Published in final edited form as:

Int J Imaging Syst Technol. 2014 December ; 24(4): 339–349. doi:10.1002/ima.22111.

A Study of Long-Term fMRI Reproducibility Using Data-Driven Analysis Methods

Xiaomu Song¹, Lawrence P. Panych², Ying-Hui Chou³, and Nan-Kuei Chen³

¹Department of Electrical Engineering, School of Engineering, Widener University, Chester, PA 19013

²Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115

³Brain Imaging and Analysis Center, Duke University Medical Center, Durham, NC 27710

Abstract

The reproducibility of functional magnetic resonance imaging (fMRI) is important for fMRI-based neuroscience research and clinical applications. Previous studies show considerable variation in amplitude and spatial extent of fMRI activation across repeated sessions on individual subjects even using identical experimental paradigms and imaging conditions. Most existing fMRI reproducibility studies were typically limited by time duration and data analysis techniques. Particularly, the assessment of reproducibility is complicated by a fact that fMRI results may depend on data analysis techniques used in reproducibility studies. In this work, the long-term fMRI reproducibility was investigated with a focus on the data analysis methods. Two spatial smoothing techniques, including a wavelet-domain Bayesian method and the Gaussian smoothing, were evaluated in terms of their effects on the long-term reproducibility. A multivariate support vector machine (SVM)-based method was used to identify active voxels, and compared to a widely used general linear model (GLM)-based method at the group level. The reproducibility study was performed using multisession fMRI data acquired from eight healthy adults over 1.5 years' period of time. Three regions-of-interest (ROI) related to a motor task were defined based upon which the long-term reproducibility were examined. Experimental results indicate that different spatial smoothing techniques may lead to different reproducibility measures, and the wavelet-based spatial smoothing and SVM-based activation detection is a good combination for reproducibility studies. On the basis of the ROIs and multiple numerical criteria, we observed a moderate to substantial within-subject long-term reproducibility. A reasonable long-term reproducibility was also observed from the inter-subject study. It was found that the short-term reproducibility is usually higher than the long-term reproducibility. Furthermore, the results indicate that brain regions with high contrast-to-noise ratio do not necessarily exhibit high reproducibility. These findings may provide supportive information for optimal design/implementation of fMRI studies and data interpretation.

Keywords

reproducibility; wavelet; support vector machine; quantitative fMRI

I. INTRODUCTION

Blood oxygenation level dependent (BOLD) contrast functional magnetic resonance imaging (fMRI), which localizes task-related or resting-state brain activation by detecting changes in paramagnetic deoxyhemoglobin, plays an important role in advancing the understanding of brain function and disorders. One of fundamental assumptions of BOLD fMRI is that the identified brain activities associated with a functional state or neuro-pathologic condition are repeatedly measurable within a subject or across multiple subjects at different time under the same experimental condition. This is usually referred to as the fMRI reproducibility, which is important for the investigation of brain plasticity, medical diagnosis, disease progression, treatment efficacy, drug induction, and neurosurgical planning (Vlieger et al., 2004; Chen et al., 2007; Fox et al., 2010).

fMRI reproducibility under a task stimulation has usually been examined on the scale of the whole brain, regions-of-interest (ROI), or active regions. A variety of studies have been conducted to examine fMRI reproducibility within a session (Tegeler et al., 1999), across multiple sessions for a single subject (McGonigle et al., 2000), and across multiple subjects and sessions (Rombouts et al., 1997; Specht et al., 2003; Wei et al., 2004; Ming et al., 2012, Kristo et al., 2014). Some studies were carried out during a relatively short period of time with two or more repeated sessions separated by a couple of hours, a few days, or up to 2 months (Rombouts et al., 1997; McGonigle et al., 2000; Specht et al., 2003; Ming et al., 2012, Kristo et al., 2014). A small number of studies were performed to examine the long-term fMRI reproducibility over 1 year, which is comparable to the time duration of clinical trials (Wei et al., 2004; Yoo et al., 2005). Although evidence has been found for fMRI reproducibility under various task paradigms, there are multiple issues that may affect the reproducibility study, such as the time interval of repeated scans, number of sessions, and data analysis methods used to identify brain functional activities. The purpose of this work is to examine how data analysis methods would affect the long-term fMRI reproducibility study with a focus on the spatial smoothing and activation detection methods.

The Gaussian kernel has been widely used for spatial smoothing in fMRI data analysis, including reproducibility studies (Yoo et al., 2005; Raemaekers et al., 2007; Caceres et al., 2009; Plichta et al., 2012). It has been shown that different full widths at half maximum (FWHM) of the Gaussian kernel lead to different levels of reproducibility even when the same set of experimental data are used (Caceres et al., 2009). Although an “optimal” FWHM parameter resulting in a high reproducibility measure could be identified, it is not clear if such a parameter is clinically meaningful because the resultant smoothing may also remove important details in fMRI data.

In previous reproducibility studies, in order to detect active brain regions under a task paradigm, certain statistical approaches, such as hypothesis test or correlation analysis, have often been used. These methods implicitly assume a probability distribution for test statistics

and identify active brain voxels using a fixed significance threshold across all subjects and/or sessions. Note that fMRI data exhibit a nonstationary nature due to long-term neurophysiological effects, such as sensory/learning-induced neuronal plasticity and functional reorganization (Yoo et al., 2005), and short-term factors, such as subjects' task strategy, attention, fatigue, mood, movement, physiological fluctuation, and machine instability. As a result, considerable variation in amplitude and spatial extent of fMRI activations exist across subjects and sessions even using a consistent task paradigm and identical imaging conditions (Liu et al., 2004). Under such circumstances, a fixed significance threshold would generate an ambiguity: it is not clear if the detected change in active brain area is resulting from a change of brain function, or an improper statistical threshold. As indicated by Voyvodic et al. (2009), the spatial extent of an active brain region identified using a statistical significance threshold depends on "the quality and quantity of signal acquired" instead of the real boundaries of brain function. To reduce the ambiguity caused by improper thresholds, different thresholds have been examined in some fMRI studies (Caceres et al., 2009; Voyvodic et al., 2009). When each threshold was tested, however, it was fixed across subjects and sessions, and was not adapted to the inter-session and inter-subject variation.

