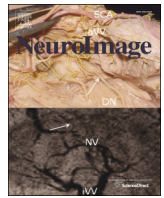


Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Review

Statistical analysis of fNIRS data: A comprehensive review

Sungho Tak^{a,b}, Jong Chul Ye^{a,*}^a Bio Imaging & Signal Processing Lab., Dept. of Bio & Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea^b Rotman Research Institute at Baycrest Centre, University of Toronto, Toronto, Ontario M6A 2E1, Canada

ARTICLE INFO

Article history:

Accepted 7 June 2013

Available online xxxx

Keywords:

fNIRS

t-Test

Correlation analysis

Spectral analysis

GLM

Statistical parameter mapping

Multi-level analysis

Group analysis

Multiple comparison

Data-driven analysis

ABSTRACT

Functional near-infrared spectroscopy (fNIRS) is a non-invasive method to measure brain activities using the changes of optical absorption in the brain through the intact skull. fNIRS has many advantages over other neuroimaging modalities such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), or magnetoencephalography (MEG), since it can directly measure blood oxygenation level changes related to neural activation with high temporal resolution. However, fNIRS signals are highly corrupted by measurement noises and physiology-based systemic interference. Careful statistical analyses are therefore required to extract neuronal activity-related signals from fNIRS data. In this paper, we provide an extensive review of historical developments of statistical analyses of fNIRS signal, which include motion artifact correction, short source-detector separation correction, principal component analysis (PCA)/independent component analysis (ICA), false discovery rate (FDR), serially-correlated errors, as well as inference techniques such as the standard *t*-test, *F*-test, analysis of variance (ANOVA), and statistical parameter mapping (SPM) framework. In addition, to provide a unified view of various existing inference techniques, we explain a linear mixed effect model with restricted maximum likelihood (ReML) variance estimation, and show that most of the existing inference methods for fNIRS analysis can be derived as special cases. Some of the open issues in statistical analysis are also described.

© 2013 Elsevier Inc. All rights reserved.

Contents

Introduction	0
Imaging physics	0
Modified Beer–Lambert law	0
Tomographic mapping	0
Signal processing of fNIRS signals	0
Motion artifact correction	0
Short-distance correction	0
PCA and ICA for skin blood flow removal	0
Removal of baseline fluctuation	0
Partitioning of physiological noise signals	0
Statistical inferences for fNIRS signals	0
Generalized linear mixed model (GLMM)	0
Derivation of classical statistical analysis for fNIRS	0
ANOVA	0
Basic <i>t</i> -test	0
GLM-based fNIRS time course analysis	0
Single-level versus multi-level approach in SPM-type group analysis	0
Standard statistics versus SPM type multi-level group analysis	0
The multiple comparison problem in fNIRS	0
Hypothesis testing	0

* Corresponding author.

E-mail address: jong.ye@kaist.ac.kr (J.C. Ye).

Multiple comparison problems	0
Multiple comparison in discrete setup	0
Multiple comparison in continuous setup	0
False discovery rate control	0
Functional connectivity analysis	0
Seed-based approaches	0
ICA-based approach	0
Conclusion and outlooks	0
Acknowledgments	0
Equivalence	0
References	0

Introduction

Functional near-infrared spectroscopy (fNIRS) is a non-invasive method to measure brain activity by measuring the absorption of the near-infrared light between 650 and 950 nm through the intact skull (Villringer and Dirnagl, 1995). As the absorption spectra of oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR) are distinct in this region, it is possible to determine the concentration changes of HbO and HbR from diffusely scattered light measurement (Ferrari et al., 2004; Jobsis, 1977; Kleinschmidt et al., 1996; Villringer et al., 1993).

fNIRS has many advantages over other neuroimaging modalities. For example, fNIRS can measure a wide range of functional contrast such as HbO, HbR, and total hemoglobin (HbT). In contrast to blood-oxygenation level-dependent (BOLD) response measured by functional magnetic resonance imaging (fMRI), which is a nonlinear function of oxygen level and cerebral blood flow (Buxton et al., 2004), a direct measurement of oxygen concentration in fNIRS can provide a potential way to unveil the mechanism of complicated neurovascular coupling (Tak et al., 2010, 2011; Yücel et al., 2012). Furthermore, fNIRS has high temporal resolution that allows us to study the temporal behavior of the hemodynamic response to neural activation. Such high temporal resolution is especially useful for functional connectivity analysis which has become increasingly important as a new brain imaging paradigm (Hall et al., 2013; Homae et al., 2011; Kozel et al., 2009; Lloyd-Fox et al., 2010; Medvedev et al., 2011; Mesquita et al., 2010; Niu et al., 2012; Quaresima et al., 2012; Y. Zhang et al., 2010). Other advantages of fNIRS include that it requires only a compact measurement system and is robust to motion artifacts, which allow brain studies during daily working, exercise, and rehabilitation (Arenth et al., 2007; Ferrari et al., 2004; Hoshi, 2003; Irani et al., 2007; Saitou et al., 2000).

However, fNIRS lacks anatomical information, making it difficult to localize the brain area where the fNIRS signal originated (Ferrari et al., 2004; Kleinschmidt et al., 1996; Lloyd-Fox et al., 2010; Villringer et al., 1993). Moreover, fNIRS has poor spatial resolution and limited penetration depth due to the high level of light scattering within tissue (Ferrari et al., 2004; Kleinschmidt et al., 1996; Lloyd-Fox et al., 2010; Villringer et al., 1993). Another important limitation is that the fNIRS signal is corrupted by measurement noise, motion artifacts, and physiological noise arising from cardiac pulsation, respiration, and blood pressure Mayer waves (Boas et al., 2004). Therefore, to improve the sensitivity and spatial specificity of neuronal activity from fNIRS data, careful statistical analysis is required (Ferrari et al., 2004; Koh et al., 2007; Lloyd-Fox et al., 2010; Schroeter et al., 2002; Ye et al., 2009).

In the early ages of fNIRS studies, brain signal detection was usually attempted through visual inspection or simple thresholding with some preprocessing steps (Benaron et al., 2000; Murata et al., 2002). However, such heuristic approaches are prone to error, especially when the noise and interference levels increase, so more rigorous statistical analyses were called for. Hence, various statistical

analysis methods such as *t*-test or analysis of variance (ANOVA) have been applied (Germon et al., 1994; Hoshi et al., 2001, 2003; Isobe et al., 2001; Kleinschmidt et al., 1996; Mehagnoul-Schipper et al., 2002; Okamoto et al., 2004; Tsujimoto et al., 2004; Young et al., 2000). These approaches often take average values during the task period as data to avoid any assumption of the exact shape or timing of the time course of changes in HbO and HbR in response to stimuli.

However, such an average sample-based statistical test also has limitations in that it does not utilize the time course of data, which is quite important in fNIRS data. Therefore, many investigators are interested in understanding the fNIRS time course. The well-known and widely used method of regression approach in this regard is the general linear model (GLM) (Friston et al., 2011), which assumes that data can be represented as a linear combination of several sources (regressors). These several sources consist of task-related regressors and non-task related so-called nuisance regressors. Schroeter et al. (2004) were the first to apply the GLM to analyze fNIRS data to overcome the uncertainty of the assumed differential pathlength factor (DPF), and numerous authors have employed GLM analysis for a variety of fNIRS experiments (Abdelnour and Huppert, 2009; Custo et al., 2010; Koh et al., 2007; Minagawa-Kawai et al., 2011; Plichta et al., 2006, 2007; Shimada and Hiraki, 2006; Singh and Dan, 2006; Ye et al., 2009). Another advantage of GLM is that group analysis can be implemented relatively simply using multi-level analysis, as investigated by many authors (Ciftci et al., 2008; Okamoto et al., 2006; Plichta et al., 2006; Singh and Dan, 2006; Ye et al., 2009).

Having reviewed the historical developments of statistical analysis briefly, in this paper we are interested in explaining these tools in more detail in a self-contained manner so that fNIRS practitioners can easily obtain the required information without searching through many papers and textbooks. In particular, rather than describing specific statistical analysis, such as one sample *t*-test, paired *t*-test, ANOVA, GLM, multi-level analysis, etc. with separate contexts, we show that all these methods can be derived from a general mixed model with restricted maximum likelihood (ReML) covariance estimation and hypothesis testing. Even though most of these materials are not novel and can be found in standard statistics textbooks, we believe that such a unified and extensive review can help students understand the existing works in a more organized way and provide them the opportunity to develop their own statistical approaches for their specific problems.

In addition, we also provide an overview of the existing data-driven approaches that have been investigated quite extensively to overcome the limitations of the classical approaches, and in particular for the new research area of functional brain connectivity using fNIRS (Hall et al., 2012; Homae et al., 2011; Kozel et al., 2009; Lloyd-Fox et al., 2010; Medvedev et al., 2011; Mesquita et al., 2010; Niu et al., 2012; Quaresima et al., 2012; Y. Zhang et al., 2010).

While this paper is mainly to review the existing works, there are a few novel contributions. In particular, a complete derivation of group analysis using multi-level analysis is performed using linear mixed effect model, which was not available before and will be included in a

new release of NIRS-SPM toolbox (<http://bisp.kaist.ac.kr/NIRS-SPM.html>).

This paper is organized as follows. The **Imaging physics** section provides the basic physics for fNIRS measurement. Existing signal processing approach for fNIRS statistical analysis is reviewed in the **Signal processing of fNIRS signals** section. The **Statistical inferences for fNIRS signals** section describes the mixed effect models and their basic mathematical tools, and we describe how these tools can be applied to fNIRS. Multiple comparison problems that are encountered in fNIRS studies are discussed in **The multiple comparison problem in fNIRS** section. The **Functional connectivity analysis** section reviews fNIRS functional connectivity analysis, which is followed by the **Conclusion and outlooks** section.

Imaging physics

Modified Beer–Lambert law

The modified Beer–Lambert law (Cope and Delpy, 1988), describing optical attenuation in a highly scattering medium, allows us to quantify the changes in the concentrations of HbO and HbR from the absorption of near-infrared light. According to the modified Beer–Lambert law, a change in optical density AOD can be described as

$$\Delta OD(r_s, r_d; \lambda, t) = -\ln \frac{\Phi(r_s, r_d; \lambda, t)}{\Phi_0(r_s, r_d; \lambda)} \approx \Delta \mu_a(r_s, r_d; \lambda, t) BL, \quad (1)$$

where r_s is the source position, r_d is the detector position, λ is the illumination length, $\Phi(r_s, r_d; \lambda, t)$ is the photon fluence at the detector position r_d at time t , generated by a source located at r_s , $\Phi_0(r_s, r_d; \lambda)$ is the incident photon fluence, $\Delta \mu_a(r_s, r_d; \lambda, t)$ is the absorption coefficient variation, B is a differential pathlength factor, and L is the distance between the source and detector. Assuming that the dominant chromophores in tissue for near-infrared wavelengths of 650 to 950 nm are HbO and HbR, the absorption coefficient variation is given by

$$\Delta \mu_a(r_s, r_d; \lambda, t) = \epsilon_{\text{HbO}}(\lambda) \Delta[\text{HbO}](r_s, r_d; t) + \epsilon_{\text{HbR}}(\lambda) \Delta[\text{HbR}](r_s, r_d; t), \quad (2)$$

where $\epsilon_{\text{HbO}}(\lambda)$ and $\epsilon_{\text{HbR}}(\lambda)$ are the extinction coefficients of HbO and HbR at wavelength of λ , and $[\text{HbO}](r_s, r_d; t)$ and $[\text{HbR}](r_s, r_d; t)$ are the concentration changes of HbO and HbR, respectively. By measuring the optical density changes at two wavelengths λ_1 and λ_2 , the concentration changes of HbO and HbR can then be determined by the following matrix formulation:

$$\begin{bmatrix} \Delta OD(r_s, r_d; \lambda_1, t) \\ \Delta OD(r_s, r_d; \lambda_2, t) \end{bmatrix} = \frac{1}{BL} \begin{bmatrix} \epsilon_{\text{HbO}}(\lambda_1) & \epsilon_{\text{HbR}}(\lambda_1) \\ \epsilon_{\text{HbO}}(\lambda_2) & \epsilon_{\text{HbR}}(\lambda_2) \end{bmatrix}^{-1} \begin{bmatrix} \Delta[\text{HbO}](r_s, r_d; t) \\ \Delta[\text{HbR}](r_s, r_d; t) \end{bmatrix}. \quad (3)$$

Tomographic mapping

Although the modified Beer–Lambert law (Cope and Delpy, 1988) has been used extensively in fNIRS studies, this is a first-order approximation of diffuse light scattering. As a more rigorous approach, the optical density changes in a scattering medium can be derived from the photon diffusion equation using the Ryto approximation (Arridge, 1999; Boas et al., 2002, 2004; Ye et al., 2009):

$$\Delta OD(r_s, r_d; \lambda, t) = -\ln \frac{\Phi(r_s, r_d; \lambda, t)}{\Phi_0(r_s, r_d; \lambda, t)} \approx \int \frac{\Phi_0(r_s, r; \lambda) \Phi_0(r, r_d; \lambda)}{\Phi_0(r_s, r_d; \lambda)} \Delta \mu_a(r; \lambda, t) dr. \quad (4)$$

Using Eq. (2), for given wavelength λ_1 and λ_2 , Eq. (4) can be represented by

$$\begin{bmatrix} \Delta OD(r_s, r_d; \lambda_1, t) \\ \Delta OD(r_s, r_d; \lambda_2, t) \end{bmatrix} \approx \mathcal{A} \cdot \begin{bmatrix} \Delta[\text{HbO}](r; t) \\ \Delta[\text{HbR}](r; t) \end{bmatrix}, \quad (5)$$

where the operator \mathcal{A} is given by

$$\mathcal{A} \equiv \begin{bmatrix} \epsilon_{\text{HbO}}(\lambda_1) \int dr \frac{\Phi_0(r, r_d; \lambda_1) \Phi_0(r_s, r; \lambda_1)}{\Phi_0(r_s, r_d; \lambda_1)} & \epsilon_{\text{HbR}}(\lambda_1) \int dr \frac{\Phi_0(r, r_d; \lambda_1) \Phi_0(r_s, r; \lambda_1)}{\Phi_0(r_s, r_d; \lambda_1)} \\ \epsilon_{\text{HbO}}(\lambda_2) \int dr \frac{\Phi_0(r, r_d; \lambda_2) \Phi_0(r_s, r; \lambda_2)}{\Phi_0(r_s, r_d; \lambda_2)} & \epsilon_{\text{HbR}}(\lambda_2) \int dr \frac{\Phi_0(r, r_d; \lambda_2) \Phi_0(r_s, r; \lambda_2)}{\Phi_0(r_s, r_d; \lambda_2)} \end{bmatrix}.$$

The concentration changes of HbO and HbR at voxel location r can then be determined by

$$\begin{bmatrix} \Delta[\text{HbO}](r; t) \\ \Delta[\text{HbR}](r; t) \end{bmatrix} = \mathcal{A}^{-1} \cdot \begin{bmatrix} \Delta OD(r_s, r_d; \lambda_1, t) \\ \Delta OD(r_s, r_d; \lambda_2, t) \end{bmatrix}. \quad (6)$$

The inverse problem is highly nonlinear, so many investigators have proposed a variety of regularization methods (Boas et al., 2001, 2004; Culver et al., 2003a, 2003b; Custo et al., 2010; Lee et al., 2011; Pogue et al., 1999; Ye et al., 1999, 2001). In particular, if we confine the reconstruction volume to the brain surface, then the inverse mapping \mathcal{A}^{-1} provides an interpolation kernel for topographic mapping, as described in Ye et al. (2009). In both cases, the location of the channels and the optical properties of the brain are irregular, so the resulting interpolation kernel is not spatially invariant, providing inhomogeneous noise statistics (Ye et al., 2009).

Signal processing of fNIRS signals

Considering the back-reflection geometry of fNIRS measurements, the fNIRS signal reflects the hemoglobin oxygenation changes in the gray matter as well as the physiology-based systemic confounds in the superficial layers of the head, including the skin and the skull (Gagnon et al., 2012). These confounds need to be regressed out before the statistical analysis of fNIRS data, so this section reviews advanced signal processing techniques to deal with this issue.

Motion artifact correction

Although fNIRS is less susceptible to motion artifacts than other neuroimaging modalities (Lloyd-Fox et al., 2010), head movement will cause a shift in the optical coupling between the fiber and the skin. These motion artifacts typically result in rapid changes (such as sharp spikes), and increased magnitude in the measured fNIRS signal, compared with tissue hemodynamics-related changes. Moreover, the motion artifact may shift the baseline values of fNIRS measurements (Cooper et al., 2012; Robertson et al., 2010).

Numerous approaches have been developed for reducing the motion artifacts from the fNIRS measurements, and the motion correction methods can be divided into two main categories: (i) approaches which use an additional sensor to detect the motion artifact and (ii) signal processing methods to identify and remove the motion artifact from the fNIRS data without an external measure of motion. In the first category, Blasi et al. (2010) used an accelerometer to measure the signal highly sensitive to motion, and identified the motion event based on the standard deviation. Motion artifacts were then removed from the fNIRS data using the recursive least squares (RLS) adaptive filter. Virtanen et al. (2011) estimated the baseline motion artifacts using an accelerometer and corrected those artifacts by scaling the amplitude of fNIRS signal. Robertson et al. (2010) used co-located channels (source–detector pairs on the same location) to identify the motion artifacts in the fNIRS signals, assuming that the detector on the same source mostly measure motion-related signal, and may not reflect the hemodynamic response. In the removal of motion artifacts, several methods were then compared, including two-input RLS adaptive filtering, wavelet-based filtering, independent component analysis (ICA), and multiple channel linear regression. They suggested that multiple-channel regression and ICA are the most promising methods to reduce motion artifacts in fNIRS on the basis of increased signal-to-noise ratio (SNR). We will describe the ICA-based motion correction method in more detail.

