NeuroImage 61 (2012) 987–999

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Inferring consistent functional interaction patterns from natural stimulus FMRI data

Jiehuan Sun [a,e,1], Xintao Hu [b,1], Xiu Huang [a], Yang Liu [a,c], Kaiming Li [b,d], Xiang Li [d], Junwei Han [b], Lei Guo [b], Tianming Liu [d,*], Jing Zhang [a,*]

[a] Department of Statistics, Yale University, CT, USA
[b] School of Automation, Northwestern Polytechnic University, China
[c] Department of Computing, The Hong Kong Polytechnic University, HK, China
[d] Department of Computer Science and Bioimaging Research Center, The University of Georgia, Athens, GA, USA
[e] Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

ARTICLE INFO

Article history:
Accepted 13 January 2012
Available online 14 March 2012

Keywords:
Natural stimulus fMRI
DTI
Functional interaction

ABSTRACT

There has been increasing interest in how the human brain responds to natural stimulus such as video watching in the neuroimaging field. Along this direction, this paper presents our effort in inferring consistent and reproducible functional interaction patterns under natural stimulus of video watching among known functional brain *regions* identified by task-based fMRI. Then, we applied and compared four statistical approaches, including Bayesian network modeling with searching algorithms: greedy equivalence search (GES), *Peter and Clark (PC) analysis*, independent multiple greedy equivalence search (IMaGES), and the commonly used Granger causality analysis (GCA), to infer consistent and reproducible functional interaction patterns among these brain *regions*. It is interesting that a number of reliable and consistent functional interaction patterns were identified by the GES, PC and IMaGES algorithms in different participating subjects when they watched multiple video shots of the same semantic category. These interaction patterns are meaningful given current neuroscience knowledge and are reasonably reproducible across different brains and video shots. In particular, these consistent functional interaction patterns are supported by structural connections derived from diffusion tensor imaging (DTI) data, suggesting the structural underpinnings of consistent functional interactions. Our work demonstrates that specific consistent patterns of functional interactions among relevant brain *regions* might reflect the brain's fundamental mechanisms of online processing and comprehension of video messages.

Introduction

FMRI leverages the coupling between neural activity and hemodynamics in the human brain to obtain non-invasive measurement of brain activity (Logothetis, 2008; Friston, 2009). In the past few decades, fMRI has revolutionized how we study the function of the human brain (Fox and Raichle, 2007; Logothetis, 2008; Friston, 2009). Specifically, task-based fMRI has been widely used as a benchmark approach to localize functionally-specialized brain regions (Logothetis, 2008; Friston, 2009). Therefore, thousands of fMRI tasks have been reported in the literature to map those functionally-specialized brain *regions* (e.g., Laird et al., 2009). Recently, resting state fMRI has been increasingly used in the literature to map resting brain *regions* based on the premise that correlations between resting state fMRI time series originate from coherency in the underlying neural activation patterns of brain regions and reflect functional connectivity (e.g., Fox and Raichle, 2007; Cohen et al., 2008; van den Heuvel et al., 2008). In comparison, natural stimulus fMRI (Bartels and Zeki, 2004, 2005; Golland et al., 2007; Hasson et al., 2010) is relatively less studied due to more complexity and variability during fMRI experiment design and imaging sessions. Nevertheless, the advantage of natural stimulus fMRI under movie watching is that the human subjects are naturally engaged in the perception and cognition of the multimedia streams. This is more in an uncontrolled natural environment, which is well suited to study the functional interactions and dynamics among brain regions in response to multimedia stimuli. *In the neuroimaging field, there have been increasing interests in investigating how the human brain responds to the natural stimulus such as image/video watching and in studying if consistent response patterns exist across individuals* (Bartels and Zeki, 2004, 2005; Golland et al., 2007; Kay et al., 2008; Hasson et al., 2010; Haxby, 2010; Said et al., 2010; Nishimoto et al., 2011). Notably, a recent work in Hasson et al. (2004) demonstrated the effectiveness of applying natural stimulus fMRI in understanding the dynamics of cognitive systems during movie watching, and showed that different brains' fMRI-derived responses to the same movie could be similar. A recent review article

* Corresponding authors.
E-mail addresses: tliu@cs.uga.edu (T. Liu), jing.zhang.jz349@yale.edu (J. Zhang).
[1] Co-first authors.

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.01.142

on natural stimulus fMRI and the reliability of brain activity during natural stimulation is provided in Hasson et al. (2010).

In addition to using video/image stimulus to study the brain's function (e.g., Bartels and Zeki, 2004, 2005; Golland et al., 2007; Hasson et al., 2010; Haxby, 2010; Said et al., 2010; Nishimoto et al., 2011), there has been *increasing effort from* the video/image content analysis field *in applying brain imaging techniques as neuro-monitoring tools to extract high-level semantics features, thus guiding and facilitating video/image analysis* (e.g., Wang et al., 2009; Hu et al., 2010; Ji et al., 2011; Hu et al., 2012). For instance, in Wang et al. (2009), an electroencephalography (EEG)-based brain-machine interface was developed to guide image annotation and retrieval task. In this work, the rapid serial visual presentation (RSVP) paradigm (Gerson et al., 2006) was used for the presentation of images to a subject. Then, the recorded EEG signals were mapped as an "interest score" that meant to reflect how much of a user's attention was directed toward an image. Finally, the interest scores assigned to training images helped to propagate the labels to retrieve relevant images from a larger pool based on a graph pattern mining subsystem, and promising results were achieved (Wang et al., 2009) in contrast to the approach of using visual image representation alone. Recently, several studies took the advantage of fMRI to advance human semantics understanding and/or video/image analysis (Walther et al., 2009; Hu et al., 2010; Ji et al., 2011; Hu et al., 2012). The authors in Walther et al. (2009) studied a network of specific brain regions of interests (ROIs) that affects the human ability of discriminating visual scenes and demonstrated that brain activity from some ROIs can guide the selection of *interesting image frames*. The studies in Hu et al. (2010) and Hu et al. (2012) demonstrated that functional brain connectivity patterns derived from relevant networks can be used as high-level semantics features to guide the selection of low-level features for the purpose of video shot classification. The authors in Ji et al. (2011) extracted meaningful features from natural stimulus fMRI data to conduct video representation and retrieval in the brain imaging space, and demonstrated promising results.

It is clear that there has been growing interest from both brain imaging and image/video analysis fields in applying natural stimulus brain imaging of image/video watching to study the brain function (Bartels and Zeki, 2004, 2005; Golland et al., 2007; Hasson et al., 2010), *or to guide video/image content representation and analysis* (Wang et al., 2009; Hu et al., 2010; Ji et al., 2011; Hu et al., 2012). In the field, people have to rely on consistent and reproducible patterns, e.g., *functional or effective connectivities* (Murre and Goebel, 1996; Friston et al., 2003; Storkey et al., 2007; van de Ven et al., 2008; Chai et al., 2009), to infer meaningful information from natural stimulus fMRI data based on existing neuroscience knowledge (Hasson et al., 2004, 2010; Hu et al., 2010, 2012). Along this direction, this paper presents our effort in inferring consistent and reproducible functional interaction patterns among known functional brain regions under natural stimulus of movie watching. Given the substantial variability of structural localization of functionally-specialized brain regions, we used extensive task-based fMRI studies to identify brain regions involved in working memory, motor, visual and auditory regions in young healthy participants. Then, the same group of participants underwent natural stimulus fMRI while watching video clips selected from the TRECVID dataset (http://trecvid.nist.gov/), which is an international consortium providing benchmark video data for content-based video retrieval. At *this* stage, the video clips are semantically categorized into sports, weather report, and commercial advertisement that were defined by the large-scale concept ontology for multimedia (LSCOM) group (http://www.lscom.org/). Afterwards, the natural stimulus fMRI BOLD signals are extracted from the mapped functional brain *regions* in each individual subject for the purpose of computational modeling of functional interaction patterns. In particular, we applied and compared four statistical approaches, including Bayesian network modeling with searching algorithms: greedy equivalence search (GES) (Meek, 1997), *Peter and Clark* (PC) *analysis* (Spirtes and Glymour, 1991), independent multiple greedy equivalence search (IMaGES) (Ramsey et al., 2010), and the popularly used Granger causality analysis (GCA) (Granger, 1969), *to infer consistent and reproducible functional interaction patterns among the above-mentioned brain regions based on the extracted natural stimulus fMRI BOLD signals*. It is interesting that we found a number of reliable and consistent functional interaction patterns by the GES, PC and IMaGES algorithms in different subjects when they watched the same category of video clips, that is, one of sports, weather report, or commercial advertisement. These interaction patterns are meaningful given current neuroscience knowledge and are reasonably reproducible across different brains. In particular, the functionally strong interaction patterns are supported by structural connection patterns derived from tractography (Behrens et al., 2003) result of diffusion tensor imaging (DTI) data, suggesting the structural substrates of consistent functional interactions.

From computational and statistical perspectives, our experimental results suggest that Bayesian network modeling approach is more reliable and meaningful than the widely used GCA for the application of modeling functional interactions in natural stimulus fMRI data for the following reasons. Bayesian network methods consider all the ROIs for inference simultaneously and thus has lower false positive rate. *GCA takes into account the temporal lag information of the time series fMRI data, which makes its inference less accurate due to the influence of the relatively low temporal resolution of MRI data, the influence of the intervention of hemodynamics on the fMRI signals* (Roebroeck et al., 2005), *and the influence of complex temporal dynamics of effective connectivity*. In comparison, Bayesian network methods do not rely on the temporal information of fMRI signals and thus tend to be more robust.

