Client-Centred Music Imagery Classification Based on Hidden Markov Models of Baseline Prefrontal Hemodynamic Responses

Tiago H. Falk, Kelly M. Paton and Tom Chau

Additional information is available at the end of the chapter

1. Introduction

A prevalent avenue in rehabilitation engineering is the development of new access technologies which enable individuals with severe and/or multiple disabilities to access or interact with their environment. Despite continued efforts to develop such technologies, many individuals still possess no means to communicate or control external devices [1]. Brain-computer interfaces (BCIs) have surfaced as promising access solutions for those with severe motor disabilities such as late-stage amyotrophic lateral sclerosis (ALS), brainstem stroke, or severe cerebral palsy [2]. Electroencephalography (EEG), a noninvasive method used in traditional BCI technologies, is safe and inexpensive, but long-term electrode fixation is difficult and the signal is inherently noisy [1, 3]. Other common noninvasive BCI technologies such as functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and positron emission topography (PET) are theoretically feasible, but are expensive and technically demanding for online applications [3]. More recently, the possibility of using near infrared spectroscopy (NIRS) as a tool in this field has been investigated due to its affordability, portability, high temporal resolution (similar to that of fMRI), and higher signal-to-noise ratio than that of EEG due to the absence of artifacts [3–9].

Near-infrared spectroscopy is an analysis technique that operates by transmitting near-infrared (650 nm – 950 nm wavelengths) electromagnetic radiation through the skull and comparing the intensities of the returning and incident light. When functional mental activity occurs, metabolic demands cause an increase in the blood flow to pre-defined regions of the brain (e.g., motor cortex during motor imagery tasks [3]). The increase in blood flow causes changes to the regional concentrations of oxygenated and deoxygenated hemoglobin,

hence altering the optical properties of the brain tissue [10, 11]. As a consequence, by measuring the intensity of the reflected light, NIRS technologies can be used to examine regional cortical activity with depths of up to 2 cm [10]. Previous research has shown that NIRS can be used to assess hemodynamic responses in the motor cortex using activities such as motor imagery [3, 7], as well as in the prefrontal cortex using higher cognitive tasks such as mental arithmetic [6, 11, 12], working memory [13, 14], and emotion induction [15, 16], both in silent (e.g., [17, 18]) and noisy environments [19].

A major drawback of measuring hemodynamic responses in the motor cortex is the interference of hair [8]. Additionally, for individuals who have congenital motor impairments or who are many years post-traumatic injury, motor imagery may be a difficult, or even impossible, task due to the absence of a somatosensory map of motor activities [20–22]. These problems can be avoided by using more intuitive tasks which activate the prefrontal cortex (PFC). In this study, we proposed to classify hemodynamic responses in the PFC resultant from an intuitive music imagery task. More specifically, we were interested in exploiting the emotional response that music imagery can have on an individual. We were motivated by previous studies which have shown that music can elicit [23] and enhance [24] intense emotional responses that activate brain regions believed to be associated with emotional behaviours, such as the PFC [25] and the orbitofrontal and frontopolar areas [26, 27]. More importantly, prefrontal hemodynamic responses due to subject-selected music imagery have been recently observed using fMRI [28]. An additional motivation for using the music imagery task was related to the simplicity of the cognitive task | relative to mental arithmetic, for example | which would allow younger children to use this classification system. Recent experiments have suggested that NIRS responses in the PFC resultant from music imagery are detectable and sufficiently different from those resultant from mental arithmetic tasks and baseline resting states [17, 18, 29].

The goal of this study was to investigate the potential of classifying rest and music imagery on the basis of non-invasively acquired NIRS signals. Automated detection of music imagery events could then be used to control a system-paced two-state BCI (e.g., a binary switch). To develop a music imagery classifier that is applicable to users with different characteristics (e.g., age, gender, disability) requires a significant BCI–user coordination effort [2]. For the task at hand, we have incorporated a client-centred paradigm where subject-specific physical and physiological factors such as movement, heartbeat, and respiration were taken into account during classifier development [10]. Additional motivations for pursuing a client-driven paradigm included *i)* dependency of the hemodynamic responses on the intensity and valence (i.e., happy or sad) of the emotions evoked [30], *ii)* the mode (major or minor) and tempo (fast or slow) of the imagined songs [31], and *iii)* the differences in age, race, gender, skull thickness and skull shapes which may lead to inter-subject variations in penetration depths of NIR light [7, 32]. In this paper, hidden Markov models (HMMs), which have been previously used for motor imagery classification [3, 33], were used to characterize the individuals' hemodynamic responses during rest (baseline). Client-centredness was achieved by optimizing HMM model parameters and classifier input parameters on a per-client basis.

2. Experimental methods

This section describes the experiment setup, execution, and data pre-processing.

2.1. Participants

Thirteen able-bodied adults (three male, mean age of 33.5 ± 12.8 years) were recruited from the University of Toronto and Bloorview Kids Rehab (Toronto). Exclusion criteria were metabolic, cardiovascular, respiratory, psychiatric, or drug- or alcohol-related conditions that could affect either the measurements or the participant's ability to follow the experimental protocol. Participants were required to have normal hearing. Ethical approval was obtained from the relevant institutions, and all participants provided informed signed consent.

2.2. Instrumentation

The hemodynamic response was recorded from each subject using the Imagent Function Brain Imaging System from ISS Inc. Two photomultiplier tube detectors were employed along with sixteen light sources — eight at 690 nm and eight at 830 nm. The sources delivered 110 MHz-modulated light to the forehead via 400 μm-diameter optical fibres, and the detectors received the returning light via 3 mm-diameter optical fibres. Light returned to the detectors was demodulated at a cross-correlation frequency (CCF) of 5 kHz. To avoid cross-signal contamination, the light sources were cyclically switched such that no two sources were on simultaneously. For one data collection cycle — defined as one complete sequence through all sixteen sources — each source remained on for eight periods of the CCF (i.e., 1.6 ms), separated by a two-period break (0.4 ms) to ensure no overlap. This resulted in an effective sampling rate of 31.25 Hz per full data collection cycle. For each cycle, the eight waveforms from each source-detector pair were averaged, and a fast Fourier transform (FFT) was applied to the result in order to obtain three output data components: AC (relative amplitude at the CCF), DC (relative amplitude at 0 Hz), and phase. In this study, only the AC and DC data were used.

