Exploring Inverse Encoding Models for decoding of visual information from fMRI data

Wolfram Höps
Supervisors: Kai Görgen, Prof. Dr. John-Dylan Haynes

January 25, 2017

Abstract

Translating brain activity into information about a person's experience has long been a goal in neuroscience. Indeed, noninvasive human neuroimaging has triggered remarkable success in brain decoding, especially in the visual domain. In this short study, an Inverse Encoding Model is used to reconstruct checkerboard stimuli observed by a participant in an fMRI experiment. The study suggests that inverse encoding is a valid approach for this task, producing informative decoding results. The positions of decoding sites furthermore allow insights into the spatial organization of the primary visual cortex, where we can confirm the concept of retinotopic mapping.

1 Introduction

Conscious experience relies on neural encoding and decoding of sensory stimuli. If neural activity holds enough information for the human brain to infer properties of the outer world, should the same not be possible for a machine, too? Although this view remains debatable, studies have demonstrated that decoding mental states is possible to at least some limited extent [1]. A branch that is particularly successful is visual decoding, which tries to establish a link between brain activity and visual stimulation [2]. People have created so-called Inverse Encoding Models, which are acquired by estimating how the human brain encodes information and using the inverse process as a decoding model [3]. These models have, among many other achievements, been able to identify observed pictures, reconstruct shapes and even classify visual imagery to a certain degree [2–5].

In this short study, an inverse encoding model is trained to reconstruct checkerboard stimuli observed by a participant in an fMRI scanning session. It is examined if and to what extent this is possible using a simple linear model. The position of decoding-relevant voxels is furthermore used to investigate the spatial organization of the primary visual cortex.

2 Methods

2.1 Experimental design and image acquisition

The fMRI images used in this study were originally recorded for another study [5]. In the scanning sessions, participants were presented a flickering checkerboard divided in 48 segments, where the contrast of each segment was randomly chosen from 4 logarithmic levels between 0.1 and 1. Every 3 seconds, new contrast values were chosen randomly and independenctly for all segments. An example stimulus is depicted in Fig. 1. A total of 8 runs were recorded per participant, where each run consisted of 100 checkerboard stimuli.

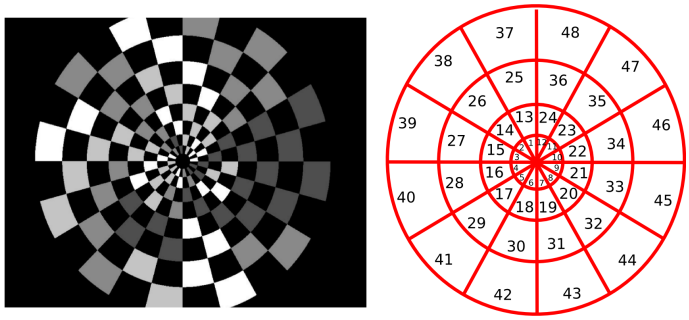

Figure 1: Example checkerboard stimulus. **Left**: Actual stimulus. The checkerboard consisted of 48 segments, where each segment showed a randomly chosen contrast value on a logarithmic value between 0.1 and 1. **Right**: Schematic structure of the 48 sectors.

The images were acquired on a 3T-scanner at a voxel size of $3mm^3$ and a repetiton time of $Tr = 1500ms$. In order to prevent saccades, participants were performing a simple fixation task. A more detailed description can be found in the original paper [5].

2.2 Inverse Encoding

For the study, an inverse encoding approach was used to reconstruct the checkerboard stimuli from recorded fMRI data. The basic workflow is depicted in Fig. 2: First, a linear least-squares regressor is established to map real-world stimuli to the measured response of single voxels. This is done for each voxel separately by least-squares regression. The mathematical inverse of the encoding model is then used for the actual "reconstruction", which is to find the stiumuls that is most likely to have caused the observed voxel activations. The scope of this report does not allow a detailed discussion of the diverse methodology

used in Inverse Encoding – for a much more in-depth view on the subject, see e.g. [6].

