6-1-2015

Fusion analysis of functional MRI data for classification of individuals based on patterns of activation.

Mahdi Ramezani
Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada & Robotics and Control Laboratory, Department of Electrical Engineering, The University of British Columbia, 2332 Main Mall, Vancouver, BC, Canada V6T 1Z4, mahdir@ece.ubc.ca

Purang Abolmaesumi
Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada

Kris Marble
Department of Psychology, Queen's University, Kingston, ON, Canada

Heather Trang
Department of Psychology, Queen's University, Kingston, ON, Canada

Ingrid Johnsrude
Centre for Neuroscience Studies, Queen's University, Kingston, ON, Canada & Department of Psychology, Queen's University, Kingston, ON, Canada & Linnaeus Centre for Hearing and Deafness, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

Fusion analysis of functional MRI data for classification of individuals based on patterns of activation

Mahdi Ramezani · Purang Abolmaesumi · Kris Marble · Heather Trang · Ingrid Johnsrude

Abstract Classification of individuals based on patterns of brain activity observed in functional MRI contrasts may be helpful for diagnosis of neurological disorders. Prior work for classification based on these patterns have primarily focused on using a single contrast, which does not take advantage of complementary information that may be available in multiple contrasts. Where multiple contrasts are used, the objective has been only to identify the joint, distinct brain activity patterns that differ between groups of subjects; not to use the information to classify individuals. Here, we use joint Independent Component Analysis (jICA) within a Support Vector Machine (SVM) classification method, and take advantage of the relative contribution of activation patterns generated from multiple fMRI contrasts to improve classification accuracy. Young (age: 19–26) and older (age: 57–73) adults (16 each) were scanned while listening to noise alone and to speech degraded with noise, half of which contained meaningful context that could be used to enhance intelligibility. Functional contrasts based on these conditions (and a silent baseline condition) were used within jICA to generate spatially independent joint activation sources and their corresponding modulation profiles. Modulation profiles were used within a non-linear SVM framework to classify individuals as young or older. Results demonstrate that a combination of activation maps across the multiple contrasts yielded an area under ROC curve of 0.86, superior to classification resulting from individual contrasts. Moreover, class separability, measured by a divergence criterion, was substantially higher when using the combination of activation maps.

Keywords fMRI · Fusion analysis · Functional image analysis · jICA · Automatic classification

M. Ramezani · P. Abolmaesumi
Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada

I. Johnsrude
Centre for Neuroscience Studies, Queen's University, Kingston, ON, Canada

K. Marble · H. Trang · I. Johnsrude
Department of Psychology, Queen's University, Kingston, ON, Canada

I. Johnsrude
Linnaeus Centre for Hearing and Deafness, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

M. Ramezani (✉)
Robotics and Control Laboratory, Department of Electrical Engineering, The University of British Columbia, 2332 Main Mall, Vancouver, BC, Canada V6T 1Z4
e-mail: mahdir@ece.ubc.ca

Introduction

Functional Magnetic Resonance Imaging (fMRI) studies are typically analyzed to reveal regional specialization for cognitive functions or tasks, or to compare patterns of activity between two groups, such as patients and normal-control participants (Coleman et al. 2007). Usually, simple comparisons of conditions are performed to reveal regions that are reliably active by the task of interest, and/or regions that differ reliably between groups (Friston et al. 1999, 1995a, b). Several studies have taken advantage of these identified regions for group classification based on fMRI data (Arribas et al. 2010; Demirci et al. 2008; Ford et al. 2003).

Although scant previous work deals with our specific problem of classification of individuals based on patterns of brain activity, the closely related goal, of use machine learning to decoding stimuli, mental states, and behaviours from fMRI data is rapidly gaining in popularity; particularly the set of methods called representational similarity analysis, or pattern-

information analysis (see (Pereira et al. 2009), (Mur et al. 2009) for tutorial reviews). In this context, Haxby and colleagues showed that fMRI activation patterns are different when viewing a photograph of a face from viewing a house, a shoe, or a chair (Haxby et al. 2001). Using a similar dataset, Cox and Savoy successfully classified patterns of fMRI activation evoked by the presentation of photographs of various categories of objects, by applying Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) (Cox and Savoy 2003). Mitchell et al. successfully trained classifiers to automatically decode the subject's cognitive state at a single time instant or interval (Mitchell et al. 2004). De Martino et al. combined multivariate voxel selection and SVM for classification of fMRI spatial patterns (De Martino et al. 2008). Pereira combined dimensionality reduction and classification into a single learning objective to achieve better learning performance (Pereira 2007). Unlike these studies that used relatively simple stimuli or images drawn from fixed categories, Kay et al. used natural receptive-field models to identify a specific image, viewed by an observer, from a large set of natural images (Kay et al. 2008). Their group further combined structural and semantic encoding models, and prior information about the structure and semantic content of natural images, to produce accurate reconstructions of observed natural images from brain activities (Naselaris et al. 2009). Schrouff et al. used feature extraction methods with different classifiers to decode semi-constrained brain activity patterns, where number and duration of mental events were not externally imposed (Schrouff et al. 2012).

In our study, the goal is to characterize activity averaged over many time intervals and to use it for classification of individuals, rather than to detect transient cognitive states. This procedure typically consists of three steps. The first step is to determine the activation maps using a data-driven method such as group-Independent Component Analysis (group-ICA) (Calhoun et al. 2008), or a model-based approach such as general linear model (Fan et al. 2011). The next step is to reduce the dimensionality of the data and compute representative features using Principal Component Analysis (PCA) (Demirci et al. 2008; Ford et al. 2003), Singular Value Decomposition (SVD) (Arribas et al. 2010), ICA (Fan et al. 2011), GLM or Recursive Feature Addition (RFA) (Schrouff et al. 2012). Finally, a classification is performed on the obtained features. These approaches have only focused on the classification based on single comparison of conditions, and are not sensitive to shared information among different contrasts generated from those comparisons.

Recently, considerable attention has been focused on combining data across multiple modalities including fMRI, or, within fMRI, across different contrasts, to assess the joint information that may exist among those sources (Sui et al. 2010b). These analyses may enhance sensitivity to group differences, and permit better understanding of complex disorders that affect many aspects of the brain (such as its structure, function, and organization (Calhoun et al. 2006a)). The premise of multi-modality approaches is that each imaging modality provides complementary information about different tissue characteristics, and at different spatial and temporal resolutions. The premise of multi-task approaches is that each fMRI contrast may provide complementary information about how the system responds to sensory and cognitive challenges, and this information may be usefully combined.

Many techniques have been proposed to combine multi-modal or multi-task fMRI information. These techniques can be categorized into two main types: data-integration and data-fusion methods (Calhoun and Adali 2009; Daunizeau et al. 2010; Savopol and Armenakis 2002). Data-integration techniques use one imaging modality to improve the results of another modality (for example, registration of EEG or MEG to MRI (Halchenko et al. 2005); and using fMRI to estimate the location of dipoles or the distribution of neural sources prior to EEG (Liu et al. 2006)). Data fusion techniques utilize multiple modalities (Calhoun et al. 2006a) or tasks (Calhoun et al. 2006b) to take advantage of combined information. Generally, due to weak cross-modality relationships and inter-subject variability, finding one-to-one correspondence in multi-modal images is difficult; however, performing fusion analysis across multiple subjects makes this an easier problem to solve. In this type of analysis, each modality is usually reduced to a feature that is a lower-dimensional representation of a selected brain structure or task-related activation pattern. Using variations across individual subjects, associations across the features can be explored (Calhoun and Adali 2009).

