33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Decoding ensemble activity from neurophysiological recordings in the temporal cortex

Gabriel Kreiman

***Abstract*—We study subjects with pharmacologically intractable epilepsy who undergo semi-chronic implantation of electrodes for clinical purposes. We record physiological activity from tens to more than one hundred electrodes implanted in different parts of neocortex. These recordings provide higher spatial and temporal resolution than non-invasive measures of human brain activity. Here we discuss our efforts to develop hardware and algorithms to interact with the human brain by decoding ensemble activity in single trials. We focus our discussion on decoding visual information during a variety of visual object recognition tasks but the same technologies and algorithms can also be directly applied to other cognitive phenomena.**

I. Introduction

THERE has been significant progress over the last several years in the possibilities of interacting with human cortex through invasive methods. Here we describe our efforts to record and decode ensemble activity from a variety of different electrodes implanted in the human brain for clinical purposes. Subjects are patients who suffer from pharmacologically intractable epilepsy. In order to localize the seizure focus for potential resection, subjects are implanted with electrodes in a semi-chronic fashion for 7-10 days. The algorithms described here are based on electrophysiological recordings performed during this period (for reviews, see [3-5]).

Recording and decoding activity from the human brain at high spatial and temporal resolution has potential for a significant impact at different levels: (i) it can help us transform our understanding of the function of different human brain areas; (ii) it can significantly enhance clinical approaches that require direct interaction with the human brain including the burgeoning field of brain prosthetic applications and (iii) it offers the possibility of examining future technologies that may depend on brain-computer interfaces.

As a proof-of-principle, here we focus on decoding visual information in a variety of tasks that involve discriminating among different types of objects (Methods). The visual system offers a number of important features as a test bed to examine and characterize our methodology. First, we can rigorously, precisely and systematically control the input. Second, humans are quite proficient at visual discrimination tasks. Finally, we can compare and relate our findings to a large body of literature describing visual responses throughout the primate visual cortex (e.g. [1, 4, 6-9]).

This paper is organized as follows. First, we introduce the methodology for recording activity from the human brain and the basic algorithms for decoding brain activity. Second, we provide examples describing decoding performance, stationarity and other relevant metrics. Finally, we provide a summary and a look ahead.

II. Methods

A. *Recording invasive data from the human brain*

We record intracranial neurophysiological signals from the human brain. Subjects are patients that undergo surgical implantation of unilateral or bilateral depth electrodes and/or cortical surface electrodes for planning of resective surgery to treat epilepsy at Children's Hospital Harvard Medical School (CH) and Brigham and Women's Hospital (BW). Subjects stay in the hospital, with the electrodes implanted, for 7-10 days, until enough seizure events are accumulated to guide surgery [3, 10, 11]. Patient participation is voluntary; a consent form is signed after making sure that the subjects understand the procedures. The research efforts are based on the existing clinical recording procedures. All recordings are overseen by the attending physician according to the protocols currently in place and approved by the Institutional Review Boards at CH and BW. Given that these recordings are performed in epileptic subjects, it is interesting to consider the extent to which the results provide general insights about normal cognitive function. We make several remarks on this question: (1) Many electrodes (typically 40-140) are used because the epileptogenic regions are not known. Consequently, most of the electrodes (>90%) are located in non-epileptogenic areas (centimeters away or in the opposite hemisphere). (2) After the epileptogenic focus is found, we can compare the epileptogenic and non-epileptogenic electrodes. (3) Although some cognitive functions might be different in some epileptic patients, the consensus is that visual cognition is normal in most cases. (4) The study of epilepsy has provided key insights into brain function.

