

# Multilevel Statistical Inference From Functional Near-Infrared Spectroscopy Data During Stroop Interference

Koray Çiftçi\*, Bülent Sankur, *Senior Member, IEEE*, Yasemin P. Kahya, *Member, IEEE*, and Ata Akin, *Member, IEEE*

**Abstract**—Functional near-infrared spectroscopy (fNIRS) is an emerging technique for monitoring the concentration changes of oxy- and deoxy-hemoglobin (oxy-Hb and deoxy-Hb) in the brain. An important consideration in fNIRS-based neuroimaging modality is to conduct group-level analysis from a set of time series measured from a group of subjects. We investigate the feasibility of multilevel statistical inference for fNIRS. As a case study, we search for hemodynamic activations in the prefrontal cortex during Stroop interference. Hierarchical general linear model (GLM) is used for making this multilevel analysis. Activation patterns both at the subject and group level are investigated on a comparative basis using various classical and Bayesian inference methods. All methods showed consistent left lateral prefrontal cortex activation for oxy-Hb during interference condition, while the effects were much less pronounced for deoxy-Hb. Our analysis showed that mixed effects or Bayesian models are more convenient for faithful analysis of fNIRS data. We arrived at two important conclusions. First, fNIRS has the capability to identify activations at the group level, and second, the mixed effects or Bayesian model is the appropriate mechanism to pass from subject to group-level inference.

**Index Terms**—General linear model (GLM), near-infrared spectroscopy, statistical inference, Stroop task.

## I. INTRODUCTION

FUNCTIONAL near-infrared spectroscopy (fNIRS) is a noninvasive method to monitor brain activation by measuring changes in the concentrations of oxy- and deoxy-hemoglobin (oxy-Hb and deoxy-Hb) [1]. It is simply based on measuring the transmitted and received near-infrared light in multiple wavelengths and calculating the relative concentrations of oxy-Hb and deoxy-Hb using modified Beer–Lambert law [2]. fNIRS has significant advantages over functional magnetic resonance imaging (fMRI) such as absence of radiation, portable nature of the device, relative user friendliness, and low cost of the

procedure. On the other hand, fNIRS has the shortcomings of low spatial resolution, shallow depth of penetration, and consequently, some uncertainty about the probed region. Although near-infrared spectroscopy has been successfully employed in a number of physiological measurements [3], [4], more progress is needed to establish standard methods for statistical inference of cognitive activity detection and assessment.

Parametric statistical analysis (PSA) of neuroimaging data typically focuses on testing of hypotheses for subject populations. Given a set of observed data, PSA tries to infer on the highest level in the hierarchy. This highest level often represents the effect analysis over some or all of several measurements, detectors, sessions, and subjects in a population. General linear model (GLM) has been the most commonly used tool to make inferences from fMRI data [5]. GLM may also be extended to a hierarchical mode to arrive at multilevel statistical inferences [6]. Apart from a few inherent differences between them, both fMRI and fNIRS aim to detect and localize brain hemodynamic activity based upon neurovascular coupling model. Thus, it would be logical to extend the GLM methodology to fNIRS signals. Schroeter *et al.* [7] were 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. In a recent study, it was shown that model-based analysis with GLM is capable of detecting event-related human brain activity recorded with fNIRS in the occipital cortex [8]. A “shift method” has also been proposed to recover small signals within the GLM framework, which exploits the higher temporal resolution of fNIRS with respect to fMRI [9].

We address in this study the multilevel inference problem for fNIRS signals using a hierarchical GLM to link the measurement space to the upper-level parameters. In the multilevel approach, the activation patterns are estimated first at the subject level, and then, these patterns are carried over to the group level. In this analysis, we use comparatively three classical methods of multilevel inference, i.e., fixed effects (FFX), random effects (RFX), mixed effects (MFX) analyses, and two Bayesian inference methods. One of the Bayesian methods also goes by the name of pseudomixed effects ( $\Psi$ FX) [10], since it employs the basic GLM at the subject level and uses the Bayesian methodology to merge the subject parameters at the group level. The second method, denoted as Bayesian posterior estimation (BPE), is a fully Bayesian one. It may be argued that the Bayesian methodology can cope better with the classical problem of within-subject and between-subject variances in a

Manuscript received October 15, 2007; revised March 1, 2008. This work was supported in part by Boğaziçi University Research Fund under Grant BURF 02S102 and Grant DPT 03K120250 and in part by the Türkiye Bilimsel ve Teknolojik Arastırma Kurumu-Consiglio Nazionale delle Ricerche (TUBITAK-CNR) under Grant 104E101. The work of A. Akin was supported by the National Institutes of Health (NIH). *Asterisk indicates corresponding author.*

\*K. Çiftçi is with the Institute of Biomedical Engineering, Boğaziçi University, Istanbul 34342, Turkey (e-mail: rciftci@boun.edu.tr).

B. Sankur and Y. P. Kahya are with the Department of Electrical and Electronic Engineering, Boğaziçi University, Istanbul 34342, Turkey (e-mail: bulent.sankur@boun.edu.tr; kahya@boun.edu.tr).

A. Akin is with the Institute of Biomedical Engineering, Boğaziçi University, Istanbul 34342, Turkey (e-mail: ata.akin@boun.edu.tr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TBME.2008.923918

more principled framework [6]. We will contrast the classical and Bayesian approaches and demonstrate the advantages of the latter.

The particular experimental protocol that we used in this study is a variant of Stroop task, which is known to be a good activator for prefrontal cortex [11], [12]. We used the version of the Stroop task introduced by Zysset *et al.* [13], since it provides a way to separate interference that takes place at the conceptual level from the response preparation, and furthermore, it is very suitable for computerized application of the test.

We obtained data from healthy subjects and determined activations in their prefrontal cortex during Stroop interference at the subject and group levels. Schroeter *et al.* [14] had investigated the same type of Stroop task with fNIRS on healthy subjects. They found that the hemodynamic response was stronger for incongruent compared to neutral and congruent trials in the lateral prefrontal cortex, bilaterally both for oxy-Hb and deoxy-Hb (and total Hb, which is their sum). Ehlis *et al.* [15] used a different kind of Stroop paradigm to monitor the hemodynamic response of healthy subjects by fNIRS. In interference trials, they revealed specific activations in the inferior–frontal areas of left hemisphere for oxy-Hb and total Hb. However, their activation for deoxy-Hb was much weaker and less conclusive. In line with these studies, we set to explore the feasibility of fNIRS to monitor the cognitive activity and search for spatially specific cognitive activations in the prefrontal cortex during Stroop interference.

An important aspect of our research is the investigation of the cortical cognitive responses in a multilevel setting, that is, both at subject and group levels. This allows us to keep track of the activations at different levels and pursue methodological discussions. Although Schroeter *et al.* [7] used random effects model for group activation analysis, in-depth analysis of fNIRS signals using hierarchical models has not been performed yet. Hence, this study must be seen as an effort to demonstrate the potential of fNIRS for cognitive activity monitoring. The second important contribution of our research is the experimental comparison of classical methodologies and Bayesian methodologies, and the assessment of the superiority of Bayesian approach in predicting activity.