Joint Independent Component Analysis (jICA) (Calhoun et al. 2006a) is one multivariate technique for fusion analysis. It is an extension of ICA that combines information from multiple modalities or functional contrasts. The simplified noise-free ICA model seems to be sufficient for many applications (Hyvärinen and Oja 2000), and has been successfully applied to fMRI data (Calhoun et al. 2001; McKeown et al. 1998; Sui et al. 2009a). In this simplified model, the ICA components that contribute the least, and which may have a "speckled" spatial distribution in contrast images, are noise of unknown origin (McKeown et al. 1998). ICA is typically used as a first-level data-driven approach to find spatially or temporally independent brain sources of activity from fMRI scans of a person's brain, while that person is performing a desired task (McKeown et al. 1998). Spatial ICA results in a set of independent components (spatial brain activation patterns), and a set of "mixing coefficients".

ICA has also been used for making inferences at a group level (Calhoun et al. 2009, 2001). Joint ICA is a group-level analysis technique that uses extracted features from individual subjects' data and tries to maximize the independence among joint components. Assuming that the features, obtained from

multiple modalities or multiple contrasts, share the same modulation profile (i.e. mixing-coefficient matrix), jICA uses more information to estimate the same number of mixing coefficients; and may therefore yield improved results compared to ICA. Joint ICA and its extensions have been successfully applied to study major depression (Choi et al. 2008), aphasia (Specht et al. 2009) and schizophrenia (Calhoun et al. 2006a, b; Liu et al. 2009; Sui et al. 2009a, b, 2010a, b). These studies demonstrate that jICA accurately identifies sources of common variance among features. An additional advantage of jICA is the computation of modulation profiles along with the identified sources. Such profiles substantially reduce the dimensionality of the data, and can be used for group classification. The ability of jICA to reduce the number of dimensions is particularly important in the context of fMRI data analysis, where the dimensionality of input feature space is high and the number of available subjects is usually low. Although the discrimination ability of joint components has been investigated (Sui et al. 2009b), to the best of our knowledge, the application of modulation profiles for classification of individuals has not, to our knowledge, been tried. Recently, Fan et al. used the modulation profile, which resulted from applying ICA to resting-state fMRI data, to classify individuals with schizophrenia and healthy controls (Fan et al. 2011). Their work did not investigate the possibility of improved classification that could be obtained by combining multiple fMRI contrasts.

In the present study, we use jICA for group classification. We first identify the modulation profile that reflects group differences by fusion analysis of multiple contrasts, and then use the resulting profile for group classification. We test this classification approach using fMRI data collected from 16 young and 16 older neurologically normal individuals who were scanned in multiple stimulus conditions in a speech perception experiment (Davis et al. 2011; MacDonald et al. 2008). The jICA modulation profile, which reflects group differences in activation patterns observed in three different functional contrasts, is used to automatically classify individuals as young or older. One contrast compares responses to unintelligible noise bursts, amplitude modulated with the temporal envelope of spoken sentences, with silence. A second contrast compares responses to sentences without coherent meaning ("anomalous" sentences, e.g., "Her good slope was done in carrot") with the unintelligible noise bursts. A third contrast compares responses to sentences with coherent meaning ("coherent" sentences, e.g., "Her new skirt was made of denim") to anomalous sentences. The intelligibility of anomalous sentences is determined by the quality of the signal, whereas the intelligibility of coherent sentences is determined both by the quality of the signal, and by semantic knowledge. At any given signal quality, comprehension is greater for the latter than for the former (referred to hereafter as "context benefit") (MacDonald et al. 2008). Although older and

younger adults do not differ in context benefit measured behaviorally (MacDonald et al. 2008), we examine whether patterns of activity in functional contrasts can be used to classify young and older people, without considering any information reflecting the distinct brain structural differences of the two groups (Chee et al. 2011; Good et al. 2001).

Joint ICA is used with data from the three contrasts to probe the unique and joint information among different contrasts and groups. First, joint independent components based on different combinations of these features, along with the mixing coefficients, are determined and then statistical differences among mixing coefficients (reflecting the network strengths) are examined using t-tests. Third, separability of the joint-source distributions is measured in order to assess the difference between distributions from different young and older participants (Hero et al. 2001). Finally, the modulation profile extracted from the three functional contrasts is used to classify individuals as young or older, and the accuracy of this classification is assessed. We demonstrate that by fusing the three contrasts with jICA, the discrimination of subjects as young or older is substantially improved compared to using each individual contrast alone.

Here we show the feasibility of this method by examining age-related differences in healthy subjects, where an independent measure (namely, age in years) can be used to differentiate between groups with 100 % certainty. This 'gold standard' allows us to validate the approach, which we expect to be applicable to real-world diagnostic problems without such a clear standard to differentiate groups.

Materials

Listening study

Subjects were asked to listen to sentences in the scanner and try to understand them as well as they could. Sentences with and without coherent sentence-level meaning ("coherent" and "anomalous" sentences, respectively) were taken from those used by Davis et al. (Davis et al. 2011) and were mixed with noise that had the same long term spectrum of the speech and the amplitude envelope of the signal to be masked (Signal-Correlated Noise: SCN; (Schroeder 1968) at six different signal to noise ratios (SNRs): −5 dB, −3.5 dB, −2.5 dB, −1 dB, 0 dB, 2.5 dB. Clear speech was also tested, making seven sentence conditions. Coherent and anomalous sentences were divided into seven sets, which were pseudorandomly assigned to conditions such that each sentence set was tested in each of the seven SNR conditions (including clear speech) an equal number of times, across participants. Over the scanning session, each participant heard 14 trials of each sentence type at each SNR, half of which were followed by 'repeat' trials requiring the participant to repeat as much of the

sentence as possible. Fourteen trials of SCN on its own and 16 trials of silence were also scanned, making 324 trials in all, distributed across four blocks of trials, each 81 trials long. The repeat trials were randomly ordered and counterbalanced across participants. Intelligibility, defined here as the proportion of words correctly reported, was obtained for each signal quality level and for each sentence type, for each participant.

Participants

Sixteen young (mean age: 21.1, range: 19–26, 11 female) and 16 older (mean age: 64.2, range: 57–73, 11 female) adults were scanned. All subjects were native speakers of English, without any history of neurological illness, head injury, or hearing impairment. This study was cleared by the Queen's University Health Sciences Research Ethics Board, and written informed consent was obtained from all participants.

Data acquisition

The fMRI data were acquired using a 3.0 Tesla Siemens Trio MRI system with a 12-channel head coil in the MRI facility at Queen's University, Kingston, Canada. Each acquisition consisted of 32 contiguous slices with 4 mm thickness, field of view 211×211 mm, in plane resolution of 3.3×3.3 mm, resulting in a grid of 64×64×32 voxels, each 3.3×3.3×4 mm in volume. The repetition time (TR) was 9 s and the acquisition time was 2 s. This sparse GE-EPI imaging technique allowed for stimuli to be presented in the silent gaps between scans. Total functional imaging time was 48 min. Auditory stimuli and the visual 'repeat' instructions were presented to the participants using E-Prime v.1.2 and a NEC LT265 DLP projector. Participants viewed the screen via a mirror system mounted on the head coil (MacDonald 2008).

