

A Bag-of-Features Approach to Predicting TMS Language Mapping Results from DSI Tractography

Mohammad Khatami¹, Katrin Sakreida², Georg Neuloh²,
and Thomas Schultz¹(✉)

¹ Department of Computer Science, University of Bonn, Bonn, Germany
schultz@cs.uni-bonn.de

² Department of Neurosurgery, RWTH Aachen University, Aachen, Germany

Abstract. Transcranial Magnetic Stimulation (TMS) can be used to indicate language-related cortex by highly focal temporary inhibition. Diffusion Spectrum Imaging (DSI) reconstructs fiber tracts that connect specific cortex regions. We present a novel machine learning approach that predicts a functional classification (TMS) from local structural connectivity (DSI), and a formal statistical hypothesis test to detect a significant relationship between brain structure and function. Features are chosen so that their weights in the classifier provide insight into anatomical differences that may underlie specificity in language functions. Results are reported for target sites systematically covering Broca’s region, which constitutes a core node in the language network.

1 Introduction

Relating the structure of the human brain to its function is one of the most fundamental challenges in neuroscience [3]. Diffusion MRI (dMRI) is an established method for *in vivo* mapping of fiber bundles based on how they affect the motion of water molecules. Reconstructed fibers from dMRI have been used to derive a structural subdivision of gray matter structures, such as the thalamus or parts of the cortex, based on their connectivity [2,4].

Transcranial Magnetic Stimulation (TMS) is a non-invasive method to investigate the function of certain brain regions. It is also referred to as a “virtual lesion technique” when inducing an acute and reversible focal dysfunction [14], and it is currently being explored as a tool for neurosurgical planning [15]. In TMS, a magnetic coil is placed near the head of a subject, and is used to induce, through the skin and skull, an electric current in the nearby part of the cortex.

This work was partly supported by the DFG under grant SCHU 3040/1-1.

Electronic supplementary material The online version of this chapter (doi:[10.1007/978-3-319-66182-7_74](https://doi.org/10.1007/978-3-319-66182-7_74)) contains supplementary material, which is available to authorized users.

Observing how such stimulation affects the subject’s ability to perform specific tasks, such as naming objects shown to them, allows us to map the function of the respective brain region [17].

In this work, we propose a novel computational method that allows us to explore the relationship between cortical connectivity, as indicated by dMRI fiber tractography, and its function, as observed in a TMS experiment. A related task on which more prior work exists is to investigate the relationship between dMRI tractography and task-based functional MRI (fMRI), which performs functional mapping by imaging changes in blood oxygenation during certain tasks. In this context, it is common to cluster cortical regions with similar connectivity, and compare the results with regions of functional activation [13], or to seed tractography in areas exhibiting distinct function in fMRI, and visually observe differences in the resulting connectivity patterns [10].

Our approach goes beyond this by quantifying the extent to which differences in connectivity make it possible to predict differences in function. We introduce a computational pipeline that combines tractography, clustering, and image registration with supervised classification and a formal test of the null hypothesis that the observed functional pattern cannot be predicted with higher accuracy than a randomly permuted one. Successful prediction indicates that gray matter function is closely related to its connectivity. Note that TMS affects the local function, and does *not* alter the structural connectivity captured by DSI.

We apply our framework to data from a language mapping task. First, we predict function at a specific target site based on observations elsewhere in the same subject. Then, we predict functional patterns purely based on training data from other subjects. In both settings, our method successfully rejects the above-mentioned null hypothesis, indicating presence of a statistically significant structure-function relationship that generalizes between subjects. Visualizing feature weights and correlations enables additional insight on specific anatomical differences that may underlie language functions.

2 Data Acquisition

Our experimental standards and all procedures performed in this study involving human participants were in accordance with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards and were approved by the local ethics committee (EK 054/13). Prior to investigation, we obtained written informed consent from all of our volunteers.

Diffusion Spectrum Imaging [18] was performed on 12 healthy, left-dominant volunteers on a 3T Prisma (Siemens, Germany) with $136 \times 136 \times 84$ voxels of 1.5 mm isotropic size, TE/TR = 69/11600 ms, including one $b = 0$ image and 128 DWIs with up to $b = 3000$ on a Cartesian grid. Anatomical T_1 weighted images with $240 \times 256 \times 176$ voxels of 1 mm isotropic size were also collected.