In this work, we revisited a set of fMRI data collected for the long-term fMRI reproducibility study in a finger-tapping motor task (Yoo et al., 2005), and examined three motor-task related regions-of-interest (ROI) in primary/premotor area, supplementary motor cortex, and cerebellum. The reason of using the finger-tapping task was that it is almost impossible to induce functional plasticity or learning effects that could be elicited by more complicated motor sequence learning tasks (Karni et al., 1998; Yoo et al., 2005). In addition to the conventional Gaussian smoothing, a wavelet domain Bayesian method was evaluated for spatial smoothing in terms of its impact on fMRI reproducibility (Song et al., 2006). In this method, a temporal prior obtained from a statistical t test on each voxel's time course in the wavelet domain is used as an initial significance estimation of wavelet coefficients, and applied to the Gaussian mixture model (GMM)-based Bayesian wavelet shrinkage in the spatial domain. To detect brain activation under the motor task stimulation without relying on a fixed statistical threshold, a data driven multivariate machine learning-based method was used (Song et al., 2009). The method aims to explore a true boundary between active and inactive voxels in a feature space, and does not depend on a threshold for the final decision, reducing the threshold-induced ambiguity. This method was compared to a widely used general linear model (GLM)-based technique included in FSL (Woolrich et al., 2009) in terms of the specificity and the long-term reproducibility at the group level. The long-term reproducibility was examined for individual subjects and across multiple subjects based on the aforementioned three ROIs. The study not only verifies existing findings, but also provides new information that was not revealed in previous analyses.

II. METHODS

A. fMRI Experiments

Eight right-handed healthy normal subjects (seven males, one female, aged from 21 to 51 with an average of 35.1 ± 11.6 years) without any neurological diseases and/or psychiatric issues were recruited as participants in the fMRI study (Yoo et al., 2005). A 1.5 Tesla GE

MR scanner was used to collect data. For each subject, nine scan sessions were implemented, and the inter-session gap ranged from 21 to 140 days, with an average of 56.9 ± 24.6 days. The entire experiment period for each subject ranged from 378 to 536 days, with an average of 454.9 ± 47.2 days. In each scan, 24 axial-plane echo planar imaging (EPI) data were acquired to cover the entire brain with 6 mm slice thickness. EPI parameters included a repetition time (TR) of 2.5 s, an echo time (TE) of 50 ms, and a flip angle of 90° . The image matrix size was 64×64 . During the scan, the subject was performing a right hand finger tapping paced at 1 Hz guided by auditory cues delivered through a MR-compatible headset. The experimental paradigm consisted of four 30-s task blocks and five 30-s off blocks. A total of 114 vol. were acquired in each scan. There was a 15-s dummy scan at the beginning of each scan, corresponding to 6 vol. of acquisitions that were removed before the analysis. Therefore 108 vol. from each scan session were finally used in the analysis.

Because of a computer storage failure, not all acquired data were retrievable. In this work, there are five subjects with all nine sessions of data available, two subjects with eight sessions, and one subject with five sessions. In addition, T1-weighted anatomical images were collected but not retrievable. Data were de-identified before they were provided for the analysis. The experiments on human subjects were compliant with the standards established by the Institutional Review Boards of Brigham and Women's Hospital, Boston, MA. Informed consent was obtained from the subjects before enrollment in the study.

B. Preprocessing

For the study of each individual subject, the data collected from all sessions were registered to a volume from the first session using a 12 degree-of-freedom registration method (Jenkinson et al., 2002). For the group study, the data of each session was first motion corrected using a six degree-of-freedom rigid body registration (Jenkinson et al., 2002), and was spatially normalized to the Montreal Neurological Institute (MNI) template. Then the data were detrended to remove low frequency drift using an in-house Matlab program.

C. Spatial Smoothing

For spatial smoothing, a wavelet domain Bayesian noise removal method was used in this study (Song et al., 2006). This method decomposes fMRI data into the wavelet domain, and implements a multiscale Gaussian mixture model (GMM)-based Bayesian wavelet shrinkage. It can efficiently attenuate spatial noise while preserving signal details without oversmoothing the data. The details of this method are described in the following.

If we assume that a wavelet coefficient y_i , with its true value x_i , contaminated with additive noise z_i : $y_i = x_i + z_i$, is statistically independent of others, then the high peak, heavy-tailed marginal distribution of x_i can be characterized by a two-component GMM:

$$p(x_i/\theta) = \sum_{n=0}^1 \alpha_n g(x_i/\theta_n),$$
 where $\theta = \{\alpha_n, \theta_n\}$, $\theta_n = \{\mu_n, \sigma_n\}$, and $n = \{0, 1\}$. $g(x_i/\theta_n)$ is a Gaussian probability density function (pdf) with mean μ_n and standard deviation σ_n . α_1 and α_0 are the prior probabilities that x_i is significant or not, which is denoted by $s_i = 1$ (significant) or 0 (insignificant), and $\alpha_0 + \alpha_1 = 1$. In the wavelet domain, $\mu_n = 0$, $n = 0, 1$. If a square loss function is used, the posterior mean of x_i is estimated as:

$$E(x_i|y_i)=p(s_i=0|y_i)\frac{\hat{\sigma}_0^2}{\hat{\sigma}_z^2+\hat{\sigma}_0^2}y_i+p(s_i=1|y_i)\frac{\hat{\sigma}_1^2}{\hat{\sigma}_z^2+\hat{\sigma}_1^2}y_i, \quad (1)$$

where $\hat{\sigma}_0^2, \hat{\sigma}_1^2, \hat{\sigma}_z^2$ are the estimations of σ_0^2, σ_1^2 , and noise variance, and:

$$p(s_i=n|y_i)=\frac{\alpha_n g(y_i|\theta_n)}{\sum_{k=0}^1 \alpha_k g(y_k|\theta_k)}. \quad (2)$$