Data preprocessing

Before preprocessing, the Siemens motion correction algorithm[1] was applied to the DICOM MR images, and then the DICOM images were converted to NIFTI format. The fMRI data were preprocessed using Statistical Parametric Mapping software (SPM8, Wellcome Department of Cognitive Neurology, London, UK). Preprocessing steps included realignment, coregistration and the segmentation-based spatial normalization of SPM8. The data were spatially smoothed using an 8-mm Gaussian kernel. The first scan of each session was discarded, and the rest were coded according to the auditory condition of the preceding stimulus and entered into a single-subject general linear model. The hemodynamic response function was selected as the basis function. Three

---

[1] Available online at http://imaging.mrc-cbu.ac.uk/imaging/DataDiagnostics

functional contrasts were calculated: SCN versus silence, identifying brain regions that process the acoustic properties of sound; anomalous sentences versus SCN to highlight speech-responsive areas; coherent versus anomalous sentences to reveal regions sensitive to coherent sentence-level semantic content.

Methods

As mentioned in Section Listening study, the behavioral performance for each subject was measured as the words correctly reported at different SNRs. Pilot work revealed that in general at a given SNR older people reported fewer words than younger people. Accordingly we altered SNRs for the two groups to match behavioral performance. The average report score for anomalous sentences, which do not provide a contextual benefit, gives a good indication of low-level speech processing. A range of SNRs for each group was chosen in order to equate the behavioral performance while hearing the anomalous sentences. The ranges were −5 dB, −3.5 dB, −2.5 dB, −1 dB, and 0 dB for younger people and −3.5 dB, −2.5 dB, −1 dB, 0 dB, and +2.5 dB for older adults.

Fusion analysis

Joint ICA assumes a noise-free generative model $X=AS$ where a source matrix $S=[s_1, s_2, \ldots, s_M]^T$ combines with the mixing coefficients matrix A (also called the ICA loading parameters matrix) to generate the observations $X=[x_1, x_2, \ldots, x_N]^T$. The jth row, s_j, of S is the jth joint independent component, and M is the number of independent components. N is the number of participants and x_i is a vector formed by concatenating the three contrasts described in Section Data preprocessing.

The jICA method, shown schematically in Fig. 1, involves finding $U=WX$, where $W=A^{-1}$ is called the unmixing matrix and U is the estimate of the joint source matrix S. In this figure A_y, A_o indicate the submatrices associated with the young and old subjects. A MATLAB implementation of jICA is provided by the FIT 2.0b software (Calhoun et al. 2006a), available online at http://mialab.mrn.org/software/.

Joint independent components were found using the Infomax algorithm (Bell and Sejnowski 1995), which is based on minimization of mutual information of components. In this algorithm, the output entropy of a neural network is adaptively maximized with as many outputs as the number of Independent Components (ICs) to be estimated.

The best way to estimate the most appropriate number of independent components is not clear. This number can affect the results of ICA, particularly if it is too small (Ma et al. 2007). To estimate the number of ICs, we first attempted to use the Minimum Description Length (MDL) criterion (Li et al. 2007), which is an information theoretic technique for
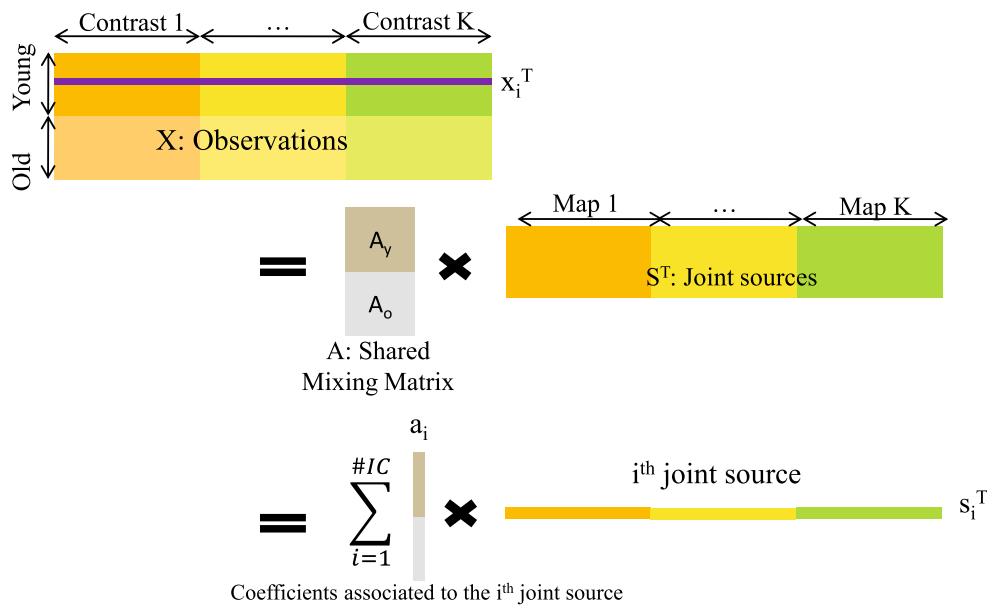
Fig. 1 Schematic of the jICA method. The observation matrix is made by stacking the SPM t-contrast maps from the experimental conditions side by side. Assuming that maps of joint sources share the same mixing-coefficient matrix, jICA tries to maximize the independence among them

model order selection. However, because of the heterogeneity of activation localization and size in both groups, this criterion did not converge. So the precedent set by (Specht et al. 2009) was followed and eight was chosen as the number of ICs. To justify this choice, the analysis was also repeated for 12 and 16 components, and the differences were negligible, as shown in the results section.

The jICA procedure generated a set of joint independent components and associated mixing coefficients. These low-dimensional coefficients model the modulation of each subject's functional contrast by a joint source, and thus can be used as a criterion for capturing group differences. To investigate whether the groups were separable by different weightings of the joint sources, unpaired two-sample t-tests with unequal variance (heteroscedastic) on the mixing coefficients were performed. The z-scaled results indicated the joint components of interest.

Selection of optimal joint sources

If two groups differ, than the distributions of their joint components should be separable (Calhoun and Adali 2009; Sui et al. 2009a, b). Separability can be quantified by computing a divergence between joint histograms. Group joint-sources are defined by $U_y = A_y^{-1} X_y = W_y X_y$, where A_y, W_y and X_y indicate the submatrices associated with the young subjects (similar for the older group). For each subject, the appropriate group joint source (e.g. U_o for an older subject) was divided into three maps, which correspond to the three contrasts used in the jICA analysis. The map elements (each one representing a specific voxel) were sorted and thresholded, leaving a set of voxels statistically relevant to the joint source. Each voxel that survived thresholding in all three maps was included in a three-

dimensional joint histogram in a bin defined by the three contrast values (from the input observation matrix X) at that voxel's location.

The group-averaged joint histograms were then calculated by taking the mean of the joint histograms of each subject in the group. The difference between the two groups was then assessed using the Renyi divergence formula (Hero et al. 2001) with $\alpha=0.5$:

$$D_\alpha(P\|Q) = \frac{1}{\alpha-1} log\left(\sum_{i=1}^{n} \frac{p_i^\alpha}{q_i^{\alpha-1}}\right) = \frac{1}{\alpha-1}\sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}$$

$$(1)$$

where P and Q are probability distributions, reflected in the group-averaged joint histograms.

The divergence is also computed for different combinations of contrasts. The higher the values of the Renyi divergence criterion, the better the discrimination between groups. Therefore, best combination of contrasts is that which yields the highest divergence value.