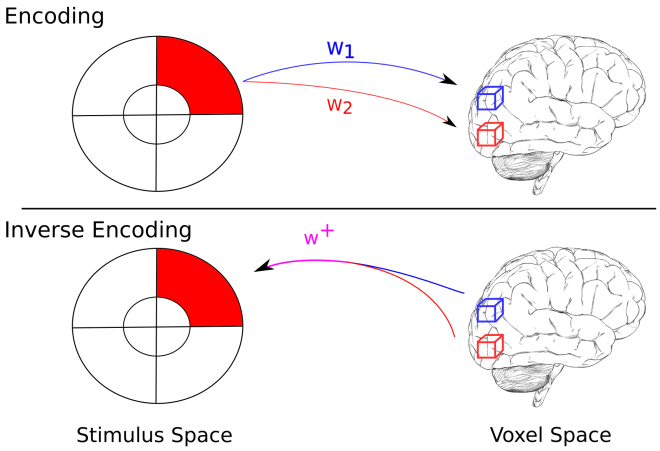

Figure 2: Inverse Encoding workflow. **Top**: In the encoding phase, weights are determined to predict every single voxel's response to a stimulus. This step is done in a univariate fashion. **Bottom**: Once it is known how each voxel reacts to a stimulus, we can invert the process: given a set of voxel activations, we can estimate the stimulus that is most likely to have caused them. Brain scheme taken from [7].

Since each stimulus consisted of 48 compartments, we decided to train an inverse encoding model for each compartment separately, ignoring the remaining ones. The procedure therefore resulted in 48 single-compartment models. The encoding model for each voxel consisted of a linear regressor that maps the four real-space contrasts values to the voxel's response. The model thus assumed a linear relationship between stimulus intensity and voxel response. For decoding, a so-called searchlight [8] was used on the primary visual cortex with a radius of two voxels. Due to the simplicity of the encoding model, we could simply take its inverse as our inverse encoding model. The analysis relied on the software package *The Decoding Toolbox (TDT)*, which specializes in multivariate decoding of neuroimaging data [9].

2.3 Permutation analysis

An important task in statistical analysis is to determine a threshold for significance. To get a sense of the distribution of resulting correlation from meaningless data, a permutation analysis was conducted. For each run, stimulus labels were switched in a randomly chosen order that changed between runs (e.g. all ones switched to four, all twos to threes,...).

We then performed searchlight-decoding on the scrambled data. To minimize statistical insecurity, the whole procedure was repeated 2000 times. The resulting correlation values are assumed to represent a baseline distribution for non-informative data.

3 Results

3.1 Permutation analysis

In order to estimate the regressor's performance distribution dealing with non-informative data, a permutation analysis was conducted. Fig. 3 shows the Z-Correlation distribution based on 2000 random permutations of V1 Voxels. The dotted line indicates the 99th percentile (0.094).

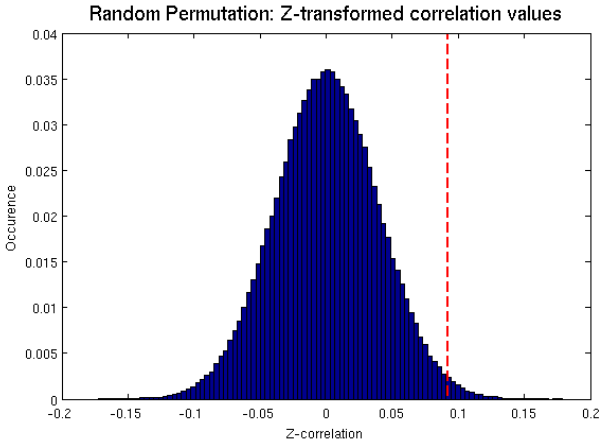

Figure 3: Result of 2000 permutation tests using the Voxels in V1. The red line indicates the 99th percentile at z-corr = 0.094 , which is furthermore used as a basic threshold of statistical significance.

3.2 Reconstruction accuracy

Each of the 48 decoders was trained to predict one sector's time course based on voxel data. An example reconstruction trace is depicted in Fig. 4. Since we were using regression rather than classification, the predictor emitted continuous values. The validity of a prediction is determined by the z-transformed correlation between true and predicted labels. The fit shown in Fig. 4 corresponds to a value of 0.83, which by far exceeds any correlations seen in the permutation data (99th percentile of permutation data: 0.094).