Materials and methods

Overview

The flowchart of the proposed work is summarized in Fig. 1. In brief, we performed block design task-based fMRI (T-fMRI), which is widely considered as a reliable and readily available approach (Logothetis, 2008; Friston, 2009), to accurately localize the functional brain *regions* engaged in video comprehension including visual, auditory and language (A&L), motor, and working memory (WM) systems (Dudai, 2008) for each participating subject. Natural stimulus fMRI (N-fMRI) dataset was then acquired when the same group of subjects was watching video clips randomly selected from three semantic categories of sports, weather report, and commercial advertisement in the TRECVID dataset. After that, we applied and compared four statistical algorithms including GES (Meek, 1997), PC (Spirtes and Glymour, 1991), IMaGES (Ramsey et al., 2010) and GCA (Granger, 1969) to estimate the consistent functional interaction patterns in the localized functional brain *regions* based on the N-fMRI signals. Moreover, diffusion tensor imaging (DTI) dataset was also acquired and structural connectivities were inferred to independently assess the structural underpinnings of those consistent functional interaction patterns identified via N-fMRI data.

Data acquisition and preprocessing

Four healthy young adults were recruited at The University of Georgia (UGA) under IRB approval to participate in this study. MRI datasets were acquired in a GE 3T Signa HDx MRI system (GE Healthcare, Milwaukee, WI) using an 8-channel head coil at the UGA Bioimaging Research Center (BIRC). The multimodal DTI, T-fMRI and N-fMRI scans were performed in three separate scan sessions for each participating subject. DTI data was acquired using the isotropic spatial

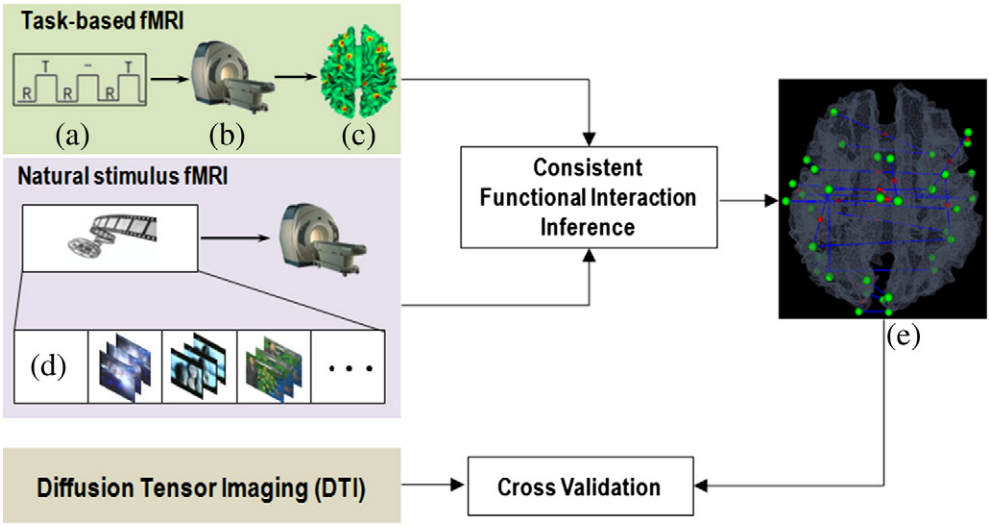

Fig. 1. The flowchart of our experimental design and computational pipeline. (a) Task paradigm. (b) fMRI scanner. (c) Functional brain regions derived from task-based fMRI data. (d) Video clips randomly selected from TRECVID. (e) Inferred consistent functional interactions during natural viewing of video clips. The interaction patterns are further cross-validated via structural connectivity derived from independent DTI dataset.

resolution 2 mm × 2 mm × 2 mm; parameters were TR 15.5 s and TE was min-full; b-value = 1000 with 30 DWI gradient directions and 3 B0 volumes acquired. DTI data preprocessing includes skull removal, motion correction and eddy current correction (Zhu et al., 2012). T1-weighted structural MRI data with 1 mm × 1 mm × 1 mm isotropic resolution was also acquired for each subject for anatomical reference.

The T-fMRI datasets were acquired to map the functionally-specialized brain regions including working memory, auditory and language, motor and visual systems. Our rationale is that the working memory, visual, auditory and language, and motor regions are among the most relevant brain systems that are involved in the perception and comprehension of video content (Dudai, 2008). Recently, a working memory theory of movie watching was developed in Dudai (2008). In brief, the working memory system combines online information (visual and aural inputs) with offline information (long-term memories) for the purpose of perception and cognition of movie watching. In this paper, we adopted this theory as the neuroscience base for modeling the brain's functional interactions in response to video stimuli, and hypothesize that consistent and meaningful interaction patterns can be inferred from the natural stimulus fMRI data.

Specifically, a modified version of the operation span (OSPAN) task (3 block types: OSPAN, Arithmetic, and Baseline) (Faraco et al., 2011) was performed to localize the working memory region. In the task, each run was preceded by visual instructions and contain 15 epochs; The OSPAN, Arithmetic, and Baseline conditions were structured so that subjects received similar amounts of visual input and gave the same amount of motor output. FMRI scans were acquired using a T2*-weighted single shot echo planar imaging (EPI) sequence and was aligned to the AC-PC line; TE = 25 ms, TR = 1500 ms, 90° RF pulse, 30 interleaved slices, acquisition matrix = 64 × 64, spacing = 0 mm, slice thickness = 4 mm, FOV = 240 × 240 mm, and ASSET factor = 2.

The visual network was mapped using a similar paradigm described in Xuan et al. (2007). The paradigm consists of cycled alternations between a fixation block and a stimulation block. During fixation block, one white dot is stationary on the black background, followed by the stimulation block in which a group of pictures are switching in a wedge-shaped area radiated from the white dot. Durations of the fixation and stimulation are both 15 s. A cycle consists of four fixation blocks and four stimulation blocks, among which the location of the white dot shifts in the order of 'left-bottom-right-top'. 144 volumes (6 cycles) of fMRI data were collected using the same parameters as above except with FOV covering only the visual cortex. The motor tasks were designed using similar paradigm in Meier et al. (2008), in which elbow, lip and ankle movements were performed respectively. The auditory/language tasks were similar to those in Fernández et al. (2001). The fMRI scan parameters used were the same as that in OSPAN T-fMRI. Notably, there are a variety of different task-based fMRI paradigms available in the literature (Laird et al., 2009) for each of the above mapped brain regions. In this work, we just used in-house verified fMRI tasks (e.g., Faraco et al., 2011) in UGA BIRC to map the working memory, visual, motor, auditory and language regions. A comprehensive examination of the activation patterns of these regions during various fMRI tasks is beyond the main focus of this paper. In addition, only those most consistent activation peaks that appear in all of the four participants' fMRI datasets were selected as functional ROIs for the following steps of functional interaction mapping, because accurate correspondence of the functional ROIs among different brains is essential for validity of the functional interaction inference.

The preprocessing of fMRI data included brain skull removal, motion correction, spatial smoothing, temporal prewhitening, slice time correction, and global drift removal (Lv et al., 2011). Then, the *general* linear model (GLM) implemented in FSL fMRI toolkits (Smith et al., 2004; Woolrich et al., 2009) was used to generate the activation map for the task-based fMRI for each subject and each task individually. Brain regions with the highest activation and high group-wise consistency were identified as the ROIs in the related brain regions via the similar methods in Zhu et al. (2012). *Specifically, we obtained group activation map through group-level fMRI data analysis, which was then linearly warped into each individual subject's space. Afterwards, we applied FSL FEAT to obtain an individual activation map for each subject. In particular, under the guidance of the initialized group-wise activation map (p-value: 0.05; multiple comparison correction: cluster-based thresholding using GRF theory,* Worsley et al., 1996*) in each individual brain, all of the consistently activated peaks that existed in both group-wise and individual maps were selected as the functional ROIs for each brain network, if they were within a neighborhood of 8 mm on the activation maps and shared similar anatomical locations on the MRI images. This procedure is similar to those in* Zhu et al. (2012) *and* Li et al. (2010). Totally, we identified 36 consistently activated brain ROIs as follows: 1) 16 ROIs in working memory region for each subject including left and right insula, left medial frontal gyrus,

left and right precentral gyrus, left and right paracingulate gyrus, left and right superior frontal gyrus, left and right supramarginal gyrus, left occipital pole, right frontal pole, right lateral occipital gyrus, left and right precuneus; 2) 6 ROIs in auditory/language region included Heschl's gyrus, *Wernicke's* area and Broca's area on both brain hemispheres; 3) 6 ROIs in motor regions that are activated in response to elbow, lip and ankle movements; 4) 8 ROIs in the visual region included the primary visual cortex (V1), the secondary visual cortex (V2), the middle temporal cortex (MT) and the middle superior temporal cortex (MST) on both brain hemispheres. Fig. 2 gives an illustration of the identified brain ROIs and regions for a subject. The names of these 36 ROIs are listed in Table 1.

To perform natural stimulus fMRI, we randomly selected 51 shots that are in the semantic categories of sports, weather report and commercial advertisement (20 sports, 19 weather reports and 12 commercial advertisements, respectively) from the TRECVID database and composed them into 8 clips. Each clip is about 11 min long. Fig. 3 shows screenshot examples of those video clips in each category. These clips were presented to the four subjects during N-fMRI scan via MRI-compatible goggles. Around 480 brain fMRI volumes were collected in each natural stimulus fMRI session, and the scan parameters were as follows: 30 axial slices, matrix size 64×64, 4 mm slice thickness, 220 mm FOV, TR = 1.5 s, TE = 25 ms, ASSET = 2. The strict synchronization between movie viewing and fMRI scan was achieved via the E-prime software (Schneider et al., 2002).

Statistical methods for functional interaction inference

In general, we use graph nodes to represent brain ROIs and infer the functional interactions among graph nodes based on the N-fMRI BOLD signals extracted for these ROIs. In the literature, directed graphs are commonly adopted to represent *functional* connectivity, among brain regions (Friston et al., 2003). *Here, directed graphs are used to model functional interactions. This methodology provides candidate models that, under certain additional assumptions, imply functional interaction relationships. The assumptions* (Neapolitan, 2004) *include directed acyclic graph and the Markov condition.* Specifically, each graph is comprised of a set of nodes (each represents a ROI here) and a set of edges. If there exists a directed edge going from node i to node j, $N_i \rightarrow N_j$, then the N_j is called the child of N_i and thus N_i is the parent of N_j. For a sequence of nodes (N_1,N_2,…,N_k), if there exists a directed edge between N_i and N_{i+1}, for each $i = 1,2,\ldots,k-1$, then the sequence of nodes is called a path. Here, k represents the total number of ROIs we used in the computation, and thus each node represents one ROI in the following paragraphs.

