MULTILEVEL STATISTICAL INFERENCE FROM FUNCTIONAL NEAR INFRARED SPECTROSCOPY SIGNALS

Koray Ciftci, Bülent Sankur**, Yasemin P. Kahya**, Ata Akın**
Boğaziçi University *Institute of Biomedical Engineering., **Department of Electrical and Electronic Engineering

ABSTRACT

Functional near infrared spectroscopy (fNIRS) is a technique that tries to detect cognitive activity by measuring changes in the concentrations of the oxygenated and deoxygenated hemoglobin in the brain. We develop Bayesian statistical tools for making multilevel inferences, that is, inferences generalizable to a population within the context of fNIRS neuroimaging problem. Specifically, we present a method for multilevel modeling of fNIRS signals using a hierarchical general linear model. The model is treated in the context of Bayesian networks. Experimental results of a cognitive task (Stroop test) are presented with comparison to classical approaches.

1. INTRODUCTION

Functional near infrared spectroscopy (fNIRS) is a non-invasive method to monitor brain activation by measuring changes in the concentrations of oxygenated and deoxygenated hemoglobin (Hb) [1]. Differences in the absorption spectra of oxy- and deoxy-Hb allows us to measure the transmitted and received near infrared light in multiple wavelengths and calculate the relative concentrations of these chromospheres. fNIRS has significant advantages like, lack of radiation, portable nature of the device, relative easiness and low cost of the procedure. Although fNIRS is successfully employed in a number of physiological measurements it is still a challenging task to detect cognitive activity by fNIRS.

Statistical analysis of neuroimaging data is interested in detecting the activation pattern of the brain under a particular kind of stimulus. Statistical inference generally handles a group of subjects to deduce results attributable to that particular population. Summary statistics approach [2] is an efficient way to arrive at results generalisable to a population. Working with summary statistics substantially decreases the computational burden of the analysis because each level accepts as input the parameters estimated from the previous level (or observed data) and each unit in a level (session, subject etc.) may be analyzed isolated from the rest of the units at the same level. However, it is not always clear what the correct statistics are that summarize the data. Indeed, whether it is possible to work with summary statistics or not is still an open problem.

General linear model (GLM) is the most commonly used tool to make inferences from functional magnetic resonance imaging data [3]. Albeit some differences between the two techniques, fNIRS, also tries to detect brain haemodynamic activity, which is based upon neurovascular coupling theory. Thus, it would be logical to test the validity of the GLM for fNIRS signals. Schroeter et al. [4] was one of the first groups to apply GLM for fNIRS signals. Using a visual stimulus, they arrived at the conclusion that GLM is feasible especially for deoxy-Hb.

The main problem we address in this study is to make multilevel inference from fNIRS signals, i.e. we want to infer on upper level (group, groups of groups etc.) parameters, having observed some set of measurement data. We use a hierarchical GLM to link measurement space to the upper level parameters. We propose to use Bayesian networks [5] to define structural and functional relationships between the variables of the model. This will allow us to arrive at posterior probability distributions for the parameters that we want to infer on.

The particular experimental protocol that we used in this study is a variant of Stroop task which is known to be a good activator of prefrontal cortex [6,7]. We used the version of the Stroop task introduced by Zysset et al. [8], since it provides a way to separate interference that takes place at the conceptual level from the response preparation and also it is very suitable for computerized application of the test. We intended to show that fNIRS can detect cognitive activity in the prefrontal cortex using suitable statistical tools. Thus, this study may be seen as an effort to make a contribution to the ongoing studies to clarify the potential of fNIRS for cognitive activity monitoring.

2. METHODS

2.1 Preliminaries

We analyzed fNIRS data with the familiar two-level GLM. In the subsequent analysis we will call first level as the "subject level" and second level as the "group level", to be more explicit. More levels can be added to the model if need be, for example, for making between-session or between-group analysis. However, the arguments will be similar. The model is mass univariate, i.e. each detector's signal is analyzed separately. The first level models the within-subject effects:

$$Y_k = X_k B_k + E_k \tag{1}$$

where Y_k is the $N \times 1$ preprocessed fNIRS data for subject k from some generic detector, X_k is the $N \times p$ design matrix, B_k is the $p \times 1$ vector of unknown parameters and E_k is the $N \times 1$ error vector. The columns of the design matrix, X_k, are constructed to model the hypothetical hemodynamic response function [3]. We are assuming that the parameter vector B_k is a realization from some population. Thus,

