NeuroImage 148 (2017) 318–329

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Semantic attributes are encoded in human electrocorticographic signals during visual object recognition

Kyle Rupp[a], Matthew Roos[b], Griffin Milsap[a], Carlos Caceres[b], Christopher Ratto[b], Mark Chevillet[b], Nathan E. Crone[c], Michael Wolmetz[b,*]

[a] Department of Biomedical Engineering, Johns Hopkins University, 720 Rutland Ave., Baltimore, MD 21205, USA
[b] The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Rd., Laurel, MD 20723, USA
[c] Department of Neurology, Johns Hopkins University, 600 N. Wolfe St., Meyer 2-161, Baltimore, MD 21287, USA

ARTICLE INFO

Keywords:
Semantics
Electrocorticography
High-gamma activity
Encoding models
Object recognition

ABSTRACT

Non-invasive neuroimaging studies have shown that semantic category and attribute information are encoded in neural population activity. Electrocorticography (ECoG) offers several advantages over non-invasive approaches, but the degree to which semantic attribute information is encoded in ECoG responses is not known. We recorded ECoG while patients named objects from 12 semantic categories and then trained high-dimensional encoding models to map semantic attributes to spectral-temporal features of the task-related neural responses. Using these semantic attribute encoding models, untrained objects were decoded with accuracies comparable to whole-brain functional Magnetic Resonance Imaging (fMRI), and we observed that high-gamma activity (70–110 Hz) at basal occipitotemporal electrodes was associated with specific semantic dimensions (manmade-animate, canonically large-small, and places-tools). Individual patient results were in close agreement with reports from other imaging modalities on the time course and functional organization of semantic processing along the ventral visual pathway during object recognition. The semantic attribute encoding model approach is critical for decoding objects absent from a training set, as well as for studying complex semantic encodings without artificially restricting stimuli to a small number of semantic categories.

1. Introduction

The view that objects are encoded according to their semantic attributes or features, while not new, has become quite practical. Under an attribute-based view, a concept can be encoded over a large set of meaningful attributes, with each attribute assigned a value or set of values related to its probability, weight, or importance (Rosch, 1978). For example, the encoding of the concept "bird" assigns high probabilities to attributes typical of birds (has beak, flies, etc.) and low or zero probabilities to attributes atypical of birds (has four legs, manmade, etc). Substantial work has been done to catalogue the attributes and weights associated with different concepts, and attribute ratings can account for a host of human judgments about the relationships between concepts and the organization of categories (Binder et al., 2016; Cree and McRae, 2003; Garrard et al., 2001; Ruts et al., 2004). In related work on vector space models of semantics, automated methods can be used in place of human annotators to learn latent semantic features from the statistical properties of words and phrases in large text corpora (Deerwester et al., 1990; Mikolov et al., 2013; Pennington et al., 2014), and these latent features are similarly useful in accounting for human judgments (Pereira et al., 2016).

Efforts to decompose concepts into their constituent attributes or features have been used to great effect in the study of knowledge representation in the human brain. Following methods pioneered by Mitchell et al. (2008) to learn relationships between individual semantic features and the neural activity patterns they evoke, subjects perform tasks that require semantic processing – viewing or naming objects (Clarke et al., 2014), reading words or sentences (Wehbe et al., 2014), considering semantic attributes (Sudre et al., 2012), generating category exemplars (Simanova et al., 2015), watching movies (Huth et al., 2012), or listening to stories (Huth et al., 2016) – while neural responses are recorded with functional magnetic resonance imaging (fMRI) or magnetoencephalography (MEG). Because stimuli can be represented in terms of their constituent semantic attributes or features, a mapping can be learned between each semantic feature and its associated neural responses (i.e. voxel intensities, MEG sensor

Abbreviations: (BOLD), Blood Oxygen Level Dependent; (ECoG), Electrocorticography; (fMRI), functional Magnetic Resonance Imaging; (MEG), magnetoencephalography; (MRA), mean rank accuracy; (RSA), Representational Similarity Analysis

* Corresponding author.
E-mail address: mikew@jhu.edu (M. Wolmetz).

http://dx.doi.org/10.1016/j.neuroimage.2016.12.074
Received 2 August 2016; Accepted 26 December 2016
Available online 11 January 2017
1053-8119/