## II. METHODS

### A. Classical Analysis of fNIRS Data

Classical analysis of multilevel functional neuroimaging data generally proceeds in a bottom-up fashion. Once the statistics that summarize the data at one level are calculated, they are carried to the upper level. The main difference among the three classical statistical inference techniques of FFX, RFX, and MFX lies in the determination of the variance estimates [16], [17]. An outline of the methods is presented in Appendix A. Briefly, FFX and RFX ignore the between-subject and within-subject variances, respectively. Note that, since it ignores the between-subject variance, the inference of FFX is limited to the particular set of subjects [16]. After calculating the subject parameter and variance estimates using GLM specifically designed for each subject, FFX takes the average variance estimate as the

group variance. On the other hand, RFX calculates the group variance over the estimated parameters of the subjects. MFX tries to integrate both within- and between-subject variances by carrying the subject variance estimates to the group level. In this study, MFX was carried on as described by Beckmann *et al.* [16] and implemented by Thirion *et al.* [10]. FFX, RFX, and MFX are all summary statistical approaches, i.e., beginning from the bottom level, each level is analyzed separately and only the parameters of interest are carried to the upper level. The main benefit of working with a summary statistics approach is its computational ease, which becomes very important for high-dimensional data like fMRI.

The statistics proposed by Neumann and Lohmann [18], called pseudomixed effects ( $\Psi$ FX), is a mixture of classical and Bayesian procedures. The parameter and variance estimates are calculated at the subject level using the GLM. Then to arrive at the group decision, the posterior distribution of a subject is taken as the prior distribution of another subject. The end result is an average of subject parameter estimates inversely weighted by their variance estimates. In essence, this is a fixed-effects approach, since it does not take into account between-subject variances. Note that, this is also a summary statistics method.

In conclusion, parameters estimated at subject level are the same for all of these four methods, namely FFX, RFX, MFX, and  $\Psi$ FX. After specifying subject-specific GLMs, one calculates subjects' parameters and variances, and continues toward average group activation calculation. Since we are generally not interested in all of the parameters but rather in a particular linear combination of them, contrast vectors are specified at the subject level and applied to the parameter and variance estimates.

### B. Bayesian Analysis of fNIRS Data

Bayesian analysis of hierarchical GLM has been applied extensively to fMRI signals [6], [17]. Our implementation of Bayesian methodology for fNIRS signals will also follow similar procedures. We specify noninformative priors as in [17], since we do not have any prior information and generally the number of subjects is so small to make the influence of the prior significant. The details of the Bayesian analysis are presented in Appendix B. Since the modes of the conditional posterior probability distribution functions can be easily calculated, we can make use of an algorithm like iterated conditional modes (ICMs) [19]. Beginning from some initial values, we can cycle through the modes until convergence. We preferred ICM over some other Monte Carlo schemes like Gibbs sampling because of its simplicity and speed, which are important criteria especially for practical purposes. For multimodal distributions, ICM has the risk of getting stuck at a local minimum or oscillating, but for unimodal distributions (as it is in our case), ICM gives quick solutions. In actual implementation, we confirmed the convergence of the algorithm to the same output by starting the chain at different initial points.

We applied the contrast vector only when all of the estimation process had ended and that group parameters were available. As in the classical analysis case, this may be achieved by specifying a contrast vector. The marginal posterior of

contrasted group parameters obeys a univariate noncentral Student's  $t$ -distribution [20]. We can make inferences using this posterior, and ask whether our contrasted parameter estimates are higher than a particular value.

The main difference between the Bayesian analysis presented here (BPE) and the methods mentioned in the previous section is that the former is not summary statistics. Bayesian analysis in our implementation incorporates the group variables into subject parameter estimation process. Hence, each subject should be analyzed simultaneously, and if a new subject is included in the group, the analysis should be repeated for every subject.

### III. EXPERIMENTS

#### A. Subjects

We recruited 12 healthy subjects (7 females, 5 males) from the university community (right-handed, mean age  $26.17 \pm 4.30$ , range 20–31). Subjects had no reported neurological, medical, and psychiatric disorders. None were taking medications at the time of measurement. All the subjects had normal or corrected-to-normal vision and normal color vision. Written informed consent was obtained from all subjects before the measurement.

#### B. fNIRS Data Acquisition

Experiments were performed using a continuous-wave near-infrared spectroscopy device (NIROSCOPE 301) built in Biophotonics Laboratory, Boğaziçi University [21], [22]. The device is capable of transmitting near-infrared light at two wavelengths (730 and 850 nm), which are known to be able to penetrate through the scalp and probe the cerebral cortex. Calculation of concentration changes of oxy-Hb and deoxy-Hb in blood is based on Beer–Lambert law. Employing four LEDs and ten detectors, the device can sample 16 different volumes (channels) in the brain simultaneously. The distance between each source and detector is 2.5 cm, which guarantees a probing depth of approximately 2.0 cm from the scalp (see Fig. 1 for the details of the probe). We have chosen rectangular probe geometry for obtaining nonoverlapping areas and limited our source-to-detector distances to 2.5 cm for a better fit to the forehead. This amount of separation has been shown to reliably probe the cortical activity [22]–[26]. LEDs and detectors were placed in a flexible printed circuit board that was specially designed to fit the curvature of the forehead. Sampling frequency of the device was 1.4 Hz.

Stimulus onset vectors for each type of stimulus (neutral, congruent and incongruent) were formed and convolved with the canonical hemodynamic response function (HRF) [27]. Canonical HRF was scaled so that it had a peak response of unity; hence, the parameter estimates directly gave peak concentration change during the stimulus. These three vectors constituted the cognitive part of the design matrix. The fNIRS data were digitally low-pass filtered with a cutoff frequency of 330 mHz. To be able to cope with various low-frequency trends, discrete cosine transform (DCT) basis functions [28] were added to the design matrix with a minimum period of 120 s. Incorrect and omitted trials were modeled separately and they, together with the trend terms, form the nuisance part of the design matrix.

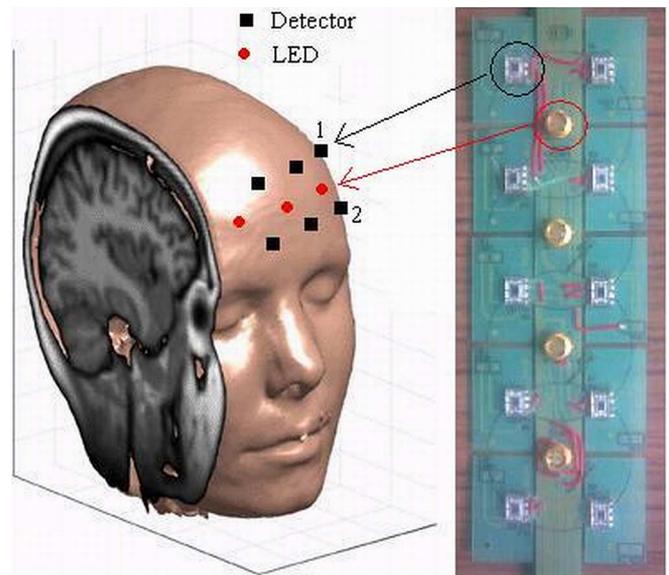


Fig. 1. NIROSCOPE 301 probe (right) is attached to the forehead. Source–detector geometry ensures probing of 16 nonoverlapping volumes when the light sources are time multiplexed. (Head image was obtained from MATLAB Central File Exchange: <http://www.mathworks.com/matlabcentral/fileexchange>.)