Automatic classification of young and older subjects

In order to overcome the classification problems caused by high dimensionality of fMRI data and the small set of available subjects (16 in each group), columns of the mixing coefficients matrix, which reflect the weighting of each joint source in a subject's contrast, were used as input features to a classification algorithm. A non-linear Support Vector Machine (SVM) was used to classify the subjects. SVM does not assume that data points conform to a specific model, but rather seeks to find the

hyperplane that separates the two classes with maximum margin (Theodoridis and Koutroumbas 2003). The hyperplane is defined by $f(x) = \omega.K_s(x) + b$, where $K_s(x) = [k(x,s_1), .., k(x,s_d)]$ is the vector of kernel functions centered at the support vectors, ω is the parameter vector and b is a scalar. Radial Basis Functions (RBF) were used as the kernels: $k(x,s) = exp(-|x-z|^2/\sigma^2)$. The data were split into a training set and a test set. In the training phase, fusion analysis was repeated on only those subjects in the training set. This produced a new mixing coefficient matrix $A_{(train)}$ and joint source matrix $S_{(train)}$ that modeled the generation of the training observations $X_{(train)}$. The columns of the mixing matrix $A_{(train)}$ were used as input features to train the classifier, which

divided the training subjects into two classes of young and older adults.

The input features for the test set, columns of $A_{(test)}$, were then found by least-squares solution of $X_{(test)} = A_{(test)}S_{(train)}$. The positions of these vectors in k-dimensional feature space, relative to the hyperplane found in running the classifier on the training set, determined the classification of each test subject. The number of columns of A, or mixing coefficients, used in the classification was k. A MATLAB implementation of the classifier provided by the Statistical Pattern Recognition Toolbox (STPRtool), available at http://cmp.felk.cvut.cz/cmp/cmp_software.html was used for this step.

Performance of the classification procedure was measured by repeatedly splitting the data into training and test sets and
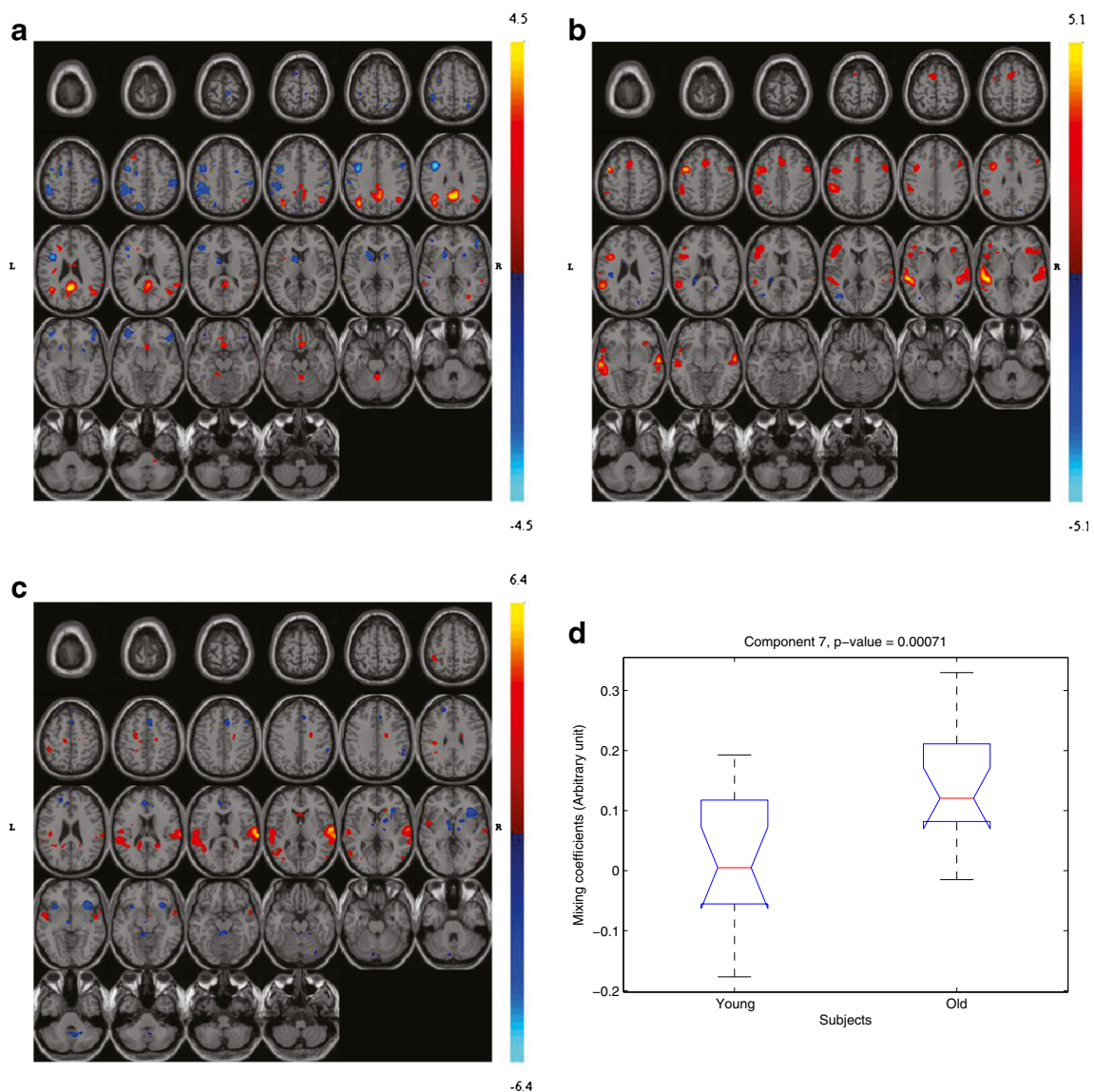

Fig. 2 Joint Independent Component Analysis (jICA) of brain patterns. The joint source map of the most significant component for the contrast of high vs anomalous sentences (a), for the contrast of anomalous sentences vs SCN (b), and for the contrast of SCN vs Silence (c), along with the mixing coefficients for the young and older subjects (d) is presented

averaging classification performance across iterations. The splitting was done 200 times (each time selecting different 11 young and 11 older subjects as the training set and the remaining five subjects in each group as the test set). The False Positive (FP), False Negative (FN), True Positive (TP) and True Negative (TN) values were calculated and the ratio between TP and TN values to the total number of outcomes was taken as the performance metric. Selecting the significant features (mixing coefficients) is an important factor in classification results. The mixing coefficients were sorted by the p-values resulting from a two-sample t-test checking for differences between groups, and the k coefficients with the most

significant difference were chosen for use in classification. Unlike some statistical classification methods, SVM does not provide posterior class probabilities (P_i). Without posterior probabilities, it is not possible to assess the performance of the classifier at other threshold values, and to measure the sensitivity and specificity of the classifier. Following the Platt (Platt 1999) approach, we trained an SVM and later the parameters of an additional sigmoid function to map the values of SVM outputs to posterior probabilities. Using the posterior probabilities, the Receiver Operating Characteristic (ROC) curves were plotted and their Area Under the Curves (AUC) calculated. The AUC metric is

Table 1 Stereotaxic coordinates for the most discriminative source map in three contrasts. MNI coordinates of voxels, which are above a threshold of $|Z|>3.5$, are shown in the table

