Statistical Study of Parameters for Deep Brain Stimulation Automatic Preoperative Planning of Electrodes Trajectories

Caroline Essert · Sara Fernandez-Vidal · Antonio Capobianco · Claire Haegelen · Carine Karachi · Eric Bardinet · Maud Marchal · Pierre Jannin

Abstract *Purpose* Automatic methods for preoperative trajectory planning of electrodes in Deep Brain Stimulation are usually based on the search for a path that resolves a set of surgical constraints to propose an optimal trajectory. The relative importance of each surgical constraint is usually defined as weighting parameters that are empirically set beforehand. The objective of this paper is to analyze the use of these parameters thanks to a retrospective study of trajectories manually planned by neurosurgeons. For that purpose we firstly retrieved weighting factors allowing to match neurosurgeons manually planned choice of trajectory on each retrospective case, secondly we compared the results from two different hospitals to evaluate their similarity, and thirdly we compared the trends to the weighting factors empirically set in most current approaches.

Methods To retrieve the weighting factors best matching the neurosurgeons manual plannings, we proposed two approaches: one based on a stochastic sampling of the parameters and the other on an exhaustive search. In each case, we obtained a sample of combinations of weighting parameters with a measure of their quality,

C. Essert, A. Capobianco
ICube, Université de Strasbourg / CNRS (UMR 7357)
300 bd Sébastien Brant - BP 10413 - 67412 Illkirch, France
E-mail: essert@unistra.fr

S. Fernandez-Vidal, E. Bardinet
ICM CENIR, CNRS (UMR 7225) / INSERM (U975) / UPMC / Hôpital Pitié-Salpêtrière
47 boulevard de l'Hôpital - 75013 Paris, France

C. Haegelen, and P. Jannin
LTSI, INSERM (UMR 1099) / Université Rennes 1
CS34317 - 35043 Rennes, France

C. Haegelen
Department of Neurosurgery
Rennes University Hospital, 35000 Rennes, France

C. Karachi
Sorbonne Universités, UPMC Université Paris 6, INSERM, CNRS, INRIA, ICM
AP-HP, Hôpital Pitié-Salpêtrière, Department of Neurosurgery
47 boulevard de l'Hôpital - 75013 Paris, France

M. Marchal
INRIA Rennes Bretagne-Atlantique/ IRISA / INSA Rennes, France
Campus de Beaulieu, 35042 Rennes Cedex, France

i.e. the similarity between the automatic trajectory they lead to and the one manually planned by the surgeon as a reference. Visual and statistical analysis were performed on the number of occurrences and on the rank means.

Results We performed our study on 56 retrospective cases from two different hospitals. We could observe a trend of the occurrence of each weight on the number of occurrences. We also proved that each weight had a significant influence on the ranking. Additionally, we observed no influence of the medical center parameters, suggesting that the trends were comparable in both hospitals. Finally, the obtained trends were confronted to the usual weights chosen by the community, showing some common points but also some discrepancies.

Conclusions The results tend to show a predominance of the choice of a trajectory close to a standard direction. Secondly, the avoidance of the vessels or sulci seems to be sought in the surroundings of the standard position. The avoidance of the ventricles seem to be less predominant, but this could be due to the already reasonable distance between the standard direction and the ventricles. The similarity of results between two medical centers tend to show that it is not an exceptional practice. These results suggest that manual planning software may introduce a bias in the planning by proposing a standard position.

Keywords Surgical Planning · Trajectory Optimization · Statistical Analysis · Deep Brain Stimulation · Neurosurgery

1 Introduction

Deep Brain Stimulation (DBS) is a reversible surgical treatment of neurological disorders such as Parkinson's disease or essential tremors, mostly proposed to patients with severe symptoms who do not respond well to medication. The intervention consists in implanting one or several electrodes into deep locations of the brain to impulse an electric stimulation causing an inhibition of the motor disorders. This treatment is very efficient but the planning is still challenging and mainly relies on the study of preoperative MRI and CT [9].

The objectives of the preoperative planning are first to accurately locate the anatomical target, and then to find a secure path for an electrode towards the selected target. Even if some commercial workstations already propose an interactive assistance to these tasks, they remain manual and tedious, and can sometimes take up to one hour. That is why many neurosurgeons expressed a need for a computer-aided automatic assistance for the whole process, and several research groups worked on this topic. The literature also underlines the importance of an accurate planning to avoid risks of side effects [15] or hemorrhage [2].

This paper focuses more particularly on the second part of the preoperative planning: the optimization of the trajectories. Most of the current approaches are based on solving surgical rules over the geometry of the anatomy: constraints to be optimized or risks to be minimized, translated into numerical cost functions. However, all surgical rules do not have the same importance to the surgeon, implying the choice of parameters values for weighting them within the search for an optimum. When proposing solutions, the algorithms usually require the combination of several cost functions into a single function with weighting factors. Most often, these factors are chosen empirically, using the expertise of the neurosurgeons.

The objective of this work is to study the use of weighting parameters by comparing trajectories automatically computed with trajectories manually planned by surgeons. To this end, this study first strives to find the weights that lead to automatic trajectories best matching the trajectories manually planned by the neurosurgeons. Two approaches are proposed to guess those weights: the first one is based on a stochastic strategy, and the second one is an exhaustive search. Both use a generic and objective *a posteriori* approach on patients images. We report the tests on 56 patients cases from two different medical centers, using an automatic planning method that we previously validated and published [7]. Finally we analyze and discuss the trends of the values and the similarities between centers.

2 Related Works

Recently, several authors reported methods for automatic computation of linear trajectories for DBS electrodes [7,3,6,13,14]. All approaches are based on constraints to be optimized or risks to be minimized. They are usually defined based on rules expressed by surgeons and translated into numerical data based on patient specific multimodal medical images. Then different approaches are defined for computing an optimal trajectory from such numerical constraints that express the strategies used by the surgeon for selecting the trajectory. Automatic methods usually require the combination of the constraints into a single cost function. The surgical rules do not have the same importance, implying the choice of different values for weighting the constraints within the search for an optimum.