$$B_k = X_{gk} B_g + E_{gk} \quad (2)$$

where X_{gk} is the $p \times q$ group level design matrix linking the subject's parameters to the group parameters, B_g is the $q \times 1$ vector of group parameters and E_{gk} is the $p \times 1$ error vector. We may also analyze all the subjects together and modify (1) as,

$$Y = XB + E \quad (3)$$

where, Y is the $KN \times 1$ concatenated data vector, X is the $KN \times Kp$ separable subject-level design matrix, B is the $Kp \times 1$ concatenated parameter vector and E is the $KN \times 1$ concatenated error. The group level model may also be written as,

$$B = X_g B_g + E_g \quad (4)$$

where X_g is the $Kp \times q$ group level design matrix, B_g is the $q \times 1$ vector of group parameters and E_g is the $Kp \times 1$ error vector. Note that, we carried all the parameters from the subject level to the group level, i.e., no contrasts were applied at the subject level. Thus, group-level model is inherently multivariate since it brings together the subject-level estimates (a vector) to arrive at a group-level decision. Probability density function (pdf) of the subject-level error vector is assumed to be Gaussian with no temporal correlation, i.e. $\mathrm{cov}(E_k) = \sigma_k^2 I_N$. The pdf of the group-level error vector is also Gaussian but this time with a general covariance structure, i.e. $\mathrm{cov}(E_{gk}) = C_g$. Thus, the only assumption about the group-level error covariance is positive definiteness.

Directed acyclic graphs (DAG) or Bayesian networks provide an easy route to define structural and functional relations within a model. To build a DAG, we must define the relevant variables, i.e. nodes and the structural relationships, i.e. edges. Then it becomes possible to define the functional relationships in terms of conditional probabilities. The hierarchical model defined by (3) and (4) may be pictured by the DAG given in Figure 1. Having defined the variables and structural relationships, functional relationships may be stated by the following rule [9]:

$$p(z_n \mid r.v.) \propto p(z_n \mid pa(z_n)) . \prod_{z_m \in ch(z_n)} p(z_m \mid pa(z_m)) \quad (5)$$

where z_n and z_m are nodes of the DAG, $pa(z_n)$ stands for the parent of z_n and $ch(z_n)$ stands for children of z_n, and *r.v.* for the remaining variables. Thus, the pdf of any node in the network conditioned over the remaining variables can be written as in (5):

$$p(B_k \mid M, r.v.) \propto p(B_k \mid M, B_g, C_g) . p(Y_k \mid M, B_k, \sigma_k^2) \quad (6a)$$

$$p(\sigma_k^2 \mid M, r.v.) \propto p(\sigma_k^2 \mid M) . p(Y_k \mid M, B_k, \sigma_k^2) \quad (6b)$$

$$p(B_g \mid M, r.v.) \propto p(B_g \mid M) . \prod_{k=1:K} p(B_k \mid M, B_g, C_g) \quad (6c)$$

$$p(C_g \mid M, r.v.) \propto p(C_g \mid M) . \prod_{k=1:K} p(B_k \mid M, B_g, C_g) \quad (6d)$$

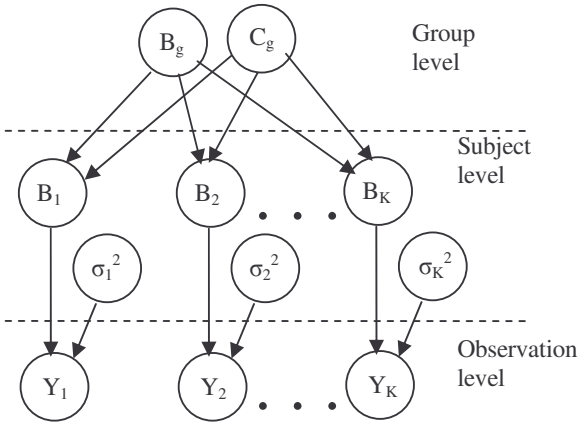

Figure 1 -Directed acyclic graph for the two-level hierarchical model.

where M stands for the model. Equations (6b-d) state that we need to specify prior distributions for σ_k^2, B_g, and C_g to be able to derive conditional posterior pdf's. Since generally we don't have a prior information about the distribution of the parameters, we have to use uninformative priors. Note that we do not have to specify a prior for B_k., since once the priors of B_g are specified, we already know the priors for B_k. Then group parameters are estimated by the combination of prior information with the information supplied by subjects' parameter estimates.