The student's t test is performed on each voxel in the wavelet domain. Let T_1 denote the time interval of a task period, T_0 indicate the resting period, and $y_{i,t}^j$ is the wavelet coefficient of the i^{th} voxel at the j^{th} scale and time t . The null hypothesis is that there is no significant difference between BOLD signals during the resting and task period, and $h_i^j=1$ or 0 indicate if the null hypothesis of the voxel is rejected or not rejected. $p(s_{i,t}^j/y_{i,t}^j)$ can be initialized as:

$$p(s_{i,t}^j=1/y_{i,t}^j)=\begin{cases} \gamma, & \text{if } t \in T_1 \text{ and } h_i^j=1 \\ 1-\gamma, & \text{OTW} \end{cases}, \quad (3)$$

where γ is close to 1, and $p(s_{i,t}^j=0/y_{i,t}^j)=1-p(s_{i,t}^j=1/y_{i,t}^j)$. At the j^{th} scale, α_n at time t is estimated as:

$$\hat{\alpha}_{n,t}^j=\frac{\sum_i p(s_{i,t}^j=n/y_{i,t}^j)}{S_j}, n=0, 1, \quad (4)$$

where S_j is the number of wavelet coefficients at scale j . The variance of each Gaussian pdf is calculated as:

$$\hat{\sigma}_{n,t}^{j^2}=\frac{\sum_i y_{i,t}^{j^2} p(s_{i,t}^j=n/y_{i,t}^j)}{\sum_i p(s_{i,t}^j=n/y_{i,t}^j)}, n=0, 1. \quad (5)$$

The posterior estimation of x_i and $p(s_{i,t}^j/y_{i,t}^j)$ can be obtained using Eqs. (1) and (2). After the wavelet domain analysis, the denoised fMRI data are obtained by performing an inverse wavelet transform.

The vanishing moment of a wavelet plays a similar role as the FWHM of a Gaussian kernel. A higher vanishing moment involves more neighboring voxels to spatial smoothing. It has been shown that different FWHMs may generate different levels of reproducibility when a Gaussian kernel is used for spatial smoothing (Caceres et al., 2009). This also applies to wavelets with different vanishing moments. There is no theoretical guidance on how to choose optimum vanishing moments for wavelet-based spatial smoothing. It should be noted

that, unlike the single scale Gaussian smoothing, the wavelet method is a multiscale analysis where a wavelet with low vanishing moments can provide sufficient noise removal while preserving details around boundaries of active brain regions (Song et al., 2006). In this work, Daubechies wavelets with vanishing moments from 1 to 10 were examined at the group level (Daubechies, 1992). For comparison, the Gaussian smoothing with FWHM values from 2 to 20 mm was also evaluated in terms of their effects on computed reproducibility measures at the group level.

D. Detection of Functional Activation

A multivariate support vector machine (SVM)-based quantitative fMRI method was used to detect brain activation associated with the motor task (Song et al., 2009). Figure 1 schematically illustrates the method. Based on preprocessed fMRI data, multiple features are calculated from each voxel and its neighboring voxels. An offline SVM-based feature selection process is performed to identify a subset of representative features to represent each voxel (Evgeniou et al., 2003). A one-class support vector machine (OCSVM) is used to provide an initial mapping of brain activation. OCSVM is an implementation of the oneclass classifier that aims to estimate an approximation function to separate a majority of the learning data from outliers in a feature space (Schölkopf et al., 2001). Given N m -dimensional feature vector $\mathbf{x}_i \in R^m, i = 1, \dots, N$, OCSVM computes a hyperplane in the feature space by minimizing:

$$\frac{1}{2}w^2 + \frac{1}{vN} \sum_{i=1}^N \xi_i - \rho, \text{ subject to } c_i(\mathbf{x}_i \cdot \mathbf{w}) \geq \rho - \xi_i, \quad (6)$$

where $\mathbf{w}$ is the normal vector defining the hyperplane, ρ/w is the margin size, i.e., the distance between the hyperplane and origin, $\xi_i \geq 0$ is the slack variable corresponding to training errors, c_i is the class label of $\mathbf{x}_i$, and v controls an upper bound of outliers. The OCSVM learning separates a pre-specified fraction $(1-v)$ of data with the maximum distance to the origin. Because the number of active voxels is usually less than that of inactive ones (Song et al., 2009), most of active voxels are included in the outliers identified by OCSVM. The parameter v is task-specific, and due to fMRI non-stationarity, it is also session- and subject-specific. An accurate setting of v is not realistic and OCSVM results may contain a significant number of false detections. A spatial domain prototype selection is performed to identify voxels that are correctly classified by OCSVM.

A prototype consists of a feature vector representing a voxel and its class label (active or inactive in this work). Active voxels are spatially grouped together at multiple anatomic sites, and mis-detections have a higher chance to be spatially isolated. In the prototype selection, the spatial relation between each voxel and its neighboring voxels is described by a Gabriel Graph (GG) (Jaromczyk et al., 1992), and the 1st-order graph neighborhood editing technique of GG with voting strategy is used to remove any voxel if its class label is not dominant in its 3×3 neighborhood (Dasarathy et al., 2000). The remaining voxels are used to train a two-class SVM (TCSVM) to reclassify all voxels. TCSVM is a supervised learning and classification tool that aims to estimate a linear classification hyperplane in a feature space so that two classes can be maximally separated (Vapnik, 1998). The TCSVM

learning usually allows training errors where a parameter C is used to control a tradeoff between the hyperplane complexity and the training error. Active and inactive voxels may overlap in the feature space constructed by the selected feature subset, especially when the BOLD increase due to the task stimulation is small. In such cases, the kernel technique can be used to implement nonlinear OCSVM and TCSVM to improve the classification performance, and the radial basis function (RBF) kernel was used in this work. It is possible that part of active voxels cannot be detected by OCSVM (false negative), or detected but removed by the prototype selection. It is also possible that a few inactive voxels are detected and selected as the training prototypes for the active ones (false positives). To re-detect false negatives and/or attenuate the effects from the remaining false positives in the training data, a high generalization performance for the TCSVM learning was considered by using a large RBF kernel width and a small C value. TCSVM results are transferred into probability estimates (Wu et al., 2004). The final decision on a voxel is made by comparing the voxel's probabilities of active and inactive. From a statistical point-of-view, this is a "soft" decision that is more reliable than a "hard" threshold that might not fit the true boundary between active and inactive voxels in the feature space. For group level activation detection, each voxel's probabilities of active and inactive are averaged across all sessions and subjects. A final decision is made by comparing the averaged probabilities.

In the SVM-based method, OCSVM is used as an unsupervised learning tool to adapt to changes of data distribution in the feature space. The prototype selection and TCSVM reclassification transfer the unsupervised learning into a self-supervised learning process, and can be repeated multiple times until there is no more change in final results. The RBF kernel width and TCSVM parameter C are usually determined experimentally or by cross validation. Once determined, they are fixed for the analysis of all datasets. It was found in our study that a moderate change of these parameters does not significantly affect the final results because the repetition of prototype selection and TCSVM reclassification may attenuate the negative impacts from an improper choice of processing parameters. For comparison, a general linear model (GLM)-based method with mixed effects was implemented using FSL to detect group level activations in the ROIs (Woolrich et al., 2009).