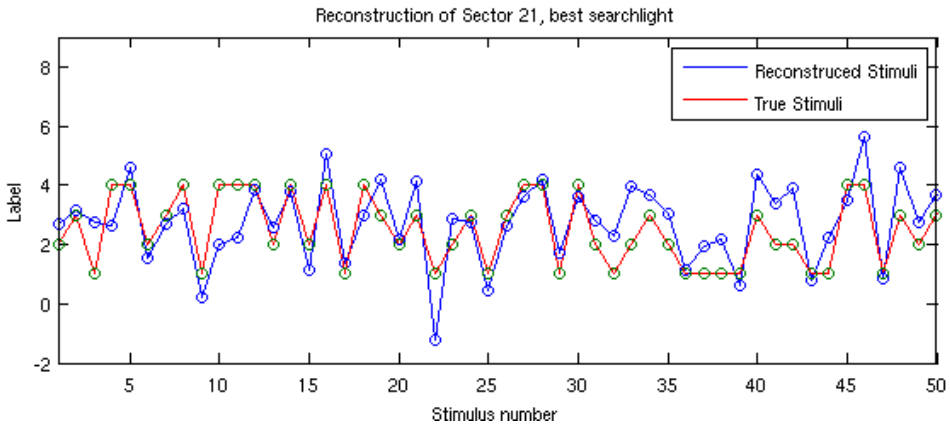

Figure 4: Example of stimulus reconstruction of segment 21 at a well performing searchlight. Due to the usage of local decoders, stimuli of each checkerboard segment are estimated individually. This reconstruction trace is shown for illustrative purposes only.

An overview over the reconstruction process is given in Fig. 5. The most trivial observation is that the regressors were indeed decoding *something*, with z-correlations of up to 0.8, which by far exceeds values seen in the permutation analysis. Furthermore, only a small subset of voxels seems to hold information about a given sector. Interestingly, the subset of informative voxels changes over sectors: apparently, different sectors are encoded at different places in the brain.

To get an estimate of the *true* decoding potential, we used nested crossvalidation: an inner loop is searching for the optimal searchlight position for each sector, which is then tested on an untouched validation set. The outcome of this procedure is relatively proof against overfitting. Fig. 6 shows the best reconstructions that appeared for each sector. The values show large variance between sectors: while some, often close to the horizontal axis, exhibit relatively high correlation, other sectors do not reach statistical significance at all. This effect will briefly be reviewed in the discussion section.

3.3 Spatial organization in V1

Due to the usage of searchlights, we can visualize which part of the brain holds information for which part of the visual stimulus. An overview over the spatial organization is given in Fig. 7. The results allow a first glance onto retinotopical organization. Two basic conclusions can be drawn: First, stimuli in one visual hemifield show good decoding results

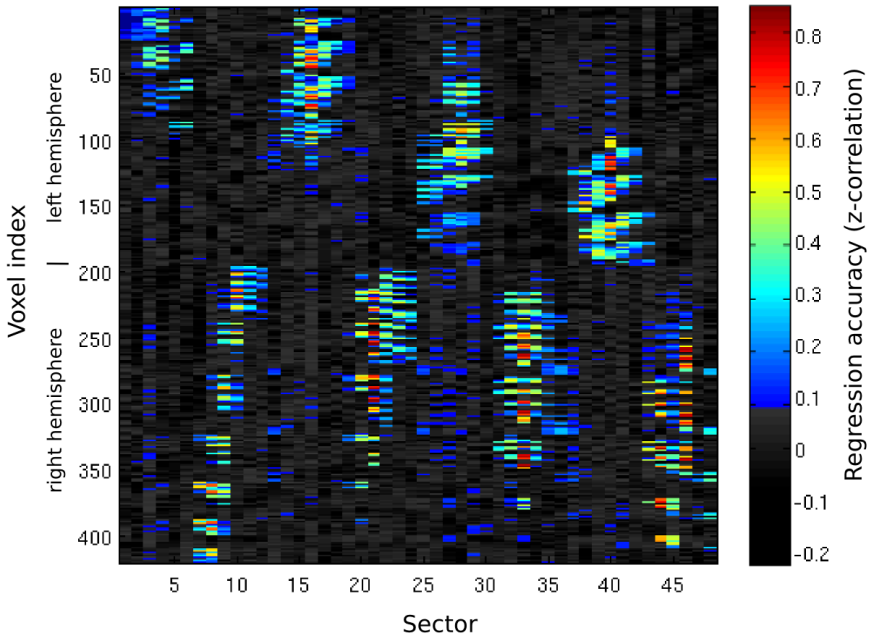

Figure 5: Prediction performance as a function of searchlight position and considered sector. Non-significant values are depicted in grey (threshold 0.094). Best performing searchlight positions do not overlap largely across many sectors: different parts of the visual field are represented in well-separated groups of voxels.

