Wavelet-MDL based detrending method for near infrared spectroscopy (NIRS)

Kwang Eun Jang, Sungho Tak, Jaeduck Jang, Jinwook Jung, and Jong Chul Ye[a]

Korea Advanced Institute of Science and Technology (KAIST)
373-1 Guseong-dong, Yuseong-gu, Daejeon, Republic of Korea.

ABSTRACT

Near infrared spectroscopy (NIRS) is a relatively new non-invasive brain imaging method to measure brain activities associated with regional changes of the oxy- and deoxy- hemoglobin concentration. Typically, functional MRI or PET data are analyzed using the general linear model (GLM), in which measurements are modeled as a linear combination of explanatory variables plus an error term. However, the GLM often fails in NIRS if there exists an unknown global trend due to breathing, cardiac, vaso- motion and other experimental errors. In order to overcome these problems, we propose a wavelet-MDL based detrending algorithm. Specifically, the wavelet transform is applied to NIRS measurements to decompose them into global trends, signals and uncorrelated noise components in distinct scales. In order to prevent the over-fitting the minimum length description (MDL) principle is applied. Experimental results demonstrate that the new detrending algorithm outperforms the conventional approaches.

Keywords: Near infrared spectroscopy, global bias, detrending, wavelet, minimum description length

1. INTRODUCTION

Near-infrared (NIR) light with a wavelength between 650nm and 950nm easily gets through biological tissue. This is because the light in NIR region is remarkably weakly absorbed by only a few biological chromophores such as hemoglobin, myoglobin and cytochrome c oxidase.[1] The relatively deep penetration depth for a human brain of NIR light allows us to measure brain activities associated with regional changes of the oxy- and deoxy-hemoglobin concentration.[2] This spectroscopic technique using NIR light for monitoring brain activities is called as functional near-infrared spectroscopy (NIRS).[3]

NIRS has a number of advantages over other neuroimaging modalities such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI).[3] NIRS has an excellent temporal resolution compared to fMRI or PET. Also, it does not require for a subject to lie on his/her back in the confined environment during experiments. Therefore, it allows us to investigate subjects who are difficult to examine using fMRI or PET, such as infants, children, and patients with psychological problems. Moreover, NIRS is highly flexible, portable and relatively low cost.

Unfortunately, compared to PET or fMRI, the quantitative statistical analysis of NIRS data is not relatively well-established. To overcome this, Schroeter et al[4] recently proposed standardized analysis for NIRS data using the generalized linear model (GLM), and Plichta et al[5] applied it for rapid event-related NIRS experiment design. In the GLM framework, which is a standard method for analyzing fMRI and PET data,[6,7] measurements are simply modelled as a linear combination of explanatory variables plus an error term. The GLM is robust to many cases even with a severe optical signal attenuation due to scattering or poor contact, since the GLM measures the temporal variational pattern of signals rather than their absolute magnitude.

However, the GLM often fails in NIRS if there is an unknown global bias or a trend due to breathing, cardiac and vaso- motion as well as patient movement during the measurement. In most cases of the brain monitoring, such long-term physiological drift are often observed. Moreover, the amplitude of the global drift is comparable to that of the signal from a brain activation.

[a] Corresponding Author: Jong Chul Ye, E-mail: yong.ye@kaist.ac.kr, Telephone: +82-42-869-4320.

Multimodal Biomedical Imaging III, edited by Fred S. Azar, Xavier Intes,
Proc. of SPIE Vol. 6850, 68500Y, (2008)
1605-7422/08/$18 · doi: 10.1117/12.764141

In order to overcome this problem, we employ a wavelet-based detrending algorithm[8] and adapt it for NIRS applications. Specifically, wavelet transform is applied to the NIRS time series to decompose them into bias, signal and uncorrelated noise components in distinct scales. This is because that trends do not vary greatly between short time intervals, so the drift signal can be modelled as a linear combination of large scale wavelets. In order to prevent the over-fitting of bias estimation in a wavelet basis, the minimum length description (MDL) principle[9] is applied. Experimental results with real NIRS measurements demonstrate that the new detrending algorithm outperforms the conventional approaches.