Area	Broadman area	R/L volume (cc)	R/L random effects: max value (x, y, z)
Higher-level cognitive patterns			
Positive			
Cingulate gyrus	31	0.5/0.5	4.2 (−4, −48, 35)/4.3 (7, −48, 35)
Precuneus	7,31,39	0.4/0.5	4.2 (−4, −48, 38)/4.1 (7, −48, 38)
Posterior cingulate	23,31	0.2/0.2	3.9 (0, −45, 31)/3.8 (3, −45, 31)
Angular gyrus	*	0.0/0.1	ns/3.6 (44, −63, 40)
Negative			
Middle frontal gyrus	*	0.0/0.2	ns/4.3 (44, 15, 32)
–	*	0.0/0.2	ns/4.0 (48, 15, 32)
Inferior frontal gyrus	9	0.0/0.3	ns/4.0 (44, 12, 32)
Precentral gyrus	6	0.0/0.1	ns/3.7 (48, 7, 39)
Mid-level cognitive patterns			
Positive			
Middle temporal gyrus	21, 22	0.7/1.6	4.8 (−64, −5, −6)/5.1 (67, −29, 8)
Superior temporal gyrus	21, 22, 42	0.3/1.7	3.8 (−55, −32, 8)/4.9 (57, −32, 6)
Middle frontal gyrus	6	0.1/1.0	3.6 (−57, 10, 55)/4.7 (51, 10, 55)
–	*	0.1/0.5	3.9 (−51, −32, 8)/4.6 (57, −35, 6)
Inferior parietal lobule	40	0.0/0.4	ns/4.2 (63, −42, 31)
Supramarginal gyrus	40	0.0/0.1	ns /4.0 (63, −43, 28)
Precentral gyrus	6	0.0/0.1	ns/3.9 (57, 7, 51)
Postcentral gyrus	*	0.0/0.1	ns /3.6 (63, −26, 45)
Inferior frontal gyrus	*	0.0/0.1	ns/3.6 (48, 12, 32)
Low-level cognitive patterns			
Positive			
Postcentral gyrus	40, 43	1.1/0.1	6.0 (−74, −16, 20)/3.6 (67, −24, 23)
Superior temporal gyrus	22, 41, 42	1.9/0.7	5.6 (−74, −22, 17)/4.4 (67, −28, 24)
Transverse temporal gyrus	41, 42	1.2/0.2	5.5 (−67, −10, 12)/4.0 (67, −13, 13)
Precentral gyrus	43	0.1/0.0	4.7 (−64, −7, 12)/ns
Insula	13	0.0/0.2	ns/4.2 (63, −31, 24)
–	*	0.1/0.0	5.9 (−78, −22, 17)/ns
Negative			
Inferior frontal gyrus	47	0.4/0.0	3.9 (−34, 28, −4)/ns
Insula	13	0.1/0.0	3.6 (−31, 22, −6)/ns

L and R show the assigned anatomical left and right hemispheres, cc stands for cubic centimeters showing the volume concentration of voxels, the coordinates and value of the maximum Z are also provided in the table. Not-significant regions are shown by ns
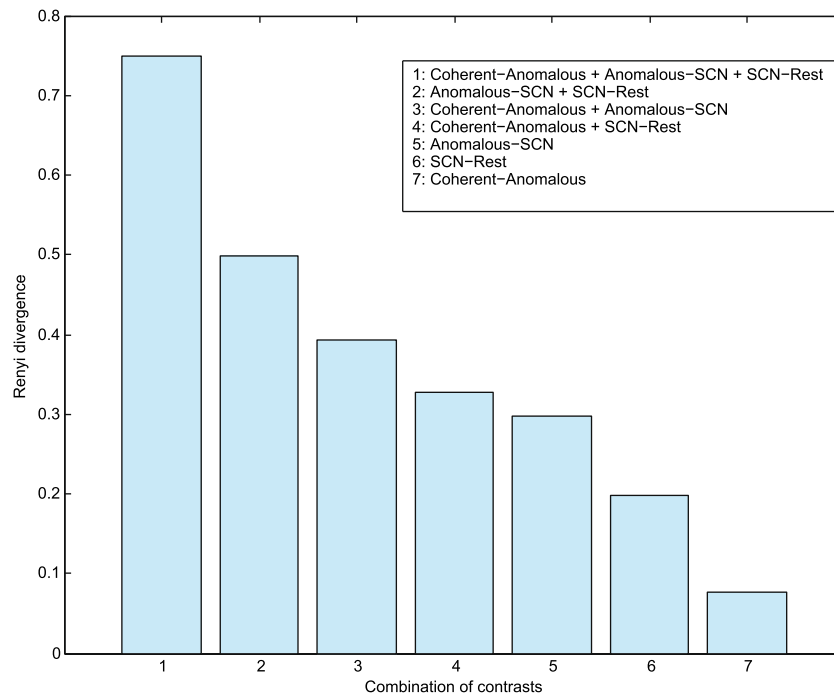
Fig. 3 The effect of combining different contrasts on differentiation between histograms. The higher the values of the Renyi divergence, the better the discrimination between groups

the most common way to compare the accuracy of classification methods in the machine learning community. Detection reliability ρ, was defined based on AUC as $\rho = 2 \times AUC-1$ (Ramezani and Ghaemmaghami 2010). Under this definition, $\rho = 1$ for perfect detection and $\rho = 0$ for failure in detection.

The joint ICA classification result was compared to ICA for each of the three contrasts separately, to examine whether the fusion analysis has advantages over analysis of the results of each contrast separately. As in the fusion analysis, the mixing coefficients were employed as input features for classification of young and older adults.
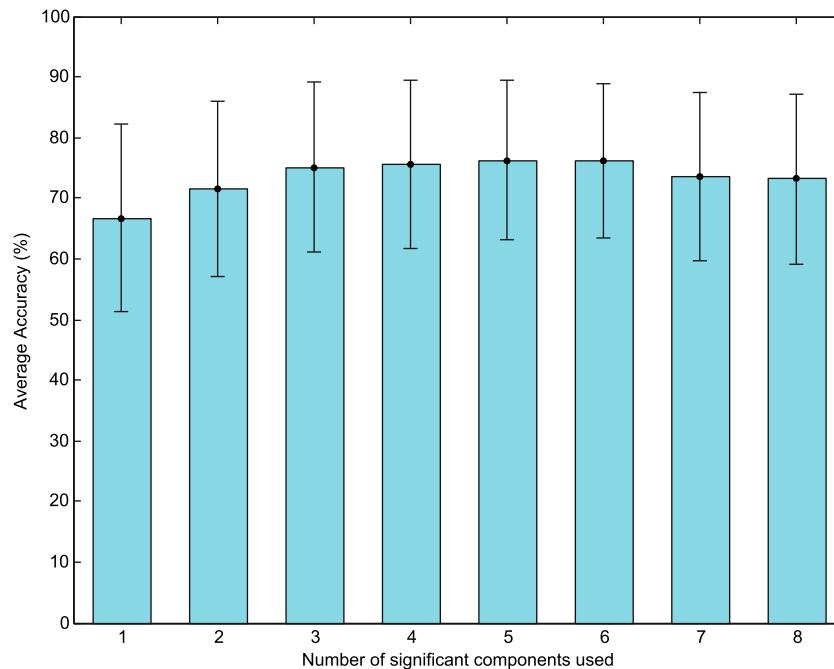