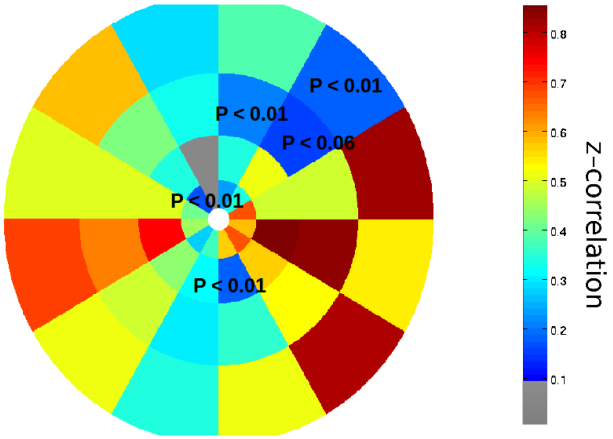

Figure 6: Correlation between true stimuli and predictions by the best performing searchlights. Reconstruction performance varies largely for different segments, reaching from values of more than 0.8 to barely significant ones of less than 0.2. Generally, segments along the horizontal plane seem to perform better than those on the vertical plane. Unless indicated otherwise, all p-values are smaller than 0.0005.

mostly in the contralateral hemisphere. Second, stimuli that are spacially close to each other are apparently also represented closely on the cortex.

