Sparse coding of ECoG signals identifies interpretable components for speech control in human sensorimotor cortex

Kristofer E. Bouchard[1,2⋆], Alejandro F. Bujan [3], Edward F. Chang[4], & Friedrich T. Sommer[3]

Abstract— **The concept of sparsity has proven useful to understanding elementary neural computations in sensory systems. However, the role of sparsity in motor regions is poorly understood. Here, we investigated the functional properties of sparse structure in neural activity collected with high-density electrocorticography (ECoG) from speech sensorimotor cortex (vSMC) in neurosurgical patients. Using independent components analysis (ICA), we found individual components corresponding to individual major oral articulators (i.e., Coronal Tongue, Dorsal Tongue, Lips), which were selectively activated during utterances that engaged that articulator on single trials. Some of the components corresponded to spatially sparse activations. Components with similar properties were also extracted using convolutional sparse coding (CSC), and required less data pre-processing. Finally, individual utterances could be accurately decoded from vSMC ECoG recordings using linear classifiers trained on the high-dimensional sparse codes generated by CSC. Together, these results suggest that sparse coding may be an important framework and tool for understanding sensory-motor activity generating complex behaviors, and may be useful for brain-machine interfaces.**

INTRODUCTION

Sparse coding was successful in elucidating neural sensory processing, serving as a normative theory and also as a data analysis tool, but it has rarely been applied to motor areas. For data analysis, sparse coding principles have induced various algorithms that decompose complex spatio-temporal patterns into 'components' that can be physically meaningful. This offers alternatives to the dominant method for decomposing spatio-temporal patterns of neural activity from motor cortex, PCA. PCA yields low-dimensional state-space descriptions that represent a large fraction of the variance in the neural measurements, but the state-space dimensions do not correspond to behaviorally meaningful quantities, hindering understanding. For example, our previous analysis of ECoG signals from vSMC with PCA [1], failed to extract components that were associated with individual speech articulators, thereby obscuring their interpretation.

Independent components analysis (ICA) [2] and sparse coding (SC) [3], [4] are related techniques that isolate signal components by imposing sparsity constraints. While PCA operates solely on the second-order covariance structure of the data, ICA and SC are able to leverage higher-order structure, perhaps more directly correlated with behavior. In fact, ICA and SC have already been applied to extract behaviorally relevant components from hippocampal local field potentials ([5]). Here, we demonstrate for the first time that ICA and SC can extract individual components from human ECoG that correspond to individual vocal tract articulators engaged in speech production. We propose the identification of distinct motor-control signals that can independently be assigned to individual articulators may be useful for future brain-machine interfaces.

[1]Lawrence-Berkeley National Laboratory, Berkeley
[2]Helen Wills Neuroscience Institute, UC Berkeley
[3]Redwood Center for Theoretical Neuroscience, UC Berkeley
[4]UCSF Epilepsy Center, UC San Francisco
⋆ Corresponding Author

METHODS

Subjects and task

The experimental protocol was approved by the Human Research Protection Program at the University of California, San Francisco. Three native English speaking human subjects underwent chronic implantation of a high-density, subdural electrocortigraphic (ECoG) array over the language dominant hemisphere as part of their clinical treatment of epilepsy. Subjects gave their written informed consent before the day of surgery. Data from two of these subjects have been used in previous studies (e.g.,[1])

Each subject read aloud consonant-vowel syllables (CVs) composed of 19 consonants followed by one of three vowels (/a/, /i/ or /u/). Each CV was produced between 15 and 100 times total.