While each subject named objects shown to them on a computer screen, TMS was performed at 30 target sites distributed on a uniform grid of size 5×6 , covering the entire pars opercularis and the pars triangularis of the left inferior frontal gyrus, as well as the anterior part of the inferior precentral gyrus,

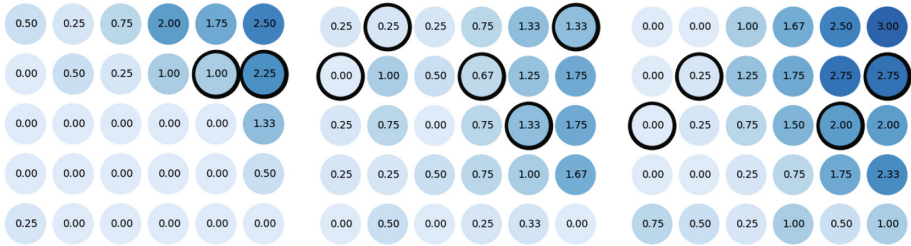


Fig. 1. TMS language scores of our three subjects. Sites that are misclassified in the within-subject analysis are marked with a black circle.

as identified in the T_1 image. Mapping was repeated five times for each target, including a sham condition. Audiovisual recordings of the experiment were analyzed off-line by two speech and language therapists, who rated the severity of language-specific errors at each site with a numerical score that accounts for all repetitions. The values indicate no significant error (0.0–0.33), slight (0.5–1.0), moderate (1.25–1.67), severe (1.75–2.33), or extreme (2.5–3.0) errors. The scale does not allow for numerical values that would fall between these classes.

We report results for direct prediction of this score, as well as for a classification task, in which we only distinguish between target sites where TMS either has a clear effect on language production (1.25–3.0), and ones where it has at most a slight effect (0.0–1.0). In several of the subjects, no or very few targets exhibited even a moderate response, which made them unsuitable for this type of analysis. Therefore, we present our initial proof-of-concept on a subset of three subjects with a clear region of at least moderate response, as shown in Fig. 1.

3 Image Analysis and Classification Pipeline

An overview of our proposed computational pipeline is shown in Fig. 2.

3.1 Preprocessing

Our navigated TMS system (LOCALITE Biomedical Visualization Systems, Germany) registers the coil to T_1 image space by establishing landmark correspondences via a pointer, and refining them via surface registration. The focus point of each TMS stimulation is estimated by projecting the position of the coil onto the brain surface. Based on the expected region of influence, we define the target site as a $(3\text{ mm})^3$ volume around the focus, take the union of the resulting target maps across runs, and register the T_1 to the Diffusion Spectrum Imaging space of the same subject using linear registration (FLIRT) from FSL [12].

Multi-fiber deterministic streamline-based tractography is performed using a previously described constrained deconvolution approach for DSI data [1]. One seed point is generated in a random location within each voxel in a volume that extends 7 voxels (in 1.5 mm DSI space) beyond the union of all target sites.

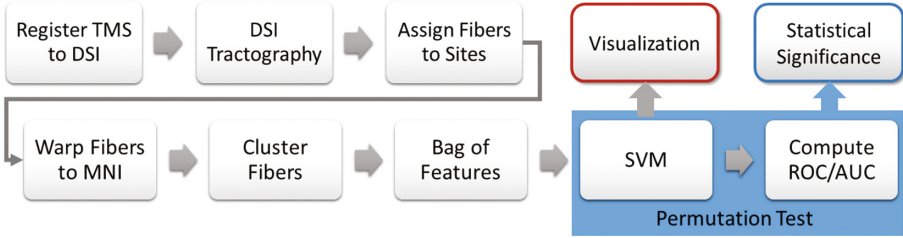


Fig. 2. Our computational pipeline combines tractography, registration, clustering, and supervised learning into a framework that enables formal hypothesis testing and visualization for anatomical interpretation.

Tracking uses step size 0.5 mm, and terminates when exceeding a maximum turning angle of 45° , or when leaving the region in which the white matter volume fraction as estimated by the multi-tissue deconvolution drops below 0.5.

Fiber tractography is performed in the scanner space of the individual DSI measurements. Fibers are subsequently mapped to a common space using an in-house script. It applies the warp that results from nonlinear registration of Fractional Anisotropy images to an MNI template, which is a standard operation (called FNIRT) within the publicly available FSL package [12].

3.2 Feature Representation

A key step in creating a method that predicts function from fiber tractography is to represent the estimated connectivity as a feature vector. We follow a “bag-of-features” approach [16], which we design so that features retain an anatomical interpretation, as will be illustrated in our experiments.