Fig. 4 Classification accuracy for different number of features used. Selecting only three significant features (*p*-value <0.05), produced an average classification accuracy of around 75 %
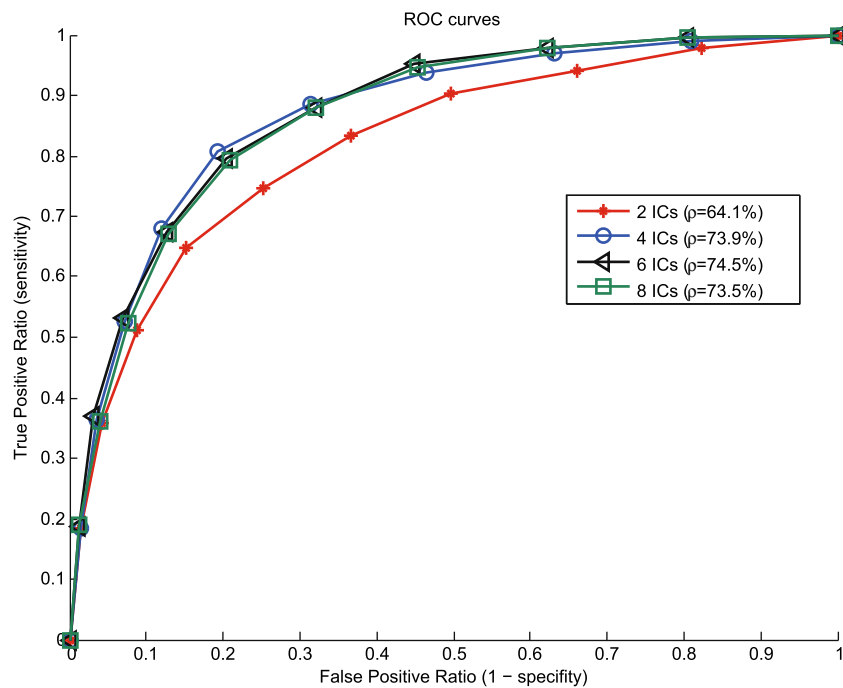
Fig. 5 ROC curves and detection reliability resulting from four different numbers of features used. Perfect detection: $\rho=1$; detection failure: $\rho=0$. Performance of the classification was measured by repeatedly splitting the data into training and test sets and averaging classification performance on a test data set. Splitting was done 200 times and the classification was repeated five times with randomized order of the subjects in the training dataset. ROC curves are computed using posterior probabilities of the SVM output

Results

Although older adults need higher SNRs to achieve the same performance scores as young adults (i.e. they do not perform as well in noise), behaviorally there is no difference in the amount of benefit they get from contextual information compared to anomalous information.

The goal of our fusion analysis is to examine whether jICA components can be used to accurately distinguish young and older adults on the basis of fMRI data from a speech perception
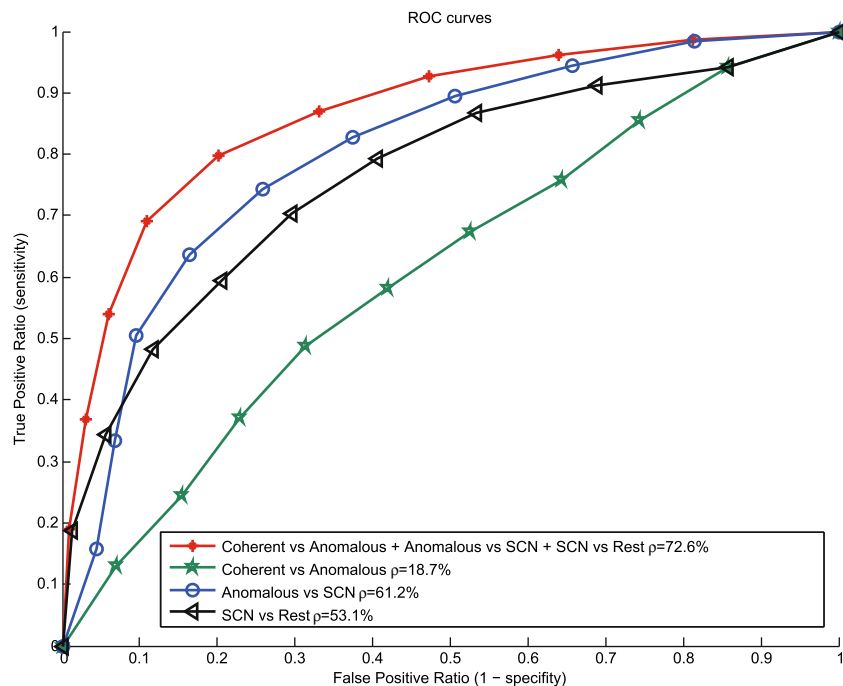

Fig. 6 ROC curves and detection reliability for different contrasts. Perfect detection: $\rho=1$; detection failure: $\rho=0$. All of the contrasts (except that of coherent vs anomalous sentences) show high detection accuracy

experiment, despite the similarity in contextual benefit between the groups. The success of this analysis is evaluated by examining the statistical difference among the mixing coefficients of joint sources, by applying the Renyi divergence criterion, and by an automatic classification method. These tests are described in the following three subsections.

Statistical difference among joint sources

Unpaired two-sample t-tests (assuming unequal variance) were performed on the mixing coefficients, and two components were found to differ significantly (p-value=0.00071 and p-value=0.0301, number of subjects=16 in each group) between the two groups. Figure 2a, b and c show the statistical Z-maps generated for the joint source (shown as rows of Map 1, Map 2, and Map 3 in Fig. 2) with the largest group difference. Figure 2d shows that the mixing coefficients for this joint source have higher values in older subjects compared to younger subjects. Table 1 shows the corresponding stereotaxic coordinates in MNI space for this source.

Selection of optimal joint sources

The sorted maximum Renyi divergence values for different combinations of contrasts are shown in Fig. 3. Higher values indicate better discrimination between the groups. It is clear that combining all three contrast images yields the best results. It can also be seen that the contrast comparing responses to anomalous sentences and to unintelligible noise is the most effective single contrast in separating the groups.

Classification

Figure 4 shows the average classification accuracy for different numbers of features. Results show that the young and older subjects can be classified based on their patterns of activity across the three contrasts of interest. Considering the fact that the number of subjects is low and the dimensionality of the input fMRI dataset to the fusion framework is quite high, the results are very promising. By selecting only the first three features, an average classification accuracy of around 75 % is obtained. Figure 5 shows the ROC curves and detection reliability for four different numbers of features. Since each additional feature is (by definition) less important than the previous one, the addition of features beyond three leads to asymptotically improving discriminability.

Using the first three features, the joint ICA classification result is compared to ICA for each of the three contrasts separately in Fig. 6, which shows the ROC curves, the detection reliability and the area under ROC curve obtained. Classification of individuals based on ICAs of the contrast comparing anomalous sentences and noise, and the contrast comparing SCN and silence, yield significantly (p-value< 0.05) better-than-chance classifications, but substantially lower than the result obtained by the jICA classification.