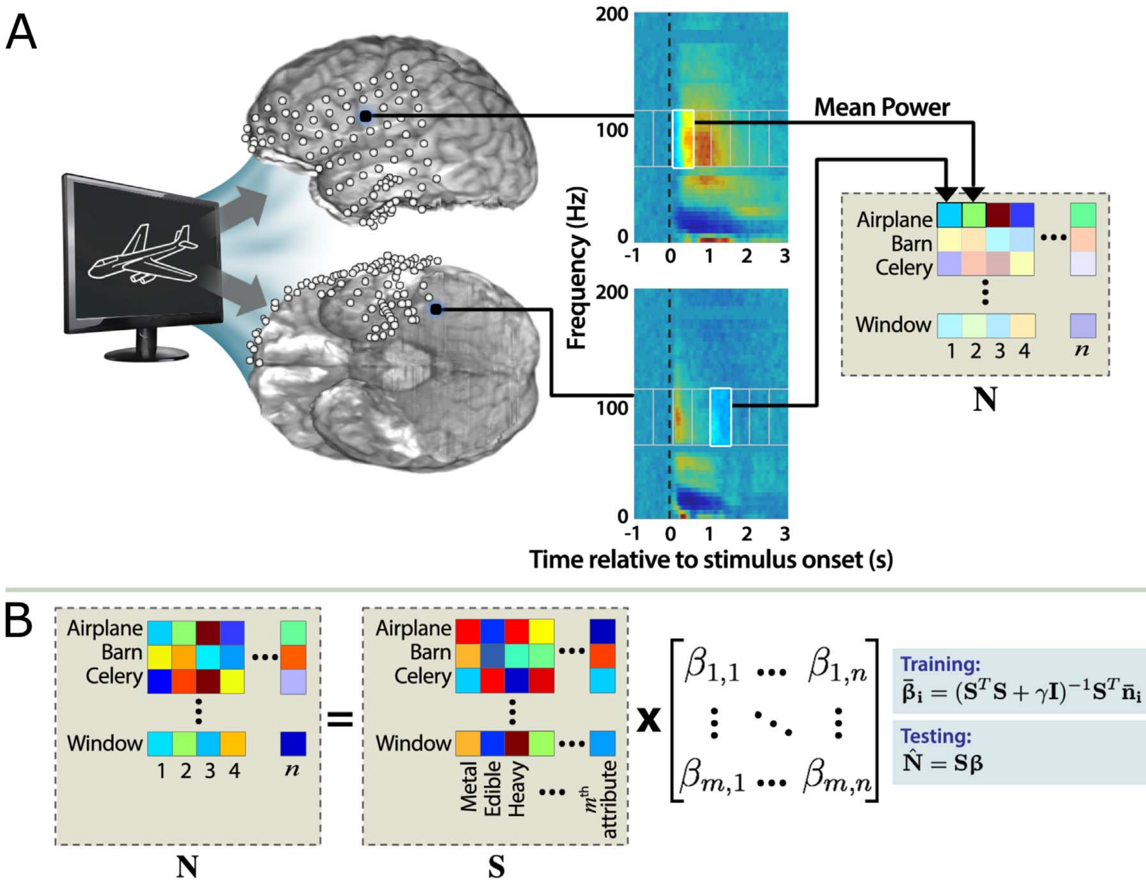

Fig. 1. Training and testing encoding models from ECoG. (A) Patients named objects and spectral estimation was performed on their neural signals to produce mean power over a variety of frequency bands and temporal windows (only high-gamma shown here). A subset of neural features (particular frequency bands at particular time windows at particular electrodes) was selected for use in the encoding model. (B) Linear ridge regression was used to learn a neural encoding model $\boldsymbol{\beta}$, which maps from semantic attribute ratings **S** to neural feature values **N**. To decode a new neural activity pattern $\bar{\mathbf{n}}$ generated by an untrained object, $\bar{\mathbf{n}}$ is compared via cosine distance to a set of predicted neural activity patterns generated by applying $\boldsymbol{\beta}$ to a catalogue of possible objects and their semantic attributes.

amplitudes), typically through linear regression. These encoding models project semantic features into a neural feature space, and similarly, decoding models can be used to project recorded neural activity patterns into a semantic feature space.

The resulting neurosemantic models have provided new insights into conceptual knowledge representation in the mind and brain. The fact that neurosemantic models can be used to successfully learn mappings between semantic and neural features suggests that the brain's representation of objects involves decomposition into semantic features. This paradigm also provides a framework for testing theories about what specific semantic features are represented in the human brain (Just et al., 2010), how they are encoded in neural activity (Huth et al., 2016), and how cognitive processes modulate neurosemantic representations (Çukur et al., 2013). Likewise, from a decoding perspective, decompositional neurosemantic models are very powerful in that they can interpret neural activity from concept classes they have not been trained on in a process termed zero-shot learning (Palatucci et al., 2009).

The impact of this approach, though, is limited by the quality and quantity of available neural data. Non-invasive neuroimaging methods are subject to lower signal-to-noise ratios, trade-offs between temporal and spatial resolution, and indirect estimates of neural activity. Invasive alternatives like electrocorticography (ECoG) can only be used in the relatively rare clinical setting when implanting electrodes on the surface of the cortex is a clinical necessity. As a result, spatial coverage is determined solely by clinical considerations, which leads to varied anatomical sampling across patients. At the same time, ECoG offers high temporal resolution, a high signal-to-noise ratio due to direct contact between electrodes and the cortical surface, and more direct observations of neural processing. Evidence of this can be found in studies showing that ECoG responses correlate well with spiking activity (Manning et al., 2009; Ray et al., 2008) and hemodynamic responses (Logothetis et al., 2001; Niessing et al., 2005), with activity in high-gamma frequencies (e.g. 70–110 Hz) serving as a particularly good index of underlying neural processing.

Despite the potential advantages, there have been relatively few studies of semantic attribute representation using ECoG. The few attempts to use ECoG for semantic decoding have relied on discriminative approaches over a small number of trained classes or categories (Liu et al., 2009; Wang et al., 2011). In one of the only published examples of semantic decoding from ECoG, Wang et al. (2011) asked patients to perform several different tasks that activated representations of semantic properties (e.g. visual object naming), and then trained Support Vector Machine (SVM) and Gaussian Naïve Bayes (GNB) classifiers to decode the evoked responses to one of the three possible target categories (i.e. foods, tools, and body parts). Performance varied across subjects, tasks, and classifier types, with mean classification rates of approximately 56% correct and a range from approximately 40% to 74% (where 33% is chance), indicating that substantial category information can be extracted from ECoG. While encouraging, conclusions drawn from a very restricted number of classes (e.g. foods, tools, and body parts) or dimensions of variation (e.g. living vs. non-living, large vs. small) may be partially confounded by expectation and perceptual set effects that cause subjects to artificially attend to and process these dimensions.