E. Reproducibility Study

Three ROIs related to the motor task were examined in this study, including ipsilateral cerebellar areas in dentate (ROI I, center MNI coordinates: $-18, -61, -26$), primary/pre-motor areas (ROI II, center MNI coordinates: $-40, -15, 56$), and supplementary motor cortex (ROI III, center MNI coordinates: $-4, 7, 53$). These ROIs were manually segmented following a method described in a previous work (Yoo et al., 2005). The reproducibility was examined for individual subjects and across subjects.

E.1. Within-subject Study—For each individual subject, multisession fMRI data were used to evaluate the reproducibility. After the activation detection in each session, the reproducibility of size and location of active regions was computed against the first session. The reproducibility of location is defined as (Rombouts et al., 1997; Yoo et al., 2005):

$$R_{overlap}^{i,j} = \frac{2V_{overlap}^{i,j}}{V_i + V_j}, \quad (7)$$

where V_i and V_j are the number of active voxels in session i and session j , and $V_{overlap}^{i,j}$ is the number of active voxels detected in both sessions. The reproducibility of size is calculated as (Rombouts et al., 1997; Yoo et al., 2005):

$$R_{size}^{i,j} = \frac{2V_{smaller}^{i,j}}{V_i + V_j}, \quad (8)$$

where $V_{smaller}^{i,j}$ is the number of voxels in the smaller active area in sessions i and j . $R_{overlap}^{i,j}$ and $R_{size}^{i,j}$ range from 0 (no overlap) to 1 (perfect overlap).

The intraclass correlation coefficient (ICC) has been widely used to quantify fMRI reproducibility (Shrout et al., 1979; Aron et al., 2006; Raemaekers et al., 2007; Caceres et al., 2009; Plichta et al., 2012). In this work, ICC was used to evaluate the inter-session reproducibility of brain activation under the motor task for each subject. There are different ICC measures, and ICC(3,1) defined in (9) was used (Shrout et al., 1979):

$$ICC(3,1) = \frac{BMS - EMS}{BMS + (k - 1)EMS}, \quad (9)$$

where BMS is the between subjects mean square, EMS is the error mean square, and k is the number of the repeated sessions. When ICC(3,1) was used to quantify the inter-session reproducibility, BMS was calculated across all voxels in each ROI using a contrast-to-noise ratio (CNR) statistic defined in the following equation (Menon et al., 1997):

$$CNR = \frac{\Delta S}{0.5(SE_{base} + SE_{task})}, \quad (10)$$

where ΔS is the change of signal intensity under the task stimulation, SE is the standard error of intensity in baseline/task period, and is defined as: $SE = SD/M^{0.5}$, with SD the standard deviation of the signal intensity during the baseline/task period, and M the number of images during the baseline/task period. Kendall's coefficient of concordance (KCC) is a measure of regional homogeneity (Zang et al., 2004), and was computed among active voxels in each ROI across all sessions for each subject.

E.2. Inter-subject Study—Average CNR maps were computed for slices covering the ROIs using all sessions of eight subjects. The brain activation under the motor task stimulation was detected at the group level in the ROIs. The effects of wavelet-based spatial smoothing on ICC were evaluated across subjects using Daubechies wavelets with vanishing moments from 1 to 10 (D1–D10). For comparison, the effects of Gaussian smoothing with FWHM ranging from 2 to 20 mm were evaluated. Active brain regions in the ROIs detected by the SVM- and GLM-based activation detection methods were compared at the group level in terms of their matching to the CNR distribution. The detection frequency of the

SVM-based method was also computed for voxels in the ROIs and compared with the CNR maps. The long-term reproducibility across subjects was examined using two different data analysis settings. One consists of the wavelet-based spatial smoothing and the SVM-based activation detection. The other uses the Gaussian smoothing and GLM-based detection. Under each data analysis setting, ICC was calculated using maximum CNR, mean CNR, and KCC of detected active voxels in each ROI over different number of sessions. Voxel level ICC maps in individual slices that cover the ROIs were generated and compared with the CNR maps.

III. RESULTS

A. Within-subject Study

Brain activation was detected using the SVM-based method, and $R_{overlap}$, R_{size} , and KCC values were computed. Figure 2 shows $R_{overlap}$ computed from active regions in the ROIs using all sessions of eight subjects, which are represented by A–H*. Above each subject's ID, there are three columns of $R_{overlap}$ values indicated by squares (ROI I), diamonds (ROI II), and triangles (ROI III). The largest variation of $R_{overlap}$ in ROI I was observed from subject H, which ranges from 0.29 to 0.88. $R_{overlap}$ in ROI III varies from 0.3 to 0.97 on subject C. Variation of $R_{overlap}$ in the three ROIs is similar in subject A. Figures 3 and 4 show R_{size} and KCC of the active regions in the ROIs. The largest variation of R_{size} was observed in ROI III from subject C, ranging from 0.31 to 0.96. Subjects A, D, F, and H show relatively smaller variation of R_{size} in the ROIs. The largest variation of KCC was from ROI I of subject E, ranging from 0.23 to 0.94. The ICC values of the ROIs across all sessions of individual subjects are listed in Table I with their 95% confidence intervals.

B. Inter-subject Study

Figures 5–7 illustrate the ICC(3,1) values computed using the maximum CNR (Fig. 5), average CNR (Fig. 6), and KCC (Fig. 7) of active voxels in the ROIs as functions of (a) the vanishing moments of Daubechies wavelets, and (b) the FWHM of the Gaussian kernel when the wavelet and Gaussian kernel-based methods were used for spatial smoothing. Active voxels were detected at the group level using the SVM-based method described in section II.D. It was observed from Figures 5a, 6a, and 7a that when the wavelet-based method was used, D1 (the lowest vanishing moment) resulted in the largest ICC in six out of nine sets of ICC values, and D2 or D3 led to the largest ICC in two of nine sets of ICC values. When the Gaussian smoothing was used, ICC values computed using maximum and mean CNR increase for ROIs II and III when FWHM = 6 mm, as shown in Figures 5b and 6b, but a decrease of ICC was also found for ROI I. When ICC was calculated based on KCC values, an increase of FWHM led to a decrease of ICC in ROIs I and III, and a fluctuation of ICC was observed for ROI II when FWHM increased from 2 to 20 mm.