2.2 Subject-level parameters: B_k & σ_k^2

Conditional posterior of subject level parameters B_k:

Let us now write (6a) explicitly using the stated Gaussian assumptions:

$$p(B_k \mid M, r.v.) \propto |C_g|^{-1/2} \exp\left\{-(B_k - X_{gk} B_g)^T C_g^{-1} (B_k - X_{gk} B_g)/2\right\} \times \sigma_k^{-N/2} \exp\left\{-(Y_k - X_k B_k)^T (Y_k - X_k B_k)/2\sigma_k^2\right\}$$

Conditional posterior pdf of subject-level parameter estimates consists of the product of two Gaussian distributions, which is also a Gaussian:

$$p(B_k \mid M, r.v.) \propto \exp\left\{-(B_k - \hat{B}_k)^T \Delta_k^{-1} (B_k - \hat{B}_k)/2\right\} \quad (7)$$

where,

$$\hat{B}_k = \left(C_g^{-1} + \sigma_k^{-2} X_k^T X_k\right)^{-1} \left(C_g^{-1} X_{gk} B_g + \sigma_k^{-2} X_k^T Y_k\right) \quad (8a)$$

$$\Delta_k = \left(C_g^{-1} + \sigma_k^{-2} X_k^T X_k\right)^{-1} \quad (8b)$$

Equation (7) states that conditional posterior pdf of the subject level parameters has a Gaussian distribution whose mean and variance are, respectively, determined by the group-level mean and variance as well as subject-level variance of the given data. Thus, it is a combination of the information provided by the data and prior information provided by the group parameters. It may be easily seen that MLE (Maximum Likelihood Estimate) results may be obtained by setting $C_g = \infty$.

Conditional posterior of subject level error variance, σ_k^2:

The uninformative prior for σ_k^2 is known to be the Jeffrey's prior, that is, $p\left(\sigma_k^2 \mid M\right) \propto \left(\sigma_k^2\right)^{-1}$. Then we can write (6b) as,

$$p\left(\sigma_k^2 \mid M, r.v.\right) \propto \left(\sigma_k^2\right)^{-1}\left(\sigma_k^2\right)^{-N/2} \exp\left[-\left(Y_k - X_k B_k\right)^T \left(Y_k - X_k B_k\right)/2\sigma_k^2\right]$$
$$\propto \left(\sigma_k^2\right)^{-(N+2)/2} \exp\left[-\left(Y_k - X_k B_k\right)^T \left(Y_k - X_k B_k\right)/2\sigma_k^2\right]$$
$$\propto \left(\sigma_k^2\right)^{-(N+2)/2} \exp\left(-G/2\sigma_k^2\right)$$

where $G = \left(Y_k - X_k B_k\right)^T \left(Y_k - X_k B_k\right)$. Thus, conditional posterior pdf of noise covariance is inverse Gamma whose mean and mode are given by,

$$E\left(\sigma_k^2 \mid M, r.v.\right) = \frac{G}{N-2},$$
$$Mode\left(\sigma_k^2 \mid M, r.v.\right) = \frac{G}{N+2}.$$

2.3 Group-level parameters: B_g & C_g

Conditional posterior of group parameters, B_g:

The uninformative prior for the second level parameters is the uniform distribution:

$$p\left(B_g \mid M\right) \propto \text{constant}.$$

Then, conditional posterior pdf of B_g becomes:

$$p\left(B_g \mid M, r.v.\right) \propto \prod_{k=1:K} p\left(B_k \mid M, B_g, C_g\right), \qquad (9)$$

which is just the likelihood for B_g. We can expand the product term on the right hand side of (9) and, adopting the Kronecker product notation $\otimes$, we can write:

$$p\left(B_g \mid M, r.v.\right) \propto \prod_{k=1:K} \left|C_g\right|^{-1/2} \exp\left[-\left(B_k - X_{gk} B_g\right)^T C_g^{-1} \left(B_k - X_{gk} B_g\right)/2\right]$$
$$\propto \left|I_K \otimes C_g\right|^{-1/2} \exp\left[-\left(B - X_g B_g\right)^T \left(I_K \otimes C_g\right)^{-1} \left(B - X_g B_g\right)/2\right]$$
$$\propto \left|I_K \otimes C_g\right|^{-1/2} \exp\left[-\left(B_g - \hat{B}_g\right)^T \Delta_g^{-1} \left(B_g - \hat{B}_g\right)/2\right]$$