So far, the weights have been chosen empirically by the developers of the methods or at best a priori in cooperation with the neurosurgeons. In order to follow a more objective approach and arguing that each surgeon has his/her own preferences, Liu et al. proposed in [10] an approach for adapting the weighting factors to a single surgeon. In [11], by studying multi-site and multi-surgeon combinations, the same authors showed that single optimal sets of factors might satisfy surgeons even more that their own selection. Both papers outlined the complexity in tuning the weighting factors and the need for further studies. In both papers, the computation of the optimal weights values was done manually by subjective analysis of the computed trajectories. The authors concluded that methods are needed to learn surgeon's preferences from training sets of retrospective clinical cases.

We present a learning based approach for automatic computation of weighting factors from retrospective analysis. Beyond validation issues and definition of imaging based constraints, the selection of the weighting is probably the most difficult remaining point in the automatic computation of straight trajectories.

3 Materials and Methods

This section first details the main principle on which this work is based. Then, the two approaches proposed to gather the combinations of weights closest to the manual plannings are exposed: stochastic and exhaustive. Finally, we explain how we set up our experiment, and the statistical analysis approach we used.

3.1 Automatic Determination of Weights on Retrospective Manual Plannings

The solving processes of automatic trajectory planning approaches usually solve a linear combination f of n cost functions f_i with weighting factors w_i assigned to each individual cost function:

$$f = \frac{\sum_{i=1}^{n} w_i.f_i}{\sum_{i=1}^{n} w_i} \tag{1}$$

When launched with fixed values for the weights, a set of candidate trajectories that satisfy the constraints can be obtained. It can be graphically displayed as a colormap, with colors representing values of f, as illustrated in Fig. 1(a) where only feasible entry points are colored and displayed. One trajectory T_{auto} providing the minimal result $r_{auto} = f(w_1, ..., w_n, T_{auto})$ is indicated as the optimal trajectory.

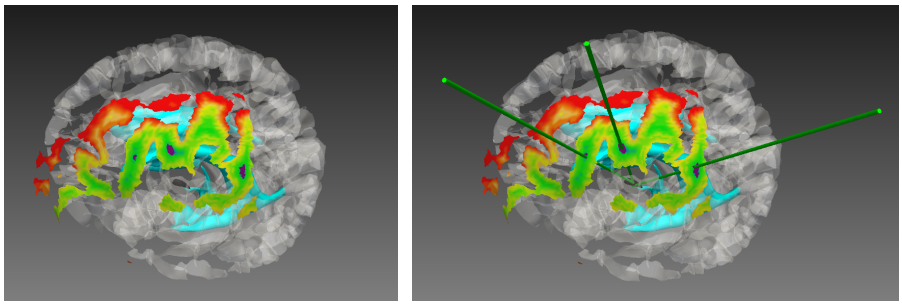
Most of automatic trajectory planning methods try to mimic the surgeons strategies. Therefore, the main hypothesis of known approaches is that, for every trajectory manually chosen by a surgeon (let's call it *reference trajectory* or T_{ref}), a set of weights that allows to find automatically this particular trajectory exists. If we want to find it, we should try all possible combinations of weights, and launch for each one an optimization process. For at least one of the experimented combinations, the optimal solution should fit T_{ref}, or at least be very close. The similarity criterion we have chosen to evaluate the proximity of an automatically computed trajectory T_{auto} to T_{ref} is the angle α between the two trajectories. The target point is fixed to the one set by the neurosurgeon *i.e.* the tip of T_{ref}.

Experimenting all possible combinations with a high precision would be very time-consuming. If we consider 3 weights between 0 and 1 to combine, and a precision of 0.01 per weight, an optimization process of about 0.1 second per combination will lead to $100^3 * 0.1$ seconds to compute, which represents nearly 28 hours for one patient case. To reduce the computation time, we could either decrease the precision or use a stochastic approach allowing an efficient browsing of the entire space of combinations while keeping a very high precision in a reasonable computation time. Section 3.3 describes the first possibility, an exhaustive search in limited precision. In the following, we describe the stochastic approach.

3.2 Stochastic Approach

This section recalls the main principles and extends the stochastic approach we previously published in [8]. The proposed algorithm consists in first defining a maximum number of combinations $iter_{max}$ that will be tested. We set $iter_{max}$ at $20,000$, leading to a computation time of approximately 30 minutes. Then, we start looping over combinations of weights. At each step j, we assign new values w_{i_j} for the n weights, randomly chosen between 0 and 1 with a precision of 3.10^{-5}. The weights are normalized so that their sum is equal to 1 using the following formula: $w'_{i_j} = \frac{w_{i_j}}{\sum_{h=1}^{n} w_{h_j}}$. Equation 1 can then be simplified to $f = \sum_{i=1}^{n} w'_{i_j}.f_i$. The optimization process is launched for step j with weights w'_{i_j}.

Our software that we used in our experiments, described in [7], finds for one patient case and one combination of weights, one optimal trajectory corresponding to the global minimum of f (see also [1]). Furthermore, it is also able to propose



(a) 3 connected components (b) 3 locally optimal trajectories

Fig. 1 Snapshots showing a typical 3D scene with the result of a resolution process, displayed as a color map where the best entry points are in green and the worst in red. In (a) three connected components, representing valleys of locally minimal values of f are displayed in purple. Corresponding locally optimal trajectories are displayed as green cylinders in (b). In both images, the target is in black, the ventricles in blue, and the sulci are white and semi-transparent for readability purposes. The MRI is not shown on these pictures. Complete 3D scene for this patient can be seen in Fig. 2

m interesting alternatives corresponding to m local mimima of f. Indeed, it can occur that the zone of possible entry points (*insertion zone*) is composed of several interesting areas located in different parts, and each area contains a locally optimal candidate trajectory, *i.e.* a trajectory for which $r > r_{auto}$ but $r < r_{auto}+\varepsilon$ (for more details, refer to [7,8]). This case is illustrated in Fig. 1, where 1(a) shows in purple three interesting connected components at different locations of the color map, and 1(b) shows their respective locally optimal trajectory. The number of alternative propositions shown to the surgeon can be chosen by adjusting parameter ε. To take this observation into account, we extended our search to the x "most optimal" trajectories, *i.e.* the global optimum and the $m = x - 1$ best alternatives. In this study, we limited x to 3, as it was observed in the preliminary study [8] that with a lower number we could too often miss interesting trajectories, and with a higher number we did not find better results.