The sixteen source fibres were grouped in pairs comprised of one source at each of the two wavelengths; this allowed a given location to be probed by both wavelengths simultaneously. Positioning four source pairs around one detector on each side of the participant's forehead allowed both the right and left prefrontal cortices to be probed, as shown in Figure 1(a). Emitters in position 3 on each side were located roughly at the left and right prefrontal cortices (FP1 and FP2 in the 10-20 system). The raw AC and DC signals obtained from each source can be represented as $x^{\lambda}_{p,i}(\gamma)$ where the subscript i indexes the source position ($i = 1, \ldots 4$), subscript p indicates the left or right side by L or R, and superscript λ indicates the wavelength ($\lambda =$690, 830 nm). The signal type (AC or DC) is indicated by the parameter γ.

On each side, the four source pairs were positioned 2.12 cm away from the detector (see Figure 1(a)). It has been shown (in the occipital region) that, although a larger source–detector distance (3.37 cm to 4.55 cm) gives a greater magnitude of response, the same features can be seen in a signal obtained at a distance of 2.25 cm [34]. While the chosen distance of 2.12 cm is slightly smaller, the region of interest is the prefrontal cortex, which previous NIRS studies have shown to have attenuation 1000 times smaller than the occipital region [34]. Such insight suggests hemodynamic responses should be observable with a source–detector separation of 2.12 cm. The promising results reported in 4, as well as the more recent results reported in [35], corroborate this assumption.
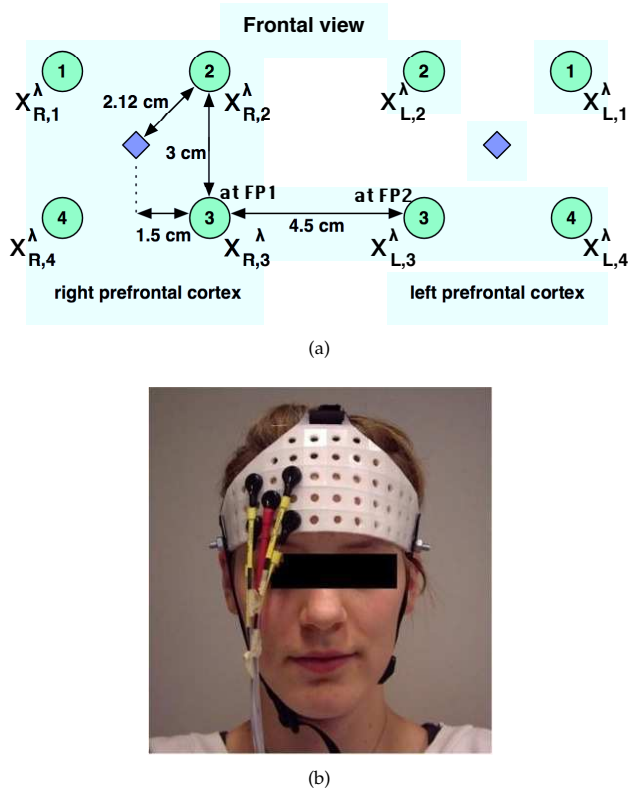
(a)

(b)

Figure 1. (a) Positioning of source pairs (circles) and detectors (diamonds). Each source pair represents one $\lambda =$690 nm source and one $\lambda =$830 nm source. The signal from each source is represented as $x_{p,i}^{\lambda}$, with $i = 1, \ldots 4$ and $p = L$ (left) or R (right). Source pairs at position 3 on each side are located, anatomically, at the prefrontal cortices FP1 and FP2. (b) Image of headband position on the forehead with only the right side detector and sources in place.

Sources and detectors were secured to the forehead with a specially designed headband constructed from 1.6 mm low density polyethylene, as shown in Figure 1(b). The area probed using this arrangement corresponds approximately to the frontopolar cortex, the superior portion of the orbitofrontal cortex, and the more medial sections of the dorsolateral prefrontal cortex. Subjects wore a blindfold and industrial earmuffs (AOSafety Professional Hearing Protector, noise rating of 30 dB) for the entire data collection duration in order to eliminate any visual and/or auditory distractions external to the experiment.

2.3. Protocol

NIRS signals were collected from each participant while they were outfitted with the headband, blindfold, and earmuffs and seated comfortably in front of a computer in a quiet room. Each subject performed four trials spread over two separate sessions (i.e., two trials were completed per session); each session lasted about 40 minutes, including the time for experiment preparation (e.g., headband placement) and NIRS sensor calibration. Each
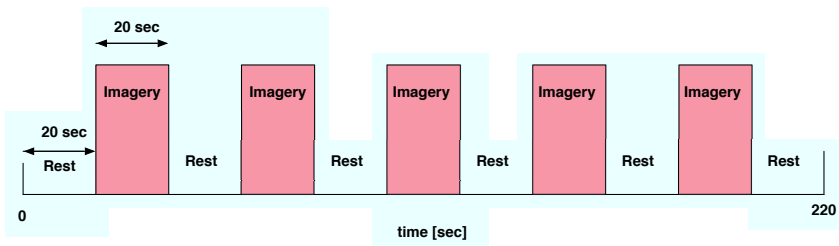
Figure 2. Stimulus pattern for the imagery trials. Imagery periods are shaded and rest intervals are not shaded.

session was performed on a separate day to ensure maximum concentration and focus and to reduce any effects of discomfort due to the headband. Participants were asked to remain as still as possible and to refrain from speaking during all experimental trials. Halfway through each session, the participant was given a short break to alleviate fatigue. While the blindfold and earmuffs were removed during the break, the NIRS sensors remained in place so as to preserve calibration and positioning.

Two of the trials consisted of the participant in a static state of rest with no music imagery (referred to as 'baseline' trials). Participants were instructed to clear their minds as much as possible, and to focus on their breathing for approximately two minutes. For the remaining two trials, participants were informed that they were required to sing *one* of their chosen songs vividly in their head when cued by the experimenter via a light tap in the shoulder (referred to as 'imagery' trials). Prior to each session, participants were asked to choose several songs of their own preference that were of the same emotional valence (i.e. happy or sad), which they felt elicited a strong emotional reaction. Participants were informed that the purpose of the music imagery task was to elicit an intense emotional reaction. As a consequence, they were instructed to switch to a new song once they began to feel emotionally habituated to the currently-used song.

