

Journal of Biomedical Optics

SPIDigitalLibrary.org/jbo

Is resting-state functional connectivity revealed by functional near-infrared spectroscopy test-retest reliable?

Han Zhang
Yu-Jin Zhang
Lian Duan
Shuang-Ye Ma
Chun-Ming Lu
Chao-Zhe Zhu

Is resting-state functional connectivity revealed by functional near-infrared spectroscopy test-retest reliable?

Han Zhang, Yu-Jin Zhang, Lian Duan, Shuang-Ye Ma, Chun-Ming Lu, and Chao-Zhe Zhu

Beijing Normal University, State Key Laboratory of Cognitive Neuroscience and Learning, Xin-Jie-Kou-Wai Street 19, Beijing 100875, China

Abstract. Recently, resting-state functional near-infrared spectroscopy (rs-fNIRS) research has experienced tremendous progress. Resting-state functional connectivity (RSFC) has been adopted as a pivotal biomarker in rs-fNIRS studies. However, it is yet to be clear if the RSFC derived from rs-fNIRS is reliable. This concern impedes extensive utilization of rs-fNIRS. We systematically address the issue of reliability. Sixteen subjects participate in two rs-fNIRS sessions held one week apart. RSFC in sensorimotor system is calculated using the seed-correlation approach. Then, test-retest reliability is evaluated at three different scales (map-, cluster-, and channelwise) for individual- and group-level RSFC derived from different types of fNIRS signals [oxygenated (HbO), deoxygenated (HbR), and total hemoglobin (HbT)]. The results show that, for HbO signals, individual-level RSFC generally has good-to-excellent map-/clusterwise reliability, while group-level RSFC has excellent reliability. For HbT signals, the results are similar. For HbR signals, the clusterwise reliability is comparable to that for HbO while the mapwise reliability is slightly lower (fair to good). Focusing on RSFC at a single channel, we report poor channelwise reliability for all three types of signals. We hereby propose that fNIRS-derived RSFC is a reliable biomarker if interpreted in map- and clusterwise manners. However, channelwise interpretation of individual RSFC should proceed with caution. © 2011 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.3591020]

Keywords: test-retest; reliability; near-infrared spectroscopy; optical imaging; functional connectivity; resting state.

Paper 10518R received Sep. 21, 2010; revised manuscript received Mar. 23, 2011; accepted for publication Apr. 26, 2011; published online Jun. 8, 2011.

1 Introduction

Functional near-infrared spectroscopy (fNIRS) is an emerging functional imaging technique with fewer physical restrictions, greater practicality, and better portability than other imaging techniques.¹ Using fNIRS, brain functional activity can be assessed by recording the concentration of both oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR) with relatively high temporal resolution (e.g., 10 Hz). In recent decades, task activation studies based on fNIRS have been extensively carried out in the field of cognitive and clinical neuroscience,^{2–6} and the results have been demonstrated to be reliable.^{7–9}

Recently, fNIRS has been adopted to investigate resting-state (i.e., task-free) brain function by White *et al.*¹⁰ and our group.^{11–13} During rest, the human brain intrinsically fluctuates with a low-frequency (<0.1 Hz) character.^{14,15} These low-frequency fluctuations (LFFs) are considered to be information rich,^{16–18} and the synchronization of LFFs within widely distributed neuroanatomical systems [i.e., resting-state functional connectivity (RSFC)] is believed to reflect a pivotal functional architecture of the brain.^{19–21} On the basis of resting-state fNIRS (rs-fNIRS), RSFC was successfully detected within the sensorimotor,^{10–12} visual,^{10,12} auditory,¹¹ and language systems,¹³ as well as within whole-brain networks.^{22,23}

Because rs-fNIRS experiments can be carried out with few requirements to the subjects, they are particularly suitable for studying special populations (e.g., neonates, infants, or hospital patients), therefore contributing to better understanding of human development and rehabilitation procedures. Furthermore, due to its low cost, rs-fNIRS is quite suitable for repetitive or continuous recording in brain plasticity studies.

However, an essential question remains to be elucidated: is the RSFC detected based on rs-fNIRS measurements and then proposed as a biomarker in previous studies reliable? In investigation of a scientific problem, reliability indicates to what extent we can trust our result. If it is not reliable, then the interpretation, inference, comparison, and integration of this result must be conducted with caution. Unfortunately, several factors may have a negative effect on the reliability of the fNIRS-based RSFC. First, the resting-state paradigm, in which subjects are instructed to keep still and relax their mind, is only a descriptive experimental paradigm without a specific task engagement or operative instruction, which makes it inherently unconstrained.²⁴ Therefore, experimental variations related to time and participants exist (e.g., different subjects have different moods during scanning, and the thoughts of a subject may vary at times), which impacts reliability. Second, the performance of fNIRS machines and the scanning environment (e.g., the illumination level) may also introduce variations between scanning sessions, which may also reduce reliability. Third, the location of the fNIRS optode may vary, though perhaps not significantly, across participants

Address all correspondence to: Chao-Zhe Zhu, Beijing Normal University, State Key Laboratory of Cognitive Neuroscience and Learning, Xin-Jie-Kou-Wai Street 19, Beijing 100875, China. Tel: 86 10 5880 2965; Fax: 86 10 5880 6154; E-mail: czzhu@bnu.edu.cn.

and scanning sessions³ and will thus reduce reliability.²⁵ Finally, other influencing factors, such as noise level and calculating parameters, may also reduce the reliability of RSFC result. Considering these factors, it is unclear whether RSFC findings based on rs-fNIRS are reliable enough to be adopted as biomarkers in cognitive neuroscience and clinical study. Thus, addressing such a question is of preeminent importance in the field of rs-fNIRS.

In this study, RSFC in the sensorimotor areas was calculated by using the seed-correlation approach utilized by the previous rs-fNIRS studies,^{10,11,13,22} and its reliability was comprehensively assessed via a test-retest experiment. On the basis of the findings, practical guidance was provided for future fNIRS-based RSFC study.

