, 1-14, 1-11, 1-12, 1-13; In 3rd row, with a strong non-linear correlation, features number: 15-18, 5-18, 15-19, 15-20, 15-2; In the 4th row, features with very low correlation.
Figure 9 – Plots of time domain features directly correlated with classification for all electrodes in hemisphere 1 (1st row) and 2 (2nd row) for patient 1, and hemisphere 1 (3rd row) and 2 (4th row) for patient 5. Chosen electrode for chronic implantation is shaded.

Figure 10 – Top: color coded (red means higher) MER signal amplitude distribution depending on the Estimated Distance to Target (EDT) of features 17 to 36 for chosen routes in two different patients. Bottom: together with respective human labeling.
Figure 11 – Top row: frequency mean and median features; middle row: frequency spectrum; bottom row: average of the bottom and top half frequency spectrum bands for chosen routes of 4 different patients.

Figure 12 – Spike amplitude features for 4 different patients. Top row: Median Spike Peak Amplitude (MAX) and Median Maximum Spike Amplitude Differential (MSA). Middle row: Spike Shape Median (SHAPE). Bottom row: Spikes Amplitude Distribution (SAD).
**Figure 13** – Binary KNN classification results.

<table>
<thead>
<tr>
<th>Feature set</th>
<th>Dim</th>
<th>Correct</th>
<th>False Pos.</th>
<th>False Neg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>[19 20]</td>
<td>2</td>
<td>83.2 ± 8.11</td>
<td>8.3</td>
<td>8.5</td>
</tr>
<tr>
<td>[19 20 14]</td>
<td>3</td>
<td>85.1 ± 7.97</td>
<td>7.6</td>
<td>7.3</td>
</tr>
<tr>
<td>[19 4 57]</td>
<td>3</td>
<td>85.5 ± 3.66</td>
<td>8.5</td>
<td>5</td>
</tr>
<tr>
<td>[19 21 12]</td>
<td>3</td>
<td>84.5 ± 6.48</td>
<td>7.6</td>
<td>7.9</td>
</tr>
<tr>
<td>[83 54 115]</td>
<td>3</td>
<td>78.8 ± 6.04</td>
<td>10</td>
<td>11.2</td>
</tr>
<tr>
<td>[19 21 14]</td>
<td>4</td>
<td>83.3 ± 6.34</td>
<td>8.9</td>
<td>7.8</td>
</tr>
<tr>
<td>[19 21 57]</td>
<td>4</td>
<td>84.1 ± 6.87</td>
<td>8.6</td>
<td>7.3</td>
</tr>
<tr>
<td>[19 20 14 114 57]</td>
<td>6</td>
<td>84.7 ± 5.54</td>
<td>8</td>
<td>7.3</td>
</tr>
</tbody>
</table>

**Figure 14** – Binary Support Vector Machine classification results.

<table>
<thead>
<tr>
<th>Feature subset</th>
<th>Acc Rate (%)</th>
<th>Comp. Time (s)</th>
<th>Confusion Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original [19 4]</td>
<td>79.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC [1 2]</td>
<td>84.9 ± 4.24</td>
<td>3.548</td>
<td></td>
</tr>
<tr>
<td>PC [1 3]</td>
<td>78.5 ± 5.92</td>
<td>6.244</td>
<td></td>
</tr>
<tr>
<td>PC [1:5]</td>
<td>89.1 ± 4.83</td>
<td>8.172</td>
<td></td>
</tr>
<tr>
<td>PC [1:10]</td>
<td>90.6 ± 5.98</td>
<td>15.679</td>
<td></td>
</tr>
<tr>
<td>PC [1:20]</td>
<td>85.9 ± 5.03</td>
<td>37.002</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 15** – Binary Linear Discriminant classification results using 10-fold cross-validation of a population of 1000.
Figure – Results for a Nominal Logistic Classifier trained with 900 samples and tested in a 100 samples set.

Figure 16 – Multinomial KNN classifier mean accuracy tested with 10-fold of 1000 sample train set. Left: Maximum of 92.40±1.48% with first 27 principal components according to 10-fold cross-validation test. Right: Evolution of the accuracy with the increase in the number of neighbors in the KNN algorithm. Maximum accuracy obtained was approximately 92.6% with 23 neighbors.
Figure 17 – Application example of the Multinomial KNN estimation. First row: chosen route from patient 1; following rows: all 5 routes for patients 1, 2, 5 and 6.