We assign a subset of fibers to each target site. Considering that streamlines sample bundles which individual axons might enter or leave along the way, we assign any streamline that passes a site within 0.5 mm to that site. The margin accounts for the fact that target sites are located in the gray matter, and streamlines may terminate before reaching them. We make sure that each site is assigned at least the 100 nearest streamlines. This is required for some sites that fall in between gyri, and that would otherwise receive very few or no fibers.

Streamlines that are not assigned to any target are discarded. The remaining N fibers are represented as 9D vectors as proposed in [5], and clustered using the k means algorithm. Given the resulting k clusters, each target site t is characterized by a k -dimensional vector \mathbf{n}^t whose entries n_i^t are the number of streamlines from cluster i that have been assigned to target t . We obtain feature vectors \mathbf{x} using a term frequency – inverse document frequency weighting

$$x_i^t := \frac{n_i^t}{n^t} \log \frac{N}{n_i}, \quad (1)$$

where n^t is the sum of all fibers assigned to t , and n_i is the sum of all fibers in cluster i .

We found that the optimum number of clusters k differs between subjects and between within- and across-subject analysis. A simple general strategy that is used throughout this paper is to construct five different feature vectors based on $k \in \{10, 20, 30, 40, 50\}$, ℓ_2 normalize each of them, and concatenate them into a single 150-dimensional feature vector.

3.3 Classification, Regression, and Evaluation

Classification is performed using a linear soft-margin support vector machine (SVM) [8]. Since our training data is quite unbalanced, we assign a higher weight to the smaller class, by setting values of the SVM parameter C to the fraction of training samples available for the opposite class. Linear regression is performed with an ℓ_2 regularizer whose weight was empirically fixed at $\alpha = 7$.

Accuracy is evaluated using two different modes of cross-validation. For within-subject analysis (leave-one-target-out), each fold trains on the features and labels from 29 sites of a single subject, and makes a prediction for the remaining target from that same subject. For across-subject analysis (leave-one-subject-out), we independently predict all 30 sites of one subject after training on all features and labels of the two remaining subjects.

For regression, we evaluate the coefficient of determination R^2 . For classification, we compute the area under the ROC curve, which is more informative than overall classification accuracy in case of unbalanced data. The ROC curve plots true positive rate over false positive rate, and is obtained by systematically varying the threshold of the SVM decision function. The area under this curve (AUC) equals the probability that, given two randomly chosen examples from different classes, the classifier will rank them correctly [9]. Random guessing would lead to a diagonal ROC curve with $\text{AUC} = 0.5$. Larger AUCs indicate predictive power above chance level.

We use a formal permutation-based test of whether our predictions are significantly better than chance. For this, we repeat the cross-validation 1 000 times, each time with a random permutation of the target labels. The same permutation is applied on training and test data. We record the R^2 or AUC values corresponding to the true labels and all random permutations. Finally, we compute a p value as the fraction of runs in which R^2/AUC was at least as large as for the true labels. This test corresponds to the null hypothesis that the observed labels cannot be predicted with higher accuracy than randomly permuted ones. Rejecting it supports the hypothesis of a structure-function relationship, which is destroyed by randomly permuting the functional labels.

4 Results and Discussion

4.1 Within- and Between-Subject Analysis

We first tried classification on three subjects in a leave-one-target-out manner. This led to the ROC curves shown in Fig. 3 (left) with areas under the curve

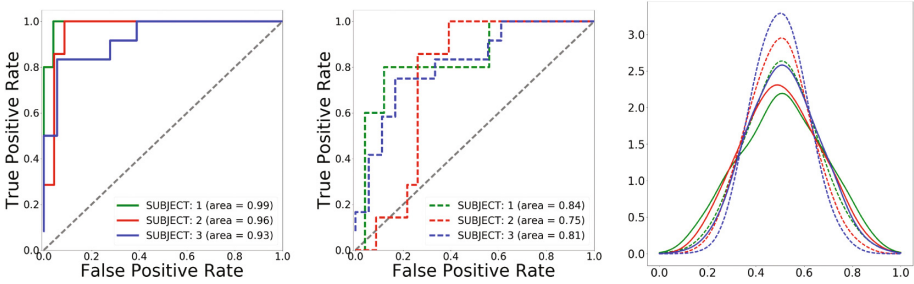


Fig. 3. For all three subjects (different colors), ROC curves indicate predictive power that is clearly above chance level both in the within-subject (left) and the between-subject setting (center). The plot on the right shows the null distributions of area under the curve, as estimated by our permutation test.