At each step j, we compute angle α_{jk} between the reference trajectory T_{ref} and each of the x locally optimal $T_{auto_{jk}}$. Then we compare α_{jk} with the smallest angle α_{min} we found so far: if α_{jk} is smaller, $T_{auto_{jk}}$ is saved as the nearest to the objective, *i.e.* weights w'_{i_j} are saved and angle α_{jk} becomes the new α_{min}. The loop stops when j reaches $iter_{max}$. The results are ranked according to angles α and the 200 best combinations of weights w'_{i_j} leading to the smallest angles are kept. Selecting a higher number of selected combinations would not be reasonable, because it would be too difficult to analyze the results, more particularly for pair-wise comparisons. The main bias of the stochastic approach is that the selection of the 200 best combinations might not be representative of the solution space.

3.3 Exhaustive Approach

The main objective of the exhaustive approach is to compare for all patients the same combinations homogeneously distributed in the solution space. As stated above, for an exhaustive approach to be computed in a reasonable time, the pre-

cision needs to be reduced. We have chosen to use a precision of 0.1 per weight, which means 11 possible values from 0.0 to 1.0. In order to be even more efficient, the normalization step has been avoided by choosing directly weights having a sum equal to 1.0: a weight w_h is chosen such that $w_h < 1 - \sum_{i=1}^{h-1} w_i$. When all weights $w_1, ..., w_{n-1}$ are chosen, the last weight w_n can be deduced by $w_n = 1 - \sum_{1}^{n-1} w_i$.

In this study, we have used $n = 3$ soft constraints (see Section 3.4). When w_1 is fixed, then w_2 needs to be smaller than $1 - w_1$, and w_3 is deduced by $w_3 = 1 - (w_1 + w_2)$. If w_1 is 0.0, then there are 11 possibilities for w_2 from 0.0 to 1.0, and then w_3 is deduced accordingly. If w_1 is 0.1 there are only 10 possible values left for w_2 from 0.0 to 0.9, and so on. When w_1 is 1.0, then 0.0 is the only possibility for w_2. The total number of possible combinations of w_1, w_2, w_3 is equal to $\sum_{1}^{11} i = \frac{11*12}{2} = 66$. With an optimization process of about 0.1 second per combination, it takes less than 7 seconds for one patient case. A precision of 0.01 would lead to a number of combinations of $\sum_{1}^{101} i = 5151$, and a precision of 0.05 would lead to a number of combinations of $\sum_{1}^{51} i = 1326$ which are too large numbers for subsequent analysis.

In the same way as the stochastic approach, an optimization process is launched for each combination of weights, where x "most optimal" trajectories are computed. The trajectories closest to T_{ref} in terms of angle are saved. Finally, the 66 combinations are sorted according to the angle criterion.

3.4 Data and Pre-processing

3.4.1 Image Processing and 3D Scene

We performed our retrospective study on 28 anonymized patients datasets, 13 from Hospital A, and 15 from Hospital B. All of these patients had undergone bilateral DBS, making a total of 56 implantation cases in our study. All were treated for Parkinson's disease, and the target was the Subthalamic Nucleus (STN). By the time of their treatment, the surgeons had not used our planning software. In the following, we explain the image processing algorithms and pipelines we used in order to perform our study.

We used preoperative images: 3T T1-weighted MRI (1.0mm x 1.0mm x 1.0mm, Philips Medical Systems) for Pontchaillou Hospital subjects, and 1.5T T1-weighted MRI (0.9373mm x 0.9375mm x 1.3mm, GE Medical systems) for Pitié-Salpêtrière subjects, acquired just before the intervention. For the solving process, we segmented 3D meshes of the ventricles and the sulci. Ventricles and vessels are the main critical structures to avoid. However, vessels could not be segmented with a sufficient quality based on MRI only, if angiography is not available as it is the case in our datasets. We used the 3D mesh of the cortical sulci instead, as surgeons often do in the current clinical routine, since most of the vessels are located inside the sulci. Finally, a 3D mesh of a selected portion of the scalp was segmented as an initial search area for the entry point of the electrode.

The segmentation of the scalp and cortical sulci and the generation of the associated triangle meshes were done automatically using the pyDBS toolbox [5]. Volumetric segmentation of the ventricular system was performed indirectly. The Parkinson model in pyDBS suite includes a representation of the two lateral ventricles and the third and fourth ventricles. This anatomical model is registered on

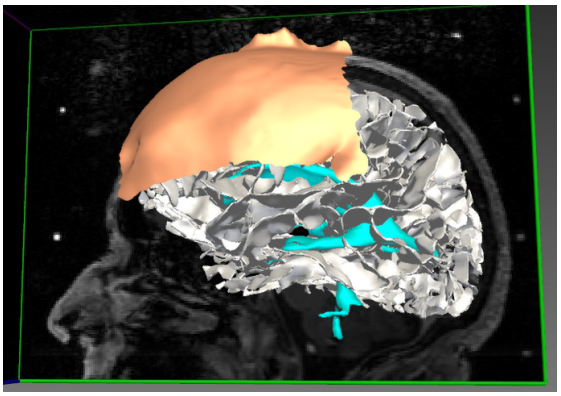


Fig. 2 Snapshot showing the data as a complete 3D scene, for the same patient as in Fig. 1. In the background, the preoperative MRI. In white, the cortical sulci. The ventricles are in blue and the target is in black. On the top of the head, the initial zone is in orange.

each antatomical subject space (the deformation strategy is described in [5]) to get the meshes of the ventricular structures in the subject space.

