

Published in final edited form as:

Med Image Anal. 2009 June ; 13(3): 392–404. doi:10.1016/j.media.2009.01.001.

Automated Classification of fMRI Data Employing Trial-based Imagery Tasks

Jong-Hwan Lee¹, Matthew Marzelli², Ferenc A. Jolesz¹, and Seung-Schik Yoo^{1,3}

¹ Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, MA, USA

² Department of Biomedical Engineering, Boston University, MA, USA

³ Department of BioSystems, Korea Advanced Institute of Science and Technology, Daejeon, Korea

Abstract

Automated interpretation and classification of functional MRI (fMRI) data is an emerging research field that enables the characterization of underlying cognitive processes with minimal human intervention. In this work, we present a method for the automated classification of human thoughts reflected on a trial-based paradigm using fMRI with a significantly shortened data acquisition time (less than one minute). Based on our preliminary experience with various cognitive imagery tasks, six characteristic thoughts were chosen as target tasks for the present work: right hand motor imagery, left hand motor imagery, right foot motor imagery, mental calculation, internal speech/word generation, and visual imagery. These six tasks were performed by five healthy volunteers and functional images were obtained using a T_2^* -weighted echo planar imaging (EPI) sequence. Feature vectors from activation maps, necessary for the classification of neural activity, were automatically extracted from the regions that were consistently and exclusively activated for a given task during the training process. Extracted feature vectors were classified using the support vector machine (SVM) algorithm. Parameter optimization, using a k -fold cross-validation scheme, allowed the successful recognition of the six different categories of administered thought tasks with an accuracy of 74.5% (mean) \pm 14.3% (standard deviation) across all five subjects. Our proposed study for the automated classification of fMRI data may be utilized in further investigations to monitor/identify human thought processes and their potential link to hardware/computer control.

Keywords

Pattern recognition; Support vector machine (SVM); Neuroimaging; functional MRI; Brain-computer interface (BCI)

1. Introduction

Automated characterization of brain activities, and a concurrent attempt to interpret corresponding human thoughts, is an emerging research field. Specifically, automatic interpretation and classification of neuroimaging data may hold important keys for understanding the human mind, which has raised interests due to the potential commercial/

Corresponding author: Seung-Schik Yoo, Ph.D. MBA, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis St., Boston, MA 02115, E-mail: E-mail: yoo@bwh.harvard.edu, Tel: 617-525-3308, Fax: 617-525-3330.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

clinical applications. These applications include motor rehabilitation (Papathanasiou et al., 2003; Sharma et al., 2006), neurosurgical planning (Gasser et al., 2005; Yoo et al., 2004a), and brain-computer interface (BCI) for quadriplegic or 'locked-in' patients (Pfurtscheller et al., 2003; Hochberg et al., 2006). Among the neuroimaging modalities, functional MRI (fMRI) measures and maps dynamic neuronal activities based on the detection of blood-oxygenation-level-dependent (BOLD) signal contrast (Kwong et al., 1992; Ogawa et al., 1992).

fMRI allows for repeated non-invasive measurements of brain function and has been widely used in various neurocognitive research studies. Most of the current fMRI-based neuroimaging techniques have been based on the generation of a parametric map of the neural activity (measured by BOLD signal) that is correlated with paradigm-driven task-conditions. Subsequent analysis of the spatial distribution of activation is typically done manually without assistance from advanced, algorithm-based, pattern classification. Meanwhile, the use of pattern classification/recognition algorithms has been widely accepted in the analysis of electroencephalography (EEG) data, such as in the classification of spatio-temporal features of EEG signal recordings (Wolpaw et al., 2002).

Recent studies have addressed the concept of automatic pattern classification of fMRI data acquired from multiple sensory, motor, and cognitive performances. For example, Cox and Savoy (2003) discriminated neural responses from the visual areas during the viewing of ten visual objects with 53% asymptotic accuracy (chance=10%). Martinez-Ramon et al. (2006), on the other hand, segmented the whole brain into functional areas based on a neuroanatomical atlas in order to increase the classification accuracy for sensory, motor, and cognitive tasks at different magnetic field strengths (1.5- and 4-T). This approach resulted in the achievement of relatively higher classification accuracy (less than 10% of misclassification rates) in comparison to the single whole brain approach. Additionally, Mourao-Miranda et al. (2005) presented binary classification of fMRI data from a visual stimulus matching test (face/location images) with an average accuracy greater than 90%. These fMRI data classifications, however, utilized a block-based design (repeated blocks of task & rest periods) as a task paradigm, which lengthens the data acquisition time and limits the flexibility of examining a trial-based neural activation (Zarahn et al., 1997). In addition, the adopted paradigms were based solely on a subject's overt performance during sensorimotor or cognitive tasks or response to external stimuli. As a result, these previous studies did not systematically test subject-driven (initiated) thought processes.

The purpose of the present study is to investigate the feasibility of automatic identification/classification of multiple human thoughts (in the form of specific imagery tasks) performed in a single trial over a short-time period (<5-sec) within a 50-sec fMRI scan including pre- and post-trial resting periods. Classification is quite challenging since BOLD signal intensities induced from imagery tasks are much weaker than those induced from actual performances, thus reducing detection sensitivity (Porro et al., 1996; Stippich et al., 2002) and resulting in a greater degree of inter-scan/inter-subject variability (Yoo et al., 2007). In addition, the short-time, trial-based task paradigm decreases the BOLD signal contrast compared to the block-based task paradigm (Jech et al., 2005). In order to address these challenges, we implemented automated procedures enabling (1) the extraction of spatial features of activation and (2) the optimization of classification parameters. Short fMRI scans (less than one minute) of a single thought-trial were repeated seven times per task to generate training data sets ($n=4$) and testing data sets ($n=3$). Using various combinations of the training/testing data sets generated for each task, the potential range of classification accuracy was systematically examined. The classification results according to either decreased scan duration or reduced spatial resolution were also investigated via simulation.

In this paper, we addressed the classification of fMRI data obtained during the performance of imagery tasks. To our knowledge, automatic classification of brain activations via fMRI from (a) trial-based and (b) covert imagery tasks has never before been demonstrated. These two elements, which are capable of distinguishing thought-processes occurring over a short-time period (less than 50-sec including 5-sec of task-period), are crucial for employing more realistic neuro-cognitive tasks in fMRI (i.e. compared to the overt tasks based on the widely used block-based task paradigm), thus enabling potential links to brain-computer interface (BCI) systems (Weiskopf et al., 2007; Wolpaw, 2007; Birbaumer and Cohen, 2007). We adopted an existing artificial neural network-based pattern classifier, support vector machine (SVM), to classify the spatial activation patterns associated with performance of imagery tasks (i.e. subject-initiated thought-processes), as detected by fMRI. In order to account for the significant inter-session and inter-subject variability of activation that is commonly observed in fMRI (Smith et al. 2005; Siedentopf et al., 2008), we developed a region-of-interest (ROI)-based feature extraction scheme to derive the feature vectors based on the individual-specific activation pattern. The scheme was further optimized by deploying a cross-validation (CV) scheme using training data within the acquired fMRI scans. The classification performance was also investigated using combinations of fMRI scans as training and testing data across the sessions and subjects.

2. Materials and Methods

2.1. Study Overview

Six mental imagery tasks were used to test the developed method for automated identification of human thoughts. An automated method for the selection of optimal feature vectors was implemented based on the individual's specific patterns of brain activity. This type of 'data-driven' selection of feature vectors is more robust considering the variability of activation in comparison to 'hypothesis-driven' approaches, whereby the feature vector is selected from anatomically segmented brain regions (Martinez-Ramon et al., 2006). As for a classification algorithm, the support vector machine (SVM) was employed due to its computational efficiency and robustness (Burges, 1998).

2.2. Study Participants and MRI Parameters

A local IRB committee on studies involving human subjects approved this study. Five right-handed healthy volunteers (22–35 years in age; two females; noted as S1 through S5; non-smokers) without a history of neurological or psychiatric conditions participated in this study. Before the experiment (i.e. 24 hours prior to the scan), the subjects were not allowed to consume nicotine and alcohol, which could have affected neural and/or cerebro-vascular activity (Perthen et al., 2007). The consumption of caffeinated beverages, such as coffee and tea, was discouraged, but not prohibited. The handedness of participants was examined according to the Edinburgh Handedness Inventory (Oldfield, 1971). The study was conducted in a 3 Tesla clinical scanner (Signa VH, GE Medical Systems) using a standard birdcage head coil. Prior to the fMRI scan, a T1-weighted 3-plane localizer was acquired to prescribe the imaging locations for the echo planar imaging (EPI) scan. An EPI sequence was then used to image a volume covering most of the brain, including the superior part of cerebellum, for the detection of BOLD signal associated with neural activity (13 axial slices, flip angle of 80°, TE/TR=40/1000msec at a resolution of 64 frequency and 64 phase encodes, 5mm thickness with a 1mm gap, 240mm square field-of-view: FOV). A total of 60 volumes were acquired in one minute. The first 10 volumes (dummy data) were excluded from further analysis to allow the T₁-related signals to equilibrate.

2.3. Task Paradigm

We employed a trial-based design in which the task started at the 15th scan for 5 seconds (Fig. 1A). The timings of the start and the end of the task were presented to the subject in the MRI system via a pre-recorded voice using an auditory headset (Avotec, Jensen Beach, FL). The participants were asked to close their eyes and generate six different types of mental tasks that the observer could not overtly detect. These six tasks were selected based on the preliminary experiment (data not shown) where we asked ten separate individuals to rate the ease and reliable reproduction of the tasks in multiple scans. The six mental tasks (with corresponding acronyms) were (1) right hand motor imagery (RH), (2) left hand motor imagery (LH), (3) right foot motor imagery (RF), (4) mental calculation (MC), (5) internal speech generation (IS), and (6) visual imagery (VI).

For the motor imagery tasks, the participants were asked to imagine clenching their hands (right hand for RH and left hand for LH) or wiggling the toes in their right foot (RF) at approximately 2 times per second (2Hz) without producing actual movement. For the MC task, subjects were asked to sequentially subtract an odd number (such as 3) from a two digit number greater than 20 as fast as they could (for example, 22-19-16-13- ...). The IS task required subjects to internally speak short sentences of choice (such as the Pledge of Allegiance or part of a rap song) without speaking out loud. Finally, the VI task asked subjects to imagine a static scene (s) as vividly as possible. These scenes could include natural static scenes, symbols, or faces; however, they were asked to only imagine a single type of scene (for example, a face and symbol should not be mixed during the session) to minimize the variations of region-specific activations according to the different types of visual imagery content (O'Craven and Kanwisher, 2000).