Figure 2 shows the stimulus pattern used for the 220-second imagery trials. As can be seen, each imagery trial consisted of eleven 20-second intervals alternating between rest (six intervals) and music imagery (five intervals). An activation window of 20 seconds was chosen to take into account the task initiation delay, which we have found in previous experiments to be as long as 5 seconds, as well as the inherent physiological latency of the hemodynamic responses, which can be in the order of 5-10 seconds [3]. For similar reasons, twenty-second rest intervals were chosen. Such choice of stimulus pattern, however, can reinforce slow 3rd-order blood pressure waves and its subharmonics (circa 0.05-0.1 Hz); thus, a wavelet-based denoising algorithm was used, as described in 2.4.

2.4. Pre-processing

The raw signals (AC and DC) were filtered to mitigate physiological noise due, primarily, to cardiac signals (0.5–2 Hz), respiration (0.2–0.4 Hz) and the Mayer wave (approximately 0.1 Hz) [10], as well as the low-frequency artifacts (e.g., 3rd-order blood pressure waves) that may have been reinforced by the selected stimulus timing pattern. Since the frequencies of physiological noise and the desired hemodynamic response may vary slightly among participants, wavelet-based filters that would remove low-powered noise without relying on a specific cutoff frequency were used. Wavelet denoising has been shown to be effective

for functional NIRS hemodynamics-driven signals [10], hence three wavelet filters were investigated which performed a 12-level decomposition using a Daubechies-12 wavelet. The filtered signals resulted from the reconstruction of the approximation wavelet coefficients and either the last three, four or five detail coefficients (henceforth referred to as 3-, 4-, and 5-detail wavelet filtered signals, respectively). A visual representation of the raw and filtered signals can be seen in Figure 3. The filtered signals are denoted as $x_{p,i}^{\lambda}(\gamma, d)$, with d indicating the 'detail' of the filtered signal ($d = 3, 4, 5$) and λ, p, i, and γ as before.

2.5. Feature extraction

The estimated changes in the concentration of oxygenated hemoglobin ($\Delta[HbO_2]$) and deoxygenated hemoglobin ($\Delta[HHb]$) were also computed, using the modified Beer-Lambert law [3] as shown in Eq. 1 and Eq. 2:

$$\Delta[HbO_2] = \frac{(\varepsilon_{Hb}^{690nm} \cdot \frac{\Delta OD^{830nm}}{DPF^{830nm}}) - (\varepsilon_{Hb}^{830nm} \cdot \frac{\Delta OD^{690nm}}{DPF^{690nm}})}{r \cdot (\varepsilon_{Hb}^{690nm} \cdot \varepsilon_{HbO}^{830nm} - \varepsilon_{Hb}^{830nm} \cdot \varepsilon_{HbO}^{690nm})}, \tag{1}$$

$$\Delta[HHb] = \frac{(\varepsilon_{HbO}^{830nm} \cdot \frac{\Delta OD^{690nm}}{DPF^{690nm}}) - (\varepsilon_{HbO}^{690nm} \cdot \frac{\Delta OD^{830nm}}{DPF^{830nm}})}{r \cdot (\varepsilon_{Hb}^{690nm} \cdot \varepsilon_{HbO}^{830nm} - \varepsilon_{Hb}^{830nm} \cdot \varepsilon_{HbO}^{690nm})},$$

where

$$\Delta OD^{\lambda} = \log \frac{DC_0^{\lambda}}{DC^{\lambda}(t)}. \tag{2}$$

Parameter ΔOD^{λ} corresponds to the change in optical density of light at wavelength λ; DC_0^{λ} to the mean DC intensity; $DC^{\lambda}(t)$ to the DC intensity at time t; and r to the source-detector separation distance. Constant DPF^{λ} corresponds to the differential pathlength factor in the human adult head at wavelength λ, and $\varepsilon_{Hb}^{\lambda}$ and $\varepsilon_{HbO}^{\lambda}$ correspond to extinction coefficients for HHb and HbO_2 cromophores, respectively. Table 1 reports the constant values for each wavelength used in the experiments described herein. Concentration signals can be represented as $c_{p,i}^{h}(d)$, where the superscript h indexes oxygenated (HbO) or deoxygenated hemoglobin (Hb), and $p=L, R$, $i = 1, \ldots 4$, and $d = 3,4,5$ as before.

In addition to each individual channel, the mean left and right responses were also computed by averaging like-wavelength raw signals or like-chromophore concentration signals over the four measurement positions on each side, resulting in:

$$\bar{x}_p^{\lambda}(\gamma, d) = \frac{1}{4} \sum_{i=1}^{4} x_{p,i}^{\lambda}(\gamma, d), \tag{3}$$

$$\bar{c}_p^{h}(d) = \frac{1}{4} \sum_{i=1}^{4} c_{p,i}^{h}(d). \tag{4}$$

Wavelength	Variable	Value
	ε_{HbO}	$0.3123\ mM^{-1}cm^{-1}$
690 nm	ε_{Hb}	$2.1382\ mM^{-1}cm^{-1}$
	DPF	6.51 cm
	ε_{HbO}	$1.0507\ mM^{-1}cm^{-1}$
830 nm	ε_{Hb}	$0.7804\ mM^{-1}cm^{-1}$
	DPF	5.86 cm

Table 1. Differential pathlength factors (DPF) and extinction coefficients (ε) used in hemoglobin concentration calculations. Values are taken from [36] and [37].
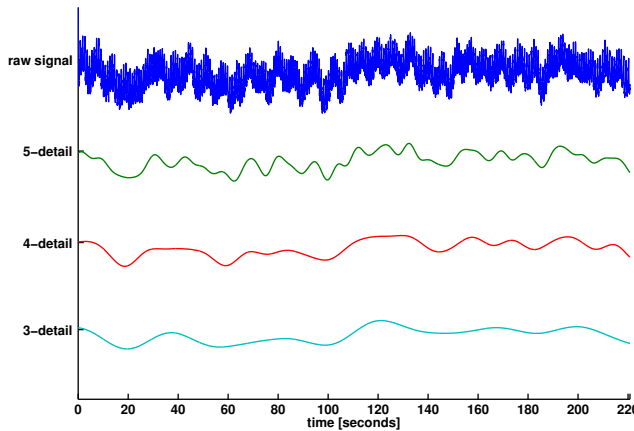

Figure 3. Comparison of raw DC data obtained during an imagery trial with its three different wavelet-filtered signals. From the top down, the signals are: raw, 5-, 4-, and 3-detail filtered.