ICA (Comon, 1994) is a technique that decomposes a linear mixture of signals Y into a set of source signals S , which are maximally statistically independent: $Y = AS$, where mixing matrix A specifies the contribution of source S to measured signal Y . Source S is then estimated by using a unmixing matrix W such that $\hat{S} = WY$, where $W = A^{-1}$ is estimated by several methods, including maximum likelihood estimation (Beckmann and Smith, 2004). In the ICA-based motion correction method (Robertson et al., 2010), the motion-specific component q in source matrix S was identified and removed if its contribution to the co-located fNIRS channels was greater than the mean contribution of q to all channels.

The second category comprises methods for motion artifact removal without requiring additional sensors, and the methods are based on Wiener filtering (Izzetoglu et al., 2005), Kalman filtering (Izzetoglu et al., 2010), correlation based signal improvement (Cui et al., 2010), wavelet-based filtering (Molavi and Dumont, 2012; Sato et al., 2006), and moving standard deviation with spline interpolation (Scholkmann et al., 2010). More specifically, Cui et al. (2010) found that head movement causes the correlation between oxy-hemoglobin and deoxy-hemoglobin to become more positive, although both signals are generally negatively correlated. By maximizing the negative correlation between HbO and HbR, spikes were removed and contrast-to-noise ratio was significantly improved. Wavelet-based filtering effectively isolated motion artifacts in the form of abrupt signal changes (e.g., sharp spikes) from the fNIRS data, because the wavelet transform provides good localization of rapid signal changes (Molavi and Dumont, 2012; Sato et al., 2006). Scholkmann et al. (2010) segmented fNIRS signals into motion artifacts and hemodynamic-related signal by calculating the moving standard deviation, and subtracted the spline interpolation of motion-related segments from the fNIRS signals. By using the simulated hemodynamic response function (HRF), Cooper et al. (2012) compared the performance of the four motion correction methods, including Kalman filtering (Izzetoglu et al., 2010), wavelet-based filtering (Molavi and Dumont, 2012), spline interpolation (Scholkmann et al., 2010), and principal component analysis (Wilcox et al., 2008, 2010; Zhang et al., 2005). They showed that spline interpolation produces the largest reduction in mean-squared error (55%) of the recovered HRF and wavelet-based filtering produces the highest increase in contrast-to-noise ratio (39%), compared with no motion correction results.

Short-distance correction

fNIRS signals can be significantly contaminated with global interference arising from superficial layers of the scalp, as the back-reflection measurement of fNIRS makes it more sensitive to the superficial layers (Gagnon et al., 2011). Short source–detector separation channels is most sensitive to the superficial layers, and less reflects the neuronal activity-related signals. Exploiting the short source–detector separation channels, numerous noise removal approaches have been developed (Gagnon et al., 2011, 2012, 2013; Saager and Berger, 2005, 2008; Saager et al., 2011; Umeyama and Yamada, 2009; Yamada et al., 2009; Zhang et al., 2007, 2009). Umeyama and Yamada (2009) and Yamada et al. (2009) corrected the scattering and absorption changes arising from superficial layers using relatively short source–detector channels (2 cm) and Monte Carlo simulations of layered media. In addition, to remove the superficial contribution from the fNIRS signals measured by long source–detector separation channels (generally 3 cm), the weighting factors for the signal of short source–detector separation channels in the linear model were estimated using least-squares minimization (Saager and Berger, 2005, 2008) and adaptive filtering techniques (Zhang et al., 2007, 2009). For visual stimulation responses, significant improvement of contrast-to-noise ratio and signal-to-noise ratio in HbO signals was achieved after least-squares minimization-based noise removal (Saager et al., 2011) and adaptive filtering

(Zhang et al., 2009). In more recent studies by Gagnon et al. (2011, 2012, 2013), state-space modeling with Kalman filter estimation was developed and optimal locations of short source–detector channels were proposed for recovering the hemodynamic responses with significant reduction of superficial contamination. We will describe the state-space modeling with Kalman filter estimation in more detail.

The fNIRS signal in the long source–detector separation channel y_{LS} was modeled by a linear combination of the signal in the short source–detector channel y_{SS} and the neuronal activity-related response y_b (Gagnon et al., 2011):

$$y_{LS}[n] = y_b[n] + a_{SS}y_{SS}, \quad (7)$$

where

$$y_b[n] = \sum_{k=-\infty}^{\infty} h[k]\mu[n-k], \quad h[n] = \sum_{i=1}^{N_w} w_i b_i[n]. \quad (8)$$

Here, a_{SS} denotes the weights for the superficial contribution, $h[n]$ denote the hemodynamic response, $h[n]$ was modeled by a linear combination of normalized Gaussian functions $b_i[n]$, N_w is the number of Gaussian functions, and w_i is the weights for the Gaussian functions. The weights used to model the hemodynamic response w_i and the weight used to model the superficial signal a_{SS} were simultaneously estimated by using the Kalman filter followed by the Rauch–Tung–Striebel smoother. The hemodynamic response $\hat{h}[n]$ with reduction of superficial contamination was then reconstructed using estimated weights w_i and temporal basis function $b_i[n]$ in Eq. (8). These superficial contamination removal approaches, based on short source–detector separation channel and Kalman filtering, significantly improved the recovery of HbO and HbR, compared with other approaches (such as adaptive filtering, least-squares minimum noise removal, and the conventional GLM technique). Gagnon et al. (2012) suggested that the distance between short separation channels and the standard fNIRS channels should be less than 1.5 cm, due to inhomogeneities of superficial contamination across the surface of scalp. In addition, Gagnon et al. (2013) showed further improvement of the performance of Kalman filtering in reduction of superficial contamination, when using two short separation measurements, one located close to the source and one located close to the detector (noise reduction of 59% and 47% for HbO and HbR, compared with the conventional GLM method without short separation measurement).

PCA and ICA for skin blood flow removal

Data-driven approaches, including principal component analysis (PCA) (Virtanen et al., 2009; Wilcox et al., 2008, 2010; Zhang et al., 2005) and temporal ICA (Akgül et al., 2006; Katura et al., 2008; Kohno et al., 2007; Markham et al., 2009; Patel et al., 2011), have been applied to the fNIRS data, for separating the component of hemodynamic response from unwanted sources. PCA and ICA methods use an assumption of orthogonality and statistical independence between components, to decompose the signals, respectively.

Zhang et al. (2005) applied PCA to the baseline of fNIRS signal, for identifying the dominant spatial eigenvectors of systemic interference. They assumed that the baseline signal primarily contains spatial patterns of global interference, and the corresponding spatial component is more global with higher energy. The spatial filtering was then performed by projecting the stimulus fNIRS data onto the orthogonal subspace of identified spatial eigenvectors. The results showed more localized fNIRS response with improved contrast-to-noise ratio. Kohno et al. (2007) used temporal ICA to remove the global physiological noise from fNIRS data. Comparing the mixing matrix which specifies the contribution of ICA components to measured fNIRS signal, they identified the most spatially uniform component, and confirmed that the corresponding component signal is significantly correlated with skin blood

flow measured with the laser Doppler tissue blood flow meter. Without noise components, a more localized activation map was reconstructed. Temporal ICA was also used to extract the cleaned hemodynamic response component by comparing the correlation with the canonical hemodynamic time courses (Akgül et al., 2006; Markham et al., 2009) and reproducibility of the component over repeated trials (Katura et al., 2008; Patel et al., 2011).

Removal of baseline fluctuation

A wavelet-based detrending algorithm was used to decompose fNIRS measurements into global trends, hemodynamic signals, and uncorrelated noise at distinct scales (Jang et al., 2009; Lina et al., 2010). The global trend was modeled as a signal restricted to a subspace spanned by coarse scale wavelets, and its contribution was determined by the maximum likelihood estimates of the GLM framework.

More specifically, the global trend is described using a wavelet basis, and the wavelet coefficients of global trends and the signal strength of predicted hemodynamic response were estimated using the maximum likelihood estimation within a GLM framework. In order to determine the number of wavelet coefficients that optimally describes the global trends, the minimum description length (MDL) principle was employed. By minimizing the MDL cost function, which consists of goodness of fit and concision of the model, we can estimate the optimal number of wavelet coefficients.

The wavelet detrending method has several advantages over the conventional high-pass filtering approaches. As the frequencies of global trends and hemodynamic responses can be overlapped, a non-negligible amount of the hemodynamic signal can be filtered out. However, the wavelet-based detrending method includes the predicted hemodynamic response in the least-square estimation process. Hence, the canonical hemodynamic time series is fully utilized in the estimation process, and the algorithm is more robust. In addition, the wavelet-based approach is more effective in removing relatively fast-varying trends due to the optimality of the wavelet transform in describing the transient changes of fast varying signals.

Partitioning of physiological noise signals

Rather than just removing these confounds as artifacts, Tong et al. (Frederick et al., 2012; Tong and Frederick, 2010; Tong et al., 2011b) conducted a series of innovative experiments to utilize these confounds. In Tong and Frederick (2010), to elucidate the physiological origin of low-frequency oscillations of BOLD responses, they employed simultaneous recording of fNIRS and fMRI. Specifically, using the low frequency oscillations of fNIRS signals measured over the right prefrontal area as a regressor, they applied the GLM analysis to the BOLD signal. Experimental results showed that the contribution of low-frequency oscillations of fNIRS to BOLD signal temporally follows the circulation of blood through the brain, which suggests that one of the major sources of low-frequency oscillations might be the fluctuations in blood and hemoglobin oxygenation at a global circulatory system level. In Frederick et al. (2012), to investigate the contribution of various physiological sources to the BOLD response, they proposed a method to extract a time-series of low-frequency oscillation, respiration, and cardiac pulsation from the fNIRS signals, and then visualize its spatio-temporal activation pattern. Experimental results suggest that the highly correlated areas between the fNIRS and BOLD signals in the low-frequency oscillation band overlap the regions of the default mode network, which contains neuronal activity. However, the dynamic evolution of these activations indicates that at least some part of the low frequency oscillations detected by both modalities is closely correlated with cerebral blood circulation, not with neuronal activity. In addition, a highly aliased heartbeat signal existed in the BOLD signal at voxels located mainly in and around the arteries and large veins (see Fig. 1).

Statistical inferences for fNIRS signals

After the aforementioned signal processing step, statistical inference are usually performed to detect brain activation.

In the early ages of fNIRS studies such as in Murata et al. (2002), the authors simply calculated the concentration changes of hemoglobin oxygenation during the task period and showed the time-series of cerebral

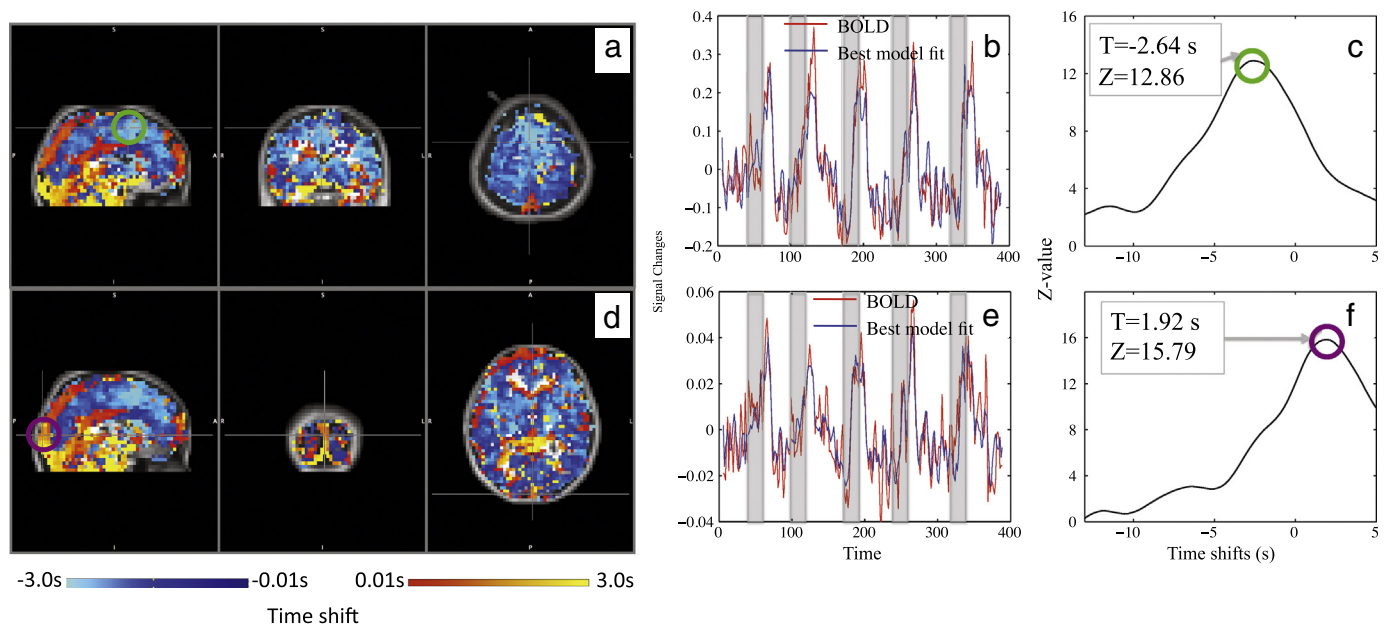


Fig. 1. Optimum time lag map between BOLD response and fNIRS regressor are shown in (a) and (d). The BOLD signal and time-shifted regressor on a voxel of (b) green circle area and (e) purple circle area are described, and the z-values with respect to time shift of fNIRS regressors are plotted on (c) and (f), respectively. The time lag map was calculated as follows: (1) GLM analysis of BOLD response was performed, using NIRS signals shifted in a range of -14.4 to 7.2 s as a regressor, and (2) the time index of the maximum z value for each voxel was calculated from the concatenated z-statistic maps. The value of time lag map at each voxel represents relative arrival times of the systemic hemodynamic response between that voxel location and the site where the fNIRS data were recorded in the prefrontal area.

Figure courtesy of Tong et al. (2011a).

oxygenation changes for visual inspection. Benaron et al. (2000) detected increases in hemoglobin oxygenation ($\text{HbO} / (\text{HbR} + \text{HbO})$) and hemoglobin volume in the motor cortex during a finger tapping task of a healthy subject after subtracting a resting-state signal. Then, increases in hemoglobin oxygenation greater than 2 SD (standard deviation) from the mean oxygenation baseline were selected as an indicator of local activation. However, such heuristic approaches are prone to error, especially when the noise and interference level increases. Hence, various statistical analysis methods have been applied. For example, simple statistics such as the t -test can be used to test the difference of means in fNIRS experiments (Germon et al., 1994, 1999; Hoshi et al., 2003; Matsuo et al., 2003; Schroeter et al., 2002). In addition, a paired t -test was performed to statistically compare the activation level between different conditions (Aldrich et al., 1994; Hoshi et al., 2001; Isobe et al., 2001; Kennan et al., 2002; Kim et al., 2009; Shibuya-tayoshi et al., 2007; Tachtsidis et al., 2004; Tsujimoto et al., 2004; Young et al., 2000).

Note that a two sample t -test corresponds to one-way ANOVA that compares the condition differences with respect to condition variance. A more general class of statistical tests using multi-way ANOVA has also been employed for fNIRS studies (Arenth et al., 2007; Bartocci et al., 2000; Fallgatter and Strik, 1998, 2000; Folley and Park, 2005; Germon et al., 1994; Herrmann et al., 2003, 2004; Hoshi, 2003; Irani et al., 2007; Kameyama et al., 2006; Kleinschmidt et al., 1996; Matsuo et al., 2002; Mehagnoul-Schipper et al., 2002; Okamoto et al., 2004; Saitou et al., 2000; Suto et al., 2004; Taga et al., 2003; Villringer et al., 1998, 1993). These approaches often take average values during the task period as data to avoid any assumption of the exact shape or timing of the time course of changes in HbO and HbR in response to the stimuli.

Recall that the time course in fNIRS is directly related to the hemodynamic response, but in the aforementioned simple statistics, such information is lost. Therefore, many investigators were interested in understanding the fNIRS time course, for example, to find the relationship between the fMRI and fNIRS time course data. To deal with this, correlation analysis has been employed (Huppert et al., 2006; Siegel et al., 2003; Strangman et al., 2002).

While these correlation analyses were successful in revealing the correlation between fMRI and fNIRS time series, it is often difficult to explain magnitude differences and task-unrelated variations in the fNIRS stand-alone test. Fourier analysis is another simple way of comparing time course data (Schroeter et al., 2004; White et al., 2009).

This analysis assumes that any task should have maximum power at the task frequency. So if we find a peak at this frequency, we can assert that this signal correlates with task. This concept is somewhat equivalent to correlation analysis in the frequency domain and can also be regarded as a kind of GLM if we use sine and cosine functions as regressors. For example, Schroeter et al. (2004) proposed spatially resolved spectral analysis to characterize the coherence and phase shift of the hemodynamic response. Similar spectral analysis techniques have been successfully employed for high resolution retinotopic mapping of the adult human visual cortex with high-density diffuse optical tomography (Zeff et al., 2007).

One of the main limitations of Fourier analysis is that it is often difficult to account for the time course with non-periodic tasks such as an event-related paradigm. Hence, researchers were interested in more advanced analysis methods that could overcome aforementioned weaknesses. The GLM (Friston et al., 2011) assumes that data can be represented as a linear combination of several sources (regressors). Schroeter et al. (2004) were the first to apply the general linear model to analyze the fNIRS data, and numerous authors have employed GLM analysis for a variety of fNIRS experiments (Abdelnour and Huppert, 2009; Custo et al., 2010; Koh et al., 2007; Minagawa-Kawai et al., 2011; Plichta et al.,

2006, 2007; Shimada and Hiraki, 2006; Singh and Dan, 2006; Ye et al., 2009).