From the triangle mesh of the scalp, pyDBS selects automatically a portion of the mesh located on the anterior part of the side of the head as the initial area on the skin. Ideally, this portion has to be wide enough not to be too restrictive. However, the insertion point should remain anterior to the precentral sulcus, posterior to the hair area, and not too close to the ear area. Therefore, those limits were used in pyDBS to provide the initial area for each side of the bilateral implantations.

A snapshot showing all the anatomical structures after they were processed and reconstructed as 3D meshes is shown in Fig. 2. The overall pyDBS pipeline takes about 20 minutes per patient dataset, producing 2 cases (one per side).

3.4.2 Experimental and validation setup

In this specific study, we have used the automatic planning approach published in [7], but it could have been applied to any other similar method using a weighted sum of soft constraints. For this test, we have used three soft constraints ST, DS and DV, associated to three cost functions f_{ST}, f_{DS}, and f_{DV} (with resulting values in range [0,1]) and respective weights w_{ST}, w_{DS}, and w_{DV} (also in range [0,1]). The first constraint, "ST" for *Standard Trajectory*, favors trajectories close to a standard direction of 30° from the sagittal plane and 30° from the coronal plane in the posterior-anterior direction, corresponding to typical angles of approach for the STN target [12]. Commercial planning consoles usually suggest this typical direction as the default proposition, and let the neurosurgeon refine it afterwards according to the risks. This constraint was added in order to study the influence of the use of commercial workstations. Its cost function f_{ST} is defined as equation (2), where T is a candidate trajectory and ST is the standard trajectory:

$$f_{ST} = \frac{angle(T, ST)}{90} \tag{2}$$

The two soft constraints most commonly used in the community have been chosen to complete the set of rules, as there seems to be a strong consensus on

them: the distance to vessels and the distance to ventricules. As explained before, vessels can not be segmented with a sufficient quality based on MRI only, so this constraint is substituted by the distance to sulci where major vessels are often located. Therefore, the second constraint "DS" (*Distance to Sulci*) is the maximization of the distance to sulci, and similarly the third constraint "DV" (*Distance to Ventricles*) is the maximization of the distance to ventricles. They are typical "risk" constraints. Their associated cost functions are built on the model of equation (3), where T is a candidate trajectory, AS is the anatomical structure to avoid (sulci or ventricles), D_{minAS} is the distance from which the trajectory is considered as safe regarding structure AS:

$$f_{D_{AS}} = Max\left(\frac{D_{minAS} - dist(T, AS)}{D_{minAS}}, 0\right) \tag{3}$$

Other constraints such as entry points restrictions for esthetic or functional reasons are handled in a preliminary step defining the insertion zone.

For each case, we have also collected information about the manually planned trajectory and target: coordinates of the entry point P_E and target point P_T. The manually planned target point P_T is used as the target for our automatic optimization process. T_{ref} is computed as the trajectory from P_E towards P_T.

3.5 Statistical Analysis Method

Let us recall that the three objectives of this study were: 1) to retrieve the weighting factors leading to T_{auto} most similar to T_{ref}, 2) to compare the results from two different hospitals, and 3) to compare the trends to the weighting factors empirically set. For each case, we have: one list of 200 combinations of weights for the stochastic approach, and one list of 66 weights for the exhaustive approach, each sorted according to the similarity criterion (angle α). The methods used to analyze the datasets from the two hospitals were the following.

First, on the results obtained thanks to the stochastic approach, we counted the number of occurrences of each weighting factor. This allowed to give a first rough idea of the trend on a reasonably large dataset. However, 200 samples is too large to perform pairwise comparisons. Moreover, as it will be shown in Section 4 the problem is that the distribution of the $20,000$ samples from which the 200 best were extracted was not homogeneous enough, due to the normalization, so reducing the number of selected samples could bias the results.

Indeed, to produce the $20,000$ samples each weight was randomly chosen between 0 and 1, which leaded to a homogeneous tridimensional draw. However, when normalized as explained in Section 3.2, the distribution is not homogeneous anymore due to probability laws related to central limit theorem. In our case, the combination of the three weights (1.0,0.0,0.0) can only happen if these exact numbers are drawn before normalization, whereas a combination of (0.25,0.25,0.5) can be obtained with the normalization of (0.125,0.125,0.25) or (0.25,0.25,0.5) or (0.5,0.5,1.0) or any multiple in which none of the weights exceeds 1. For this reason, the density of the draws is higher for medium weights.

Then in a second study, we performed a statistical analysis using the results of the exhaustive experiment. We first calculated the mean ranking for each combination of weights in order to identify if any trend connecting these factors could

be identified. A correlation analysis between the mean ranking and each of the weights was performed to evaluate its importance and statistical significance.

Based on this first analysis, we chose to select a subset of 20 constraints combinations: the 10 that led to the best ranking and the 10 that led to the worst ranking. Using this subset, we performed an analysis of variance using a non-parametric method, the Kruskal-Wallis analysis of variance by ranks, using the *labels* as an independent variable and the rankings as a dependent variable.

Labels are strings containing numbers. They were used for convenience to represent combinations of weights as single values. They were built as follows: label $l = w_{ST} * 1000 + w_{DS} * 10$. In case $l < 1000$, some zeros were added at the beginning of the number, so that the label always contains 4 digits. The two first digits represent the first weight, and the two last digits represent the second weight. The third weight can be deduced by a simple subtraction of their sum from 1. For instance, label 0205 represents the combination (0.2,0.5,0.3), label 0004 represents the combination (0.0,0.4,0.6), and label 1000 represents the combination (1.0,0.0,0.0) for weights (w_{ST}, w_{DS}, w_{DV}).

4 Results

All experiments were performed on an Intel Core i7 CPU at 3.4 GHz and 8GB RAM, equipped with a NVIDIA GeForce GTX 560 Ti GPU which is used to speed up occlusion computation. The computation times for each case are: around 20 minutes per patient for the pyDBS pre-processing of images; around 2 to 5 seconds to compute the insertion zone, which has to be done once for each case (twice per patient); around 0.1 second to produce an optimal trajectory for one combination of weights, leading to a processing time of 30 mn for the stochastic approach and about 7 seconds for the exhaustive approach.

