Voxel-wise Modeling with Spatial Regularization: Application to Semantic Mapping during Natural Listening

Özgür Yılmaz (ozguryilmaz@ee.bilkent.edu.tr)
Department of Electrical and Electronics Engineering, Bilkent University
Ankara, TR 06800 Turkey

Emin Çelik (ecelik@bilkent.edu.tr)
Neuroscience Program, Bilkent University
Ankara, TR 06800 Turkey

Tolga Çukur (cukur@ee.bilkent.edu.tr)
Department of Electrical and Electronics Engineering, Bilkent University
Ankara, TR 06800 Turkey

Abstract:

Voxel-wise modeling (VM) is a powerful tool that is used to estimate responses of individual voxels evoked by features of complex natural stimuli. Still, VM discards spatial correlations across functional selectivities of neighboring voxels, and this can lead to decreased sensitivity during model estimation with noisy measurements. Here, we describe a spatially-informed voxel-wise modeling (SPIN-VM) method that utilizes response correlations in spatial neighborhoods of voxels. SPIN-VM performs regularization both across spatial neighborhoods of voxels and across model features of individual voxels. We compared the performance of SPIN-VM to regular VM on a dataset collected from a natural listening experiment. Compared to VM, SPIN-VM leads to higher prediction performances and better capturing of local coherence of semantic representations with SPIN-VM.

Keywords: fMRI, Voxel-wise modeling, Spatial Correlation

Introduction

In naturalistic fMRI experiments, voxel-wise modeling (VM) is a powerful tool that is used to assess cortical representations with improved sensitivity (Kay et al., 2008). For fMRI experiments, the aim in VM is to estimate voxel-wise BOLD responses in terms of model features of naturalistic stimuli (Naselaris et al., 2011). VM uses regularization across the model weights to alleviate noise and ensure generalizability. Since VM models each voxel independently without any spatial smoothing, it increases sensitivity for detecting voxel-wise functional selectivity (Dumoulin and Wandell, 2008). Still, VM discards information from spatial correlations across model weights of neighboring voxels and this leads to decreased sensitivity with noisy measurements.

Here, we describe a spatially informed voxel-wise modeling (SPIN-VM) method that better utilizes response correlations among neighboring voxels. To do that, we implement a weighted graph Laplacian that uses the distance between neighboring voxels (Grosenick et al., 2013; Penny et al., 2005). Like VM, SPIN-VM performs regularization across model features but with an additional regularization across spatial neighborhoods of voxel-wise model weights. SPIN-VM still maintains high sensitivity of detecting differences among functional selectivity of individual voxels since it still predicts BOLD responses in single voxels. To test whether SPIN-VM improves modeling performances, we implemented both methods on an fMRI dataset collected in a natural story listening experiment (Huth et al., 2016). In the experiment, we used a voxel-wise semantic model that estimates BOLD responses in terms of the semantic features of the stories. Regularized linear regression was used to fit voxel-wise models that optimally predict the measured BOLD responses in terms of the semantic features. Models obtained using VM and SPIN-VM were compared in terms of single-voxel prediction accuracy and smoothness of semantic maps.

Methods

MRI protocols

MRI data were collected on a 3T Siemens Tim Trio scanner at the University of California, Berkeley using a 32-channel head coil. Gradient EPI sequence was used with repetition time = 2.00 s, echo time = 33 ms, flip angle = 70°, voxel size = 2.24 x 2.24 x 4.13 mm^3, slice thickness = 3.5 mm with 18% slice gap, field of view = 224 x 224 mm^2 and 32 axial slices to cover the entire cortex. Anatomical data were collected using T_1-weighted multi-echo MP-RAGE sequence with voxel size = 1 x 1 x 1 mm^3 and field of view = 256 x 212 x 256 mm^3.

Experimental procedures

Subjects listened to naturally spoken stories that were selected from The Moth Radio Hour which covers a wide range of topics. In each story, a single speaker tells a memoir to a live audience. Whole-brain blood-oxygen-level-dependent (BOLD) responses were recorded via functional magnetic resonance imaging (fMRI) from five human subjects during listening. However, in this paper, we compared VM and SPIN-VM performances for a single subject.