The participants were asked to perform each task seven times (42 performances total) randomly throughout 42 separate scans (one imagery task per scan). Randomization of the task order minimizes potential learning of a task strategy through sequential administration of the task, which tends to decrease/inhibit the activations via habituation (Loubinoux et al., 2001). The number of trials for each of the six tasks was decided based on the restriction of the fMRI study logistics (less than 1 hour to maintain a subject's comfort level). The subjects were also asked to use the same task strategy for each imagery task in order to minimize task-dependent variability in performance (deCharms et al., 2004; Yoo et al., 2007). The absence of any actual motion during the motor imagery tasks was constantly monitored during the scans. After the scanning, subjects informed us of the strategies they used during the six imagery tasks. Since we were also interested in the reproducibility of the method, all participants were asked to return to perform the same tasks on a separate visit.

2.4. Data Analysis and SVM Classification

2.4.1. Overview—The overall flow diagram of fMRI data processing is shown in Fig. 1B. After image reconstruction and realignment (motion correction), for a given fMRI scan, a cross-correlation (CC) analysis was performed between each voxel's BOLD time series and a reference hemodynamic response function (HRF) to obtain the voxel-wise correlation coefficient. The reference HRF was designed using SPM2 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm) to reflect the study design (Fig. 1A). Note that low-frequency drift, which is the largest source of fMRI noise and is typically observed below 0.02 Hz (Smith et al., 1999; Tanabe et al., 2002), was not considered in the model because the duration of the BOLD time series used in our data (50-sec) represents the minimum frequency of 0.02 Hz. Since the only signal components measured were those showing the frequency characteristics above 0.02 Hz, the presence of noise due to low-frequency drift is not likely to be a key contributing factor to the analysis results. Moreover, the drift noise usually seen at non-cortical areas (e.g. brain edges, brain-cerebrospinal fluid

interfaces, and the anterior skull base; Tanabe et al., 2002; Lund et al., 2006) would appear for all tasks; therefore, the spurious activation patterns underlying these regions would be removed during our feature extraction technique whereby only exclusively activated voxels among the tasks were selected as a ROI (see '2.4.3. Feature Vector Extraction across the Six Tasks' for the details).

The resulting voxel-wise CC maps were generated upon the acquisition of fMRI data from the subject ($n=42$). Then, the CC maps from seven scans per task were divided into training data sets ($n=4$) and testing data sets ($n=3$) for the classification experiment. The number of combinations resulting from four training and three testing scans per task was 35 ($=7C_4$) and thus the total number of combinatorial choices of training and testing data sets across the six tasks was 35^6 . The classification test for all of these sets would be too computationally exhaustive. Therefore, 30 randomly sampled combinations of training/testing data sets across all six tasks were prepared instead, resulting in 30 values of hit rates. Due to the small sample size, the resulting sample mean of the hit rates would follow the t -distribution with a mean,

m (unknown population mean), and standard deviation, $\frac{\sigma_s}{\sqrt{n}}$ (i.e. standard error of the mean), n where σ_s is the sample standard deviation and n (i.e. $n=30$ in our case) is the number of samples (Upton and Cook, 2001). The confidence interval for the population mean is

$m_s \pm t^* \frac{\sigma_s}{\sqrt{n}}$, where m_s is a sample mean, t^* is an upper $(1-\alpha)/2$ critical value (α : a confidence level) for the t -distribution with $(n-1)$ degrees-of-freedom (Upton and Cook, 2001). For our adopted number of samples ($n=30$; d.f.=29), an interval for a population mean with a 95% confidence level was approximately $m_s \pm 0.37\sigma_s$.

For each data set, there are two stages of data processing - (1) a stage for the processing of training data to find optimized parameters and (2) a stage for the classification of testing data. The training process includes the application of a threshold p -value, the selection of task-specific ROIs, the extraction of feature vectors based on the task-specific ROIs, SVM training, and a subsequent iteration for ranges of p -values. The testing process is aimed at identifying a type of task (out of the six tasks used) based on the SVM classification.

2.4.2. Selection of Region-of-Interest (ROI) Representing Each Task Type—A functional brain map with activation information in a 3-dimensional (3-D) space needs to be transformed into a 1-dimensional (1-D) feature vector so that the resulting vector can be utilized as an input for a classifier. In order to achieve maximum classification accuracy for different tasks, the feature vectors should represent maximal variability for inter-class and minimum variability for intra-class. The removal of redundant information is also desired for computational efficiency. Meanwhile, fMRI experiments cast unique sets of challenges since large variations in activations are commonly encountered, especially in studies involving short scan times (thus smaller sample sizes during statistical calculation) and short trial durations due to the inconsistency in task performances (Yoo et al., 2007). In order to account for these types of variability among trials, as well as under- and over-activation in the context of a short, trial-based design, an automated construction of the functional map reflecting the underlying spatial distribution of activation was needed. In this context, a voxel was included within a ROI of each task if the voxel was active in three or more scans in the training data. For example, Fig. 2 shows a defined ROI from the RF task scanned from one female subject (S1). The 3rd and 4th scans correspond to under- and over-activated scans, respectively. By applying the presented ROI definition, task-related activation regions were successfully extracted without manual intervention.

2.4.3. Feature Vector Extraction across the Six Tasks—Since the neural substrates that were activated during the six mental tasks could involve common functional areas (Mellet

et al., 1998; Yoo et al., 2004b) (example shown in Fig. 3A~F), overlapping regions of activation such as the basal ganglia and auditory areas may exist. For classification purposes, however, it is important to separate out task-specific activation areas so that they are as exclusive as possible. Therefore, we obtained a binary mask by including exclusively activated regions among the six ROIs. Consequently, the voxels within the mask were assigned as elements of a feature vector, which is an input for a classifier. The exclusively activated regions ($p < 0.01$; $z\text{-score} > 2.58$; $d.f. = 49$) for the six imagery tasks, as in the case of Fig. 3A~F, were color-coded and are shown in Fig. 3G. These regions were selected, without manual intervention, as the mask for a feature vector representation.

2.4.4. Classification using Support Vector Machine (SVM)—The SVM algorithm was deployed as a classifier. Among many supervised (with target values) classification algorithms for automatic pattern recognition, SVM is computationally efficient and has a good generalization capability due to its robustness in analyzing high dimensional data, even for a small training set (Burges, 1998). In this study, an extension of voting strategies based on the ‘one-against-one’ scheme (Hsu and Lin, 2002) was adopted in order to apply the SVM to the classification of more than two categories (six imagery tasks in this work). In addition, a Gaussian radial basis function (RBF) was selected as a nonlinear transformation kernel due to its superior performance over other types of functions, such as polynomial and sigmoidal functions (Burges, 1998). In this study, we utilized the LIBSVM (www.csie.ntu.edu.tw/~cjlin/libsvm/) library, which supports these functions in the MATLAB (Mathworks, Natick, MA) computing environment.

2.4.5. Parameter Optimization using Cross Validation Scheme—A k -fold cross validation (CV) scheme during the SVM training (Fan et al., 2005) was used to determine the optimal p -value and concurrent optimization of classification parameters for each combination of training/testing data sets. This k -fold CV scheme, which is widely utilized in parameter optimization of classifiers (Huang and Change, 2006; Goulermas et al., 2005), was originally aimed at finding optimal SVM parameters (i.e. cost function parameter C and kernel parameter γ for the RBF kernel; Fan et al., 2005) using the extracted feature vectors for training data. However, due to the inter-session/subject variations in activation patterns of fMRI data (again, we are using a short trial-based task design), the statistical threshold value (i.e. p -value used to determine activation) is another important variable affecting the selection of feature vectors. As a result, the p -value needed to be optimized depending on the session/subject in order to find a suitable ‘strength’ of activation pattern.

Figure 4 shows the block diagram of our employed parameter optimization method utilizing the k -fold CV scheme with the ‘grid-search’ method (i.e. search within 2-D parameters). According to the k -fold CV scheme, actual training data was divided into k folds of equal size. The data of each fold was tested using the SVM classifier trained on the remaining data of $k-1$ folds. In our training data size (24 scans from six ‘classes’ of the tasks), we conducted a 5-fold CV scheme. The use of a 5-fold CV scheme was also adopted in the study by the SVM developers (Fan et al, 2005). For our data, in order to divide the 24 scans of training data into 5 folds with an equal size of six scans (i.e. same as the number of classes) for each fold, it was necessary to include duplicate scans among the folds (i.e. maximum of four duplicate scans in a single-fold). On the other hand, since the classifier could be over-fitted toward the duplicate training data, it may also be necessary to define the validation sets without duplicate data. Thus, by considering our training data ($n=24$), a 4-fold CV scheme was also adopted whereby each fold consisted of six non-duplicate scans.

The test for each parameter was denoted as a ‘validation test’ and the resulting accuracy was referred to as the ‘validation accuracy’. Therefore, by using the k -fold CV scheme, there were k validation accuracies corresponding to k validation tests. Throughout the k validation tests,

the averaged value of the k validation accuracies was regarded to as the CV accuracy. The optimal C and γ parameters, representing the maximum CV accuracy, were searched by comparing the CV accuracies obtained from the varying C and γ values of the SVM. The range of tested pairs of C and γ were in exponentially growing sequences (i.e. $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ & $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$). Using the p -values of $p < 0.05$, $p < 0.01$, $p < 5 \times 10^{-3}$, $p < 10^{-3}$, $p < 5 \times 10^{-4}$, and $p < 10^{-4}$, the CV accuracies were iteratively obtained. As a result, an optimal p -value and its subsequent ROIs, along with SVM parameters (C & γ), were automatically selected from the maximum CV accuracy and utilized in the following testing process.

2.4.6. Classification Experiment during Testing Process—After the optimization of the ROIs (for feature vector selection) and SVM parameters during the training process, we tested the classification performance using the remaining testing data (3 scans per task). To do so, the feature vector of the CC map was extracted from the optimized ROIs and fed into the trained SVM classifier. Finally, from the total one-against-one binary classification results of the SVM, the maximally voted task was regarded as an identified task. Since the type of intended task during the testing process was known, the hit rate was calculated as the percentage of correctly classified scans out of 18 total testing scans in each session (chance $\approx 16.7\%$; i.e. one out of six tasks can be chosen). Also, the relative accuracies for each of the six types of imagery tasks were pooled over all sessions and subjects.