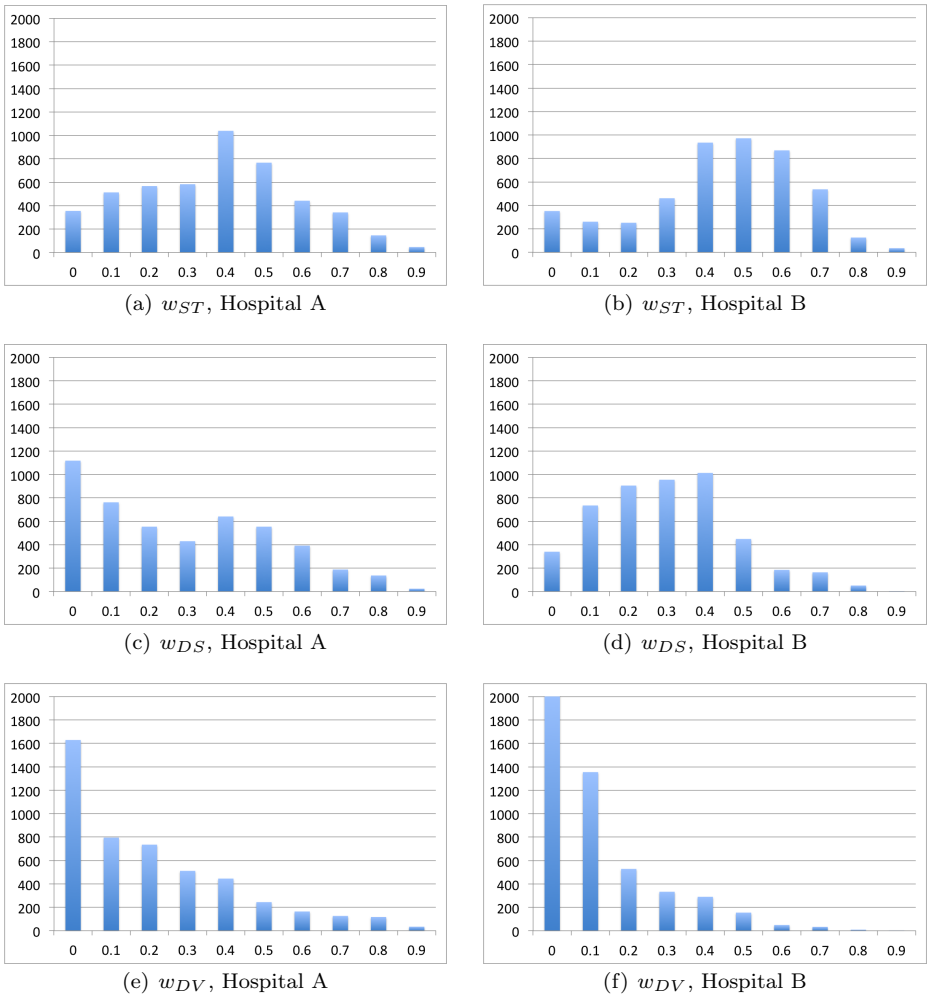
4.1 Stochastic Approach

First, we counted the number of occurrences separately for the three weighting factors and both medical centers. The results are presented in Fig. 3, where the values are discretized into ten lower precision values for readability purposes.

As mentioned in Section 3.5, these results have to be moderated by the possible bias due to the non-homogeneity of the distribution. The distribution of the $20,000$ samples of combinations of weights (w_{ST}, w_{DS}, w_{DV}) after normalization is illustrated in Fig. 4, where the three axis correspond to the three weights. The points are located in the triangle representing the intersection between the plane of equation $x + y + z - 1 = 0$ (where the sum of all point's coordinates is equal to 1) and the unity cube (that contains only coordinates that do not exceed 1). It can be observed the density is higher with medium values in the combination.

In Fig. 3 we first observe that the results are quite comparable between both sites. A preliminary trend can be observed, with a predominance of values in a range [0.4;0.5] for w_{ST}, a range [0.0;0.5] for w_{DS}, and a range [0.0;0.1] for w_{DV}, *i.e.* high values of w_{ST}, medium values of w_{DS} and low values of w_{DV}.

This trend is also illustrated in Fig. 5 where the 200 combinations of weights (w_{ST}, w_{DS}, w_{DV}) leading to trajectories T_{auto} closest to T_{ref} for three different



(a) w_{ST}, Hospital A (b) w_{ST}, Hospital B

(c) w_{DS}, Hospital A (d) w_{DS}, Hospital B

(e) w_{DV}, Hospital A (f) w_{DV}, Hospital B

Fig. 3 Count of the occurrences of the different weights for the two hospital's datasets. On the x-axis, 10 different discretized values of weights $0.0, ..., 0.9$ represent weights comprised respectively in ranges $[0.0, 0.1], ..., [0.9, 1.0]$. On the y-axis, the sum of occurrences of the weights within the defined ranges on a scale varying from 0 to 2000.

patients are displayed over the distribution. As we can see, most of the selected combinations of weights represented by blue circles, green squares and pink triangles are located in the above mentioned ranges. The statistical analysis presented below intends to validate this result.

4.2 Exhaustive approach and statistical analysis

In Fig. 4, we superimposed the distribution of the 66 samples of combinations of weights for the exhaustive method for a better understanding of their locations.

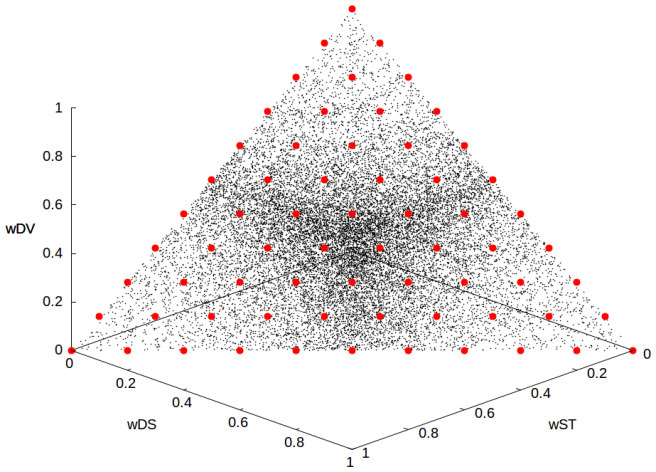


Fig. 4 Distribution of the 20,000 samples of combinations of weights (w_{ST}, w_{DS}, w_{DV}) for the stochastic approach after normalization (small black dots). The evenly spaced red spheres are the 66 samples of the exhaustive method. All of them are located within a triangle representing the intersection between plane of equation $x + y + z - 1 = 0$ and the unity cube.