Thus, group level parameter vector has conditionally a Gaussian posterior distribution with conditional mean and variance:

$$\hat{B}_g = \left[X_g^T \left(I_K \otimes C_g\right)^{-1} X_g\right]^{-1} X_g^T \left(I_K \otimes C_g\right)^{-1} B,$$
$$\Delta_g = \left[X_g^T \left(I_K \otimes C_g\right)^{-1} X_g\right]^{-1}.$$

Here B denotes the concatenated B_k vectors, that is, $[B_1\ B_2 \ldots B_K]^T$).

Conditional posterior of group error covariance, C_g:

The uninformative prior for group level error covariance is,

$$p\left(C_g \mid M\right) \propto \left|C_g\right|^{-1}.$$

Then (6d) may be written as,

$$p\left(C_g \mid M, r.v.\right) \propto \left|C_g\right|^{-1} \times \prod_{k=1:K} p\left(B_k \mid M, B_g, C_g\right)$$
$$\propto \left|C_g\right|^{-1} \times \left|I_K \otimes C_g\right|^{-1/2} \exp\left[-\left(B - X_g B_g\right)^T \left(I_K \otimes C_g\right)^{-1} \left(B - X_g B_g\right)/2\right]$$

We may switch to a multivariate expression by defining B^m such that each of its rows is the vector of parameter estimates for a subject. Design matrix may also be arranged respectively. Then we can write,

$$p\left(C_g \mid M, r.v.\right) \propto \left|C_g\right|^{-(K+2)/2} \exp\left(-trC_g^{-1}W/2\right)$$

where "*tr*" stands for the trace operator and $W = \left(B^m - X_g^m B_g^m\right)^T \left(B^m - X_g^m B_g^m\right)$. Thus, group-level error variance has a conditional inverted Wishart distribution with mean and mode,

$$E\left(C_g \mid M, r.v.\right) = \frac{W}{K-2q},$$
$$Mode\left(C_G \mid M, r.v.\right) = \frac{W}{K+2}.$$

Mean value is defined for K > 2q, i.e., the number of subjects must be bigger than twice the length of the group-level parameter vector. If this is not the case, then a more structured covariance matrix should be chosen.

2.4 Inference via posterior modes

Since we have already calculated the modes of the conditional posterior pdf of all the variables, we may proceed with an algorithm like iterated conditional modes (ICM) [10]. Beginning from some initial values we may cycle through the modes until convergence. The algorithm is summarized below:

$$\hat{B}_k = \left(C_g^{-1} + \sigma_k^{-2} X_k^T X_k\right)^{-1} \left(C_g^{-1} X_{gk} B_g + \sigma_k^{-2} X_k^T Y_k\right), \qquad (10a)$$

$$\hat{\sigma}_k^2 = \frac{\left(Y_k - X_k B_k\right)^T \left(Y_k - X_k B_k\right)}{N+2}, \qquad (10b)$$

$$\hat{B}_g = \left[X_g^T \left(I_K \otimes C_g\right)^{-1} X_g\right]^{-1} X_g^T \left(I_K \otimes C_g\right)^{-1} B, \qquad (10c)$$

$$\hat{C}_g = \frac{\left(B^m - X_g^m B_g^m\right)^T \left(B^m - X_g^m B_g^m\right)}{K+2}. \qquad (10d)$$

We may run (10a) for each of the K subjects and continue with the rest 3 equations. After convergence we may calculate $\text{var}\left(\hat{B}_g\right) = \left[X_g^T \left(I_K \otimes C_g\right)^{-1} X_g\right]^{-1}$. In this way, the mean and variance of the posterior Gaussian distribution of the group parameters are found. Generally, we are interested in the probability distribution of a linear combination of the parameters (for instance, difference between the parameters of two types of stimuli). This may be achieved by specifying a contrast vector, c^T. Then the mean and variance for $c^T B_g$ are simply:

$$E\left\{c^T B_g\right\} = \mu = c^T \hat{B}_g$$

$$\mathrm{var}\left\{c^T B_g\right\} = \sigma^2 = c^T \mathrm{var}\left(B_g\right) c = c^T \left[X_s^T \left(I_K \otimes C_s\right)^{-1} X_s\right]^{-1} c$$