Specifically, we used *three* different search algorithms, GES (Meek, 1997) and PC, which are two methods for searching Bayesian network structure, and an extended version of GES (IMaGES) (Ramsey et al., 2010) for the following functional interaction inference. Intuitively, Bayesian network posits a joint probability distribution on nodes of a directed graph so that it could represent probabilistic relations among those nodes. *The directed graphs in Bayesian networks methods we used here are confined to acyclic graphs, i.e., directed acyclic graph (DAG), which could satisfy the causal condition* (Glymour, 2003; Spirtes et al., 2000), and thus Markov

Table 1
The list of nodes in the identified functional brain *regions*.

Working memory		Auditory and language	
1	Left insula	1	Left Heschl's gyrus
2	Right insula	2	Right Heschl's gyrus
3	Left medial frontal gyrus	3	Left *Wernicke's* area
4	Left precentral gyrus	4	Right *Wernicke's* area
5	Right precentral gyrus	5	Left Broca's area
6	Left paracingulate gyrus	6	Right Broca's area
7	Right paracingulate gyrus		
8	Left superior frontal gyrus	Visual	
9	Right superior frontal gyrus	1	Left primary visual cortex
10	Left supramarginal gyrus	2	Right primary visual cortex
11	Right supramarginal gyrus	3	Left secondary visual cortex
12	Left occipital pole	4	Right secondary visual cortex
13	Right frontal pole	5	Left middle temporal
14	Right lateral occipital gyrus	6	Right middle temporal
15	Left precuneus	7	Left medial superior temporal cortex
16	Right precuneus	8	Right medial superior temporal cortex
Motor			
1	Left primary motor cortex	4	Right primary motor cortex
2	Left supplementary motor area	5	Right supplementary motor area
3	Left premotor cortex	6	Right premotor cortex

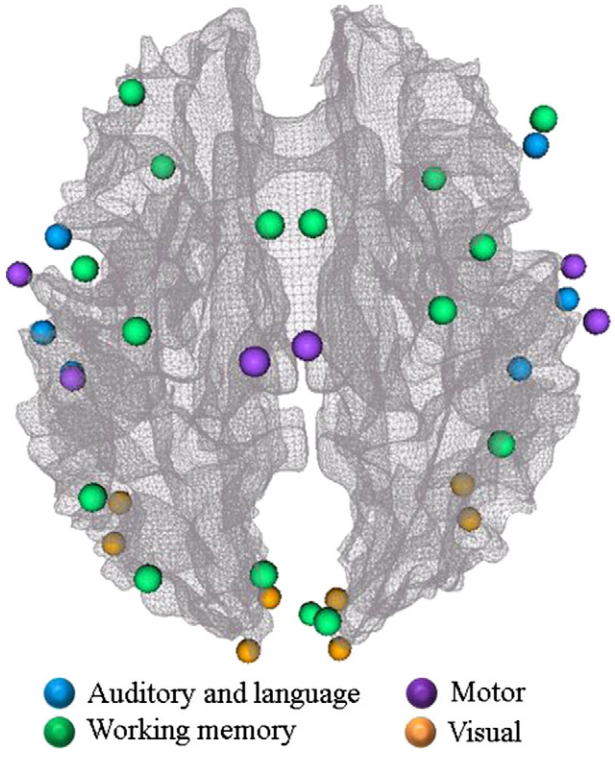

Fig. 2. Illustration of the identified brain ROIs and *regions* via task-based fMRI. It is noted that the cortical surface used in this rendering was the white matter/gray matter boundary. Thus some of the ROIs in this figure are slightly outside the cortical surface.

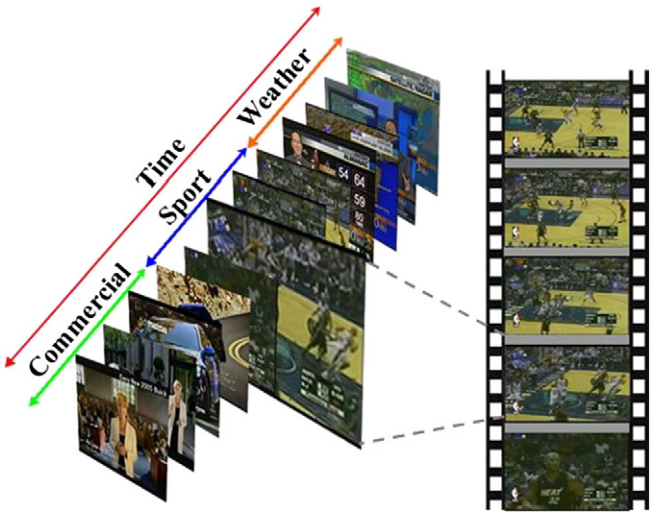

Fig. 3. Illustration of selected video clips from TRECVID 2005. 8 clips composed of 51 video shots were used as the natural stimuli.

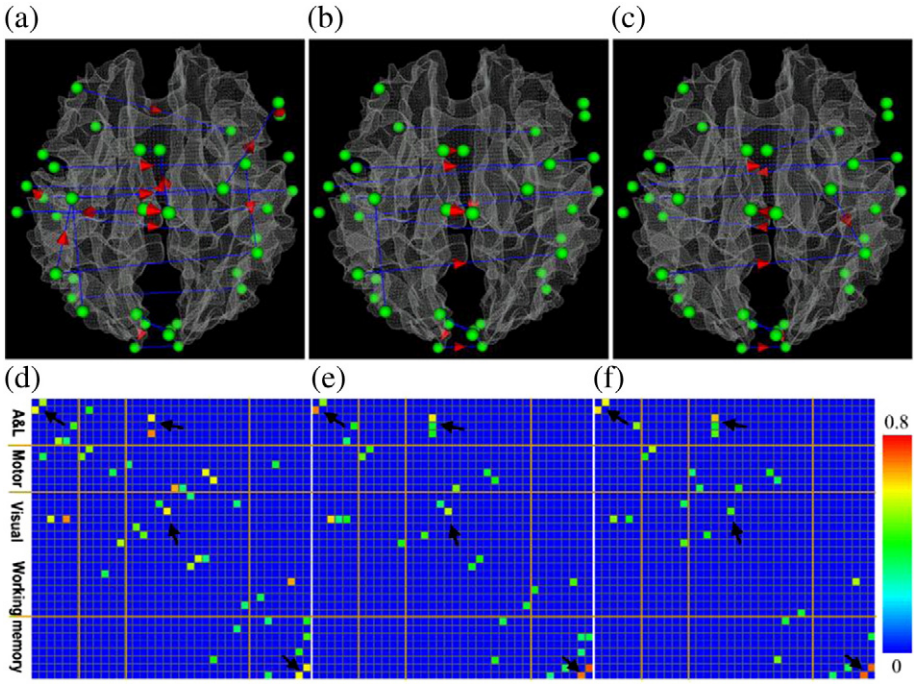

Fig. 4. Visualization of the consistent interaction patterns extracted via GES algorithm (pd = 1) in the functional brain *regions* include auditory/language, motor, working memory and visual systems when subjects watched video clips. Panels (a), (b) and (c) show the consistent connectivity patterns for video categories of advertisement, sports and weather report, respectively. The nodes of the functional networks are represented as green spheres. The blue lines without arrows and with red arrows represent non-directional and directional connections, respectively. Panels (d), (e) and (f) show the significance of consistency of the interaction connections demonstrated in panels (a), (b) and (c), respectively. The color bar is on the right side.

factorization of the distribution works, i.e., the joint distribution of the nodes in a DAG is equal to the product of probability of individual node conditional on their parent nodes.

In GES (Meek, 1997), which is a search and score approach, we score each DAG by using the Bayesian Information Criterion (BIC) (Schwarz, 1978) *under the normal distribution assumption on the data*:

$$-2\ln(ML) + k\ln(n),$$

where *ML* is the maximum likelihood estimate, *k*is the dimension of the model (i.e. total number of free parameters) and *n* is the sample size (i.e. the number of time points in each fMRI signal we used). *Other assumptions* (Neapolitan, 2004) *are: we assume the observed data are from a joint probability P (multivariate Gaussian in our case), and (G, P) satisfies the Markov condition where G is a DAG.* We then define the DAG with the best score as the most consistent model given the data. However, the total number of different DAG's increases exponentially with the increase of nodes in the data, and thus searching over the whole space of DAG's becomes computationally inefficient. Markov Equivalence Class, defined as a collection of DAG's whose factorization of joint distribution is mathematically equivalent, helps to solve the dimension problem of the space of DAG's. Since the score used to quantify each model only depends on ML, *k* and *n*, where the joint distribution determines the ML and DAG determines the *k* and *n*, the DAG's in the Markov equivalence class will have the same score and thus we only need to search over the space of Markov equivalence classes, which reduces the dimension of the searching space to a great *extent*.