2. GENERAL LINEAR MODEL - A REVIEW

All classical analysis of functional data such as fMRI, PET, EEG and MEG are based on the general linear model (GLM).[6,7] In this section, a brief review for GLM is presented. For more detailed description, see.[10]

GLM describes a measurement y in terms of a linear combination of L explanatory variables plus an error term:

$$y = x_1\beta_1 + \cdots + x_L\beta_L + \epsilon. \tag{1}$$

Here the β_i's denote unknown strengths of response, x_i's are explanatory variables, which are models of hemodynamic responses. The error ϵ is assumed to be normally distributed with zero mean but possibly temporally correlated.

Since we have N sampled measurements along the time axis, a set of simultaneous equations is written as:

$$\begin{aligned} y_1 =& x_{11}\beta_1 + \cdots + x_{L1}\beta_L + \epsilon_1 \\ &\vdots \\ y_N =& x_{1N}\beta_1 + \cdots + x_{LN}\beta_L + \epsilon_N \end{aligned} \tag{2}$$

which can be written in a matrix notation as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3}$$

where $\mathbf{y}$ is a N-dimensional column vector whose elements are the sampled NIRS data at N time points, $\boldsymbol{\epsilon}$ denotes an error vector and $\boldsymbol{\beta}$ is a L dimensional column vector which represents unknown strengths of response. Usually, the $N \times L$ matrix $\mathbf{X}$ is called as a design matrix and serves as a predictor for the measured signal.[10]

Usually, the stick function or the boxcar function is used for stimulus function. For the HRF, there are a number of possible models. In this paper, we follow the tradition of fMRI and employ so called canonical HRF, which is characterized by two gamma functions.[10] Since the precise shape of the HRF varies across the brain, the two derivatives of HRF with respect to delay and dispersion has been used together to mitigate that problem.[11]

After the model specification, a least-squares parameter estimator is derived using ordinary least squares. If the design matrix X is of full rank, the least squares estimates are:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^-\mathbf{X}^T\mathbf{y} \tag{4}$$

where $\mathbf{X}^-$ denotes the pseudoinverse of $\mathbf{X}$. With the obtained least squares estimates, one can construct a statistics for the statistical inference. In most cases, we consider a linear combination of the parameter estimates:

$$\lambda_1\hat{\beta}_1 + \cdots + \lambda_L\hat{\beta}_L = \boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} \tag{5}$$

where the vector $\boldsymbol{\lambda}^T$ is called as a contrast vector.[10] Since the parameter estimates are normally distributed, a contrast is a random variable which is normally distributed.

Thus t-statistics for the hypothesis $H : \boldsymbol{\lambda}^T\boldsymbol{\beta} = c$ is given by

$$T = \frac{\boldsymbol{\lambda}^T\hat{\boldsymbol{\beta}} - c}{\sqrt{\hat{\sigma^2}\boldsymbol{\lambda}^T(\mathbf{X}^T\mathbf{X})^-\boldsymbol{\lambda}}} \tag{6}$$

where T denotes a random variable with a Student's t-distribution with $N - rank(\mathbf{X})$ degrees of freedom.[10]

3. WAVELET-MDL DETRENDING

In many cases, there often exists global drifts in the NIRS measurements due to variety reasons, including subject movement during the experiment, blood pressure variation, long-term physiological changes or instrumental instability. Moreover, the amplitude of a global drift is often comparable to that of the signal from a brain activation. In order to eliminate the global trend to improve the signal-to-noise ratio, the previous studies of NIRS often employed high-pass filtering[4] or autoregressive model.[5] However, the signals from brain activations can be often degraded by the simple filtering since the frequency response of HRF is concentrated in only extremely low frequency domain (0Hz $\sim$ 0.15Hz).[10]