We record three different types of signals: multi-unit spiking activity (MUA) through microwires [4], local field potentials (LFPs) through microwires [12] and intracranial field potentials (IFPs) recorded through grid and strip electrodes [2]. We record spike activity through Ad-tech microwires (Pt/Ir, 40 μm diameter, ~1 MΩ impedance, Racine, WI) to monitor electrical activity at high spatial and

Manuscript submitted April 14, 2011. This work was supported in part by NSF (1010109 and 0954570), NIH (R21EY019710 and DP2OD005393) and USAF (FA8750-11-2-009)

G. Kreiman is with Swartz Center for Theoretical Neuroscience at Harvard University and Children's Hospital, Harvard Medical School, Boston, MA 02115 (phone: 617-919-2530; e-mail: gabriel.kreiman@tch.harvard.edu)

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

temporal resolution (~100 μm and <1 ms). Nine microwires (8 recording microwires + local reference) are inserted through the lumen of the main electrode. The signals are amplified and split into two streams: (i) We high-pass filter the signal (>600 Hz) to obtain MUA; (ii) We low-pass filter the signal (<100 Hz) to obtain local field potentials. We routinely compare MUA against single-unit activity (SUA) by using a spike-sorting algorithm. We use two different types of "macro/micro" electrodes from Ad-Tech (Racine, WI) to record local field potentials. In the depth-electrode cases, we use the electrodes described above to record both action potentials and LFPs. In non-depth cases, we use microwires attached to the standard clinical electrodes described below. These microwires enable us to record LFPs but not spike data. We record LFPs from tens to hundreds of microwires, depending on the total number of electrodes. We use grid and strip electrode to record intracranial field potentials (IFP) using Ad-Tech epilepsy electrodes (2 mm diameter, 1 cm separation, Pt) and amplified with a Bio-Logic system (Knoxville, TN) with 500 Hz sampling rate, a bandpass between 0.1-100 Hz and a notch at 60 Hz. We use a non-invasive infrared scanning system to monitor eye movements with a spatial resolution of ~1 deg and a temporal scanning frequency of 75 Hz (ISCAN DTL-300, Woburn, MA). We monitor eye movements continuously during the experiments.

The number and location of the electrodes are decided based on pre-surgical clinical evaluation. In the depth electrode cases, the number of electrodes is typically between 6 and 10 (each electrode has 9 microwires). In the intracranial grid electrode cases, the number of electrodes is typically between 40 and 140. In the hybrid macro/micro electrodes, there are 9 microwires per grid electrode). To localize the electrodes on the brain surface, we integrate the anatomical information provided by Magnetic Resonance Imaging (MRI), the spatial information of the electrodes provided by Computer Tomography (CT) and high-resolution images taken during electrode implantation. MR images are acquired before electrode implantation and CT images are acquired ~1 day after implantation. The 3D brain surface is reconstructed for each subject. An automatic parcellation is performed using Freesurfer [13]. A preliminary co-registration of CT and MRI is implemented using SPM [14]. Particular care must be taken in the co-registration because the brain may swell, shifting the CT and MR images in non-trivial ways. We often find that the effect of swelling is particularly severe (up to ~2 mm) only for a few of the tens of electrodes implanted. Because of these potential distortions, the co-registration is fine-tuned using Freesurfer in a manually intensive effort. Finally, we record the Talairach coordinates of each electrode. Our analyses are restricted to the non-epileptogenic locations.

B. Decoding information from electrophysiological recordings

Consider a situation where images containing different objects were presented to the subject and we are interested in decoding the image content based on the ensemble activity. Let ${}_is_r(T)$ denote the response of electrode i during time interval T in repetition r (i=1,..,N_e where N_e is the number of electrodes and r=1,…,R where R is the number of repetitions). We routinely compare different possible definitions for ${}_is_r(T)$. How ${}_is_r(T)$ defined is an interesting question in itself, i.e., what aspects of the neural signals carry relevant information about the stimulus. For simplicity, for LFPs and IFPs, here we define s as the signal power in the time interval T. We also separately examine the signal power in different frequency bands: 0.1-10 Hz, 10-20 Hz, 21-55 Hz, 65-100 Hz. For the spike trains, we define s as the spike count within T. We perform spike sorting on the MUA recordings to obtain single-unit activity (SUA) and separately analyze the MUA and SUA responses. The analysis interval $T=[t_b;t_a]$ (with respect to stimulus onset) provides information about the temporal evolution of the neurophysiological responses. We use a sliding window to characterize the responses as a function of time. $\mathbf{s}$ can also be a vector that includes a combination of multiple features extracted from the response waveform. We consider the following possible features: voltage at each time point; average voltage in a window of size τ; projection onto the first principal components accounting for 80% of the variance, time of minimum voltage, time of maximum voltage, response amplitude; projection onto the first independent components, spike counts, interspike interval code. To shorten the discussion, we refer to $\mathbf{s}$ as the "response".