2 Materials and Methods

2.1 Subjects

Twenty-one college students were enrolled in the first session of the rs-fNIRS scan (this data were previously reported in Zhang *et al.*¹²), and 17 of them were rescanned in a following session after an interval of about one week (5–8 days). After further exclusion of a left-handed subject,²⁶ 16 right-handed subjects (ages 21.44 ± 1.82 , seven females and nine males) were involved. All procedures were conducted in accordance with the Declaration of Helsinki and were approved by the Ethics Committee at State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University. Written informed consent was obtained from all subjects before the experiment.

2.2 Protocols and fNIRS Measurements

In both sessions, each subject underwent an 11-min resting-state scan and a subsequent 6-min bilateral finger-tapping task scan. During the resting-state experiment, subjects were required to be seated in a chair and to keep still with eyes closed. In the finger-tapping task, subjects tap their bilateral index fingers in different sequences as indicated by visual instruction in a block-design experiment (see detailed descriptions in Lu *et al.*¹¹). The fNIRS scanning was conducted with a 52-channel ETG-4000 Optical Topography System (Hitachi Medical Company, Tokyo) with 17 emitters and 16 detector optodes (interoptode distance = 30 mm) in a holder cap located above the sensorimotor area, based on the international 10–20 system for electroencephalogram electrode placement (the optode between channels 47/48 was in Cz, and channels 32 and 42 were placed in T3 and T4, respectively). The channel locations were illustrated in Fig. 1(a). The absorption of near-infrared light at two wavelengths (695 and 830 nm) was measured with a 10-Hz sampling rate. The HbO and HbR hemoglobin signals were computed with the modified Beer–Lambert law,²⁷ and their sum produced the total hemoglobin (HbT) concentration signal.

2.3 Seed-Region Definition

Choosing a seed region is an important issue for the seed-correlation approach, because the results may depend (more or less) on *a priori* seed-region definition. In fNIRS studies, it is commonly difficult to define the seed region according to anatomical markers in the cortex. Thus, we used the task-session data and the resultant peak-activation channels to define

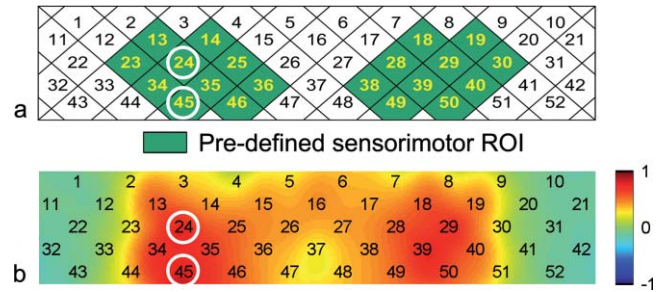


Fig. 1 Seed-region definition: (a) labeling map of the sensorimotor areas for all channels and (b) the averaged group-level task-activation map with normalized t -values ranging from -1 to 1 across two fNIRS recording sessions and three different hemoglobin concentration signals. The channels in white circles represent the seed channels.

the seed region in a more objective way.¹¹ Because specifying seed regions based only on a specific hemoglobin concentration signal or on a specific session can lead to bias, we took all the data (from two sessions and for three hemoglobin types) into consideration.

Specifically, six group-level task-activation maps (t -maps) for two sessions and for the three hemoglobin types were calculated based on the general linear model that was previously used in Lu *et al.*¹¹ with our in-house-developed scripts and the NIRS-SPM toolkit.²⁸ To equivalently take all activation maps into consideration, each of the group-level t -maps was first normalized to the same range (from -1 to 1). During normalization, the t value at each channel was divided by the maximum of the absolute t value in each map. The normalized maps were then averaged to form a single map [Fig. 1(b)]. The two most activated channels in Fig. 1(b), channel 45 (with a value of 0.845) and channel 24 (0.772), were selected as the seed region. Both channels were located at the left side of the predefined sensorimotor region [Fig. 1(a)]. As the data from two sessions were merged, a reasonable concern may rise if the near-infrared spectroscopy (NIRS) probe displacement between sessions affects the merging procedure and subsequently affects the seed-region definition. To address this concern, we quantified the extent to which the probe-location changed between sessions, using the distance between the task-activation centers derived from two sessions' data, the averaged value of which was only 1.79 cm and was smaller than the nearest channel-to-channel distance (~ 2.12 cm). Thus, we estimated that the probe displacement error was likely to have a small effect on the seed-region definition.

2.4 RSFC Calculation

The RSFC within the sensorimotor system was detected for each session and for each type of hemoglobin signal by means of seed correlation. The process, as defined in Lu *et al.*,¹¹ involves the following steps:

- (1) A visual inspection demonstrated that, for most of the subjects, the first 40 s and the last 20 s signals were unstable. Therefore, for all subjects, these data points were discarded from the study, leaving 10 min of data (6000 time points).
- (2) A bandpass filter (0.01–0.08 Hz)^{10,11,13,19} was applied to the rs-fNIRS time series for all channels to extract the

- LFFs and to remove noise and artifacts with extremely low or high frequencies.
- (3) As defined above, the seed time course was defined as the averaged LFFs signal in the seed region.
 - (4) Individual-level analysis was calculated channel by channel in the framework of a general linear model (GLM), with the seed time course evaluated as an independent variable and the filtered signals evaluated as dependent variables. The time series' autocorrelation was accounted for.²⁸ The resultant t -map for each subject, with t -value at each channel representing temporal resemblance between the seed time course and the signal at this channel, was considered to be an individual-level RSFC map.
 - (5) To make a group inference, a one-sample t -test was performed on all subjects' GLM-produced β maps (i.e., the maps consisting of the regression parameter corresponding to the seed-time-course regressor generated by GLM parameter estimation for all channels) in a random-effect framework. The resultant t -map was considered to be a group-level RSFC map (while from another view angle, it reflects both the resemblance of the individual RSFC map across all subjects and the overall RSFC strength taking all subjects into account).