#### C. Experimental Paradigm

Subjects were asked to perform color–word matching Stroop task whose trials are the Turkish versions of Zysset *et al.* [13]. Subjects were presented with two words, one written above the other. The top one was written in ink color whereas the bottom 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 color and word match, then subjects were to press on the left mouse button with their forefinger, and if not, on the right mouse button with their middle finger. Subjects were informed to perform the task as quickly and correctly as possible. The words stayed on the screen until the response was given with a maximum time of 3 s. The screen was blank between the trials. 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. The trials were presented in a semiblocked manner. Each block consisted of six trials. Interstimulus interval within the block was 4.5 s and the blocks were placed 20 s apart in time. The trial type within a block was homogeneous (but the arrangements of false and correct trials were altering). There were ten blocks of each type. Experiments were performed in a silent, lightly dimmed room. Words were presented via an LCD screen that was 0.5 m away from the subjects. The task protocol is approved by the Ethics Review Board of Boğaziçi University.

## IV. RESULTS

#### A. Behavioral Results

Reaction times (RTs) were calculated only from the correctly answered trials. The first and second subjects responded slower

to congruent stimuli in comparison to incongruent stimuli. Subject 6 responded slightly slower to neutral trials than congruent trials. For the rest of the subjects, the ordering of RTs is neutral–congruent–incongruent. The average RTs to neutral, congruent, and incongruent trials are  $1029.3 \pm 277.1$ ,  $1183.9 \pm 370.5$ , and  $1308.8 \pm 367.1$  ms, respectively.

Comparing the RTs, two-tailed paired  $t$ -test revealed significant differences among all three trial types: incongruent vs. neutral:  $t(11) = 7.042$ ,  $p = 0.000$ ; incongruent vs. congruent:  $t(11) = 2.882$ ,  $p = 0.015$ ; congruent vs. neutral:  $t(11) = 4.351$ ,  $p = 0.001$ . There were two common effects in the Stroop task: first, the interference effect refers to the observation that subjects have more difficulty in answering incongruent trials with respect to neutral trials. Second, the facilitation effect comes from the observation that subjects respond quicker to congruent trials compared to neutral trials [29]. Although the interference effect was evident in RTs, we could not observe a facilitation effect. Using the same kind of stimuli, Zysset *et al.* [13] have not observed facilitation effect either. It has been pointed out that facilitation was not a necessary concomitant of interference and it played a much lesser role than interference [12]. It was asserted that the missing facilitation was due to trying to speed up an already rapid response [12], [14]. Additionally, as pointed out by an anonymous reviewer, the slower response to congruent trials may be related with the observation that the subjects try to judge whether the trial is congruent or incongruent, which puts an extra cognitive load with respect to neutral trials.

Error rates were generally small, and most of the subjects did not make any mistakes for neutral and congruent trials. Mean error rates (in percentage) were  $0.56 \pm 1.92$ ,  $0.56 \pm 1.30$ , and  $4.31 \pm 5.97$ . We did not make any statistical test in terms of error rates, since their distributions are clearly not Gaussian. However, it can be said that interference effect also manifests itself in error rates.

## B. fNIRS Results

Our fNIRS device provides us with a set of time series recorded over 16 channels over the scalp. The locations of the probed regions are shown in Fig. 1. Note that the ordering of the channels is from left to right, i.e., “1” is on the left and “16” is on the right. Oxy-Hb and deoxy-Hb data were analyzed separately.

1) *Oxy-Hb Results*: The subject-level and group-level activation patterns for interference effect (incongruent–neutral) are shown in Fig. 2(a). These patterns and the others presented in the following figures result from the thresholded  $z$ -scores at 0.05 significance level (that is,  $z_{\text{thresh}} = 1.65$  and  $p = 0.05$ , adjusted for multiple comparisons by Bonferroni correction). We also converted the posterior probabilities given by the  $\Psi$ FX and BPE to  $z$ -statistics. Recall that subject-level activations are common for FFX, RFX, MFX, and  $\Psi$ FX, and estimated by ordinary least squares (OLS) in a single step, whereas BPE iteratively estimates both subject and group parameters. Our first observation is that there is activation widespread over channels for most of the subjects. Furthermore, all subject activations resemble each other for both OLS and BPE approaches. This is

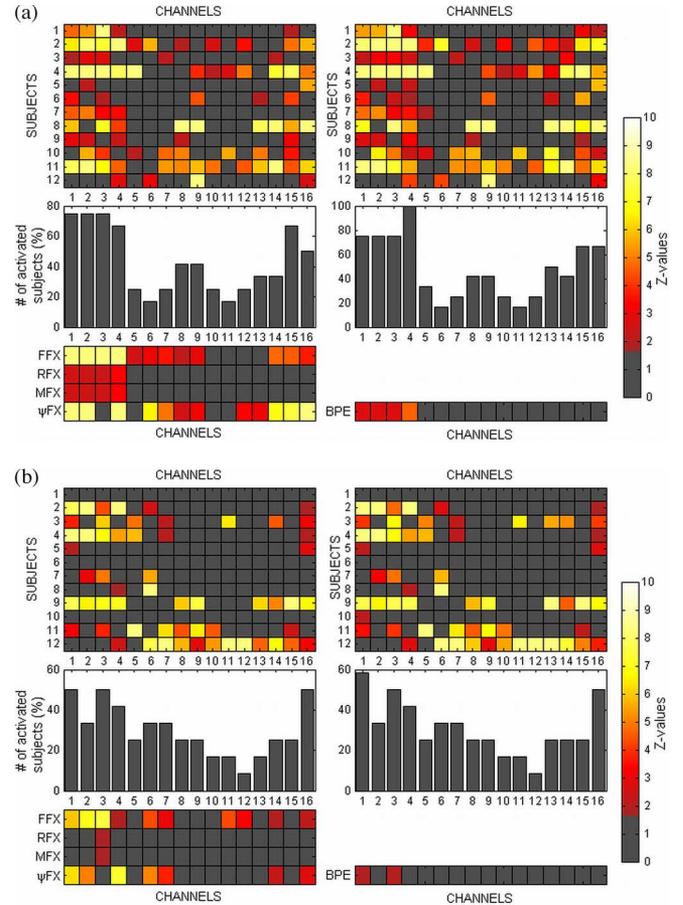


Fig. 2. (a) Activation patterns for oxy-Hb for “incongruent–neutral” contrast. Top: Subject-level activations detected by OLS (left) and BPE (right). Middle: Activated subject count (%) for OLS (left) and BPE (right). Bottom: Group-level activations for FFX, RFX, MFX,  $\Psi$ FX (left), and BPE (right). (b) Activation patterns for oxy-Hb for “incongruent–congruent” contrast. Organization of the figure is the same as that of Fig. 2(a).