(AUC) 0.99, 0.96, and 0.93, respectively. At a level of $\alpha = 0.05$, permutation-based testing rejected the null hypothesis that the same AUC can be achieved for a random rearrangement of labels, with p values 0.001, 0.002, and 0.003. Estimated null distributions are illustrated in the right panel of Fig. 3.

We also performed leave-one-subject-out cross-validation on the same three subjects. Inter-subject variability makes this case more challenging, which is reflected in the ROC curves in the center of Fig. 3, with reduced AUC (0.84, 0.75, and 0.81). Corresponding p values were greater (0.039, 0.017, 0.010), but still below the level of $\alpha = 0.05$.

Similarly, regression worked better within subjects (R^2 values 0.66, 0.57, 0.78) than between subjects (0.42, 0.30, 0.36). In all cases, predicting the true numerical values gave significantly better results than trying to learn randomly permuted ones ($p = 0.001$).

4.2 Visualization of Feature Weights and Correlations

In order to gain insight into the specific structural differences that allow us to predict the functional classification, Fig. 4 visualizes streamlines representing the 150 cluster centers used in the leave-one-subject-out experiment, colored according to the weight of the respective feature in the linear SVM (top), or the Pearson correlation between the feature and the (non-thresholded) TMS language score (bottom), averaged over subjects.

Fibers that contribute to a classification as “clearly language active” or which are positively correlated with the TMS score are shown in red; blue fibers contribute to the opposite classification or indicate a negative correlation. Each of the three cross-validation folds leads to its own weights and correlations. To focus on the effects that could be reproduced across all subjects, we color streamlines as white if the direction of the effect is not the same in all three cases.

Figure 4 helps confirm the anatomical plausibility of the learned classifier, since fiber systems known to be relevant for speech production, such as parts of

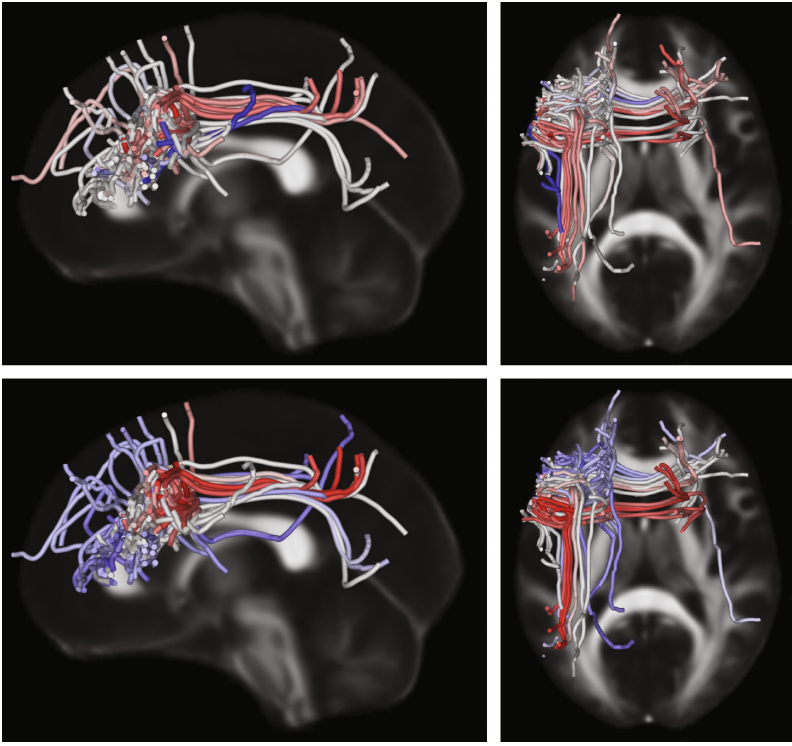


Fig. 4. Visualization of the cluster centers used for prediction. On the top, colors indicate SVM weights whose signs were identical for all three subjects; on the bottom, they show correlation coefficients with consistent signs. Red indicates a positive correlation with language impairment, blue a negative one.

the arcuate fasciculus [7], and connections to the supplementary motor area [6], are shown to lead the SVM to classify corresponding TMS target sites as having a clear effect on speech. Somewhat surprisingly, the SVM regards two lateral cluster centers that appear to run at the outer boundary of the arcuate fasciculus as evidence against a strong effect on speech. As shown in the bottom image, these particular tracts do not have a clear correlation with the TMS scores. It has been pointed out that linear classifiers may assign negative weights to features that are statistically independent from the label in order to cancel out distractors [11], which in our case might result from errors in the fiber tractography.