For the case of fMRI, Meyer proposed a new detrending algorithm that includes the baseline drift as part of the linear model.[8]

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{7}$$

where $\boldsymbol{\theta}$ denotes the baseline drift and $\boldsymbol{\epsilon}$ is the additive noise with the temporal covariance matrix $\boldsymbol{\Sigma}$, respectively. According to the wavelet theory which allows a multi-resolution analysis, the unknown trend can be modeled as a signal which belongs to a subspace spanned by large scale wavelets since the trend varies slowly. In other words, the global drift is restricted in the low-resolution signal subspace. Base on this observation, a model for the trend is given by:[8]

$$\theta(t) = s\theta_0^J \Phi(2^{-J}t) + \sum_{j=J_0}^{J} \sum_{k=0}^{2^{-j}N-1} d\theta_k^j \psi(2^{-j}t - k) \tag{8}$$

where Φ, ψ denote the scaling function and the wavelet function associated with a multiresolution analysis, respectively. Here, we assumed $N = 2^J$ for the simplicity, and J_0 determines the complexity of the global drift.[12]

More instinctive notation can be achieved if we use the discrete wavelet transform. Let $\mathbf{W}$ denote the discrete wavelet transform, then the model for the trend is given by

$$\mathbf{W}\boldsymbol{\theta} = [s\theta_0^J, d\theta_0^J, \cdots, d\theta_{2^{-J_0}N-1}, 0, \cdots, 0]^T = [\boldsymbol{\Theta}^T, \mathbf{0}]^T, \tag{9}$$

where the number of non-zero coefficients that describe the trend is $2^{-J_0+1}N = n_0$. Under the constraint, the best fitting trend is derived in sense of the maximum likelihood estimation. Specifically, the maximum likelihood estimates for the trend θ and the unknown β is given by:[8]

$$[\hat{\boldsymbol{\Theta}}^T, \hat{\boldsymbol{\beta}}]^T = [\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A}]^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{y}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{I}_{n_0 \times n_0} & \\ & \mathbf{WX} \\ \mathbf{0}_{(N-n_0)\times n_0} & \end{bmatrix}. \tag{10}$$

where $\mathbf{I}$ denotes a $n_0 \times n_0$ identity matrix, $\mathbf{0}_{(N-n_0)\times n_0}$ denotes a matrix whose elements are zero, $\mathbf{A}$ is a $(N \times (n_0 + L))$ matrix, and $\boldsymbol{\Sigma}$ denotes the $N \times N$ noise covariance matrix, respectively.

To complete the estimation of the unknown drift, the complexity of the global drift, J_0 in Eq. (8) and Eq. (9), should be estimated. More specifically, in order to prevent the over-fitting or under-fitting of bias estimation in a wavelet basis, an appropriate complexity of the global drift - i.e. the number of non-zero coefficients (n_0)- should be determined. For the case of over-fitting, the hemodynamic response due to the brain activation is damaged during a detrending process, whereas in the under-fitted model the unknown trend is erroneously classified as a brain activation. The problem of estimating the correct order is called the model order selection problem.[9,13,14]

In general, a criterion for model order selection consists of two parts: a distance between data and a model, a penalty for the over-fitting. Balancing these two terms is crucial in preventing the over- or under-fitting. Akaike's information criterion (AIC) and Schwartz information criterion (SIC) are two popular models and they were originally used for wavelet-based detrending algorithm.[8] The improved version of AIC[13] and SIC[14] are given by

$$AIC_C = \log\hat{\sigma}^2 + \frac{N + n_0}{N - n_0 - 2}, \quad SIC = \frac{N}{2}\log\hat{\sigma}^2 + \frac{1}{2}n_0 \log N \tag{11}$$

where $\hat{\sigma}^2, N$ denotes the residual sum of squares and the number of data points respectively.