2.5 Assessment of Test-Retest Reliability

To comprehensively evaluate the test-retest reliability of the fNIRS-based, seed-correlation-derived RSFC, we assessed at three spatial scales, namely, mapwise (with the largest spatial scale including all channels), clusterwise (a medium spatial scale including tens of channels), and channelwise (the smallest spatial scale focusing on a single channel).

Initially, the mapwise assessment was carried out to investigate the between-session reproducibility of the global RSFC map, both at the individual and group levels, using the Pearson correlation coefficient (r). Second, clusterwise reliability was assessed, both at the individual and group levels, using three different indices for reliability: (i) the reproducibility of the RSFC cluster size (R_{size}), (ii) the spatial overlap (R_{overlap}) of the RSFC cluster(s); and (iii) the reproducibility of the averaged RSFC strength within the RSFC cluster(s) [clusterwise intraclass correlation coefficient ($\text{ICC}_{\text{cluster}}$)]. The specific forms of the indices 1–2 are

$$R_{\text{size}} = 1 - |C1 - C2| / (C1 + C2) \quad (1)$$

$$R_{\text{overlap}} = 2 \times C_{\text{overlap}} / (C1 + C2) \quad (2)$$

according to Rombouts *et al.*²⁹ and Plichta *et al.*,^{7,8} where $C1$ and $C2$ denote the size of significant RSFC cluster(s) for both sessions, and C_{overlap} denotes the size of overlap between them. The third index, $\text{ICC}_{\text{cluster}}$, was calculated based on a two-way random effect model for consistency measurements³⁰ in the form of

$$\text{ICC} = \frac{\text{MS}_s - \text{MS}_e}{\text{MS}_s + (k - 1)\text{MS}_e}, \quad (3)$$

where MS_s and MS_e are the between-subject mean square and error mean square, respectively, k is the number of measurements ($k = 2$ in case of two sessions).

Finally, in the most localized view, channelwise reliability was assessed using an index of channelwise ICC ($\text{ICC}_{\text{cluster}}$), calculated similar to $\text{ICC}_{\text{cluster}}$. The only differences involved using the individual-level RSFC strength at each channel and calculating in a channelwise manner.^{31,32}

All the reliability indices listed above were evaluated according to the criteria proposed by Cicchetti and Sparrow,³³ wherein a value ≥ 0.75 indicates that reliability is “excellent,” 0.59–0.75 suggests “good,” 0.40–0.58 is “fair,” and ≤ 0.40 is “poor.”

3 Results

Figure 2 shows the group-level RSFC (t -maps) derived from step 5 in the RSFC calculation for the HbO, HbR, and HbT signals (first three rows, respectively) and for two sessions (left and right panels). Both were symmetrically distributed within the bilateral sensorimotor areas. The spatial patterns of the group-level RSFC from two sessions were highly similar, and both resembled the predefined sensorimotor region-of-interest [(preROI), last row of Fig. 2]. The t -statistical values, measuring the group-level RSFC strength, were generally higher for the HbO and HbT than those for the HbR signals.

3.1 Mapwise Reliability Assessment

To quantitatively evaluate the similarity between the group-level RSFC maps for two sessions, we plotted the group-level RSFC strength (t value) derived from session 1 against the corresponding RSFC strength from session 2 for each channel, producing the scatter plots depicted in Fig. 3. The scatter plots for the HbO, HbR, and HbT signals are depicted from left to right, respectively. Each data point represents the strength of the group-level RSFC at a single channel. Good-to-excellent mapwise reliability is indicated by the data points closely distributed near the fitted line with the spatial correlation coefficients $r = 0.78, 0.70,$ and 0.88 for the HbO, HbR, and HbT signals, respectively. It is noteworthy that all the significantly functional connected channels (at the top right of the two gray threshold lines with $p < 0.01$, uncorrected) for the HbO and HbT signals, and most for the HbR signals, were within the preROI (depicted by red dots). Such a result indicates our justifiable preROI definition, as well as the good quality of our rs-fNIRS data.

Mapwise reliability was also evaluated at the individual level. Table 1 summarizes the spatial correlation coefficients (r) between individual RSFC maps derived from two sessions for all 16 subjects and for the three hemodynamic parameters. For HbO, of the total 16 subjects, four showed excellent, nine with good, two with fair, and one with poor reliability. In this case, the reliability can be considered to be good to excellent. For HbR, there was one with excellent, three with good, eight with fair, and four with poor reliability. In this case, the reliability can be considered to be fair. For HbT, reliability data demonstrates three with excellent, four with good, five with fair, and four with poor. From this, it is difficult to classify the reliability.