Figure 7 shows the ROC curves and the corresponding detection reliability when the analysis is run with different numbers of ICs. It is easily seen that adding more ICs has little effect on the performance. Comparing the classification errors, there is really no major difference between different numbers of ICs.
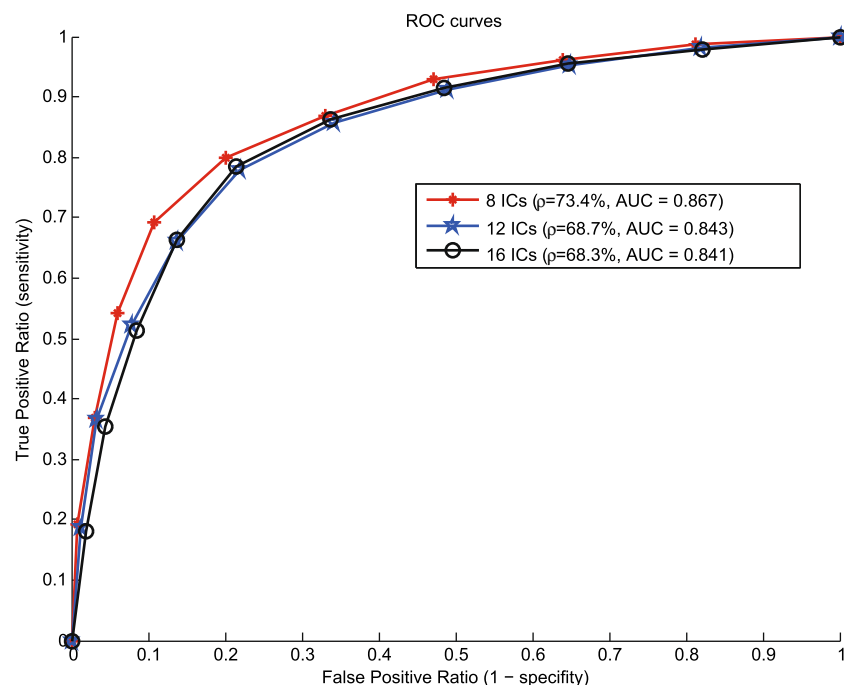

Fig. 7 ROC curves and detection reliability for different numbers of ICs. Perfect detection: $\rho=1$; detection failure: $\rho=0$

Discussion

In this paper we used GLM to generate activation maps for multiple fMRI conditions related to speech perception and comprehension, and jICA to decompose the activation maps into independent maps that share modulation profiles. This is similar to assuming a fixed hemodynamic response for each subject and modeling the amplitude differences in modulation profiles (Calhoun et al. 2006b). The modulation profiles were used within a non-linear SVM framework to classify individuals. Our major findings are: (1) brain functional patterns of activation permit classification of individuals as younger or older; (2) combining these patterns improves the separability of joint sources and the accuracy of classification of individuals.

The joint source that differed the most between the two groups appeared to reflect activation differences at multiple levels of processing, including in left inferior frontal cortex for the contrast between coherent and anomalous sentences, in both temporal lobes for the contrast between anomalous sentences and SCN, and in primary auditory cortex bilaterally for the contrast of SCN vs. silence. This reflects the well-known hierarchy of speech processing in which low-level acoustic features are analyzed in auditory cortex, superior and middle temporal gyri are sensitive to processing of auditorily presented sentences, and left inferior frontal gyrus activity reflects higher-level linguistic (possibly semantic) processing (Davis and Johnsrude 2003; Davis et al. 2011; Okada et al. 2010; Peelle et al. 2010; Rodd et al. 2005).

The relation between the three maps of the joint sources (see Fig. 1) were investigated by looking back to the SPM contrast images and examining regions that contributed significantly in the joint source, i.e. computing the joint histogram. The divergence criterion derived from the joint histograms was used to measure the separability of the two groups based on the joint sources. This criterion confirmed that the fusion of contrasts improved separability, compared to the consideration of each contrast separately.

Results demonstrate that individuals can be classified relatively accurately into young and older age groups by combining functional contrasts sensitive to the processing of noise vs. silence, anomalous sentences vs. noise, and coherent vs. anomalous sentences. Note that although the brain imaging data permit this classification, behavioral data did not: the ability to report words from the anomalous and coherent sentences was matched between young and older listeners for the contrasts examined. Figure 4 shows that using only three coefficients of the mixing matrix, a classification accuracy of around 75 % can be obtained, albeit with a high standard deviation. The high standard deviation on the classification performance might be due to the small number of datasets, or because our method of controlling for hearing ability based on the behavioural performance did not work as well as we had hoped.

Although using all three contrasts resulted in the best detection reliability, i.e. highest area under the ROC curve, the contrast of anomalous sentences vs. unintelligible noise had the most impact on separability of the groups, with a detection reliability of around 60 % (AUC of 80 %) by itself. This may be because, in order to match intelligibility, older adults heard sentence materials at more advantageous SNRs, and the acoustic differences between these more positive SNRs and those experienced by the younger listeners may be reflected in different patterns of activity in auditory regions in the two groups. The analysis did not appear sensitive to the number of independent components (8, 12, or 16) included. Also, the number of features, as long as it was three or more, had relatively little impact on classification accuracy.

This was simply a validation study to demonstrate that information across multiple functional contrasts can be usefully combined for classification. Although here we differentiate young and older people, using age as an observable 'gold standard' way to discriminate groups, we anticipate that this method will be useful to aid in classification of individuals to clinical groups using objective, quantitative, criteria. Future work will apply this method in order to classify individuals with mood or other mental disorders.

In summary, using the joint ICA method together with an SVM classification algorithm, we have demonstrated that cognitive patterns can be used to classify individuals in the absence of behavioral differences. Feasibility of the proposed framework is shown by demonstrating that functional activity maps can be used to classify subjects accurately. The best combination of contrasts and optimal components are identified. We showed that by combining three different functional contrasts, revealing three different patterns of brain activity, the overall performance of the classification improves. This paper is a proof of concept that we hope to extend to diagnosis of brain disorders in future studies.

References

Arribas, J. I., Calhoun, V. D., & Adali, T. (2010). Automatic bayesian classification of healthy controls, bipolar disorder, and schizophrenia using intrinsic connectivity maps from fMRI data. *Biomedical Engineering, IEEE Transactions On, 57*(12), 2850–2860.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*(6), 1129–1159.

Calhoun, V. D., & Adali, T. (2009). Feature-based fusion of medical imaging data. *Information Technology in Biomedicine, IEEE Transactions On, 13*(5), 711–720.

Calhoun, V., Adali, T., Pearlson, G., & Pekar, J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping, 14*(3), 140–151.

Calhoun, V. D., Adali, T., Giuliani, N., Pekar, J., Kiehl, K., & Pearlson, G. (2006a). Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data. *Human Brain Mapping, 27*(1), 47–62.

Calhoun, V. D., Adali, T., Kiehl, K. A., Astur, R., Pekar, J. J., & Pearlson, G. D. (2006b). A method for multitask fMRI data fusion applied to schizophrenia. *Human Brain Mapping, 27*(7), 598–610. doi:10.1002/hbm.20204.

Calhoun, V. D., Maciejewski, P. K., Pearlson, G. D., & Kiehl, K. A. (2008). Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. *Human Brain Mapping, 29*(11), 1265–1275.

Calhoun, V. D., Liu, J., & AdalI, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage, 45*(1), S163–S172.

Chee, M. W. L., Zheng, H., Goh, J. O. S., Park, D., & Sutton, B. P. (2011). Brain structure in young and old East asians and westerners: comparisons of structural volume and cortical thickness. *Journal of Cognitive Neuroscience, 23*(5), 1065–1079.