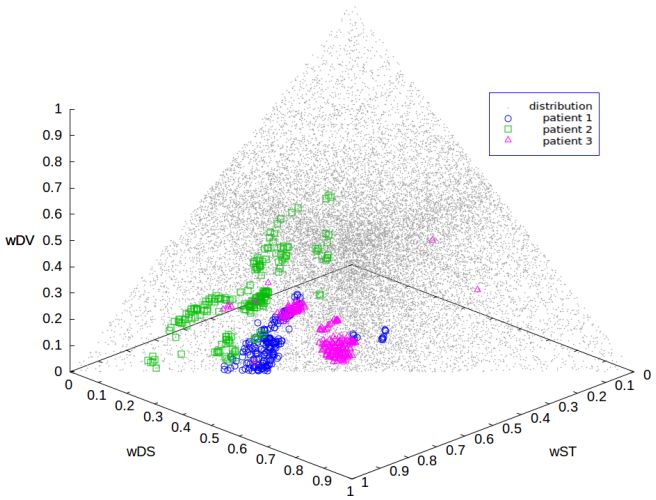


Fig. 5 Among the 20,000 random samples tested, the 200 combinations of weights (w_{ST}, w_{DS}, w_{DV}) that provided the T_{auto} closest to T_{ref} for 3 different patients are displayed in green, blue and pink.

As presented in section 3.5 we first calculated the mean ranking for each combination of weights. As shown in Fig. 6, we can observe that the mean ranking seems to decrease when the first weight increases. The labels leading to the best results are shown in the upper part of Table 2. These labels, in which the two first digits are close to 10, correspond to high values of the first weight w_{ST}. Similarly, the labels with the worst results, shown in the lower part of Table 2, have their two first digits close to 00. These labels correspond to low values of w_{ST}. We thus made the assumption that higher values of w_{ST} weight led to simulated trajec-

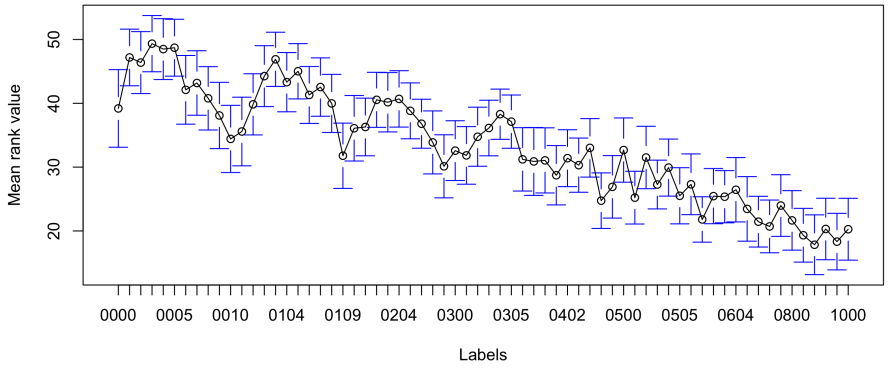


Fig. 6 Mean ranking for each combination of weights with a 95% confidence interval. The x-axis represents the combination of weights expressed as labels. The y-axis shows the mean ranking values. A lower value of mean ranking suggests that the corresponding combination of weights tends to lead to trajectories closer to the reference trajectory.

tories that were closer to those actually chosen by the surgeons. The oscillatory trend that can also be observed is linked to the way the combinations of weights are labeled.

To confirm this hypothesis, we ran separate correlation analysis between the weights and the rank using Spearman's rank correlation test. These three correlation analysis were all statistically significant ($p<0.001$), showing that the results were not due to random sampling. For the first weight w_{ST}, the calculated correlation is ρ = -0.406, meaning that the rank is inversely correlated to the weight value. The mean rank decreases while the value of w_{ST} increases (see figure 7). For the second and the third constraints, the correlation values are $\rho = 0.144$ and $\rho = 0.263$ respectively, meaning that the mean rank increases when w_{DS} or w_{DV} increase (see figures 8 and 9). These results are summarized in Table 1.

In order to further validate these results, we also ran a Kruskal-Wallis analysis of variance on the rankings, comparing the 20 labels: the 10 combinations leading to the best rank and the 10 combinations leading to the worst rank. These labels are summarized in Table 2. We chose to ran the analysis on this subset, since our objective was not to compare the all set of labels, but to validate the trend we identified with these previous results. Comparing all 66 labels would have led to

Table 1 Spearman's correlation coefficients

	w_{ST}	w_{DS}	w_{DV}
p-value	$< 2.2e^{-16}$	$< 2.2e^{-16}$	$< 2.2e^{-16}$
ρ-value	-0.406	0.144	0.2630

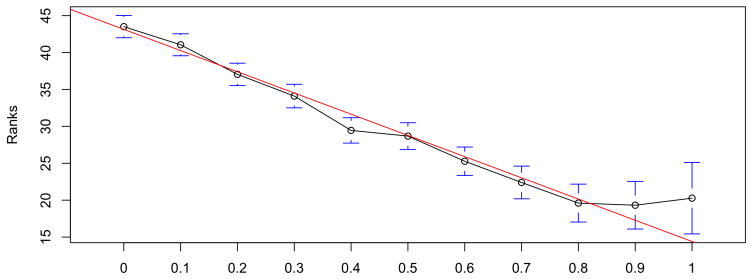


Fig. 7 Mean ranks for the values of the first weight w_{ST}, and regression line (in red).