The same decoding algorithms are applied to LFPs, IFPs and spike trains. The ensemble analyses provide a natural way of thinking about the computations that neurons must solve. The ensemble analyses also allow a rigorous quantitative comparison among alternative neural codes. For a population of N_e electrodes, as a first approximation we assume independence and construct a population vector by concatenating the responses of each electrode: $\mathbf{p}_r(T)=[\,{}_1\mathbf{s}_r(T), {}_2\mathbf{s}_r(T),\ldots, {}_{N_e}\mathbf{s}_r(T)]$. The dimensionality of this vector is N_e times the number of features per electrode. The input to the classifier consists of the vectors $\mathbf{p}_r(T)$. As a control, we compare the results against the classification performance (CP) obtained during the baseline interval T=[-200;0] ms (before stimulus onset there should be no information about the objects due to the random order). Each trial is associated with a label (λ_r) that indicates the category or exemplar of the object shown in trial r. In a binary classifier (e.g. when comparing one category versus the rest), λ_r takes the values "+1" or "-1". In the case of multiclass classifiers, λ_r indicates the object category or exemplar. We randomly subsample the data to have the same number of repetitions with "+1" and "-1" labels. In an abuse of notation, we sometimes refer to "training a classifier with image X and evaluating the CP with image Y". What we mean is: training the classifier with the neural ensemble responses obtained upon presenting image X and labels corresponding to image X and evaluating

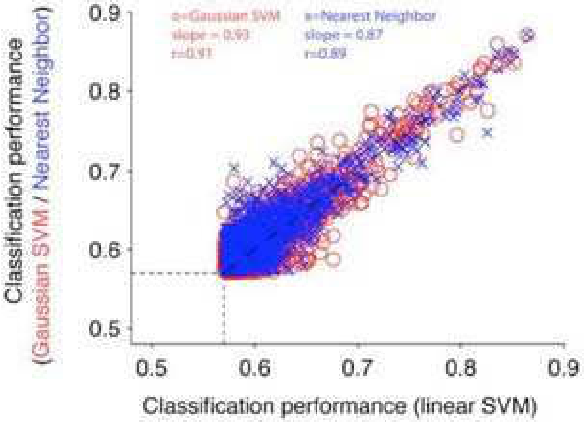

Fig 1. **Comparison among different statistical classifiers and selectivity criteria.** Comparison of classification performance levels obtained using different statistical classifiers for individual electrodes (based on the data in [2]. Throughout the text, we report the performance of a Support Vector Machine (SVM) classifier with a linear kernel. Here we show a direct comparison between the linear SVM (x-axis) and an SVM classifier with a Gaussian kernel (red) or a nearest neighbor classifier (blue). The diagonal dashed line would correspond to identical performance across classifiers. We compare the values only for those electrodes and conditions that yielded a performance above 3 standard deviations of the null hypothesis defined by shuffling the object labels.

the CP using the neural ensemble responses obtained upon presenting image Y and labels corresponding to image Y. A binary support vector machine (SVM) classifier is trained to discriminate between the "+1" and "-1" examples, that is, to quantify whether the neural ensemble activity can distinguish between the two conditions [1, 2, 15]. In the case of multiclass problems we use a one-versus-all approach. By default, we use a linear kernel where the classifier boundary can be expressed as ***w.p*** where the weights ***w*** are learnt during training. We routinely test SVMs with more complex kernel functions including polynomial kernels and radial basis function kernels and we also compare different statistical classifiers [16-18]. Importantly, the CP is always evaluated with independent data that are not seen by the classifier during training. We report the classification performance (CP) as the proportion of test trials that are correctly classified. CP ranges from 0 (no trial correctly labeled) to 1 (all trials correctly labeled). The chance level is 1 divided by the number of possible trial labels (e.g. 0.5 in the case of binary classifiers). To evaluate the statistical significance of the CP values obtained, we define a null hypothesis where we randomly shuffle the trial labels [19]. This procedure is repeated 10000 times and we compute the *p* value by comparing the actual CP against the distribution of CP values based on this null hypothesis.