Figure 18 – Multinomial KNN classifier results: example of a drastic feature mismatch in one patient chosen route.
Figure 19 – Test results for a Multinomial Logistic Classifier trained with 900 samples and tested in the 100 left from a set of 1000, using only the 1st principal component of the features as predictor.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MER signal band</td>
<td>500 to 3000 Hz (HFB)</td>
<td>500 to 8000 Hz</td>
<td>Not stated</td>
<td>Not stated</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Signal enhancing</td>
<td>X</td>
<td>Moving Average</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Filters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Threshold-based methods (denoising and smoothing)</td>
<td>X (40μV)</td>
<td>Adaptive</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normalization Artifacts Remotion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Feature extraction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time/amplitude analysis</td>
<td>RMS</td>
<td>X</td>
<td>Spike firing Rate</td>
<td>Inter spike interval</td>
<td>Spike firing Rate</td>
<td></td>
</tr>
<tr>
<td>Frequency analysis</td>
<td>PSD (Welch’s method)</td>
<td>PSD</td>
<td>X</td>
<td>PSD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Neural noise background | X | Adaptive Wavelet | X | Wavelet transform | Feature identification (classifiers) | Not stated | Bayesian | Gaussian; Kernel | Naive; LDC; QDC | K-NN | Euclidean; DTW | X | Classification Tree | Gini; Max Deviance | HMM | Viterbi algorithm | General/best method accuracy | 89.60% | 88.00% | Not stated | Online implementation | No | No | Yes | Yes | Region discriminated | Not stated | Thalamus - TAL | X | X | Zona Incerta – ZI | X | X | X | Subthalamic Nucleus – STN | Dorsal and Ventral borders | X | X | X | STN dorsolateral oscillatory region (DLOR) | X | Substantia Nigra Reticulata – SNR | X | X | X

Table 1 – Comparison between published DBS targeting automation studies.

### TIME DOMAIN FEATURES

<table>
<thead>
<tr>
<th>NR.</th>
<th>ABBR.</th>
<th>DESCRIPTION</th>
<th>EQUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INT</td>
<td>Integrated Absolute Value.</td>
<td>$MAV = \frac{1}{N} \sum_{i=1}^{N}</td>
</tr>
<tr>
<td>2</td>
<td>MAV</td>
<td>Mean Absolute Value. N is the total number of samples in the signal and $x_i$ is the i-th sample.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>MMAV1</td>
<td>Modified Mean Absolute Value 1, which correspond to a windowed MAV adds a weighing window to the MAV, where the start and end 25% is weighted only half.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>MMAV2</td>
<td>Modified Mean Absolute Value 2, which adds a different weighting window $w_i$ to MAV.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>MED</td>
<td>Median Absolute Value.</td>
<td>$MED = \frac{1}{N} \sum_{i=1}^{N} x_i$</td>
</tr>
<tr>
<td>6</td>
<td>RMS</td>
<td>Root Mean Square.</td>
<td>$RMS = \sqrt{\frac{1}{N} \sum_{j=1}^{N} x_j^2}$</td>
</tr>
</tbody>
</table>
### Analysis and Classification of Microelectrode Recordings in Deep Brain Stimulation Surgery

7  VAR Variance. \[
VAR = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2
\]

8  WL Waveform or Curve Length. \[
WL = \sum_{i=1}^{N-1} |x_{i+1} - x_i|
\]

9  ZC Zero Crossings. A threshold \( \varepsilon \) is introduced to reduce the noise associated with segments of the signal sustained close to zero. \[\{x_i > 0 \text{ and } x_{i+1} < 0\} \text{ or } \{x_i < 0 \text{ and } x_{i+1} > 0\} \text{ and } |x_i - x_{i+1}| \geq \varepsilon\]

10 SSC Slope Sign Changes. Again a threshold \( \varepsilon \) is necessary to reduce noise. \[\{x_i > x_{i-1} \text{ and } x_i < x_{i+1}\} \text{ or } \{x_i < x_{i-1} \text{ and } x_i > x_{i+1}\} \text{ and } |x_i - x_{i+1}| \geq \varepsilon \text{ or } |x_i - x_{i-1}| \geq \varepsilon\]

11 WAMP Willson Amplitude, which correspond to a windowed WL, counts the number of times the signal increases by a predetermined threshold value \( \varepsilon \). \[
WAMP = \sum_{i=1}^{N} f(x_i - x_{i+1})
\]

\[f(x) = \begin{cases} 1 & , x > \varepsilon \\ 0 & , otherwise \end{cases}\]