Inspired by the success of GLM as well as the standard sample-wise statistical testing, Huppert et al. (2009) introduced the first public domain fNIRS analysis package, called HomER. HomER is a MATLAB-based graphical user interface program for dealing with fNIRS time-series and functional analysis techniques using the aforementioned GLM and standard statistics. Moreover, HomER has various signal processing techniques such as subspace-based filtering for removing physiological, instrumental, and motion-artifact noise from optical data, Koh et al. (2007) introduced another software package, fOSA (functional Optical Signal Analysis), which incorporates the statistical parametric mapping (SPM) method for the analysis of fNIRS data. Several features for physiological removal and statistical analysis were included. More specifically, to remove the noise, fOSA incorporated three types of digital filters (elliptic filter, Chebyshev filter, Butterworth filter). For making inference of brain activation, fOSA incorporated two different t -statistical approaches: channel-wise Student t -test and the SPM method.

While these are extensive statistical tools for fNIRS analysis, several fundamental issues still remain. For instance, even though stationary Gaussian random field theory has been used and justified in fOSA as a tool for family-wise error rate (FWER) control during inference, the basic assumption for the stationary Gaussian random field model breaks down in fNIRS. Recall that SPM for fMRI analysis assumes that the residuals after the GLM fitting are *dense* samples on lattice representations from an underlying continuous Gaussian random field (Friston et al., 1996) due to the Gaussian kernel smoothing. However, because the distance between each channel of fNIRS is far and the number of measurements is sparse, it is not reasonable to use stationary Gaussian random field theory in making inferences about fNIRS data. To address this issue, Ye et al. (2009) introduced a new public domain statistical toolbox called NIRS-SPM for the quantitative analysis of fNIRS signals. More specifically, they interpolated fNIRS channel measurements on a topographic surface and statistically analyzed the interpolated fNIRS data based on the GLM, proposing a novel theory for making inference using Sun's tube formula for excursion statistics.

Strangman (2009) recently introduced a software package, called NinPy, that provides a range of computational tools for an analysis of fNIRS data. NinPy was developed using the Python programming language and uniquely offers functions to control the stimulus presentation and data acquisition, compared with other non-commercial software packages such as HomER and NIRS-SPM. While those features are customized for the author's fNIRS instruments, other powerful functions in NinPy, including optical image reconstruction and hierarchical statistical analysis may contribute to the fNIRS community as one of reliable analysis tools.

In the following, we are interested in explaining these tools in more detail in a self-contained manner. In particular, we show that all these methods can be derived from a general mixed model with ReML covariance estimation and hypothesis testing.

Generalized linear mixed model (GLMM)

For a matrix A , $\text{Tr}(A)$ is the trace of a matrix A , A' is its adjoint, A^\dagger denotes the Penrose–Moore pseudo-inverse, and for a square matrix A , A^- denotes a generalized inverse. $|A|$ refers to the determinant, $R(A)$ denotes the range space of A , and P_A (or $P_{R(A)}$) and P_A^\perp (or $P_{R(A)}^\perp$) are the projection on the range space and its orthogonal complement, respectively.

We denote I_a and $\mathbf{1}_a$ as $a \times a$ identity matrix and a -dimensional vector with 1's, respectively; similarly, $\mathbf{0}_{a \times b}$ and $\mathbf{0}_a$ refer to $a \times b$ zero matrix and a -dimensional zero vector, respectively. The notation $E\{\cdot\}$ denotes an expectation for random variable or fields.

In statistics, a general linear mixed model (GLMM) extends a GLM by allowing a more flexible description of measurement. We use the following notation for GLMM:

- \mathbf{y} : an observation vector ($\in \mathbb{R}^N$)
- $\boldsymbol{\alpha}$: a fixed effect parameters ($\in \mathbb{R}^p$)
- X : a known fixed effect matrix ($\in \mathbb{R}^{N \times p}$),
- p^* : the degree of freedom for fixed effect, i.e. $\text{rank}(X) = p^* \leq p < N$,
- $\boldsymbol{\beta}$: a vector of q unobservable random effects ($\in \mathbb{R}^q$)
- Z : a random effect matrix ($\in \mathbb{R}^{N \times q}$)
- ϵ : an additive noise term ($\in \mathbb{R}^N$)
- R : covariance matrix for noise ϵ ($\in \mathbb{R}^{N \times N}$)
- G : covariance matrix for random effect parameter $\boldsymbol{\beta}$ ($\in \mathbb{R}^{q \times q}$)
- V : covariance matrix for combined random effect and noise
- Ω : empirical Bayesian measurement covariance matrix
- L : contrast matrix ($\in \mathbb{R}^{p_1 \times p}$)
- S : test statistics.

Then, a general linear mixed model is represented as (SAS Institute, 1985; Searle, 1979)

$$\mathbf{y} = X\boldsymbol{\alpha} + Z\boldsymbol{\beta} + \epsilon. \quad (9)$$

The GLM is a special case of the general linear mixed effect model where the random effect matrix $Z = 0$. Note that we often consider a rank deficient fixed effect design matrix such that $\text{rank}(X) = p^* < p$ (in particular, for ANOVA analysis).

In a general linear mixed effect model, residual error terms and random effect parameters are assumed to have normal distribution:

$$\epsilon \sim N(\mathbf{0}, R), \quad \boldsymbol{\beta} \sim N(\mathbf{0}, G). \quad (10)$$

(Note that the difference between a *generalized* linear model and general linear model is that a generalized linear model extends GLM by allowing non-Gaussian noise (SAS Institute, 1985)). Our goal is then to estimate the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, R and G from the measurement \mathbf{y} . The estimates for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, G , and R are denoted by $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, \hat{G} and \hat{R} , respectively.

The maximum likelihood (ML) principle is to obtain parameter estimates as values that yield a large value for the likelihood of the observed sampled for the chosen probability density functions, thereby maximizing the likelihood (Searle, 1979). Henderson (1973, 1982) obtained the estimates of the mixed model parameters by maximizing the likelihood of the joint distribution: which results in a closed form estimate for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (Henderson, 1973; Henderson, 1982; Robinson, 1991; SAS Institute, 1985; Searle, 1979):

$$\hat{\boldsymbol{\alpha}} = (X'V^{-1}X)^{-1}X'V^{-1}\mathbf{y} \quad (11)$$

$$\hat{\boldsymbol{\beta}} = GZ'V^{-1}(\mathbf{y} - X\hat{\boldsymbol{\alpha}}), \quad (12)$$

where V is given by

$$V = ZGZ' + R. \quad (13)$$

In practice, the covariances R and G are not known, so they need to be estimated. To address this issue, Patterson and Thompson (1971) proposed a method which takes into account the loss in degrees of freedom resulting from estimating fixed effects. More specifically, they restricted their attention to a set of such contrast that is invariant to the fixed effect parameter and then estimated the covariance for such *restricted* cases of maximum likelihood. This idea is the so-called *restricted maximum likelihood* (ReML) method (Friston et al., 2011; Graser et al., 1987; Harville, 1977; Kenward and Roger, 1997; Searle, 1979). More specifically, if we are interested in finding the restricted maximum likelihood that is invariant to the fixed effect parameter, we need to find a full rank matrix $K \in \mathbb{R}^{N \times (N-p)}$ such that $K'X = 0$. Then, we deal with maximum likelihood for a new variable $K'y$ in the ReML

method, which can be equivalently represented by

$$L_{\text{ReML}}(R, G) = -\frac{1}{2} \log |V| - \frac{1}{2} \log |X'V^{-1}X| - \frac{1}{2} \mathbf{y}'\Omega\mathbf{y}, \quad (14)$$

where V is defined in Eq. (13) and Ω is defined as:

$$\Omega = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}. \quad (15)$$

In a linear mixed effect model, we are often interested in an inference concerning the fixed effect by considering the following linear combinations of the fixed effect parameter:

$$\hat{\mathbf{v}} = L(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}), \quad (16)$$

where $L \in \mathbb{R}^{p_1 \times p}$ denotes the *contrast* matrix (SAS Institute, 1985; Searle, 1979). One of the often discussed issues regarding the contrast matrix is that not every contrast can be used. More specifically, if X is rank deficient, non-trivial null space of X exists. Therefore, for some contrast L , there are chances that $L\boldsymbol{\alpha}$ lies in the null space of X . However, if $L\hat{\boldsymbol{\alpha}}$ is a linear functions of $\hat{\boldsymbol{\alpha}}$ such that $R(L') \subset R(X')$, implying that $L' = X'A$ for some matrix A , then we can guarantee that the function is not in the null space, so it is an *estimable* function (Rao and Toutenburg, 1999). Furthermore, for such contrast L , we have $L\hat{\boldsymbol{\alpha}} = L(X'V^{-1}X)^{-1}X'V^{-1}\mathbf{y}$, and the corresponding error covariance is given by (Rao and Toutenburg (1999))

$$C_V = E\{(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})'\} = L(X'V^{-1}X)^{-1}L' \in \mathbb{R}^{p_1 \times p_1}.$$

Accordingly, we have the following test statistic (SAS Institute, 1985):

$$\hat{\mathbf{z}} = C_V^{-1}\hat{\mathbf{v}} \sim \mathcal{N}(\mathbf{0}, I_{p_1 \times p_1}). \quad (17)$$

In practice, the true covariance R and Z are unknown and need to be estimated using ML or ReML. In this case, the test statistic becomes

$$S = \begin{cases} \frac{L(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})}{\sqrt{L(X'\hat{V}^{-1}X)^{-1}L'}}, & p_1 = 1 \\ \frac{(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})'L'(L(X'\hat{V}^{-1}X)^{-1}L')^{-1}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})}{p_1}, & p_1 > 1, \end{cases} \quad (18)$$

where \hat{V} denotes the empirical covariance matrix. In this case, the statistics become t_{ν} and $F_{p_1, \nu}$ -statistics, respectively, where ν denotes the degree of freedom for the denominator (SAS Institute, 1985).

The degree of freedom can be calculated in various ways. For example, if the noise is spherically distributed, then we can use the residual degree of freedom, i.e. $\nu = N - p^*$. However, when the residual is not spherical, the denominator is not exactly χ^2 distributed, so one often uses an approximate (effective) degree of freedom (DOF). In SPM, Friston et al. (2011) proposed the following Satterthwaite approximation.

However, there are some concerns that the Satterthwaite correction may not correct for fNIRS, because it still grossly overestimates the degree of freedom in the presence of serially correlated errors which are very high in fNIRS data due to the high over-sampling rates. The issue of correct DOF estimation for fNIRS is still an open problem, and needs to be investigated further. In statistics, the degree of freedom is very important since it determines the sensitivity of the test. For example, as shown in Fig. 2, the t_{ν} statistics with lower degree of freedom has heavier tails, so in hypothesis testing, it provides higher threshold values for the same statistical significance. This implies that the detection of activation becomes more conservative.

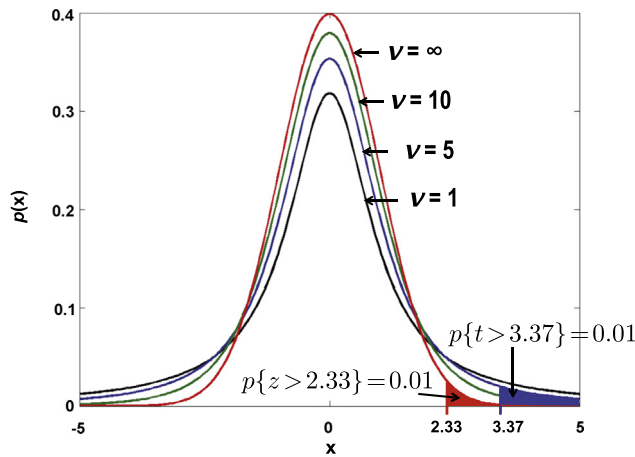


Fig. 2. t_ν statistics with various degrees of freedom ν . Note that a t -statistics with a large degree of freedom results in a lower threshold value.

In fNIRS analysis, the degree of freedom can increase by various means. For example, using longer temporal time trace, or group data, we can increase the degree of freedom. On the other hand, if the temporal time traces are highly correlated or within subject variation needs to be considered as in repeated measure ANOVA, the degree of freedom decreases (Friston et al., 2011).

Derivation of classical statistical analysis for fNIRS

In classical statistics, such as the one sample t -test, two sample t -test, paired t -test, and ANOVA, which have been extensively used for fNIRS analysis, the time course data are segmented into different condition periods, such as baseline and activations, and the averaged values during this period are used as data for standard statistical analysis.

ANOVA

Extensive applications of ANOVA (Scheffe, 1959) exist for fNIRS data analysis. A partial list of studies that employ ANOVA includes Mehagnoul-Schipper et al. (2002), Okamoto et al. (2004), Kleinschmidt et al. (1996), Suto et al., 2004, Saitou et al. (2000), Irani et al. (2007), Hoshi (2003), Arenth et al. (2007), Villringer et al. (1993), Kleinschmidt et al. (1996), Germon et al. (1994), Suto et al. (2004), Fallgatter and Strik (1998), Herrmann et al. (2003), Bartocci et al. (2000), Fallgatter and Strik (2000), Herrmann et al. (2004), Kameyama et al. (2006), Matsuo et al. (2002), Folley and Park (2005) and Taga et al. (2003).

To show how a specific fNIRS data analysis can be mapped into this framework, we use the study by Fallgatter and Strik (1998) as an example, where 2×2 ANOVAs for repeated measurements [hemispheres (left vs right) \times segments (activation vs baseline)] were computed for the parameters HbR and HbO (relative concentrations) during the Wisconsin Card Sorting Test (WCST). The analyses consider spatial location (hemisphere, electrode number) to be a factor in a multi-factorial analysis of variance, and one looks for interactions of conditions with locations as well as main effects. Here, a two channel measurement system was used to measure left and right hemisphere fNIRS data from 10 subjects. For quantification, the average values of HbR and HbO for prestimulus baseline (−25 to −1 s) and WCST activation (0–216 s) were computed separately for the left and right hemispheres.

More specifically, the two-way between-subject ANOVA is described by

$$y_{ijl} = \mu_{ij} + e_{ijl}, \quad i = 1, \dots, a; \quad j = 1, \dots, b; \quad l = 1, \dots, n_{ij},$$

where y_{ijl} denotes the l th data sample at level i of factor A and level j of factor B , μ_{ij} is the average for observations at level i of factor A and level

j of factor B , and n_{ij} is the number of observations at that combination of levels. The saturated model is then given by

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij},$$

where $(\alpha\beta)_{ij}$ denotes a two-way interaction. Here, the notation (\cdot) is a single symbol for an interaction, not as the product of multiple parameters within the parenthesis (Christensen, 2011).

This model can be mapped to the problem in Fallgatter and Strik (1998) as follows:

- y_{11l} : l th subject left hemisphere average value for prestimulus baseline
- y_{12l} : l th subject left hemisphere average value for WCST activation
- y_{21l} : l th subject right hemisphere average value for prestimulus baseline
- y_{22l} : l th subject right hemisphere average value for WCST activation
- α_i : main effect from hemispheres (left vs. right)
- β_j : main effect from segments (baseline vs. activation)
- $(\alpha\beta)_{ij}$: interaction between hemispheres and segments
- n_{ij} : the number of subjects ($n_{ij} = n = 10$).

For a balanced case, where $n_{ij} = n$ for all i, j, k and $\sum n_{ij} = N$, two-way ANOVA can be represented in the following GLM form (Christensen, 2011):

$$y = X\alpha + \epsilon$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_N)$ and

$$\mathbf{y} = [y_{111}, y_{112}, \dots, y_{11N}, y_{121}, \dots, y_{12N}, \dots, y_{ab1}, \dots, y_{abN}]' \in \mathbb{R}^N$$

$$\alpha = [\mu, \alpha_1, \dots, \alpha_a, \beta_1, \dots, \beta_b, (\alpha\beta)_{11}, \dots, (\alpha\beta)_{ab}]'$$

and a design matrix

$$X = [\mathbf{1}_N, X_A, X_B, X_C, X_{A \times B}], \quad (19)$$

where it can be represented as sub-design matrices:

$$\mathbf{1}_N = \mathbf{1}_a \otimes \mathbf{1}_b \otimes \mathbf{1}_n$$

$$X_A = I_a \otimes \mathbf{1}_b \otimes \mathbf{1}_n$$

$$X_B = \mathbf{1}_a \otimes I_b \otimes \mathbf{1}_n$$

$$X_{A \times B} = I_a \otimes I_b \otimes \mathbf{1}_n.$$

Now, to test the main or interaction effect, we need to design contrast matrix L . To make the test function estimable, we need to design L by considering the nesting relationship. Indeed, we do not need to calculate the explicit contrast matrix L since the corresponding test statistic is equivalent to the standard F -statistic to test main effect of A in ANOVA:

$$S = \frac{\hat{\alpha}' L' (L' X' X L)^{-1} L \hat{\alpha}}{\sigma^2 \text{rank}(L)}$$

$$= \frac{y' (P_{X_0}^\perp - P_X^\perp) y}{y' P_X^\perp y} \frac{N - ab}{a - 1} \quad (20)$$

$$= \frac{SSA / (a - 1)}{SSE / (N - ab)}$$

$$\sim F_{a-1, N-ab},$$

where SSA denotes the factor A sum of squares and SSE is the error sum of squares (Scheffe, 1959), and X_0 denotes the reduced model by excluding the factor A main effect. Similarly, we can test other main and interaction effects.

Note that the mean values from each subject can be different during baseline and WCST activations. Hence, the test should be done using two-way ANOVA with repeated measurement. Two-way within-subject ANOVA or two-way repeated measurement ANOVA is different from two-way between subject ANOVA by modeling the

mean as

$$\mu_{ijn} = \mu_n + \alpha_i + \beta_j + (\alpha\beta)_{ij},$$

where we assume that the mean value μ_n can be different for each measurement. In this case, from the design matrix of X in Eq. (19) in the two-way between-subject ANOVA, the first column should be replaced by $\mathbf{1}_{ab} \otimes I_n$ to account for between-subject mean variation. Then, our test statistic becomes $F_{a-1, N-ab-n}$ due to the reduction of the degrees of freedom to $N - ab - n$. This implies that, for the same p -value, the threshold value become higher than between-subject ANOVA, which makes the hypothesis testing more conservative (Christensen, 2011; Friston et al., 2011).