256 channel high-density electrocorticography

We used electrocorticography (ECoG) arrays (4mm pitcth, 16x16 electrodes) implanted subdurally to record cortical field potentials (FPs) directly from the surface of the brain. FPs were recorded with a multi-channel amplifier optically connected to a digital signal processor [3052 Hz] (Tucker-Davis Technologies [TDT], Alachua, FL). The spoken syllables were recorded with a microphone, digitally amplified, and recorded inline with the ECoG data [acquired at 22kHz]. The time series from each channel was inspected for artifacts or 60 Hz line noise, and these channels were excluded from all subsequent analysis. The raw recorded voltage signal of the remaining channels were common average referenced and used for spectro-temporal analysis. For each (usable) channel, the time-varying analytic amplitude was extracted from eight bandpass filters (Gaussian filters, logarithmically increasing center frequencies [70-150 Hz] and semi-logarithmically increasing band-widths) with the Hilbert transform. The high-gamma($H\gamma$) activity was calculated by averaging the analytic amplitude across these eight bands. This signal was down-sampled to 200 Hz and z-scored relative to baseline activity for each channel.

978-1-5090-2809-2/17/$31.00 ©2017 IEEE

Independent components analysis

Independent components analysis (ICA) is a technique to identify the non-Gaussian and mutually independent sources that produced a signal [2]. In ICA models, each data sample x is expressed as a weighted linear combination of a set of basis vectors A with the weights s:

$$x^{(i)} = As^{(i)}, \tag{1}$$

where x is an n-dimensional signal vector, $A \in \mathbb{R}^{n \times k}$ is a set of k basis vectors embedded in the n-dimensional space, also known as the mixing matrix, and s is k-dimensional vector of coefficients or source values. When $k = n$, the exact recovery of s is possible since the mixing matrix can be inverted. The goal of ICA is to find the unmixing matrix W such that s can be recovered: $s = Wx$ with $W = A^{-1}$. An estimate of the unmixing matrix W can be obtained by solving the following constrained optimization problem:

$$\begin{aligned} \arg\min_{W} \quad & \tfrac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} ||s_j^{(i)}||_1 \\ \text{s.t.} \quad & WW^T = I \end{aligned} \tag{2}$$

where I is the identity matrix. The constraint $WW^T = I$ prevents the vector solutions from co-aligning, i.e., becoming identical, and limits the solution to the space of orthonormal matrices.

Sparse coding and convolutional sparse coding

Similar to ICA, sparse coding (SC) seeks to represent a dataset as a linear combination of a number of features or basis functions [3]. Contrary to ICA, in which the sparse code is generated by applying a linear transformation to the data, in SC the values of the coefficients s are obtained through inference. Since SC is not restricted to a linear transformation of the data, the prior belief on the distribution (i.e., sparsity) of the latent variables can be enforced more strictly. The SC models were generated by solving the following optimization problem:

$$\begin{aligned} \arg\min_{A,s} \quad & \tfrac{1}{2m} \sum_{i=1}^{m} ||x^{(i)} - As^{(i)}||_2^2 + \tfrac{\lambda}{m} \sum_{i=1}^{m} \sum_{j=1}^{k} ||s_j^{(i)}||_1 \\ \text{s.t.} \quad & ||A_j||_2 = 1 \quad \forall j = 1, \ldots, k. \end{aligned} \tag{3}$$

Here, λ controls the degree of sparsity and was determined by cross-validation (80:10:10, train-validate-test).

In convolutional sparse coding (CSC), a similar technique to SC, the data is expressed as a temporal convolution between a spatio-temporal basis A and a sparse code s, i.e., $\sum_{j}^{k} a_j * s_j^{(i)}$. Due to the presence of the convolution operation in the objective, CSC models are able to learn spatio-temporal features that have translation invariance in the time dimension. This is potentially useful, as it can account for differences in the temporal alignment between neural signals and behavioral outputs. The dimensionality of the CSC model was chosen to be 150 (which corresponds to representation that is approximately 1.75 times overcomplete) by using a ten-fold cross-validation procedure.

Classification analysis using SVMs

Using the sparse codes obtained with CSC, we trained linear support vector machines (SVMs) [6] to classify syllables in different speech related classification tasks.

