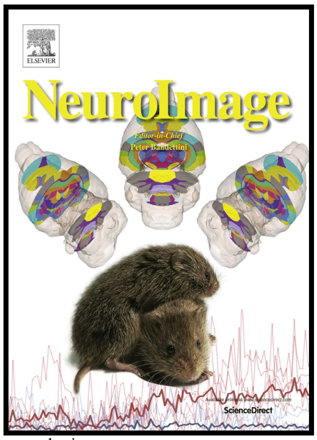
Author's Accepted Manuscript



www.elsevier.com

Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity

Jacek P. Dmochowski, Jason Ki, Paul DeGuzman, Paul Sajda, Lucas C. Parra

PII: S1053-8119(17)30429-9
DOI: http://dx.doi.org/10.1016/j.neuroimage.2017.05.037
Reference: YNIMG14049

To appear in: *NeuroImage*
Accepted date: 17 May 2017

Cite this article as: Jacek P. Dmochowski, Jason Ki, Paul DeGuzman, Paul Sajda and Lucas C. Parra, Extracting multidimensional stimulus-response correlation using hybrid encoding-decoding of neural activity, *NeuroImage* http://dx.doi.org/10.1016/j.neuroimage.2017.05.037

Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity

Jacek P. Dmochowski[a], Jason Ki[a], Paul DeGuzman[b], Paul Sajda[c], and Lucas C. Parra[a]

[a]Department of Biomedical Engineering, City College of New York, New York, NY, 10031
[b]Neuromatters LLC, New York, NY 10038
[c]Department of Biomedical Engineering, Columbia University, New York, NY, 10027

Abstract

In neuroscience, stimulus-response relationships have traditionally been analyzed using either encoding or decoding models. Here we propose a hybrid approach that decomposes neural activity into multiple components, each representing a portion of the stimulus. The technique is implemented via canonical correlation analysis (CCA) by temporally filtering the stimulus (encoding) and spatially filtering the neural responses (decoding) such that the resulting components are maximally correlated. In contrast to existing methods, this approach recovers multiple correlated stimulus-response pairs, and thus affords a richer, multidimensional analysis of neural representations. We first validated the technique's ability to recover multiple stimulus-driven components using electroencephalographic (EEG) data simulated with a finite element model of the head. We then applied the technique to real EEG responses to auditory and audiovisual narratives experienced identically across subjects, as well as uniquely experienced video game play. During narratives, both auditory and visual stimulus-response correlations (SRC) were modulated by attention and tracked inter-subject correlations. During video game play, SRC varied with game difficulty and the presence of a dual task. Interestingly, the strongest component extracted for visual and auditory features of film clips had nearly identical spatial distributions, suggesting that the predominant encephalographic response to naturalistic stimuli is supramodal. The diversity of these findings demonstrates the utility of measuring multidimensional SRC via hybrid encoding-decoding.

Introduction

Understanding the relationship between a sensory stimulus and the resulting neural response is a fundamental goal of neuroscience. Two distinct paradigms have shaped the pursuit of the neural code. The *encoding* approach attempts to explain neural responses from features of the stimulus, typically via linear filtering [1]. Examples include receptive fields and spike-triggered averages in single-unit electrophysiology [1], the generalized linear model (GLM) in functional magnetic resonance imaging (fMRI) [2, 3], spectrotemporal response functions (STRF) in electrocorticograms [4], and temporal response functions in encephalographic recordings [5, 6, 7]. In contrast to encoding, the *decoding* approach is to predict the stimulus by filtering over an array of neural responses. Decoding techniques have been shown to reconstruct experienced stimuli in a large number of findings spanning animal [8, 9, 10, 11] and human investigations of both visual [12, 13, 14, 15, 16, 17] and auditory stimuli [18, 19, 20, 21, 22].

The encoding and decoding approaches possess contrasting strengths and weaknesses: whereas encoding models operate on the stimulus and are thus easily interpretable [23], they generally predict the responses of individual data channels (i.e, neurons, voxels, or electrodes) and do not efficiently recover distributed neural representations. Decoding techniques filter neural activity over multiple channels and are therefore naturally suited to capturing distributed representations, but at the expense of models that are often difficult to interpret and prone to overfitting. Therefore, an approach that efficiently captures distributed neural representations and is readily interpretable in the stimulus space is lacking.

Here we propose a hybrid approach that combines the strengths of encoding and decoding. The technique integrates neural responses across space while filtering the stimulus in time, i.e. it "decodes" neural activity to recover an "encoded" version of the stimulus. By jointly learning decoding and encoding models, distributed neural representations are identified and explicitly linked to portions of the stimulus. In contrast to existing paradigms, this approach decomposes the neural representation of stimuli into multiple dimensions, with each dimension defined by a (spatial) response component and a (temporal) stimulus component.

To validate the ability of the proposed technique to recover multiple simultaneous stimulus-driven components, we first conducted a simulation study using data generated from a finite element model (FEM) of the head. The recovered components matched the ground-truth activations in both spatial topography and time course. We then evaluated the technique on recordings of neural activity in response to various naturalistic audiovisual stimuli and found multiple significant dimensions of stimulus-response correlation (SRC) for both auditory and visual features. These multiple dimensions were modulated by the attentional state of the observer. Interestingly, we found that independent visual and auditory features possessed a common response component, suggesting that the dominant EEG