Data-driven Multisubject Neuroimaging Analyses for Naturalistic Stimuli

Felix Biessmann[§†], Michael Gaebler[*], Jan-Peter Lamke[*], Ui Jong Ju[†], Stefan Hetzer[*], Christian Wallraven[†], and Klaus-Robert Müller[†‡]

[§]Amazon Development Center, Berlin, Germany [*]Charité – Universitätsmedizin Berlin, Germany [†]Korea University, Seoul, Republic of Korea [‡]Technische Universität Berlin, Germany

***Abstract*—A central question in neuroscience is how the brain reacts to real world sensory stimuli. Naturalistic and complex (e.g. movie) stimuli are increasingly used in empirical research but their analysis often relies on considerable human efforts to label or extract stimulus features. Here we present data-driven analysis strategies that help to obtain interpretable results from multisubject neuroimaging data when complex movie stimuli are used. These analyses a) enable localization and visualization of brain activity using standard statistical parametric maps in the subspace of brain activity shared between subjects and b) facilitate interpretation of intersubject correlations. We show experimental results obtained from 50 subjects.**

***Index Terms*—Multisubject Neuroimaging, Hyperscanning, Canonical Correlation Analysis, CCA, Intersubject Correlation**

I. Introduction

REALISTIC stimuli have become increasingly popular in neuroimaging, see e.g. [1], [3]. Many of the analysis approaches to neuroimaging data acquired during stimulation with naturalistic stimuli however require intensive human work, such as for the design of appropriate filters for feature extraction [2] or labeling of objects and scenes [3]. For large scale analyses this approach quickly becomes infeasible. Unsupervised analysis methods, which do not require manually created labels or stimulus regressors, have proven useful in this setting. These methods allow to find structure in data sets in an explorative fashion. Supervised methods usually find those aspects of brain activity that correlate well with a regressor. Which aspects of brain activity they find depends on the objective function of the analysis method. Many approaches find linear approximations of the data matrix such that rows or columns (corresponding to time and space of the brain data) have maximal variance or are statistically independent from each other. Another useful criterion to optimize is the shared covariance or correlation between pairs of subjects, that is to maximize *intersubject correlations*. Intersubject correlations are correlations of neuroimaging time series between pairs of subjects participating in the same experiment (exposed to the same stimulus or involved in direct interaction). The hypothesis is that finding locations or networks that exhibit large intersubject correlations allows to find those aspects of brain activity that are shared between multiple subjects. While early approaches were restricted to mass-univariate correlation coefficients [1], later work made use of multivariate methods in order to analyze networks of activity, rather than single locations in the brain [4]. If the objective is to find the most correlated subspaces in the brain imaging data of multiple subjects, the most straightforward analysis is canonical correlation analysis (CCA) [5]. The rationale for multisubject neuroimaging studies is illustrated in Figure 1. CCA assumes that any variance in the data that is shared amongst all subjects is reflecting brain processes associated with a complex stimulus. For a more formal definition see section II-D. In the following we will refer to analyses that are based on intersubject correlations as *ISC* methods and to analyses based on the multivariate extension as canonical ISC or *CISC* approaches.

One major problem with ISC-based approaches is that the results can be difficult to interpret. Often it is necessary to apply extensive pre- and postprocessing to the data, especially when working with mass-univariate approaches: the strongest shared activations (among subjects) are typically rather unspecific. For instance in visual paradigms all visual cortices are activated. If one is interested in more specific aspects of brain activation, these unspecific activations have to be subtracted [1]. Here we present two simple ways of extending (C)ISC based analyses that allow for a better interpretation of brain activation shared amongst multiple subjects. The first approach is a combination of the results presented in [6] to the latent variable model estimated by multiway CCA and large-scale data mining techniques as presented in [7]. This allows to visualize and interpret the common networks and localize differences in experimental conditions. The second approach is based on multivariate decoding of stimulus conditions from intersubject correlations. This allows to draw conclusions about the nature of changes in intersubject synchronization, for instance by relating these changes to single scenes of a movie as done in the mass-univariate reverse correlation approach in [1] or the multivariate approach in [4].