12 SSI Simple Square Integral. \[
SSI = \sum_{i=1}^{N} |x_i^2|
\]

13 PWRA Average Power. \[
PWRA = \frac{1}{N} \sum_{i=1}^{N} y_i^2
\]

14 NM Noise mode.

15 AKUR Amplitude Distribution Kurtosis. \[
Kurtosis = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^4}{(N-1)s^4}
\]

16 ASKW Amplitude Distribution Skewness. \[
Skewness = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^3}{(N-1)s^3}
\]

17-36 PDF Amplitude probability distribution or Histogram. \[
HIST \equiv \{bin_1(|y|), bin_2(|y|),..., bin_k(|y|)\}
\]

\[bin_k(|y|) \equiv \# \ y_i: b_j < y_i \leq b_{j+1} \] and \( b = \{0.5,...100\} \)

<table>
<thead>
<tr>
<th>NR.</th>
<th>ABBR.</th>
<th>DESCRIPTION</th>
<th>EQUATION</th>
</tr>
</thead>
</table>
| 37  | FMD   | Frequency Median. fMDI is the median frequency when the discrete frequency scale index or bin MDI divides the power spectrum sum in two. | \[
FMD = \left\{MDI_{\text{MdI}}: \sum_{i=1}^{MDI} PSD = \frac{1}{2} \sum_{i=1}^{N} PSD \right\}
\]

| 38  | FMN   | Frequency Mean. | \[
FMN = \frac{1}{N} \sum_{i=1}^{N} f_i PSD_i
\]

Table 2 – Synthesis of extracted time domain features.
### Analysis and Classification of Microelectrode Recordings in Deep Brain Stimulation Surgery

#### Table 3 – Synthesis of extracted frequency domain features.

$$PSD_x(f_x) = \frac{1}{N f_x} \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi}{f_x} n}$$

#### Table 4 – Synthesis of action potential domain features.

<table>
<thead>
<tr>
<th>NR.</th>
<th>ABBR.</th>
<th>DESCRIPTION</th>
<th>EQUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>59</td>
<td>CNT</td>
<td>Spike Count.</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>MSA</td>
<td>Median spike amplitude differential.</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>ISI_E</td>
<td>Inter-spike Interval Distribution Entropy.</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>ISI_MN</td>
<td>Inter-spike Interval Mean.</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>ISI_MD</td>
<td>Inter-spike Interval Median.</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>ISI_STD</td>
<td>Inter-spike Interval Standard Deviation.</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>MBI</td>
<td>Modified burst index, defined as the number of inter-spike intervals (ISI) greater than 10ms divided by the number of ISI lesser than 10ms.</td>
<td>$$MBI = \frac{#(ISI &lt; 10ms)}{#(ISI &gt; 10ms)}$$</td>
</tr>
<tr>
<td>66</td>
<td>PI</td>
<td>Pause index, corresponding to the number of ISI greater than 50ms divided by the number of ISI lesser than 50ms.</td>
<td>$$PI = \frac{#(ISI &gt; 50ms)}{#(ISI &lt; 50ms)}$$</td>
</tr>
<tr>
<td>67</td>
<td>PR</td>
<td>Pause ratio, corresponding to the cumulative ISI greater than 50ms divided by the cumulative ISI lesser than 50ms.</td>
<td>$$PR = \frac{\sum (ISI &gt; 50ms)}{\sum (ISI &lt; 50ms)}$$</td>
</tr>
<tr>
<td>68</td>
<td>DEP</td>
<td>Median Depolarization Time.</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>REP</td>
<td>Median Repolarization Time.</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>REF</td>
<td>Median Refractory Time.</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>SPK_MAX</td>
<td>Spike Shape Maximum.</td>
<td></td>
</tr>
<tr>
<td>72-81</td>
<td>SAD</td>
<td>Spike amplitude distribution. 10 bins.</td>
<td></td>
</tr>
<tr>
<td>82-101</td>
<td>ISI_H</td>
<td>Interspike time distribution histogram. 20 bins.</td>
<td></td>
</tr>
<tr>
<td>102-139</td>
<td>SHAPE</td>
<td>Median Action Potential Shape. The segments corresponding to detected spikes of 1.4ms with 3600 samples were downsampled to 20 bins.</td>
<td></td>
</tr>
</tbody>
</table>