NeuroImage 72 (2013) 106–119

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Technical Note

A reference-channel based methodology to improve estimation of event-related hemodynamic response from fNIRS measurements

Fabio Scarpa [a,*], Sabrina Brigadoi [a], Simone Cutini [b], Pietro Scatturin [a], Marco Zorzi [b,c,d], Roberto Dell'Acqua [a,c], Giovanni Sparacino [e]

[a] Department of Developmental Psychology and Socialization, University of Padova, Padova, Italy
[b] Department of General Psychology, University of Padova, Padova, Italy
[c] Center for Cognitive and Brain Science, University of Padova, Padova, Italy
[d] IRCCS San Camillo Hospital, Lido-Venice, Italy
[e] Department of Information Engineering, University of Padova, Padova, Italy

ARTICLE INFO

Article history:
Accepted 13 January 2013
Available online 26 January 2013

Keywords:
fNIRS
Near-infrared spectroscopy
Reference-channel
Bayesian filtering
Physiological noise

ABSTRACT

Functional near-infrared spectroscopy (fNIRS) uses near-infrared light to measure cortical concentration changes in oxygenated (HbO) and deoxygenated hemoglobin (HbR) held to be correlated with cognitive activity. Providing a parametric depiction of such changes in the classic form of stimulus-evoked hemodynamic responses (HRs) can be attained with this technique only by solving two problems. One problem concerns the separation of informative optical signal from structurally analogous noise generated by a variety of spurious sources, such as heart beat, respiration, and vasomotor waves. Another problem pertains to the inherent variability of HRs, which is notoriously contingent on the type of experiment, brain region monitored, and human phenotype. A novel method was devised in the present context to solve both problems based on a two-step algorithm combining the treatment of noise-only data extrapolated from a reference-channel and a Bayesian filter applied on a per-trial basis. The present method was compared to two current methods based on conventional averaging, namely, a typical averaging method and an averaging method implementing the use of a reference-channel. The result of the comparison, carried out both on artificial and real data, revealed a sensitive accuracy improvement in HR estimation using the present method relative to each of the other methods.

Introduction

Functional near-infrared spectroscopy (fNIRS) is a low-cost neuroimaging technique that allows neuroscientists to study stimulus-evoked functional and neural activity by monitoring non-invasively hemodynamic changes occurring in specific brain regions (Boas et al., 2002; Bunce et al., 2006; Gervain et al. 2011; Jöbsis, 1977). Sources of near-infrared light and light-detectors are positioned in a principled way on the scalp of human participants at a distance of approximately 3 cm. The light reaches the cerebral cortex, and is subject to known optical phenomena, like refraction, scattering, and absorption by two specific subcomponent of the blood tissue, oxy- and deoxy-hemoglobin (HbO and HbR, respectively). Part of the light is redirected to the scalp, where HbO/HbR concentration changes are in turn detected and translated quantitatively based on the Modified Beer Lambert Law (MBLL) (Sassaroli and Fantini, 2004). Although reliable measurements of HbO/HbR concentration are confined to the outer 2.5 cm of the cerebral tissue underlying the scalp, fNIRS provides several advantages with respect to other current neuroimaging technologies, the most important being the higher sampling rate relative to functional magnetic resonance imaging, as well as safety, portability and increased resistance to movement artifacts. The fNIRS technique has in fact been employed in a variety of empirical investigations in cognitive neuroscience (for a review, see e.g. Cutini et al., 2012; Ferrari and Quaresima, 2012). Like all other techniques and instruments, a correct parameter estimation of the so-called hemodynamic response (HR; spectral band centered at ≈.1 Hz) using fNIRS is pivotal to inform theories of human brain functioning, and not devoid of the problems pervading every direct imaging effort to isolate informative bio-signals from noise due to structurally analogous changes modulated by physiological components other than those of interest. Considering hemodynamics, sources of physiological noise range from heart beat (spectral band centered at ≈1 Hz), respiration (≈.2 Hz), vasomotor (or Mayer's) wave (≈.1 Hz), and generators of low-frequency oscillations (<.1 Hz) (Zhang et al., 2007), which jointly exert a particularly strong camouflaging effect on brain-driven fNIRS signal. The HR elicited by a single briefly presented stimulus has an amplitude typically of <500 nM, which represents a relatively small fraction of total raw signal amplitudes recorded by detectors (>2000 nM).