For fNIRS studies, two-way ANOVAs with repeated measurement have been used for various studies. In Herrmann et al. (2003), frontal activation during a verbal-fluency task was measured by near-infrared spectroscopy. The data collection was similar to Fallgatter and Strik (1998), and 2×4 (hemisphere \times segment) ANOVA for repeated measurements was calculated for the different variables (HbO, HbR), where segments consist of baseline and activations from “A,” “F,” and “S” characters. In Bartocci et al. (2000), activation of the olfactory cortex in newborn infants after odor stimulation was studied. Here, the mean values of each subject during the three different smell conditions (control, colostrum, and vanilla) were recorded. A two-way ANOVA for repeated measurements were used to compare the changes of HbO in response to the three different stimuli, which showed significant effects of the smell condition. In Fallgatter and Strik (2000), frontal functional asymmetry in schizophrenia during a cued continuous performance test were assessed using 2×3 repeated measures ANOVA (2 hemispheres \times 3 segments) of HbR and 2×3 ANOVA (2 groups \times 3 segments) of the normalized (\log_e) left/right raw HbR ratios to test group differences in relative hemisphere activation.

Basic t -test

A one-sample t -test can be used to test the null hypothesis that the mean of $N = n$ data is zero. Here, for given measurements $\{y_i\}_{i=1}^n$ such that $y_i = \mu + \epsilon_i$, we test the null hypothesis that the mean is zero, i.e. $H_0: \mu = 0$. Assuming that the noise is i.i.d. Gaussian with variance σ^2 , we can easily see that this is zero-way ANOVA by excluding all interactions and the main effects from the two-way ANOVA. Therefore, we have

$$X = \mathbf{1}_n, \quad L = 1.$$

A two-sample t -test allows one to test the null hypothesis that the means of two groups $\{y_i^A\}_{i=1}^n$ and $\{y_i^B\}_{i=1}^n$ are equal. We can easily see that this is in fact, one-way between subject ANOVA with a factor of $A = \text{group}$ with a level of $a = 2$. Therefore, we can see

$$X = [\mathbf{1}_2 \otimes \mathbf{1}_n, I_2 \otimes \mathbf{1}_n], \quad L = [0, 1, -1].$$

Finally, a paired t -test is an extension of the two-sample t -test model by assuming that the means over pairs are not equal. More specifically, consider a two group where the measurement y_i^A, y_i^B comes in pairs. By inspection, we can see that this is indeed a one-way within-subject ANOVA, so the corresponding parameters are

$$X = [\mathbf{1}_2 \otimes I_n, I_2 \otimes \mathbf{1}_n], \quad L = [\mathbf{0}_n', 1, -1].$$

While the resulting test statistics for a two-sample t -test and a paired t -test appear identical, the main difference between the two comes from the loss of degrees of freedom in the case of the paired t -test. This is because the paired t -test takes into account different means over pairs, whereas the two-sample t -test assumes that the group mean is the same. This makes the paired t -test more conservative.

The basic t -test has been widely used in other fNIRS studies to detect activation or compare a condition against the control. For instance, Hoshi et al. (2003) investigated changes of oxy-hemoglobin concentration in the prefrontal cortex during n -back and random number generation tasks. Here, the channel-wise student t -test was employed to statistically compare two different states of the hemoglobin concentration changes (task vs. control).

In Tsujimoto et al. (2004), to examine the role of the lateral prefrontal cortex of preschool children in working memory, an item-recognition task with different memory-loads was performed on adult and child subjects. The Student's t -test was employed for finding the significantly activated channel during the task period. In addition, a paired t -test was performed to statistically compare the activation level between different memory-loads.

GLM-based fNIRS time course analysis

In time GLM-based fNIRS time course analysis, tests are repeated at each spatial location and one plots a map of e.g. condition effects.

In Schroeter et al. (2004), the design matrix was generated with a boxcar function convolved with a Gaussian kernel. Temporal correlation and effective degrees of freedom were estimated by the method in Worsley and Friston (1995). To evaluate the methods, they performed two visual tasks and hypothesized that stimulation with a checkerboard activates the visual cortex (V1–V3), whereas stimulation with rotating ‘L’s additionally involves the motion area V5. Analysis with the general linear model showed that the activated area was localized on the visual cortex for the first paradigm and was a nearby region to the motion area for the second paradigm. Hofmann et al. (2008) applied a general linear model to statistically analyze fNIRS data to investigate hemoglobin oxygenation changes during lexical decision making on words and pseudowords. In a design matrix, the predicted response was generated by convolving a Gaussian function with each event, and its first and second derivative terms were included. To estimate the autocorrelation matrix, a first-order autoregressive process was performed. In Plichta et al. (2007), the applicability of the GLM analysis to a rapid event related-fNIRS design was tested.

SPM-type analysis using GLM was first proposed by Koh et al. (2007) for channel-wise analysis. In this paper, they introduced a software package, fOSA, that incorporates the SPM method for the analysis of fNIRS data. For making inference of brain activation, fOSA applied the SPM method to fNIRS signals as follows: 1. configure the fNIRS measured hemoglobin concentration changes as topographic images such that each channel can be represented as a pixel in the SPM map and 2. perform the SPM analysis that consists of the general linear model and random field theory on a two-dimensional patch. Canonical HRF and the inverse HRF were used for modeling the predicted response of HbO and HbR, respectively. Furthermore, the ReML was employed for estimating the error covariance matrix. These have been significantly generalized by Ye et al. (2009), Jang et al. (2009), Tak et al. (2011), Li et al. (2012) and Tak et al. (2010) for interpolated maps.

In most existing fNIRS literature, two-level analysis using summary statistics has been extensively used for group analysis. Okamoto et al. (2006) examined the role of the lateral prefrontal cortex (LPFC) in taste encoding using a two-level summary statistics approach for random effects analysis. Specifically, a random effect group analysis with a one-tailed t -test was performed for making inferences about the activated channel during the task period. In addition, a paired t -test was performed to compare the contrast of all subjects between the encoding and control conditions.

In Plichta et al. (2007), ANOVA was applied at the second level of GLM. The effects of two factors (parametric factor contrast with four levels and dichotomous factor region with two levels) were investigated.

One of the technical issues in fNIRS group analysis is that, unlike fMRI, where group analysis is usually done after all the individual

session data are aligned on a global template, only a few optodes exist and the global alignments of the channel positions between individual session are often difficult. Hence, rather than using inter-subject alignments of optodes, Ye et al. (2009) and Li et al. (2012) proposed a voxel based global alignment between the *interpolated* maps before facilitating group analysis. However, the group analysis on the interpolated map in Li et al. (2012) and Ye et al. (2009) was still incomplete, so we provide here a more comprehensive derivation of multi-level group analysis using linear mixed effect model and ReML. This part is novel and will be included in the next release of NIRS-SPM (<http://bisp.kaist.ac.kr/NIRS-SPM.html>).

Suppose there are n subjects in a group. Then, we have

$$\mathbf{y}_l(r) = D\boldsymbol{\alpha}_l(r) + \epsilon_l(r), \quad l = 1, \dots, n,$$

where $\mathbf{y}_l(r)$ denotes the l th subject HbO or HbR time series at r location. Note that r could be either optode number, sensor space location, or location on the cortical surface depending on whether the analysis is performed in channel wise, or interpolated maps using topographic or tomographic mapping. However, for topographic mapping with random field analysis for inference later as in NIRS-SPM, r should be interpreted as a location on the cortical surface.

Then, we have the following concatenated expression:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1(r) \\ \vdots \\ \mathbf{y}_n(r) \end{bmatrix} = \begin{bmatrix} D & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & D \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1(r) \\ \vdots \\ \boldsymbol{\alpha}_n(r) \end{bmatrix} + \boldsymbol{\epsilon}. \quad (21)$$

If individual variation in a group is modeled as a random effect, then the regression coefficients are modeled as follows:

$$\boldsymbol{\alpha}_l(r) = \boldsymbol{\mu}(r) + \boldsymbol{\beta}_l(r), \quad l = 1, \dots, n, \quad (22)$$

where $\boldsymbol{\mu}(r)$ denotes the group mean and $\boldsymbol{\beta}_l(r)$ denotes a between-subject random effect parameter, respectively. Our goal is then to test whether the group mean is statistically significant. In fNIRS, activated channels are usually non-zero mean, so this test corresponds to the group-level activation detection.

Note that this is basically a 1-sample t -test and can be expressed as

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1(r) \\ \vdots \\ \boldsymbol{\alpha}_n(r) \end{bmatrix} = [\mathbf{1}_n \otimes I_p] \boldsymbol{\mu}(r) + \boldsymbol{\beta}.$$

Then, by plugging in Eq. (22) to Eq. (21), we have the following mixed effect model:

$$\mathbf{y} = Z\mathbf{X}_G\boldsymbol{\alpha} + Z\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (23)$$

where $\boldsymbol{\beta}(r) = [\boldsymbol{\beta}_1(r)', \dots, \boldsymbol{\beta}_n(r)']'$ and

$$\boldsymbol{\alpha} = \boldsymbol{\mu}, \mathbf{X}_G = [\mathbf{1}_n \otimes I_p], Z = I_n \otimes D. \quad (24)$$

By inspection, we know $\mathbf{X} = Z\mathbf{X}_G$, $\mathbf{V} = Z\mathbf{G}Z' + \mathbf{R}$. In a group analysis, we usually assume that the residual errors are uncorrelated such that the corresponding covariance matrix is given by

$$\mathbf{R} = E\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\} = \begin{bmatrix} R_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & R_n \end{bmatrix}, \quad (25)$$

where R_j denotes the j th subject covariance matrix. Using the following matrix equalities that can be proven using matrix inversion lemma

$$Z'V^{-1} = \left((Z'R^{-1}Z)^{-1} + \mathbf{G} \right)^{-1} (Z'R^{-1}Z)^{-1} Z'R^{-1} \quad (26)$$

$$Z'V^{-1}Z = \left((Z'R^{-1}Z)^{-1} + \mathbf{G} \right)^{-1}, \quad (27)$$

we have

$$\mathbf{X}'V^{-1}\mathbf{X} = \mathbf{X}'_G\mathbf{V}_G^{-1}\mathbf{X}_G \quad (28)$$

$$\mathbf{X}'V^{-1} = \mathbf{X}'_G\mathbf{V}_G^{-1}(Z'R^{-1}Z)^{-1}Z'R^{-1}, \quad (29)$$

where

$$\begin{aligned} \mathbf{V}_G &= (Z'V^{-1}Z)^{-1} = (Z'R^{-1}Z)^{-1} + \mathbf{G} \\ &= \begin{bmatrix} (D'R_1^{-1}D) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & (D'^{-1}D)^{-1} \end{bmatrix} + \mathbf{G}. \end{aligned} \quad (30)$$

Furthermore, using Eq. (11), we have the following estimate of the fixed effect parameter:

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}'_G\mathbf{V}_G^{-1}\mathbf{X}_G)^{-1}\mathbf{X}'_G\mathbf{V}_G^{-1}(Z'R^{-1}Z)^{-1}Z'R^{-1}\mathbf{y}. \quad (31)$$

Note that the form in Eq. (31) allows a multi-level analysis similar to (Beckmann et al., 2003). More specifically, if we define $\hat{\boldsymbol{\beta}} = (Z'\hat{R}^{-1}Z)^{-1}Z'\hat{R}^{-1}\mathbf{y}$, then using the separable variance assumption in Eq. (25), we have

$$\hat{\boldsymbol{\beta}} = (Z'\hat{R}^{-1}Z)^{-1}Z'\hat{R}^{-1}\mathbf{y} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1(r) \\ \vdots \\ \hat{\boldsymbol{\beta}}_n(r) \end{bmatrix}$$

where the summary statistics are given by

$$\hat{\boldsymbol{\beta}}_l(r) = (D'\hat{R}_l^{-1}D)^{-1}D'\hat{R}_l^{-1}\mathbf{y}_l(r). \quad (32)$$

Therefore, the fixed effect parameter estimate is given by

$$\hat{\boldsymbol{\alpha}}(r) = (\mathbf{X}'_G\hat{\mathbf{V}}_G^{-1}\mathbf{X}_G)^{-1}\mathbf{X}'_G\hat{\mathbf{V}}_G^{-1}\hat{\boldsymbol{\beta}}(r),$$

which is equivalent to performing a second level GLM

$$\hat{\boldsymbol{\beta}}(r) = \mathbf{X}_G\boldsymbol{\alpha}(r) + \boldsymbol{\eta}(r), \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{V}}_G). \quad (33)$$

Under the uncorrelated noise assumption in Eq. (25), the group-level fixed effect parameter estimate is given by performing the first-level analysis using Eq. (32) and then performing the second level GLM using Eq. (18). This is another derivation of multi-level analysis by Beckmann et al. (2003), but the derivation using Eqs. (28) and (29) using linear mixed effect model is novel, since it can be easily extended to more general group analysis such as group level t -test or ANOVA.

Specifically, the aforementioned summary statistics approach can be easily extended for group level two sample t -test to see the difference between two groups or for general ANOVA analysis. For example, for two-sample t -test of control and patient groups, we can generalize Eq. (21) across the groups:

$$\mathbf{y}_{il}(r) = D\boldsymbol{\alpha}_{il}(r) + \epsilon_{il}(r), \quad i = 1, 2, \quad l = 1, \dots, n,$$

where $\mathbf{y}_{il}(r)$ denotes the l th subject time series at r location from the i th group. Then, regression coefficients are modeled as a random effect:

$$\boldsymbol{\alpha}_{il}(r) = \boldsymbol{\mu}(r) + \boldsymbol{\tau}_i(r) + \boldsymbol{\beta}_{il}(r), \quad (34)$$

where $\mu(r)$ denotes the global mean, $\tau_i(r)$ is the group effect, and $\beta_{il}(r)$ denotes between the subject random effect parameter, respectively. Then, very similar multi-level analysis can be easily derived, and we need to perform one-way ANOVA for the summary statistics:

$$\hat{\beta}(r) = X_G \alpha(r) + \eta(r), \quad \eta \sim \mathcal{N}(\mathbf{0}, \hat{V}_G), \quad (35)$$

where

$$\alpha = \begin{bmatrix} \mu(r) \\ \tau_1(r) \\ \tau_2(r) \end{bmatrix}, X_G = [\mathbf{1}_2 \otimes \mathbf{1}_n \otimes I_p, I_2 \otimes \mathbf{1}_n \otimes I_p], Z = I_{2n} \otimes D. \quad (36)$$

and V_G is given by

$$V_G = (Z'V^{-1}Z)^{-1} = (Z'R^{-1}Z)^{-1} + G \\ = \begin{bmatrix} (D'R_{11}^{-1}D)^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & (D'R_{2n}^{-1}D)^{-1} \end{bmatrix} + G, \quad (37)$$

where R_{il} denotes the first-level individual covariance estimation for the i th group l th subject. Similar multi-level analysis can be easily derived for multi-way ANOVA, so we will skip the derivation here.

Single-level versus multi-level approach in SPM-type group analysis

From the previous discussion, we know that, for group analysis, multi-level analysis appears equivalent to a single-level mixed effect approach that processes all group data together. This implies that we only need to store *summary statistics* from each individual data for group analysis, which makes the analysis much simpler. Indeed, this is the standard method in SPM, FSL (FMRIB Software Library), and FMRISTAT for fMRI processing (Beckmann et al., 2003; Friston et al., 2011; Smith et al., 2004; Woolrich et al., 2009).

However, one should be aware that such equivalence is not a true equivalence, and indeed the multi-level analysis is an approximation of the full mixed effect model unless the second-level covariance estimate \hat{V}_G is accurately estimated across the individual and group level. More specifically, as shown in Eqs. (30) and (37), the estimation of \hat{V}_G requires the estimation of \hat{R}_l (or \hat{R}_{il}) as well as \hat{G} . The technical difficulty lies in that ReML covariance estimation needs to compute \hat{R}_l (or \hat{R}_{il}) as well as \hat{G} simultaneously using the full data, whereas the multi-level analysis estimates \hat{R}_l (or \hat{R}_{il}) without considering other individual data and transfers them to the second level. Hence, the only estimable covariance from the second-level analysis is only for \hat{G} , which makes the overall covariance estimation sub-optimal. Aside from the difficulty of estimating \hat{V}_G for the second level analysis, there exists another important difference since the degrees of freedom now increase to $nN - p$ rather than $n - p$. As described before, this reduces the threshold values for hypothesis testing for the same statistical significance, which makes the detection more sensitive.

In fMRI analysis, the covariance estimation problem has been addressed using a few different approaches. For example, in FSL, a Bayesian approach is taken to decouple the first-level and the second-level covariance estimation (Woolrich et al., 2009). In FMRISTAT (Worsley et al., 2002), which is the closest to the approach of this review, the second term of \hat{V}_G is estimated within ReML framework using expectation-maximization (EM) algorithm. More specifically, starting with an initial value of $G = \hat{\beta}'P_{X_G}^{\perp}\hat{\beta}/\nu$ where $\nu = n - \text{rank}(X_G)$, the updated estimate is given by

$$\hat{G} = \left(G(p + \text{tr}(SR_{V_G})) \right) + G^2 \hat{\beta}'R_{V_G}^2\hat{\beta}/n, \quad (38)$$

where $p = \text{rank}(X_G)$, $S = \text{diag}((D'R_1^{-1}D)^{-1}, \dots, (D'R_n^{-1}D)^{-1})$, and

$$R_{V_G} = V_G^{-1} - V_G^{-1}X_G(X'_GX_GV_G^{-1}X_G)^{-1}X'_GX_GV_G^{-1}. \quad (39)$$

Replacing G with \hat{G} in Eq. (30), the processes from Eqs. (38) and (39) are iterated until it converges. Note that the EM procedure is, however, still an approximation of ReML, since the first level individual variances are not updated during EM procedure.