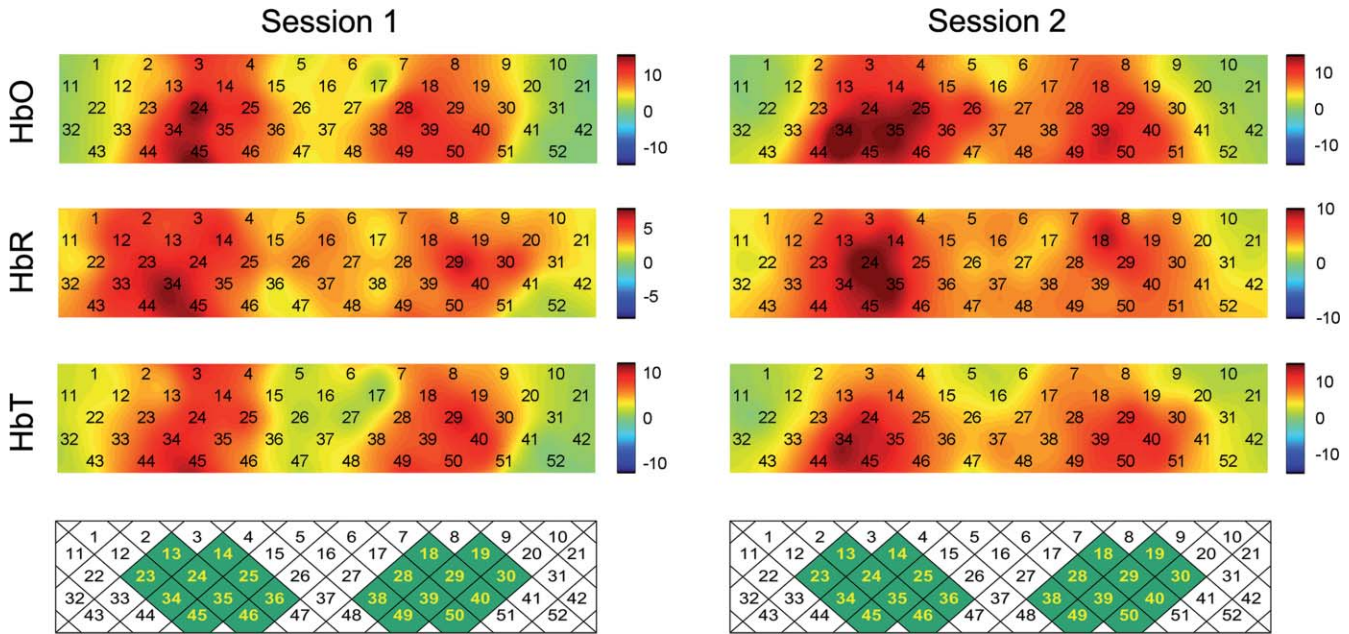


Fig. 2 Group-level RSFC maps for the HbO, HbR, and HbT signals (first three rows) derived from session 1 (left panel) and session 2 (right panel), together with the predefined sensorimotor region-of-interest (last row). Note that the value of each channel in those group-level RSFC maps equals the original t -value generated by the group analyses without normalized to -1 to 1 .

3.2 Clusterwise Reliability Assessment

We found that clusterwise reliability was even better than mapwise reliability. Both the R_{size} and $R_{overlap}$ evaluations for the group-level RSFC clusters at various threshold levels (from $p < 0.001$ to $p < 0.05$, all uncorrected) uniformly demonstrate excellent reliability (Table 2), indicating that the size and location of the RSFC clusters were largely reproducible.

The individual-level evaluation of the clusterwise reliability (Table 3) exhibited reproducibility at a weaker but still tangible level when compared to the previous group-level results, though under a more stringent threshold ($p < 0.001$, uncorrected) for cluster definition. Specifically, for HbO, 15 out of 16 subjects showed excellent reliability, with the one remaining being good, using R_{size} ; nine subjects demonstrate excellent reliability while the rest demonstrating good, using $R_{overlap}$. Therefore, we concluded that the individual-level, clusterwise reliability for HbO signal was good to excellent. For HbR, reliability data are as follows: nine, three, one, and two subjects have excellent, good, fair, and poor R_{size} , respectively, indicating good-to-excellent

reliability. For the result of $R_{overlap}$, however, it is difficult to classify the reliability as the big variance. For HbT, 13 subjects showed excellent reliability (with the rest being good) using R_{size} , and 14 showed good-to-excellent reliability (with the rest being fair) using $R_{overlap}$. This indicates a generally good-to-excellent reliability. Additionally, the previously mentioned analysis were also performed with less stringent threshold levels ($p < 0.005$, 0.01 , and 0.05), revealing more favorable levels of reliability.

Another index of clusterwise reliability, the $ICC_{cluster}$, demonstrated acceptable reliability for the HbO (good-to-excellent, with value being up to 0.77) and HbT (fair-to-good, up to 0.58) signals (Table 4). Here, both of the two types of the ICC , which evaluates reliability for single measures [$ICC(C, 1)$] and average measures [$ICC(C, k)$],^{7,8} were shown to give a comprehensive assessment. Please also note that the $ICC_{cluster}$ was calculated using two cluster definition strategies: the pre-ROI [Fig. 1(b)], and the *post hoc*-defined region-of-interest [(postROI), using significant group-level RSFC at session 1 with

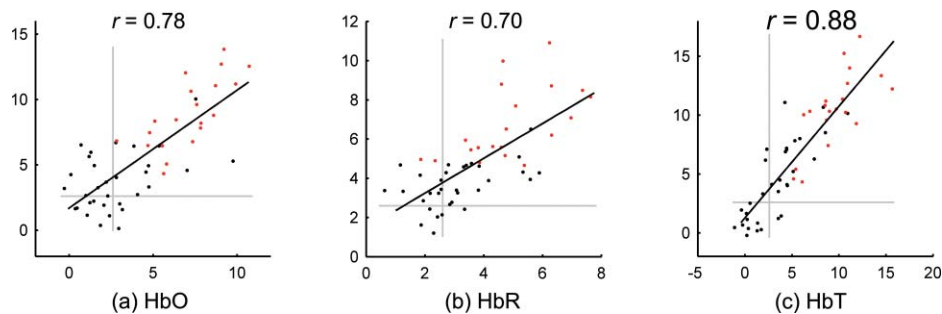


Fig. 3 Scatter plots of the group-level RSFC for the (left) HbO, (middle) HbR, and (right) HbT signals for the mapwise reliability assessment. Each data point represents the strength of the group-level RSFC at a single channel derived from sessions 1 and 2. The red data points indicate the channels in the predefined sensorimotor region. The solid black line is the fitted line, and the grey lines indicate the threshold of $p < 0.01$, uncorrected.

Table 1 Mapwise reliability at the individual level.