usual and points to the fact that group-level variance is higher than subject-level variance, which causes the effect of group parameters being weighted down in the estimation of subject-level parameters (see Section II-B). Despite the apparent similarity between OLS and BPE methods, the consistent activation in channel 4 revealed by BPE is worth noticing. BPE finds that channel 4 is activated for all of the subjects, while this is not the case for single-level GLM. Our second important observation is that the percentage of activated subjects per channel indicates that activation is dominantly left lateral [Fig. 2(a), middle row]. When group-level inference is inspected [Fig. 2(a), bottom row], this left laterality is especially evident with RFX, MFX, and BPE. Channels 1–4 are found to be active, with channel 4 giving the highest  $z$ -value and consistency. Third, it can be seen that the widespread activation at the subject level is carried over to the group level with FFX and  $\Psi$ FX. This is to be expected because these two methods do not consider the between-subject variance. The consequence is that FFX and  $\Psi$ FX have higher sensitivity but at the risk of high false positive rates.

We also investigated whether there is a significant activation difference between incongruent and congruent trials. Our

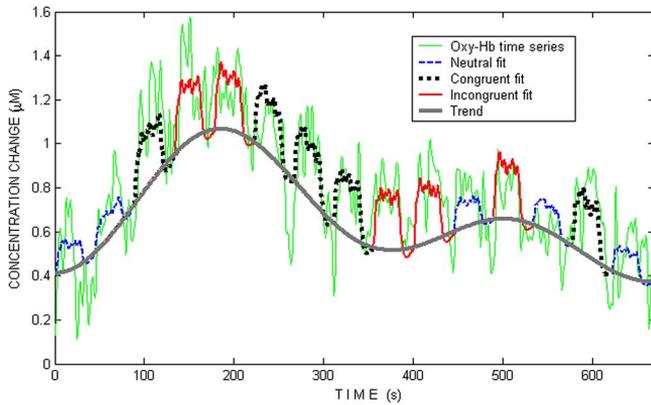


Fig. 3. oxy-Hb time series with fitted cognitive waveforms and trend component.

behavioral results have shown that there was no facilitation effect, i.e., subjects had more difficulty with congruent trials with respect to neutral trials. This also manifested itself in fNIRS findings and the activations both at the subject and group levels are less pronounced this time [we could not find any activation for subjects 1 and 6, see Fig. 2(b), top row]. FFX and  $\Psi$ FX again found higher number of activated channels compared to the other three methods [Fig. 2(b), bottom row]. The activations of RFX, MFX, and BPE are confined to the left lateral channels.

We would suggest that the medial activations detected by FFX and  $\Psi$ FX may be due to anterior cingulate cortex (ACC), which has been identified as a region involved in Stroop-like inhibition paradigms [30]. However, it has been shown that ACC is not specifically involved in interference processes, but rather in motor preparation processes [13]. Hence, ACC should not be substantially activated when comparing neutral and incongruent conditions, as the motor response preparation process, once the decision is taken, is the same for both conditions in color-word matching Stroop task [13]. Additionally, considering the penetration depth of near-infrared light [1], it is doubtful if fNIRS would be able to capture the activations in ACC with source-detector separation of 2.5 cm. Hence, we conclude that the medial activations detected by FFX and  $\Psi$ FX are false activations.

Since subjects had more difficulty with answering congruent trials with respect to neutral trials, we also investigated the group-level activation for the difference between these two trial types. Although there was some activation at the subject level, we could not find any activation at the group level.

It is possible to present the fitted cognitive waveforms to the measured signal as in Fig. 3. The large slow trend over the signal may be seen in this figure. For the case of this subject, the contrast of “incongruent vs. neutral” trials is significant while “incongruent vs. congruent” contrast is not.

Up to this point, we considered the activation detection problem. In other words, given a canonical HRF signal model, we check whether there is activation or not in our measurements. The complementary problem would be the estimation of this HRF signal. To this effect, we applied a second GLM where we modeled the HRF as successive time bins, that is, as a finite

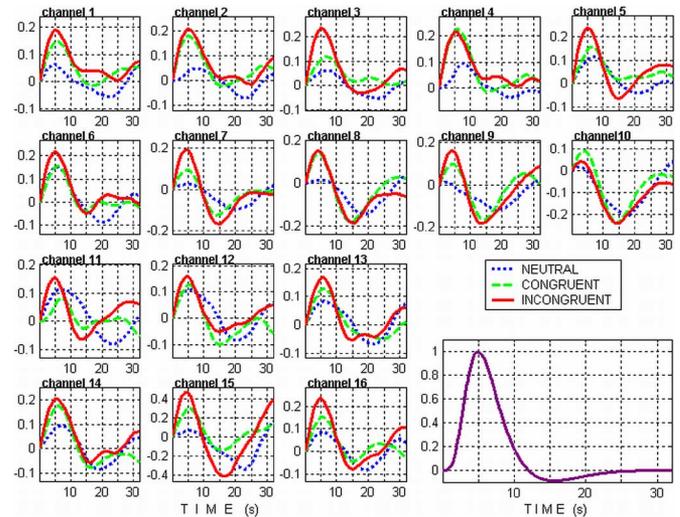


Fig. 4. Estimated hemodynamic response function waveforms averaged over subjects (running averages over 3 s), with hypothetical HRF at the bottom right.

impulse response filter. In this setting of the problem, the coefficients of the filter should give us the HRF waveform. Note that this approach does not put any constraints over the HRF, and effectively it averages the event-related responses for each subject [22]. Fig. 4 demonstrates the HRF waveforms for each type of stimulus averaged over subjects. For most of the channels, the end result is a plausible HRF waveform. We want to examine especially the waveforms acquired from channels 1–4, since BPE identified channels 1–4 as activated for “incongruent vs. neutral” contrast and channels 1 and 3 for “incongruent vs. congruent” contrast. The resulting waveforms from these channels are also consistent with this result. A caveat is that average waveforms are by no means a direct indication of group activation, but we want to point out to the consistency between the detection and estimation procedures.

2) *Deoxy-Hb Results*: Our analysis of deoxy-Hb signals did not discover as strong activation patterns as those of oxy-Hb. Fig. 5(a) shows the activations for “incongruent vs. neutral” contrast. In fact, there are activations at the subject level [Fig. 5(a), top row], and these are carried to the group level by FFX and  $\Psi$ FX; however, RFX, MFX, and BPE do not identify any of the channels as significantly activated [Fig. 5(a), bottom row]. This is a consequence of the fact that deoxy-Hb exhibits a greater variability among the subjects. To demonstrate this variability, consider Fig. 6. What we present in this figure are the subjects’ parameter estimates for the third channel of deoxy-Hb for “incongruent vs. neutral” contrast and again the third channel of oxy-Hb for “incongruent vs. congruent” contrast. We chose these combinations because deoxy-Hb shows activation for 7 subjects (out of 12) but with no group activation for RFX, MFX, and BPE, whereas oxy-Hb shows activation for 6 subjects along with group activation by the aforementioned methods. The reason for this lies in the greater variance (mainly due to the first and third subjects) exhibited by deoxy-Hb. We present the resulting activations of deoxy-Hb for “incongruent vs. congruent” contrast in Fig. 5(b).