The basic algorithm of GES is briefly discussed below. GES starts with an empty graph and then search forwards over the space of Markov equivalence classes with one additional edge each time until the model is not improved (i.e. the BIC does not change or become worse) by adding a new edge. Then, GES continues by searching backwards (i.e. with one edge removed each time) until no improvement by removing an edge. To find the consistent pattern across multiple datasets, we use IMaGES (Ramsey et al., 2010), which extends the GES algorithm by generalizing the BIC score to IMscore:

$$IM\ \ score = -\left(\frac{2}{m}\right)\sum_i \ln(L(D_i, G)) + ck\ln(n)$$

where *m* is the number of datasets (D_i represents dataset *i* in our case and *m* represents the number of total subjects used in the computation), *G is the graph the algorithm is scoring over,* $\ln(L(D_i, G))$ *is natural log of the likelihood of the dataset* D_i *calculated by using the maximum likelihood estimate of the free parameters in graph G, k is the dimension of the model and n is the sample size and c is the penalty parameter. IMaGES is run with a range of penalty parameters until the generated graph has no triangles and the penalty parameter used to produce this graph is the penalty parameter we use in calculating the IMscore, since triangulated nodes can contain a false positive edge, due to the reason that the measurements are noisy effects of the neural processes in the fMRI studies, whose functional interaction relations we are trying to estimate* (Ramsey et al., 2010).

Another main approach of learning graph from data we used here is PC algorithm (Spirtes and Glymour, 1991), which is a constraint-based approach and thus does not use the BIC score. PC algorithm is comprised of two steps: finding the skeleton and orienting the edges. The skeleton of a graph is the graph with only undirected edges, and thus the graph becomes a DAG after orienting the edges. PC begins with the undirected complete graph (i.e. every pair of nodes in the graph has an edge) and iteratively removes the edges based on the tests of independence or conditional independence with the pre-specified type one error rate alpha. Since the data used here is continuous *and we assume normality on the data*, the test of independence or conditional independence is equal to test of zero

correlation or zero partial correlation respectively. Remove the edge between node N_1 and node N_2, if the tests of conditional independence of N_1 and N_2 conditional on all subsets of V with cardinality of n fail to reject the null hypothesis until the cardinality of V is less than n, where n is increased by 1 each time unless the cardinality of V is less than n or one test rejects the null hypothesis, and V is the set of nodes that satisfy two conditions (1. There exists an edge between the node and N_1 or N_2; 2. The node is on the path between N_1 and N_2). After finding the skeleton of the graph, orientation of the undirected edges in the graph will be conducted. For each triple of nodes (N_1, N_2, N_3) with N_1 adjacent to N_2 (i.e. there exists an edge between N_1 and N_2) and N_2 adjacent to N_3 but no edge existing between N_2 and N_3, the edge in N_1–N_2–N_3 should be orientated as $N_1 \rightarrow N_2 \leftarrow N_3$, if and only if N_1 and N_3 are dependent on all subsets containing N_2 of V (V is obtained in the same way as above, except it is based on the skeleton obtained in the first step).

For comparison purpose, we applied Granger causality analysis (GCA) on the same dataset to infer the causalities among brain ROIs. Proposed by Clive Granger in 1969 (Granger, 1969), GCA investigates the causality from time series X to time series Y (in this study, the fMRI data from each pair of ROIs), and has been applied in brain imaging field in recent years (e.g., Roebroeck et al., 2005). Briefly, GCA first performs auto-regression of X with its P-order lagged values:

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + e1_t$$

to obtain the prediction errors e1, which is a $P \times 1$ vector. Then it adds the lagged values of time series Y into the regression to get prediction errors e2, which is also a $P \times 1$ vector:

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{i=1}^{p} b_i Y_{t-i} + e2_t$$

As the new regression term Y has been added, the change of prediction error from e1 to e2 could indicate the Granger-causal influence from Y to X, and an F-statistics is constructed accordingly:

$$F_{Y \rightarrow X} = \frac{\mathrm{var}(e1) - \mathrm{var}(e2)}{\mathrm{var}(e2)}$$

When there is no causality caused by Y to X, the value of F will approach zero since the additional Y terms will not influence the explanation power. While if the value is greater than the given threshold, we will reject the null hypothesis and add that Granger-causal edge to the result DAG. The Granger causality from X to Y could be obtained similarly. By repeating the pairwise analysis for all the ROIs, we could obtain the final DAG representing the Granger-causal connectivity of the whole brain.

*In this paper, we used the TETRAD software (*http://www.phil.cmu.edu/projects/tetrad/*) to apply the GES, IMaGES and PC algorithms. Also, it is worth noting that there are undirected edges in the output graphs generated by GES, IMaGES and PC, which do not exist in the GCA model. The undirected edge ($N_1 - N_2$) is the hypothesis that N_1 causes N_2 ($N_1 \rightarrow N_2$) or N_2 causes N_1 ($N_1 \leftarrow N_2$), but the algorithms cannot determine which is better.*

Results

Based on the datasets in Data acquisition and preprocessing section and statistical methods in Statistical methods for functional interaction inference section, we performed four experiments to demonstrate that consistent and meaningful functional interaction patterns can be inferred from natural stimulus fMRI data under movie watching. *In the PC algorithm, we used the type one error 0.01 as the threshold. In the GES algorithm, we use the penalty parameter $c = 1$. These are the meanings of PC = 0.01 and GES = 1 labels in the following sections.*

Consistent functional interaction patterns for three video categories

The identified functional connectivity patterns using the GES algorithm are shown in Fig. 4. In each scanned TRECVID video category, we estimated the functional interactions for all of the video clips for 4 subjects. In total, we performed 80, 76 and 48 estimations for sport, weather report and commercial advertisement, respectively. The number of the occurrence of a given interaction connection was counted for each video category separately. In each video category, the consistency of an interaction connection between any possible ROI pair was measured as the ratio between the number of occurrence and the total number of estimations in this video category. Fig. 4 shows the interaction connections with consistency above 0.35. In Figs. 4(a), (b) and (c) show the consistent interaction patterns for video categories of advertisement, sports and weather report, respectively. The nodes in the functional brain *regions* are represented by green spheres. The blue lines without arrows and with red arrows represent non-directional and directional interaction connections, respectively. In Figs. 4(d), (e) and (f) show the consistency (*measured by the percentage of occurrences of a specific edge over all scanned video shots among all of the 4 subjects*) of the interactions demonstrated in Figs. 4(a), (b) and (c), respectively.

It is interesting that for each video category, the GES method can identify reasonably consistent and meaningful interaction patterns. For instance, the working memory *region* has more functional interactions in the advertisement video category, meaning that video content in this category might need more online information processing during the video watching. *This result is consistent with the working memory theory of movie watching proposed in Dudai (2008), that is, the working memory system is responsible for online information processing for the perception and cognition of movie watching.* Examination of the videos of three categories (see supplemental example video at: http://www.cs.uga.edu/~tliu/TRECVID/positive_commercial_7.mpg) suggests that advertisement videos indeed have more complex content, more rapidly changing shots, and thus demand more online information processing. In contrast, the weather report videos (see supplemental example video at: http://www.cs.uga.edu/~tliu/TRECVID/positive_weather_5.mpg) contain simpler content, have less motion and video shot changes, and therefore demand less online information processing. Hence, the results in Figs. 4a–f are reasonable considering that working memory system is the brain hub for online information processing that interacts with online visual and auditory information processing (Dudai, 2008). *Also, it is reasonable to observe that functional interactions with the visual regions are stronger in advertisement and sports (see supplemental example video at:* http://www.cs.uga.edu/~tliu/TRECVID/positive_sport_9.mpg*) videos than those in weather report videos, as the weather report video contains much less motion.* Finally, quite a few interaction items are consistent and reproducible in multiple categories of videos, such as those highlighted by corresponding black arrows in Figs. 4d–f, and their consistencies are quite high (e.g., in orange or yellow colors). This result suggests that these common interaction patterns might be fundamental for online processing and comprehension of three different types of video messages. *We performed additional statistical analysis of the GES outputs. Specifically, we calculated the within-video variation and between-video variation. The ratio of within-video variation over between-video variation is approximately 13.5 and this ratio is quite consistent across video categories.*

Similarly, the identified consistent functional interaction patterns using PC algorithm is shown in Fig. 5. The definition of consistency of the identified functional interaction patterns is the same as that in Fig. 4. Also, only those interaction connections with consistency above 0.35 are shown in Fig. 5. It is interesting that most of consistent

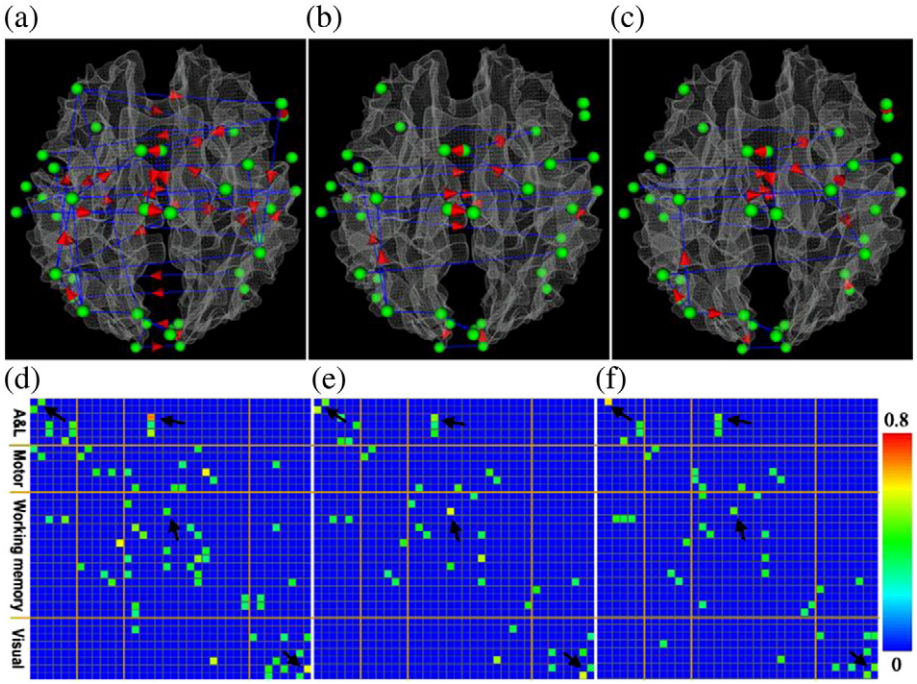

Fig. 5. Visualization of the consistent connectivity patterns inferred via PC algorithm (pc = 0.01) in the functional brain *regions* include auditory and language, motor, working memory and visual systems when subjects watched video clips. Panels (a), (b) and (c) show the consistent functional interaction patterns for video categories of advertisement, sports and weather reports, respectively. The nodes of the functional networks are represented as green spheres. The blue lines without arrows and with red arrows represent non-directional and directional interaction connections, respectively. Panels (d), (e) and (f) show the significance of consistency of the interaction connections demonstrated in panels (a), (b) and (c), respectively. The color bar is on the right side.