Another simple approach, which has been taken in SPM, is to assume homoscedastic variance across subjects. More specifically, let us consider Eq. (30). If we assume that $R_l = \sigma_l^2 I$ (which is usually the case after the prewhitening) and $G = \sigma_G^2 I$, then we have

$$V_G = \begin{bmatrix} \sigma_1^2(D'D)^{-1} + \sigma_G^2 I & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^2(D'D)^{-1} + \sigma_G^2 I \end{bmatrix}.$$

In designing the design matrix D for the first level analysis, the regressors are usually made orthogonal to each other. This is because, in addition to the constant regressor to regress out the mean, the other regressors are a canonical hemodynamic response and its first derivatives; so as long as they are de-trended to have zero mean, they are nearly orthogonal to each other. Under the orthogonality condition for the regressors, i.e. $D'D \approx I$, V_G becomes a diagonal matrix. Furthermore, under the homoscedastic variance assumption across subjects, we have $\sigma_1^2 = \dots = \sigma_n^2 = \sigma_F^2$; hence, we have $V_G = \sigma^2 I$ for some $\sigma^2 = \sigma_G^2 + \sigma_F^2$. Therefore, under the aforementioned assumptions, we can show that

$$S = \frac{\hat{\alpha}(r)'L'(L(X'\hat{V}^{-1}X)^{-1}L')^{-1}L\hat{\alpha}(r)}{p_1} \\ = \frac{\hat{\alpha}(r)'L'(L(X'_GX_G^{-1}X_G)^{-1}L')^{-1}L\hat{\alpha}(r)}{p_1} \\ = \frac{\hat{\alpha}(r)'L'(L(X'_GX_G)^{-1}L')^{-1}L\hat{\alpha}(r)}{\sigma^2 p_1} \\ = \frac{\hat{\beta}'(P_{X_{G0}}^{\perp} - P_{X_G}^{\perp})\hat{\beta} N - p}{\hat{\beta}'P_{X_G}^{\perp}\hat{\beta}} \frac{N - p}{p_1} F_{p_1, \nu},$$

where X_{G0} denotes the reduced model by excluding the effect estimated by contrast L and where $\hat{\beta}$ is summary statistics. Here, the second equality comes from the matrix equality $X'\hat{V}^{-1}X = X'_G\hat{V}_G^{-1}X_G$, the third from $\hat{V}_G = \hat{\sigma}^2 I$ and $\hat{\sigma}^2 = \hat{\beta}'P_{X_G}^{\perp}\hat{\beta}/(N - p)$, the fourth one from the equivalence relationship in the [Equivalence](#) section, which is another important novel contribution of this paper. Since these F -statistics are standard statistics for ANOVA analysis, the result indicates that we can perform classical ANOVA analysis using the summary statistics, and such analysis is equivalent to the inference in a mixed model as long as our assumption holds. Moreover, we do not need to perform computationally expensive ReML covariance estimation since the ReML variance estimation parts are already built-in within the resulting F -statistics.

In practice, however, the homoscedastic variance assumption can be easily violated, which often results in $\hat{\sigma}^2 \hat{\beta}'P_{X_G}^{\perp}\hat{\beta}/(N - p) < \hat{\sigma}_F^2 \hat{\beta}'\hat{\beta}/n \sum_i \sigma_i^2$ which is contradictory to the basic model $\sigma^2 = \sigma_G^2 + \sigma_F^2$ from the homoscedastic variance assumption. In this case, the variance estimation from the second level should be constrained such that $\hat{\sigma}^2 \geq \hat{\sigma}_F^2$. Under

the constraint, it is easy to show that the ML estimation of the variance is given by

$$\hat{\sigma}_{ML}^2 = \max \left\{ \frac{\hat{\mathbf{B}}^T \mathbf{P}_{X_c}^\perp \hat{\mathbf{B}}}{(N-p)}, \frac{1}{n} \sum_i \sigma_i^2 \right\}. \quad (40)$$

The constrained form (Eq. (40)) is very important to make the variance estimation under homoscedastic variance assumption works. Fig. 3 shows the group one-sample *t*-test results using NIRS-SPM. Datasets were acquired from 3 subjects performing a finger tapping task. Here, we used a block design sequence consisting of 15 s of task and 72 s of rest in one cycle, and the full experimental run consisted of four task and rest cycles. Activation patterns of HbO and HbR were localized on the primary motor cortex, which is the main region of interest of this experiment. Although *t*-statistic values were decreased under the homoscedastic variance model, overall activation areas were nearly identical to those from ReML model with EM procedure after thresholding for a given *p*-value.

The two versions of the variance estimation – one using ReML with EM procedure and homoscedastic assumption have been implemented in NIRS-SPM and will be available in the next release of NIRS-SPM.

Standard statistics versus SPM type multi-level group analysis

At a glance, a standard statistic for fNIRS analysis appears very different from SPM-type multi-level group analysis. However, a careful observation reveals that the standard statistics using *t*-test and ANOVA are equivalent to the second-level SPM type group analysis using *t*-test and ANOVA when the first-level design matrix is composed of a single boxcar regressor that represents each segment (baseline, conditions) and if we assume homoscedastic variance across subjects. This is because the

boxcar regressors average the data when we apply the ordinary least square to find the regression coefficients. However, SPM-type multi-level analysis is more effective than the standard statistics since we can use nuisance regressors to remove drift, cardiac artifacts, etc. Furthermore, the total degree of freedom for SPM-type analysis is usually larger than that of standard statistics since the first level degrees of freedom are transferred to the second-level, which makes the test more sensitive.

The multiple comparison problem in fNIRS

Hypothesis testing

One of main goals of statistical analysis in fNIRS is to decide whether the statistic represents convincing evidence of the effect we are interested in. A usual approach is that we test the statistic against the null hypothesis H_0 , which is the hypothesis that there is no fixed effect:

$$H_0: \mathbf{L}\boldsymbol{\alpha} = \mathbf{0}.$$

Under the null hypothesis, a corresponding random variable for the test statistics is given by

$$\mathcal{S} = \frac{\hat{\boldsymbol{\alpha}} \mathbf{L}' \left(\mathbf{L} (\mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X}) - \mathbf{L}' \right)^{-1} \mathbf{L} \hat{\boldsymbol{\alpha}}}{p_1}, \quad (41)$$

which becomes $F_{p_1, r}$ -statistics. Then, to control the type I error for a given statistical significance α , we should calculate a threshold u such that

$$\Pr\{\mathcal{S} > u\} = \alpha. \quad (42)$$

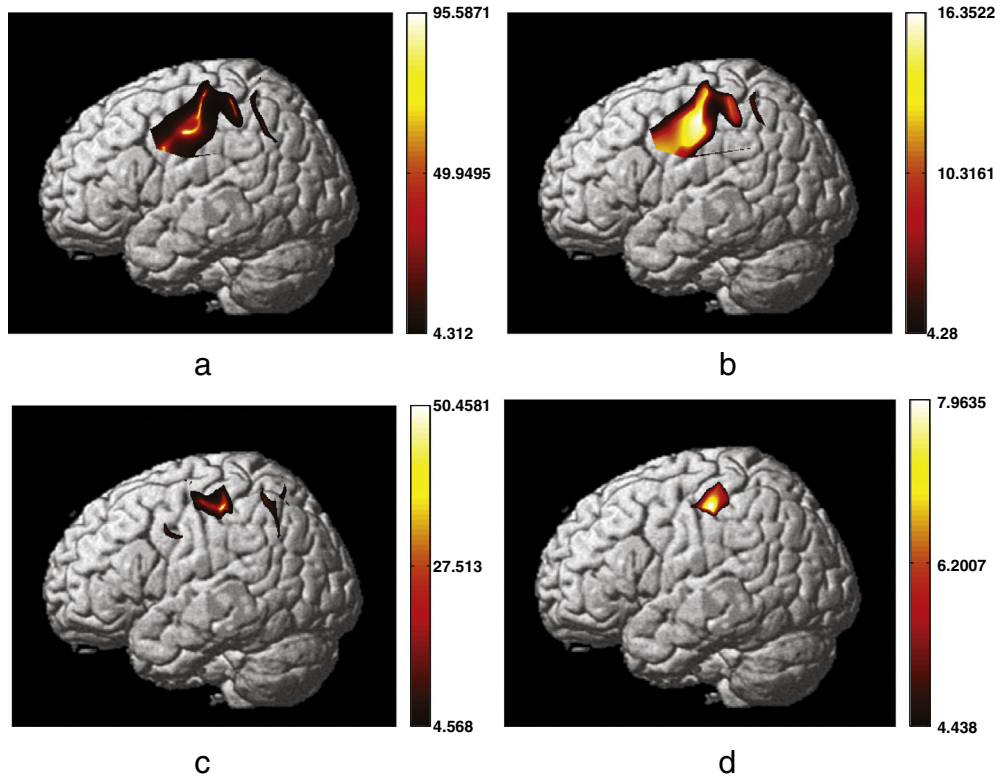


Fig. 3. Group activation maps estimated from (a, b) HbO and (c, d) HbR responses during finger tapping task (3 subjects, $p < 0.001$, the expected Euler characteristic (EC) correction based on Lipschitz–Killing curvature). In the left column, restricted maximum likelihood estimation (ReML) model with expectation maximization (EM) procedure was used for second-level covariance estimate $\hat{\mathbf{V}}_G$. In the right column, homoscedastic variance model was used for estimate of $\hat{\mathbf{V}}_G$. Both activation patterns of HbO and HbR were localized on the primary motor cortex, which is the main region of interest of this experiment. Moreover, two variance models give similar activation regions after thresholding for a given *p*-value.

The meaning of this test is that even though there are chances that no effect exists, the probability that we can erroneously reject the null hypothesis, is not greater than α (hence, we can control the type I error to α).

Multiple comparison problems

However, in many statistical fNIRS analyses, multiple comparison problems are often encountered. In this situation, errors in inference are more likely to increase when we consider the entire test as a whole. Therefore, in multiple comparison problems, we generally require a stronger level of evidence to be observed for each individual comparison to control the errors in simultaneous inference. This implies that we need to adjust the statistical significance, i.e. p -value, to reduce the type I errors that occur when statistical tests are performed simultaneously.

One of the intuitive examples of such family-wise error rate (FWER) control is Bonferroni correction (Dunn, 1961). Let H_1, \dots, H_m be a family of hypotheses, and p_1, \dots, p_m the corresponding p -values. The family-wise error rate is then the probability of rejecting at least one of the hypotheses. The Bonferroni correction states that rejecting all hypotheses using $p_i < \alpha/m$ will control the FWER up to α . This can be proven using Boole's inequality (Miller, 1981). In addition to Bonferroni correction, there are many variations of controlling FWER in multiple comparison problems, which include Least Significant Differences (LSD), Tukey, Scheffe, random field correction, etc. (Miller, 1981).

The multiple comparison problems have been frequently encountered in fNIRS studies in both discrete and continuous contexts, so we review them here.

Multiple comparison in discrete setup

In fNIRS, representative examples of multiple comparison problems in discrete setup come from ANOVA (Arenth et al., 2007; Bartocci et al., 2000; Fallgatter and Strik, 1998; Fallgatter and Strik, 2000; Folley and Park, 2005; Germon et al., 1994; Herrmann et al., 2003, 2004; Hoshi, 2003; Irani et al., 2007; Kameyama et al., 2006; Kleinschmidt et al., 1996; Matsuo et al., 2002; Mehagnoul-Schipper et al., 2002; Okamoto et al., 2004; Saitou et al., 2000; Suto et al., 2004; Taga et al., 2003; Villringer et al., 1993). As a simple example, in one way ANOVA with at least three levels, an omnibus F -test is first used to assess any of several possible differences, and then post hoc analysis is performed. More specifically, if the omnibus F -test rejects the null hypothesis and concludes that at least one mean is different from the others, then the next question is how they are different. This is post hoc analysis, and we tested multiple hypotheses at the same time.

For instance, in Bartocci et al. (2000), to detect the activation of the olfactory cortex in newborn infants after odor stimulation, one-way ANOVA was performed with three levels of the stimuli in the following order: (1) control, (2) colostrum, and (3) vanilla. They first found a significant difference between all three test conditions and performed post hoc comparison to test the changes between each smell. Because post hoc analysis requires multiple comparisons (control versus colostrum; control versus vanilla; colostrum versus vanilla), p -value correction is necessary, and they used the Newman-Keuls method (Keuls, 1952; Newman, 1939). In Herrmann et al. (2004), to assess bilaterally reduced frontal activation during a verbal fluency task in depressed patients, two-by-four-by-two (two hemispheres, four task conditions, two diagnostic groups) ANOVA for repeated measurements was calculated for the variables HbO and HbR. They first observed statistically significant main effects of the factor "condition" as well as a significant main effect of the factor "diagnosis" and a significant interaction between these two variables. They performed subsequently-conducted post hoc tests and p -value correction to reveal that the mean HbO concentration in each of the three conditions was significantly increased in the healthy subjects when compared to the depressive subgroup.

Other instances of multiple comparison problems in discrete setup can be also found in fNIRS analysis, especially when the measurements from each channels are not interpolated but used as a discrete set of time course data. For example, in Plichta et al. (2006), to estimate the statistical significance of fNIRS response during task periods, the Bonferroni correction and the Dubey/Armitage-Parmar alpha boundary (Sankoh et al., 1997) were used for statistical inference of activated channels. Hofmann et al. (2008) applied a general linear model to statistically analyze fNIRS data and applied the partial Bonferroni correction for making inference of the activation for each channel using the Dubey/Armitage-Parmar alpha boundary.

Multiple comparison in continuous setup

Rather than using the discrete set of fNIRS channels as they are, if interpolation (or tomographic inversion) is used to obtain a topographic (or tomographic) map, then we encounter the multiple comparison problem in a continuous setup. More specifically, we need to test a random process (rather than a random variable) \mathcal{S} in Eq. (41) that is continuously distributed across the volume of interest. In this case, we need to test the following null hypothesis:

$$H_0: \mathbf{L}\alpha(r) = \mathbf{0}, \quad \forall r \in \Psi,$$

where Ψ denotes a search volume. The meaning of this null hypothesis is that we consider $\{\mathcal{S}(r), r \in \Psi\}$ as a whole; hence, the problem becomes multiple comparison problems with a continuously indexed infinite number of hypothesis. For such hypothesis testing, we need to calculate the following statistical significance

$$Pr\left\{\max_{r \in \Psi} \mathcal{S}(r) > u\right\} = \alpha, \quad (43)$$

where u denotes a corresponding threshold value.

In general, spatial correlation is very common in fNIRS topographic/tomographic maps, which need to be exploited for inference. The direct calculation of Eq. (43) under spatial correlation is, however, extremely difficult, even for a 1-dimensional random process (Adler and Taylor, 2007), so this problem is usually addressed by the expected Euler characteristic (EC) method (Adler and Taylor, 2007; Friston et al., 2011). The Euler characteristic of an excursion set is a geometric property of a volume which can be calculated as the number of connected excursion set minus the number of holes plus the number of hollows (Worsley, 1995). More specifically, we define the excursion set as follows:

$$A_u = A_u(\mathcal{S}, \Psi) = \{r \in \Psi: \mathcal{S}(r) \geq u\}.$$

One of the most important discoveries in random field theory is that for a large u values, the excursion probability can be approximated using expected EC as follows:

$$Pr\left\{\max_{r \in \Psi} \mathcal{S}(r) \geq u\right\} \approx E\{\varphi(A_u(\mathcal{S}, \Psi))\}, \quad (44)$$

where $\varphi(A_u)$ denotes the Euler characteristics of a set A_u .

The intuition behind Eq. (44) is the so-called "Poisson clumping heuristics" saying that, for a large threshold value u , the excursion set usually has one cluster (i.e. EC value becomes one) without holes; so the mean value is a good approximation of the excursion probability (Adler, 2000; Adler and Taylor, 2007).

Hence, rather than directly calculating Eq. (43), we are interested in finding the mean EC density; and for a smooth random field, it has been shown (Adler and Taylor, 2007) that

$$E\{\varphi(A_u(\mathcal{S}, \Psi))\} = \sum_{j=0}^N \rho_j(u) \mathcal{L}_j(\Psi), \quad (45)$$

where $\mathcal{L}_j(\Psi)$ is the j -dimensional Lipschitz-Killing curvatures (LKC) of

the manifold Ψ representing the intrinsic volume of Ψ in a Riemannian metric, and $\rho_j(\cdot)$ denotes the j -dimensional EC density function determined by the type of random fields (Adler and Taylor, 2007).

A method to estimate LKC was developed in Taylor and Worsley (2007), and we review it here briefly. First, we calculate the least-square residual $\eta(r)$:

$$\eta(r) = P_X^\perp \mathbf{y}(r), \quad (46)$$

where $\mathbf{y}(r)$ is the measurement at $r \in \Psi$ and X is a fixed effect design matrix. Then, the corresponding normalized residual is as follows:

$$\bar{\eta}(r) = \eta(r) / \|\eta(r)\|_2 \in \mathbb{R}^N.$$

Then, for example, the 2-dimensional LKC of Ψ can be estimated as follows:

$$\hat{\mathcal{L}}_2(\Psi) = \sum_{r \in \Psi} |\nabla \eta(r)' \nabla \eta(r)|^{\frac{1}{2}},$$

where $|\cdot|$ denotes the determinant and $\nabla \eta$ is a gradient with respect to r .