We present preliminary results obtained in two studies with 25 subjects each. We investigated the shared brain activation in response to naturalistic movie stimuli, which were shown with stereoscopic depth (3D condition) and without this information (2D condition). We found that canonical ISC analysis can reliably detect and localize those cortical networks that share activation across subjects – per condition and differentially between conditions. In addition it can identify movie scenes in which this shared activation carries information about the stimulus condition. In our setting standard SPM type mass-univariate analyses with the stimulus condition as regressor

978-1-4799-4149-0/14/$31.00 2014 IEEE

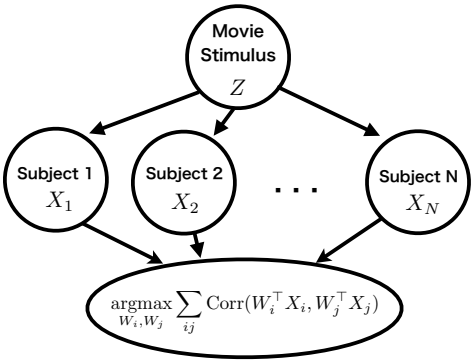

Fig. 1. The model underlying canonical intersubject correlation studies: the same complex stimulus (such as a movie) is presented to a group of subjects. A good approximation of the brain activity elicited by the stimulus is that brain activity that is shared amongst all subjects. This aspect of brain activation is extracted by multiway CCA, which finds a subspace W_s of each subject's brain activation X_s such that the pairwise canonical correlation across subjects is maximized.

TABLE I
DESCRIPTION OF THE MOVIE CONTENT.

No.	Description
1	Ride through a city in an oldtimer car
2	Time lapse movie of a pink flower opening and closing its bloom
3	Flock of dolphins swimming through underwater plants
4	Police sheriff and woman exploring a dark alley
5	Skateboarders doing tricks in a skateboard hall
6	Mountainbikers jumping over gaps in a dirt course
7	Three people fishing and exchanging money, two leaving in a canoo
8	Race car rallye through the woods
9	Roller coaster ride
10	Scenes from a Graffiti and BMX event
11	Surfer standing on his board and riding a wave
12	Individual manatee, then a flock of manatees under water
13	People jumping over a cliff in wingsuit costumes
14	Skydive with the jump from the plane, free fall, and landing
15	Race car rallye through the woods (longer version)
16	Fight between mantis shrimp and octopus
17	Walk through cherry blossoms

failed to capture significant differences between conditions. Furthermore, combining our results with data-driven analyses of a large body of neuroimaging studies [7] we can map the localization results to psychological concepts associated with the differential activation of the networks found. Importantly the two proposed approaches are fully automatized and do not require human interaction for feature extraction or labelling.

II. METHODS

In two experiments, 50 subjects (experiment 1: 13 female, 12 male, age 26.7±3.5 years, range 21-35; experiment 2: 12 female, 13 male, age 26.6±5.1 years, range 19-38) were shown movie clips of varying content while we recorded their brain activity using fMRI. Each participant saw each movie with and without stereoscopic depth information. Participants were naïve with respect to content and category of the stimuli, had normal or corrected-to-normal vision, and could perceive stereoscopic depth cues.

A. Stimuli

Stimuli 1-14 (experiment 1) were videos of 42.5 s length each: content length was 40.5 s, preceded by 2 s of black screen without fixation cross for visual adjustment and to avoid distortions induced by codec and presentation software. Stimuli 15-17 (experiment 2) were movie clips of 120 s length. All videos were presented at 30 frames per second, resulting in a total number for each stimulus of 1275 frames at size 768×576 pixels on each eye. The videos were acquired over the internet and video content varied from, for example, a calm time lapse montage of a blossoming flower (http://www.stereomaker.net/sample/index.html, accessed March 20, 2013) to a rapid car rallye, filmed by onboard cameras (http://alesco.cz/, accessed December 8, 2012; see Table I). Videos were edited using VirtualDub 1.9.11 (http://www.virtualdub.org/) and encoded using the XVID codec. Every movie was shown twice: In the 3D condition, stereoscopic depth was induced by presenting the two binocular perspectives of the scene to the corresponding eyes, while in the 2D condition, the same stimulus (left eye) was delivered to both eyes. As stimulus order was pseudo-randomized for each participant, novelty effects were balanced and stimulus characteristics were controlled for. The latter involve statistical properties like brightness, contrast, color, or motion but also more subjective stimulus features like personal preference for the movie content. Videos were interspersed with 20 s blocks of fixation and the first video presentation was preceded by a 30 s baseline fixation block. Presentation software version 14.9 (Neurobehavioral Systems, Inc., Albany, CA), a stereo adapter, and MR-compatible video goggles with a native resolution of 800×600 pixels and a color depth of 32 bit (VisualSystem, NordicNeuroLab; Bergen, Norway) were used for stimulus presentation. Careful adjustment of the goggle system and its built-in dioptric correction prior to scanning ensured optimal stimulus visibility.