Tables II–IV list several statistics calculated from the ICC values illustrated in Figures 5–7, including the average, standard deviation (Std), and maximum difference (Max Diff: the absolute difference between the maximum and minimum) of each of the two groups of ICC values, where one group was obtained after the wavelet-based spatial smoothing with different vanishing moments, and the other was computed after the Gaussian smoothing with

different FWHM values. A t test was performed to evaluate the difference between the two groups of ICC values, and their variances were also compared using an F test. The P values of t test and F test are listed in the tables.

It was observed from Table II that for ROI I, the average ICC values calculated using maximum, average CNR, and KCC after the wavelet-based spatial smoothing are greater than those computed after using the Gaussian kernel-based smoothing, with significance levels below 1.5%. The variation of ICC with respect to different wavelet vanishing moments is less than that with respect to different FWHM, and the F test results show that the difference is significant at levels below 4.6% when average CNR and KCC were used to compute ICC, and is not significant when Maximum CNR was used to calculate ICC. These observations are also reflected via Std and Max Diff of ICC values listed in this table. For ROI II as shown in Table III, the average ICC values computed using maximum CNR and KCC after the wavelet-based smoothing are greater than those after the Gaussian smoothing. The difference is not significant when the maximum CNR was used to compute ICC, and is significant when KCC was used. The average ICC computed using mean CNR after the wavelet-based smoothing is less than that calculated after the Gaussian smoothing at a significance level of 0.014%. The variation of ICC with respect to different wavelet vanishing moments and Gaussian kernel FWHM are not significantly different. Table IV shows the comparison in ROI III. The average ICC values computed using maximum and mean CNR after the wavelet-based smoothing are less than those after the Gaussian smoothing at significance levels below 0.0014%, and the average ICC computed using KCC after the wavelet-based smoothing is greater than that obtained after the Gaussian smoothing at a significance level of 0.0011%. The variation of ICC with respect to different wavelet vanishing moments is less than that with respect to different FWHM, and the F -test results show that the difference is significant at levels below 2.6% when average CNR and KCC were used to compute ICC, and is not significant when Maximum CNR was used to calculate ICC.

Figure 8a shows average CNR maps overlaid on the three slices covering the three ROIs. The slices from left to right contain ROI I, ROI III, and ROI II respectively. Regions encircled by the dark lines are group level activations detected by the SVM-based method, and regions encircled by the grey lines are group level activations identified by the GLM-based method. Multiple thresholds were examined for the group level activation detection using the GLM-based method. To minimize the bias in the comparison study, a threshold of $Z > 2.5$ and a significance level of 2% was finally chosen because the correspondingly detected activations generated the highest ICC values. Figure 8b shows the SVM-based detection frequencies of active voxels in ROI I, ROI III, and ROI II normalized between 0 and 1. Only voxels with a normalized frequency > 0.4 are shown here. The number of active voxels with a detection frequency > 0.5 is 69 in ROI I, 108 in ROI II, and 45 in ROI III.

Table V lists the ICC values computed using maximum CNR, mean CNR, and KCC of active voxels in the ROIs when the wavelet-based spatial smoothing was performed using the Daubechies wavelet of vanishing moment one (D1), and the SVM-based method was used to detect brain activation. Two sets of ICC values were obtained. The first set was computed using the first four sessions of the subjects, and the second was obtained using the

eight sessions. Table VI lists same sets of ICC values computed from the group activations identified by the GLM-based detection method after a Gaussian smoothing with FWHM = 6 mm. The group level activation detection was made at the same threshold that was identified to generate activation maps shown in Figure 8a.

Figure 9a shows the ICC maps overlaid on the same three slices as shown in Figures 8a and 8b obtained using the first two adjacent sessions of subjects A–E, G and H with an average interval about 57 days. Figure 9b shows ICC maps computed using the first and eighth sessions of the subjects with an average interval of 403 days. A significant difference can be observed between Figures 9a and 9b. Regions showing high reproducibility in 9(a) exhibit decreased reproducibility in 9(b), and some regions with low reproducibility in 9(a) show increased reproducibility in 9(b).

IV. DISCUSSION

$R_{overlap}$, R_{size} , and KCC shown in Figures 2–4 are considerably different between subjects and sessions. Compared to $R_{overlap}$, the variation of R_{size} and KCC is smaller across subjects and sessions. It was also observed that higher levels of $R_{overlap}$ and R_{size} were achieved than those obtained previously (Yoo et al., 2005). A possible reason is that the SVM-based method can better adapt to the inter-session and inter-subject variation, and provide a more reliable detection of active voxels than t test (Song et al., 2009). Most ICC values listed in Table I indicate a moderate inter-session reproducibility ($0.4 < ICC < 0.59$) for individual subjects, and some of them show a substantial reproducibility with ICC values above 0.6 (Landis et al., 1977).

From ICC values computed after the wavelet-based smoothing as shown in Figures 5a, 6a, and 7a, it was found that with a low vanishing moment, the wavelet-based method can lead to higher or equivalent ICC values than those obtained using greater vanishing moments. It is a favorable finding that the Daubechies wavelet with a low vanishing moment can result in a higher reproducibility than others because such a wavelet can preserve more details of the spatial structure during the smoothing, and is computationally more efficient.

It was observed from Tables (II–IV) that in most cases there is a significant difference between the ICC averaged over wavelet vanishing moments and that averaged over Gaussian kernel FWHM except for the one calculated using maximum CNR in ROI II. This indicates that the spatial smoothing does affect the computation of the long-term fMRI reproducibility measures, and different methods may generate different levels of reproducibility. Among nine pairs of mean, Std, and Max Diff of ICC values computed after the wavelet- and Gaussian kernel-based spatial smoothing, as listed in Tables (II–IV), six pairs show that the mean ICC calculated after the wavelet-based smoothing is higher than that obtained after the Gaussian smoothing, and seven pairs show that the variation of ICC with respect to wavelet vanishing moments is less than that with respect to Gaussian kernel FWHM. An ideal spatial smoothing should provide sufficient noise attenuation while preserving details in image slices. Compared to Gaussian smoothing, it was shown that wavelet-based techniques better preserve the sharpness of image slices and retain the original shapes of active regions (Wink et al., 2004). It has also been known that even with a

low vanishing moment, the multiscale wavelet analysis may provide a sufficient smoothing without losing image details (Crouse et al., 1998). Therefore, the wavelet-based method was used for spatial smoothing in the long-term reproducibility study of this work.