3. Data analysis

This section discusses basic HMM theory and the application of HMMs to characterize baseline signals and detect music imagery events.

3.1. Hidden Markov models

A hidden Markov model is a statistical model which examines a Markov process wherein the observable outputs are dependent upon the unobservable states. Here, only a brief description of HMMs is given and the reader is referred to [38] for further details. An HMM representing the feature vector $\vec{u}$ (see 3.2.1) is completely characterized by the number of Q discrete states, a transition matrix $A = \{a_{ij}\}$ of transition probabilities between states i and j, an initial state distribution π and a state-dependent observation probability distribution $B = \{b_j(\vec{u})\}$ for state j. Commonly, Gaussian mixture models (GMM) are employed as observation probability distributions. A GMM is given by a weighted sum of M component densities,

$$P(\vec{u}) = \sum_{k=1}^{M} \alpha_k p_k(\vec{u}) \tag{5}$$

where $\alpha_k \geq 0$ are the mixture weights with $\sum_{k=1}^{M} \alpha_k = 1$, and $p_k(\vec{u})$ are N-dimensional Gaussian densities with mean $\vec{\mu}_k$ and covariance matrix Σ_k, given by:

$$p_k(\vec{u}) = \frac{1}{(2\pi)^{N/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(\vec{u} - \vec{\mu}_k)^\top \Sigma_k^{-1}(\vec{u} - \vec{\mu}_k)\right). \tag{6}$$

A GMM with $M = 1$ indicates a conventional Gaussian density.

In this study, model parameters, such as state transition probabilities, initial state probabilities, and output distribution parameters, were computed using a recursive greedy-EM algorithm where model parameters and the number of Gaussian components were estimated simultaneously using a Bayesian information criterion to avoid overfitting [39]. Mean left and right concentration and raw features extracted from the imagery data were assessed against the baseline reference models via the log likelihood measure (see 3.2.1). Log likelihood values were computed using the so-called forward-backward procedure described in [38]. Normalization was performed based on the number of data points in the sampled signal under test. As mentioned previously, client-centredness was achieved by optimizing HMMs parameters for each individual participant. More details about HMM training and testing are described in 3.2.1 and 3.2.2, respectively. The publicly available HMM Toolbox for Matlab was used for the simulations [40].

3.2. Modeling process

3.2.1. Training stage

For each participant, HMMs were trained for different combinations of AC, DC, or concentration features using the signals obtained during the baseline trials. More specifically, the four-dimensional feature vectors included:

$$\vec{u}_{DC,d} = [\bar{x}_L^{690}(DC,d), \bar{x}_L^{830}(DC,d), \bar{x}_R^{690}(DC,d), \bar{x}_R^{830}(DC,d)], \tag{7}$$

$$\vec{u}_{AC,d} = [\bar{x}_L^{690}(AC,d), \bar{x}_L^{830}(AC,d), \bar{x}_R^{690}(AC,d), \bar{x}_R^{830}(AC,d)], \tag{8}$$

$$\vec{u}_{conc,d} = [\bar{c}_L^{Hb}(d), \bar{c}_L^{HbO}(d), \bar{c}_R^{Hb}(d), \bar{c}_R^{HbO}(d)], \tag{9}$$

where d indexes 3-, 4-, or 5-detail wavelet filters.

During hidden Markov model training, the number of parameters that need to be estimated depends on the number of HMM states Q, type of HMM (e.g., fully connected, left-right),

number of Gaussian components M, data dimensionality K, and GMM covariance matrix type (i.e., diagonal or full). In this study, fully-connected HMMs and full covariance Gaussian components were used to explore correlations between NIRS signals measured from neighboring channels. Preliminary experiments with the recursive EM algorithm showed that the following HMM-GMM configurations were used most often:

- $Q=4$, $M=1$,
- $Q=2$, $M=1$,
- $Q=2$, $M=2$.

Such configurations were consistent with those reported in previous HMM-based BCI studies (e.g., [3, 33]). Participant-specific models were trained using the feature vectors $\vec{u}_{DC,d}$, $\vec{u}_{AC,d}$, and $\vec{u}_{conc,d}$, for all three types of filtered data (d = 3, 4, and 5) and all three combinations of HMM parameters listed above. This resulted in 27 (3 data types $\times$ 3 filters $\times$ 3 model parameter sets) different models for each of the thirteen participants, or 9 models per data type per participant.

3.2.2. Testing stage

For testing, consecutive, overlapping data samples from the imagery trials were scored against the per-subject baseline reference HMMs via the log likelihood measure. Sliding-window samples of various lengths (l = 1, 3, 5, 7, 10, and 15 seconds) were investigated with 0.5–second overlap and the log likelihood for a window was normalized by the length of the window. Window lengths of different sizes were investigated for two reasons. First, NIRS signal propagations have varied delay times ranging from 5–8 seconds [3]. Second, reaction times differ between subjects due to various external factors such as mental alertness, innate reflex reaction time, and familiarity with the procedure.

Higher log likelihood values suggest hemodynamic responses akin to those observed during the baseline trials (i.e., rest). Lower log likelihood values, in turn, indicate responses different from rest, i.e, music imagery events. Given this interpretation of the log likelihoods, a decrease in the log likelihood function was expected during imagery periods and either an increase or a constant value during the rest periods. The plot depicted in Figure 4 shows one subject's representative log likelihood temporal series clearly illustrating the expected increases and decreases during rest (un-shaded intervals) and imagery (shaded intervals), respectively.

3.3. Music imagery detection

Given the nature of the log likelihood functions, complex classifier training was not needed in order to detect music imagery events. Instead, changes from positive to negative slope in the log likelihood function were used for classification. In order to remove possible artifacts due to e.g., head movement, which may produce a momentary positive-to-negative slope change, music imagery events were only detected if a decrease in the log likelihood function persisted for at least five seconds after a slope change was detected.