Choi, K., Yang, Z., Hu, X., & Mayberg, H. (2008). *A combined functional-structural connectivity analysis of major depression using joint independent components analysis.* Toronto: Psychiatric MRI/MRS.

Coleman, M. R., Rodd, J. M., Davis, M. H., Johnsrude, I. S., Menon, D. K., Pickard, J. D., et al. (2007). Do vegetative patients retain aspects of language comprehension? Evidence from fMRI. *Brain, 130*(10), 2494–2507.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage, 19*(2), 261–270.

Daunizeau, J., Laufs, H., & Friston, K. J. (2010). EEG + fMRI information fusion: Biophysics and data analysis. In C Mulert & L Lemieux (Eds.), *EEG - fMRI* (pp. 511–526). Springer Berlin Heidelberg. doi:10.1007/978-3-540-87919-0_25

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *The Journal of Neuroscience, 23*(8), 3423.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "Top–Down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience, 23*(12), 1–3932 (Early Access).

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage, 43*(1), 44–58.

Demirci, O., Clark, V. P., & Calhoun, V. D. (2008). A projection pursuit algorithm to classify individuals using fMRI data: application to schizophrenia. *NeuroImage, 39*(4), 1774–1782.

Fan, Y., Liu, Y., Wu, H., Hao, Y., Liu, H., Liu, Z., et al. (2011). Discriminant analysis of functional connectivity patterns on grassmann manifold. *NeuroImage, 56*(4), 2058–2067.

Ford, J., Farid, H., Makedon, F., Flashman, L. A., McAllister, T. W., Megalooikonomou, V., et al. (2003). Patient classification of fMRI activation maps. *Medical image computing and computer-assisted intervention-MICCAI* 2003 (pp. 58–65) Springer. doi:10.1007/978-3-540-39903-2_8

Friston, K., Frith, C., Turner, R., & Frackowiak, R. (1995a). Characterizing evoked hemodynamics with fMRI. *NeuroImage, 2*(2PA), 157–165.

Friston, K., Holmes, A., Poline, J., Grasby, P., Williams, S., Frackowiak, R., et al. (1995b). Analysis of fMRI time-series revisited. *NeuroImage, 2*(1), 45–53.

Friston, K. J., Holmes, A. P., Price, C., Büchel, C., & Worsley, K. (1999). Multisubject fMRI studies and conjunction analyses. *NeuroImage, 10*(4), 385–396.

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N. A., Fristen, K., & Frackowiak, R. S. J. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage, 14*(1), 21–36.

Halchenko, Y. O., Hanson, S. J., & Pearlmutter, B. A. (2005). Multimodal integration: fMRI, MRI, EEG, MEG. *Advanced Image Processing in Magnetic Resonance Imaging* (223–265).

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425–2430.

Hero, A. O., Ma, B., Michel, O., & Gorman, J. (2001). Alpha-divergence for classification, indexing and retrieval. *Communication and Signal Processing Laboratory, Technical Report CSPL-328, U.Mich.*

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks, 13*(4), 411–430.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452*(7185), 352–355.

Li, Y. O., Adali, T., & Calhoun, V. D. (2007). Estimating the number of independent components for functional magnetic resonance imaging data. *Human Brain Mapping, 28*(11), 1251–1266.

Liu, Z., Ding, L., & He, B. (2006). Integration of EEG/MEG with MRI and fMRI. *Engineering in Medicine and Biology Magazine, IEEE, 25*(4), 46–53.

Liu, J., Pearlson, G., Windemuth, A., Ruano, G., Perrone-Bizzozero, N. I., & Calhoun, V. (2009). Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. *Human Brain Mapping, 30*(1), 241–255.

Ma, L., Wang, B., Chen, X., & Xiong, J. (2007). Detecting functional connectivity in the resting brain: a comparison between ICA and CCA. *Magnetic Resonance Imaging, 25*(1), 47–56.

MacDonald, H. (2008). *Behavioural and neuroimaging studies of the influence of semantic context on the perception of speech in noise.* (Master of Science, Department of Neuroscience, Queen's University). Thesis (Master, Neuroscience Studies), Queen's University.

MacDonald, H., Davis, M. H., Pichora-Fuller, K., & Johnsrude, I. S. (2008). Contextual influences: perception of sentences in noise is facilitated similarly in young and older listeners by meaningful semantic context; neural correlates explored via functional magnetic resonance imaging (fMRI). *The Journal of the Acoustical Society of America, 123*(5), 3887.

McKeown, M. J., Makeig, S., Brown, G. G., Jung, T. P., Kindermann, S. S., Bell, A. J., et al. (1998). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping, 6*(3), 160–188.

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M. A., et al. (2004). Learning to decode cognitive states from brain images. *Machine Learning, 57*, 145–175.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social Cognitive and Affective Neuroscience, 4*(1), 101–109.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron, 63*(6), 902–915.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex, 20*(10), 2486–2495.

Peelle, J. E., Johnsrude, I. S., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience, 4*, 51.

Pereira, F. (2007). *Beyond brain blobs: Machine learning classifiers as instruments for analyzing functional magnetic resonance imaging data* ProQuest.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage, 45*(1), S199–S209.

Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers 10*(3), 61-74.

Ramezani, M., & Ghaemmaghami, S. (2010). Towards genetic feature selection in image steganalysis. *Consumer Communications and Networking Conference (CCNC)*, IEEE, 1–4.

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15*(8), 1261.

Savopol, F., & Armenakis, C. (2002). Merging of heterogeneous data for emergency mapping: data integration or data fusion? *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 34*(4), 668–674. doi:10.3389/fnhum.2010.00051

Schroeder, M. R. (1968). Reference signal for signal quality studies. *The Journal of the Acoustical Society of America, 44*(6), 1735–1736.

Schrouff, J., Kussé, C., Wehenkel, L., Maquet, P., & Phillips, C. (2012). Decoding semi-constrained brain activity from fMRI using support vector machines and gaussian processes. *PLoS One, 7*(4), e35860.

Specht, K., Zahn, R., Willmes, K., Weis, S., Holtel, C., Krause, B. J., et al. (2009). Joint independent component analysis of structural and functional images reveals complex patterns of functional reorganisation in stroke aphasia. *NeuroImage, 47*(4), 2057–2063.

Sui, J., Adali, T., Pearlson, G. D., Clark, V. P., & Calhoun, V. D. (2009a). A method for accurate group difference detection by constraining the mixing coefficients in an ICA framework. *Human Brain Mapping, 30*(9), 2953–2970.

Sui, J., Adali, T., Pearlson, G. D., & Calhoun, V. D. (2009b). An ICA-based method for the identification of optimal FMRI features and components using combined group-discriminative techniques. *NeuroImage, 46*(1), 73–86.

Sui, J., Adali, T., Pearlson, G., Yang, H., Sponheim, S. R., White, T., et al. (2010a). A CCA ICA based model for multi-task brain imaging data fusion and its application to schizophrenia. *NeuroImage, 51*(1), 123–134. doi:10.1016/j.neuroimage.2010.01.069

Sui, J., Adali, T., Li, Y., Yang, H., & Calhoun, V. D. (2010b). A review of multivariate methods in brain imaging data fusion. *Medical Imaging 2010 - Biomedical Applications in Molecular, Structural, and Functional, Imaging, 7626* (pp. 1–11). doi:10.1117/12.843922

Theodoridis, S., & Koutroumbas, K. (2003). *Pattern recognition*. New York: Academic Press.