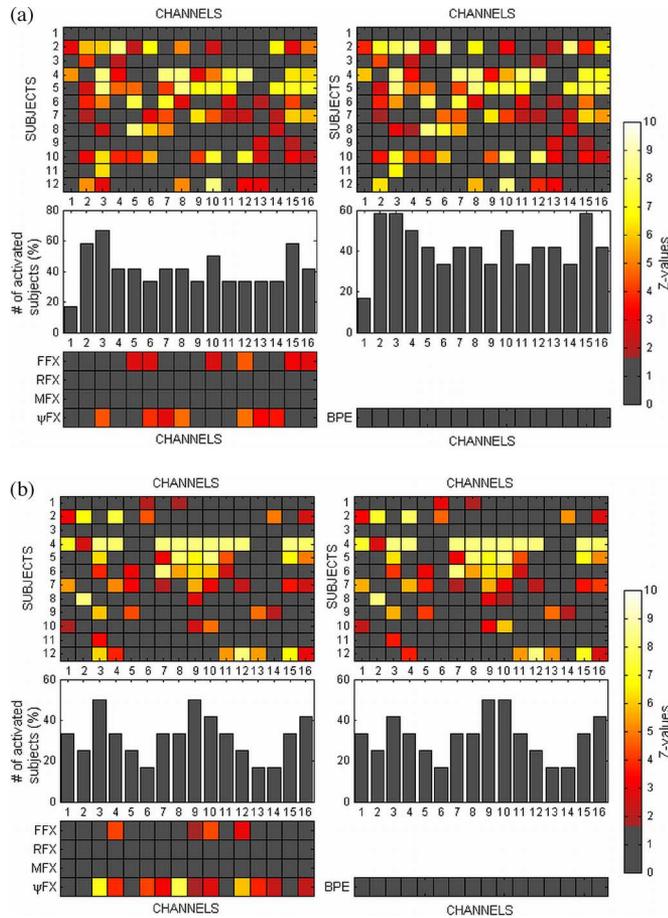


Fig. 5. (a) Activation patterns for deoxy-Hb for “incongruent–neutral” contrast. Organization of the figure is the same as that of Fig. 2(a). (b) Activation patterns for deoxy-Hb for “incongruent–congruent” contrast. Organization of the figure is the same as that of Fig. 2(a).

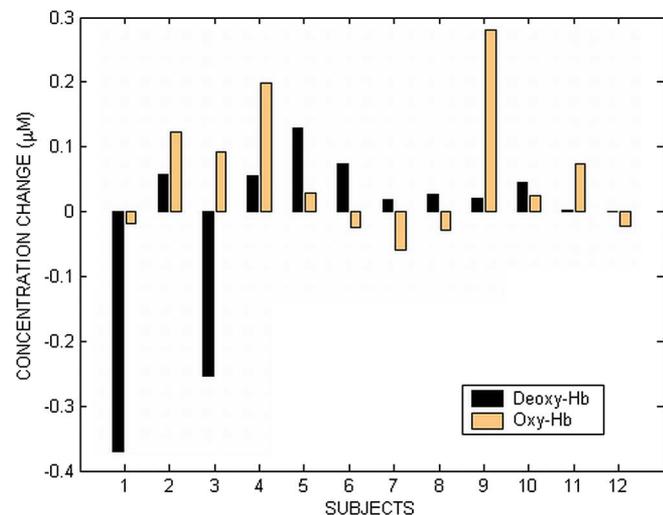


Fig. 6. Example set of contrasted subject-level parameters (see text for detailed explanation).

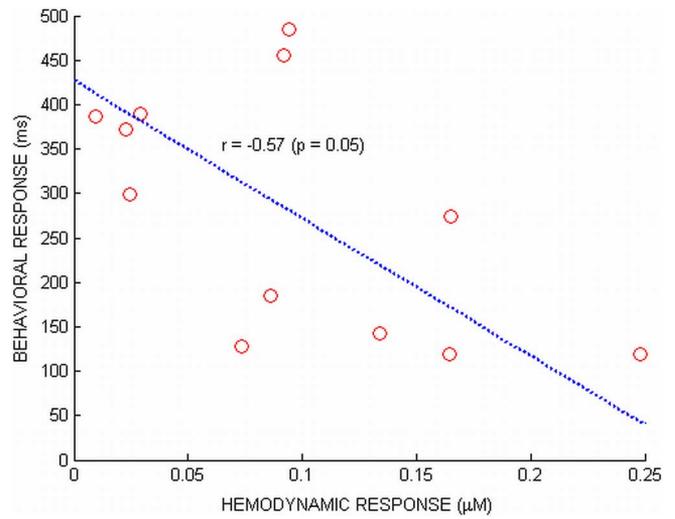


Fig. 7. Correlation between the hemodynamic and behavioral responses for oxy-Hb in the fourth channel during interference (“incongruent–neutral”) condition.

3) *Relation Between Hemodynamic and Behavioral Responses:* We investigated the relation between hemodynamic and behavioral responses by finding the channel-by-channel correlation coefficients between the interference effects measured by the difference in concentration changes and reaction times of incongruent and neutral trials. We found correlation for oxy-Hb in the fourth channel ( $r = -0.57, p = 0.05$ ). Let us remind that the fourth channel was the most consistently activated channel across subjects. Scatter plot of behavioral vs. hemodynamic response for this channel is shown in Fig. 7. Note that the correlation is negative, i.e., hemodynamic response is smaller for higher behavioral interference effect. This finding supports the hypothesis that “higher Stroop-specific brain activation leads to more successful inhibition of competing responses and hence, a smaller behavioral interference effect” [31]. Not very surprisingly, we did not find any correlation between reaction times and hemodynamic responses for the “incongruent vs. congruent” contrast of oxy-Hb and for both of the contrasts of deoxy-Hb.

## V. DISCUSSION

The main finding of this study is that fNIRS data lend itself to multilevel statistical inference. We arrived at a consistent activation pattern during Stroop interference, particularly for oxy-Hb. We recommend that the application of multilevel statistical inference to fNIRS data should always include random effects, and propose the usage of MFX or Bayesian methods. The problem with fixed effects models is that it ignores between-subject variability, and since within-subject variance is much smaller, it becomes possible for the channels to have illusory activation. To overcome this risk, extensions to FFX, like conjunction analysis, may be pursued [32].

We assert that Bayesian methodology may have a number of advantages over classical procedures in analyzing multilevel GLMs. First of all, it can cope better with the classical problem of within-subject and between-subject variances in a more

principled framework [6]. Bayesian analysis also enabled us to include the information obtained from the rest of the group in the analysis of the particular subject.