ROI abbreviations

Regions of interest (ROIs) used in this paper were: Auditory cortex (AC), Broca's area (Broca), Wernicke's area (WER), angular gyrus (AngG), supramarginal gyrus (SupMG), frontal operculum (FO), intraparietal sulcus (IPS), frontal eye fields (FEF), superior/middle/inferior frontal gyri (SFG, MFG, IFG), and retinotopic early visual areas (RET: V1-V4).

Semantic model construction

To estimate semantic tuning of single voxels, we constructed a semantic model that explains BOLD responses in terms of the semantic features of the story stimuli. To extract semantic features, a word embedding space was formed based on word co-occurrence statistics. In this space, co-occurrences of a word with 985 basic words were taken as the features of that word (Huth et al., 2016). The assumption here is that words with similar meaning tend to occur in nearby positions in text. We then projected each word in the stories onto this space. This procedure yielded a stimulus matrix of 985 x number of TRs. Additionally, to rule out possible biases due to correlations between semantic features and word-rate, phoneme rate, and phoneme types, we included these features as nuisance regressors in our semantic model.

Voxel-wise model estimation and validation

We fit voxel-wise models between the stimulus matrix and BOLD responses using L2-regularized ridge regression. We performed model fitting and performance evaluations via a nested cross-validation procedure. The regularization parameters were determined in inner folds and prediction performances were calculated by using the selected parameters in outer folds. Separate models were fit for 10 different parameters in the range 10^3 to 10^8 (spaced logarithmically). We calculated prediction scores as the Pearson's correlation between measured and predicted BOLD responses.

Construction of the semantic space

To visualize semantic tuning of cortical voxels, we implemented principal component analysis (PCA) on the tuning vectors obtained from the listening task. We selected the first three PCs that captured 32.25 ± 2.33% (mean ± s.d. across subjects) of the variance in tuning vectors. We then projected voxel-wise tuning vectors onto three PCs and assigned RGB colors to the voxels according to those projections. We then visualized semantic tuning of cortical voxels on flat maps.

Results

We have obtained voxel-wise prediction performances and semantic maps for VM and SPIN-VM, separately. In Fig. 1, the prediction performances of the two methods across important ROIs are shown. Except for RET, prediction performances significantly increase in all ROIs (Wilcoxon signed-rank test, $P < 0.05$). In Fig. 2, we see that semantic maps for VM have patchy regions across many auditory and attention control areas. However, for SPIN-VM, the resulting semantic maps are smoother. SPIN-VM also presents semantic maps that are locally more coherent. Especially auditory areas (AC, Broca, WER) and attention-control areas (FEF, IPS) have relatively smoother semantic maps in SPIN-VM compared to VM. These results indicate that SPIN-VM increases prediction performance and better captures locally coherent semantic maps.

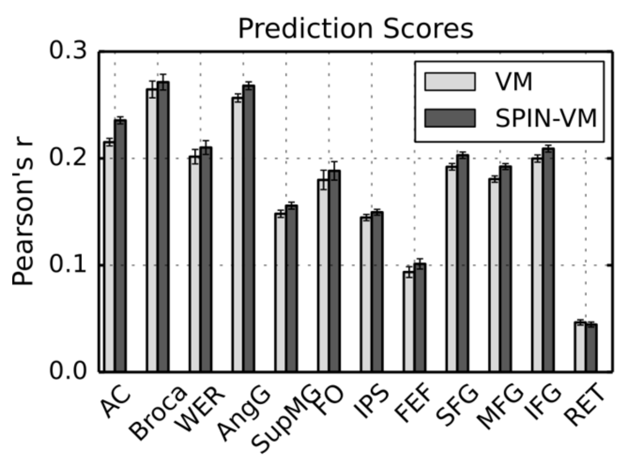

Figure 1: Comparison of prediction performances of VM and SPIN-VM across important regions of interest (ROIs). Except for RET, prediction performances increase significantly in all ROIs (Wilcoxon signed-rank test, $P < 0.05$).