An expected imagery event (true positive) was identified with the occurrence of an activation within the first thirteen seconds of a twenty-second imagery interval. This time period
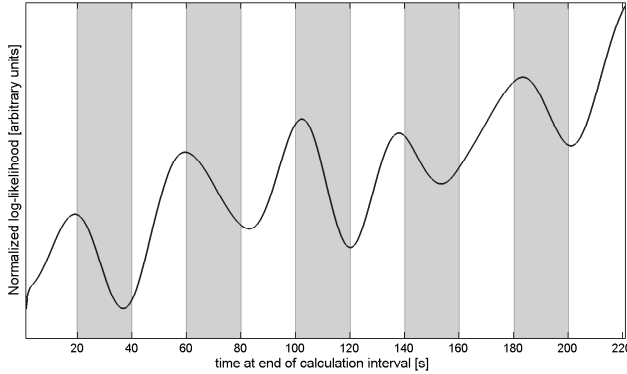
Figure 4. Normalized log likelihood function for a single subject, clearly showing the anticipated pattern for an imagery trial. Computed using $\vec{u}_{DC,3}$ data and tested against a 4-state 1-Gaussian HMM using l=3-second sliding windows. Imagery intervals are shaded. The X axis is time [sec] and the Y axis is normalized log likelihood [arbitrary units].

allowed for both the delay in the change of the hemodynamics-driven signal (up to eight seconds [3]) and the delay on the part of the subject to start music imagery once tapped on the shoulder, which, in turn, was found to be between 1–5 seconds in pilot experiments. If no activation occurred during this first portion of an imagery interval, the interval was deemed to be a false negative, or incorrectly classified as a rest interval. Intervals were also labeled false negatives if an activation occurred after thirteen seconds into the beginning of the imagery interval. A rest interval with no activations was considered a true negative, as was a rest interval with an activation within the first eight seconds after the previous imagery interval — this margin was motivated again by the delay in the NIRS signal, and the observed few seconds of delay in the participant's cessation of the music imagery [3, 10]. Rest intervals with an activation after the first eight seconds were deemed false positives or incorrectly classified as imagery intervals. The number of true/false positives (TP/FP) and true/false negatives (TN/FN) were computed and summed over the two imagery trials, for a total of 12 rest intervals and 10 imagery intervals, and were used to compute the performance metrics sensitivity and specificity, given by:

$$Sensitivity = \frac{TP}{TP + FN}, \tag{10}$$

$$Specificity = \frac{TN}{TN + FP}. \tag{11}$$

Sensitivity is a measure of how many actual activations were correctly identified, while the specificity indicates how many rest intervals were correctly identified as baseline.

4. Experimental results

In this section, calculated statistics and optimal parameters are reported.

Subject	DC					AC					Hemoglobin Concentration				
	Sens	Spec	Q–M	d	l	Sens	Spec	Q–M	d	l	Sens	Spec	Q–M	d	l
1	0.80	0.92	4–1	4	7	0.70	1.00	4–1	3	7	1.00	1.00	2–1	3	10
2	1.00	0.58	2–1	5	3	1.00	0.67	4–1	5	5	0.70	1.00	2–1	5	7
3	0.90	0.83	4–1	4	1	0.90	0.75	4–1	4	1	0.60	0.92	2–2	4	10
4	0.90	1.00	2–1	3	7	1.00	0.92	2–1	4	10	1.00	0.70	2–2	3	7
5	0.80	0.75	4–1	4	3	0.90	0.75	2–1	5	10	0.80	0.83	2–2	5	7
6	0.80	0.58	2–2	5	10	0.60	0.83	2–1	5	15	1.00	0.83	4–1	4	1
7	1.00	0.67	2–1	5	5	0.80	0.92	4–1	3	5	0.80	0.86	2–2	3	1
8	0.90	0.92	2–2	4	3	0.80	1.00	2–2	4	1	0.70	0.92	2–1	4	3
9	0.60	1.00	2–1	3	10	0.70	0.92	4–1	4	1	0.80	0.92	2–2	4	3
10	0.50	1.00	4–1	5	10	0.80	0.67	4–1	5	1	1.00	0.64	4–1	4	5
11	0.90	0.92	4–1	4	1	0.90	1.00	4–1	4	5	0.90	0.75	2–1	5	10
12	0.80	0.83	2–1	5	10	0.70	0.92	4–1	5	15	1.00	1.00	4–1	3	3
13	0.90	0.75	4–1	5	5	0.80	0.83	2–1	4	1	0.90	1.00	4–1	3	1
Mean	**0.83**	**0.83**	–	–	–	**0.82**	**0.86**	–	–	–	**0.86**	**0.87**	–	–	–
(SD)	**(0.14)**	**(0.15)**	–	–	–	**(0.12)**	**(0.12)**	–	–	–	**(0.14)**	**(0.12)**	–	–	–

Table 2. Optimal performance results for all three data types, by subject. Columns labeled 'Sens' and 'Spec' indicate sensitivity and specificity, respectively. The HMM parameters, namely, number of states and number of Gaussian mixtures, are shown by Q and M, respectively. Parameter d indicates the 'detail' of the filter and l the window size [seconds]. SD is standard deviation.

4.1. Quantitative results

Table 2 reports sensitivity (column labeled "sens" in the table) and specificity ("spec") for the HMM–filter–window combinations that resulted in the best per-user performance (defined as the highest average of sensitivity and specificity) out of the 27 possible combinations described in 3.2.1. The average over all participants is recorded in the last row, along with the standard deviation (SD). As can be seen, average sensitivities and specificities were all greater than 82%, and in many individual instances all imagery trials were correctly detected (100% sensitivity). These results are promising for the prefrontal cortical task of music imagery as they are comparable to previously reported results from a group of 3 subjects using motor imagery [8], but were obtained without having to compensate for the interference of hair. Additionally, as observed in the table, all three data types exhibited comparable average performances over all subjects. Although a slight improvement was observed with the hemoglobin concentration data, performance gains were not significant (p-values ≥ 0.38, t-test). This suggests that the additional computational effort required to calculate the concentrations may not be justified. Nonetheless, if computational power can be afforded, some subjects were shown to benefit greatly from using the concentration data (e.g., subjects 1 and 6).

