NeuroImage 163 (2017) 244–263

Contents lists available at ScienceDirect

NeuroImage

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

Robust inter-subject audiovisual decoding in functional magnetic resonance imaging using high-dimensional regression

Gal Raz [a,b,c,*], Michele Svanera [d], Neomi Singer [a,c,e], Gadi Gilam [a,e], Maya Bleich Cohen [a], Tamar Lin [a], Roee Admon [f], Tal Gonen [a,g], Avner Thaler [a,c,h,i], Roni Y. Granot [j], Rainer Goebel [k], Sergio Benini [d], Giancarlo Valente [k]

[a] Functional Brain Center, Wohl Institute for Advanced Imaging, Tel Aviv Sourasky Medical Center, 64239 Tel Aviv, Israel
[b] Film and Television Department, Tel Aviv University, 69978 Tel Aviv, Israel
[c] Sagol School of Neuroscience, Tel Aviv University, 69978 Tel Aviv, Israel
[d] Department of Information Engineering, University of Brescia, 38, 25123 Brescia, Italy
[e] School of Psychological Sciences, Tel Aviv University, 69978 Tel Aviv, Israel
[f] Department of Psychology, University of Haifa, 3498838 Haifa, Israel
[g] Department of Neurosurgery, Tel Aviv Sourasky Medical Center, 64239 Tel Aviv, Israel
[h] Movement Disorders Unit, Neurological Institute, Tel-Aviv Sourasky Medical Center, 64239 Tel Aviv, Israel
[i] Sackler Faculty of Medicine, Tel Aviv University, 69978 Tel Aviv, Israel
[j] Musicology Department, Hebrew University of Jerusalem, 9190501 Jerusalem, Israel
[k] Department of Cognitive Neuroscience, Maastricht University, 6211 LK Maastricht, The Netherlands

ARTICLE INFO

Keywords:
fMRI
Audiovisual decoding
Motion pictures
Kernel ridge regression
Sound loudness
Optical flow
Face
Motion pictures

ABSTRACT

Major methodological advancements have been recently made in the field of neural decoding, which is concerned with the reconstruction of mental content from neuroimaging measures. However, in the absence of a large-scale examination of the validity of the decoding models across subjects and content, the extent to which these models can be generalized is not clear. This study addresses the challenge of producing generalizable decoding models, which allow the reconstruction of perceived audiovisual features from human magnetic resonance imaging (fMRI) data without prior training of the algorithm on the decoded content. We applied an adapted version of kernel ridge regression combined with temporal optimization on data acquired during film viewing (234 runs) to generate standardized brain models for sound loudness, speech presence, perceived motion, face-to-frame ratio, lightness, and color brightness. The prediction accuracies were tested on data collected from different subjects watching other movies mainly in another scanner.

Substantial and significant (Q_{FDR}<0.05) correlations between the reconstructed and the original descriptors were found for the first three features (loudness, speech, and motion) in all of the 9 test movies ($\overline{R}$=0.62, $\overline{R}$ = 0.60, $\overline{R}$ = 0.60, respectively) with high reproducibility of the predictors across subjects. The face ratio model produced significant correlations in 7 out of 8 movies ($\overline{R}$=0.56). The lightness and brightness models did not show robustness ($\overline{R}$=0.23, $\overline{R}$ = 0). Further analysis of additional data (95 runs) indicated that loudness reconstruction veridicality can consistently reveal relevant group differences in musical experience.

The findings point to the validity and generalizability of our loudness, speech, motion, and face ratio models for complex cinematic stimuli (as well as for music in the case of loudness). While future research should further validate these models using controlled stimuli and explore the feasibility of extracting more complex models via this method, the reliability of our results indicates the potential usefulness of the approach and the resulting models in basic scientific and diagnostic contexts.

* Corresponding author. Kfar Daniel 253, 7312500, Israel.
E-mail address: galraz@post.tau.ac.il (G. Raz).

https://doi.org/10.1016/j.neuroimage.2017.09.032
Received 3 May 2017; Received in revised form 14 September 2017; Accepted 17 September 2017
Available online 20 September 2017
1053-8119/© 2017 Elsevier Inc. All rights reserved.

1. Introduction

"Mind reading" based on neural decoding is an ambitious line of research within contemporary neuroscience. Assuming that certain psychological processes and mental contents may be encoded in the brain as specific and consistent patterns of neural activity, researchers in this field aim to decode and reconstruct them given only the neuroimaging data. In order to "read" stimuli out of one's brain, researchers adopt different machine learning approaches and apply various pattern analysis methods that link local or distributed neural activity patterns with specific audiovisual features.

Neural decoding refers to the prediction of a stimulus features from measured brain activity (Schoenmakers et al., 2013) (fMRI data in our case). Several notable neural decoding achievements have been reported so far, mainly in studies employing functional magnetic resonance imaging (fMRI), but also in intracranial recording and electro- and magneto-encephalography experiments (for review, see Chen et al., 2013; Haxby, 2012). Reported decoding classification accuracies for out-of sample data commonly range between 70 and 90% (see Poldrack et al., 2009), and correlation as high as 0.99 between predicted and observed continuous descriptors was demonstrated (Valente et al., 2011). Decoding targets vary and include mental states such as action intentions (Haynes et al., 2007), reward assessment (Kahnt et al., 2011) and response inhibition (Cohen et al., 2010; Poldrack et al., 2009); low-level features such as visual patterns in dynamic video (Nishimoto et al., 2011), geometrical patterns, text (Fujiwara et al., 2009; Miyawaki et al., 2008; van Gerven et al., 2010), and optical flow acceleration in a video game (Chu et al., 2011; Valente et al., 2011); and semantic elements such as animal and objects categories (Connolly et al., 2012; Haxby et al., 2001, 2011), objects and actions in a hierarchical semantic space (Huth et al., 2012), visual imagery content during sleep (van Gerven et al., 2010), and actions and events in a video game (Chu et al., 2011; Valente et al., 2011).