Then, the posterior probability of the contrasted parameters having a value *d* is:

$$p\left(d \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(d-\mu\right)^2}{\sigma^2}\right]$$

3. EXPERIMENTAL RESULTS

3.1 Experimental setup and data collection

Subjects: We recruited 9 healthy male subjects from the university community. Subjects had no reported neurological disorder and informed consent was obtained before the measurement.

fNIRS data acquisition: Experiments were performed using a continuous wave near-infrared spectroscopy device (NIROXCOPE 201) built in Biophotonics Laboratory of Bogazici University [11]. The device is capable of transmitting near-infrared light in two wavelengths (730 nm, and 850 nm), which are known to be able to penetrate through the scalp and probe the cerebral cortex. Calculation of concentration changes in oxygenated and deoxygenated (HbO2 and HbH) blood is based on Beer-Lambert law. Employing four light emitting diodes (LEDs) and 10 detectors, the device can sample 16 different volumes in the brain simultaneously. The distance between each source and detector is 2.5 cm, which guarantees a probing depth of approximately 2 cm from the scalp. LEDs and detectors were placed in a rubber band that was specially designed to fit the forehead. Sampling frequency of the device was 1.7 Hz.

Experimental Paradigm: Subjects were asked to perform color-word matching Stroop task [8]. They were presented with two words one written on top of the other. The top one was written in ink-color whereas the below one was in white (over a black background). Subjects were asked to judge whether the word written below correctly denotes the color of the upper word or not. If so, subjects were to press the left mouse button with their forefinger, and if not to the right mouse button with their middle finger. The stimuli stayed on the screen until the subject responded or at the end of three seconds in case the subject did not respond.

The experiment consisted of neutral, congruent and incongruent trials. In the neutral condition upper word consisted of four X's (XXXX) in ink-color. In the congruent condition ink-color of the upper word and the word itself were the same whereas in the incongruent condition they were different. Inter-stimulus interval was 16 seconds and stimuli were presented in a random order. Experiments were performed in a silent, lightly-dimmed room.

We expect the subjects to have more difficulty with the incongruent stimulus, and thus an increase in the cognitive activity with respect to neutral and congruent stimulus.

3.2 Behavioral Results

Reaction times (RT) were calculated only from the correctly answered trials. RTs to neutral, congruent and incongruent stimuli were 988.5±180.7, 1042.3±272.1 and 1190.6±369.1 ms, respectively. Percentage error rates were 6.25±9.92, 7.92±10.97 and 16.25±16.54, respectively. Comparing the RTs, there was no significant difference between neutral and congruent stimuli. Two-tailed paired t-test revealed a significant difference between incongruent and neutral ($t(8) = 2.449$, $p = 0.040$) and incongruent and congruent ($t(8) = 3.476$, $p = 0.008$) trials which point to a clear interference effect.

3.3 fNIRS results

fNIRS data were digitally low-pass filtered with a cut-off frequency of 330 mHz. Low frequency trends were eliminated in the temporal domain with 120 s cut-off. Stimulus onset vectors for each type of stimuli were formed and convolved with the canonical haemodynamic response function (HRF). Correctly answered, incorrectly answered and omitted stimuli were modeled separately. Inference was based on the correctly answered stimuli. We eliminated some of the detectors which exhibit excessive noise and use 12 of the 16 channels such that we covered the lateral and medial parts of the prefrontal cortex. The oxy- and deoxy-Hb data have been analyzed separately. fNIRS results for interference (incongruent – neutral) are summarized in Table I. We compare four methods, namely: Fixed effect analysis (FFX); Randox effect analysis (RFX); Mixed-effect analysis (MFX); Bayesian estimation of group parameters (BPE).