However, we observed that AIC and SIC tend to give a over-fitted model for NIRS signal. This is because NIRS measurements have a large number of data points thanks to the excellent temporal resolution of NIRS, and it is known that in AIC the probability of over-fitting is more than zero under quite general conditions as $N \to \infty$.[15] In fact, the number of temporal NIRS sampling is much larger than that of fMRI data, which may leads AIC and SIC to fail in NIRS problem.

We have observed that a model selection criterion based on minimum description length (MDL) principle,[9] which is widely used in denoising and compression,[16] gives desired tendency. According to MDL, the best model is the one whose description permits the shortest code length.[9,17] Under appropriate assumptions, the a criterion based on MDL principle for defining the complexity of the global trend is given as

$$MDL = \frac{N}{2}\log_2 \hat{\sigma^2} + \frac{1}{2}n_0 \log_2 N + (-n_0 \log_2 P(n_0)), \tag{12}$$

where $\hat{\sigma^2}$ denotes the ML estimate of unknown variance, $\log_2 P(n_0)$ denotes the *prior* distribution of the coefficients of the unknown trend along the wavelet domain. In the point of the noise suppression and the data compression view, the wavelet coefficients are uniformly distributed, i.e., $P(n_0) = 1/N$. However, as described before, coefficients in the coarser scale has higher probability to be included to the unknown trend signal. Therefore, it is possible to impose the characteristic of the trending signal to the model order selection criteria by using the non-increasing *prior* distribution. Instead of using subjective distributions, we employed so called universal prior for integers proposed by Rissanen.[17]

$$P(n) = 2^{-L^0(n)}, \quad n > 0, \quad L^0(n) = \log_2^* n + \log_2 c, \quad c \approx 2.865064. \tag{13}$$

where $\log_2^* n = \log_2 n + \log_2(\log_2 n) + \log_2(\log_2(\log_2 n)) + \cdots$, where the sum involves only the non-negative terms.

Since for NIRS measurements, the number of wavelet coefficients is quite different between scales, we calculated Eq. (10) by adding each coefficient within the scale to maximize the detection of the global minimum in the model order selection.

4. EXPERIMENTAL RESULTS

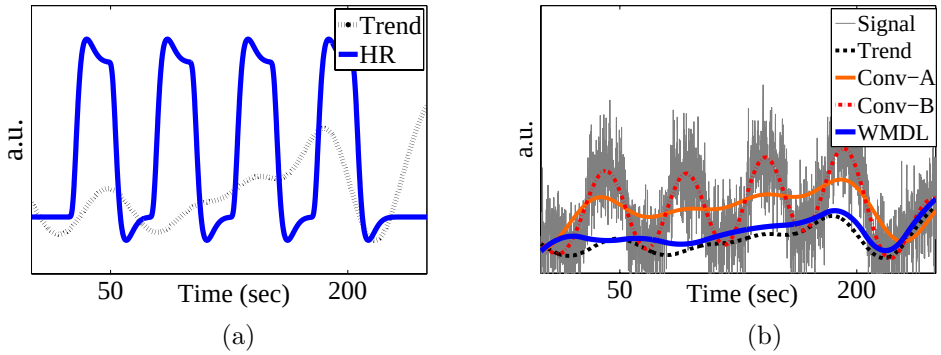
We performed NIRS experiments using Oxymon Mk III (Artinis, Netherlands) that has eight laser diodes and four detectors. In this system, two continuous wave lights (856nm and 781nm) are emitted at each source fiber. A suitable arrangement constructs 16 optodes and 24 pairs of a source and a detector. The source and the detector were attached by optical fibers on the scalp which covers motor cortex. The distance between source and detector was 3.5cm.