2.5. Simulation of Effect of the Reduced Acquisition Time and Spatial Resolutions

We also examined the effect of scan time (i.e. number of volumes) and spatial resolution (i.e. size of voxel) on the classification performance. Using the original 50-sec task paradigm for each trial (excluding a 10-sec dummy data acquisition in the beginning), it took about one hour to obtain fMRI data corresponding to all 42 trials (7 trials for each task, plus scanner operation time and a short relaxation time between scans). Accordingly, shortening the scan time for each trial may be desired without compromising the quality of the classification. In order to simulate different lengths of scan time, the time-series data from each scan was truncated to 40-, 30-, and 20-sec from the original 50-sec EPI data acquisition. Subsequently, the classification experiments were repeated and accuracies were measured. The effect of spatial resolution (in terms of the voxel size) on the classification performance was also examined. For every acquisition time, the EPI data (volume= $64 \times 64 \times 13$ voxels; size of voxel= $3.75 \times 3.75 \times 5.5 \text{ mm}^3$; denoted as R1) was sub-sampled in the in-plane dimension to create lower resolution data (size of voxel= $7.5 \times 7.5 \times 5.5 \text{ mm}^3$; R2). This data was then smoothed using a 3-D $15 \times 15 \times 11 \text{ mm}^3$ full-width-at-half-maximum (FWHM) Gaussian kernel and re-sampled to further reduce spatial resolution of the data (size of voxel= $15 \times 15 \times 11 \text{ mm}^3$; R3). The resulting low-resolution (R2 & R3) data via simulation was processed using the same training and testing processes for classification.

2.6. Frequency of Activation (FOA) Depending on Cortical Areas

In order to qualitatively evaluate the functional activations acquired from two fMRI sessions for each of the five participants, a frequency of activation (FOA) was counted for different cortical areas. These cortical areas were automatically segmented into 16 sub-areas according to their functional relevance (listed in Fig. 8A) using the Brodmann's Area (BA) and Automated Anatomical Labeling (AAL) indices (Lee et al., 2008). Among the ROIs defined at the level of the optimized p -values (i.e. the ROIs defined from the CV scheme; see section 2.4.5 for definition), the ROI showing the maximum hit rate from the 30 sets of training and testing data (see section 2.4.1) was delineated (the max. hit rate and the corresponding optimal p -value are shown in Table 1). If any activation cluster of the delineated ROI with a volume greater than $5 \times 5 \times 5 \text{ mm}^3$ was detected, the corresponding anatomical area was considered activated for a given task. Since the ROI that resulted in a maximum hit rate of each session was used for the calculation of the FOA, the maximum number of counts was 10 across the

two sessions for all five subjects. Compared to reporting the activation counts from each of the fMRI scans, this FOA counting strategy would be less susceptible to inter-scan variability of activation due to, for example, delay and maximum intensity of hemodynamic response and sensitivity toward the selected p -value for definition of an activated voxel (Smith et al., 2005; Siedentopf et al., 2008).

2.7. Cross-Session Test (CST) and Within-Session Test (WST)

In the previous analysis, the classification experiments were separately conducted using the data acquired within each session for each subject. On the other hand, in order to show the robustness of our classification method regarding the data obtained from different sessions, we also analyzed the classification performance across sessions for each subject. Since each subject underwent two separate fMRI sessions, the data from each of the two sessions were alternately used as training and testing data. By employing the classification procedure illustrated in Fig. 1B, the hit rates corresponding to each of the two sessions were obtained from each subject.

Potential misalignment of imaging volumes from different sessions may severely degrade the classification accuracy. Therefore, the EPI volumes obtained from all sessions and subjects were normalized into a standardized Montreal Neurological Institute (MNI) space. In detail, the realigned volumes in an individual space ($X:Y:Z=64:64:13$, voxel size= $3.75 \times 3.75 \times 5.5$ mm³) were transformed into the standardized MNI space based on a local optimization of the 12-parameter affine transformation (Ashburner and Friston, 1997). The resulting volumes were re-sampled into 4mm isotropic voxels ($X:Y:Z=41:48:35$) to reduce the computational load. Finally, the spatially normalized EPI volumes were smoothed using an 8mm isotropic FWHM Gaussian kernel to reduce the spatial noise. This anatomical normalization process also enabled testing of classification performance between/across the subjects.

An activation pattern of each normalized scan was then calculated from the general linear model (GLM), which utilizes a global minimization of linear least-squares error between the BOLD time series' within a brain region and a reference HRF following a task-paradigm (Worsley and Friston, 1995). This part of the processing was done using the SPM2 toolbox available in the MATLAB computing environment. In order to extract a feature vector representing each task from the activation patterns, a ROI was defined from the voxels that were active during at least three of the seven scans performed for each task from one of the two sessions (i.e. training session). Finally, only exclusively activated voxels among the ROIs representing the six tasks were used to extract the feature vectors of the 42 scans (i.e. seven scans per task) in the remaining session (i.e. testing session). The classification results from a within-session test (WST) were also obtained using the normalized data. For the WST, 30 randomly selected sets of training data (4 scans) and testing data (3 scans) per task were again used to calculate the range of hit rates.

2.8. Between-Subject Test (BST)

The prediction of thought processes from a new subject without previous training data sets, based on the knowledge of 'grouped' activation patterns, would be of interest since the individual-specific activation pattern may be classified without the need for individual training. In order to show the feasibility of making this type of prediction, the classification performance of a between-subject test (BST) was assessed. Given that five subjects participated in the study, there are four choices on the selection of subjects for training and testing data (i.e. one, two, three, and four subjects' data for training, and the remaining subjects' data for testing). The classification tests for these four choices were again conducted following the procedure shown in Fig. 1B. More specifically, the GLM-based activation maps of the spatially normalized fMRI scans were employed (see the section 2.7). The voxels of the ROIs used for the cross-session

tests (i.e. active for at least three scans and exclusive across all six tasks) were included in the ROI for feature vector extraction of this BST. The data from the remaining subjects' sessions were then classified and hit rates were subsequently calculated.

3. Results

3.1. Selected Exclusively Task-Specific ROIs for the Extraction of Feature Vectors

The exclusively task-specific ROIs for the selection of the feature vector are shown in Fig. 5. Since these exclusive ROIs were based on the maximum classification accuracy in each session for each subject, the corresponding optimal p -values were slightly different (Table 1). From the examination of the two sessions for all the subjects, there were degrees of variations in terms of spatial distribution and the size of the ROIs. For example, the activations of motor tasks (RH, LH, & RF) for S1 moved to the posterior part of the brain for the 2nd session as compared to the 1st session. For S4, the ROI for the left-hand imagery task appeared on the left inferior frontal and parietal areas, ipsilateral to the task, but was later detected on the contralateral motor area in the 2nd session. During the 2nd session for S5, the activations for the LH and RF tasks became less-apparent, compared to the 1st session. The IS and VI tasks were also involved in the areas where hand motor imagery tasks were activated.

3.2. Results of Classification

The classification results for each subject's session, measured by the k -fold leave-one-out CV accuracies as well as hit rates across 30 combinations of training-testing data sets, are summarized as a box-whisker plot (shown in Fig. 6). Overall, the resulting CV accuracies and hit rates across all of the subjects' sessions were qualitatively similar. The CV accuracies were $82.0\% \pm 10.3\%$ for the 4-fold CV and $82.5\% \pm 10.4\%$ for the 5-fold CV and the hit rates were $74.3\% \pm 14.2\%$ for the 4-fold CV and $74.5\% \pm 14.3\%$ for the 5-fold CV, respectively. Depending on the selection of training and testing data sets, there were variations of classification accuracies, even for the same subject in the same session. Among all 30 hit rates calculated for each session, a box represents the 25th, 50th, and 75th percentiles and whiskers denote the 10th and 90th percentiles. An average accuracy is marked with a black square dot and minimum- and maximum-accuracies are shown as 'x' marks. Overall, the worst average classification accuracy, observed from the 2nd session of S5, showed a hit rate of approximately 50%, which is still higher than a hit rate based on the probability of guessing (i.e. 1/6 16.7%). Four subjects, except one male subject (S5), showed higher than 80% average classification performance in at least one session, and all of the subjects showed higher than 80% in their maximum accuracies. During the 2nd session from S4, 23 sets out of 30 sets showed 100% hit rates. In general, however, the distribution of hit rates (between the 25th and 75th percentiles: shown as a box) varied about 5% to 20% across the subjects. The hit rates for each imagery task are shown in Fig. 7A. Although the average hit rates per task were within 70~80% across the tasks, the LH task showed relatively larger variations in hit rate which is shown as a box.

3.3. Classification from Simulated Reduced Acquisition Time and Spatial Resolutions

Classification accuracies for the EPI data simulating reduced acquisition time (40-, 30-, and 20-sec, excluding dummy scans) are shown in Fig. 7B. The simulated EPI data with a data acquisition time of 40-sec did not degrade the classification performance (examined by a paired t -test; $p > 0.05$; t -score=1.37; d.f.=299); however, a data acquisition time shorter than 30-sec began to severely reduce the hit rates for all of the subjects (a paired t -test of 50-sec vs. 30-sec: $p < 10^{-5}$; t -score=4.26; d.f.=299 and a paired t -test of 50-sec vs. 20-sec: $p < 10^{-32}$; t -score=13.53; d.f.=299). For two participants (S1 & S4), however, 80% average classification performance was maintained for simulated data sets spanning a duration as short as 30-sec. The classification accuracies examining three different voxel sizes are shown in Fig. 7C. For all subjects, the activation maps obtained from the spatial resolution of R2 (size of voxel

$=7.5 \times 7.5 \times 5.5 \text{mm}^3$) showed comparable classification results to the case of original spatial resolution of R1 (size of voxel $=3.75 \times 3.75 \times 5.5 \text{mm}^3$). However, the classification accuracies began to drastically degrade for the spatial resolution of R3 (size of voxel $=15 \times 15 \times 11 \text{mm}^3$) across all the subjects.

3.4. Cortical Areas of Activations as Defined by the SVM Training across the Subjects

The brain areas of activation showed large inter-subject variations. Figure 8B shows the FOA within the ROI which was defined by the optimized p -values while showing a maximum hit rate (Table 1). A different gray scale was applied with a FOA inside a circle (10 was the maximum number of occurrences of activation from 2 sessions across 5 subjects) for 16 representative cortical areas. The bar graphs of average FOA across the tasks for each region along with the standard deviation (whisker) are shown in Fig. 8C. Based on this FOA analysis, several anatomical areas, including the SMA (d), the left auditory area (h), the inferior frontal cortices (i), and the cingulate gyri (l), were more commonly activated than other parts of the brain throughout the six imagery tasks.

On the other hand, the dominant trends of region-specific activation were noted depending on the characteristics of imagery tasks. For example, the left primary motor area (b) was more frequently activated for the RH, RF and IS tasks than it was for the MC and VI tasks. Superior frontal gyri (e) and dorsolateral prefrontal gyri (j), along with cingulate gyri (l), were selectively activated during the MC task (Cowell et al., 2000) while left middle and inferior temporal gyri (m) were dominantly activated during the IS task (Shergill et al., 2004). The superior parietal lobe (c) and occipital lobe (g) were more frequently activated for the MC (FOA ≥ 6 in both hemispheres) and VI (FOA ≥ 5 in both hemispheres) tasks, which indicates the involvement of visual memory (Tang et al., 2006; Kosslyn and Thompson, 2003). Among the motor imagery tasks (RH, LH, & RF), leftward laterality of RH and RF was observed at the primary motor area (b); however, this trend was obscured during the motor imagery tasks of the left hand (Stinear et al., 2006).