III. RESULTS

A. Decoding performance

We compared different machine learning algorithms to decode visual information in single trials from an experiment in which subjects had to categorize images from 5 different natural categories (Fig. 1). Overall, the results show concordance across different machine learning algorithms, highlighting the high information content in the recordings. We built a pseudopopulation concatenating recordings in different patients to extrapolate to the performance that one might expect to obtain in large-scale recordings across multiple electrodes (Fig. 2). This extrapolation revealed that even from coarse intracranial field potential recordings, we can achieve significant classification performance levels in single trials in complex object categorization both for binary classification (Fig. 2A) as well as multiclass classification (Fig. 2B).

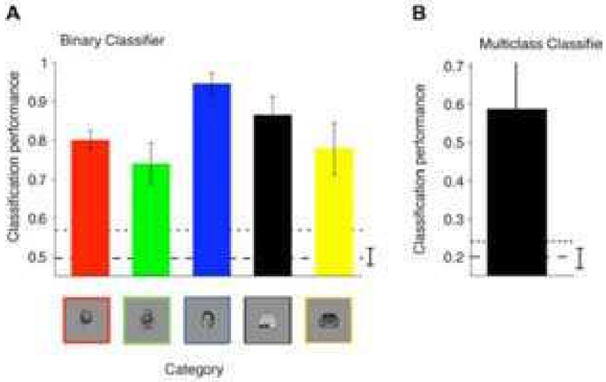

Fig 2. **Pseudo-population analysis showing classification performance using an ensemble of 11 electrodes**. For each subject, one electrode was chosen based on the rank of *rv* values. *rv* is the ratio of the variance across categories divided by the variance within categories [1]. rv was computed using only the training data. **A**. Binary classification performance. The colors correspond to different object categories. The horizontal dashed lines denote the chance performance value of 0.5 and the significance threshold value. Next to the chance level line we show the range of classification performance values obtained after randomly shuffling the object category labels (100 iterations). **B**. Multiclass classification performance. Here the chance level is 0.2 (5 object categories). The data for this figure are taken from ref. [2].

B. Tolerance to object transformations

A key challenge in decoding approaches involves not only being able to extract information but also assessing how tolerant the decoding procedure is to different sources of variation. In the particular context of visual recognition, it is not sufficient to show selectivity to object exemplars or object categories; it is critical to consider how well the algorithms extrapolate across different transformations in the image. We have shown that the single trial decoding procedure illustrated here can extrapolate across transformations in object scale and rotation [2], image clutter [15] and object occlusion [20].

C. Stationarity of the responses

Another important aspect of decoding information from human the human brain concerns the long-term stability of the recordings. In Fig. 3A we show the responses of an example electrode across sessions, illustrating the degree of stationarity in the visually selective responses over scales of >24 hrs. We compared decoding performance *within* sessions to decoding performance *across* sessions and observed that there was a significant correlation. These results emphasize that an algorithm to extract information from neurophysiological recordings can show a strong degree of stationarity.

IV. Summary

Recent developments in hardware, algorithms and computational resources make it possible to decode information from the activity of ensembles in single trials and even in real time. Here we provide proof-of-principle evidence for the possibility of decoding information focusing on the visual system and describing the performance of machine learning algorithms, their tolerance to stimulus transformations and their temporal stability. Neurotechnological advances make it possible to interact with the human brain at unprecedented spatial and temporal resolution. Invasive recordings from the human brain offer many clinical, scientific and engineering opportunities.

Acknowledgment

All the recordings are based on collaborative efforts with the following neurosurgeons: Itzhak Fried, Joseph R. Madsen, Alex Golby and Stanley Anderson. We acknowledge help from the clinical team at UCLA, Children's Hospital, Brigham and Women's Hospital and Johns Hopkins Medical School, without whom none of these recordings would be feasible. The recordings and analyses described here were performed by Yigal Agam, Arjun Bansal, Calin Buia, Hesheng Liu, Radhika Madhavan, Daniel Millman and Jedediah Singer.

References

[1] 1. Hung, C., et al., *Fast Read-out of Object Identity from Macaque Inferior Temporal Cortex.* Science, 2005. **310**: p. 863-866.

[2] 2. Liu, H., et al., *Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex.* Neuron, 2009. **62**(2): p. 281-290.

