A Brief Introduction to the Forward Encoding Model

Presented by Qing Yu
05/30/2014
Outline

• Introduction
• Example (Brouwer & Heeger, 2011)
• Methodology
• Application
• Discussion
Introduction

Encoding and Decoding in fMRI (Naselaris et al., 2011)

The brain can be viewed as a system that nonlinearly maps stimuli into brain activity. According to this perspective, a central task of systems and cognitive neuroscience is to discover the nonlinear mapping between input and activity. The relationship between encoding and decoding can be described in terms of a series of abstract spaces. In experiments using visual stimuli, the axes of the input space are the luminance of pixels, and each point in the space represents a different image. Brain activity measured in each voxel is represented by an activity space. The axes of the activity space correspond to the activity of different voxels, and each point in the space represents a unique pattern of activity across voxels. In between the input and activity spaces is a feature space. The mapping between the input space and the feature space is nonlinear, and the mapping between the feature space and activity space is linear. This linear mapping is called linearizing, because the nonlinear mapping into feature space linearizes the relationship between the stimulus and the response (Wu et al., 2006). Encoding models based on linearizing feature spaces are referred to as linearizing encoding models. Linearizing encoding models have a simple interpretation and are relatively easy to estimate. The mapping between the input space and the feature space is assumed to be nonlinear because most of the interesting computations performed by the brain are nonlinear. The mapping between feature space and activity space is assumed to be linear because the features that are represented by an ROI should have the simplest possible relationship to its activity. The nonlinear mapping is the same for each voxel; only the linear mapping has to be estimated from measured voxel activity. Thus, linearizing encoding models require only linear estimation. This can be performed by readily available algorithms for linear regression (Wu et al., 2006).

Once estimated, the linear mapping between feature space and activity space describes the particular mix of features that evoke activity in each voxel. As far as we know all of the encoding models that have been published in the field of fMRI thus far make use of a linearizing feature space. That is, they assume that there is a nonlinear mapping from the stimulus space to the feature space, and a linear mapping between the feature space and the activity space. We have already discussed the study of Kay et al. (2008) in detail. A subsequent study by Naselaris et al. (2009) reanalyzed the data collected as part of the Kay et al. study. However, Naselaris et al. constructed two different models for each voxel: a model based on phase-invariant Gabor wavelets, and a semantic model that was based on a scene category label for each natural scene. Naselaris et al. showed that the Gabor wavelet and semantic models performed similarly, and that the best model for each voxel was determined by cross-validation. The Gabor wavelet model was more flexible and could fit the data better, but the semantic model was easier to interpret and more generalizable.
Introduction

Encoding Model: general approach (Naselaris et al., 2011)

(1) Collect data and divide into training and validation data sets

(2) Use the training data to estimate one or more encoding models for each voxel

(3) Apply the estimated encoding models to the validation data and evaluate prediction accuracy
Introduction

The forward encoding model (Brouwer & Heeger, 2009; 2011)

- A small number of hypothesized feature-selective channels in each voxel
- Compute the weights for each hypothesized channel
- Reconstruct population-level orientation tuning responses in each brain region using the weights
Cross-orientation suppression in human visual cortex

Gijs Joost Brouwer and David J. Heeger
Department of Psychology and Center for Neural Science, New York University, New York, New York
Stimuli & Procedures

A Weight estimation experiment
- 120°
- Stimulus (1.5s)
- ISI (3 - 6s)

B Contrast suppression experiment
- Target only 25% contrast
- Target + Mask 50% contrast
- Stimulus (1.5s)
- ISI (3 - 6s)

C Control experiment

Example (Brouwer & Heeger, 2011)
Results

There was little, if any, interaction between conditions, such that the presence of the mask did not lead to an observable suppression of the mean responses to the target. If it had, the two contrast-response functions should have converged at high contrasts. To test this formally, we fitted the mean responses of the main cross-orientation suppression experiment and the control experiment to the normalization model with \((\text{Eq. 7})\) and without \((\text{Eq. 8})\). Both models used the \(a\) and \(n\) parameters obtained in fitting the channels separately, allowing only \(r_{\text{max}}\) and \(b\) to vary. For the cross-orientation suppression experiment, the model without cross-orientation suppression provided a better fit to the mean responses (mean \(r^2 = 0.84\)) than the model with cross-orientation suppression (mean \(r^2 = 0.71\)). Cross-validation of the fits confirmed that this difference was statistically significant \((P < 0.01)\). This is in agreement with the observation that there appears to be little interaction between the conditions, as shown in Fig. 9A. A small but significant \((P < 0.05)\) difference was found for the fits to the mean responses of the control experiment (normalization model including suppression: mean \(r^2 = 0.94\); without suppression: mean \(r^2 = 0.98\)).

DISCUSSION

We used fMRI, in combination with a forward modeling analysis (Brouwer and Heeger 2009; Kay et al. 2008), to measure cross-orientation suppression in human primary visual cortex (V1) and test the normalization model. We found that for the channel tuned to the orientation of the target grating, responses to the target grating were suppressed when a second, orthogonal mask grating was added. The remaining channels, tuned to either intermediate orientations or the mask orientation, also showed clear evidence of suppression. We found this suppression to be implemented as a change in the contrast gain of the channel responses. When the target and mask were temporally interleaved, no suppression was observed.
Methodology

- **Training dataset:** $B_1$ (Weight estimation experiment)
- **Test dataset:** $B_2$ (contrast suppression & control experiments)
- **Hypothesized channel responses for $B_1$:** $C_1$
- **Weight matrix:** $W$
- **Actual channel responses for $B_2$:** $C_2$

- **Step 1:** $B_1 = WC_1$
- **Step 2:** $\hat{W} = B_1C_1^T(C_1C_1^T)^{-1}$
- **Step 3:** $\hat{C}_2 = (\hat{W}^T\hat{W})^{-1}\hat{W}^TB_2$
Application

- **Orientation** (Anderson et al., 2013; Brouwer & Heeger, 2011; Garcia et al., 2013; Ho et al., 2012; Scolari et al., 2012)
- **Color** (Brouwer & Heeger, 2009; 2013)
- **Motion direction** (Kok et al., 2013)
- **Spatial location** (Sprague & Serences, 2013)
Discussion

• Circular feature space?

• Separate training dataset?

• Technical issues:
  • Preprocessing steps
  • Raw data or beta weights?
  • Voxel selection?
  • ......
Thanks!