3.5. Classification from the Cross-Session Test

The hit rates from the cross-session test (CST) along with the within-session test (WST) are summarized in Table 2. Firstly, the average hit rate from the WST using the normalized volumes (voxel size $=4 \times 4 \times 4 \text{mm}^3$; 77.6%) was comparable to that obtained from the original (unsmoothed) volumes ($3.75 \times 3.75 \times 5.5 \text{mm}^3$; 74.5%) shown in Fig. 6B and D. Overall, the classification performance of the CST was not significantly decreased compared to that of the WST in which the average hit rates (WST: CST; %) for each subject were (92.0: 81.0) for S1, (71.0: 69.1) for S2, (76.6: 71.4) for S3, (91.0: 95.3) for S4, and (57.5: 38.1) for S5. For S5, although the average hit rate from the 1st session was 71.9%, the 2nd session resulted in a 43% average hit rate, which suggests a deterioration of the classification performance of the CST. The average hit rates along with the standard deviation of the WST and CST across all subjects' normalized data were $77.6\% \pm 16.2\%$ and $71.0\% \pm 20.2\%$, respectively.

3.6. Classification from the Between-Subject Test (BST)

Figure 9 shows the average (bar) and standard deviation (whisker) of the resulting hit rates across all six tasks (A) and for each individual task (B). The index of the four choices on the number of training subjects (i.e. one, two, three, & four subjects) is denoted as 'a' through 'd'. The numbers of combinatorial selections for training and testing subjects corresponding to the choices 'a' through 'd' are 5 ($=_5C_1$), 10 ($=_5C_2$), 10 ($=_5C_3$), and 5 ($=_5C_4$), respectively. Higher hit rates were expected for a larger number of training subjects' data since more variations of feature vectors for each task could be modeled into the classifier. Again, the overall hit rates (i.e. $68.3\% \pm 10.1\%$) were still higher than the hit rate by chance. In the results from the four subjects' data as training data and one remaining subject's data as testing data (i.e. 'd' of Fig.

9A), the lowest hit rate of 51.2% was obtained from the subject S5, who also showed the lowest average hit rate for both within- and cross-session tests. The variation of classification performance depending on the type of imagery task is shown in Fig. 9B. Three motor imagery tasks (i.e. RH, LH, & RF) demonstrated consistently higher hit rates compared to the remaining non-motor imagery tasks (i.e. MC, IS, & VI). Overall, the visual imagery task (VI) showed the lowest hit rate. Information regarding computer hardware and computational time of each analytical step is summarized in Table 3.

4. Discussion

In this study, we reported an automated classification of six imagery tasks employing a short-time trial-based paradigm design. Our data-driven, feature selection scheme from the task-specific ROIs was applied to the classification of the six different tasks in the presence of large variability attributed to the short trial-based imagery tasks. A cross validation scheme was also employed for the optimization of the p -value, consequent feature selection, and SVM parameters and attained an average accuracy of $74.5\% \pm 14.3\%$ across all five subjects. Compared to the classification results obtained from overt or passive stimulation paradigms based on a long, block-based fMRI design (Cox and Savoy, 2003; Martinez-Ramon et al., 2006), we believe that imagery tasks based on a short, trial-based design can produce reasonable classification accuracy.

In terms of the study design, the six types of imagery tasks were performed in a random sequence to minimize the potential learning and adaptation effects. On the other hand, sequential performance of tasks and investigation on its effects toward the classification performance may potentially help to address important questions, such as which areas activate, and how commonly/exclusively they are activated, if a certain task order is selected, including the temporal dynamics of functional connectivity in the context of the habituation or memory consolidation (Loubinoux et al., 2001; Kemere et al., 2008).

Inter-Session/Subject Variability and Confounding Effects

While functional MRI is a noninvasive tool used to observe cortical activities, the large variations across sessions/subjects remain an unsettled issue (Aguirre et al, 1998; deCharms et al. 2004; Yoo et al. 2006). Consequently, the reproducibility of fMRI data is important in terms of data classification, which assumes consistent intra- and inter-class activation patterns, even though MR data in a high field magnet (3.0-T and 4.0-T) is reported to generate reasonable reproducibility in the mean sensitivities (Zou et al., 2005). The transitions/variability of activations was observed even for a single imagery task, which demonstrated that the thought process may not be exactly reproducible among individuals or even between sessions. As shown in Fig. 5, we have seen significant inter-subject variability, even with motor imagery tasks. The motor imagery task is believed to involve the primary motor areas. However, in reality, inter-session and inter-subject variability was observed (deCharms et al. 2004; Yoo et al., 2007). This observation is similar to the previous studies as reviewed by Sharma et al. (2006) in which the neural mechanism of motor imagery and the associated cortical activation patterns were investigated. It is conjectured that transient thought and emotional status may account for the variability of the activation characteristics as a confounder (Pessoa and Padmala, 2005; Giardino et al., 2007).

Moreover, imagery tasks may potentially be more vulnerable to these confounded effects than real stimulus- or execution-based tasks due to the relatively weak level of neuronal activity of imagery tasks (Grossman and Blake, 2001; Dechent et al., 2004). The weak level of activation may also be attributed to the single short-time (<5-sec) trial-based task paradigm (Buckner, 1998). On the other hand, the neuro-substances such as caffeine or nicotine may impose confounding effects on BOLD fMRI data such as reduced coupling of cerebral blood flow and

oxygen metabolism (Perthen et al., 2007) or declined learning of cognitive tasks (Bendlin et al., 2007). We put forth an effort to minimize these negative effects by defining consistently task-exclusive regions for the feature vector selection. Augmentation of the current classification approach by using multiple classifiers (Chou and Shen, 2006; Martinez-Ramon et al., 2006) may enhance the performance of the proposed implementation. In addition, the effect of the neuro-substances on the classification performance of fMRI data reflecting thought processes needs to be elucidated.

During the fMRI classification of multiple thought processes, the inherent similarity and subsequent confounding nature of the tasks posed multiple challenges. For example, the subjects may form a visual representation of the numbers used during the mental calculation (MC) task, and hence the activation patterns from the MC task may overlap with the activation patterns from the visual imagery (VI) task. This may lead to simultaneous VI task-related activation. Similarly, while performing the internal speech generation task (IS), one may visualize the singer of the song or the subject matter of the speech (e.g. the flag while reciting 'The Pledge of Allegiance'). As another example, the subject may visualize a physical object in the grasp of the hand while performing the hand motor imagery tasks (e.g. squeezing a tennis ball). In order to address these potential confounding effects among the tasks, only exclusively activated voxels across the tasks were chosen as the ROI for extraction of feature vectors. For example, the analysis of the frequency-of-activation (FOA) shown in Fig. 8B may exemplify the efficacy of the deployed method, whereby the occipital visual area 'g' was dominantly included in the ROIs only for the MC and VI tasks.

Effects of the Simulated Reduction in Volume Acquisition/Spatial Resolution

Examination of the effect of data acquisition time on classification performance showed that the accuracy was not reduced in most subjects for periods as short as 40-sec. For shorter data acquisition periods (simulating 30-sec and 20-sec data), however, the classification results were severely degraded for all subjects. This could be attributed to the small amount of available temporal data (thus reducing the statistical significance), combined with a failure to capture characteristics of the hemodynamic responses, which generally evolve about 10~12-sec following as short as 1-sec of a single trial (Buckner, 1998).

The classification accuracy attainable from only 40-sec of data acquisition (approximately 73% from the WST results of the six imagery tasks) was reasonable compared to that of EEG-based BCI experiments (e.g. about 80% adopting left- and right-hand motor imagery; 0.7~1 letter/min; Neuper et al., 2006). However, it is important to note that the 'simulated' shortening of the paradigm was truncated from the original fMRI data consisting of a 40-sec rest period following the task period. If a given task paradigm is presented in a relatively short time window to expedite the experiment, each scan of a single trial would possibly be corrupted by a non-linear adverse effect (i.e. carry over effect) from the previous trial (Aguirre, 2007). Therefore, this inherent delay in hemodynamic responses may potentially limit the temporal resolution for interpretation of thought processes in comparison to EEG-based BCI methods.

Regarding the effect of voxel size on classification performance, we found that reduced spatial resolution, simulated by a $32 \times 32 \times 13$ data matrix (size of voxel = $7.5 \times 7.5 \times 5.5 \text{mm}^3$; denoted as R2), could provide enough spatial resolution to discriminate the six selected imagery tasks. This indicates that the time saved by fewer frequency- and phase-encodes can be used to cover a larger brain volume during the same 1-sec TR period than is presently covered (13 slices). Consequently, the degree of freedom of activation maps across different tasks may increase and classification accuracies could potentially benefit from the expanded spatial distribution of activation maps. In addition, the saved time can be spent on increasing the number of data acquisitions per scanning period (thus more data are available for statistical processing), while maintaining the number of slice encodings or the SNR of the image. This may enhance the

temporal characteristics so that the temporal trajectory of neuronal activities could be better represented. As a result, the different onsets of a neural event can be detected and can lead to the characterization of functional connectivity, thus enhancing our understanding of human thought processes (Hernandez et al., 2002; Nair, 2005).

Laterality of the Hand Motor Imagery Tasks Based on the FOA

The FOA results revealed that the task-specific spatial pattern of activation observed from the subjects was in close agreement with previous studies on similar tasks. However, several unanticipated activations were also observed. Interestingly, for the non-dominant hand motor imagery task, the contralateral primary motor area showed a relatively lower activation than that of the dominant hand motor imagery task (index: b in Fig. 8B; the FOA in left M1 for RH = 7 vs. the FOA in right M1 for LH = 4). Both right-handedness (Edinburgh Handedness score > +80%) and the requirement for the FOA (i.e. consistently activated in three or more scans within each task as well as exclusively activated across the six tasks) might have contributed to the obscured laterality for the LH task. For example, handedness alters the laterality of activation depending on whether or not the subject is engaging their dominant hand (Begliomini et al. 2008). Group inference from multiple subjects is warranted to justify the observed laterality for the left-hand motor imagery task. Instead of the primary motor areas, a left dorsolateral prefrontal area (index: j) was dominantly activated for the LH task (FOA=7). This may be related to the role of the dorsolateral prefrontal cortex in the temporal storage of sensorimotor associations so that the selection of movement can be determined (Halsband and Lange, 2006). We conjecture that strong right-handedness may have required more effort to prepare for the LH task compared to the RH task, thus inducing recruitment of the prefrontal area by recalling the stored left-hand motor function during the task period.