The classification with multiple labels was implemented using a *one-against-rest* scheme. Classifiers were trained solving the following optimization problem:

$$\arg\min_{\theta} \ \frac{1}{2}||\theta||_2 + \frac{C}{m} \sum_{i=1}^{m} H[y^{(i)}(\theta x^{(i)} + b)] \tag{4}$$

where θ are the parameters of the classifier, b is the bias term, $x^{(i)}$ are the sparse code vectors, $y^{(i)}$ are the class labels, $H(z) = max(0, 1-z)^2$ is the squared hinge-loss and C is a weight on the error penalty. The hyper-parameter C was determined by cross-validation (80:10:10, train-validate-test).

Algorithm implementation

For ICA, we used the FastICA package in MATLAB, which implements the fast fixed-point algorithm for independent component analysis and projection pursuit (Homepage: http://www.cis.hut.fi/projects/ica/fastica/index.shtml).

For the CSC and SC analysis, we used the algorithm introduced in [4]. This implementation uses the orthant-wise l-BFGS method [7] for the inference of the coefficients and truncated projected gradient descent for the learning of the basis.

RESULTS

ICA extracts components that are reliably activated on single-trials

It is unknown whether sparse coding methods are suited to identify individual speech control signals from ECoG recordings of sensorimotor cortex. To investigate this possibility, we trained ICA and CSC models using ECoG recordings from vSMC that were collected while subjects uttered consonant-vowel (CV) syllables (see Methods).

Figures 1 and 2 show the magnitude of the sparse coefficients associated with three different ICA and CSC features (figure rows) across 40 trials of utterances /gi/, /fi/, and /zi/ (figure columns). The consonant production in each of these utterances is dominated by a different articulator: the lips are the predominant articulator during the production of /f/, and the coronal and dorsal parts of the tongue are the most dominant articulators involved in the generation of /z/ and /g/ respectively.

Each of the features shown in the Figures 1 and 2 has large coefficients in association with only one of the utterances, which suggests that these features are encoding signals controlling individual articulators.The presence of large coefficients is largely consistent across trials. Similar features can be found both in ICA and CSC models, as can be observed by visually comparing both figures. The coefficients associated with the CSC components are more sparse (have more values set to exactly zero), which is a non-linear effect induced by the regularization term which is part of the CSC objective (see Methods).

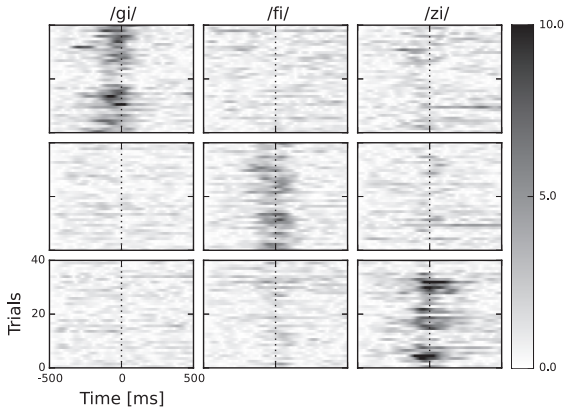

Fig. 1: Trial-by-trial activation patterns of three independent components during the production of syllables /gi/, /fi/ and /zi/.

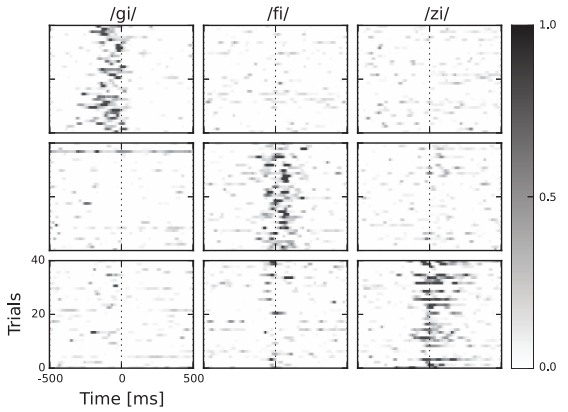

Fig. 2: Trial-by-trial activation patterns of three CSC components during the production of syllables /gi/, /fi/ and /zi/.