functional interaction patterns identified in Fig. 4 are well reproduced in this experiment by using the PC algorithm. For instance, most of the functional interaction items highlighted by black arrows in Fig. 4 are replicated in Fig. 5. The similar patterns of stronger working memory and visual interactions in advertisement and sports videos than those in weather report videos are also replicated in Fig. 5. This result suggests that there are truly reproducible and consistent meaningful functional interaction patterns among the relevant functional brain *regions* when viewing natural stimulus videos. We premise that those consistent functional interaction patterns across three video categories reflect the common functional architecture of the human brain, while the functional interaction pattern differences demonstrate the variation of the individual brain's responses to different content and/or types of videos. *Similarly, we calculated the within-video variation and between-video variation in PC outputs. The ratio of within-video variation over between-video variation is approximately 15.3 and this ratio is quite consistent across video categories.*

IMaGES takes multiple datasets into *consideration* and will find only one graph, which tells whether each edge (interaction connection) occurs or not. Therefore, the result for IMaGES has only one graph for all the video data in each category and the result is shown in Fig. 6. It is evident that the results obtained by IMaGES are *quite* similar to those by the GES method shown in Fig. 5, further demonstrating that consistent functional interaction patterns can be inferred from natural stimulus fMRI data using different statistical methods. *Notably, for weather report/commercial videos, the IMaGES-derived graphs are much denser compared to the graphs obtained by PC or GES methods. Specifically, the numbers of edges by the IMaGES method for weather report/commercial videos are 52/51, while the numbers of edges are 22/41 and 17/29 by GES and PC methods respectively. This could be attributable to the fact that the IMaGES method effectively multiplies the sample size by 4 (we have 4 subjects here), and thus weaker connections should be identified. This is consistent with the fact that the agreement of IMaGES with diffusion tensor imaging results in* Fig. 9

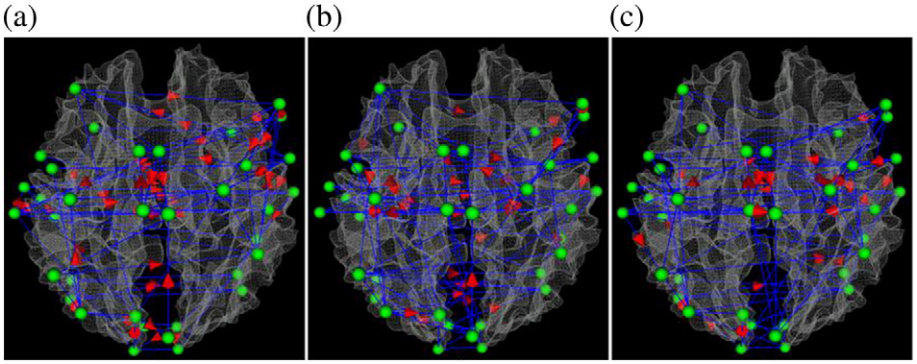

Fig. 6. Visualization of the consistent functional interaction patterns inferred via the IMaGES algorithm in the functional brain *regions* include auditory and language, motor, working memory and visual systems when subjects watched video clips. Panels (a), (b) and (c) show the consistent interaction patterns for video categories of advertisement, sports and weather reports, respectively.

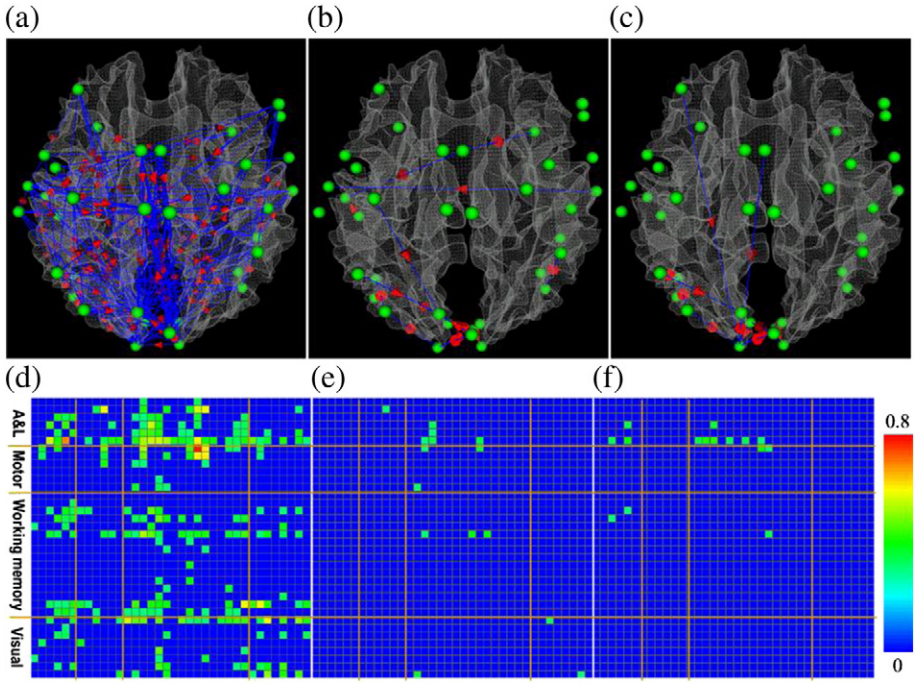

Fig. 7. Visualization of the consistent interaction patterns inferred via GCA in the functional brain networks include auditory/language, motor, working memory and visual systems when subjects watched video clips. Panels (a), (b) and (c) show the consistent interaction patterns for video categories of advertisement, sports and weather report, respectively. Panels (d), (e) and (f) show the significance of consistency of the interaction connections demonstrated in panels (a), (b) and (c), respectively. *Specifically, we first used a threshold to remove those inconsistent edges (p-value > 0.01). That is, the blue cells in panels d–f represent inconsistent edges. Then, we visualized the frequency of occurrence in the panels d–f for those consistent edges. The color of frequency of edge occurrence is color-coded by the color bar on the right.*

is much better at the lower threshold (50) for structural connections. Also, the similar functional interaction patterns for three categories of videos revealed in Figs. 4–5 are replicated here via the IMaGES algorithm.

Comparison with GCA

To compare the GES, PC and IMaGES algorithms with GCA, Fig. 7 shows the consistent functional interaction patterns obtained by GCA for three video categories respectively. It is apparent that the results inferred by GCA are quite different compared to those by GES, PC and IMaGES algorithms. Actually, the results by GES and PC algorithms are much more consistent within and across video categories, and Fig. 8 provides quantitative comparisons of the significance of functional interaction consistence among GCA, GES and PC algorithms. It is evident that the GES and PC algorithms *provide more consistent results* than the GCA method, in particular for the sports and weather report video categories. Specifically, for each video category in each subject, we conducted GCA, GES and PC respectively. Then, for each algorithm and each video category, we summarized the proportion of all the interaction connections that appear in each subject over all the subjects, and selected the interaction connections whose proportion is higher than 35%, which is the same as those used in Figs. 4–5. As we can see from the histograms (particularly Figs. 8b and c), GES and PC algorithms produce more consistent interaction connections than that of GCA, e.g., most of the histogram bins with higher consistency have green (GES) and blue (PC) bars. It is noted that for the commercial video category, all of GES, PC and GCA produced many more interaction connections (Fig. 8a), and the GCA method produced even more connections than GES and PC methods. Our interpretation is that there are more rapidly changing shots in the commercial videos, and the GES and PC methods tend to be more conservative in inferring the interaction patterns as they consider all the ROIs for inference simultaneously. Since the GCA method only considers a pair of two ROIs and does not consider the whole network, this method tends to have more false positives. Another possibility is that the length of fMRI scan data for the commercial video category is 64% larger than the other two video categories on average, and the multiple testing problem in the identification of significant causality in GCA could produce more false positives from the commercial video category. The above interpretations will be further verified by independent, external evaluation via DTI data in Structural connectivity between functionally interacting brain regions section.

> *In addition to the above pairwise GCA analyses between functional ROIs, in this paper, we also applied the conditional Granger Causality Analysis among those ROIs in our datasets, to examine possible multivariate relationships. We used the toolbox implemented by Seth in MATLAB* (Seth, 2010) *to perform the conditional GCA analysis. Our results showed that very few consistent edges could be detected by conditional GCA. For instance, 4 subjects watched a total of 30 weather video shots, and only one specific connection was detected by the conditional Granger Causality Analysis 1 time for all the 30 video shots.*

Also, the functional interaction patterns inferred by GES, PC and IMaGES algorithms have better structural connectivity support, which will be shown in Fig. 9. Based on the above results, we hypothesize that GES, PC and IMaGES algorithms outperform GCA in the application of inferring consistent functional interaction patterns during natural stimulus fMRI of video watching. From computational perspective, our interpretations are as follows. *GCA makes the inference of causality based on the temporal lags between time series fMRI data, which possibly makes its inference less accurate due to the relatively low temporal resolution of fMRI, the influence of the intervention of temporal hemodynamics on the fMRI signals* (Roebroeck et al., 2005), *and the complex temporal dynamics of effective connectivity between fMRI signals* (Li et al., 2011). *In comparison, Bayesian network methods do*