Cross-Session and Between-Subject Tests

The fact that the average accuracies obtained from the cross-session tests ($71.0\% \pm 20.2\%$) were not significantly degraded compared to those from the within-session tests ($77.6\% \pm 16.2\%$) indicates that the thought-processes of each subject are consistent over a few weeks and even several months (e.g. the 2nd session of S2 was conducted seven-months later). Furthermore, the hit rates from the between-subject tests ($68.3\% \pm 10.1\%$) that are much higher than the hit rate by pure chance suggest that fMRI can be used as a reasonable predictor for classifying thought processes even when training sets obtained from other subjects are used to define a ROI for feature extraction and are subsequently used to train a classifier. This classification performance, however, needs to be corroborated by a further study that involves a greater number of participants.

The between-subject test showed that the classification accuracy of each image task (shown in Fig. 9B) decreased for non-motor tasks (e.g. the visual imagery task). We hypothesize that the complexity as well as the variability in the execution of non-motor imagery tasks may be responsible for this reduced accuracy. For example, during the visual imagery task, participants adopted various types of visualizations such as buildings and faces. Since the regional BOLD activation patterns for distinct categories of visual stimuli are highly specific within the early visual areas (Kay et al., 2008), the control/regulation of the categories of visual stimuli may enhance the performance of the VI task. We believe the real-time feedback of one's own brain function, with its utility to modulate region-specific neural activations and associated BOLD signals (Weiskopf et al., 2003; Yoo and Jolesz, 2002; Yoo et al., 2006), may reduce the variability in the task performance, thus improving the classification results. The utility of neurofeedback-enabled fMRI in this context requires further investigation.

5. Conclusion

We proposed a technique that allows the processing of patterns of neuronal activities as discrete units. The results indicate that fMRI can be successfully employed to detect subjects' intentions reflected on a short-time imagery trial without a previous training period even though the classification accuracy was diminished in non-motor imagery tasks (e.g. the visual imagery task). Due to the challenging task paradigm (i.e. six imagery tasks based on a short-time single-trial), the classification performance of the resulting fMRI data was vulnerable to the variability among scans. Despite this variability, the average accuracies from the cross-session and between-subject tests were reasonable and were again much higher than the hit rate by chance. The reasonable classification accuracy of fMRI data without a previous training period may create new opportunities in BCI, whereby fMRI can be used to optimize the EEG electrode configuration of a BCI. The adjunctive use of fMRI for EEG may reduce the time necessary to operate an EEG-based BCI system, which took several weeks to adapt its imagery performances (Wolpaw et al., 2002). The presented study of automatic identification of human thoughts may contribute to future studies that address the emerging area of fMRI data classification research as well as the potential implementation of fMRI-guided BCI systems.

Acknowledgements

Authors appreciate the technical support of Ms. Heather O'Leary for the data acquisition, the logistic support of Mr. Dong-Woo Hahn, and the editorial support of Mr. Samuel Polio. This work was partially supported by grants from NIH (R01-NS048242 to Yoo, SS and NIH U41RR019703 to Jolesz FA), the Korean Ministry of Commerce, Industry, and Energy (grant No. 2004-02012 to S.S. Yoo), and Gachon Neuroscience Research Institute Grant (to Yoo SS).

References

- Aguirre GK, Zarahn E, D'Esposito M. The variability of human, BOLD hemodynamic responses. *NeuroImage* 1998;8(4):360–369. [PubMed: 9811554]
- Aguirre GK. Continuous carry-over designs for fMRI. *Neuroimage* 2007;35:1480–1494. [PubMed: 17376705]
- Ashburner, J.; Friston, KJ. Spatial transformation of images. In: Frackowiak, RSJ.; Friston, KJ.; Frith, CD.; Dolan, RJ.; Mazziotta, JC., editors. *Human Brain Function*. Academic Press; USA: 1997. p. 43-58.
- Begliomini C, Nelini C, Caria A, Grodd W, Castiello U. Cortical activations in humans grasp-related areas depend on hand used and handedness. *PLoS ONE* 2008;3(10):e3388. [PubMed: 18846222]
- Birbaumer N, Cohen LG. Brain-computer interfaces: communication and restoration of movement in paralysis. *J Physiol* 2007;579:621–636. [PubMed: 17234696]
- Buckner RL. Event-related fMRI and hemodynamic response. *Hum Brain Mapp* 1998;6:373–377. [PubMed: 9788075]
- Burges CJC. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 1998;2:121–167.
- Carlson TA, Schrater P, He S. Patterns of activity in the categorical representations of objects. *J Cogn Neurosci* 2003;15(5):704–717. [PubMed: 12965044]
- Chou KC, Shen HB. Predicting protein subcellular location by fusing multiple classifiers. *J Cell Biochem* 2006;99(2):517–527. [PubMed: 16639720]
- Cowell SF, Egan GF, Code C, Harasty J, Watson JD. The functional neuroanatomy of simple calculation and number repetition: A parametric PET activation study. *Neuroimage* 2000;12(5):565–573. [PubMed: 11034863]
- Cox DD, Savoy RL. Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage* 2003;19:261–270. [PubMed: 12814577]

- de Charms RC, Christoff K, Glover GH, Pauly JM, Whitfield S, Gabrieli JD. Learned regulation of spatially localized brain activation using real-time fMRI. *Neuroimage* 2004;21(1):436–443. [PubMed: 14741680]
- Dechent P, Merboldt KD, Frahm J. Is the human primary motor cortex involved in motor imagery? *Cognitive Brain Research* 2004;19:138–144. [PubMed: 15019710]
- Egner T, Gruzeliier JH. EEG Biofeedback of low beta band components: frequency-specific effects on variables of attention and event-related brain potentials. *Clin Neurophysiol* 2004;115(1):131–139. [PubMed: 14706480]
- Fan RE, Chen PH, Lin CJ. Working set selection using second order information for training support vector machines. *J Machine Learning Research* 2005;6(12):1889–1918.
- Gasser T, Ganslandt O, Sandalcioglu IE, Stolke D, Fahlbusch R, Nimkysky C. Intraoperative Functional MRI: Implementation and Preliminary Experience. *NeuroImage* 2005;26(3):685–693. [PubMed: 15955478]
- Giardino ND, Friedman SD, Dager SR. Anxiety, respiration, and cerebral blood flow: implications for functional brain imaging. *Compr Psychiatry* 2007;48(2):103–112. [PubMed: 17292699]
- Goulermas JY, Findlow AH, Nester CJ, Howard D, Bowker P. Automated design of robust discriminant analysis classifier for foot pressure lesions using kinematic data. *IEEE Trans Biomed Eng* 2005;52(9):1549–1562. [PubMed: 16189968]
- Grossman ED, Blake R. Brain activity evoked by inverted and imagined biological motion. *Vision Res* 2001;41(10–11):1475–1482. [PubMed: 11322987]
- Halsband U, Lange RK. Motor learning in man: A review of functional and clinical studies. *J Physiol – Paris* 2006;99:414–424. [PubMed: 16730432]
- Hernandez L, Badre D, Noll D, Jonides J. Temporal sensitivity of event-related fMRI. *NeuroImage* 2002;17:1018–1026. [PubMed: 12377175]
- Hochberg LR, Serruya MD, Friehs GM, Mukand JA, Saleh M, Caplan AH, Branner A, Chen D, Penn RD, Donoghue JP. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 2006;442:164–171. [PubMed: 16838014]
- Hsu CW, Lin CJ. A comparison of methods for multi-class support vector machines. *IEEE Trans Neural Netw* 2002;13:415–425. [PubMed: 18244442]
- Huang HL, Chang FL. ESVM: Evolutionary support vector machine for automatic feature selection and classification of microarray data. *Biosystems*. 2006 Dec 16;[Epub ahead of print]
- Jech R, Dusek P, Wackermann J, Vymazal J. Cumulative blood oxygenation-level-dependent signal changes support the ‘time accumulator’ hypothesis. *Neuroreport* 2005;16(13):1467–1471. [PubMed: 16110273]
- Kay KN, Naselaris T, Prenger RJ, Gallant JL. Identifying natural images from human brain activity. *Nature* 2008;452:352–355. [PubMed: 18322462]
- Kemere C, Santhanam G, Yu BM, Afshar A, Ryu SI, Meng TH, Shenoy KV. Detecting neural state transitions using hidden Markov models for motor cortical prostheses. *J Neurophysiol*. 2008 2008 Jul 9;[Epub ahead of print]
- Kosslyn SM, Thompson WL. When is early visual cortex activated during visual mental imagery? *Psychological Bulletin* 2003;129(5):723–746. [PubMed: 12956541]
- Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R, Cheng HM, Brady TJ, Rosen BR. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci USA* 1992;89(12):5675–5679. [PubMed: 1608978]
- Lee JH, O’Leary HM, Park H, Jolesz FA, Yoo SS. Atlas-based multichannel monitoring of functional MRI signals in real-time: Automated approach. *Hum Brain Mapp* 2008;29(2):157–166. [PubMed: 17370340]
- Loubinoux I, Carel C, Alary F, Boulanouar K, Viallard G, Manelfe C, Rascol O, Celsis P, Chollet F. Within-session and between-session reproducibility of cerebral sensorimotor activation: a test–retest effect evidenced with functional magnetic resonance imaging. *J Cereb Blood Flow Metab* 2001;21:592–607. [PubMed: 11333370]
- Lund TE, Madsen KH, Sidaros K, Luo WL, Nichols TE. Non-white noise in fMRI: does modelling have an impact? *Neuroimage* 2006;29(1):54–66. [PubMed: 16099175]