B. Data acquisition

MR imaging in two separate experiments was performed on two different Siemens TIM Trio 3T MR scanners with standard 12-channel head coils (Siemens Medical Solutions, Erlangen, Germany). For each participant, a T_1-weighted image was acquired as high-resolution anatomical reference. T_2^*-weighted gradient-echo echo-planar images (EPI) were collected for whole-brain functional imaging with voxels of 2.5×2.5×2.5 mm^3 and 2×2×4 mm^3 in experiments 1 and 2, respectively.

C. Imaging data preprocessing

Image preprocessing and statistical analyses were carried out using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/) and Matlab (MathWorks, Natick, MA, USA). Image series were inspected for excessive head movements but no subject exceeded the threshold of 1 mm/*TR*. After realignment to the first image and T_1 coregistration onto the mean EPI, rigidly aligned tissue-class images for gray and white matter and cerebrospinal fluid were generated from the coregistered T_1 images employing the "New Segment" function. Functional images were then normalized to MNI space and smoothed with a Gaussian

kernel of 6 mm FWHM using the normalization function of the DARTEL toolbox. For further analysis we extracted the grey matter voxels using the respective template contained in SPM8 after binarizing it with a threshold of .5.

D. fMRI data analysis

In order to find brain networks of activation that are common to all subjects, we used canonical correlation analysis (CCA) [5]. The assumption of CCA is that a set of K networks of brain activation for each subject $s \in \{1, 2, \ldots, S\}$ can be modeled as a linear subspace $W_s = [w_{s1}, w_{s2}, \ldots, w_{sK}] \in \mathbb{R}^{V \times K}$ (V denotes the number of voxels) of the multivariate voxel time series $X_s \in \mathbb{R}^{V \times T}$ (T denotes the number of fMRI volumes). The column vectors w_{s1} to $w_{sK} \in \mathbb{R}^V$ are called *canonical directions*; the subscript s refers to a specific subject. We can obtain the time courses, also called *canonical components*, of these brain networks for subject s by computing $W_s^\top X_s$. The goal of CISC analysis is to find canonical directions W_s such that the sum over all pairwise correlations (for all pairs of subjects) between the canonical components is maximized, with the constraint that the time courses of two different networks w_{su} and w_{sv} be uncorrelated for all subjects s and all components $u \neq v$. The objective function of CCA can be formulated as

$$\underset{W_i, W_j}{\operatorname{argmax}} \sum_i \sum_j Trace\big(W_i^\top X_i X_j^\top W_j\big), \quad \forall i, j \qquad (1)$$
$$\text{subject to } W_i^\top X_i X_i^\top W_i = \mathbf{I}, \quad \forall i,$$

where $\mathbf{I}$ is the identity matrix. In general if there are N multivariate variables and corresponding centered data matrices $\{X_1, X_2, \ldots, X_N\}$, the basis vectors of the canonical subspace of each variable $\{W_1, W_2, \ldots, W_N\}$ can be found by solving the generalized eigenvalue problem

$$\begin{bmatrix} 0 & C_{12} & \ldots & C_{1N} \\ C_{21} & 0 & \ldots & C_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \ldots & 0 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix} = \begin{bmatrix} C_{11} & 0 & \ldots & 0 \\ 0 & C_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & C_{NN} \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_N \end{bmatrix} \Lambda. \qquad (2)$$

where C_{ij} denotes the covariance matrix between the ith and jth variable. For a more in-depth treatment of multi-way CCA objectives see e.g. [8]. The dimensionality of the data was reduced using PCA to keep only as many principal components as are needed to cover 99.9 percent of the variance in all voxels. This resulted in 20 to 30 principal components. We performed all analyses in a leave-one-movie-out cross-validation manner. For each movie, we estimated the PCA subspace as well as the canonical directions on all but this movie (the *training data set*). The canonical components for the fMRI data recorded during the held-out movie were computed by projecting them onto the PCA space and the canonical directions computed on the training data set. The cross-validated CISCs reported here are computed on these canonical components.