The modified Beer-Lambert law (MBLL) which describes an optical attenuation in a highly scattering medium like biological tissue[18] allows the transformation from raw optical density (OD) data to changes of chromophore concentrations. According to the MBLL, the change in the $OD(\lambda, r, t)$ at the wavelength λ from the cerebral cortex position r at time t due to the N_c number of chromophore concentration changes $\{\Delta c^{(i)}(r,t)\}_{i=1}^{N_c}$ is described as

$$\triangle OD(\lambda, r, t) = -ln(\frac{I_F}{I_o}) = \sum_{i=1}^{N_c} a_i(\lambda) \triangle c^{(i)}(r,t) d(r) l(r) \tag{14}$$

where I_F denote the final measured optical intensity, I_o denotes the initial measured optical intensity, $a_i(\lambda)$ is the extinction coefficient of the i-th chromophore at the wavelength λ, $d(r)$ is the differential pathlength factor (DPF) at the position r, $l(r)$ is the distance between the source-detector separation at the position r, respectively. By measuring $\triangle OD$ at one or more wavelengths per chromophore and using known extinction coefficients[19] $a(\lambda)$ of chromophores, concentration changes ΔC can be estimated.

One healthy, right-handed male adults were examined under right-finger tapping (RFT) task. The 21 seconds periods of activation were alternated with 30 seconds periods of rest. During the activation periods, subjects were asked to tap right hand fingers. Total recoding time was 552 seconds. Also, the same subject were examined



(a) (b)

Figure 1. (a) Synthetic hemodynamic response(HR) and global trend. The trend was extracted from the real NIRS experiment under the null-hypothesis condition. (b) Synthetic NIRS time series and the estimated global trends. Black line is the true bias. Blue line is the result of the wavelet-MDL based detrending method. Red and orange lines are from the conventional detrending methods.

under the same condition. In this case, during the recording period, the subject did not receive any stimulus. This data set which is called as noisy data is used for simulation study in the following. No subject had a history of any neurological disorder. All subjects were informed about the whole experiment process as well as environment. The investigation was approved by Institutional Review Board of KAIST.

4.1 Simulation Study

For the comparison, the conventional detrending based on high-pass filtering in[4] was implemented. We designed FIR low-pass filters by employing the Kaiser window which is known as a near-optimal method for constructing FIR filters.

A simulated NIRS time series consisted of the global bias, the hemodynamic response and the Gaussian noise. The global bias was extracted from the real NIRS experiment under the null-hypothesis condition using the conventional detrending method whose passband edge and the stopband edge was 0.02Hz and 0.025Hz, respectively. The hemodynamic response was modelled by the canonical HRF with its derivatives. The extracted trend and the synthetic brain signal was normalized to [0,1] and [0, 1.5]. The power ratio between the two signals was 2.25. The standard deviation of the Gaussian noise was 0.5. The simulated NIRS time series and its components are given in Fig. 1.

In Fig. 1(b), 4 global trends are illustrated. The black dashed line denotes the extracted trend which can be regarded as a true bias. The orange line (Conv-A, in the figure) is the result from the conventional detrending method whose passband edge and the stopband edge was 0.015Hz and 0.02Hz, respectively. The red dashed line (Conv-B, in the figure) is also from the exactly same low-pass filter which was used for synthesizing the *true* trend. The blue line is estimated by the wavelet-MDL based detrending method.

Even though we know the exact cutoff frequency for the low-pass filtering, it is hard to distinguish between the brain signal and the global trend when the both components are mixed together in the low-frequency band. Indeed, the hemodynamic response is a extreme low-frequency signal similar to the global bias. Conv-B shows this limitation clearly; it can remove the true bias perfectly, however, also it removes the most part of brain signals. If we adjust the cutoff frequency, such as Conv-A, better estimation is possible. However, we can easily see that the wavelet-MDL based detrending method gives much closer estimate for the unknown bias. Also, there is little damage to the brain signal. The strength of the proposed algorithm can be more clearly elucidated in the real application.