It was found from Figure 8a that the active regions detected by the SVM-based method better match the high CNR regions than those detected by the GLM-based method. The threshold of the later method could be increased so that the detected activations are closer to the high CNR regions. However, it was found that a slight increase of the threshold led to a significant under-detection of active voxels in ROI III. These findings indicate that the SVM-based method provides a more reliable detection of active voxels than the GLM-based method. The detection frequency maps obtained from the SVM-based method shown in Figure 8b also indicate good matches with high CNR regions in the ROIs. Thus the SVM-based method was chosen to detect active voxels in the ROIs for the long-term reproducibility study in this work.

We noted from Table V that ICC values of ROIs I and II are higher than those of ROI III. A possible reason is that there are more active voxels in ROIs I and II involved to the computation of ICC than that in ROI III, and effects from noise and other artifacts could be relatively more smoothed out. It was also noted from Tables V and VI that ICC values computed using maximum and mean CNR of the first four sessions are usually greater than those calculated using all eight sessions. If we compare these ICC values to those from short-term studies (Ming et al., 2012, Kristo et al., 2014), similar observations can be made. This is reasonable considering the long-term and short-term factors described in the introduction. But a few ICC values computed using KCC indicate increased reproducibility when all eight sessions' data were used. This is possible if the last four sessions have a more consistent regional homogeneity than the first four sessions. In addition, the ICC values computed using KCC are usually higher than those using maximum and mean CNR. This indicates that regional homogeneity is more consistent across sessions than maximum and mean CNR. The ICC values calculated after performing the wavelet-based spatial smoothing and the SVM-based detection (Table V) are greater than those computed after the Gaussian smoothing and the GLM-based detection (Table VI) in all three ROIs except for those computed from maximum and average CNR in ROI III. Because the GLM-based detections do not well match the high CNR regions in ROIs I and II, as shown in Figure 8a, the variation of maximum CNR, mean CNR and KCC in these regions would be more significant across subjects than those from the active regions detected by the SVM-based method.

The ICC maps in Figures 9a and 9b demonstrated a substantial difference of reproducibility when different pairs of sessions with different time intervals were used. Compared to the regions showing decreased reproducibility from 9(a) to 9(b), those showing increased reproducibility are not consistent when different pairs of sessions (e.g., sessions 2 and 7, sessions 1 and 4, etc.) were evaluated. The reason of this difference might be session specific, and caused by various factors, such as imaging artifacts, physiological fluctuations and residual subject movement artifacts. If we compare Figures 8 and 9, it was found that regions with high CNR do not necessarily have high ICC. Similar observations were made in other short-term reproducibility studies (Aron et al., 2006; Caceres et al., 2009).

V. CONCLUSION

In this work, we have studied the long-term fMRI reproducibility using data-driven analysis methods. The effects of spatial smoothing on the reproducibility were evaluated based on a comparison study between a wavelet-domain Bayesian method and the widely used Gaussian smoothing. Experimental results indicate that the wavelet-based smoothing results in less variation of ICC when different wavelet vanishing moments are used, as compared to the variation of ICC when different FWHM values are used for the Gaussian smoothing. In addition, the wavelets with low vanishing moments have a higher chance to provide greater ICC values while being able to preserve image sharpness. Thus, the wavelet method with a low vanishing moment would be considered as an alternative option for spatial smoothing in long-term fMRI reproducibility study.

A multivariate SVM-based quantitative fMRI method was used to identify active brain voxels. This method can explore a true boundary between active and inactive voxels in a feature space, and adapt to intra- and inter-subject variation. Its performance was evaluated and validated at the group level based on a comparison study with a widely used GLM-based method. Experimental results show that the SVM-based method can provide better mapping of truly active brain regions under the motor task stimulation.

For the long-term fMRI reproducibility study, multiple numerical criteria were used to quantify the reproducibility in three predefined ROIs for individual subjects and across multiple subjects. The results of individual subjects indicate a moderate to substantial inter-session reproducibility. The inter-subject study shows that a high level of reproducibility could be achieved during a relatively short period of time. With increased time intervals, the reproducibility will decrease but a reasonable level of reproducibility can still be observed. Similar to previous short-term reproducibility studies, it was observed that regions exhibiting high long-term reproducibility are not necessarily those showing high CNR. In addition, regions showing high or low short-term reproducibility do not necessarily overlap with regions showing high or low long-term reproducibility.

Acknowledgments

Grant sponsor: This research was partially funded by NIH grants R01-NS037922 (LPP) and R01-NS074045 (NKC)

REFERENCES

- Aron AR, Gluck MA, Poldrack RA. Long-term test-retest reliability of functional MRI in a classification learning task. *NeuroImage*. 2006; 29:1000–1006. [PubMed: 16139527]
- Caceres A, Hall DL, Zelaya FO, Williams S, Mehta MA. Measuring fMRI reliability with the intra-class correlation coefficient. *NeuroImage*. 2009; 45:758–768. [PubMed: 19166942]
- Chen CH, Ridler K, Suckling J, Williams S, Fu CH, Merlo-Pich E, Bullmore E. Brain imaging correlates of depressive symptom severity and predictors of symptom improvement after antidepressant treatment. *Biol Psychiatry*. 2007; 62:407–414. [PubMed: 17217921]
- Crouse M, Nowak R, Baraniuk R. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans Signal Process*. 1998; 46:886–902.
- Dasarathy B, Sanchez J, Townsend S. Nearest neighbour editing and condensing tools-synergy exploitation. *Pattern Anal Appl*. 2000; 3:19–30.