Now, let us discuss how these equations relate to a practical analysis of fNIRS data. For example, let us compare two cases where the interpolated fNIRS on the cortical surface is similar (CASE I), and very dissimilar (CASE II). For CASE I, $\eta(r)$ should be similar and the coefficient $\hat{\mathcal{L}}_2(\Psi)$ tends to smaller, whereas for CASE II, the resulting coefficient should be larger. This implies that for the same Type I error α value in Eq. (43), due to Eqs. (45) and (44), the corresponding threshold u for CASE I is smaller than that of CASE II.

Estimating the lower dimensional LKC is more complicated, so a short-cut method is often used in practice. Specifically, the D -dimensional LKC, which makes the largest contribution to the p -value approximation, is first estimated. Then, the lower dimensional terms are estimated by simply assuming that Ψ is a D -dimensional ball (Adler and Taylor, 2007). In fNIRS topographic mapping, the dimension is $D = 2$; therefore, we assume that Ψ is a disk, and then a short-cut calculation of $\hat{\mathcal{L}}_1(\Psi)$ and $\hat{\mathcal{L}}_0(\Psi)$ can be calculated as follows:

$$\hat{\mathcal{L}}_0(\Psi) = 1, \quad \hat{\mathcal{L}}_1(\Psi) = \sqrt{\pi \hat{\mathcal{L}}_2(\Psi)}. \quad (47)$$

On the other hand, for a Gaussian random field $S(r)$, Sun's formula can be also used to calculate the excursion probability if the underlying random process can be represented using *finite* term Karhunen–Loève expansion; i.e.

$$S(r, w) = \sum_{j=1}^I \xi_j(w) \psi_j(r).$$

Then, the excursion probability is given by

$$Pr \left\{ \sup_{r \in \Psi} S(r) \geq u \right\} \approx \sum_{j=0}^N \rho_j(u) \mathcal{L}_j(\psi(\Psi)) \quad (48)$$

where

$$\psi = [\psi_1, \psi_2, \dots, \psi_I]^T. \quad (49)$$

Note that an excursion probability using the volume of the tube formula is the same as Eq. (45) except the following two differences: 1) the zeroth order EC density function $\rho_0(u)$ (see Theorem 10.6.1 in Adler and Taylor, 2007), and 2) the LKC \mathcal{L}_j is now for $\psi(\Psi) = [\psi_1(\Psi), \dots, \psi_I(\Psi)]$.

A method to estimate LKC for the tube in the fNIRS problem was developed in Ye et al. (2009) for the case of fNIRS, and we review it here briefly. As the degree of freedom is quite large due to the long time trace, t_ν converges to the Z statistics asymptotically:

$$S = \frac{L\hat{\alpha}}{\sqrt{L(X'\hat{V}^{-1}X)^{-1}L'}} \rightarrow \frac{L\hat{\alpha}}{\sqrt{L(X'V^{-1}X)^{-1}L'}},$$

which can be represented as a finite term Karhunen–Loève expansion:

$$\frac{L\hat{\alpha}}{\sqrt{L(X'V^{-1}X)^{-1}L'}} = \mathcal{Z}'\psi(r),$$

where $\mathcal{Z} \sim \mathcal{N}(0, I)$ and ψ is given by

$$\psi(r) = \frac{(\hat{\Sigma}^{1/2} \mathbf{b}(r) \otimes \hat{\Lambda}^{1/2}) L'}{\sqrt{\mathbf{b}'(r) \hat{\Sigma} \mathbf{b}(r) (L(D'\hat{\Lambda}^{-1}D)^{-1}L')}}. \quad (50)$$

(Note that this expression is slightly different from the original derivation in Ye et al. (2009), i.e. $(D'\hat{\Lambda}^{-1}D)^{-1}$ instead of $D'\hat{\Lambda}D^{-1}$ since the original derivation was based on the least square fitting rather than the maximum likelihood estimation used here.) Then, the corresponding 2-dimensional LKC of $\psi(\Psi)$ can be estimated as follows:

$$\hat{\mathcal{L}}_2(\psi(\Psi)) = \sum_{r \in \Psi} E |\nabla S(r)' \nabla S(r)|^{\frac{1}{2}} = \sum_{r \in \Psi} |\nabla \psi(r)' \nabla \psi(r)|^{\frac{1}{2}}.$$

Again, let us discuss the meaning of Eq. (50). Similar to EC analysis, the spatial similarity of the random field ψ is an important factor to determine the threshold values for activation detection. Here, there are three important factors: channel-wise noise variance $\hat{\Sigma}$, channel-by-channel correlation $\hat{\Lambda}$ as well as the interpolation kernel \mathbf{b} . This implies that the tube formula tries to find the optimal threshold value for activation detection by considering both geometric factors as well as noise statistics of measurements.

Lower dimensional LKC can be found similarly as in EC formula. One of the advantages of using a tube formula is that we often have an explicit representation of ψ , which makes the resulting computation simple. Also, for many fNIRS data with large degree of freedom (i.e. $\nu \gg 1$), the calculation of t statistics using the EC formula is numerically unstable due to the division by ν . However, the proposed tube formula is not a function of ν , which makes the calculation numerically stable. Hence, NIRS-SPM exploits the equivalence between the tube and expected EC formula to switch between tube and EC formula in calculating the p -value.

Another interesting observation made in Li et al. (2012) is that, using the equivalence between the tube and EC formula, the authors identified that the channel-wise variance matrix $\hat{\Sigma}$ should not be diagonal and needs to be estimated by considering the channel-wise correlation. This is a fundamental departure from fMRI analysis, where the voxel-wise noise is assumed independent but correlation is inserted later using Gaussian smoothing (Friston et al., 2011). The inherent correlation in fNIRS may be due to skin blood flow, intrinsic connectivity, etc., but with the consideration of that correlation, the EC and tube formula provides nearly identical p -value correction despite its very different formula, which indicates the importance of considering channel-wise correlation in fNIRS analysis (Li et al., 2012).

False discovery rate control

Another different way of approaching the multiple comparison problem is using false discovery rate control (Benjamini and Hochberg, 1995). Rather than controlling type-I error, false discovery rate (FDR) controls the expected proportion of falsely declared-active detections among the total declared-active hypotheses (Benjamini and Hochberg, 1995).

FDR has been applied to channel-wise analysis (Singh and Dan, 2006) and source localization problem (Jung et al., 2012) of fNIRS data, for making inferences about the activated region. As FDR approach does not require spatial smoothing process in calculation of p -value, it might be an optimal method in channel-wise statistics of fNIRS (Singh and Dan, 2006). In addition, compared with the Bonferroni correction (Dunn, 1961), FDR approach provides higher power (defined as average proportion of detected activation) while still adjusting the balance with specificity. Using the simulated and real fNIRS data, Singh and Dan (2006) showed that FDR offers more power and robust against changing number of channels within region of interest (ROI) than Bonferroni correction. FDR method was also applied in the study of the lateral prefrontal function during taste encoding and reliably detected the activation pattern (Okamoto et al., 2006). Another interesting application of FDR control is a source localization approach using the multiple signal classification (MUSIC) algorithm (Jung et al., 2012). They used FDR to find the threshold values of the MUSIC spectrum for given q -values, and reconstructed the activated region which is very close (6 mm to 18 mm) to the Montreal Neurological Institute (MNI) coordinates of target regions during finger tapping. We will describe the FDR control in more detail.

The total n voxels are first classified into one of four types, as shown in Table 1. The FDR is then defined as follows:

$$\text{FDR} = \frac{FP}{FP + TP} = \frac{FP}{T_1}, \quad (51)$$

which is the proportion of the false-positives (FP) among only the rejected null hypotheses ($FP + TP$).

Inference procedure based on FDR consists of the following three steps (Benjamini and Hochberg, 1995). First, specify a desired q value ($0 \leq q \leq 1$), which ensures that the expected FDR is less than or equal to q :

$$E\{\text{FDR}\} \leq \frac{n_0}{n} q \leq q. \quad (52)$$

Here, to compute the threshold based on FDR method and associated q -value, one needs uncorrected p -values, i.e. p_1, \dots, p_n for the hypothesis H_1, \dots, H_n . Secondly, sort the uncorrected p -values into ascending order, $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$, when the subscript (i) denotes the indices for the reordered sequence. The final step is to evaluate the following inequality in reverse order:

$$p(i) \leq \frac{q}{n}, \quad i = n, n-1, \dots, 1. \quad (53)$$

Let k be the largest i that satisfies Eq. (53). Then we reject all the hypotheses, $H_{(1)}, H_{(2)}, \dots, H_{(k)}$. That is, we threshold at the p_k value

and declare the corresponding voxels active. Hence, we investigate the significantly active voxels controlling the expectation of FDR less than q -level.

Functional connectivity analysis

Compared to fMRI, the temporal resolution of fNIRS is significantly higher. This gives us an opportunity to investigate the functional connectivity of the brain by exploiting the temporal correlations between multiple channels. In calculating the connectivity maps, two different approaches exist: seed-based analysis and independent component analysis. This area of research is still evolving, so we briefly review some of the representative works.

Seed-based approaches

The simplest seed-based functional connectivity analysis can be conducted by selecting a specific channel as a seed and calculating the correlation value between its time series and the time courses of all other source-detector combinations (Mesquita et al., 2010). More specifically, to remove the physiology-based systemic confounds from fNIRS signals, an auxiliary measure of blood pressure fluctuations is used as a regressor. Then, a filtered signal is then calculated by regressing out the confounds. For a correlation analysis, one channel within a region of interest is selected as a seed, and the Pearson correlation coefficient r between the time course of the seed and the time course of all other channels are calculated.

In White et al. (2009), functional connectivity maps over the motor and visual cortices were investigated using diffuse optical tomography (DOT). DOT channel arrays were placed over the visual and sensory motor cortices. Seed-based correlation maps were calculated. A 1 cm^3 volume was chosen as a seed region and the time traces within the region were averaged. The correlation coefficient was then calculated between the seed signal and every other voxel, using the Pearson correlation coefficient. To examine the significance of the inter-hemispheric correlations within the motor and visual networks, two sets of inter-hemispheric r -values (motor-to-motor, and visual-to-visual) were calculated. The null hypothesis was then that each inter-hemispheric set was indistinguishable from the background set. This hypothesis was tested with a paired Student's t -test. The t -statistics were converted to p -value using the right-tail of the distribution. Functional connectivities within the visual and motor networks were observed from DOT measurements during visual and motor tasks: both visual and motor networks showed high levels of inter-hemispheric correlation. Moreover, with the resting-state analysis of DOT, the researchers found high correlations bilaterally in the same regions as in the task-related responses.

Sasai et al. (2012) tested whether the resting-state networks constructed by the fNIRS and fMRI measurements are consistent. Three types of correlations maps were calculated using BOLD signals and the following different seed signals: 1) fNIRS-measured oxy- and deoxy-hemoglobin signals of all fNIRS channels positioned over the bilateral frontal, temporal, and occipital regions, 2) BOLD time-series extracted from the brain regions corresponding to the fNIRS channel positions, and 3) fNIRS and BOLD time-series within the predefined regions of a resting-state network (dorsal attention, fronto-parietal control, and default mode networks). They observed that correlation maps calculated with fNIRS signals as a seed were clearly similar to the correlation maps with BOLD signals. These suggest that fNIRS signals obtained at several cortical regions during the resting state mainly reflect regional spontaneous hemodynamic fluctuations that originate from spontaneous cortical activity and include information that characterizes the resting-state networks.

In Sasai et al. (2011), to investigate the frequency-specific characteristics of functional connectivity, they decomposed the oxy-hemoglobin and deoxy-hemoglobin signals into band-pass filtered signals with narrow

Table 1
Voxel classification in multiple testing N hypotheses.

	Declared non-significant	Declared significant	Total
Truly non-significant	TN	FP	n_0
Truly significant	FN	TP	n_1
Total declared	T_0	T_1	n

frequency bands (0.009–0.02, 0.02–0.04, 0.04–0.06, 0.06–0.08, and 0.08–0.1 Hz). To clarify the coherence of the functional connectivity, the cross-spectral density and power spectral density between all of the channel pairs were calculated using Welch's averaged modified periodogram method.

The results of HbO signals showed that homologous connectivity is subtended in the broader frequency bandwidth (0.009–0.1 Hz), while the fronto-posterior connectivity is present in the narrower bandwidth (0.04–0.1 Hz). Moreover, the averaged coherence of the homologous connectivity was high over a wide frequency range (~0.1 Hz), while the coherence of the fronto-posterior connectivity was high only within a specific frequency range (0.04–0.1 Hz).

Homae et al. (2010) investigated the developmental changes of a brain network from several days to 6 months after birth by using spontaneous fluctuation of HbO signals. Here, temporal correlation between all pairs of measurement fNIRS channels was calculated. They found that in the temporal, parietal, and occipital regions, the homologous functional connectivity between and within hemispheres was increased, whereas in the frontal regions, it decreased progressively (see Fig. 4). To reveal the developmental changes in functional connectivity, they analyzed the differences among the infant groups on a channel pair basis. The correlation coefficient r was converted to z scores by Fisher's z transform. Then, they

evaluated the individual z scores as random effects and performed ANOVA analysis with subject groups as a factor.

ICA-based approach

In H. Zhang et al. (2010), ICA was applied to estimate the resting-state brain connectivities from fNIRS measurements. As the fNIRS data has a large number of temporal samples but a relatively small number of channels, temporal ICA is more suitable.

ICA analysis was separately applied to HbO and HbR time courses for the sensorimotor and visual systems with the following procedures: 1) the detrending was applied to the data using the first and second order polynomial functions to remove the linear and bilinear trends, 2) PCA reduction was performed to reduce the data dimensionality for each subject, and 3) ICA decomposition was applied. From ICA, the sensory-motor component and visual component were visually identified. To derive group-level results, the individual ICA results were transformed into a z -map. A two-tailed one sample t -test was then performed on the individual z -maps to generate random-effect group-level analysis. Compared with the conventional seed-based approach, the superior performance of ICA with higher sensitivity and specificity of functional connectivity maps was observed in the case of higher noise levels.

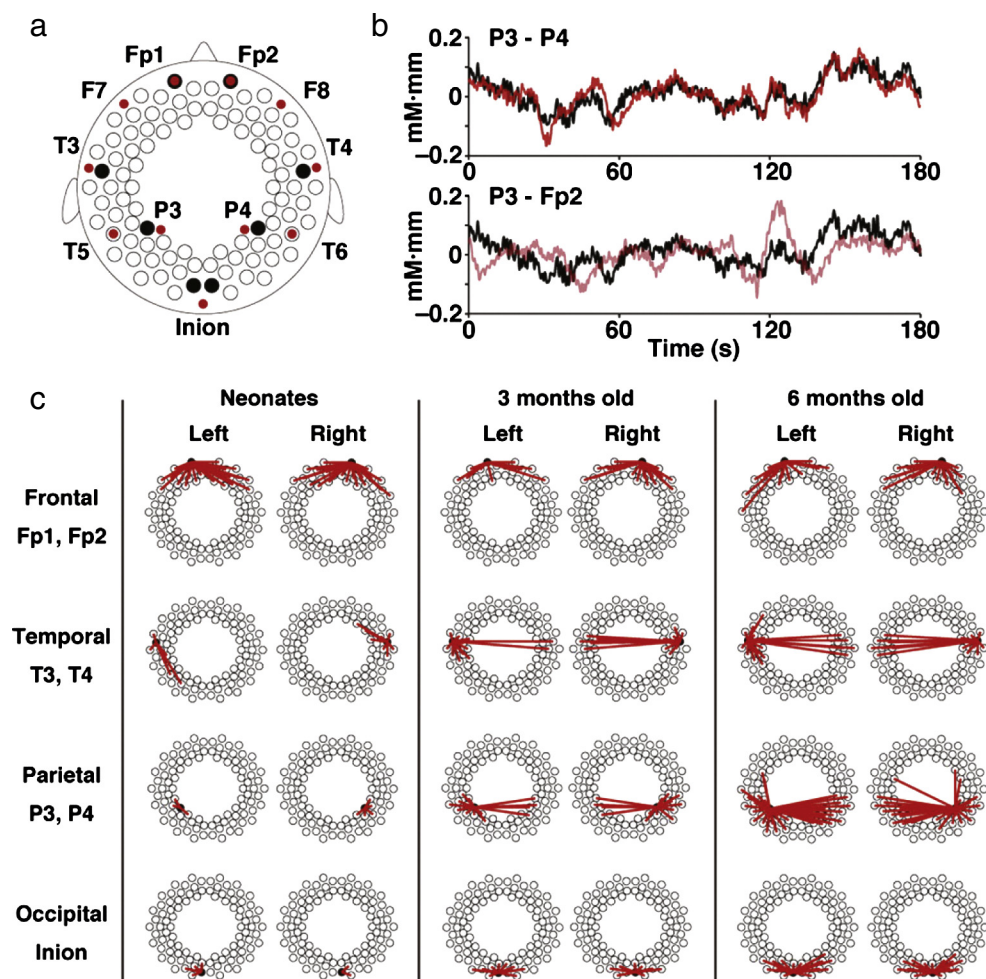


Fig. 4. Developmental changes of functional connectivities from neonates to 6 months old infants, characterized by spontaneous fluctuations of HbO signals. (a) Probe settings of the 94-channel fNIRS system. (b) Representative time courses of HbO measured from a 6 month old infant. Correlation between HbO signals from P3 and P4 was higher than that from P3 and Fp2. (c) Significant correlations between all pairs of measurement channels for neonate, 3 months old, and 6 months old. Note that in the temporal, parietal, and occipital regions, temporal correlations of HbO signals between the hemispheres were increased as the developmental changes of brain, however in the frontal region functional connectivities between the hemisphere were progressively decreased. Figure courtesy of Homae et al. (2010).

Conclusion and outlooks

In this paper, we have provided an extensive review of statistical analyses of fNIRS data. Applications of classical statistical analysis such as *t*- and *F*-statistics as well as more advanced SPM-type approaches have been discussed.