Subject	HbO	HbR	HbT
1	0.64	0.54	0.70
2	0.78	0.52	0.79
3	0.70	0.57	0.43
4	0.62	0.65	0.20
5	0.75	0.73	0.80
6	0.70	0.11	0.50
7	0.62	0.54	0.53
8	0.60	0.32	0.55
9	0.63	0.58	0.56
10	0.68	0.54	0.62
11	0.41	0.58	0.31
12	0.76	0.64	0.77
13	0.10	0.22	0.34
14	0.77	0.76	0.65
15	0.46	0.09	0.00
16	0.64	0.41	0.71
Mean (Std)	0.62 (0.17)	0.49 (0.20)	0.53 (0.23)

the threshold level of $p < 0.01$, uncorrected]. However, for the HbR signals, only the $ICC(C,k)$ showed fair reliability.

3.3 Channelwise Reliability Assessment

As shown in Table 4, reliability was determined using the averaged ICC_{channel} (outside the parentheses) across all channels, across the channels in preROI, and in postROI. All results were < 0.4 (indicating poor reliability). However, the maximum

Table 2 Clusterwise reliability at the group level.

		$p < 0.001$	$p < 0.005$	$p < 0.01$	$p < 0.05$
HbO	R_{size}	0.90	0.90	0.91	0.93
	R_{overlap}	0.84	0.78	0.83	0.83
HbR	R_{size}	0.81	0.85	0.88	0.99
	R_{overlap}	0.78	0.83	0.85	0.95
HbT	R_{size}	0.96	0.97	0.99	1.00
	R_{overlap}	0.93	0.92	0.94	0.95

Note: All the p values are the uncorrected ones.

ICC_{channel} value (inside the parentheses in Table 4) across all channels demonstrated acceptable reliability for all three hemodynamic contrasts (up to 0.54, 0.66, and 0.59 for the HbO, HbR, and HbT signals, respectively). This result indicates that, although the channelwise reliability is quite good at some channels, the overall channelwise reliability is low due to a large portion of channels exhibiting low reliability. Additionally, it is pertinent to mention that some negative ICC values were encountered when calculating ICC_{channel} . Such a situation is theoretically impossible,³⁴ and the interpretation of these values is still controversial.³⁵ Thus, we set the negatives to zero (i.e., completely not reliable).³⁶

4 Discussion

In this study, the test-retest reliability of the fNIRS-based, seed-correlation-derived RSFC was assessed at different spatial scales: map-, cluster-, and channelwise reliability. In all cases, we found acceptable map- and clusterwise reliability, but generally lower channelwise reliability (see Fig. 3 and Tables 1–4).

The different reliabilities identified among the map-/cluster- and channelwise assessments suggests that reliability should be interpreted with different levels of confidence for different spatial scales of the RSFC. Taking HbO data for example, both the overall pattern of the RSFC map and the quantitative/distributive characters of the RSFC clusters demonstrate good-to-excellent reliability. These results can thus be confidently trusted. In contrast, the individual RSFC for a single channel should likely be interpreted with caution due to a lower channelwise reliability. Practically, instead of being interested in a RSFC result at a single channel, we suggest researchers turn to the averaged RSFC value within a round cluster centered at the interested channel [as frequently performed in region-of-interest (ROI) analysis often adopted in the community of functional magnetic resonance imaging (fMRI)] to increase the reliability of the final result. In other words, it is favorable to construct an ROI and place the more reliable cluster-level result back to the specific channel to replace the less reliable channel-level result. Such an approach is actually a trade-off, balancing between the spatial resolution and reliability.

Previously, investigations have been carried out on test-retest reliability of the task activation based on fNIRS during a similar motor,⁸ as well as for other,^{7,9} task protocols. Those studies, together with ours, are both important and significant for the fNIRS community: together they contribute to the knowledge on fNIRS-based results' reliability or reproducibility. By now, it has been extended to include the measurement of the intrinsic functional architecture within resting-state brain.

Compared to the reliability of the task activations,^{7–9} the reliability of RSFC demonstrates comparable (range from fair to excellent) map- and clusterwise, but lower channelwise, reliability. The difference in channelwise reliability between task activations and RSFC may lie in the different signal generation biomechanics evaluated (i.e., task-evoked signals or spontaneous fluctuations). As speculated in Sec. 1, in contrast to the more consistent hemodynamic response during engagement of a specific task, the unconstrained nature of the resting state²⁴ makes it more difficult to replicate the channel-level result between sessions and across subjects. Such an effect may be

Table 3 Clusterwise reliability at the individual level.

Subject	HbO		HbR		HbT	
	R_{size}^*	R_{overlap}^*	R_{size}^*	R_{overlap}^*	R_{size}^*	R_{overlap}^*
1	0.91	0.91	0.67	0.63	0.98	0.91
2	0.97	0.92	0.48	0.40	0.99	0.87
3	0.90	0.84	0.96	0.68	0.95	0.87
4	0.98	0.73	0.39	0.33	0.85	0.59
5	0.63	0.59	0.73	0.73	0.62	0.53
6	0.82	0.79	0.89	0.80	0.83	0.76
7	0.99	0.88	0.79	0.79	0.79	0.76
8	0.96	0.74	0.84	0.68	0.91	0.69
9	0.81	0.81	0.82	0.71	0.73	0.73
10	0.97	0.91	0.81	0.74	0.90	0.87
11	0.90	0.66	0.27	0.27	0.96	0.67
12	0.82	0.64	0.90	0.80	0.95	0.89
13	0.92	0.66	0.94	0.38	0.90	0.60
14	0.84	0.75	0.65	0.53	0.76	0.70
15	0.82	0.68	0.89	0.43	0.71	0.55
16	0.91	0.88	0.32	0.32	0.89	0.89
Mean (Std)	0.88 (0.09)	0.77 (0.11)	0.71 (0.23)	0.58 (0.19)	0.86 (0.11)	0.74 (0.13)

* R_{size} and R_{overlap} at the threshold of $p < 0.001$, uncorrected.

further magnified due to other factors that include the potential placement-variability of the NIRS optodes between sessions and across subjects, absence of an optode coregistration algorithm based on craniocerebral correlation, and the low spatial resolution of the fNIRS. When all these factors are considered together, the exact channel-to-channel match for both RSFC strength and location is impossible, leading to the lower channelwise reliability observed.