- Martinez-Ramon M, Koltchinskii V, Heileman GL, Posse S. fMRI pattern classification using neuroanatomically constrained boosting. *NeuroImage* 2006;31:1129–1141. [PubMed: 16529955]
- Mellet E, Petit L, Mazoyer B, Denis M, Tzourio N. Reopening the mental imagery debate: lessons from functional anatomy. *NeuroImage* 1998;8(2):129–139. [PubMed: 9740756]
- Mourao-Miranda J, Bokde AL, Born C, Hampel H, Stetter M. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. *Neuroimage* 2005;28(4):980–995. [PubMed: 16275139]
- Nair DG. About being BOLD. *Brain Res Brain Res Rev* 2005;50:229–243. [PubMed: 16213027]
- Neuper C, Müller-Putz GR, Scherer R, Pfurtscheller G. Motor imagery and EEG-based control of spelling devices and neuroprostheses. *Prog Brain Res* 2006;159:393–409. [PubMed: 17071244]
- O’Craven KM, Kanwisher N. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J Cogn Neurosci* 2000;12(6):1013–1023. [PubMed: 11177421]
- Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci USA* 1992;89(13):5951–5955. [PubMed: 1631079]
- Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971;9(1):97–113. [PubMed: 5146491]
- Papathanasiou I, Filipovic SR, Whurr R, Jahanshahi M. Plasticity of motor cortex excitability induced by rehabilitation therapy for writing. *Neurology* 2003;61:977–980. [PubMed: 14557572]
- Pessoa L, Padmala S. Quantitative prediction of perceptual decisions during near-threshold fear detection. *Proc Natl Acad Sci U S A* 2005;102(15):5612–5617. [PubMed: 15800041]
- Pfurtscheller G, Müller GR, Pfurtscheller J, Gerner HJ, Rupp R. ‘Thought’ – control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. *Neurosci Lett* 2003;351(1):33–36. [PubMed: 14550907]
- Porro CA, Francescato MP, Cettolo V, Diamond ME, Baraldi P, Zuiani C, Bazzocchi M, di Prampero PE. Primary motor and sensory cortex activation during motor performance and motor imagery: A functional magnetic resonance imaging study. *J Neurosci* 1996;16(23):7688–7698. [PubMed: 8922425]
- Sharma N, Pomeroy VM, Baron JC. Motor imagery: A backdoor to the motor system after stroke? *Stroke* 2006;37:1941–1952. [PubMed: 16741183]
- Shergill SS, Brammer MJ, Amaro E, Williams SC, Murray RM, McGuire PK. Temporal course of auditory hallucinations. *Br J Psychiatry* 2004;185:516–517. [PubMed: 15572744]
- Siedentopf CM, Heubach M, Ischebeck A, Gallasch E, Fend M, Mottaghy FM, Koppelstaetter F, Haala IA, Krause BJ, Felber S, Gerstenbrand F, Golaszewski SM. Variability of BOLD response evoked by foot vibrotactile stimulation: influence of vibration amplitude and stimulus waveform. *Neuroimage* 2008;41:504–510. [PubMed: 18424181]
- Smith AM, Lewis BK, Ruttimann UE, Ye FQ, Sinnwell TM, Yang Y, Duyn JH, Frank JA. Investigation of low frequency drift in fMRI signal. *Neuroimage* 1999;9(5):526–33. [PubMed: 10329292]
- Smith SM, Beckmann CF, Ramnani N, Woolrich MW, Bannister PR, Jenkinson M, Matthews PM, McGonigle DJ. Variability in fMRI: a re-examination of inter-session differences. *Hum Brain Mapp* 2005;24:248–257. [PubMed: 15654698]
- Stinear CM, Fleming MK, Byblow WD. Lateralization of unimanual and bimanual motor imagery. *Brain Res* 2006;1095(1):139–147. [PubMed: 16713588]
- Stippich C, Ochmann H, Sartor K. Somatotopic mapping of the human primary sensorimotor cortex during motor imagery and motor execution by functional magnetic resonance imaging. *Neurosci Lett* 2002;331:50–54. [PubMed: 12359321]
- Tanabe J, Miller D, Tregellas J, Freedman R, Meyer FG. Comparison of detrending methods for optimal fMRI preprocessing. *Neuroimage* 2002;15(4):902–907. [PubMed: 11906230]
- Tang Y, Zhang W, Chen K, Feng S, Ji Y, Shen J, Reiman EM, Liu Y. Arithmetic processing in the brain shaped by cultures. *Proc Natl Acad Sci USA* 2006;103(28):10775–10780. [PubMed: 16815966]
- Upton, G.; Cook, I. *Introducing Statistics*. Vol. 2. Oxford University Press; 2001.

- Weiskopf N, Veit R, Erb M, Mathiak K, Grodd W, Goebel R, Birbaumer N. Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): methodology and exemplary data. *NeuroImage* 2003;19(3):577–586. [PubMed: 12880789]
- Weiskopf N, Sitaram R, Josephs O, Veit R, Scharnowski F, Goebel R, Birbaumer N, Deichmann R, Mathiak K. Real-time functional magnetic resonance imaging: methods and applications. *Magn Reson Imaging* 2007;25:989–1003. [PubMed: 17451904]
- Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol* 2002;113(6):767–791. [PubMed: 12048038]
- Wolpaw JR. Brain-computer interfaces as new brain output pathways. *J Physiol* 2007;579:613–619. [PubMed: 17255164]
- Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited—again. *Neuroimage* 1995;2(3):173–181. [PubMed: 9343600]
- Yoo SS, Jolesz FA. Functional MRI for neurofeedback: feasibility study on a hand motor task. *Neuroreport* 2002;13(11):1377–1381. [PubMed: 12167756]
- Yoo SS, Talos IF, Golby AJ, Black PM, Panych LP. Evaluating requirements for spatial resolution of fMRI for neurosurgical planning. *Hum Brain Mapp* 2004a;21(1):34–43. [PubMed: 14689508]
- Yoo SS, Fairney T, Chen NK, Choo SE, Panych LP, Park H, Lee SY, Jolesz FA. Brain-computer interface using fMRI: spatial navigation by thoughts. *Neuroreport* 2004b;15(10):1591–1595. [PubMed: 15232289]
- Yoo SS, O’Leary HM, Fairney T, Chen NK, Panych LP, Park H, Jolesz FA. Increasing cortical activity in auditory areas through neurofeedback functional magnetic resonance imaging. *Neuroreport* 2006;17(12):1273–1278. [PubMed: 16951568]
- Yoo SS, O’Leary HM, Lee JH, Chen NK, Panych LP, Jolesz FA. Reproducibility of trial-based functional MRI on motor imagery. *Int J Neurosci* 2007;117:1–13. [PubMed: 17365096]
- Zarahn E, Aguirre G, D’Esposito M. A trial-based experimental design for fMRI. *NeuroImage* 1997;6:122–138. [PubMed: 9299386]
- Zou KH, Greve DN, Wang M, Pieper SD, Warfield SK, White NS, Manandhar S, Brown GG, Vangel MG, Kikinis R, Wells WM. For the FIRST BIRN Research Group. Reproducibility of functional MR imaging: preliminary results of prospective multi-institutional study performed by Biomedical Informatics Research Network. *Radiology* 2005;237(3):781–789. [PubMed: 16304101]

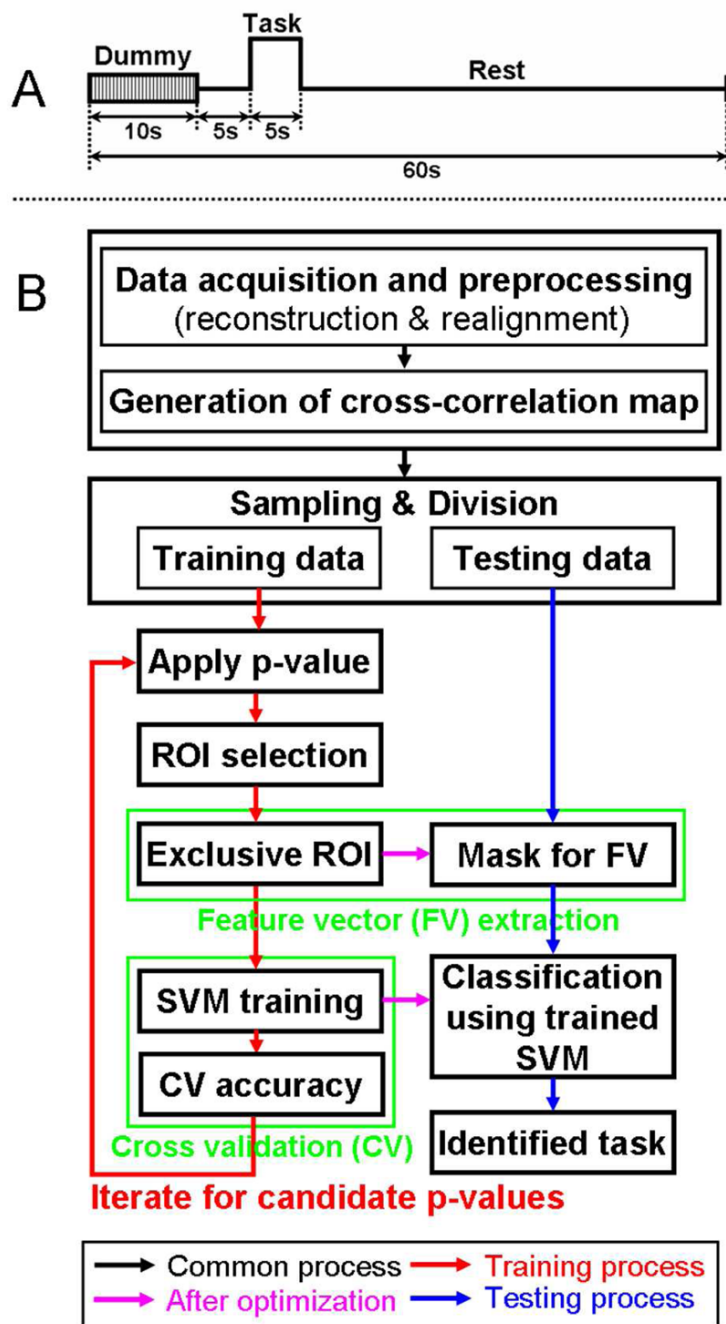


Figure 1. (A) A timing diagram of the proposed fMRI study based on single trial-based paradigm and (B) a flow diagram of data processing for automated classification. First, reconstruction, realignment, and generation of cross-correlation maps were commonly applied (black arrow) for all fMRI data acquisitions (42 scans). Then, 30 different combinations of training data (24 scans) and testing data (18 scans) were prepared to examine a range of classification results. Using the 24 scans during the training process (red arrow), the optimal ROIs and SVM parameters were determined from the maximum cross validation accuracy throughout p -values ($p < 0.05$, $p < 0.01$, $p < 5 \times 10^{-3}$, $p < 10^{-3}$, $p < 5 \times 10^{-4}$, and $p < 10^{-4}$). These optimized parameters were applied to the subsequent classification of the testing data (remaining 18 scans; magenta

arrow). During the consequent testing process (blue arrow), the testing data was entered into the trained SVM. Then, a hit or miss was determined by comparing the index of the identified (classified) task and type of original testing data.