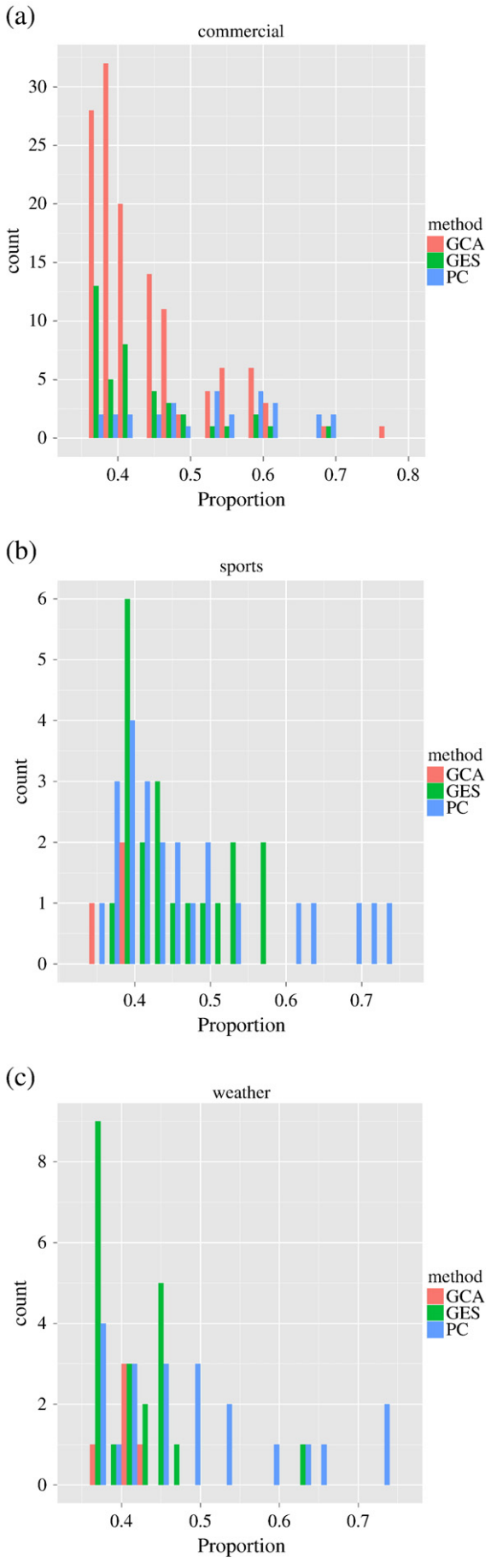

Fig. 8. Quantitative comparison of GES, PC, and GCA methods. (a)–(c): histograms of significances of consistency of the interaction connections obtained by three methods of GES (green bars), PC (blue bars) and GCA (pink bars) for three video categories, respectively. It is noted that the three histograms have different scales on the y-axis.

not rely on the precise temporal lag information and thus tend to be more robust.

Structural connectivity between functionally interacting brain regions

It is believed that the functional role of the cortical region of the human brain has its structural underpinnings (Passingham et al., 2002). In this paper, we measured the structural connectivity derived from DTI data to evaluate the structural substrate of consistent functional interactions identified from natural stimulus fMRI data by the proposed methods. Specifically, the stochastic DTI tractography algorithm (Behrens et al., 2003) implemented in FSL (Smith et al., 2004; Woolrich et al., 2009) was adopted to derive the structural connectivity from the DTI data for each subject individually. By specifying a seed ROI and a terminate ROI, the stochastic fiber tracking resulted in a volume image, in which the intensity of the voxel indicates the probability of the fibers penetrating the corresponding voxel. Fig. 9(a) shows an exemplar joint visualization of the cortical surface and the volume rendering of the resulted probability map. The above procedure was repeated for all possible ROI pairs in the identified *thirty-six* brain *regions* via task-based fMRI in Data acquisition and preprocessing section. Denote the probability volume for ROI pair R_i and R_j as I_{ij} and the structural connectivity matrix of the identified functional brain *region* as C. After applying a predefined threshold T on I_{ij}, the existence of structural connection between R_i and R_j was determined by a connectedness analysis described in (Shapiro, 1992). That is, $C(i, j) = 1$ when they were connected, otherwise $C(i, j) = 0$. We varied T from 0.01 to 1 with the step length of 0.01, resulted in 100 times of voting for each element in C. The element with higher number of votes indicates the stronger direct structural connection between the corresponding ROI pair.

The measured structural connectivity derived from DTI data using the above-mentioned stochastic fiber tracking is shown in Fig. 9b for a randomly selected subject. A threshold is applied on the structural connectivity matrix to remove the non-significant elements. The consistency between the structural connections and the inferred functional interactions is measured by the ratio of the functional connections with direct structural connection against all the identified consistent functional interactions. The ratios for GCA, PC, GES and IMaGES methods are summarized in Figs. 9c–d, in which (c) and (d) correspond to different structural connectivity thresholds of 50 and 70, respectively. It can be seen that the number of identified consistent functional interactions associated with structural connections is much higher in PC, GES and IMaGES compared with that in GCA. This result suggests that the functional interactions inferred by PC, GES and IMaGES can achieve more structural connectivity support from *independent* DTI data. *Considering that structural connection is closely correlated with functional connectivity* (Honey et al., 2009), *our results here indicate the superiority of the proposed Bayesian methods for functional interaction inference in natural stimulus fMRI.* This experiment of validation using structural connectivity was repeated for other 3 subjects and similar results were obtained.

Consistency of inferred functional interaction patterns across three video categories

To showcase the consistency of the inferred functional interaction patterns across different videos and different subjects, we ex-

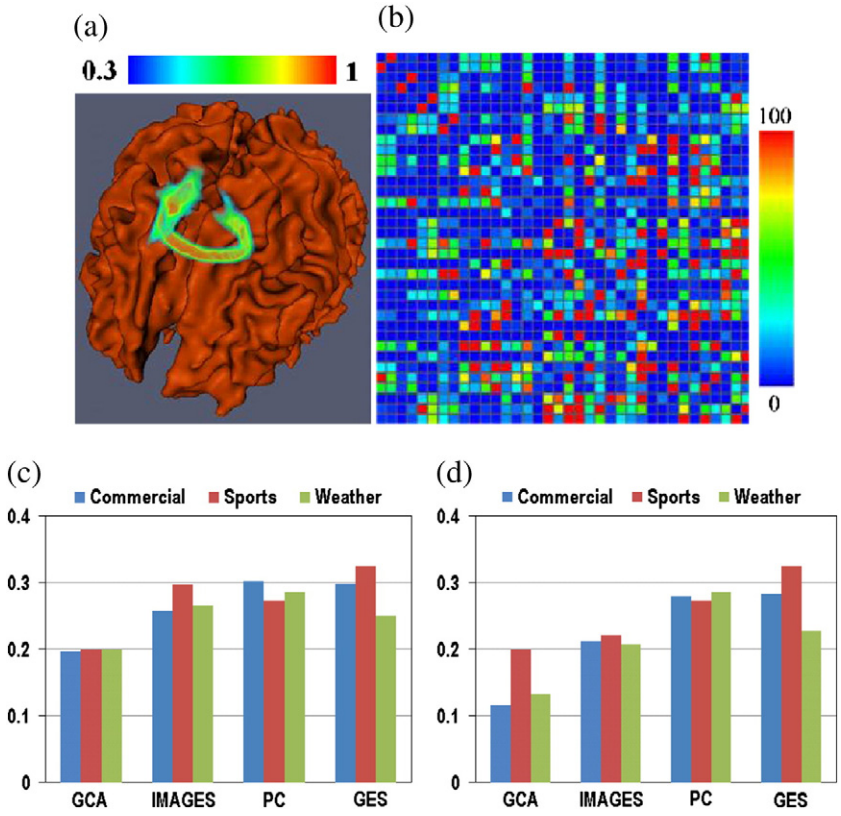

Fig. 9. (a) Joint visualization of the cortical surface and the probability map resulted from stochastic fiber tracking. (b) The structural connectivity matrix for one subject. (c)–(d): Comparison of the structural connectivities of the consistent functional interactions identified by different inference methods. The *x* axis is the index of methods. The *y*-axis is the ratio of the identified consistent interaction connections with direct structural connection against all the identified consistent functional connections. (c) Threshold of 50. (d) Threshold of 70.

amined the distributions of a few consistent functional interaction patterns. The consistent interaction means that it is inferred in most of the subjects when watching videos of the same semantic category, including weather report, sports and commercial advertisement. For instance, in the top subplot in Fig. 10a for weather report video, the title "right Heschl's gyrus to right *Wernicke's* area prob = 0.72, pc = 0.01" means that the interaction connection from the right Heschl's gyrus to right *Wernicke's* area is detected in 72% of the 72 samples by the PC algorithm. It is noted that one subject watching one video is considered as one sample, and totally we had 72 samples. In more details, for the same top subplot in Fig. 10a, all 4 subjects exhibited the same functional interaction connection when watching weather report video #5, #7, #9, and #12, which is 21% out of the total 19 videos. As an example, the video #5 can be viewed online at: http://www.cs.uga.edu/~tliu/TRECVID/positive_weather_5.mpg. Also, 3 out of 4 subjects showed the same functional interaction connection when watching video #1, #2, #3, #8, #10, #14, #15, #17, #18 and #19, which is 53% of the 19 videos. There is only one video (video #16) that activated functional interaction in only one subject. The video #16 is put online for viewing: http://www.cs.uga.edu/~tliu/TRECVID/negative_weather_16.mpg. It can be seen that this negative example of video shot (video #16) in the category of weather report is not a typical one as that in video #5 in the sense that there are more heterogeneous contents and it has much less interactive weather reports by the meteorologist. We reason that this might be the reason that video #16 induced less consistent responses by the four subjects.