Regarding the different hemodynamic parameters measured by fNIRS, HbO and HbT provided better RSFC reliability than HbR. Specifically, the map-, cluster- and channelwise reliability of the individual-level RSFC derived from HbO signals was higher than that derived from HbT and HbR signals (where result showed the weakest reliability) (see Tables 1–3, and 4). For the group-level RSFC (Fig. 3 and Table 2), the result was a little different, with reliability for HbT slightly being higher than those for HbO and HbR. However, in all cases, the HbR signal-derived RSFC demonstrated the lowest reliability compared to the other two hemodynamic variables, especially at the individual level. Such a result is in line with the weaker RSFC strength derived from the HbR signals (see Fig. 2; also reported in previous rs-fNIRS studies),^{3,11} which has been speculated to be related to the lower signal-to-noise ratio of the HbR signal

than that of the HbO and HbT signals.³ This speculation also came from previous reliability studies based on task fNIRS,^{8,9} in which similar result to ours were found. These results suggest that in subsequent studies, it would be better to use HbO signals to conduct a seed-correlation-based RSFC analysis than HbR or HbT data.

Concerning the generally poor (except for a small number of channels with acceptable) channelwise reliability (averaged ICC < 0.4 but maximum ICC > 0.4 in Table 4), we had hoped to demonstrate that the channels with high reliability were those with high RSFC strength. However, further examination revealed that the channels with high channelwise reliability did not specifically lie within the predefined sensorimotor areas or the *post hoc*-defined RSFC clusters. Additionally, there was non-significant correlation ($p > 0.05$) between ICC values and group-level RSFC strength across channels. This result is in agreement with previous findings,^{37,38} which identified highly functional connected channels with low reliability and subthreshold channels having high reliability (see comparison between maximum ICC across channels in the entire probe, preROI, and postROI in Table 4).

Because of the fact that the optical pathway of the near-infrared light also includes superficial nonbrain tissues (e.g.,

Table 4 Channelwise and clusterwise intraclass correlation coefficients (ICCs).

		Channelwise		Clusterwise	
		ICC(C,1) ^d	ICC(C,k) ^d	ICC(C,1)	ICC(C,k)
HbO	Whole ^a	0.19 (0.54)	0.29 (0.70)		
	preROI ^b	0.24 (0.48)	0.36 (0.65)	0.61	0.76
	postROI ^c	0.20 (0.48)	0.31 (0.65)	0.63	0.77
HbR	Whole	0.17 (0.66)	0.26 (0.80)		
	preROI	0.20 (0.66)	0.29 (0.80)	0.31	0.47
	postROI	0.19 (0.66)	0.29 (0.80)	0.25	0.41
HbT	Whole	0.18 (0.59)	0.28 (0.74)		
	preROI	0.20 (0.50)	0.31 (0.66)	0.41	0.58
	postROI	0.20 (0.59)	0.31 (0.74)	0.40	0.57

^aSearch within all channels in the whole probe.

^bSearch within the predefined sensorimotor ROI.

^cSearch within the *post hoc* defined ROI (significant group-level RSFC at session 1, $p < 0.01$, uncorrected).

^dReported values include mean value (max value).

skull, scalp, cerebrospinal fluid, etc.), fNIRS measurements also contain signals related to systemic activity and/or superficial signals.¹⁰ Therefore, it would be natural for concerns to rise if our findings were affected by systemic activity and/or superficial signals. However, we argue that such an effect could not be dominant based on the direct and indirect evidences from the literature as noted below. A new study, which also used a continuous-wave single distance fNIRS device similar to us, demonstrated that the RSFC between homologous regions was significantly larger than the RSFC between “control” regions, where the RSFC was not expected.³⁹ The result plus the previous findings from our papers^{11,13} suggested that the RSFC detected was nondominantly affected by superficial signals and/or systemic noise. Direct evidence was from the study of Katura *et al.*,⁴⁰ where systemic cardiovascular dynamics (as one of the sources of systemic noise) have a nondominant contribution to the hemodynamic changes in the low-frequency range (with contributions of 35% for HbO and 7% for HbR). Moreover, in a previous study,⁴¹ the effect of the skin blood-flow fluctuation, as one of the sources of the superficial signals, on the low-frequency signal has been demonstrated to be nondominant. Attributed to this finding, this type of the superficial signal may thus have a small effect on the RSFC derived from the low-frequency signal in our study. In addition, the high-frequency components of the systemic noises, contributed by cardiac pulsations (0.6–1.2 Hz) and respiration fluctuation (0.1–0.5 Hz),¹⁴ have been reduced by the bandpass filter in our study with an upper-limit frequency of 0.08 Hz.

Despite the interesting findings from this test-retest study, we must note its limitations as discussed here. First, this study is mainly designed to test the reliability of a new technical development rather than the validity of it. Specifically in this study, reliability assesses the consistency of the fNIRS-based,

seed-correlation–derived RSFC between sessions. However, reliability does not necessarily present implications regarding the validity, which assesses the accuracy of the RSFC. As an important characteristic, validity deserves full investigation in future. As such, our ongoing study is underway with a goal of determining the validity of the fNIRS-based, seed-correlation–derived RSFC by using simultaneous recording of fNIRS and fMRI.