Moreover, Bayesian statistics yield posterior distributions for the parameters of interest. This enriches our statistical test dictionary, which means that we are no longer limited with just null hypothesis significance test procedure (NHSTP). Hence, we are able to test whether the effect is greater than a meaningful size in relation to underlying physiology [6]. This is important because the statistical significance obtained by NHSTP in classical statistics does not truly reflect the magnitude of the effect [33]. For example, a very small but consistent effect might be found to be statistically significant. Although a small but very reliable activation may be interesting, neuroimaging is generally interested in activations of nontrivial magnitude, and this speaks for the usefulness of Bayesian inference [6]. We want to point out that using noninformative priors carries the Bayesian inference closer to classical inference, as has also been pointed out in [34] in a different context. The rationale for our use of noninformative priors is that cognitive fNIRS studies are in their early stage of development, and we do not want to commit ourselves prematurely; furthermore, generally the number of subjects is small so as to make the influence of the prior a lot significant. One of our goals in this study was to compare classical and Bayesian inference methods for fNIRS data. Since classical procedures work with the null hypothesis and ask whether the effect size is greater than zero or not, we assigned the same threshold for the Bayesian analysis for comparison.

We tried to circumvent the multiple comparison problem arising from the simultaneous testing of a number of channels by Bonferroni correction. It is known that Bonferroni correction is too conservative, especially when there is spatial correlation between the measurements [35]. A promising method for NIRS signals was put forward using the “false discovery rate” procedure [36]. However, as also noted in that study, multiple comparison correction of multichannel NIRS studies is still an open problem. In this study, rather than using or proposing new techniques, we used the traditional Bonferroni method, and leave this problem as a further study topic.

Our Stroop findings are generally consistent with the literature, though our results are not as strong and conclusive as those of Schroeter *et al.* [14]. They showed activation bilaterally for oxy-, deoxy-, and total Hb. However, we could find activation only for oxy-Hb in the left lateral prefrontal cortex and failed to find any activation (at the group level) for deoxy-Hb. These results coincide more with those of Ehlis *et al.* [15]. They also found only left lateral activation for oxy-Hb, and that the activations for deoxy-Hb were much weaker. In a comprehensive review [12], it was concluded that the left hemisphere generally showed more interference than the right. These findings also point to an important aspect of fNIRS data analysis: the consistencies and controversies between the results obtained by oxy-Hb and deoxy-Hb. Using a visual stimulus, Schroeter *et al.* [7] arrived at the conclusion that deoxy-Hb is more amenable to GLM. However, Hoshi *et al.* [37] concluded that cortical activation could lead to different patterns in deoxy-Hb and proposed oxy-Hb as the best indicator of regional cerebral

blood flow changes. Findings of Ehlis *et al.* [15] also support this hypothesis. On a reproducibility study of event-related fNIRS, it was stated that deoxy-Hb was associated with lower  $t$ -values at single subjects’ level as well as at the second level if compared to oxy-Hb [38]. In another study on false memory on the prefrontal cortex [39], deoxy-Hb did not show any significant activations and the authors stated that this might be attributable to the instability of deoxy-Hb concentration that was largely determined by the washout effect of the regional cerebral blood flow increase [40]. In a simultaneous fMRI-fNIRS study [41], it was found that oxy-Hb was a more robust hemodynamic signal and correlated more with fMRI-BOLD response. This was attributed to the lower signal-to-noise ratio of deoxy-Hb signal. However, in another study [42], using an experimental design that increased the signal-to-noise ratio of NIRS signals, it was found that deoxy-Hb was more correlated with fMRI-BOLD signal. When evaluated together, these findings point to the fact that although oxy-Hb is more dominantly labeled as the carrier of cognitive information, the potential of NIRS for measuring cognitive activity and the interpretation of deoxy-Hb and oxy-Hb still need further research. The results of our study indicate that oxy-Hb is more sensitive to regional blood flow changes in the prefrontal cortex caused by cognitive stimulus. We found consistent left prefrontal activation for oxy-Hb during Stroop interference. The activation patterns at the subject level are more structured and the hemodynamic results show a better correlation with the behavioral results for oxy-Hb than deoxy-Hb.

## APPENDIX A

### CLASSICAL INFERENCE

We describe briefly the basic concepts of classical inference used in our research. We will mainly adopt the expositions of [16] and [43]. For the  $k$ th of  $K$  subjects, we write  $Y_k = X_k \beta_k + e_k$ , where  $Y_k$  is the  $N$ -sample fNIRS data for subject  $k$ ,  $X_k$  is the  $N \times p$  design matrix,  $\beta_k$  is the  $p$  vector of unknown parameters, and finally,  $e_k$  is the  $N$ -long error vector. If the error vector is Gaussian distributed with no temporal autocorrelation, then ordinary least squares estimate of  $\beta_k$  is given by  $\hat{\beta}_k = (X_k^T X_k)^{-1} X_k^T Y_k$ . This estimate has variance  $\text{cov}(\hat{\beta}_k) = \sigma_k^2 (X_k^T X_k)^{-1}$ , where the noise variance  $\sigma_k^2$  is estimated from the residuals. The second level of the model links the subjects’ parameters to the group parameters:  $\beta = X_G \beta_G + e_G$ , where  $\beta = [\beta_1^T \dots \beta_K^T]^T$  is  $Kp$ -dimensional concatenated parameter vector,  $X_G$  is the  $Kp \times q$  group-level design matrix,  $\beta_G$  is the  $q$  vector of group parameters, and  $e_G$  is the  $Kp$  error vector. In the summary-statistics approach to multilevel GLM, the second level of the model takes as input the estimates of the first level but not the true (and unobservable) parameters. Hence, the second-level model is modified as  $\hat{\beta} = X_G \beta_G + e_G + (\hat{\beta} - \beta) = X_G \beta_G + \hat{e}_G$ . Then, the variance of the error vector,  $\hat{e}_G$ , is  $V_{\hat{e}_G} = \text{diag}(\{\sigma_k^2 (X_k^T X_k)^{-1}\}) + \sigma_G^2 V_G$ , where  $V_G$  is the covariance matrix of the group parameters. The first component of the variance specifies within-subject variance–covariance of the parameter vector (fixed effects) and the second component indicates the between-subject

variance (random effects). Since generally the desired inference is on a particular contrast of parameters,  $c\beta_k$ ,  $\hat{\beta}$  becomes  $\hat{\beta}_{\text{cont}} = [c\hat{\beta}_1 \cdots c\hat{\beta}_K]^T$ . Subject-level error variances then become  $\text{cov}(c\hat{\beta}_k) = \sigma_k^2 c(X_k^T X_k)^{-1} c^T$ , and  $V_G$  has a simple form, typically  $I_K$ . Summary-statistics MFX procedure accounts for both of these sources of variance whereas FFX and RFX ignore the second and first components of the variance, respectively.