- Daubechies, I. Ten lectures on wavelets; CBMS-NSF Regional Conference Series in Applied Mathematics; 1992. p. 1-377.
- Evgeniou T, Pontil M, Papageorgiou C, Poggio T. Image representations and feature selection for multimedia database search. *IEEE Trans Knowledge Data Eng.* 2003; 15:911–920.
- Fox M, Greicius M. Clinical applications of resting state functional connectivity. *Front Syst Neurosci.* 2010; 4:1–13. [PubMed: 20204156]
- Jaromczyk J, Toussaint G. Relative neighborhood graphs and their relatives. *Proc IEEE.* 1992; 80:1502–1517.
- Jenkinson M, Bannister P, Brady J, Smith S. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. *NeuroImage.* 2002; 17:825–841. [PubMed: 12377157]
- Karni A, Meyer G, Rey-Hipolito C, Jezard P, Adams MM, Turner R, Ungerleider LG. The acquisition of skilled motor performance: fast and slow experience driven changes in primary motor cortex. *Proc Natl Acad Sci USA.* 1998; 95:861–868. [PubMed: 9448252]
- Kristo G, Rutten GJ, Raemaekers M, Gelder B, Rombouts S, Ramsey NF. Task and task-free fMRI reproducibility comparison for motor network identification. *Hum Brain Mapp.* 2014; 35:340–352. [PubMed: 22987751]
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–174. [PubMed: 843571]
- Liu J, Zhang L, Brown R, Yue G. Reproducibility of fMRI at 1.5 T in a strictly controlled motor task. *Magn Reson Med.* 2004; 52:751–760. [PubMed: 15389959]
- McGonigle D, Howseman A, Athwal B, Friston K, Frackowiak R, Holmes A. Variability in fMRI: an examination of intersession differences. *NeuroImage.* 2000; 11:708–734. [PubMed: 10860798]
- Menon RS, Thomas CG, Gati JS. Investigation of BOLD contrast in fMRI using multi-shot EPI. *NMR Biomed.* 1997; 10:179–182. [PubMed: 9430345]
- Ming J, Thulborn KR, Szlyk JP. Reproducibility of activation maps for longitudinal studies of visual function by functional magnetic resonance imaging. *Investig Ophthalmol Vis Sci.* 2012; 53:6153–6163. [PubMed: 22879425]
- Plichta MM, Schwarz AJ, Grimm O, Morgen K, Mier D, Haddad L, Gerdes AB, Sauer C, Tost H, Esslinger C, Colman P, Wilson F, Kirsch P, Meyer-Lindenberg A. Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. *NeuroImage.* 2012; 60:1746–1758. [PubMed: 22330316]
- Raemaekers M, Vink M, Zandbelt B, van Wezel R, Kahn RS, Ramsey NF. Test-retest reliability of fMRI activation during prosaccades and antisaccades. *NeuroImage.* 2007; 36:532–542. [PubMed: 17499525]
- Rombouts S, Barkhof F, Hoogenraad F, Sprenger M, Valk J, Scheltens P. Test-retest analysis with functional MR of the activated area in the human visual cortex. *AJNR Am J Neuroradiol.* 1997; 18:1317–1322. [PubMed: 9282862]
- Schölkopf B, Platt J, Shawe-Taylor J, Smola A, Williamson R. Estimating the support of high-dimensional distribution. *Neural Comput.* 2001; 13:1443–1471. [PubMed: 11440593]
- Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979; 86:420–428. [PubMed: 18839484]
- Song, X.; Murphy, M.; Wyrwicz, A. Spatiotemporal denoising and clustering of fMRI data; *Proc IEEE Int Conf Image Process*; 2006. p. 2857-2860.
- Song X, Wyrwicz A. Unsupervised spatiotemporal fMRI data analysis using support vector machines. *NeuroImage.* 2009; 47:204–212. [PubMed: 19344772]
- Specht K, Willmes K, Shah N, Jäncke L. Assessment of reliability in functional imaging studies. *J Magn Reson Imaging.* 2003; 17:463–471. [PubMed: 12655586]
- Tegeler C, Strother S, Anderson R, Kim SG. Reproducibility of BOLD-based functional MRI obtained at 4T. *Hum Brain Mapp.* 1999; 7:267–283. [PubMed: 10408770]
- Vapnik, V. Statistical learning theory. New York/Chichester/Weinheim/Brisbane/Singapore/Toronto: Wiley-Interscience; 1998.

- Vlieger E, Majoie C, Leenstra S, Den Heeten G. Functional magnetic resonance imaging for neurosurgical planning in neurooncology. *Eur Radiol.* 2004; 14:1143–1153. [PubMed: 15148622]
- Voyvodic J, Petrella J, Friedman A. fMRI activation mapping as a percentage of local excitation: consistent presurgical motor maps without threshold adjustment. *J Magn Reson Imaging.* 2009; 29:751–759. [PubMed: 19306363]
- Wei X, Yoo SS, Dickey CC, ZouH KH, Guttman C, Panych LP. Functional MRI of auditory verbal working memory: Long-term reproducibility analysis. *NeuroImage.* 2004; 21:1000–1008. [PubMed: 15006667]
- Wink A, Roerdink J. Denoising functional MR images: A comparison of wavelet denoising and Gaussian smoothing. *IEEE Trans Med Imaging.* 2004; 23:374–387. [PubMed: 15027530]
- Woolrich M, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith S. Bayesian analysis of neuroimaging data in FSL. *NeuroImage.* 2009; 45:S173–S186. [PubMed: 19059349]
- Wu TF, Lin CJ, Weng RC. Probability estimates for multi-class classification by pairwise coupling. *J Machine Learn Res.* 2004; 5:975–1005.
- Yoo SS, Dickey CC, Guttman C, Panych LP. Long-term reproducibility analysis of fMRI using hand motor tasks. *Int J Neurosci.* 2005; 115:55–77. [PubMed: 15768852]
- Zang Y, Jiang T, Lu Y, He Y, Tian L. Regional homogeneity approach to fMRI data analysis. *NeuroImage.* 2004; 22:394–400. [PubMed: 15110032]

Figure 1.

The procedure of brain activation detection under task stimulation. Multiple features are extracted from preprocessed fMRI data to represent each voxel. An offline feature selection is performed to identify most representative features. The OCSVM learning provides an initial mapping of active brain voxels. Prototype selection is used to identify voxels that are correctly classified by OCSVM. The identified voxels are used to train a TCSVM to reclassify all voxels to obtain refined activation maps. The prototype selection and TCSVM

learning/reclassification steps can be repeated multiple times till there is no more change in the final mapping results.

Figure 2.

R_{overlap} values of the active regions in the three ROIs computed using all sessions of subjects A–H. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 3.

R_{size} values of the active regions in the three ROIs computed using all sessions of subjects A–H. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 4.

KCC values of the active regions in the three ROIs computed using all sessions of subjects A–H. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 5. ICC(3,1) calculated based on the maximum CNR of active voxels in ROIs I, II, and III when (a) Daubechies wavelets with different vanishing moments, and (b) Gaussian kernels with different FWHM values were used. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 6.

ICC(3,1) calculated based on the average CNR of active voxels in ROIs I, II, and III when (a) Daubechies wavelets with different vanishing moments, and (b) Gaussian kernels with different FWHM values were used. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 7.