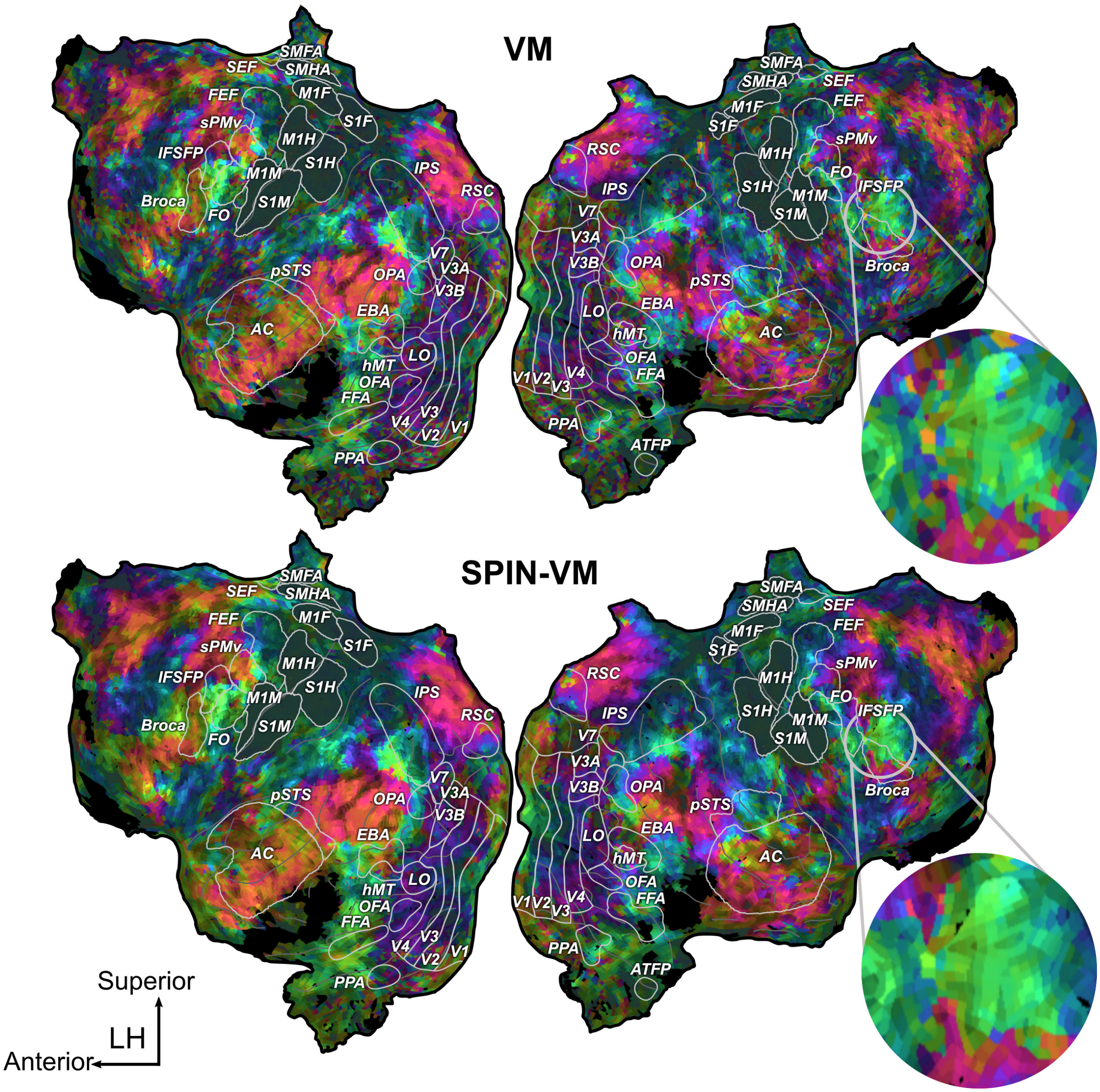

Figure 2: Comparison of semantic maps measured with VM and SPIN-VM. SPIN-VM provides smoother and locally coherent semantic maps.

Acknowledgments

The work was supported in part by grants from the National Eye Institute (EY019684), XX. The work was supported in part by a Marie Curie Actions Career Integration Grant (PCIG13-GA-2013-618101), by a European Molecular Biology Organization Installation Grant (IG 3028), by a TUBA GEBIP 2015 fellowship, by an ASELSAN PhD scholarship and by a BAGEP 2017 fellowship. We thank Alex G. Huth and Jack L. Gallant for their help in various aspects of this research.

References

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452*, 352-355.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage, 56*, 400–410.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage, 39*, 647–660.

Grosenick, L., Klingenberg, B., Katovich, K., Knutson, B., & Taylor, J. E. (2013). Interpretable whole-brain prediction analysis with GraphNet. *Neuroimage, 72*, 304–321.

Huth, A. G., Heer, W. A. De, Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*, 453–458.