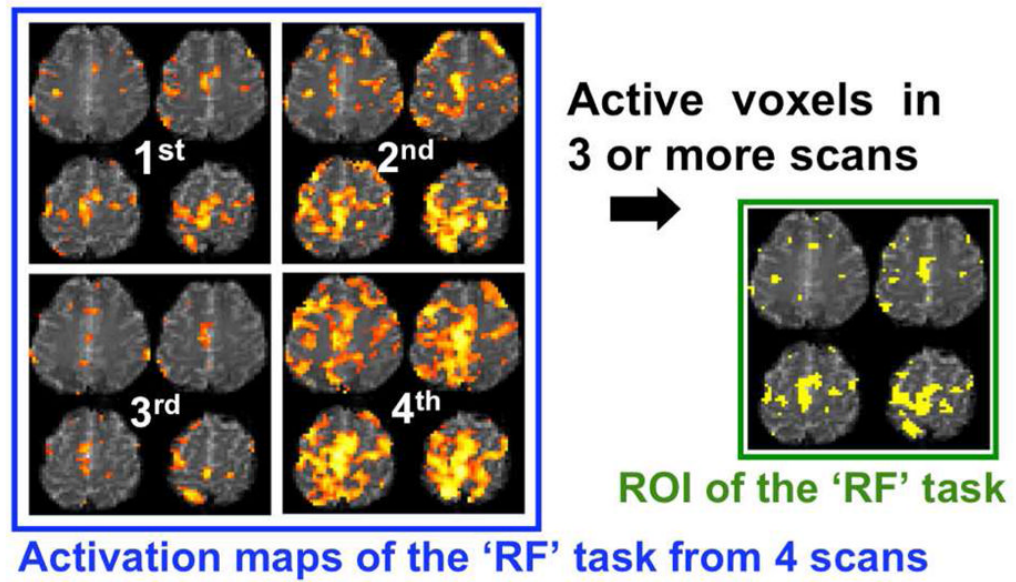


Figure 2.

An example of an ROI definition for a right-foot motor imagery (RF) task obtained from one female subject (S1). From the obtained activation maps (four slices out of 13 slices; $p < 10^{-3}$; $z\text{-score} > 3.29$; $d.f. = 49$) for training data (4 scans), the voxels that were active in 3 or more scans were included within the ROI, representing the activation feature.

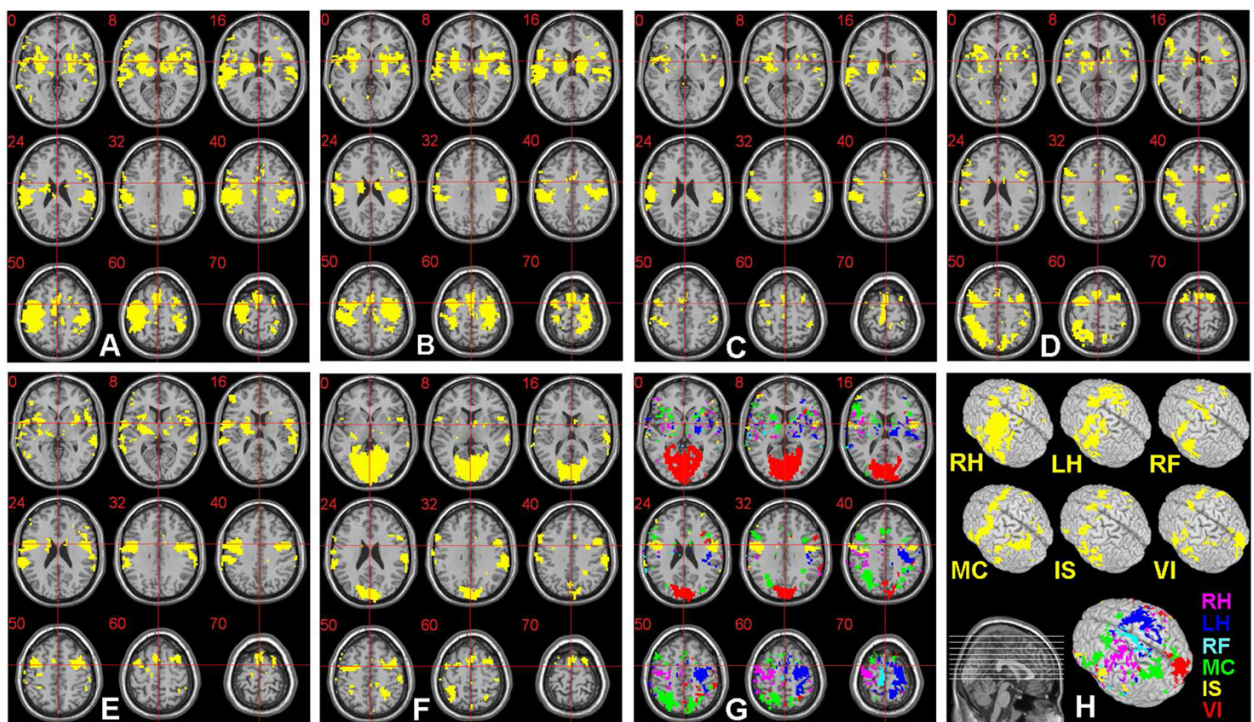


Figure 3.

The ROIs ($p < 0.01$; $z\text{-score} > 2.58$; $d.f. = 49$) obtained from the six imagery tasks performed by a male subject (S4) are shown as yellow-colored regions: (A) ROI for the right-hand motor imagery (RH), (B) left-hand motor imagery (LH), (C) right-foot motor imagery (RF), (D) mental calculation (MC), (E) internal speech generation (IS), and (F) visual imagery (VI). (G) The exclusively activated regions across (A) ~ (F) are displayed in different colors (labels in H) and are overlaid on the anatomical template (provided by MRIcro; www.mricro.com v. 1.39) after the normalization process. Slice range: $Z = 0$ to $+70$ mm (Talairach coordinates). The voxels within these exclusive ROIs were utilized as elements of a feature vector necessary for the SVM classification. (H) The original ROIs and exclusive ROI, rendered on a 3-D volume, are shown.

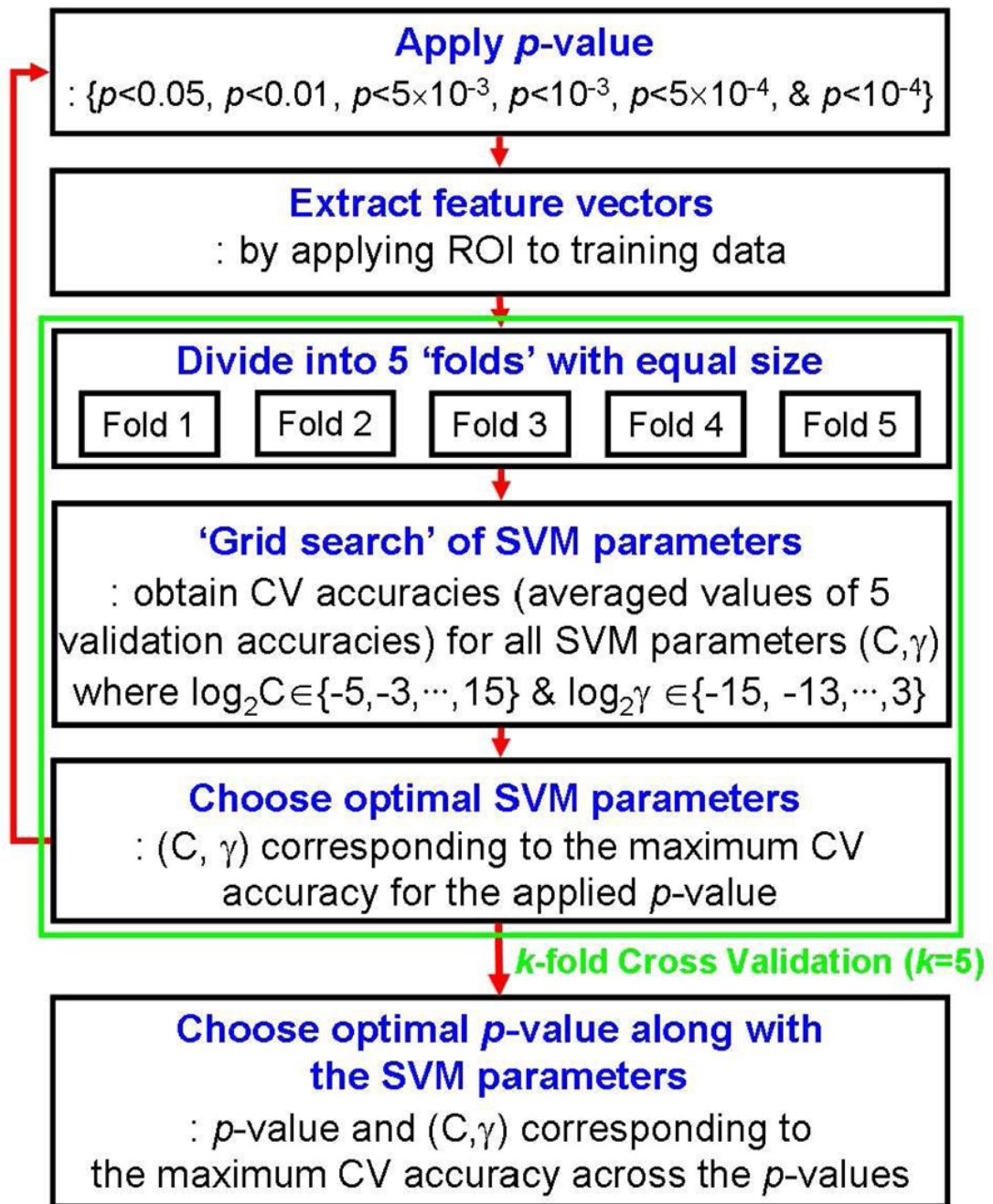


Figure 4. The detailed procedure for the optimization of both the SVM parameters (C & γ) and the p -value by using the k -fold cross validation (CV) scheme, with a 'grid-search', during the training process shown in Fig. 1B.

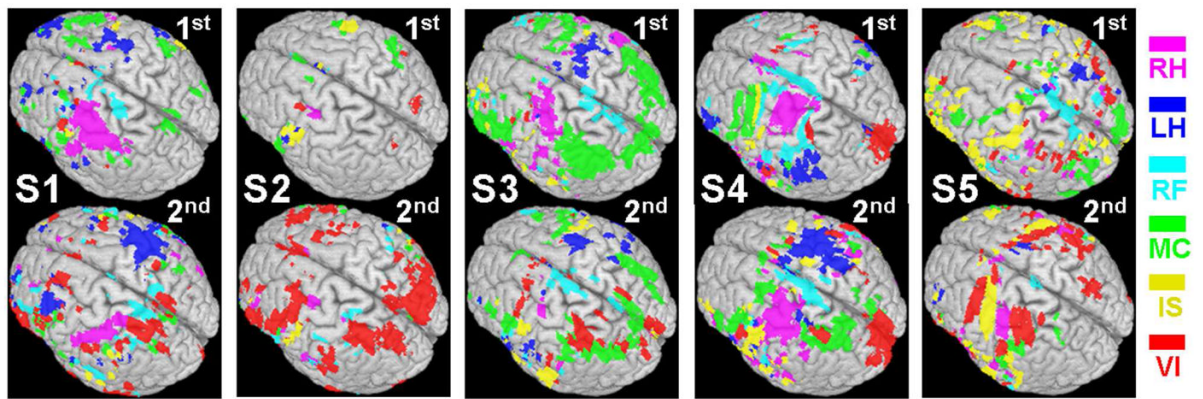


Figure 5. The exclusively task-specific ROIs obtained from the two sessions (1st and 2nd) for five subjects (S1 ~ S5; the p -value was different depending on the outcome of optimization, and listed in Table 1).