E. Localization of differential CISC strength

The spatial activation pattern A_s of a canonical component can be obtained by

$$A_s = W_s^\top X_s X_s^\top. \qquad (3)$$

For a detailed derivation see [6]. Each column of the matrix $A_s \in \mathbb{R}^{V \times K}$ contains the spatial pattern of activation corresponding to one canonical component. For better interpretability the patterns were related to psychological concepts using the *decode* function of the online database neurosynth (version 0.3.0 dev) [7]. In an automatized and unbiased manner, this function assesses the spatial similarity between an input image and all concept-based meta-analysis maps in its database. If different stimulus conditions are available, such as repeated presentations of the same stimulus with and without stereoscopic depth, subtle differences in activation patterns can be found using standard mass-univariate t-tests.

F. Classification of stimulus condition by CISCs

In order to investigate what information about the stimulus is contained in the intersubject correlations, we decoded the stimulus condition from the correlations in common networks. If we were to investigate one component at a time, we could employ a standard univariate test, but it is more likely that the relevant information is spread across common brain networks. We predicted stimulus condition (2D or 3D) from CISC values in a leave-one-movie-out cross-validation. For each movie, we trained a regularized linear discriminant classifier (LDA) on the CISC values of the 10 most strongly synchronized brain networks computed during all but one movie. LDA finds the normal vector $w_{LDA} \in \mathbb{R}^K$ of a linear decision boundary by

$$w_{LDA} = ((1-\lambda)S + \lambda\nu I)^{-1}(\mu_+ - \mu_-) \qquad (4)$$

where μ_+ and μ_- are the means of the positive and negative class, respectively, S is the sum of the within-class covariance matrices, λ is a regularization parameter that is estimated using the analytical solution provided by [9] and ν is the average eigenvalue. For the prediction of stimulus conditions, the positive class was the 3D condition and the negative class was the 2D condition. In order to obtain continous accuracy values we multiplied the LDA output values by their true labels, such that high positive values indicate correct predictions and negative values incorrect predictions. These outputs can then be averaged across all stimulus conditions and all stimulus repetitions.

III. Results

Inspection of the activation patterns for the top 5 canonical components, computed by eq. 3, showed that the common brain networks are primarily located in sensory and particularly in visual regions – as expected for movie stimuli (without audio). Also the correlations of these patterns with

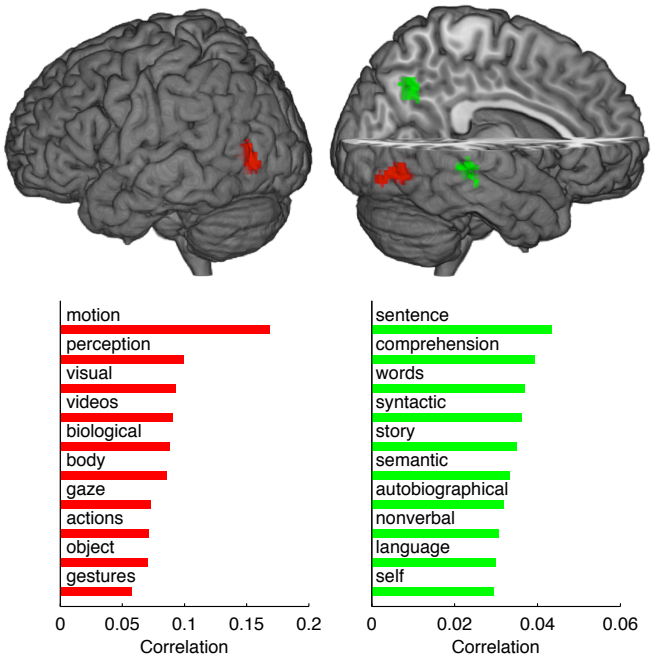

Fig. 2. Clusters of significant contrast (3D>2D) between activation patterns of canonical components computed on data (from experiment 1) of the 3D and 2D condition for first (red) and second (green) canonical component. Patterns were compared in paired t-tests using SPM8. Results were thresholded at $p<.005$ and corrected for multiple comparisons (resulting in a whole-brain correction threshold of $p<.05$) by determining individual cluster extent k thresholds with the calculated intrinsic smoothness of the individual T-value image, a cluster connection radius of 3 mm, and a 1000-iteration Monte Carlo simulation, using AlphaSim as implemented in the REST Toolbox 1.8 (http://www.restfmri.net/). Using the decode function of neurosynth [7] we recovered the most common psychological concepts associated with the contrast activation patterns.