4.2 Experimental NIRS data

To gather the spatial information and to evaluate our NIRS experimental system, we simultaneously recorded fMRI and NIRS data. From an anatomical T1 image from MRI, we mapped locations of NIRS signals to the cerebral cortex. Specifically, we attached 4 tocopherol pills which make conspicuous spots in MR T1 image. A three-dimensional digitizer, MicroScribe G2 (Immersion corp., USA), was used to register the locations of

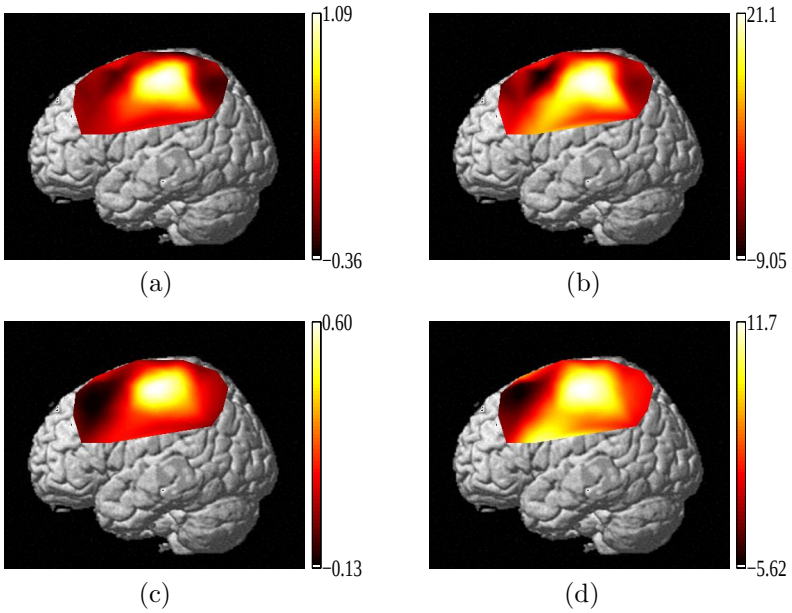


Figure 2. (a) and (b): Beta map and T map from the wavelet-MDL detrending, (c) and (d): Beta map and T map from the conventional detrending method. $\hat{\beta}_{max}$ is 1.0959 and 0.6035, respectively, in arbitrary units. $\hat{t}_{max}$ is 21.1554, 11.7819, respectively.

tocopherol and optodes. Finding the relationship between the real 3-D space and the MR image domain using pairs of measurements of the coordinates of a number of points in both systems is a well known problem called as absolute orientation. A closed-form, least-square solution for this problem is given by Horn.[20]

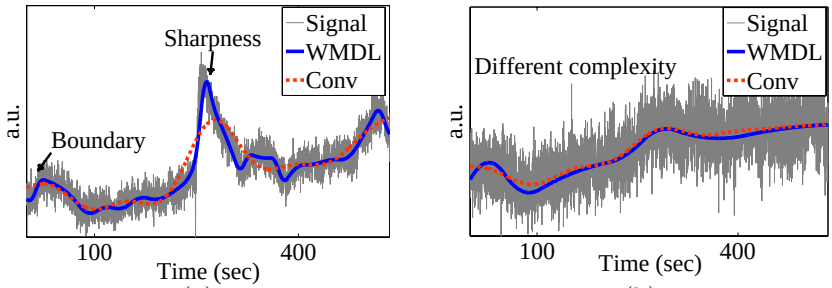
After the relation between the MR coordinate and the real 3-D coordinate is elicited based on the measured coordinates of tocopherol pills, the locations of optodes in the MR coordinate are calculated. The closest points on the cortical cortex to the optodes are found using the segmented T1 MR image.

Based on the estimated spatial coordinate, we constructed so called beta-maps and t-maps to see the brain activation under the right finger tapping task. Beta values are calculated by Eq. (4). T values are obtained by Eq. (6). The cubic interpolation is used for constructing beta- and T-maps. In Fig. 2, Beta- and T-map from both detrending methods are illustrated. In Fig. 2 task-related activation in the right primary motor cortex is clearly observable.