## APPENDIX B

### BAYESIAN INFERENCE

The Bayesian analysis (BPE) we used is summarized here. BPE takes into consideration the fact that the second level of the model imposes a prior distribution on the parameters of the first level. Consequently, the conditional posterior pdfs of the variables can be derived using the Bayes rule (posterior  $\propto$  likelihood  $\times$  prior):

$$p(\beta_k | M, \text{r.v.}) \propto p(\beta_k | M, \beta_G, V_G) p(Y_k | M, \beta_k, \sigma_k^2)$$

$$p(\sigma_k^2 | M, \text{r.v.}) \propto p(\sigma_k^2 | M) p(Y_k | M, \beta_k, \sigma_k^2)$$

$$p(\beta_G | M, \text{r.v.}) \propto p(\beta_G | M) \prod_{k=1:K} p(\beta_k | M, \beta_G, V_G)$$

$$p(V_G | M, \text{r.v.}) \propto p(V_G | M) \prod_{k=1:K} p(\beta_k | M, \beta_G, V_G)$$

where  $M$  and r.v. stand for the *model* and *remaining variables*, respectively; all the other variables being as in Appendix A. The consequence of assuming Gaussian distributions for noise vectors and using noninformative priors is that conditional posterior pdfs have analytical forms. The conditional posterior of subjects' parameter vectors are proportional to the product of two Gaussians; hence, they are also Gaussian. Actually, subjects' parameters are estimated from data and instantaneous group parameter estimates inversely weighted with their corresponding variance estimates. Subjects' variances have, with a noninformative prior, a Gamma conditional posterior pdf. Noninformative prior for group parameter vector is the uniform distribution; hence, its conditional posterior is just the likelihood formed by subjects' parameter vectors, and consequently, has a Gaussian distribution. Finally, group covariance matrix has conditionally an inverse-Wishart distribution (for more information about these distributions and hierarchical models, see [20]).

## REFERENCES

- [1] A. Villringer and B. Chance, "Noninvasive optical spectroscopy and imaging of human brain function," *Trends Neurosci.*, vol. 20, pp. 4435–442, 1997.
- [2] M. Cope and D. T. Delpy, "A system for the long-term measurement of cerebral blood and tissue oxygenation in newborn infants by near infra-red transillumination," *Med. Biol. Eng. Comput.*, vol. 26, pp. 289–294, 1988.
- [3] I. Palada, A. Obad, D. Bakovic, Z. Valic, V. Ivancev, and Z. Dujic, "Cerebral and peripheral hemodynamics and oxygenation during maximal dry breath-holds," *Respir. Physiol. Neurobiol.*, vol. 157, no. 2/3, pp. 374–381, Aug. 2007 (doi: 10.1016/j.resp.2007.02.002).
- [4] D. T. Ubbink and B. Koopman, "Near-infrared spectroscopy in the routine diagnostic workup of patients with leg ischaemia," *Eur. J. Vasc. Endovasc. Surg.*, vol. 31, pp. 394–400, 2006.
- [5] K. Friston, A. Holmes, J. B. Poline, C. Frith, and R. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapp.*, vol. 2, pp. 189–210, 1995.
- [6] K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: Theory," *Neuroimage*, vol. 16, pp. 465–483, 2002.
- [7] M. L. Schroeter, M. M. Bücheler, K. Müller, K. Uludag, H. Obrig, G. Lohmann, M. Tittgemeyer, A. Villringer, and D. Y. von Cramon, "Toward a standard analysis for functional near-infrared imaging," *Neuroimage*, vol. 21, pp. 283–290, 2004.
- [8] 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: A parametric validation study," *Neuroimage*, vol. 35, no. 2, pp. 625–634, Apr. 2007 (doi: 10.1016/j.neuroimage.2006.11.028).
- [9] J. Cohen-Adad, S. Chapuisat, J. Doyon, S. Rossignol, J.-M. Lina, H. Benali, and F. Lesage, "Activation detection in diffuse optical imaging by means of the general linear model," *Med. Image Anal.*, vol. 11, no. 6, pp. 616–629, Jun. 2007 (doi: 10.1016/j.media.2007.06.002).
- [10] B. Thirion, P. Pinel, S. Meriaux, A. Roche, S. Dehaene, and J. B. Poline, "Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses," *Neuroimage*, vol. 35, pp. 105–120, 2007.
- [11] G. K. Büyükkaksoy, N. S. Şengör, H. Gürvit, and C. Güzeliş, "Modelling the stroop effect: A connectionist approach," *Neurocomputing*, vol. 70, no. 7/9, pp. 1414–1423, Mar. 2007 (doi: 10.1016/j.neucom.2006.05.009).
- [12] C. MacLeod, "Half a century of research on the Stroop effect: An integrative review," *Psychol. Bull.*, vol. 109, pp. 163–203, 1991.
- [13] S. Zysset, K. Müller, G. Lohmann, and D. Y. von Cramon, "Color–word matching Stroop task: Separating interference and response conflict," *Neuroimage*, vol. 13, pp. 29–36, 2001.
- [14] 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," *Human Brain Mapp.*, vol. 17, pp. 61–71, 2002.
- [15] A.-C. Ehlis, M. J. Herrmann, A. Wager, and A. J. Fallgatter, "Multichannel near-infrared spectroscopy detects specific inferior-frontal activation during incongruent Stroop trials," *Biol. Psychol.*, vol. 69, pp. 315–332, 2005.
- [16] 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.
- [17] M. W. Woolrich, T. E. J. Behrens, C. F. Beckmann, M. Jenkinson, and S. B. Smith, "Multilevel linear modeling for fMRI group analysis using Bayesian inference," *Neuroimage*, vol. 21, pp. 1732–1747, 2004.
- [18] J. Neumann and G. Lohmann, "Bayesian second-level analysis of functional magnetic resonance images," *Neuroimage*, vol. 20, pp. 1346–1355, 2003.
- [19] D. B. Rowe, *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*. Boca Raton, FL: Chapman & Hall/CRC, 2003.
- [20] A. B. Gelman, J. S. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC, 1995.
- [21] A. Akin, D. Bilensoy, U. E. Emir, M. Gülsoy, S. Candansayar, and H. Bolay, "Cerebrovascular dynamics in patients with migraine: Near-infrared spectroscopy study," *Neurosci. Lett.*, vol. 400, pp. 86–91, 2006.
- [22] 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.
- [23] C. B. Akgül, B. Sankur, and A. Akin, "Spectral analysis of event-related hemodynamic responses in functional near infrared spectroscopy," *J. Comput. Neurosci.*, vol. 18, pp. 67–83, 2005.
- [24] D. A. Boas, K. Chen, D. Grebert, and M. A. Franceschini, "Improving the diffuse optical imaging spatial resolution of the cerebral hemodynamic response to brain activation in humans," *Opt. Lett.*, vol. 29, no. 1, pp. 1506–1508, Jan./Feb. 2004.
- [25] F. Fabbri, A. Sassaroli, M. E. Henry, and S. Fantini, "Optical measurements of absorption changes in two-layered diffusive media," *Phys. Med. Biol.*, vol. 49, pp. 1183–1201, 2004.
- [26] M. Firbank, E. Okada, and D. T. Delpy, "A theoretical study of the signal contribution of regions of the adult head to near-infrared spectroscopy studies of visual evoked responses," *Neuroimage*, vol. 8, pp. 69–78, 1988.
- [27] K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg, and R. Turner, "Event-related fMRI: Characterizing differential responses," *Neuroimage*, vol. 7, pp. 30–40, 1998.