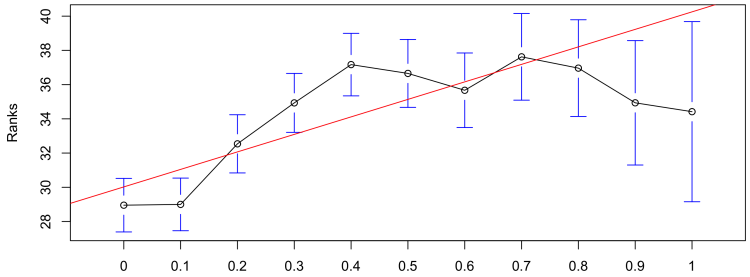


Fig. 8 Mean ranks for the values of the second weight w_{DS}, and regression line (in red).

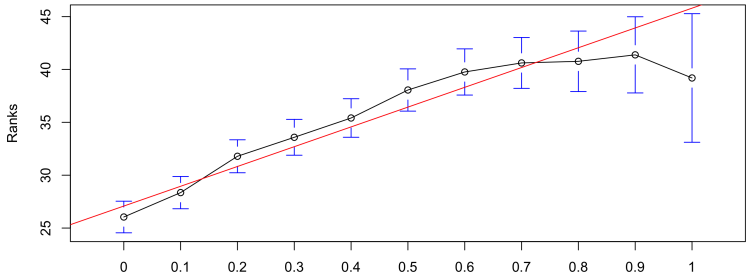


Fig. 9 Mean ranks for the values of the third weight w_{DV}, and regression line (in red).

confused results, and would have been hard to evaluate and interpret. Thus, we eliminated from this analysis the labels that led to intermediary results.

The analysis of variance showed a significant influence of the labels on the rank values ($p<0.001$). We thus ran a pairwise comparison using the Wilcoxon rank sum test. All inter-group pairwise comparisons appeared to be significant ($p<0.001$). On the other hand, all intra-group pairwise comparisons were not significant with $p>0.05$. Again, this result validates that the differences between the 2 groups of selected labels were not due to random sampling.

To determine if the medical center had an influence on the results, we also ran an analysis of variance using site as an independent variable. The results showed that the site had no significant influence on the rankings, with $p>0.05$.

Group 1	0802	0901	0801	1000	0900	0702	0701	0800	0601	0700
Means	17.83	18.32	19.32	20.27	20.30	20.70	21.45	21.65	21.80	23.45
STD	20.47	16.41	17.95	16.29	17.63	16.48	19.91	18.68	18.37	19.20
Group 2	0003	0005	0004	0001	0103	0002	0105	0102	0104	0007
Means	49.34	48.70	48.49	47.18	46.89	46.38	45.01	44.25	43.30	43.18
STD	17.92	16.38	17.81	17.32	15.56	17.24	17.86	15.28	14.70	18.70

Table 2 Means of ranks for two subsets of labels (representing combinations of weights), along with their standard deviations: Group 1 contains the 10 best labels leading to the lower mean rank values. Group 2 contains the 10 worst labels leading to the higher mean rank values. Both groups were found to be statistically different.

5 Discussion and Conclusion

Planning preoperatively a safe and optimal trajectory for DBS is a difficult exercise which requires a lot of expertise. Many research groups have implemented their own algorithms to help the neurosurgeons in this task, most of the time using a weighted sum of cost functions representing rules, with weights empirically set. In all cases, the approach aims at mimicking the surgeons decision-making process. However, it has been highlighted in the literature, for instance by Bériault *et al.* [4] and Liu *et al.* [11], that capturing the weights surgeons use implicitly is really difficult. Even the constraints that each research group use cannot be considered as unanimously accepted. Liu *et al.* even mention in [11] that surgeons of the same site did not report using the same constraints. Moreover, they observed that when asked to set manually the weights they intuitively use, both surgeons first tuned the weights in a different manner, and could later be even more satisfied with another setting. As there is a lack of ground truth and consensus, this study tried to enlighten some implicit practices with an objective study.

The first observation is that there is a similarity of the results between both hospitals, where surgeons are not used to work together. This statistically significant result strengthens our conclusions and allows us to hypothesize that this study can be generalized, even if a larger study including more sites could help in confirming this statement. At least, we can deduce that the results were not biased nor influenced by a single site.

Secondly, the trends clearly show that when a high value is given to weight w_{ST}, the rank is lower. It means that with a high value for the constraint of proximity to a standard position, the automatically computed trajectory T_{auto} gets closer to the manually planned trajectory T_{ref}. Let us note that in many patient cases, it was possible to find another combination of weights leading to an optimal trajectory for which the distance to sulci and ventricles was larger, but a little more distant to the standard direction. This means that the surgeon chose a trajectory that our algorithm considered as suboptimal regarding the risk rules.

This result can be quite surprising at first, as we could expect a higher value for the weighting factors associated to the risks (distance to sulci and ventricles). Indeed, the default weighting proposed in our software, set in cooperation with the neurosurgeons, was w_{ST}=0.2, w_{DS}=0.4, w_{DV}=0.4. In most other studies, ST

constraint is not mentioned or used. Liu et al. mention in [11] that some surgeons prefer the entry point to be near the coronal suture, which can somewhat be the most comparable constraint. In their study, the weights chosen by two neurosurgeons are around 0.56 for the proximity to the vessels and sulci together, 0.22 for the proximity to the ventricles, and only 0.18 for the proximity to the suture. In this system as in ours, the importance of this constraint compared to the others might be underestimated and seems unforeseen.

A first hypothesis to explain this phenomenon might be a strong influence of the use of the commercial planning workstations in the current practice of manual planning. The surgeons often start the search by displaying the standard direction proposed by the workstation and browse in the vicinity to find an acceptable path optimizing the other two rules. This could be due to a lack of time or low ergonomics of existing interactive software to allow for a larger search in an easy and efficient way. Another hypothesis is that other implicit rules are missing to imitate more accurately the decision making process. However, this is more unlikely as it would mean that there are many missing rules constraining the path to stay always in the immediate vicinity of the standard axis. In any case, this result suggests that further studies are needed, where this rule is suppressed and the trajectories computed with/without the rule are rated by several surgeons to determine if this rule is useful or if the standard axis proposed by the workstation is biasing the search.