We compare the maximum value of $\hat{\beta}$ and $\hat{t}$ to evaluate the performance between the wavelet-MDL based detrending method and the conventional method. $\hat{\beta}_{max}$ is 1.0959 and 0.6035, respectively, in arbitrary units. Similarly, $\hat{t}_{max}$ is 21.1554, 11.7819, respectively. The higher $\hat{\beta}_{max}$ and $\hat{t}_{max}$ value indicates that the proposed method makes lower damage to the brain signal and performs more reliable detrending. In addition, more localized activation is observable especially in T-maps in Fig 2.

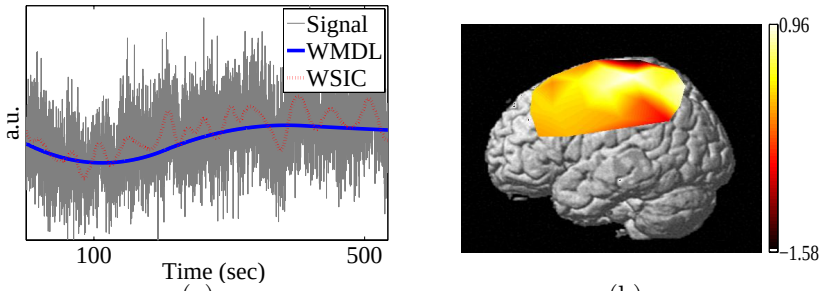
The NIRS time series in Fig. 3(a) contains a rapidly varied bias which is hard to remove by the simple low-pass filtering. If we use the conventional detrending method, the residual of the fast-varying trend can be transformed to a task-related signal, which is a general side-effect of the detrending process. We can show that the wavelet-MDL based detrending method allows to remove the unknown trend even if it has high-frequency components in the certain time point. In addition, there is no error for the boundary problem in the case of proposed method. Usually, the conventional approach suffers the boundary effect because of Gibbs phenomena. In Fig. 3(b), the estimated trend varies slowly, which shows that the complexity is successfully adjusted *automatically*. Also, we can see that the complexity of the estimated trend varies along the time axis.

We compared our model order selection criteria to SIC which is originally employed for the wavelet-based detrending method. In Fig. 4(a), it can be easily see that the estimated trend based on SIC does the overfitting. This overfitting makes it hard to localize the brain activation as illustrated in Fig. 4(b).



(a) (b)

Figure 3. Two NIRS measurements and their estimated global trends. (a) Wavelet-MDL based detrending method allows to remove the unknown trend even if it has high-frequency components in the certain time point. In addition, there is no error for the boundary problem which frequently occurs for conventional detrending methods. (b) Compare to (a), the estimated trend varies slowly, which shows that the complexity is successfully adjusted *automatically*.



(a) (b)

Figure 4. (a) For NIRS measurements, SIC tends to overfit. (b) Beta-map based on SIC. We can see the false localization of the brain activation due to the overfitting.

5. DISCUSSION

The GLM allows a robust analysis for NIRS time series. The power of the GLM is based on *a priori* knowledge of the pattern of NIRS signal. Therefore, the predictor of the GLM should be suitably chosen. The canonical HRF is used for this study and it shows a desired performance.

For complete construction of a binary hypothesis testing framework, a noise covariance matrix Σ should be included in t-statistics. The contrast t-statistics using in fMRI domain is given by[6,7]

$$t_{df} = \frac{\lambda^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 \lambda^T (X^T X)^- X^T \Sigma X (X^T X)^- \lambda}}. \tag{15}$$

In additional, the degree of freedom and the way to adjust the false-positive rate of t-statistics are required. In this study, noise of NIRS was assumed to be independent, identically distributed. However, in fact, there is a significant temporal correlations due to various physiological fluctuations. Even though the wavelet based detrending algorithm successfully removes the slow fluctuation, physiological noise still remains. The effect of temporal correlation due to physiological noise should be analyzed and removed. This is beyond the scope of this paper and will be reported elsewhere.