4.2. Client-centred design

Table 2 further indicates the per-user and per-data type optimal HMM–filter–window combinations. The variation between parameters for all participants supports a client-centred design. For raw data (AC and DC), HMMs with observation probabilities represented by only one Gaussian ($M = 1$) were selected most often, whereas for all three data types, window sizes of $l \leq 10$ seconds were selected.

Subject	Window size l [seconds]											
	1		3		5		7		10		15	
	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
1	0.50	1.00	0.67	1.00	0.67	1.00	0.83	0.67	1.00	1.00	0.50	1.00
2	0.90	0.67	0.80	0.75	0.90	0.67	0.70	1.00	0.80	0.75	0.60	1.00
3	0.78	0.55	0.78	0.55	0.33	0.88	0.40	0.58	0.60	0.92	0.50	1.00
4	1.00	0.67	1.00	0.67	1.00	0.58	1.00	0.70	0.67	0.90	0.67	0.86
5	0.67	0.63	0.83	0.57	0.75	0.50	0.80	0.83	0.40	1.00	0.60	0.83
6	1.00	0.83	1.00	0.83	1.00	0.83	1.00	0.67	0.67	1.00	0.60	0.83
7	0.80	0.86	0.80	0.86	0.80	0.86	0.80	0.86	0.50	1.00	0.67	0.83
8	0.60	0.83	0.70	0.92	0.60	0.92	0.70	0.75	0.70	0.58	0.80	0.75
9	0.70	0.83	0.80	0.92	0.70	0.83	0.60	0.92	0.70	0.83	0.50	0.75
10	0.63	0.91	0.63	0.91	1.00	0.64	0.70	0.75	0.44	0.82	0.50	0.83
11	0.80	0.75	0.63	0.90	0.63	0.90	0.71	0.89	0.90	0.75	0.70	0.83
12	1.00	0.83	1.00	1.00	1.00	0.83	0.60	1.00	0.83	0.83	0.60	1.00
13	0.90	1.00	0.80	1.00	0.80	0.92	0.70	0.92	1.00	0.75	0.80	0.92
Mean	**0.79**	**0.80**	**0.80**	**0.84**	**0.78**	**0.80**	**0.73**	**0.81**	**0.71**	**0.86**	**0.62**	**0.88**
(SD)	**(0.16)**	**(0.14)**	**(0.13)**	**(0.16)**	**(0.20)**	**(0.15)**	**(0.16)**	**(0.13)**	**(0.19)**	**(0.13)**	**(0.11)**	**(0.10)**

Table 3. Time-dependent performance by subject for hemoglobin concentration data. Columns labeled 'Sens' and 'Spec' are sensitivity and specificity, respectively. SD is standard deviation.

Table 3 reports the effects of window size on system performance. For brevity, only the concentration data is included. As can be seen in the table, the best average performance occurred at a window size of 3 seconds, and sensitivity decreased as the window size increased. However, it is noted that the optimal window size varied between subjects. For example, subject 1 experienced the best performance at a window size of 10 seconds, while subject 12 displayed the best results for a 3-second window. This variability is likely due to one or more of the following factors: differences in reaction time, intensity of emotions evoked by the music imagery task, ability to concentrate on the imagery, probe location, and/or anatomical differences, all of which could have affected the propagation delay of the hemodynamic signal. This per-user effect of window size is taken into consideration via a client-centred approach.

5. Discussion

This section provides a comparison with previous work, discusses the use of music imagery for BCI control and the advantages of a client-centred approach, and considers the study's limitations.

5.1. Comparison to past work

The proposed mental activity classifier was exclusively based on statistical models of *baseline* temporal hemodynamic responses. This is a distinct advantage over other HMM-based classification systems (e.g., [3, 17, 33]) which require models to be trained on both mental

states in order to perform classification. Such an approach requires the collected data to be separated for training and testing, hence limiting the amount of information available. The proposed method, in comparison, allows the use of all collected baseline data to train HMMs and all collected imagery data to serve as test data, thus reducing possible model overfitting artifacts.

The majority of past work has focused on converting light intensity signals into hemoglobin concentrations, with a few studies focusing on the AC or DC components directly (e.g., [6, 41]).   Results observed here suggest that neither method is more discriminating than the other for the task of music imagery classification.   As a consequence, the computationally-intense intensity-to-concentration conversion functions may be avoided, thus making it more conducive for future adaptation into a binary BCI.

5.2. Music imagery classification for BCI control

Music imagery classification may be used to control a binary BCI operating in a system-paced mode, akin to single-switch scanning found in conventional electronic augmentative and alternative communication systems. In system-paced mode, options would be highlighted sequentially at a fixed rate. The user would make a selection by performing music imagery, which in turn would be used to activate the BCI. As mentioned in 3.3, music imagery classification was performed once a sustained decrease in the log-likelihood function was observed for five consecutive seconds. This decision rule limits the information transfer rate (ITR) of the BCI to a maximum of 12 bits/min and the system scan rate (SR) to a minimum of 5 seconds. Such values are somewhat lower than those achieved with existing EEG-based BCIs which can be in the order of $ITR = 25$ bits/min and $SR = 2.4$ s [4]. However, it is emphasized that participants had no prior training in using the technology; user training may lead to improvements in system response times [2], thus allowing for a faster decision rule to be implemented.

5.3. Client-centred paradigm

In this study, a client-driven approach was used and HMM model parameters were optimized on a per-participant basis. The optimal number of HMM states Q and Gaussian components M varied between subjects and also between data types (i.e., AC, DC, and hemoglobin concentrations), suggesting that throughout the course of the trials participants did not transition directly from rest to an active mental state. The intermediate activity between rest and music imagery likely varied significantly between participants due to irregularities in level, temporal and spatial distribution of the activity within the task interval. Such irregularities could be present due to specific participant-specific factors such as mind wandering, mental alertness, innate reflex reaction time, familiarity with the procedure, and NIRS signal levels and spatial distributions caused by varying emotional intensities of user-selected songs [25]. It is also observed that for the majority of the participants, the temporal light intensity signals after the intensity-to-concentration mapping are smoother than their original counterparts. This smoothing behaviour is likely the reason that a lower number of states were required, in general, for hemoglobin concentration data relative to AC or DC data.