For a unified understanding of various classical statistical analyses, we provided a linear mixed model with ReML covariance estimation and showed that most of the classical analyses can be explained within this framework. We also provided a complete derivation of group analysis using general linear mixed model with ReML framework, which is novel and was not available before.

As for preprocessing steps for statistical analysis, methods to estimate and remove physiological confounds were discussed. We also discussed innovative approaches that utilize the confounds to estimate the various components of fNIRS signal and utilize them to improve resting-state fMRI analysis. We also provided a brief review of the functional connectivity analysis using fNIRS data. The present works have mainly focused on verifying that fNIRS analysis can also provide meaningful functional connectivity maps. We reviewed the two representative approaches: one using seed-based analysis and the other using ICA analysis. Both of the works provide promising results that confirm the efficacy of fNIRS for functional connectivity analysis.

Similar to other neuroimaging modalities, one of the exciting topics in fNIRS research is to develop widely accepted statistical analysis for functional connectivity analysis. Even though we briefly reviewed the existing literature, open problems still remain. First, finding the correct seed is one of the important factors in seed-based analysis since the connectivity maps are highly dependent on the choice of seeds. Furthermore, unlike the fMRI, the fNIRS signals are more prone to physiological artifacts, such as blood-flow, so effectively regressing out the confounds is extremely important for seed-based analysis. Second, the statistically rigorous choice of threshold values for the adjacency matrix is still an open problem, and currently most of the works choose heuristic method to calculate the threshold-values. One promising direction in this regards would be to exploit the topological structure of the adjacency matrix using the persistent homology (Lee et al., 2012), which totally removes the necessity of the thresholding.

Acknowledgments

This work was supported in part by Korea Science and Engineering Foundation, grant number 2009-0081089, and in part by a grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare (A121987-1211-0000300).

Equivalence

We are only interested in estimable function, so we assume $L' \subset R(X')$ (Rao and Toutenburg, 1999). Then, the test statistics is

$$S = \frac{\hat{\alpha}' L' (L(X'X)^{-1} L)^{-1} L \hat{\alpha}}{\hat{\sigma}^2 \text{rank}(L)} = \frac{y' X' (X'X)^{-1} L' (L(X'X)^{-1} L)^{-1} L (X'X)^{-1} X' y}{\hat{\sigma}^2 p_1}.$$

From the definition of the generalized inverse, we can find the full rank matrix X^* such that $R(X^*) = R(X)$ and $(X'X)^{-} = (X^*X^*)^{-1}$. Therefore, using the definition Eq. (15), we can define Ω such that

$$\begin{aligned} \Omega &= (X^*X^*)^{-1} - (X^*X^*)^{-1} L' (L(X'X)^{-1} L)^{-1} L (X^*X^*)^{-1} \\ &= \left(P_L^\perp (X^*X^*)^{-1} P_L^\perp \right). \end{aligned}$$

Then, we have

$$\begin{aligned} S &= \frac{y' X' ((X'X)^{-1} - \Omega_{X^*X^*}) X' y}{\hat{\sigma}^2 p_1} \\ &= \frac{y' X' ((X'X)^{-1} - (P_L^\perp (X'X)^{-1} P_L^\perp)^\dagger) X' y}{\hat{\sigma}^2 p_1} \\ &= \frac{y' X' ((X'X)^{-1} - Q(Q'XQ)^{-1}) X' y}{\hat{\sigma}^2 p_1} \\ &= \frac{y' (P_X - P_{X_0}) y}{\hat{\sigma}^2 p_1} \\ &= \frac{y' (P_{X_0}^\perp - P_X^\perp) y N - p_1}{y' P_X^\perp y \quad p_1} \\ &\sim F_{a-1, N-abc}. \end{aligned}$$

where $X_0 = XQ$ denotes the reduced model by excluding the effect of the contrast matrix L .

Conflict of interest

The authors declare no conflict of interest.

References

- Abdelnour, A.F., Huppert, T., 2009. Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model. *Neuroimage* 46 (1), 133–143.
- Adler, R.J., 2000. On excursion sets, tube formulas, and maxima of random fields. *Ann. Appl. Probab.* 10 (1), 1–74.
- Adler, R., Taylor, J., 2007. *Random Fields and Geometry*, vol. 115. Springer Verlag.
- Akgül, C.B., Akin, A., Sankur, B., 2006. Extraction of cognitive activity-related waveforms from functional near-infrared spectroscopy signals. *Med. Biol. Eng. Comput.* 44 (11), 945–958.
- Aldrich, C., Wyatt, J., Spencer, J., Reynolds, E., Delpy, D., 1994. The effect of maternal oxygen administration on human fetal cerebral oxygenation measured during labour by near infrared spectroscopy. *BJOG* 101 (6), 509–513.
- Arenth, P., Ricker, J., Schultheis, M., 2007. Applications of functional near-infrared spectroscopy fNIRS to neurorehabilitation of cognitive disabilities. *Clin. Neuropsychol.* 21 (1), 38–57.
- Arridge, S.R., 1999. Optical tomography in medical imaging. *Inverse Probl.* 15 (2), R41.
- Bartocci, M., Winberg, J., Ruggiero, C., Bergqvist, L., Serra, G., Lagercrantz, H., 2000. Activation of olfactory cortex in newborn infants after odor stimulation: a functional near-infrared spectroscopy study. *Pediatr. Res.* 48 (1), 18–23.
- Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* 23 (2), 137–152.
- Beckmann, C., Jenkinson, M., Smith, S., 2003. General multilevel linear modeling for group analysis in fMRI. *Neuroimage* 20 (2), 1052–1063.
- Benaron, D., Hintz, S., Villringer, A., Boas, D., Kleinschmidt, A., Frahm, J., Hirth, C., Obrig, H., van Houten, J., Kermit, E., Cheong, W., Stevenson, D., 2000. Noninvasive functional imaging of human brain using light. *J. Cereb. Blood Flow Metab.* 20 (3), 469–477.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 289–300.
- Blasi, A., Phillips, D., Lloyd-Fox, S., Koh, P.H., Elwell, C.E., 2010. Automatic detection of motion artifacts in infant functional optical topography studies. *Adv. Exp. Med. Biol.* 662, 279–284.
- Boas, D., Brooks, D., Miller, E., DiMarzio, C., Kilmer, M., Gaudette, R., Zhang, Q., 2001. Imaging the body with diffuse optical tomography. *IEEE Signal Proc. Mag.* 18 (6), 57–75.
- Boas, D., Culver, J., Stott, J., Dunn, A., 2002. Three dimensional Monte Carlo code for photon migration through complex heterogeneous media including the adult human head. *Opt. Express* 10 (3), 159–170.
- Boas, D.A., Dale, A.M., Franceschini, M.A., 2004. Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *Neuroimage* 23 (Suppl. 1), S275–S288.
- Buxton, R., Uludağ, K., Dubowitz, D., Liu, T., 2004. Modeling the hemodynamic response to brain activation. *Neuroimage* 23, 220–233.
- Christensen, R., 2011. *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer.
- Ciftci, K., Sankur, B., Kahya, Y.P., Akin, A., 2008. Multilevel statistical inference from functional near-infrared spectroscopy data during Stroop interference. *IEEE Trans. Biomed. Eng.* 55 (9), 2212–2220.
- Comon, P., 1994. Independent component analysis, a new concept? *Signal Process.* 36 (3), 287–314.
- Cooper, R.J., Selb, J., Gagnon, L., Phillip, D., Schytz, H.W., Iversen, H.K., Ashina, M., Boas, D.A., 2012. A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. *Front. Neurosci.* 6.

- Cope, M., Delpy, D., 1988. System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Med. Biol. Eng. Comput.* 26 (3), 289–294.
- Cui, X., Bray, S., Reiss, A.L., 2010. Functional near infrared spectroscopy (fNIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *Neuroimage* 49 (4), 3039–3046.
- Culver, J., Durduran, T., Furuya, D., Cheung, C., Greenberg, J., Yodh, A., 2003a. Diffuse optical tomography of cerebral blood flow, oxygenation, and metabolism in rat during focal ischemia. *J. Cereb. Blood Flow Metab.* 23 (8), 911–924.
- Culver, J., Siegel, A., Stott, J., Boas, D., 2003b. Volumetric diffuse optical tomography of brain activity. *Opt. Lett.* 28 (21), 2061–2063.
- Custo, A., Boas, D., Tsuzuki, D., Dan, I., Mesquita, R., Fischl, B., Grimson, W., Wells, W., 2010. Anatomical atlas-guided diffuse optical tomography of brain activation. *Neuroimage* 49 (1), 561–567.
- Dunn, O., 1961. Multiple comparisons among means. *J. Am. Stat. Assoc.* 56 (293), 52–64.
- Fallgatter, A., Strik, W., 1998. Frontal brain activation during the wisconsin card sorting test assessed with two-channel near-infrared spectroscopy. *Eur. Arch. Psychiatry Clin. Neurosci.* 248 (5), 245–249.
- Fallgatter, A., Strik, W., 2000. Reduced frontal functional asymmetry in schizophrenia during a cued continuous performance test assessed with near-infrared spectroscopy. *Schizophr. Bull.* 26 (4), 913–919.
- Ferrari, M., Mottola, L., Quaresima, V., 2004. Principles, techniques, and limitations of near infrared spectroscopy. *Can. J. Appl. Physiol.* 29 (4), 463–487.
- Folley, B., Park, S., 2005. Verbal creativity and schizotypal personality in relation to prefrontal hemispheric laterality: a behavioral and near-infrared optical imaging study. *Schizophr. Res.* 80 (2–3), 271.
- Frederick, B., Nickerson, L., Tong, Y., 2012. Physiological denoting of BOLD fMRI data using Regressor Interpolation at Progressive Time Delays (RIPTiDe) processing of concurrent fMRI and near-infrared spectroscopy (NIRS). *Neuroimage* 60, 1913–1923.
- Friston, K.J., Holmes, A.P., Poline, J.-B., Price, C.J., Frith, C.D., 1996. Detecting activations in PET and fMRI: levels of inference and power. *Neuroimage* 40, 223–235.
- Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W., 2011. *Statistical Parametric Mapping: The Analysis of Functional Brain Images: The Analysis of Functional Brain Images*. Academic Press.
- Gagnon, L., Perdue, K., Greve, D., Goldenholz, D., Kaskhedikar, G., Boas, D., 2011. Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state-space modeling. *Neuroimage* 56 (3), 1362–1371.
- Gagnon, L., Cooper, R.J., cel, M.A.Y., Perdue, K.L., Greve, D.N., Boas, D.A., 2012. Short separation channel location impacts the performance of short channel regression in NIRS. *Neuroimage* 59 (3), 2518–2528.
- Gagnon, L., Yücel, M.A., Boas, D.A., Cooper, R.J., 2013. Further improvement in reducing superficial contamination in nirs using double short separation measurements. *Neuroimage*. <http://dx.doi.org/10.1016/j.neuroimage.2013.01.073>.
- Germon, T., Kane, N., Manara, A., Nelson, R., 1994. Near-infrared spectroscopy in adults: effects of extracranial ischaemia and intracranial hypoxia on estimation of cerebral oxygenation. *Br. J. Anaesth.* 73 (4), 503–506.
- Germon, T., Evans, P., Barnett, N., Wall, P., Manara, A., Nelson, R., 1999. Cerebral near infrared spectroscopy: emitter-detector separation must be increased. *Br. J. Anaesth.* 82 (6), 831–837.
- Graser, H., Smith, S., Tier, B., 1987. A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J. Anim. Sci.* 64 (5), 1362–1370.
- Hall, M., Chaudhary, U., Rey, G., Godavarty, A., 2013. Fronto-temporal mapping and connectivity using NIRS for language-related paradigms. *J. Neurolinguistics* 26 (1), 178–194.
- Harville, D., 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* 72 (358), 320–338.
- Henderson Jr., C., 1973. Sire evaluation and genetic trends. *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr J. Lush*, pp. 10–41.
- Henderson Jr., C., 1982. Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. *Biometrics* 623–640.
- Herrmann, M., Ehli, A., Fallgatter, A., 2003. Frontal activation during a verbal-fluency task as measured by near-infrared spectroscopy. *Brain Res. Bull.* 61 (1), 51–56.
- Herrmann, M., Ehli, A., Fallgatter, A., 2004. Bilaterally reduced frontal activation during a verbal fluency task in depressed patients as measured by near-infrared spectroscopy. *J. Neuropsychiatry Clin. Neurosci.* 16 (2), 170–175.
- Hofmann, M., Herrmann, M., Dan, I., Obrig, H., Conrad, M., Kuchinke, L., Jacobs, A., Fallgatter, A., 2008. Differential activation of frontal and parietal regions during visual word recognition: an optical topography study. *Neuroimage* 40 (3), 1340–1349.
- Homae, F., Watanabe, H., Otake, T., Nakano, T., Go, T., Konishi, Y., Taga, G., 2010. Development of global cortical networks in early infancy. *J. Neurosci.* 30 (14), 4877–4882.
- Homae, F., Watanabe, H., Nakano, T., Taga, G., 2011. Large-scale brain networks underlying language acquisition in early infancy. *Front. Psychol.* 2.
- Hoshi, Y., 2003. Functional near-infrared optical imaging: utility and limitations in human brain mapping. *Psychophysiology* 40 (4), 511–520.
- Hoshi, Y., Kobayashi, N., Tamura, M., 2001. Interpretation of near-infrared spectroscopy signals: a study with a newly developed perfused rat brain model. *J. Appl. Physiol.* 90 (5), 1657–1662.
- Hoshi, Y., Tsou, B., Billock, V., Tanosaki, M., Iguchi, Y., Shimada, M., Shinba, T., Yamada, Y., Oda, I., 2003. Spatiotemporal characteristics of hemodynamic changes in the human lateral prefrontal cortex during working memory tasks. *Neuroimage* 20 (3), 1493–1504.
- Huppert, T., Hoge, R., Diamond, S., Franceschini, M., Boas, D., 2006. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* 29 (2), 368–382.
- Huppert, T., Diamond, S., Franceschini, M., Boas, D., 2009. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.* 48 (10), D280–D298.
- Irani, F., Platek, S.M., Bunce, S., Ruocco, A.C., Chute, D., 2007. Functional near infrared spectroscopy (fNIRS): an emerging neuroimaging technology with important applications for the study of brain disorders. *Clin. Neuropsychol.* 21 (1), 9–37.
- Isobe, K., Kusaka, T., Nagano, K., Okubo, K., Yasuda, S., Kondo, M., Itoh, S., Onishi, S., 2001. Functional imaging of the brain in sedated newborn infants using near infrared topography during passive knee movement. *Neurosci. Lett.* 299 (3), 221–224.
- Izzetoglu, M., Devaraj, A., Bunce, S., Onaral, B., 2005. Motion artifact cancellation in NIR spectroscopy using Wiener filtering. *IEEE Trans. Biomed. Eng.* 52 (5), 934–938.
- Izzetoglu, M., Chitrappu, P., Bunce, S., Onaral, B., 2010. Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. *Biomed. Eng. Online* 9 (1), 16.
- Jang, K.E., Tak, S., Jung, J., Jang, J., Jeong, Y., Ye, J.C., 2009. Wavelet minimum description length detrending for near-infrared spectroscopy. *J. Biomed. Opt.* 14 (3), 034004.
- Jobsis, F.F., 1977. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* 198, 1264–1267.
- Jung, J., Lee, O., Ye, J., 2012. Source localization approach for functional DOT using MUSIC and FDR control. *Opt. Express* 20 (6), 6267–6285.
- Kameyama, M., Fukuda, M., Yamagishi, Y., Sato, T., Uehara, T., Ito, M., Suto, T., Mikuni, M., 2006. Frontal lobe function in bipolar disorder: a multichannel near-infrared spectroscopy study. *Neuroimage* 29 (1), 172–184.
- Katura, T., Sato, H., Fuchino, Y., Yoshida, T., Atsumori, H., Kiguchi, M., Maki, A., Abe, M., Tanaka, N., 2008. Extracting task-related activation components from optical topography measurement using independent components analysis. *J. Biomed. Opt.* 13 (5) (054008-054008).
- Kennan, R.P., Kim, D., Maki, A., Koizumi, H., Constable, R.T., 2002. Non-invasive assessment of language lateralization by transcranial near infrared optical topography and functional MRI. *Hum. Brain Mapp.* 16 (3), 183–189.
- Kenward, M., Roger, J., 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 983–997.
- Keuls, M., 1952. The use of the studentized range in connection with an analysis of variance. *Euphytica* 1 (2), 112–122.
- Kim, M.N., Durduran, T., Frangos, S., Edlow, B.L., Buckley, E.M., Moss, H.E., Zhou, C., Yu, G., Choe, R., Maloney-Wilensky, E., Wolf, R.L., Grady, M.S., Greenberg, J.H., Levine, J.M., Yodh, A.G., Detre, J.A., Kofke, W.A., 2009. Noninvasive measurement of cerebral blood flow and blood oxygenation using near-infrared and diffuse correlation spectroscopies in critically brain-injured adults. *Neurocrit. Care* 12 (2), 173–180.
- Kleinschmidt, A., Obrig, H., Requardt, M., Merboldt, K., Dirnagl, U., Villringer, A., Frahm, J., 1996. Simultaneous recording of cerebral blood oxygenation changes during human brain activation by magnetic resonance imaging and near-infrared spectroscopy. *J. Cereb. Blood Flow Metab.* 16 (5), 817–826.
- Koh, P.H., Glaser, D.E., Flandin, G., Kiebel, S., Butterworth, B., Maki, A., Delpy, D.T., Elwell, C.E., 2007. Functional optical signal analysis: a software tool for near-infrared spectroscopy data processing incorporating statistical parametric mapping. *J. Biomed. Opt.* 12, 1–13.
- Kohno, S., Miyai, I., Seiyama, A., Oda, I., Ishikawa, A., Tsuneishi, S., Amita, T., Shimizu, K., 2007. Removal of the skin blood flow artifact in functional near-infrared spectroscopic imaging data through independent component analysis. *J. Biomed. Opt.* 12 (6), 062111.
- Kozel, F.A., Tian, F., Dhamne, S., Croarkin, P.E., McClintock, S.M., Elliott, A., Mapes, K.S., Husain, M.M., Liu, H., 2009. Using simultaneous repetitive Transcranial Magnetic Stimulation/functional Near Infrared Spectroscopy (rTMS/fNIRS) to measure brain activation and connectivity. *Neuroimage* 47 (4), 1177–1184.
- Lee, O., Kim, J.M., Bresler, Y., Ye, J.C., 2011. Compressive diffuse optical tomography: noniterative exact reconstruction using joint sparsity. *IEEE Trans. Med. Imaging* 30 (5), 1129–1142.
- Lee, H., Kang, H., Chung, M.K., Kim, B.-N., Lee, D.S., 2012. Persistent brain network homology from the perspective of dendrogram. *IEEE Trans. Med. Imaging* 31 (12), 2267–2277.
- Li, H., Tak, S., Ye, J.C., 2012. Lipschitz–Killing curvature based expected Euler characteristics for p-value correction in fNIRS. *J. Neurosci. Methods* 204 (1), 61–67.
- Lina, J., Matteau-Pelletier, C., Dehaes, M., Desjardins, M., Lesage, F., 2010. Wavelet-based estimation of the hemodynamic responses in diffuse optical imaging. *Med. Image Anal.* 14 (4), 606–616.
- Lloyd-Fox, S., Blasi, A., Elwell, C., 2010. Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neurosci. Biobehav. Rev.* 34 (3), 269–284.
- Markham, J., White, B.R., Zeff, B.W., Culver, J.P., 2009. Blind identification of evoked human brain activity with independent component analysis of optical data. *Hum. Brain Mapp.* 30 (8), 2382–2392.
- Matsuo, K., Kato, N., Kato, T., 2002. Decreased cerebral haemodynamic response to cognitive and physiological tasks in mood disorders as shown by near-infrared spectroscopy. *Psychol. Med.* 32 (06), 1029–1037.
- Matsuo, K., Taneichi, K., Matsumoto, A., Ohtani, T., Yamasue, H., Sakano, Y., Sasaki, T., Sadamatsu, M., Kasai, K., Iwanami, A., et al., 2003. Hypoactivation of the prefrontal cortex during verbal fluency test in PTSD: a near-infrared spectroscopy study. *Psychiatry Res. Neuroimaging* 124 (1), 1–10.
- Medvedev, A., Kainerstorfer, J., Borisov, S., VanMeter, J., 2011. Functional connectivity in the prefrontal cortex measured by near-infrared spectroscopy during ultrarapid object recognition. *J. Biomed. Opt.* 16 (1), 016008.
- Mehagnoul-Schippier, D., van der Kallen, B., Colier, W., van der Sluijs, M., van Erning, L., Thijssen, H., Oeseburg, B., Hoefnagels, W., Jansen, R., 2002. Simultaneous measurements of cerebral oxygenation changes during brain activation by near-infrared spectroscopy and functional magnetic resonance imaging in healthy young and elderly subjects. *Hum. Brain Mapp.* 16 (1), 14–23.