Second, in this study we focused on consistency between repeat measurements (i.e., test-retest reliability) rather than the reproducibility between laboratories, other NIRS systems, other RSFC calculation methods or, even other neuroanatomical systems. These problems are equally important and need further study. As such, when interpreting our results one should always consider the specific data recording and analyzing methods we used. In changing these parameters, the result itself may be also altered. We suggest that the reliability or reproducibility be taken as a “golden standard” to compare difference in data measuring, analyzing methods, and parameters (e.g., different frequency filtering bands) to figure out the most optimized RSFC-detection pipeline.

Third, the poor channelwise reliability we identified may be somewhat connected to absence of optode coregistration during data preprocessing. In future studies, several existing algorithms and procedures (such as probabilistic registration⁴² and virtual registration)⁴³ should be borrowed and utilized in fNIRS-based RSFC studies. In summary, reliability characterization is the foundation of a novel scientific researching tool under development. It is important, necessary, and cannot be overlooked.

5 Conclusions

The test-retest assessment of the seed-correlation–derived RSFC based on rs-fNIRS demonstrates acceptable map- and clusterwise reliability for both individual- and group-level RSFC (and even better for HbO signal-derived RSFC). However, the assessment does not demonstrate adequate channelwise reliability. Such a result suggests that the fNIRS-based, seed-correlation–derived RSFC can be treated as a reliable biomarker if interpreted at a larger scale, but the channelwise interpretation of individual RSFC should be conducted with caution.

Acknowledgments

This work was supported by the Natural Science Foundation of China (Grant No. 30970773), the Foundational Research Funds for the Central Universities, and the Engagement Fund of Outstanding Doctoral Dissertation of Beijing Normal University (Grant No. 08048).

References

1. A. Villringer and B. Chance, “Non-invasive optical spectroscopy and imaging of human brain function,” *Trends Neurosci.* **20**(10), 435–442 (1997).
2. I. Miyai, H. Yagura, M. Hatakenaka, I. Oda, I. Konishi, and K. Kubota, “Longitudinal optical imaging study for locomotor recovery after stroke,” *Stroke* **34**(12), 2866–2870 (2003).
3. Y. Hoshi, “Functional near-infrared spectroscopy: current status and future prospects,” *J. Biomed. Opt.* **12**(6), 062106 (2007).
4. F. Irani, S. M. Platek, S. Bunce, A. C. Ruocco, and D. Chute, “Functional near infrared spectroscopy (fNIRS): an emerging neuroimaging