ICC(3,1) calculated based on the KCC value of active regions in ROIs I, II, and III when (a) Daubechies wavelets with different vanishing moments, and (b) Gaussian kernels with different FWHM values were used. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 8.

(a) Average CNR maps overlaid on the slices covering the three ROIs. Regions encircled by dark lines are group level activations detected by the SVM-based method, and regions encircled by the grey lines are group level activations identified by the GLM-based method.

(b) Voxels with 0.4 and higher detection frequency in the ROIs resulting from the SVM-based detection. Value 1 in the color bar corresponds to a 100% detection frequency. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Figure 9.
 (a) ICC maps overlaid on the same slices as shown in Figure 8 computed from the first two sessions of the seven subjects. (b) ICC maps overlaid on the slices covering the ROIs calculated using the first and eighth sessions of the seven subjects. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table I

ICC values of individual subjects calculated for ROIs I, II, and III with their 95% confidence intervals.

Subjects	ROI I		ROI II		ROI III	
	ICC	95% Confidence Interval	ICC	95% Confidence Interval	ICC	95% Confidence Interval
A	0.5	0.41~0.61	0.57	0.49~0.69	0.63	0.55~0.72
B	0.47	0.35~0.61	0.52	0.43~0.62	0.42	0.29~0.56
C	0.52	0.41~0.65	0.62	0.52~0.73	0.72	0.59~0.83
D	0.56	0.46~0.66	0.44	0.35~0.55	0.59	0.5~0.68
E	0.39	0.24~0.59	0.44	0.32~0.59	0.45	0.33~0.59
F	0.59	0.45~0.73	0.57	0.46~0.68	0.66	0.53~0.78
G	0.45	0.35~0.57	0.58	0.49~0.68	0.65	0.53~0.79
H	0.55	0.44~0.68	0.61	0.52~0.69	0.67	0.57~0.77

Table II

Statistics of ICC values computed after the spatial smoothing using the wavelet- and Gaussian kernel-based methods in ROI I.

ICC	Max CNR			Mean CNR			KCC	
	Wavelet	Gaussian	Wavelet	Wavelet	Gaussian	Wavelet	Wavelet	Gaussian
Mean	0.1338	0.1213	0.1619	0.5703	0.1559	0.5703	0.5482	0.0184
Std	0.0059	0.0006	0.0017	0.009	0.0068	0.009	0.0184	0.0541
Max Diff	0.017	0.0221	0.0046	0.0313	0.0223	0.0313	0.0541	
<i>t</i> test					0.0143		0.003	
<i>F</i> test		1.6×10^{-4}			0.0003		0.0457	

Table III
Statistics of ICC values computed after the spatial smoothing using the wavelet- and Gaussian kernel-based methods in ROI II.

ICC	Max CNR			Mean CNR			KCC	
	Wavelet	Gaussian	Wavelet	Gaussian	Wavelet	Gaussian	Wavelet	Gaussian
Mean	0.269	0.2653	0.2188	0.2646	0.4552	0.4198		
Std	0.0107	0.0084	0.0097	0.0113	0.0114	0.0082		
Max Diff	0.0345	0.0254	0.0336	0.0382	0.039	0.0285		
<i>t</i> test	0.407			1.4×10^{-4}		2.6×10^{-7}		
<i>F</i> test	0.4899			0.6644		0.3356		

Table IV
Statistics of ICC values computed after the spatial smoothing using the wavelet- and Gaussian kernel-based methods in ROI III.

ICC	Max CNR			Mean CNR			KCC	
	Wavelet	Gaussian	Wavelet	Wavelet	Gaussian	Wavelet	Wavelet	Gaussian
Mean	0.1048	0.1293	0.0645	0.3098	0.1285	0.3098	0.1910	0.0569
Std	0.0068	0.0101	0.0092	0.0255	0.0332	0.0255	0.1445	
Max Diff	0.0236	0.0254	0.0335	0.066	0.0827	0.066		
<i>t</i> test		5.3×10^{-6}		1.4×10^{-5}		1.1×10^{-5}		
<i>F</i> test		0.2556		7.4×10^{-4}		0.0255		

Table V

ICC values of the maximum CNR, mean CNR, and KCC of group level activations detected by the SVM-based method in ROIs I, II, and III.

	No. of Sessions	Maximum CNR			Mean CNR			Regional Homogeneity (KCC)		
		ICC	95% Confidence Interval	ICC	95% Confidence Interval	ICC	95% Confidence Interval	ICC	95% Confidence Interval	ICC
ROI I	4	0.48	0.09~0.85	0.33	−0.02~0.79	0.45	0.07~0.84	0.45	0.07~0.84	0.45
	8	0.14	−0.02~0.57	0.16	−0.01~0.59	0.54	0.27~0.87	0.54	0.27~0.87	0.54
ROI II	4	0.45	0.07~0.84	0.41	0.05~0.82	0.62	0.25~0.9	0.62	0.25~0.9	0.62
	8	0.29	0.07~0.72	0.25	0.04~0.68	0.49	0.21~0.84	0.49	0.21~0.84	0.49
ROI III	4	0.23	−0.08~0.72	0.18	−0.12~0.68	0.31	−0.03~0.77	0.31	−0.03~0.77	0.31
	8	0.12	−0.03~0.54	0.09	−0.04~0.49	0.32	−0.08~0.74	0.32	−0.08~0.74	0.32

Table VI

ICC values of the maximum CNR, mean CNR, and KCC of group level activations detected by the GLM-based method in ROIs I, II and III.

	No. of Sessions	Maximum CNR			Mean CNR			Regional Homogeneity (KCC)		
		ICC	95% Confidence Interval	ICC	95% Confidence Interval	ICC	95% Confidence Interval	ICC	95% Confidence Interval	ICC
ROI I	4	0.42	0.05~0.83	0.3			-0.04~0.77	0.36		0.004~0.8
	8	0.12	-0.03~0.54	0.12			-0.03~0.54	0.47		0.19~0.83
ROI II	4	0.39	0.02~0.81	0.39			0.02~0.81	0.49		0.12~0.86
	8	0.21	0.02~0.65	0.21			0.02~0.65	0.37		0.12~0.77
ROI III	4	0.28	-0.05~0.76	0.33			-0.02~0.78	0.26		-0.07~0.74
	8	0.16	-0.01~0.59	0.27			0.05~0.7	0.28		0.06~0.71