6. CONCLUSION

We applied the GLM to NIRS signal processing and confirmed its usefulness. We also observed that the wavelet-MDL based detrending method is a robust tool for analyzing NIRS time series. Specifically, wavelet transform was applied to the NIRS time series to decompose them into bias, signal and uncorrelated noise components in distinct scales. With the MDL criterion, the unknown drift signal in NIRS data was successfully removed. Experimental results with real and synthetic NIRS measurements demonstrated that the new detrending algorithm outperforms the conventional approaches.

7. ACKNOWLEDGMENTS

This research was supported by a Brain Neuroinformatics Research program by Korean Ministry of Commerce, Industry, and Energy. Also it was supported in part by Radiation and Nuclear Medical Engineering Research Center (Globalization Project), Korea Advanced Institute of Science and Technology.

REFERENCES

1. F. F. Jobsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science* **198**(4323), pp. 1264–1267, 1977.
2. S. G. Diamond, T. J. Huppert, V. Kolehmainen, M. A. Franceschini, J. P. Kaipio, S. R. Arridge, and D. A. Boas, "Dynamic physiological modeling for functional diffuse optical tomography," *NeuroImage* **30**, pp. 88–101, March 2006.
3. Y. Hoshi, "Functional near-infrared optical imaging: Utility and limitations in human brain mapping," *Psychophysiology* **40**, pp. 511–520, July 2003.
4. M. L. Schroeter, M. M. Bucheler, K. Muller, K. Uludag, H. Obrig, G. Lohmann, M. Tittgemeyer, A. Villringer, and D. Y. von Cramon, "Towards a standard analysis for functional near-infrared imaging,"
5. M. M. Plichta, S. Heinzel, A. C. Ehlis, P. Pauli, and A. J. Fallgatter, "Model-based analysis of rapid event-related functional near-infrared spectroscopy (nirs) data: A parametric validation study," *NeuroImage* **35**, pp. 625–634, April 2007.
6. K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. J. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping* **2**(4), pp. 189–210, 1995.
7. K. J. Worsley and K. J. Friston, "Analysis of fmri time-series revisited-again," *NeuroImage* **2**, pp. 173–181, September 1995.
8. F. G. Meyer, "Wavelet-based estimation of a semiparametric generalized linear model of fmri time-series," *IEEE Trans. Med. Imag.* **22**, pp. 315–322, March 2003.
9. J. Rissanen, "Modeling by shortest data description," *Automatica* **41**, pp. 465–471, 1978.
10. K. J. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Academic Press, London, 2006.
11. R. N. A. Henson, C. J. Price, M. D. Rugg, R. Turner, and K. J. Friston, "Detecting latency differences in event-related bold responses: Application to words versus nonwords and initial versus repeated face presentations," *NeuroImage* **15**(1).
12. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New york, 1999.
13. K. L. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," *Biometrika* **76**(2), pp. 297–307, 1989.
14. G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.* **6**, pp. 461–464, 1978.
15. P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine* **21**, pp. 36–47, July 2004.
16. N. Saito, "Simultaneous noise suppression and signal compression using a library of orthonormal bases and the minimum description length criterion," *Wavelets in Geophysics* , pp. 299–324, 1994.
17. J. Rissanen, "A universal prior for integers and estimation by minimum description length," *The Annals of Statistics* **11**(2).
18. M. Cope and D. T. Delpy, "System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination," *Medical and biological engineering and computing* **26**(3).
19. S. Wray, M. Cope, D. T. Delpy, J. S. Wyatt, and E. O. Reynolds, "Characterization of the near infrared absorption spectra of cytochrome aa3 and haemoglobin for the non-invasive monitoring of cerebral oxygenation," *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **933**, pp. 184–192, March 1988.
20. B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Amer. A* **4**, pp. 629–642, Apri 1987.