In order to customize the proposed NIRS-driven classifier for an individual user for BCI control, a short calibration session can be undertaken with known imagery and rest intervals

— much like the imagery trials of this experiment — in order to choose the optimal parameters for the data (data type, filter), the HMM (states, Gaussians) and the response time of the switch (window size). The optimization process can be hardcoded into a calibration program, and executed with minimal intervention by an outside party, thus likely will not pose a burden on the user. This would be akin to the training phase required by widely-used speech recognition engines.

5.4. Study limitations

Experimental conditions were restrictive in the sense that participants were required to remain quiet and immobile for the duration of the data recording sessions; all focus was placed on the given task. All auditory and visual distractions were suppressed as much as possible by the earmuffs and blindfold. If BCI technologies are to be used in everyday settings, however, movement artifacts [10], auditory and visual distractions [42], and mind-wandering still have to be detected and accounted for (e.g., as in [19]). Additionally, experimental sessions lasted less than one hour. For BCIs, however, systems may be used for multiple consecutive hours. As a consequence, the effects of fatigue on classifier performance also need to be quantified. However, multiple calibration sessions may be performed throughout the day such that the detrimental effects of fatigue on the harnessed signals can be incorporated directly into the reference baseline hidden Markov models. Lastly, further work must be done to investigate the feasibility of music imagery classification for individuals in the target population. The prospects are promising, however, as hemodynamic responses due to mental activity have been shown to be detectable by NIRS for individuals with ALS [6].

6. Conclusions

This study proposed a client-centred NIRS-driven system that uses HMM as a classification tool to differentiate between music imagery and baseline (rest). The classifier performance, reported both in terms of classifier sensitivity and specificity, averaged over thirteen subjects was greater than 82%, and in many cases exhibited successful detection of all imagery events (100% sensitivity). To allow for a client-centred design, three different wavelet filters (3-, 4-, and 5–detail), three HMM parameter sets (Q–M values of 4–1, 2–2, and 2–1), and six different sliding window sizes (1, 3, 5, 7, 10 and 15 seconds) were considered. The resultant wide range of optimal parameters suggests that parameters be selected for each individual via a short, simple calibration session wherein the user alternates between rest and music imagery in a predetermined pattern. Moreover, since concentration calculations require additional computational power, the comparable results obtained here between all data types suggest that NIRS raw data (AC, DC) should be considered in future studies as an alternative to hemoglobin concentrations. The promising performance of the system suggests that a user-centered NIRS-based BCI is a concept that deserves further feasibility studies, as does the use of HMM as a classifier of cognitive activity versus rest.

Acknowledgements

Author details

Tiago H. Falk[1,*], Kelly M. Paton[2] and Tom Chau[3]

* Address all correspondence to: tiago.falk@ieee.org

1 Institut National de la Recherche Scientifique (INRS-EMT), University of Quebec, Montreal, Canada
2 Institute of Applied Mathematics, University of British Columbia, British Columbia, Canada
3 Bloorview Research Institute, University of Toronto, Toronto, Canada

References

[1] K Tai, S Blain, and Tom Chau. A review of emerging access technologies for individuals with severe motor impairments. *Assistive Technology*, 20:204–219, 2008.

[2] Jonathan R Wolpaw, Niels Birbaumer, William J Heetderks, Dennis J McFarland, P Hunter Peckham, Gerwin Schalk, Emanuel Donchin, Louis A Quatrano, Charles J Robinson, and Theresa M Vaughan. Brain-Computer Interface Technology: A Review of the First International Meeting. *IEEE Transactions on Rehabilitation Engineering*, 8(2):164–173, 2000.

[3] Ranganatha Sitaram, Haihong Zhang, Cuntai Guan, Manoj Thulasidas, Yoko Hoshi, Akihiro Ishikawa, Koji Shimizu, and Niels Birbaumerb. Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *NeuroImage*, 34:1416–1427, 2007.

[4] Marcel van Gerven, Jason Farquhar, Rebecca Schaefer, Rutger Vlek, Jeroen Geuze, Anton Nijholt, Nick Ramsey, Pim Haselager, Louis Vuurpijl, Stan Gielen, and Peter Desain. The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4), 2009.

[5] R Sitaram, Y Hoshi, and C Guan. Near infrared spectroscopy based brain-computer interface. *Proceedings of SPIE - The International Society for Optical Engineering*, 5852(1):434–442, 2005.

[6] M Naito, Y Michioka, K Ozawa, Y Ito, M Kiguchi, and T Kanazawa. A Communication Means for Totally Locked-in ALS Patients Based on Changes in Cerebral Blood Volume Measured with Near-Infrared Light. *IEICE Transactions on Information and Systems*, E90-D(7):1028–1037, 2007.

[7] S Coyle, T Ward, C Markham, and G McDarby. On the suitability of near-infrared (NIR) systems for next-generation brain-computer interfaces. *Physiological Measurements*, 25:815–822, 2004.

[8] S M Coyle, T E Ward, and C M Markham. Brain-computer interface using a simplified functional near-infrared spectroscopy system. *Journal of Neural Engineering*, 4:219–226, 2007.

[9] R Sitaram, A Caria, and N Birbaumer. Hemodynamic brain-computer interfaces for communication and rehabilitation. *Neural Networks*, 22(9):1320–1328, November 2009.

[10] Fiachra Matthews, Barak A Pearlmutter, Tomas E Ward, Christopher Soraghan, and Charles Markham. Hemodynamics for Brain-Computer Interfaces. *IEEE Signal Processing Magazine*, 25(1):87–94, January 2008.

[11] Arno Villringer and Britton Chance. Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, 20(10):431–433, October 1997.

[12] Gert Pfurtscheller, G′unther Bauernfeind, Selina Christin Wriessnegger, and Christa Neuper. Focal frontal (de)oxyhemoglobin responses during simple arithmetic. *International Journal of Psychophysiology*, 76(3):186–192, 2010.

[13] H. Ogata, T. Mukai, and T. Yagi. A study on the frontal cortex in cognitive tasks using near-infrared spectroscopy. In *Proc. IEEE Conf. of the Engineering in Medicine and Biology Society*, pages 4731–4734, 2007.