This productive stream of research supports the appealing vision of generating a repertoire of "fMRI fingerprints" for a wide range of mental states and perceptual processes (or "cognitive ontology", see Poldrack et al., 2009). Ideally, such repertoire will facilitate robust and rich neural decoding for any subject independently of prior training and using any standard MRI scanner. The generation of a reliable repertoire of this kind is valuable both in terms of basic science (providing a reproducible and comprehensive ground truth for brain-function mapping) and applicable technology (in diagnosis, for instance; see Cohen et al., 2011).

Strong evidence for the generalizability of such repertoire of functional models of the brain can be gained by demonstrating their performance under conditions of high heterogeneity across the training and the test data. For this end, it is necessary to show that these models facilitate successful decoding also when analyzing stimuli that are different from those on which the algorithm was trained. However, eminent neural decoding achievements were gained using a within-subject design including only five subjects or less (e.g., Horikawa et al., 2013; Huth et al., 2012; Miyawaki et al., 2008; Nishimoto et al., 2011), which limits the examination of the reproducibility of the results. Thus, the key aspect of inter-subject generalizability of neural decoding has yet to be systematically investigated (Chen et al., 2013).

Confronting the limitations of the within-subject design in neural decoding, Haxby and colleagues have recently demonstrated the feasibility of between-subject classification. This group developed methods for cortical anatomy alignment for different subjects based on the maximization of the inter-subject similarity of blood oxygen level dependent (BOLD) reaction patterns (Haxby et al., 2011; Sabuncu et al., 2010) and functional connectivity structures (Conroy et al., 2013). These studies have demonstrated that between-subjects classification may yield success rates equivalent to those of within-subject classification. Successful decoding of data of out-of-sample individual was also reported in few other studies that did not implement inter-subject alignment methods that rely on functional data. Shinkareva et al. (2008) and Poldrack et al. (2009), reached average accuracy rates of ~80% in classifying visual input and task type, respectively. Cohen and colleagues (Cohen et al., 2011) decoded response inhibition related variables with above-chance correlation values of 0.4–0.5 between the predicted and real parametric values.

In keeping with the notion that a compelling validation of neural decoding method relies on its success under highly heterogeneous conditions, the current work introduces a markedly increased variability across several experimental dimensions. First, we aimed to decode continuous time-varying features, which change on a moment-to-moment basis. Second, we tested the decoding reliability on a set of different movies, comprising highly heterogeneous, complex, and naturalistic stimuli. Lastly, the validation of the function-brain models was performed using movies that were not employed in the training procedure with data collected in a different MRI scanner from un-tested individuals. An overview of the study is given in Fig. 1.

We combined data from 234 movie-viewing sessions (with 5 different clips) for the training of our algorithm and the cross-validation of the resulting model (Table 1). The validity of the models was tested using an independent sample of 63 sessions (with 9 other clips). We selected relatively coarse features across three elementary perceptual domains: audio, vision and motion. The selected features were sound loudness (loudness), speech presence (speech), detected motion (motion), face-to-frame dimension ratio (face ratio), perceived lightness, and brightness. These audiovisual features were extracted using both manual and automatic annotation tools.

In order to decode these continuous features from the fMRI data we used linear kernel ridge regression (KRR) with generalized cross-validation (GCV) (Golub et al., 1979). We chose a kernel version since it is particularly efficient when the number of data points is considerably lower than the number of measurable properties, or features (in our case time repetitions and number of voxels, respectively; see Golub et al., 1979). The combination of L2-norm penalization with GCV is relatively computationally inexpensive when compared with iterative kernel methods such as Relevance Vector Regression and Gaussian Processes, while still achieving good performance, and it allows for fast permutations in order to ascertain non-parametrically statistical significance (Valente et al., 2014). We used a linear kernel for several reasons. First, by using a linear model we allow for the reconstruction of descriptors of specific features as linear combinations the weighted BOLD time series, with the advantage of a straightforward interpretation of the fMRI models relative to non-linear kernels and other complex pattern recognition methods such as artificial neural networks. Second, the optimization of non-linear kernel hyperparameters would increase the computational time of several orders of magnitude. Finally, the large number of dimensions, compared to the available samples in our problem, makes it difficult to exploit the increased flexibility provided by non-linearities, increasing the risk of overfitting. An alternative to linear kernel ridge regression could be to use a linear ridge regression after projecting the data onto the subspace spanned by the principal components, which would result in similar computational burden if all the principal components are retained.

In addition to the extraction of spatial brain models of continuous audiovisual features, we temporally optimized the models. In specific, we applied time-lag optimization following evidence that multi-voxel pattern analysis (MVPA) classification may be improved by fitting different temporal hemodynamic response models to different brain regions (Kohler et al., 2013). Due to the high dimensionality of the problem we used simulated annealing (Kirkpatrick et al., 1983), a heuristic algorithm based on thermodynamic principles, to optimize the temporal parameters of the decoding models. This procedure was performed on a cross-validation subset of the data.

Thus, we produced spatio-temporal decoding maps, which assign optimized lag and weight values to every voxel in the brain to reconstruct specific features. Our study examined the extent to which various audiovisual features can be robustly and reliably reconstructed by these

Download English Version:

https://daneshyari.com/en/article/5630793

Download Persian Version:

https://daneshyari.com/article/5630793

Daneshyari.com