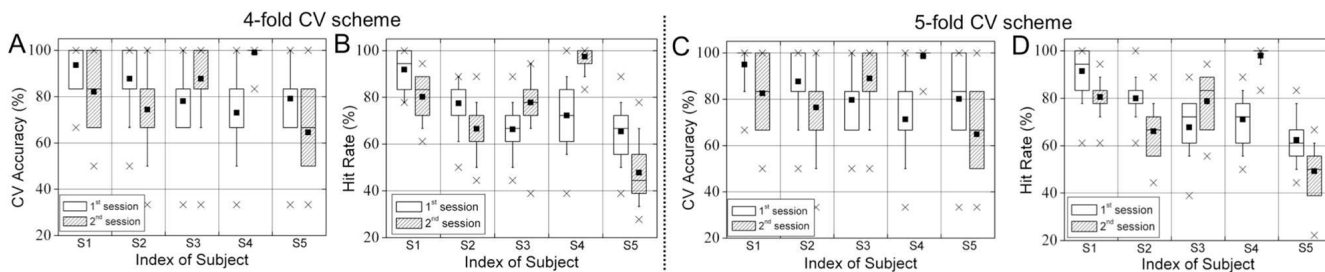


Figure 6. The box and whisker plots of the leave-one-out CV accuracies (A & C) from training data (n=24) and hit rates (B & D) from testing data (n=18) for the 4- and 5-fold CV schemes (a box: 25th, 50th, and 75th percentiles, whisker: 10th and 90th percentiles, black square dot: mean accuracy, and 'x' mark: min- and max-accuracies from 30 randomly selected sets of training and testing data).

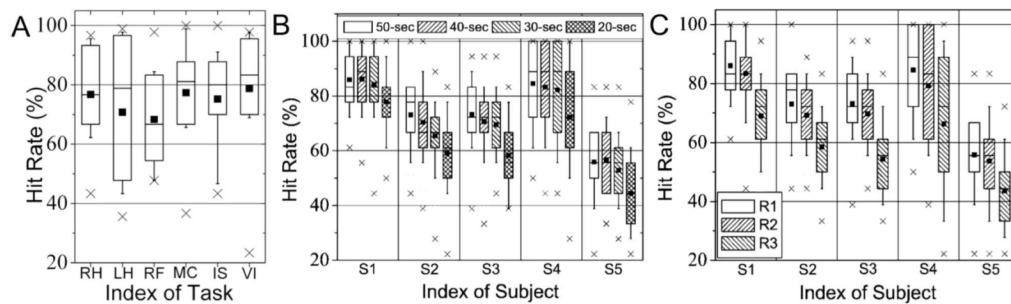


Figure 7.

Box-whisker plots of the range of classification results, as measured by the hit rate for (A) each task pooled over all the subjects and sessions, for (B) the variable acquisition times (50-, 40-, 30-, and 20-sec excluding the dummy volumes), and for (C) the variable spatial resolutions (R1: $64 \times 64 \times 13$, size of voxel = $3.75 \times 3.75 \times 5.5 \text{ mm}^3$, R2: $32 \times 32 \times 13$, size of voxel = $7.5 \times 7.5 \times 5.5 \text{ mm}^3$, and R3: $16 \times 16 \times 6$, size of voxel = $15 \times 15 \times 11 \text{ mm}^3$) based on the simulation (a box: 25th, 50th, and 75th percentiles, whisker: 10th and 90th percentiles, black square dot: mean accuracy, and 'x' mark: min- and max-accuracies).

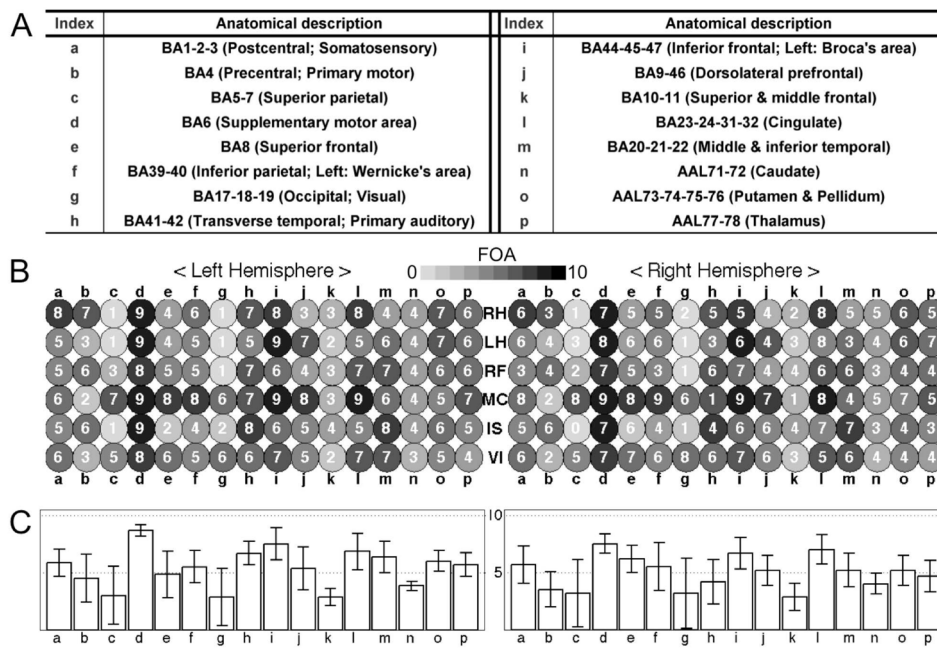


Figure 8. Results from the frequency of activation (FOA) analysis: (A) A table of brain regions that are grouped into 16 sub-regions ('a' through 'p') according to their functional relevance, whereby the Brodmann's Area (BA) and Automated Anatomical Labeling (AAL) indices are utilized as segmented anatomical templates. (B) An illustration of individual FOA counts (max=10) from 16 representative cortical areas (after SVM training). The FOA values are presented in gray scale as indicated by the bar, along with a number inside the circle (i.e. 'darker' indicates 'higher' FOA). (C) A bar graph of the average value and standard deviation of the FOA values in (B) across the six tasks for each of the 16 brain regions.

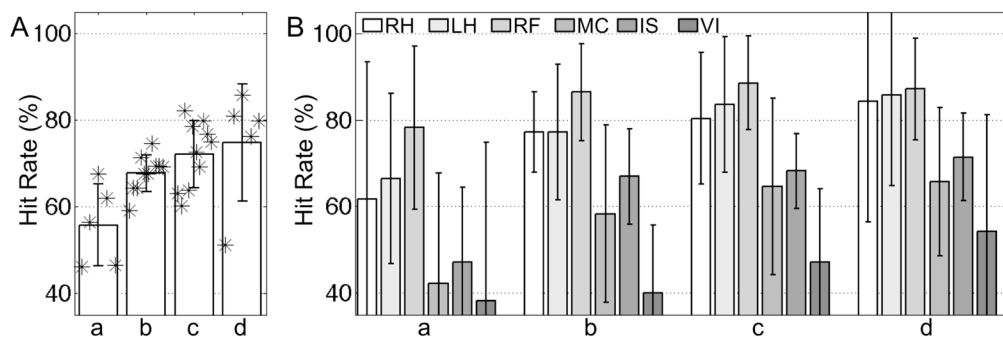


Figure 9. From the between-subject test, the average (bar) and standard deviation (whisker) values of the hit rates obtained (A) across all the tasks, and (B) for each individual task (**a**: one subject’s data for training & the remaining four subjects’ data for testing; **b**: two subjects’ data for training & the remaining three subjects’ data for testing; **c**: three subjects’ data for training & the remaining two subjects’ data for testing; **d**: four subjects’ data for training & the remaining one subject’s data for testing). The ‘*’ symbol indicates an actual hit rate.

Table 1

The maximum hit rate (among 30 sets of training/testing data combinations in each session from five subjects) and the corresponding optimal p -value.

		Hit Rate (%)	p -value
S1	1 st	100	$p < 10^{-4}$
	2 nd	94.4	$p < 10^{-4}$
S2	1 st	100	$p < 10^{-4}$
	2 nd	88.9	$p < 5 \times 10^{-3}$
S3	1 st	88.9	$p < 0.01$
	2 nd	94.4	$p < 5 \times 10^{-3}$
S4	1 st	88.9	$p < 5 \times 10^{-4}$
	2 nd	100	$p < 10^{-3}$
S5	1 st	83.3	$p < 0.05$
	2 nd	66.7	$p < 0.01$

The hit rates from the within-session test (WST) and the cross-session test (CST) for each subject. The hit rates from the WST were shown as a baseline performance of the normalized fMRI data.

Table 2

	S1	S2	S3	S4	S5
WST					
1 st session	95.4 ± 3.6	74.6 ± 11.5	68.1 ± 10.0	84.4 ± 9.9	71.9 ± 14.3
2 nd session	88.5 ± 8.9	67.4 ± 8.2	85.0 ± 11.5	97.6 ± 8.2	43.0 ± 11.9
CST					
1 st sess. (Train) – 2 nd sess. (Test)	85.7	73.8	71.4	97.6	35.7
2 nd sess. (Train) – 1 st sess. (Test)	76.2	64.3	71.4	92.9	40.5

* For WST, the averaged hit rates and standard deviation from the randomly selected 30 training and testing data sets are shown for each subject.

Table 3

Average computational time (h: hour; m: minute; s: second) for each analytical step (Intel Pentium D CPU 3GHz; 3.5GB RAM; Linux Slackware v12.1; MATLAB v7.1).

Type of test	Activation maps (per map)	Training phase (CV phase)	Testing phase
Within-Session	8m 22s (11.9s)	40s (21s)	1.0s
Cross-Session	1h 33m 48s (1m 7s)	2m 30s (1m 59s)	2.9s
		(1 subject's data for training) 9m 4s (8m 35s)	3.7s
Between-Subject	7h 49m 1s (1m 7s)	(2 subjects for training) 41m 2s (40m 4s)	3.2s
		(3 for training) 1h 33m 53s (1h 32m 26s)	2.6s
		(4 for training) 2h 45m 25s (2h 43m 18s)	1.5s

* Note that the number of activation maps of the within-session, cross-session, and between-subject tests are 42 (24 for training; 18 for testing), 84 (42 for training; 42 for testing), and 420, respectively. In the cross-session and between-subject tests, the fMRI data sets in the original space were normalized and smoothed after realignment. Subsequently, a GLM-based analysis was applied to find activation patterns (all functions were adopted from SPM2). A 5-fold cross-validation (CV) was used.