Table IA: Oxy-Hb results where the cells show *parameter estimate / variance of the estimate* (* denotes statistically significant activation, $p<0.05$)

Detector No:	FFX	RFX	MFX	BPE
1	0.574 / 0.144*	0.574 / 0.564	0.535 / 0.515	0.539 / 0.409
2	1.187 / 0.210*	1.187 / 0.612*	1.001 / 0.554*	1.015 / 0.358*
3	-0.521 / 1.038	-0.521 / 0.908	-0.051 / 0.151	-0.015 / 0.412
4	-0.090 / 0.120	-0.090 / 0.388	-0.065 / 0.354	-0.092 / 0.278
7	0.454 / 0.166*	0.454 / 0.515	0.485 / 0.468	0.496 / 0.340
8	10.073 / 6.420	10.073 / 9.708	2.535 / 7.501	7.923 / 6.258
9	0.128 / 0.191	0.128 / 0.259	0.297 / 0.205	0.166 / 0.151
10	-0.437 / 0.233	-0.437 / 0.495	-0.281 / 0.434	-0.408 / 0.358
13	0.769 / 0.220*	0.769 / 0.410*	0.643 / 0.353*	0.731 / 0.262*
14	0.150 / 0.161	0.150 / 0.663	0.131 / 0.604	0.146 / 0.488
15	-1.333 / 0.236	-1.333 / 1.140	-1.171 / 1.040	-1.253 / 0.827
16	-0.715 / 0.168	-0.715 / 0.733	-0.614 / 0.669	-0.624 / 0.495

It can be observed that the parameter estimates are close to each other. The real differences of the methods show in the variances. While none of these variance estimates can de said to be wrong, their differences stem from differing assumptions on the data model. Notice that the smallest variances are obtained by the FFX which is expected for an autocorrelated signal. It is interesting to note that RFX, MFX and BPE exhibited the same activation pattern for both oxy-Hb and deoxy-Hb. The reason for this is that between-subject vari-

ance dominates over within-subject variance. These three methods detected activation in detectors 2 and 13 for oxy-Hb and in detector 14 for deoxy-Hb. FFX has additionally detected activation in detectors 1 and 7 for oxy-Hb and in detectors 2,3,4,8 and 13.. This difference again stems from the inclusion of between subject variance.

Table IB: Deoxy-Hb results (as above)

Detector No:	FFX	RFX	MFX	BPE
1	0.175 / 0.165	0.175 / 0.743	0.158 / 0.677	0.126 / 0.541
2	-0.570 / 0.116*	-0.570 / 0.492	-0.541 / 0.449	-0.555 / 0.350
3	-0.526 / 0.145*	-0.526 / 0.526	-0.427 / 0.480	-0.433 / 0.352
4	-0.732 / 0.210*	-0.732 / 0.658	-0.733 / 0.597	-0.745 / 0.467
7	0.567 / 0.258	0.567 / 0.787	0.546 / 0.711	0.527 / 0.566
8	-0.873 / 0.338*	-0.873 / 0.947	-0.485 / 0.852	-0.539 / 0.487
9	3.144 / 0.614	3.144 / 2.385	2.854 / 2.172	2.891 / 1.619
10	8.830 / 1.806	8.830 / 8.208	6.971 / 7.483	6.337 / 4.661
13	-0.389 / 0.150*	-0.389 / 0.533	-0.391 / 0.485	-0.391 / 0.381
14	-1.714 / 0.444*	-1.714 / 0.830*	-1.318 / 0.714*	-1.451 / 0.482*
15	-0.421 / 0.332	-0.421 / 0.468	-0.254 / 0.355	-0.355 / 0.278
16	-0.291 / 0.226	-0.291 / 0.335	-0.297 / 0.258	-0.332 / 0.209

4. DISCUSSION

Schroeter *et al*. [13] used the Stroop test to detect prefrontal activation by fNIRS. For the normal subjects, they showed activation in lateral prefrontal cortex bilaterally both as an increase in oxy-Hb and tot-Hb and as a decrease in deoxy-Hb. Ehlis *et al*. [14] showed specific activation for interference trials for oxy-Hb and total hemoglobin (tot-Hb) in inferior-frontal areas of the left hemisphere, whereas the statistical results for deoxy-Hb were much weaker and less conclusive than for the oxy-Hb and tot-Hb signals. Comparing their findings with those of Schroeter *et al*. they argued that the differences might be due to differences in the task paradigm, response type or measurement device. We used the same version of the Stroop test as Schroeter *et al*. Our method has only detected activation for oxy-Hb in the left lateral prefrontal cortex and none for deoxy-Hb. Thus our results are more in line with those of Ehlis *et al*.