- technology with important applications for the study of brain disorders," *Clin. Neuropsychol.* **21**(1), 9–37 (2007).
5. T. Nakano, H. Watanabe, F. Homae, and G. Taga, "Prefrontal cortical involvement in young infants' analysis of novelty," *Cereb. Cortex* **19**(2), 455–463 (2009).
 6. H. Watanabe, F. Homae, and G. Taga, "General to specific development of functional activation in the cerebral cortexes of 2- to 3-month-old infants," *Neuroimage* **50**(4), 1536–1544 (2010).
 7. M. M. Plichta, M. J. Herrmann, C. G. Baehne, A. C. Ehlis, M. M. Richter, P. Pauli, and A. J. Fallgatter, "Event-related functional near-infrared spectroscopy (fNIRS): are the measurements reliable?," *Neuroimage* **31**(1), 116–124 (2006).
 8. M. M. Plichta, M. J. Herrmann, C. G. Baehne, A. C. Ehlis, M. M. Richter, P. Pauli, and A. J. Fallgatter, "Event-related functional near-infrared spectroscopy (fNIRS) based on craniocerebral correlations: reproducibility of activation?," *Hum. Brain Mapp.* **28**(8), 733–741 (2007).
 9. M. Schecklmann, A. C. Ehlis, M. M. Plichta, and A. J. Fallgatter, "Functional near-infrared spectroscopy: a long-term reliable tool for measuring brain activity during verbal fluency," *Neuroimage* **43**(1), 147–155 (2008).
 10. B. R. White, A. Z. Snyder, A. L. Cohen, S. E. Petersen, M. E. Raichle, B. L. Schlaggar, and J. P. Culver, "Resting-state functional connectivity in the human brain revealed with diffuse optical tomography," *Neuroimage* **47**(1), 148–156 (2009).
 11. C. M. Lu, Y. J. Zhang, B. B. Biswal, Y. F. Zang, D. L. Peng, and C. Z. Zhu, "Use of fNIRS to assess resting state functional connectivity," *J. Neurosci. Methods* **186**(2), 242–249 (2009).
 12. H. Zhang, Y. J. Zhang, C. M. Lu, S. Y. Ma, Y. F. Zang, and C. Z. Zhu, "Functional connectivity as revealed by independent component analysis of resting-state fNIRS measurements," *Neuroimage* **51**(3), 1150–1161 (2010).
 13. Y. J. Zhang, C. M. Lu, B. B. Biswal, Y. F. Zang, D. L. Peng, and C. Z. Zhu, "Detecting resting-state functional connectivity in the language system using functional near-infrared spectroscopy," *J. Biomed. Opt.* **15**, 047003 (2010).
 14. D. Cordes, V. M. Haughton, K. Arfanakis, J. D. Carew, P. A. Turski, C. H. Moritz, M. A. Quigley, and M. E. Meyerand, "Frequencies contributing to functional connectivity in the cerebral cortex in "resting-state" data," *AJNR Am. J. Neuroradiol.* **22**(7), 1326–1333 (2001).
 15. D. Cordes, V. M. Haughton, K. Arfanakis, G. J. Wendt, P. A. Turski, C. H. Moritz, M. A. Quigley, and M. E. Meyerand, "Mapping functionally related regions of brain with functional connectivity MR imaging," *AJNR Am. J. Neuroradiol.* **21**(9), 1636–1644 (2000).
 16. M. D. Fox, A. Z. Snyder, J. L. Vincent, and M. E. Raichle, "Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior," *Neuron* **56**(1), 171–184 (2007).
 17. T. Kenet, D. Bibitchkov, M. Tsodyks, A. Grinvald, and A. Arieli, "Spontaneously emerging cortical representations of visual attributes," *Nature* **425**(6961), 954–956 (2003).
 18. M. Tsodyks, T. Kenet, A. Grinvald, and A. Arieli, "Linking spontaneous activity of single cortical neurons and the underlying functional architecture," *Science* **286**(5446), 1943–1946 (1999).
 19. B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echoplanar MRI," *Magn. Reson. Med.* **34**(4), 537–541 (1995).
 20. M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nat. Rev. Neurosci.* **8**(9), 700–711 (2007).
 21. M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, "Functional connectivity in the resting brain: a network analysis of the default mode hypothesis," *Proc. Natl. Acad. Sci. U S A* **100**(1), 253–258 (2003).
 22. F. Homae, H. Watanabe, T. Otobe, T. Nakano, T. Go, Y. Konishi, and G. Taga, "Development of global cortical networks in early infancy," *J. Neurosci.* **30**(14), 4877–4882 (2010).
 23. R. C. Mesquita, M. A. Franceschini, and D. A. Boas, "Resting state functional connectivity of the whole head with near-infrared spectroscopy," *Biomed. Opt. Express* **1**(1), 324–336 (2010).
 24. Z. Shehzad, A. M. Kelly, P. T. Reiss, D. G. Gee, K. Gotimer, L. Q. Uddin, S. H. Lee, D. S. Margulies, A. K. Roy, B. B. Biswal, E. Petkova, F. X. Castellanos, and M. P. Milham, "The resting brain: unconstrained yet reliable," *Cereb. Cortex* **19**(10), 2209–2229 (2009).
 25. S. Kohno, I. Miyai, A. Seiyama, I. Oda, A. Ishikawa, S. Tsuneishi, T. Amita, and K. Shimizu, "Removal of the skin blood flow artifact in functional near-infrared spectroscopic imaging data through independent component analysis," *J. Biomed. Opt.* **12**(6), 062111 (2007).
 26. R. C. Oldfield, "The assessment and analysis of handedness: the Edinburgh inventory," *Neuropsychologia* **9**(1), 97–113 (1971).
 27. M. Cope and D. T. Delpy, "System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infrared transillumination," *Med. Biol. Eng. Comput.* **26**(3), 289–294 (1988).
 28. J. C. Ye, S. Tak, K. E. Jang, J. Jung, and J. Jang, "NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy," *Neuroimage* **44**(2), 428–447 (2009).
 29. S. A. Rombouts, F. Barkhof, F. G. Hoogenraad, M. Sprenger, J. Valk, and P. Scheltens, "Test-retest analysis with functional MR of the activated area in the human visual cortex," *AJNR Am. J. Neuroradiol.* **18**(7), 1317–1322 (1997).
 30. K. O. McGraw and S. P. Wong, "Forming Inferences About Some Intraclass Correlation Coefficients," *Psychol. Methods* **1**(1), 30–46 (1996).
 31. M. Raemaekers, M. Vink, B. Zandbelt, R. J. van Wezel, R. S. Kahn, and N. F. Ramsey, "Test-retest reliability of fMRI activation during prosaccades and antisaccades," *Neuroimage* **36**(3), 532–542 (2007).
 32. X. N. Zuo, C. Kelly, J. S. Adelstein, D. F. Klein, F. X. Castellanos, and M. P. Milham, "Reliable intrinsic connectivity networks: test-retest evaluation using ICA and dual regression approach," *Neuroimage* **49**(3), 2163–2177 (2010).
 33. D. V. Cicchetti and S. A. Sparrow, "Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior," *Am. J. Ment. Defic.* **86**(2), 127–137 (1981).
 34. V. Rousson, T. Gasser, and B. Seifert, "Assessing intrarater, inter-rater and test-retest reliability of continuous measurements," *Stat. Med.* **21**(22), 3431–3446 (2002).
 35. R. Muller and P. Buttner, "A critical discussion of intraclass correlation coefficients," *Stat. Med.* **13**(23–24), 2465–2476 (1994).
 36. J. Kong, R. L. Gollub, J. M. Webb, J. T. Kong, M. G. Vangel, and K. Kwong, "Test-retest study of fMRI signal change evoked by electroacupuncture stimulation," *Neuroimage* **34**(3), 1171–1181 (2007).
 37. C. M. Bennett and M. B. Miller, "How reliable are the results from functional magnetic resonance imaging?," *Ann. NY Acad. Sci.* **1191**(1), 133–155 (2010).
 38. A. Caceres, D. L. Hall, F. O. Zelaya, S. C. Williams, and M. A. Mehta, "Measuring fMRI reliability with the intraclass correlation coefficient," *Neuroimage* **45**(3), 758–768 (2009).
 39. S. Sasai, F. Homae, H. Watanabe, and G. Taga, "Frequency-specific functional connectivity in the brain during resting state revealed by NIRS," *Neuroimage* **56**(1), 252–257 (2011).
 40. T. Katura, N. Tanaka, A. Obata, H. Sato, and A. Maki, "Quantitative evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics," *Neuroimage* **31**, 1592–1600 (2006).
 41. H. Obrig, M. Neufang, R. Wenzel, M. Kohl, J. Steinbrink, K. Einhaupl, and A. Villringer, "Spontaneous low frequency oscillations of cerebral hemodynamics and metabolism in human adults," *Neuroimage* **12**, 623–639 (2000).
 42. A. K. Singh, M. Okamoto, H. Dan, V. Jurcak, and I. Dan, "Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI," *Neuroimage* **27**, 842–851 (2005).
 43. D. Tsuzuki, V. Jurcak, A. K. Singh, M. Okamoto, E. Watanabe, and I. Dan, "Virtual spatial registration of stand-alone fNIRS data to MNI space," *Neuroimage* **34**, 1506–1518 (2007).