mentions of psychological concepts in the literature (retrieved by neurosynth) reflect primarily visual functions (*visual*: 0.55, *object*: 0.43, *motion*: 0.28, *shape*: 0.27). However this activation pattern is not unspecific. The differential contrast (3D>2D) between the activation patterns obtained from the first two canonical components in the two stimulus conditions shows significantly higher activations in the 3D condition, see figure 2. We investigated how the strength of intersubject correlations is related to single movie scenes by decoding the stimulus condition (3D or 2D) from the CISC values[1]. An example of this decoding approach is presented in figure 3, the output of the LDA decoder (see eq. 4) multiplied by the label $y \in \{-1, 1\}$ (indicating the stimulus condition) is computed in sliding windows of 15 s (or 5 volumes) and aligned with the movie scenes. This particular stimulus featured a mantis shrimp, a colorful animal with remarkable characteristics[2], that none of our subjects had seen before, and an octopus being attacked by the shrimp. In the scenes in which the mantis shrimp appears for the first time (around 20 seconds after movie onset) and when the shrimp stands up to fight the

[1]The drawback of this approach is that the CISCs might decrease with repeated movie presentations [4] – we did not find strong evidence for this effect in our data but alternative stimulus presentation strategies could counteract this effect, e.g. manipulating the stimulus condition *within* a movie.

[2]http://theoatmeal.com/comics/mantis_shrimp

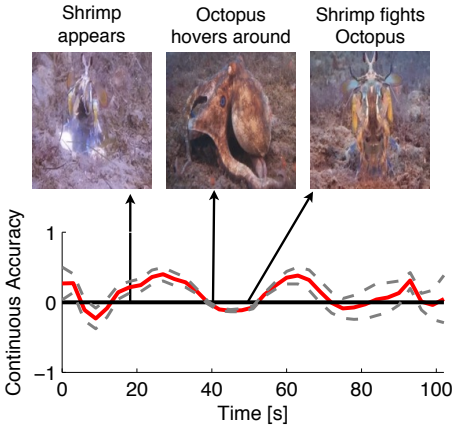

Fig. 3. Decoding accuracy of depth (from experiment 2), multiplied by stimulus label and averaged across condtions (shown are means±s.e.m.). Decoding was performed on CISCs in 10 canonical components during a movie showing a mantis shrimp fighting an octopus. No subject knew a mantis shrimp before the experiment. During its first appearance and the fight, decoding of depth cues from CISCs is significantly above chance.

octopus, the stimulus condition could be decoded reliably from the CISC values. An interesting topic of future research is to relate this increase in stimulus information to systematic changes in intersubject synchronization.

References

[1] U Hasson, Y Nir, I Levy, G Fuhrmann, and R Malach, "Intersubject synchronization of cortical activity during natural vision," *Science*, vol. 303, no. 5664, pp. 1634–40, 2004.

[2] A Bartels, S Zeki, and N K Logothetis, "Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain," *Cerebral Cortex*, vol. 18, no. 3, pp. 705–17, 2008.

[3] AG Huth, S Nishimoto, AT Vu, and JL Gallant, "A continuous semantic space describes the representation of thousands of object and action categories across the human brain," *Neuron*, vol. 76, no. 6, pp. 1210–24, 2012.

[4] JP Dmochowski, P Sajda, J Dias, and LC. Parra, "Components of ongoing eeg with high correlation point to emotionally-laden attention – a possible marker of engagement?," *Frontiers in Human Neuroscience*, vol. 6, no. 112, 2012.

[5] H Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3, pp. 321–377, 1936.

[6] S Haufe, F Meinecke, K Görgen, S Dähne, JD Haynes, B Blankertz, and F Biessmann, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *NeuroImage*, vol. 87, pp 96–110, 2013.

[7] T Yarkoni, RA Poldrack, TE Nichols, DC Van Essen, and TD Wager, "Large-scale automated synthesis of human functional neuroimaging data," *Nature Methods*, vol. 8, no. 8, pp. 665–70, 2011.

[8] JR Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[9] O Ledoit and M Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365 – 411, 2004.

Acknowledgements

This work was supported by the World Class University Program through the National Research Foundation of Korea, by the German Ministry of Education, Science, and Technology (R31-10008), by the BMBF project ALICE, "Automonous Learning in Complex Environments" (01IB10003B) and by the European Commission's Seventh Framework Programme FP7/2007-2013 (PlanetData, Grant 257641). FB was at Korea University while working on this project.