This study treated the detection problem for fNIRS signals. It is a model-based approach since it assumes an a priori model for the hemodynamic response function. In this sense, its success depends on how well the practical HRF coincides with the theoretical HRF. We used the well known canonical HRF [3] which is shown to be adequate for fMRI data. Since fMRI and fNIRS both measure the hemodynamic activity related with the cognitive activity it is plausible to use the same HRF. However, estimation of the HRF for fNIRS signals is a topic of further study. Marrelec *et al.* [15] have shown that it was possible to estimate HRF within the framework of Bayesian network. Our analysis showed that within-subject variance is much lower than between-subjects variance for fNIRS signals, as attested by the variances in columns 3 of the Tables 1A and 1B vis-à-vis columns 1 or 2. This is also the case with fMRI and other types of signals with serial correlation. Thus, it is important for fNIRS to make a mixed effects analysis, i.e., to take into account both types of variances. An important difference of our method with respect to classical ones was the implementation of a multivariate GLM in the second level. It enabled us to include the correlations between the estimated parameters. Our assumption of sphericity of the error covariance in the subject level is clearly a simplification. Thus, further studies regarding the temporal structure of the error sequence for fNIRS signals should be performed.

5. CONCLUSION

We presented a Bayesian method for making statistical inference in a multilevel setting. The main virtue of the method is its capability to produce posterior pdf's of the parameters at any level by taking into account causal relationships among the variables of the model. It provides a general framework to tackle the problem of deriving generalizable results for a population. The method was applied successfully to fNIRS signals. Results demonstrate that fNIRS has the capacity to monitor cognitive activity with suitable statistical tools.

REFERENCES

[1] A. Villringer, and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends Neurosci*, vol. 20, pp. 4435-442, 1997.

[2] A.P. Holmes, and K.J. Friston, "Generalisability, random effects and population inference," *NeuroImage,* vol. 7, S754, 1998.

[3] K. Friston, A. Holmes, J.B. Poline, C. Frith, and R. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Hum Brain Map,* vol. 2, pp. 189–210, 1995.

[4] M.L. Schroeter, M.M. Bücheler, K. Miller, K. Uludag, H. Obrig, G. Lohmann, M. Tittgemeyer, A. Villringer, and D.Y. von Cramon, "Towards a standard analysis for functional near-infrared imaging," *NeuroImage,* vol. 21, pp. 283-290, 2004.

[5] F.V. Jensen, *Bayesian networks and decision graphs*, Springer-Verlag, New York, 2001.

[6] G.K. Büyükaksoy, N.S. Şengör, H. Gürvit, and C. Güzeliş, "Modelling the Stroop effect: A connectionist approach." *Neurocomputing*. doi: 10.1016/j.neucom.2006.05.009.

[7] A.W. MacDonald, J.D. Cohen, A. Stenger, and C.S. Carter, "Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control" *Science, New Series,* no. 288, pp. 1835-1838. 2000

[8] S. Zysset, K. Muller, G. Lohmann, and D.Y. von Cramon, "Color–word matching Stroop task: separating interference and response conflict" *Neuroimage,* vol. 13, pp. 29–36, 2001.

[9] J. Pearl, *Causality: Models, reasoning, and inference*, Cambridge University Press, 2000.

[10] D.B. Rowe, *Multivariate Bayesian Statistics,* Chapman & Hall/CRC, 2002.

[11] C.B. Akgül, A. Akin, and B. Sankur, "Extraction of cognitive activity-related waveforms from functional near-infrared spectroscopy signals," *Med.&Biol.Eng.&Comput*., vol. 44, pp. 945-958, 2006.

[12] C.F. Beckmann, M. Jenkinson, and S.M. Smith, "General multilevel linear modeling for group analysis in FMRI.," *NeuroImage,* vol. 20, pp. 1052-1063, 2003.

[13] M.L. Schroeter, S. Zysset, T. Kupka, F. Kruggel, and D.Y. von Cramon, "Near-infrared spectroscopy can detect brain activation during a color-word matching Stroop task in an event-related design" *Hum. Brain Mapp*, vol. 17, pp. 61-71, 2002

[14] A.-C. Ehlis, M.J. Herrmann, M.J., A. Wagener, and A.J. Fallgatter, "Multi-channel near-infrared spectroscopy detects specific inferior-frontal activation during incongruent Stroop trials" *Biological Psychology*, vol. 69, pp. 315-332, 2005.

[15] G. Marrelec, P. Ciuciu, M. Pelegrini-Issac, and H. Benali, "Estimation of the hemodynamic response in event-related functional MRI: Bayesian networks as a framework for efficient Bayesian modeling and inference," *IEEE Trans. Med. Imag*., vol. 23, pp. 959-967, 2004.