- Mesquita, R., Franceschini, M., Boas, D., 2010. Resting state functional connectivity of the whole head with near-infrared spectroscopy. *Biomed. Opt. Express* 1 (1), 324–336.
- Miller, R., 1981. *Simultaneous Statistical Inference*. Springer-Verlag Inc.
- Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., Dupoux, E., 2011. Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cereb. Cortex* 21 (2), 254–261.
- Molavi, B., Dumont, G.A., 2012. Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiol. Meas.* 33 (2), 259–270.
- Murata, Y., Sakatani, K., Katayama, Y., Fukaya, C., 2002. Increase in focal concentration of deoxyhaemoglobin during neuronal activity in cerebral ischaemic patients. *Br. Med. J.* 73 (2), 182.
- Newman, D., 1939. The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. *Biometrika* 31 (1/2), 20–30.
- Niu, H., Wang, J., Zhao, T., Shu, N., He, Y., 2012. Revealing topological organization of human brain functional networks with resting-state functional near infrared spectroscopy. *PLoS One* 7 (9), e45771 (Sep.).
- Okamoto, M., Dan, H., Shimizu, K., Takeo, K., Amita, T., Oda, I., Konishi, I., Sakamoto, K., Isobe, S., Suzuki, T., et al., 2004. Multimodal assessment of cortical activation during apple peeling by NIRS and fMRI. *Neuroimage* 21 (4), 1275–1288.
- Okamoto, M., Matsunami, M., Dan, H., Kohata, T., Kohyama, K., Dan, I., 2006. Prefrontal activity during taste encoding: an fNIRS study. *Neuroimage* 31 (2), 796–806.
- Patel, S., Katura, T., Maki, A., Tachtsidis, I., 2011. Quantification of systemic interference in optical topography data during frontal lobe and motor cortex activation: an independent component analysis. *Oxygen Transport to Tissue XXXII*. Springer, pp. 45–51.
- Patterson, H., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58 (3), 545–554.
- Plichta, M., Herrmann, M., Baehne, C., Ehli, A., Richter, M., Pauli, P., Fallgatter, A., 2006. Event-related functional near-infrared spectroscopy (fNIRS): are the measurements reliable? *Neuroimage* 31 (1), 116–124.
- Plichta, M.M., Heinzel, S., Ehli, A.C., Pauli, P., Fallgatter, A.J., 2007. Model-based analysis of rapid event-related functional near-infrared spectroscopy (NIRS) data: a parametric validation study. *Neuroimage* 35, 625–634.
- Pogue, B., McBride, T., Prewitt, J., Österberg, U., Paulsen, K., 1999. Spatially variant regularization improves diffuse optical tomography. *Appl. Opt.* 38 (13), 2950–2961.
- Quaresima, V., Bisconti, S., Ferrari, M., 2012. A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain Lang.* 121 (2), 79–89.
- Rao, C., Toutenburg, H., 1999. *Linear Models: Least Squares and Alternatives*. Springer.
- Robertson, F.C., Douglas, T.S., Meintjes, E.M., 2010. Motion artifact removal for functional near infrared spectroscopy: a comparison of methods. *IEEE Trans. Biomed. Eng.* 57 (6), 1377–1387.
- Robinson, G., 1991. That BLUP is a good thing: the estimation of random effects. *Stat. Sci.* 6 (1), 15–32.
- Saager, R., Berger, A., 2005. Direct characterization and removal of interfering absorption trends in two-layer turbid media. *JOSA A* 22 (9), 1874–1882.
- Saager, R., Berger, A., 2008. Measurement of layer-like hemodynamic trends in scalp and cortex: implications for physiological baseline suppression in functional near-infrared spectroscopy. *J. Biomed. Opt.* 13 (3) (034017-034017).
- Saager, R.B., Telleri, N.L., Berger, A.J., 2011. Two-detector corrected near infrared spectroscopy (c-NIRS) detects hemodynamic activation responses more robustly than single-detector NIRS. *Neuroimage* 55 (4), 1679–1685.
- Saitou, H., Yanagi, H., Hara, S., Tsuchiya, S., Tomura, S., 2000. Cerebral blood volume and oxygenation among poststroke hemiplegic patients: effects of 13 rehabilitation tasks measured by near-infrared spectroscopy. *Arch. Phys. Med. Rehabil.* 81 (10), 1348–1356.
- Sankoh, A., Huque, M., Dubey, S., 1997. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Stat. Med.* 16 (22), 2529–2542.
- SAS Institute, 1985. *SAS User's Guide: Statistics*, vol. 2. SAS Inst.
- Sasai, S., Homae, F., Watanabe, H., Taga, G., 2011. Frequency-specific functional connectivity in the brain during resting state revealed by NIRS. *Neuroimage* 56 (1), 252–257.
- Sasai, S., Homae, F., Watanabe, H., Sasaki, A.T., Tanabe, H.C., Sadato, N., Taga, G., 2012. A NIRS-fMRI study of resting state network. *Neuroimage* 63 (1), 179–193.
- Sato, H., Tanaka, N., Uchida, M., Hirabayashi, Y., Kanai, M., Ashida, T., Konishi, I., Maki, A., 2006. Wavelet analysis for detecting body-movement artifacts in optical topography signals. *Neuroimage* 33 (2), 580–587.
- Scheffe, H., 1959. *The Analysis of Variance*. Wiley.
- Scholkopf, F., Spight, S., Muehleemann, T., Wolf, M., 2010. How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation. *Physiol. Meas.* 31 (5), 649–662.
- Schroeter, M., Zysset, S., Kupka, T., Kruggel, F., Von Cramon, D., 2002. Near-infrared spectroscopy can detect brain activity during a color-word matching Stroop task in an event-related design. *Hum. Brain Mapp.* 17 (1), 61–71.
- Schroeter, M.L., Bucheler, M.M., Müller, K., Uludag, K., Obrig, H., Lohmann, G., Tittgemeyer, M., Villringer, A., von Cramon, D.Y., 2004. Towards a standard analysis for functional near-infrared imaging. *Neuroimage* 21, 283–290.
- Searle, S., 1979. Notes on variance component estimation: a detailed account of maximum likelihood and kindred methodology. *Biometrics Unit, Cornell University*.
- Shibuya-tayoshi, S., Sumitani, S., Kikuchi, K., Tanaka, T., Tayoshi, S., Ueno, S., Ohmori, T., 2007. Activation of the prefrontal cortex during the trail-making test detected with multichannel near-infrared spectroscopy. *Psychiatry Clin. Neurosci.* 61 (6), 616–621.
- Shimada, S., Hiraki, K., 2006. Infant's brain responses to live and televised action. *Neuroimage* 32 (2), 930–939.
- Siegel, A., Culver, J., Mandeville, J., Boas, D., 2003. Temporal comparison of functional brain imaging with diffuse optical tomography and fMRI during rat forepaw stimulation. *Phys. Med. Biol.* 48 (10), 1391–1403.
- Singh, A., Dan, I., 2006. Exploring the false discovery rate in multichannel NIRS. *Neuroimage* 33 (2), 542–549.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niaz, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23, S208–S219.
- Strangman, G., 2009. Near-infrared neuroimaging with NinPy. *Front. Neuroinform.* 3.
- Strangman, G., Culver, J., Thompson, J., Boas, D., 2002. A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *Neuroimage* 17 (2), 719–731.
- Suto, T., Fukuda, M., Ito, M., Uehara, T., Mikuni, M., 2004. Multichannel near-infrared spectroscopy in depression and schizophrenia: cognitive brain activation study. *Biol. Psychiatry* 55 (5), 501–511.
- Tachtsidis, I., Elwell, C., Leung, T., Lee, C., Smith, M., Delpy, D., 2004. Investigation of cerebral haemodynamics by near-infrared spectroscopy in young healthy volunteers reveals posture-dependent spontaneous oscillations. *Physiol. Meas.* 25 (2), 437.
- Taga, G., Asakawa, K., Maki, A., Konishi, Y., Koizumi, H., 2003. Brain imaging in awake infants by near-infrared optical topography. *Proc. Natl. Acad. Sci. U. S. A.* 100 (19), 10722.
- Tak, S., Jang, J., Lee, K., Ye, J.C., 2010. Quantification of CMRO₂ without hypercapnia using simultaneous near-infrared spectroscopy and fMRI measurements. *Phys. Med. Biol.* 55 (11), 3249–3269.
- Tak, S., Yoon, S.J., Jang, J., Yoo, K., Jeong, Y., Ye, J.C., 2011. Quantitative analysis of hemodynamic and metabolic changes in subcortical vascular dementia using simultaneous near-infrared spectroscopy and fMRI measurements. *Neuroimage* 55 (1), 176–184.
- Taylor, J., Worsley, K., 2007. Detecting sparse signals in random fields, with an application to brain mapping. *J. Am. Stat. Assoc.* 102 (479), 913–928.
- Tong, Y., Frederick, B., 2010. Time lag dependent multimodal processing of concurrent fMRI and near-infrared spectroscopy (NIRS) data suggests a global circulatory origin for low-frequency oscillation signals in human brain. *Neuroimage* 53 (2), 553–564.
- Tong, Y., Bergethon, P.R., Frederick, B., 2011a. An improved method for mapping cerebrovascular reserve using concurrent fMRI and near-infrared spectroscopy with Regressor Interpolation at Progressive Time Delays (RIPTiDe). *Neuroimage* 56 (4), 2047–2057.
- Tong, Y., Lindsey, K., Frederick, B., 2011b. Partitioning of physiological noise signals in the brain with concurrent near-infrared spectroscopy and fMRI. *J. Cereb. Blood Flow Metab.* 31 (12), 2352–2362.
- Tsujimoto, S., Yamamoto, T., Kawaguchi, H., Koizumi, H., Sawaguchi, T., 2004. Prefrontal cortical activation associated with working memory in adults and preschool children: an event-related optical topography study. *Cereb. Cortex* 14 (7), 703–712.
- Umeyama, S., Yamada, T., 2009. Monte carlo study of global interference cancellation by multidistance measurement of near-infrared spectroscopy. *J. Biomed. Opt.* 14 (6) (064025-064025).
- Villringer, A., Dirnagl, U., 1995. Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging. *Cerebrovasc. Brain Metab. Rev.* 7, 240–276.
- Villringer, A., Planck, J., Hock, C., Schleinkofer, L., Dirnagl, U., 1993. Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neurosci. Lett.* 154 (1), 101–104.
- Virtanen, J., Noponen, T., Meriläinen, P., 2009. Comparison of principal and independent component analysis in removing extracerebral interference from near-infrared spectroscopy signals. *J. Biomed. Opt.* 14 (5) (054032-054032).
- Virtanen, J., Noponen, T., Kotilahti, K., Virtanen, J., Ilmoniemi, R.J., 2011. Accelerometer-based method for correcting signal baseline changes caused by motion artifacts in medical near-infrared spectroscopy. *J. Biomed. Opt.* 16 (8), 087005.
- White, B., Snyder, A., Cohen, A., Petersen, S., Raichle, M., Schlaggar, B., Culver, J., 2009. Resting-state functional connectivity in the human brain revealed with diffuse optical tomography. *Neuroimage* 47 (1), 148–156.
- Wilcox, T., Bortfeld, H., Woods, R., Wruck, E., Boas, D.A., 2008. Hemodynamic response to featural changes in the occipital and inferior temporal cortex in infants: a preliminary methodological exploration. *Dev. Sci.* 11 (3), 361–370.
- Wilcox, T., Haslup, J.A., Boas, D.A., 2010. Dissociation of processing of featural and spatiotemporal information in the infant cortex. *Neuroimage* 53 (4), 1256–1263.
- Woolrich, M., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S., et al., 2009. Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45 (1 Suppl.), S173–S186.
- Worsley, K., 1995. Boundary corrections for the expected euler characteristic of excursion sets of random fields, with an application to astrophysics. *Adv. Appl. Probab.* 943–959.
- Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited-again. *Neuroimage* 2, 173–181.
- Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A.C., 2002. A general statistical analysis for fMRI data. *Neuroimage* 15 (1), 1–15.
- Yamada, T., Umeyama, S., Matsuda, K., 2009. Multidistance probe arrangement to eliminate artifacts in functional near-infrared spectroscopy. *J. Biomed. Opt.* 14 (6) (064034-064034).
- Ye, J., Webb, K., Bouman, C., Millane, R., 1999. Optical diffusion tomography by iterative-coordinate-descent optimization in a Bayesian framework. *JOSA A* 16 (10), 2400–2412.
- Ye, J.C., Bouman, C.A., Webb, K.J., Millane, R.P., 2001. Nonlinear multigrid algorithms for Bayesian optical diffusion tomography. *IEEE Trans. Image Process.* 10 (6), 909–922.
- Ye, J., Tak, S., Jang, K., Jung, J., Jang, J., 2009. NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *Neuroimage* 44, 428–447.

- Young, A., Germon, T., Barnett, N., Manara, A., Nelson, R., 2000. Behaviour of near-infrared light in the adult human head: implications for clinical near-infrared spectroscopy. *Br. J. Anaesth.* 84 (1), 38–42.
- Yücel, M.A., Huppert, T.J., Boas, D.A., Gagnon, L., 2012. Calibrating the BOLD signal during a motor task using an extended fusion model incorporating DOT, BOLD and ASL data. *Neuroimage* 61 (4), 1268–1276.
- Zeff, B., White, B., Dehghani, H., Schlaggar, B., Culver, J., 2007. Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. *Proc. Natl. Acad. Sci.* 104 (29), 12169–12174.
- Zhang, Y., Brooks, D.H., Franceschini, M.A., Boas, D.A., 2005. Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *J. Biomed. Opt.* 10 (1), 011014.
- Zhang, Q., Brown, E., Strangman, G., 2007. Adaptive filtering for global interference cancellation and real-time recovery of evoked brain activity: a Monte Carlo simulation study. *J. Biomed. Opt.* 12 (4) (044014-044014).
- Zhang, Q., Strangman, G., Ganis, G., 2009. Adaptive filtering to reduce global interference in non-invasive nirs measures of brain activation: how well and when does it work? *Neuroimage* 45 (3), 788–794.
- Zhang, H., Zhang, Y.-J., Lu, C.-M., Ma, S.-Y., Zang, Y.-F., Zhu, C.-Z., 2010a. Functional connectivity as revealed by independent component analysis of resting-state fNIRS measurements. *Neuroimage* 51 (3), 1150–1161.
- Zhang, Y., Lu, C., Biswal, B., Zang, Y., Peng, D., Zhu, C., 2010b. Detecting resting-state functional connectivity in the language system using functional near-infrared spectroscopy. *J. Biomed. Opt.* 15 (4) (047003-047003).