Over the past decade, neuroscientists interested in localizing brain activity underpinning the execution of specific cognitive tasks have

* Corresponding author at: Department of Developmental Psychology and Socialization (DPSS), Via Venezia 8, 35131, Padova, Italy. Fax: +39 049 8276547.
E-mail address: fabio.scarpa@unipd.it (F. Scarpa).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.01.021

progressively abandoned blocked designs, where hemodynamic changes are tracked as temporally extended continua, in favor of event-related designs (Rosen et al., 1998; Dale, 1999; see also Cutini et al., 2011a), where the signal of interest is partitioned in time-windows locked to the onset of each stimulus. Several methods have been proposed to derive parametrically HRs from fNIRS signals recorded using event-related designs (Huppert et al., 2009), a large number of which resort to conventional averaging (CA) (Cutini et al., 2008, in press; Näsi et al., 2010; Taga et al., 2011). Succinctly, in CA, the HR is determined by averaging fNIRS signals time-locked to stimuli falling in one particular condition of the experimental design. Implicit in the use of CA are two questionable assumptions, that is, the independence of the activity elicited by each stimulus from physiological noise and the difference in phase between physiological noise and stimuli presentation rate. Other methods employed for HR estimation are based on band-pass filtering (Jasdzewski et al., 2003), principal component analysis (PCA) (Franceschini et al., 2006), and general linear model (GLM) (Ye et al., 2009). Although all these methods have proved beneficial in terms of signal gain, they have also been associated with some disadvantages, sometimes involving technical aspects of data acquisition and experiment duration (e.g., the acquisition of resting state data before each stimulus, or the registration of hemodynamic activity of the whole head), but also aspects with potential repercussions on data interpretation. Band-pass filtering, for instance, reduces noises and HR to a similar extent, given the above mentioned overlap in frequency spectra. PCA is also effective in reducing noise. However, it does so by decreasing the amplitude of HR in the activated regions and by propagating noise from perturbed channel(s) to remnant channels. GLM methods rest heavily on a-prioristic choices of basis functions, leaving aside that identical basis functions are often used unwarrantedly for both HbO and HbR without an appropriate treatment of variability ensuing from factors naturally covarying with the factors manipulated in an experimental design, such as inter-individual differences, sensory nature of stimuli used, monitored brain areas, and inter-stimulus interval (ISI).

More promising for improving estimation of HR from fNIRS measurements appear in present methods based on the use of the so-called "reference-channel" (Gagnon et al., 2011; Saager et al., 2011; Zhang et al., 2009). A reference-channel is denoted by a source-detector distance shorter than 1 cm. Owing to the direct relation between source-detector distance and depth reached by photons within tissues underlying the scalp, fNIRS signal detected by the reference-channel is unlikely to reflect hemodynamic activity other than that taking place at the subdural tissues. In this vein, signal from the reference-channel is ideal for deriving information about noise generated by physiological sources but not influenced by stimulus-evoked brain activity. Removing the noise recorded from the reference-channel from standard-channels held to record a combination of noise and brain activity is the first step of the method presented herein. More specifically, the reference-channel signal was used to estimate a parametric model of physiological noise, which was in turn used to correct the signal recorded from standard-channels. Our algorithm included a second step, in which residual random noise was further reduced by adopting the Bayesian filtering method described by Scarpa et al. (2010). HR estimates were finally obtained by averaging stimulus-evoked HRs to stimuli falling into one of two levels of a finger-tapping design composed of 40 trials. The presented method was evaluated using synthetic and real data, and compared with two CA methods for HR estimation, one of which implements a solution based on the use of the reference-channel. Results on 30 simulated datasets revealed the ability of the present method to improve the accuracy of HR estimates, despite their relatively small amplitudes (<500 nM for HbO and >-200 nM for HbR), a finding that was corroborated after the results were submitted for testing via a receiver operating characteristic (ROC) analysis. A sizable improvement of HR estimation accuracy was also observed following its application on data acquired from 10 human adults who carried out the finger-tapping task.