From our study, we also conclude that among the two remaining rules, the risk due to the proximity to the sulci seems to be the most critical compared to the proximity to the ventricles. This is consistent with the weights authors usually empirically set for these two consensual rules, such as in [4, 11] where higher values are given to the maximization of the distance to vessels or sulci than the distance to ventricles. This result confirms this intuitive setting. However, let us mention that the very low weight indicated by the trend for DV rule can also be in some way a consequence of the high weight ST, as the standard direction tends to be already far enough from the ventricles.

Overall, the trends shown in this study confirmed the relevance of settings for some weights such as distance to vessels/sulci and distance to ventricles, regarding the surgeons' past choices. But they also highlighted a strong influence of the routinely proposed standard direction, which seem to be a starting point for the path search and to constitute the central axis of a comfort zone. Further study certainly need to be performed to confirm this hypothesis. If it is verified, an interesting challenge would be to check if the use of preoperative assistance without the use of any rule advantaging a proximity to a standard axis allows to modify the decision making process compared to the current practice, while maximizing the other rules, which would point out another benefit of using such tools.

Acknowledgments

The authors thank the French National Research Agency (ANR) for funding this work through the ACouStiC project grant (ANR 2010 BLAN 0209 02).

Conflict of Interest / Ethical Approval / Informed Consent

The authors declare that they have no conflict of interest. In this paper, a retrospective study has been performed on a dataset of anonymized images, but informed consent was obtained from all individual participants included in the study.

References

1. Baegert, C., Essert-Villard, C., Schreck, P., Soler, L., Gangi, A.: Trajectory optimization for the planning of percutaneous radiofrequency ablation of hepatic tumors. Computer Aided Surgery **12**(2), 82–90 (2007)
2. Benabid, A., Chabardes, S., Mitrofanis, J., Pollak, P.: DBS of the subthalamic nucleus for the treatment of parkinson's disease. The Lancet Neurology **8**(1), 67–81 (2009)
3. Bériault, S., Drouin, S., Sadikot, A.F., Xiao, Y., Collins, D.L., Pike, G.B.: A prospective evaluation of computer-assisted deep brain stimulation trajectory planning. In: In proceedings of CLIP'12, *Springer LNCS*, vol. 7761, pp. 42–49 (2013)
4. Bériault, S., Subaie, F.A., Collins, D.L., Sadikot, A.F., Pike, G.B.: A multi-modal approach to computer-assisted deep brain stimulation trajectory planning. International Journal of Computer Assisted Radiology and Surgery **7**(5), 687–704 (2012)
5. D'Albis, T., Haegelen, C., Essert, C., Fernandez-Vidal, S., Lalys, F., Jannin, P.: PyDBS: an automated image processing workflow for deep brain stimulation surgery. International journal of computer assisted radiology and surgery pp. 1–12 (2014)
6. D'Haese, P.F., Pallavaram, S., Li, R., Remple, M.S., Kao, C., Neimat, J.S., Konrad, P.E., Dawant, B.M.: CranialVault and its CRAVE tools: A clinical computer assistance system for deep brain stimulation (DBS) therapy. Medical Image Analysis **16**(3), 744 – 753 (2012)
7. Essert, C., Haegelen, C., Lalys, F., Abadie, A., Jannin, P.: Automatic computation of electrodes trajectories for deep brain stimulation: A hybrid symbolic and numerical approach. Int. J. Comput. Assist. Radiol. Surg. **7**(4), 517–532 (2012)
8. Essert, C., Marchal, M., Fernandez-Vidal, S., D'Albis, T., Bardinet, E., Haegelen, C., Welter, M.L., Yelnik, J., Jannin, P.: Automatic Parameters Optimization for Deep Brain Stimulation Trajectory Planning. In: proceedings of MICCAI workshop DBSMC'12, pp. 20–29 (2012)
9. Limousin, P., Krack, P., Pollak, P., Benazzouz, A., Ardouin, C., Hoffmann, D., Benabid, A.L.: Electrical stimulation of the subthalamic nucleus in advanced parkinson's disease. New England Journal of Medicine **339**(16), 1105–1111 (1998)
10. Liu, Y., Dawant, B.M., Pallavaram, S., Neimat, J.S., Konrad, P.E., D'Haese, P.F., Datteri, R.D., Landman, B.A., Noble, J.H.: A surgeon specific automatic path planning algorithm for deep brain stimulation. In: proceedings of SPIE Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling, p. 83161D (2012)
11. Liu, Y., Konrad, P., Neimat, J., Tatter, S., Yu, H., Datteri, R., Landman, B., Noble, J., Pallavaram, S., Dawant, B., D'Haese, P.F.: Multisurgeon, multisite validation of a trajectory planning algorithm for deep brain stimulation procedures. IEEE Transactions on Biomedical Engineering **61**(9), 2479–2487 (2014)
12. Machado, A., Rezai, A.R., Kopell, B.H., Gross, R.E., Sharan, A.D., Benabid, A.L.: Deep brain stimulation for Parkinson's disease: surgical technique and perioperative management. Movement disorders **21**(S14), S247–S258 (2006)
13. Shamir, R., Tamir, I., Dabool, E., Joskowicz, L., Shoshan, Y.: A method for planning safe trajectories in image-guided keyhole neurosurgery. In: proceedings of MICCAI'10, vol. 6363, pp. 457–464. Springer LNCS (2010)
14. Tirelli, P., de Momi, E., Borghese, N., Ferrigno, G.: An intelligent atlas-based planning system for keyhole neurosurgery. In: Computer Assisted Radiology and Surgery supplemental, pp. S85–S91 (2009)
15. York, M.K., Wilde, E.A., Simpson, R., Jankovic, J.: Relationship between neuropsychological outcome and DBS surgical trajectory and electrode location. Journal of the Neurological Sciences **287**(1-2), 159 – 171 (2009)