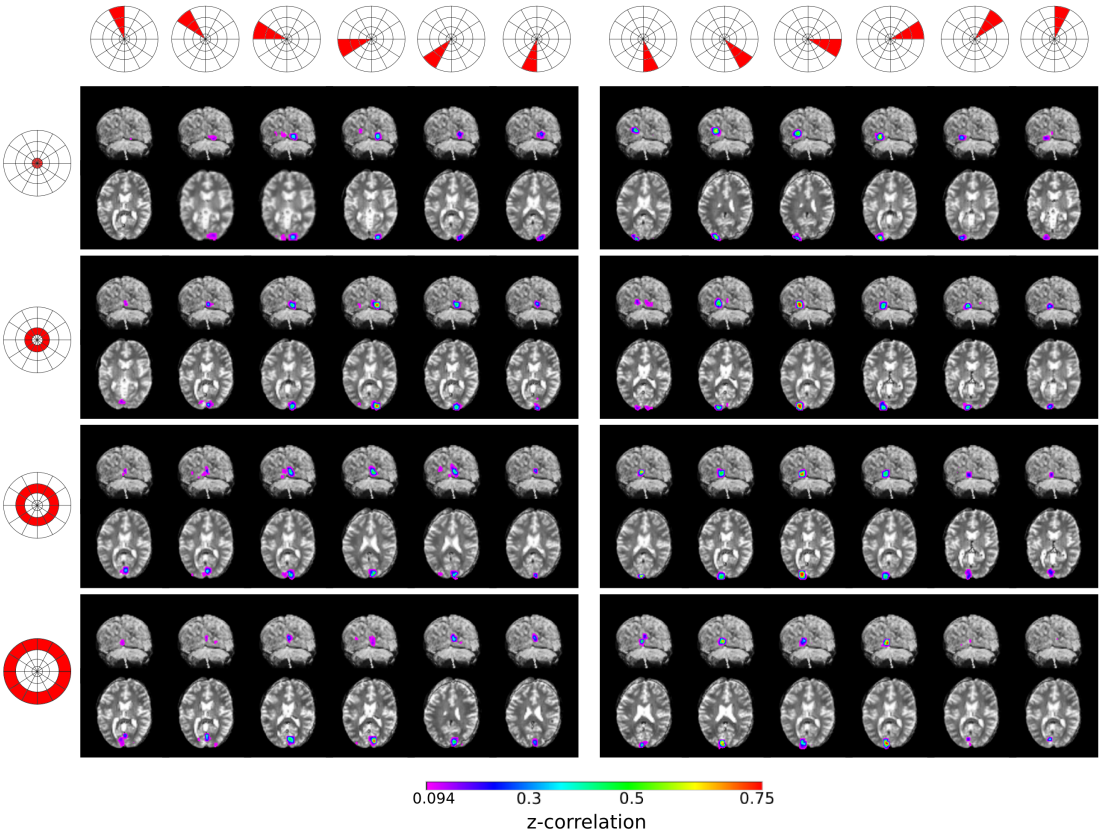

Figure 7: Searchlight positions with above-chance regression accuracy. Only V1 Voxels are considered. Rows: stimulus excentricity. Columns: stimulus angle. Stimuli that are spacially close are also represented closely on the cortex.

We then tried to understand the principle mapping structure in more detail. For this, the "top five" voxels of each segment, derived by a crossvalidation procedure, were colored according to eighter their angle or excentricity in the visual field (Figure 8). The saggital view indeed allows deeper insights into the organizational structure. Upon varying the excentricity (Figure 8A, top row), layers at different position, yet similar orientation become visible. Strikingly, varying the angle of a stimulus also results in distinct layers (second row) - perpendicular to those seen before. It seems as if any position in the visual field is encoded via its excentricity and polar angle, where both of these traits are encoded along different, perpendicular axes. Figure 8B sums up the findings in a schematic view.