The evolution of the coefficients over time provides information about the temporal dynamics of the speech signals. In Figures 1 and 2, the coefficients are shown as a function of time (indicated in the x-axis) from 500 ms before the consonant-vowel transition (dashed line) to 500 ms after. Although there is some expected trial-by-trial variability, the timing of the increase in the magnitude of the coefficients is remarkably preserved across repetitions of the same utterance (y-axis).

Figure 3 shows the evolution of the coefficients over time (color saturation) during all CV utterances (derived from ICA on mean activity across trials) in the subspace spanned by the three independent components which showed the highest correlation with the dominant consonantal articulators, i.e., the lips, and the dorsal and coronal regions of the tongue (color coded). These trajectories are clearly dominated by one of the components. These results are inline with analysis of PCA components, but in contrast to those previous analyses, here the functional interpretation of each component is layed bare by the algorithm.

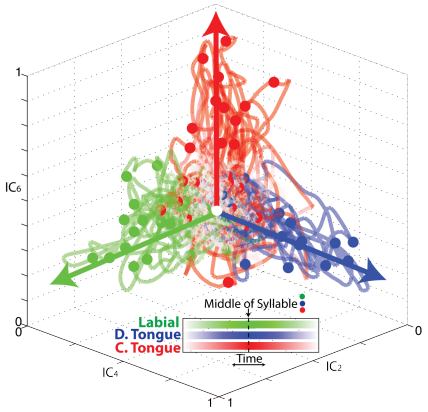

Fig. 3: Individual CV trajectories in the subspace spanned by the three independent components which showed the highest correlation with the dominant consonantal articulators (lips, dorsal tongue, and coronal tongue).

Speech control signals are associated with spatially sparse activation patterns

We next examined the spatial structure of the speech production signals extracted using ICA and CSC. Figure 4 shows four example activation patterns or features (these activation patterns correspond to the rows of the unmixing matrix W) learned using ICA. The color saturation indicates the degree to which the activity of the electrode contributes to that specific feature. The color itself indicates the sign of the feature element, which in ICA can be positive (black) or negative (red). The sign is relative to the sign of the coefficients and cannot be interpreted without taking into account the information contained in the coefficients. In these examples, three features correspond to activation patterns involving mostly one electrode and therefore are very spatiall sparse. Similar spatial stucture was observed with CSC basis, but space-time basis are more challenging to visualize.

Our results show that the structure of the learned features is in good agreement with the known functional anatomy of the sensorimotor cortex [1]. For instance, features whose coefficients are highly correlated with the production of labial utterances (e.g., /ba/, /fi/, etc) show high values in electrodes that are located in the anatomical regions of vSMC known to play an important role in the control of the lips. In summary the learned features using ICA (and CSC) correspond to electrode activation patterns that are spatially sparse and in good agreement with the functional anatomy of vSMC.

CSC components can be used to classify produced speech

In the field of computer vision, it has been shown that object classification algorithms can improve performance when trained using sparse high-dimensional feature representations of the data [8], [9]. To test whether classification of ECoG signals might also benefit from using high-dimensional and sparse representations, we trained linear Support Vector Machines (SVMs; see Methods) to classify sparse codes (generated with overcomplete $k > n$ CSC; see table I), in

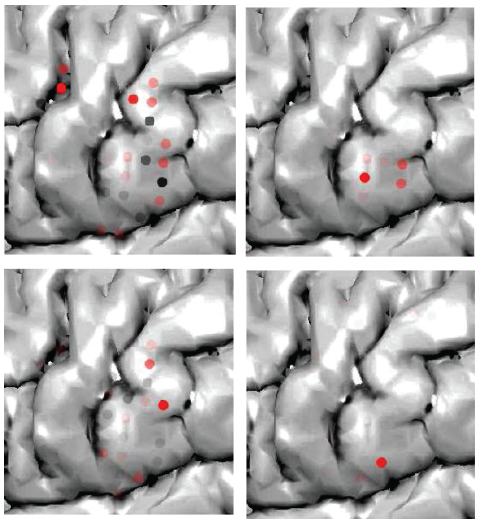

Fig. 4: Spatial filters learned with ICA. Colored circles indicate the location of the electrodes in the vSMC area.

four different classification tasks: (1) to identify the uttered syllable, (2) to identify the uttered consonant, and to identify the major oral articulator (e.g.: lips, tongue, etc) engaged during the production of (3) the consonant or (4) the vowel.