A summary of the consistent patterns in Fig. 10a for other video categories is as follows. In the sports video category by PC algorithm, all four subjects exhibited the same functional interaction connection from the right *Wernicke's* area to the right Heschl's gyrus in 20% (#7, #9, #11, #12) of the 20 videos. As an example, the video #9 can be viewed online at: http://www.cs.uga.edu/~tliu/TRECVID/positive_sport_9.mpg. Also, 3 out of 4 subjects showed the same functional interaction connection in 40% of the 20 videos. Only one subject showed one interaction pattern in 10% of the 20 videos. For the commercial category by the PC algorithm, in 25% (#7, #10, #12) of the 12 videos, all subjects exhibited the same interaction pattern from the left occipital pole to the right primary visual cortex. The video #7 can be viewed online at: http://www.cs.uga.edu/~tliu/TRECVID/positive_commercial_7.mpg. 2 out of 3 subjects showed the same functional interaction connection in 58% of the 12 videos. In 17% of the 12 videos, only one subject has one interaction connection.

The results for GES algorithm are provided in Fig. 10b. Specifically, in 16% of the 19 weather report videos, all four subjects exhibited the same functional interaction from the left medial temporal to the left medial superior temporal. In 20% of the 20 sports videos, all subjects exhibited the same interaction from the right precuneus to the left precuneus. The sport video #15 induced only one subject to have the interaction pattern by the GES algorithm, and it is put online for reviewing: http://www.cs.uga.edu/~tliu/TRECVID/negative_sport_15.mpg. For GES results of commercial videos, in 33% of the 12 advertisement videos, all four subjects had the interaction from the left occipital pole to the left primary visual cortex.

In general, results obtained from both the PC and GES algorithms are consistent to some extent, as shown in Figs. 10a–b. In comparison, the results from PC are slightly more consistent than those from GES. These results further demonstrate that consistent and meaningful functional interaction patterns can be inferred from natural stimulus fMRI data when subject watched videos, despite the complexity and variability in fMRI scans and video stimuli.

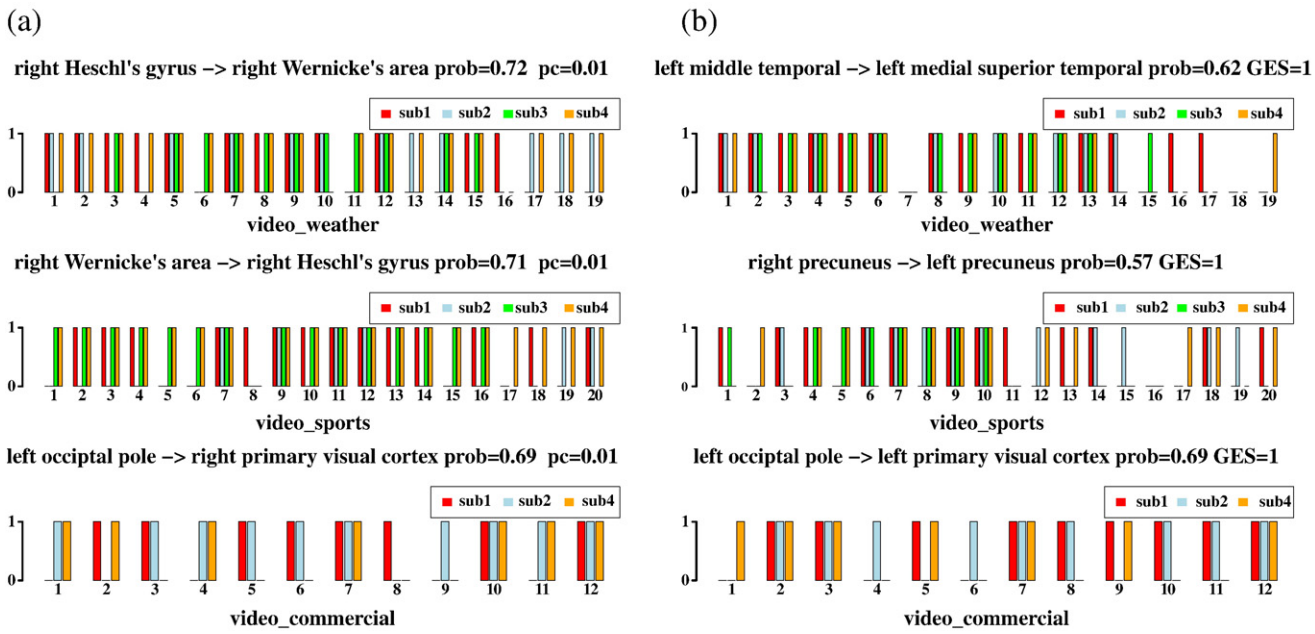

Fig. 10. The distributions of the most consistent interaction connections in each video category from the results of both PC (a) and GES (b) algorithms. For each subplot, the horizontal axis represents the index of video shots for each semantic category. The name of video category is at the bottom of each subplot, The ROIs' names for the interaction pair, probability of the occurrence of the interaction patter, and threshold used are at the top of the subplot. The black arrows represent the interaction directionality. Four subjects are represented by the colors of red, blue, green and orange, respectively. The horizontal axis stands for binary decision of existence (1) or non-existence (0). It is noted that for the video category of commercial advertisement, we only have datasets from three subjects. *The label pc = 0.01 means that the PC algorithm is run under the type one error 0.01 and the label GES = 1 means that we used the penalty parameter c = 1.*

Discussion and conclusion

Inferring meaningful information from natural stimulus fMRI data has been a challenging problem due to the complexity and variability during fMRI designs and scans and in natural video stimuli. In this work, we demonstrated that consistent and meaningful functional interaction patterns can be identified among relevant brain *regions* from different brains when they watched different videos shots of the same semantic category. These reproducible and consistent functional interaction patterns are consistent with current neuroscience knowledge, and might suggest the possible existence of a common functional brain mechanism that is responsible for video message comprehension. *For example, the results in* Figs. 4–7 *demonstrated that the working memory system plays a key role in online information processing for the perception and cognition of movie watching. This result is in agreement with the existing working memory theory of movie watching proposed in* Dudai (2008). *As another example, the inferred functional interaction patterns among the visual regions are reasonable, e.g., the interactions are stronger in advertisement and sports videos than those in weather report videos. Our structural connectivity analysis results from independent DTI data* (Fig. 9) *suggest the possible structural underpinnings of such consistent functional interaction patterns, as it has been reported that there is close relationship between structural and functional connectivities* (Honey et al., 2009). Our extensive quantitative comparison results demonstrated that the PC and GES algorithms outperform the traditionally used GCA method. Our interpretation is that Bayesian network methods consider all the ROIs for inference simultaneously and do not rely on the temporal information of noisy fMRI signals.

It should be noted that the directed graphs in Bayesian networks methods used here are confined to acyclic graphs. Our rationale is that this constraint could satisfy the causal condition (Glymour, 2003; Spirtes et al., 2000) *such that Markov factorization of the distribution works, i.e., the joint distribution of the nodes in an acyclic graphs is equal to the product of probability of individual node conditional on their parent nodes. Though this constraint facilitated the identification of the consistent functional interaction patterns in this paper, its possible limitations should be examined in the future, e.g., on complete sub-graphs of structurally-connected brain regions. Also, acyclic Bayes nets are a special case of graphical causal models, which include cyclic graphs, the linear versions of which satisfy a generalization of the Markov property, and for which there are search algorithms* (Richardson et al., 1997), *e.g., algorithms that were applied to fMRI data recently* (Ramsey et al., 2011).

In the future, we plan to perform larger scale natural stimulus fMRI scans with more subjects and more categories of video shots so that we can test and potentially replicated the findings in this paper in larger samples. Also, additional brain *regions* such as attention and emotion systems that might have important roles in video perception and cognition will be added into the functional interaction pattern analysis. Certainly, these additional functional networks have to be mapped via task-based fMRI to ensure accurate localization of the relevant brain regions. *Also, more methods for inferring functional network structures* (Smith et al., 2011) *will be compared in the future, and controls for multiple comparisons* (Li and Wang, 2009) *should be considered.* In the long-run, it would be very helpful to look into the functional interaction patterns among large-scale brain *regions* that cover the whole brain including both cortical and subcortical regions. However, accomplishment of such a goal entails quantitative representation of a common structural and functional human brain architecture in the future. We envision that large-scale inference of consistent functional interaction patterns from natural stimulus fMRI can provide fundamental insight into the functional architecture and working mechanism of the brain, because the brain is naturally engaged in the perception and cognition of the video streams under an uncontrolled environment.

At the same time, the inferred consistent functional interaction patterns from fMRI data reflect the brain's responses to natural *stimuli* of video to *a* certain extent. In the future, it might be useful

to apply those consistent functional interaction patterns to guide the selection of low-level video features (Hu et al., 2010; Ji et al., 2011; Hu et al., 2012) and the construction of computational models that are predictive of brain's responses. This fMRI-guided video content representation could potentially significantly bridge the semantic gaps in video analysis and retrieval field for years (Ji et al., 2012). Therefore, in summary, inference of consistent and meaningful functional interaction patterns from natural stimulus fMRI of video watching can only not contribute to the understanding of human brain function, but also advance digital video representation and analysis for human-centered video management and retrieval in the future.