[14] H Ayaz, M Izzetoglu, S Bunce, T Heiman-Patterson, and B Onaral. Detecting cognitive activity related hemodynamic signal for brain-computer interface using functional near-infrared spectroscopy. In *Proceedings of the IEEE Conference of the Engineering in Medicine and Biology Society*, pages 342–345, 2007.

[15] MJ Herrmann, A.C. Ehlis, and AJ Fallgatter. Prefrontal activation through task requirements of emotional induction measured with NIRS. *Biological Psychology*, 64(3):255–263, 2003.

[16] J. Leon-Carrion, J. Damas, K. Izzetoglu, K. Pourrezai, J.F. Martín-Rodríguez, J.M.B. Martin, and M.R. Dominguez-Morales. Differential time course and intensity of PFC activation for men and women in response to emotional stimuli: A functional near-infrared spectroscopy (fNIRS) study. *Neuroscience Letters*, 403(1-2):90–95, 2006.

[17] S Power, T Falk, and T Chau. Classification of prefrontal activity due to mental arithmetic and music imagery using hidden markov models and frequency domain near-infrared spectroscopy. *Journal of Neural Engineering*, 7:026002, 2010.

[18] Sarah D Power, Azadeh Kushki, and Tom Chau. Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: toward a three-state nirs-bci. *BMC Research Notes*, 5(141), March 2012.

[19] Tiago Falk, M Guirgis, S Power, and T Chau. Taking nirs-bcis outside the lab: Achieving robustness against ambient noise. *IEEE Trans Neural Syst Rehab Eng*, 19(2):136–146, 2011.

[20] Boris Kotchoubey, Anna Dubischar, Herbert Mack, Jochen Kaiser, and Niels Birbaumer. Electrocortical and behavioral effects of chronic immobility on word processing. *Cognitive Brain Research*, 17(1):188–199, June 2003.

[21] Toshiki Endo, Teiji Tominaga, and Lars Olson. Cortical Changes Following Spinal Cord Injury with Emphasis on the Nogo Signaling System. *The Neuroscientist*, 15(3):291–299, 2009.

[22] S B Frost, S Barbay, K M Friel, E J Plautz, and R J Nudo. Reorganization of Remote Cortical Regions After Ischemic Brain Injury: A Potential Substrate for Stroke Recovery. *Journal of Neurophysiology*, 89:3205–3214, 2003.

[23] E Altenmüller, K ASch*ü*rmann, V K Lim, and D Parlitz. Hits to the left, flops to the right: different emotions during listening to music are reflected in cortical lateralisation patterns. *Neuropsychologia*, 40(13):2242–2256, 2002.

[24] C L Krumhansl. An Exploratory Study of Musical Emotions and Psychophysiology. *Canadian Journal of Experimental Psychology*, 51(4):336–353, 1997.

[25] M Boso, P Politi, F Barale, and E Enzo. Neurophysiology and neurobiology of the musical experience. *Functional Neurology*, 21(4):187–191, 2006.

[26] A J Blood, R J Zatorre, P Bermudez, and A C Evans. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature Neuroscience*, 2(4):382–387, 1999.

[27] A J Blood and R J Zatorre. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences U.S.A.*, 98:11818–11823, 2001.

[28] B Kleber, N Birbaumer, R Veit, T Trevorrow, and M Lotze. Overt and imagined signing of an Italian aria. *Neuroimage*, 36:889–900, 2007.

[29] Sarah D Power, A Kushki, and Tom Chau. Towards a system-paced near-infrared spectroscopy brain-computer interface: differentiating prefrontal activity due to mental arithmetic and mental singing from the no-control state. *Journal of Neural Engineering*, 8(6):066004, December 2011.

[30] C D Tsang, L Trainor, D L Santesso, S L Tasker, and L A Schmidt. Frontal EEG responses as a function of affective musical features. *Annals of the New York Academy of Sciences*, 930:439–442, 2001.

[31] S Dalla Bella, I Peretz, L Rousseau, and N Gosselin. A developmental study of the affective value of tempo and mode in music. *Cognitio*, 80:B1–10, 2001.

[32] S K Law. Thickness and Resistivity Variations over the Upper Surface of the Human Skull. *Brain Topography*, 6(2):99–109, 1993.

[33] B Obermaier, C Guger, C Neuper, and G Pfurtscheller. Hidden Markov models for online classification of single trial EEG data. *Pattern Recognition Letters*, 22:1299–1309, 2001.

[34] I Sase, H Eda, A Seiyama, H C Tanabea, A Takatsuke, and T Yanagida. Multi-channel optical mapping: Investigation of depth information. *Proceedings of SPIE*, 4250, 2001.

[35] K. Tai and T. Chau. Single-trial classification of nirs signals during emotional induction tasks: towards a corporeal machine interface. *Journal of NeuroEngineering and Rehabilitation*, 6(39), 2009.

[36] Arlene Duncan, Judith H Meek, Matthew Clemence, Clare E Elwell, Lidia Tyszczuk, Mark Cope, and David T Delpy. Optical pathlength measurements on adult head, calf and forearm and the head of the newborn infant using phase resolved optical spectroscopy. *Physics in Medicine and Biology*, 40:295–304, 1995.

[37] Mark Cope. *The development of a near infrared spectroscopy system and its application for non invasive monitoring of cerebral blood and tissue oxygenation in the newborn infant*. Ph.d., University of London, 1991.

[38] Lawrence R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

[39] R. Hu, X. Li, and Y. Zhao. Acoustic model training using greedy EM. In *proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'05, March 18-23 2005, Philadelphia*, volume 1, 2005.

[40] Kevin Murphy. Hidden Markov Model (HMM) Toolbox for Matlab. http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html.

[41] S. Luu and T. Chau. Decoding subjective preference from single-trial near-infrared spectroscopy signals. *Journal of Neural Engineering*, 6(4), 2009.

[42] V. Gumenyuk, O. Korzyukov, K. Alho, C. Escera, and R. Naatanen. Effects of auditory distraction on electrophysiological brain activity and performance in children aged 8-13 years. *Psychophysiology*, 41(1):30–36, 2004.