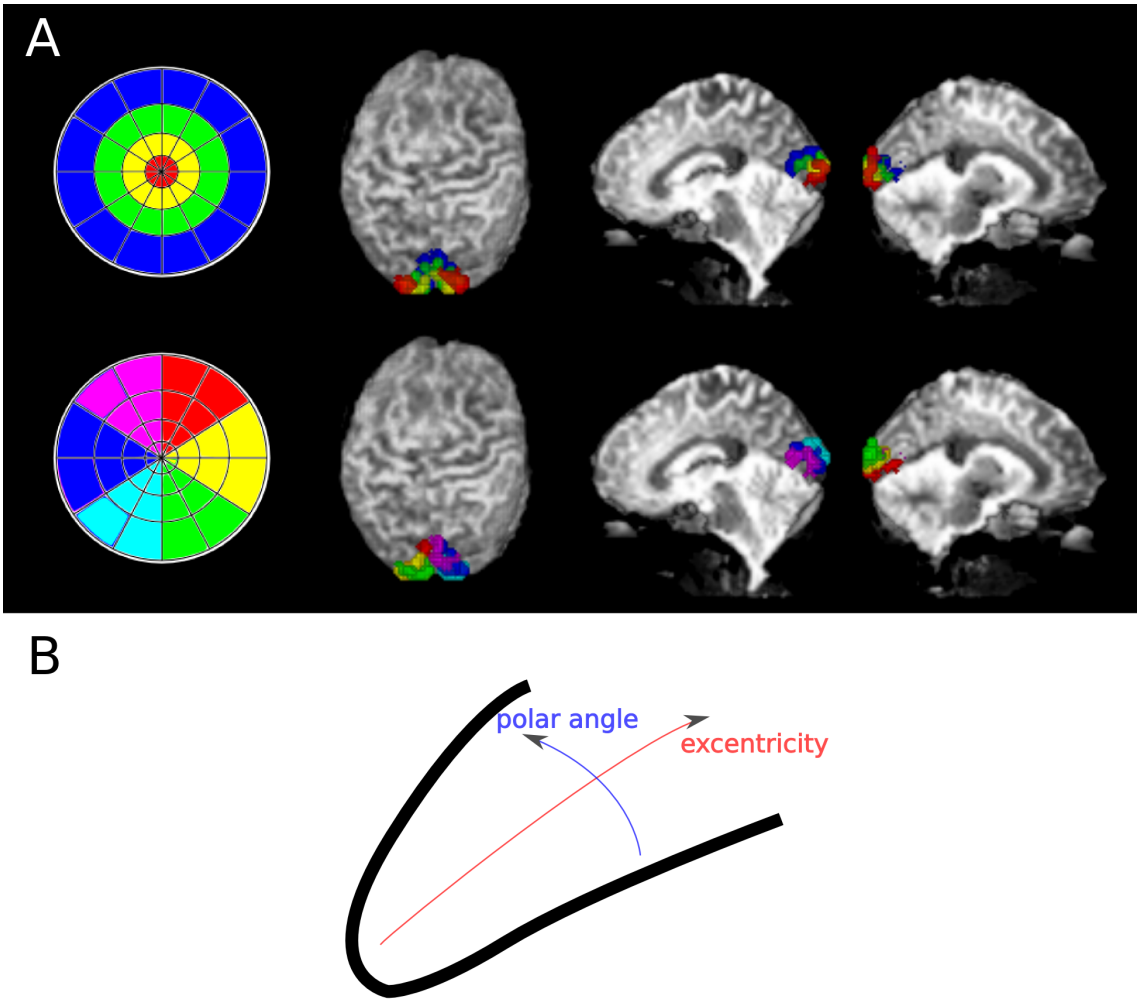

Figure 8: The top-five searchlight positions of each segment are colored according to the scheme on the left side. **A**: Varying sector positions in terms of angle and distance from the center ("excentricity") results in well-defined layers - apparently, angle and excentricity are encoded along perpendicular axes, maintaining the original structure of visual stimuli in a distorted way. **B**: Schematic view on the findings from A.

4 Discussion

In the study, we demonstrated that inverse encoding of visual stimuli by a linear regression approach is well possible and reasonable results can be produced.

A surprising finding is that decoding performance is highly dependent on stimulus position. For segments on the top or bottom of the visual field, the correlations are generally very low, while those on the horizontal axis tend to be much higher. Although the precursing study by Oliver Eberle and Hanna Röhling showed some variance in performance as well, the effect in their study was much weaker. It could be speculated that the representation of the badly reconstruced sectors is not linear and can therefore not be captured well by our linear regressor. Another, more trivial explanation could be that these sectors are just represented outside of what we define as V1 in our brain mask.

The retinotopic organization, on the other hand, was surprisingly clear and seems to be in good agreement with common literature. We demonstrated that excentricity and polar angle of stimuli are encoded along different axes and that the spatial structure of stimuli is to some degree sustained in the visual cortex. Also the concept of contralateral encoding is easily confirmed. The fact that the mapping structure was found as a byproduct of decoding furthermore confirms the sanity of the whole approach.

In a wider study, it would be interesting to consider the whole brain, rather than just V1. This would not only allow us to check for cofounds, correlated brain structures and the sanity of the whole approach, but also to see if, as speculated earlier, any segments are indeed encoded outside of our original brain mask.

Also, since the study focusses on one participant only, it is dangerous to draw hasty conclusions – it would surely be informative to repeat the procedure on different participants, to compare results and find comprehensive principles.

References

[1] John-Dylan Haynes and Geraint Rees. Decoding mental states from brain activity in humans. *Nat Rev Neurosci*, 7(7):523–534, jul 2006.

[2] Bertrand Thirion, Edouard Duchesnay, Edward Hubbard, Jessica Dubois, Jean-Baptiste Poline, Denis Lebihan, and Stanislas Dehaene. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, 33(4):1104–1116, 2006.

[3] Thomas C Sprague, Edward F Ester, and John T Serences. Reconstructions of information in visual spatial working memory degrade with memory load. *Curr Biol*, 24(18):2174–2180, sep 2014.

[4] Kendrick N Kay, Thomas Naselaris, Ryan J Prenger, and Jack L Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, 2008.

[5] Jakob Heinzle, Thorsten Kahnt, and John-Dylan Haynes. Topographically specific functional connectivity between visual field maps in the human brain. *Neuroimage*, 56(3):1426–1436, jun 2011.

[6] John Dylan Haynes. A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives, 2015.

[7] Drawing distributed under a GNU Free Documentation License, https://commons.wikimedia.org/wiki/File:Hersenen.png#file. Last access: 7.1.2017.

[8] Nikolaus Kriegeskorte, Rainer Goebel, and Peter Bandettini. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103:3863–3868, 2006.

[9] Martin N Hebart, Kai Görgen, and John-Dylan Haynes. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front Neuroinform*, 8:88, 2014.