Acknowledgments

T Liu was supported by the NIH Career Award (NIH EB 006878), NIH R01 HL087923-03S2 and The University of Georgia start-up research funding. J Zhang was supported by start-up funding from Yale University. X Hu was supported by the National Natural Science Foundation of China (61103061) and the Postdoctoral Foundation of China (20110490174). Y Liu was supported by Seessel Postdoctoral Fellowship from Yale University. The computation was done with the help from the Yale University Biomedical High Performance Computing Center, which was supported by the NIH grant RR19895. X Huang was supported by the China Scholarship Council-Yale World Scholars Program. *We would like to thank the anonymous reviewers for their constructive comments that have significantly helped to improve this paper.*

References

Bartels, Andreas, Zeki, Semir, 2004. Functional brain mapping during free viewing of natural scenes. Hum. Brain Mapp. 21, 75–85.
Bartels, Andreas, Zeki, Semir, 2005. Brain dynamics during natural viewing conditions—a new guide for mapping connectivity in vivo. NeuroImage 24, 339–349.
Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077–1088.
Chai, B., Walther, D.B., Beck, D.M., Fei-Fei, L., 2009. Exploring Functional Connectivity of the Human Brain using Multivariate Information Analysis. NIPS.
Cohen, Alexander L., Fair, Damien A., Dosenbach, Nico U.F., Miezin, Francis M., Dierker, Donna, Van Essen, David C., Schlaggar, Bradley L., Petersen, Steven E., 2008. Defining functional areas in individual human brains using resting functional connectivity MRI. NeuroImage 41 (1), 45–57.
Dudai, Y., 2008. Enslaving central executives: toward a brain theory of cinema. Projections 2 (2), 12–24.
Faraco, C.C., Unsworth, N., Lagnely, J., Terry, D., Li, K., Zhang, D., Liu, T., Miller, L.S., 2011. Complex span tasks and hippocampal recruitment during working memory. NeuroImage 55 (2), 773–787 (15).
Fernández, G., de Greiff, A., von Oertzen, J., Reuber, M., Lun, S., Klaver, P., Ruhlmann, J., Reul, J., Elger, C.E., 2001. Language mapping in less than 15 minutes: real-time functional MRI during routine clinical investigation. NeuroImage 14 (3), 585–594.
Fox, M.D., Raichle, M.E., 2007. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. 8, 700–711.
Friston, K.J., 2009. Modalities, modes, and models in functional neuroimaging. Science 326, 399–403.
Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modeling. NeuroImage 19, 1273–1302.
Gerson, A.D., Parra, L.C., Sajda, P., 2006. Cortically coupled computer vision for rapid image search. IEEE Trans. Neural. Syst. Rehabil. Eng. 14 (2), 174–179.
Glymour, C., 2003. The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology. MIT Press, New York.
Golland, Y., Bentin, S., Gelbard, H., Benjamini, Y., Heller, R., Nir, Y., Hasson, U., Malach, R., 2007. Extrinsic and intrinsic systems in the posterior cortex of the human brain revealed during natural sensory stimulation. Cereb. Cortex 17 (4), 766–777 (Apr).
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 414.
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640.
Hasson, U., Malach, R., Heeger, D.J., 2010. Reliability of cortical activity during natural stimulation. Trends Cogn. Sci. 14, 40–48.
Haxby, J.V., 2010. Multivariate analysis of fMRI data: high-dimension spaces for neural and cognitive representations. In: Hanson, S.J., Bunzl, M. (Eds.), Foundational Issues of Neuroimaging. MIT Press, Boston, pp. 55–68.
Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P., 2009. Predicting human resting-state functional connectivity from structural connectivity. PNAS 106 (6), 2035–2040.
Hu, X., Deng, F., Li, K., Zhang, T., Chen, H., Jiang, X., Lv, J., Zhu, D., Faraco, C., Zhang, D., Mesbah, A., Han, J., Hua, X., Xie, L., Miller, S., Guo, L., Liu, T., 2010. Bridging low-level features and high-level semantics via fMRI brain imaging for video classification. Proceedings of the international conference on Multimedia. ACM, Firenze, Italy, pp. 451–460.
Hu, X., Li, K., Han, J., Hua, X., Guo, L., Liu, T., 2012. Bridging the semantic gap via functional brain imaging. IEEE Trans. Multimed. 14 (2), 314–325.
Ji, Xiang, Han, Junwei, Hu, Xintao, Li, Kaiming, Deng, Fan, Fang, Jun, Guo, Lei, Liu, Tianming, 2011. Retrieving video shots in semantic brain imaging space using manifold-ranking. Proceeding of 18th IEEE International Conference on Image Processing (ICIP). Brussels, pp. 352–355. Belgium, September 11–14, 2011.
Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. Nature 452, 352–355.
Laird, A.R., Eickhoff, S.B., Kurth, F., Fox, P.M., Uecker, A.M., Turner, J.A., Robinson, J.L., Lancaster, J.L., Fox, P.T., 2009. ALE meta-analysis workflows via the brainmap database: progress towards a probabilistic functional brain atlas. Front. Neuroinformatics 3, 23.
Li, Junning, Wang, Z. Jane, 2009. Controlling the false discovery rate of the association/causality structure Learned with the PC algorithm. J. Mach. Learn. Res. 10, 475–514 (Feb).
Li, K., Guo, L., Faraco, C., Zhu, D., Deng, D., Zhang, T., Jiang, X., Zhang, D., Chen, H., Hu, H., L., Miller, S., Liu, T., 2010. Individualized ROI optimization via maximization of group-wise consistency of structural and functional profiles. Adv. Neural Info. Proc. Syst. 23.
Li, Xiang, Lim, Chuwoo, Li, Kaiming, Guo, Lei, Liu, Tianming, 2011. Fiber-centered Granger causality analysis. Med. Image Comput. Comput. Assist. Interv. 14 (Pt 2), 251–259.
Logothetis, Nikos K., 2008. What we can do and what we cannot do with fMRI. Nature 453, 869–878.
Lv, Jinglei, Guo, Lei, Hu, Xintao, Li, Kaiming, Liu, Tianming, 2011. Activated fibers: fiber-centered activation detection in task-based fMRI. Information processing in medical imaging (IPMI), volume 6801/2011, pp. 574–587.
Meek, C., 1997. Graphical Models: Selecting Causal and Statistical Models. Carnegie Mellon University, Pittsburgh, PA.
Meier, J.D., et al., 2008. Complex organization of human primary motor cortex: a high-resolution fMRI study. J. Neurophysiol. 100, 1800–1812.
Murre, J., Goebel, R., 1996. Connectionist models. In: Dijkstra, T., De Smedt, K. (Eds.), Computational Psycholinguistics. Taylor & Francis, London, pp. 49–81.
Neapolitan, Richard E., 2004. Learning Bayesian Networks. Prentice Hall.
Nishimoto, Shinji, Vu, An T., Naselaris, Thomas, Benjamini, Yuval, Yu, Bin, Gallant, Jack L., 2011. Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 21 (19), 1641–1646 (22 September).
Passingham, R.E., Stephan, K.E., Kotter, R., 2002. The anatomical basis of functional localization in the cortex. Nat. Rev. Neurosci. 3, 606–616.
Ramsey, J.D., Hanson, S.J., Hanson, C., Halchenko, Y.O., Poldrack, R.A., Glymour, C., 2010. Six problems for causal inference from fMRI. NeuroImage 49, 1545–1558.
Ramsey, J.D., Hanson, S.J., Glymour, C., 2011. Multi-subject search correctly identifies causal connections and most causal directions in the DCM models of the Smith et al. simulation study. NeuroImage 58 (3), 838–848 (1 October).
Richardson, T.S., Spirtes, P., Glymour, C., 1997. A note on cyclic graphs and dynamical feedback systems. Proceedings of Artificial Intelligence and Statistics.
Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25, 230–242.
Said, C.P., Moore, C.D., Engell, A.D., Todorov, A., Haxby, J.V., 2010. Distributed representations of dynamic facial expressions in the superior temporal sulcus. J. Vis. 10, 1–12.
Schneider, W., Eschman, A., Zuccolotto, A., 2002. E-Prime Reference Guide. Psychology Software Tools, Inc., Pittsburgh.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.
Seth, A.K., 2010. A MATLAB toolbox for Granger causal connectivity analysis. J. Neurosci. Methods 186 (2), 262–273.
Shapiro, R.M.H.a.L.G., 1992. Computer and Robot Vision. Addison-Wesley.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y.Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23, S208–S219.
Smith, Stephen M., Miller, Karla L., Salimi-Khorshidi, Gholamrez, Webster, Matthew, Beckmann, Christian F., Nichols, Thomas E., Ramsey, Joseph D., Woolrich, Mark W., 2011. Network modelling methods for FMRI. NeuroImage 54, 875–891.
Spirtes, P., Glymour, C., 1991. An algorithm for fast recovery of sparse causal graphs. Soc. Sci. Comput. Rev. 9, 62–72.
Spirtes, P., Glymour, C., Scheines, R., 2000. Causation, Prediction and Search, 2nd ed. MIT Press, Cambridge, MA.
Storkey, Amos J., Simonotto, Enrico, Whalley, Heather, Lawrie, Stephen, Murray, Lawrence, McGonigle, David, 2007. Learning Structural Equation Models for fMRI. NIPS.
van de Ven, V., Bledowski, C., Prvulovic, D., Goebel, R., Formisano, E., Di Salle, F., Linden, D.E., Esposito, F., 2008. Visual target modulation of functional connectivity networks revealed by self-organizing group ICA. Hum. Brain Mapp. 29, 1450–1461.
van den Heuvel, Martijn, Mandl, Rene, Hulshoff Pol, Hilleke, 2008. Normalized cut group clustering of resting-state fMRI data. PLoS One 3 (4), e2001.

Walther, D., Caddigan, E., Fei-Fei, L., Beck, Diane M., 2009. Natural scene categories revealed in distributed patterns of activity in the human brain. J. Neurosci. 29 (34), 10573–10581.

Wang, Jun, Pohlmeyer, Eric, Hanna, Barbara, Jiang, Yu-Gang, Sajda, Paul, Chang, Shih-Fu, 2009. Brain state decoding for rapid image retrieval. ACM Multimed. 945–954.

Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M., 2009. Bayesian analysis of neuroimaging data in FSL. NeuroImage 45, S173–S186.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant voxels in images of cerebral activation. Hum. Brain Mapp. 4, 58–73.

Xuan, B., Zhang, D., He, S., Chen, X., 2007. Larger stimuli are judged to last longer. J. Vis. 7 (10), 1–5 (2).

Zhu, Dajiang, Li, Kaiming, Faraco, Carlos, Deng, Fan, Zhang, Degang, Jiang, Xi, Chen, Hanbo, Guo, Lei, Miller, Stephen, Liu, Tianming, 2012. Optimization of functional brain ROIs via maximization of consistency of structural connectivity profiles. NeuroImage 59 (2), 1382–1393 (16).