[3] 3. Engel, A.K., et al., *Invasive recordings from the human brain: clinical insights and beyond.* Nat Rev Neurosci, 2005. **6**(1): p. 35-47.

[4] 4. Kreiman, G., *Single neuron approaches to human vision and memories.* Current Opinion in Neurobiology, 2007. **17**(4): p. 471-475.

[5] 5. Fried, I., et al., *Cerebral microdialysis combined with single-neuron and electroencephalographic recording in neurosurgical patients.* Journal of Neurosurgery, 1999. **91**: p. 697-705.

[6] 6. Felleman, D.J. and D.C. Van Essen, *Distributed hierarchical processing in the primate cerebral cortex.* Cerebral Cortex, 1991. **1**: p. 1-47.

[7] 7. Logothetis, N.K. and D.L. Sheinberg, *Visual object recognition.* Annual Review of Neuroscience, 1996. **19**: p. 577-621.

[8] 8. Tanaka, K., *Inferotemporal cortex and object vision.* Annual Review of Neuroscience, 1996. **19**: p. 109-139.

[9] 9. Connor, C.E., S.L. Brincat, and A. Pasupathy, *Transformation of shape information in the ventral pathway.* Curr Opin Neurobiol, 2007. **17**(2): p. 140-7.

[10] 10. Engel, J., *Surgery for seizures.* New England Journal of Medicine, 1996. **334**(10): p. 647-652.

[11] 11. Ojemann, G.A., *Treatment of temporal lobe epilepsy.* Annual Review of Medicine, 1997. **48**: p. 317-328.

[12] 12. Kreiman, G., et al., *Object selectivity of local field potentials and spikes in the inferior temporal cortex of macaque monkeys.* Neuron, 2006. **49**: p. 433-445.

[13] 13. Dale, A.M., B. Fischl, and M.I. Sereno, *Cortical surface-based analysis. I. Segmentation and surface reconstruction.* Neuroimage, 1999. **9**(2): p. 179-94.

[14] 14. Ashburner, J. and K. Friston, *Multimodal image coregistration and partitioning--a unified framework.* Neuroimage, 1997. **6**(3): p. 209-17.

[15] 15. Agam, Y., et al., *Robust selectivity to two-object images in human visual cortex.* Current Biology, 2010. **20**: p. 872-879.

[16] 16. Bishop, C.M., *Neural Networks for Pattern Recognition.* 1995, Oxford: Clarendon Press.

[17] 17. Vapnik, V., *The Nature of Statistical Learning Theory.* 1995, New York: Springer.

[18] 18. Poggio, T. and S. Smale, *The mathematics of learning: dealing with data.* Notices of the AMS, 2003. **50**: p. 537-544.

[19] 19. Efron, B. and R.J. Tibshirani, *An Introduction to the Bootstrap.* 1993, London: Chapman & Hall/CRC.

[20] 20. Kreiman, G., et al. *Top-down signals are needed for object completion in the human visual cortex.* in *Visual Science Society.* 2011. Naples: VSS.

[21]

[1]

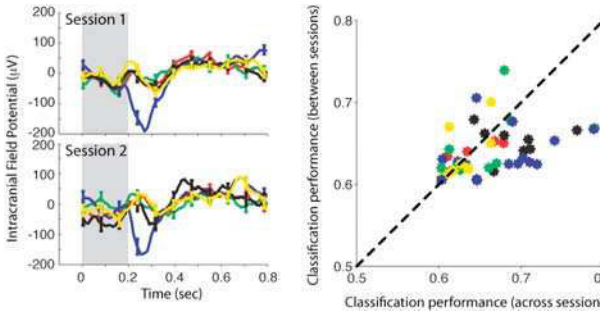

Fig 3. **Stationarity across recording sessions**. On the left we show an example intracranial field potential recording from two different sessions separated by more than 24 hours. Each color represents the average IFP (error bars denote SEM) across 5 different exemplar objects belonging to a given object category (each color represents a separate object category). The gray rectangle denotes the image presentation interval. The similarity in the responses across sessions highlights the stability of the recordings over time. On the right, we compare the output of the decoder within session (x-axis) to the output of the decoder across sessions (y-axis). Each color represents a different object category. Data taken from [2].