- [28] K. J. Friston, O. Josephs, E. Zarahn, A. P. Holmes, S. Rouquette, and J.-B. Poline, "To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis," *Neuroimage*, vol. 12, pp. 196–208, 2000.
- [29] R. L. C. Mitchell, "BOLD response during Stroop task-like inhibition tasks: Effects of task difficulty and task-relevant modality," *Brain Cogn.*, vol. 59, pp. 23–37, 2005.
- [30] C. Bench, C. Frith, P. Grasby, K. Friston, E. Paulesu, R. Frackowiak, and R. Dolan, "Investigations of the functional anatomy of attention using the Stroop test," *Neuropsychologia*, vol. 31, pp. 907–922, 1993.
- [31] M. L. Schroeter, S. Cutini, M. M. Wahl, R. Scheid, and D. Y. von Cramon, "Neurovascular coupling is impaired in cerebral microangiopathy—An event related Stroop study," *Neuroimage*, vol. 34, pp. 26–34, 2007.
- [32] R. Heller, Y. Golland, R. Malach, and Y. Benjamini, "Conjunction group analysis: An alternative to mixed/random effect analysis," *Neuroimage*, vol. 37, pp. 1178–1185, 2007.
- [33] S. L. Chow, *Statistical Significance: Rationale, Validity and Utility*. London, U.K.: Sage, 1996.
- [34] M. W. Woolrich, T. E. J. Behrens, and S. M. Smith, "Constrained linear basis sets for HRF modeling using variational Bayes," *Neuroimage*, vol. 21, pp. 1748–1761, 2004.
- [35] B. R. Logan and D. B. Rowe, "An evaluation of thresholding techniques in fMRI analysis," *Neuroimage*, vol. 22, pp. 95–108, 2004.
- [36] A. K. Singh and I. Dan, "Exploring the false discovery rate in multichannel NIRS," *Neuroimage*, vol. 33, pp. 542–549, 2004.
- [37] Y. Hoshi, N. Kobayashi, and M. Tamura, "Interpretation of near-infrared spectroscopy signals: A study with a newly developed perfused rat brain model," *J. Appl. Physiol.*, vol. 90, no. 5, pp. 1657–1662, 2001.
- [38] M. M. Plichta, M. J. Herrmann, C. G. Bachne, A.-C. Ehlis, M. M. Richter, P. Pauli, and A. J. Fallgatter, "Even-related functional near-infrared spectroscopy (fNIRS): Are the measurements reliable?," *Neuroimage*, vol. 31, pp. 116–124, 2006.
- [39] Y. Kubota, M. Toichi, M. Shimizu, R. A. Mason, R. L. Findling, K. Yamamoto, and J. R. Calabrese, "Prefrontal hemodynamic activity predicts false memory—A near-infrared spectroscopy study," *Neuroimage*, vol. 31, pp. 1783–1789, 2006.
- [40] R. B. Buxton, E. C. Wong, and L. R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: The balloon model," *Magn. Reson. Med.*, vol. 39, pp. 855–864, 1998.
- [41] G. Strangman, J. P. Culver, J. H. Thompson, and D. A. Boas, "A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation," *Neuroimage*, vol. 17, pp. 719–731, 2002.
- [42] T. J. Huppert, R. D. Hoge, S. G. Diamond, M. A. Franceschini, and D. A. Boas, "A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans," *Neuroimage*, vol. 29, pp. 368–382, 2006.
- [43] J. A. Mumford and T. Nichols, "Modeling and inference of multisubject fMRI data," *IEEE Eng. Med. Biol. Mag.*, vol. 25, no. 2, pp. 42–51, Mar./Apr. 2006.



**Koray Çiftçi** received the B.S. degree in electrical and electronic engineering in 1996 and the M.S. and Ph.D. degrees in biomedical engineering in 2001 and 2008, respectively, from Boğaziçi University, Istanbul, Turkey.

He is with the Lung Acoustics Laboratory and Biophotonics Laboratory, Boğaziçi University. His current research interests include applied biomedical signal processing and biomedical instrumentation.



**Bülent Sankur** (S'70–M'76–SM'90) received the B.S. degree in electrical engineering from Robert College, Istanbul, Turkey, and the M.Sc. and Ph.D. degrees from Rensselaer Polytechnic Institute, Troy, NY.

He is currently in the Department of Electrical and Electronic Engineering, Boğaziçi (Bosphorus) University, Istanbul, where he founded and has been the Leader of the Image and Signal Processing Laboratory. He has also been serving in various industrial consulting tasks. His current research interests include digital signal processing, image and video compression, biometry, cognition, and multimedia systems. He has held visiting positions at the University of Ottawa, Technical University of Delft, and Ecole Nationale Supérieure des Télécommunications, Paris, France.

Dr. Sankur was the Chairman of the International Conference on Telecommunications (ICT 1996) and the European Conference on Signal Processing (EUSIPCO 2005). He has also been the Technical Chairman of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2000).



**Yasemin P. Kahya** (M'92) received the B.S. degree (with highest honors) in electrical engineering and physics and the Ph.D. degree in biomedical engineering from Boğaziçi University, Istanbul, Turkey, in 1980 and 1987, respectively, and the M.S. degree in engineering and applied science from Yale University, New Haven, CT, in 1981.

She has been in the Department of Electric and Electronic Engineering, Boğaziçi University. Her current research interests include biomedical instrumentation, respiratory acoustics, and biomedical signal processing applications.

Dr. Kahya was the Program Co-Chair of the IEEE Engineering in Medicine and Biology Society (EMBS) 2001 Conference, Istanbul.



**Ata Akin** (M'94) received the B.S. degree in electronics and telecommunication engineering and the M.S. degree in biomedical engineering from Istanbul Technical University, Istanbul, Turkey, in 1993 and 1995, respectively, and the Ph.D. degree in biomedical engineering from Drexel University, Philadelphia, PA, in 1998.

From 1999 to 2002, he was a Research Assistant Professor in the School of Biomedical Engineering, Drexel University. In 2002, he joined the Institute of Biomedical Engineering, Boğaziçi University, Istanbul, as an Assistant Professor, and co-founded the Biophotonics Laboratory. His current research interests include biomedical optics and functional neuroimaging. In biomedical optics, he designs and develops medical instruments to monitor the brain and muscle oxygenation activity during health and disease, and in neuroimaging, he develops algorithms and techniques in modeling and analyzing the signals and images obtained from the body.

Dr. Akin is a Fellow of the National Institutes of Health–Fogarty International Research Collaboration Award (NIH-FIRCA).