5 Conclusion

We have proposed the first computational pipeline that predicts TMS mapping results from DSI tractography and used it to investigate structure-function relationship in a language mapping task. Our predictions were significantly above

chance level both within and between subjects, and we visualized the features to gain insights into what anatomical differences drive the prediction. Future work will evaluate and discuss a larger set of subjects.

References

1. Ankele, M., Lim, L.H., Groeschel, S., Schultz, T.: Versatile, robust, and efficient tractography with constrained higher-order tensor fODFs. *Int. J. Comput. Assist. Radiol. Surg.* doi:[10.1007/s11548-017-1593-6](https://doi.org/10.1007/s11548-017-1593-6) (2017)
2. Anwender, A., Tittgemeyer, M., von Cramon, D., Friederici, A., Knösche, T.: Connectivity-based parcellation of Broca's area. *Cereb. Cortex* **17**(4), 816–825 (2007)
3. Behrens, T.E.J., Johansen-Berg, H.: Relating connectional architecture to grey matter function using diffusion imaging. *Philos. Trans. Royal Soc. B Biol. Sci.* **360**(1457), 903–911 (2005)
4. Behrens, T.E.J., Johansen-Berg, H., Woolrich, M.W., Smith, S.M., Wheeler-Kingshott, C.A.M., Boulby, P.A., Barker, G.J., Sillery, E.L., Sheehan, K., Ciccarelli, O., Thompson, A.J., Brady, J.M., Matthews, P.M.: Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat. Neurosci.* **6**(7), 750–757 (2003)
5. Brun, A., Knutsson, H., Park, H.-J., Shenton, M.E., Westin, C.-F.: Clustering fiber traces using normalized cuts. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) *MICCAI 2004. LNCS*, vol. 3216, pp. 368–375. Springer, Heidelberg (2004). doi:[10.1007/978-3-540-30135-6_45](https://doi.org/10.1007/978-3-540-30135-6_45)
6. Catani, M., Dell'Acqua, F., Vergani, F., Malik, F., Hodge, H., Roy, P., Valabregue, R., de Schotten, M.T.: Short frontal lobe connections of the human brain. *Cortex* **48**(2), 273–291 (2012)
7. Catani, M., Jones, D.K., ffytche, D.H.: Perisylvian language networks of the human brain. *Ann. Neurol.* **57**(1), 8–16 (2004)
8. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
9. Fawcett, T.: An introduction to ROC analysis. *Pattern Recogn. Lett.* **27**(8), 861–874 (2006)
10. Friederici, A.D., Bahlmann, J., Heim, S., Schubotz, R.I., Anwender, A.: The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci.* **103**(7), 2458–2463 (2006)
11. Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.D., Blankertz, B., Bießmann, F.: On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage* **87**, 96–110 (2014)
12. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: *NeuroImage. FSL* **62**(2), 782–790 (2012)
13. Johansen-Berg, H., Behrens, T.E.J., Robson, M.D., Drobnjak, I., Rushworth, M.F.S., Brady, J.M., Smith, S.M., Higham, D.J., Matthews, P.M.: Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc. Natl. Acad. Sci.* **101**(36), 13335–13340 (2004)
14. Pascual-Leone, A., Walsh, V., Rothwell, J.: Transcranial magnetic stimulation in cognitive neuroscience - virtual lesion, chronometry, and functional connectivity. *Current Opin. Neurobiol.* **10**(2), 232–237 (2000)

15. Picht, T.: Current and potential utility of transcranial magnetic stimulation in the diagnostics before brain tumor surgery. *CNS Oncol.* **3**(4), 299–310 (2014)
16. Sivic, J., Zisserman, A.: Video google: a text retrieval approach to object matching in videos. In: *Proceedings of IEEE International Conference on Computer Vision*, pp. 127–144 (2003)
17. Tarapore, P.E., Findlay, A.M., Honma, S.M., Mizuiri, D., Houde, J.F., Berger, M.S., Nagarajan, S.S.: Language mapping with navigated repetitive TMS: proof of technique and validation. *NeuroImage* **82**, 260–272 (2013)
18. Wedeen, V.J., Hagmann, P., Tseng, W.Y.I., Reese, T.G., Weisskoff, R.M.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magn. Reson. Med.* **54**(6), 1377–1386 (2005)