TABLE I: CSC-SVM classification accuracy.

Task	S1	S2	S3	Chance
Syllable	0.06	0.20	0.11	0.018
Consonant	0.14	0.34	0.24	0.053
Vowel	0.51	0.68	0.44	0.33
Articulator	0.58	0.80	0.62	0.33

Table I summarizes the results obtained for different classification tasks for three subjects (S1, S2, S3), as well as the chance levels for the tasks. For all subjects and tasks, classificaiton performance from SVMs trained on CSC were well above chance levels. While lower in absolute value, the syllable classificaiton task, which had 58 possible classes, had the best performance relative to chance (3.3-11.1x over chance) for all subjects. In contrast, the vowel classificaiton task was the most difficult, with performance ranging from 1.3x to 2.1x over chance performance. Overall, these classificatin results are in agreement with previous results showing that both the volume of the vowel state-space, and the state-space distances between vowels, is small compared to the major oral articulators involved in consonant production [1]. However, it is important to note that our previous results were derived from mean activity over trials. Furthermore, the improved performance of classifying syllables compated to the phoneme constiuents (relative to chance levels) suggests that the syllable is the more 'atomic unit', likely due to co-articulation [10].

CONCLUSIONS

Our results indicate that ICA or CSC are well suited to identify speech control signals in ECoG recordings from sensorimotor cortex. These methods offer an advantage over feature learning methods such as PCA in that they learn features that are not mixtures of different speech signals, and are hence easier to interpret. We find that the activation patterns associated with the features that are relevant for speech production are spatially sparse. Finally, the classification performance when using linear decoders was well above chance for all of the tasks and subjects examined here, and suggest that syllables may be a better targer for BMI applications. We further note that our classificaiton results could likely be improved by increaseing the sample size of the data set relative to the number of classes (which, because of the nature of collecting human data, is relatively modest). This improvement could be realized as enhanced performance in the convolutional sparse coding reconstructions and with the subsequent SVM classifications. We propose that the sparse/unmixed features learned by sparse coding methods may be beneficial to learning brain-machine interfaces by providing a more natural bases. Furthermore, sparse signals are more compressible, allowing for reduced overhead on data I/O and calculation.

ACKNOWLEDGMENTS

This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, the Lawrence Berkeley National Laboratory Laboratory-Directed Research and Development (LDRD) project 'Nano-NeuroTech for BRAIN', PI Peter Denes, and NSF award 1219212.

REFERENCES

[1] K. E. Bouchard, N. Mesgarani, K. Johnson, and E. F. Chang, "Functional organization of human sensorimotor cortex for speech articulation," *Nature*, vol. 495, no. 7441, pp. 327–332, 2013.

[2] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[3] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[4] A. Khosrowshahi, "The laminar organization of v1 neural activity in response to dynamic natural scenes," 2011.

[5] G. Agarwal, I. H. Stevenson, A. Berényi, K. Mizuseki, G. Buzsáki, and F. T. Sommer, "Spatially distributed local fields in the hippocampus encode rat position," *Science*, vol. 344, no. 6184, pp. 626–630, 2014.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[7] G. Andrew and J. Gao, "Scalable training of l 1-regularized log-linear models," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 33–40.

[8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[9] A. Szlam, K. Gregor, and Y. LeCun, "Fast approximations to structured sparse coding and applications to object classification," in *European Conference on Computer Vision*. Springer, 2012, pp. 200–213.

[10] K. E. Bouchard and E. F. Chang, "Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex," *The Journal of Neuroscience*, vol. 34, no. 38, pp. 12 662–12 677, 2014.