Material and methods

Database

In the forthcoming sections, a description of the real database generated by administering the finger-tapping task is described first, followed by a description of the method application on synthetic data adherent to key features of the real data.

Real data

Ten right-handed participants (5 females, mean age 28, from 24 to 37) performed the experiment after providing informed consent. Each participant was seated on a comfortable chair in a dimly lit room in front of a LCD monitor placed at a viewing distance of 60 cm. The index fingers of the right and left hands were placed on the "A" and "L" key respectively. Each trial began with a central fixation cross, followed 2 s later by a white arrow head pointing unpredictably and with equal probability to the right (Condition 1) or left (Condition 2) side of the monitor. Participants had to press the "A" key twice if the white arrow pointed to the left, or the "L" key twice if the white arrow pointed to the right. Each participant performed a total of 80 trials, which were organized in two consecutive sessions of 40 trials. An ISI ranging from 12 to 15 s elapsed between consecutive trials. The choice of this double-press finger-tapping task was largely motivated by the expected small HR amplitude (Brigadoi et al., 2012; Cutini et al., 2008, 2011a) characterized however by well-established features replicated in several prior fNIRS studies (Franceschini et al., 2006; Holper et al., 2009; Leff et al., 2011; Lutz et al., 2005; Sato et al., 2007). No resting state intervals were added between stimuli in order to reduce experimental time, thereby limiting phenomena like fatigue, habituation, and reduction in sustained attention influencing the participants' performance.

The fNIRS signal was acquired with a multi-channel frequency-domain NIR spectrometer (ISS Imagent™, Champaign, Illinois), equipped with 40 laser diodes (20 emitting light at 690 nm, and 20 at 830 nm) and 4 photo-multiplier tubes. Sources and detectors were held in place on the scalp using a custom-made holder and velcro straps. Each source location comprised two source optical fibers, one for each wavelength. Sources and detectors were positioned using a probe-placement method based on a physical model of the ICBM152 template's head surface (Cutini et al., 2011b) and their position is illustrated in Fig. 1: detectors A and B were placed 1 cm behind C3 and C4 (according to the international 10–20 system), respectively. There were 10 standard-channels (characterized by a source-detector distance of 3 cm): 1 (detector A — source 1), 2 (A-2), 3 (A-3), 4 (A-4), 5 (A-5) for the left and 6 (B-1), 7 (B-2), 8 (B-3), 9 (B-4), 10 (B-5) for the right hemisphere. In addition, there were 4 reference-channels (characterized by a source-detector distance of .7 cm): 11 (C-1), 12 (C-2) for the left and 13 (D-1), 14 (D-2) for the right hemisphere. The probe arrangement even provides 4 additional channels (sources 1 and 2 with detector C on the left hemisphere and sources 1 and 2 with detector D on the right hemisphere) with a source-detector distance equal to 1.5 cm: since they are too close to be considered as standard-channels and, at the same time, too far to be considered as reference-channels, they are discarded from analyses. From each channel the signal relative to HbO and the corresponding signal relative to HbR was derived. To be noticed, only two reference-channels for each hemisphere were placed. The dataset for each channel contains about 12,000 time-points, roughly corresponding to 25 min (1500 s). The sampling frequency was 7.8125 Hz.

A representative example of the signal acquired by a reference-channel and by a standard-channel is reported in Fig. 2a.

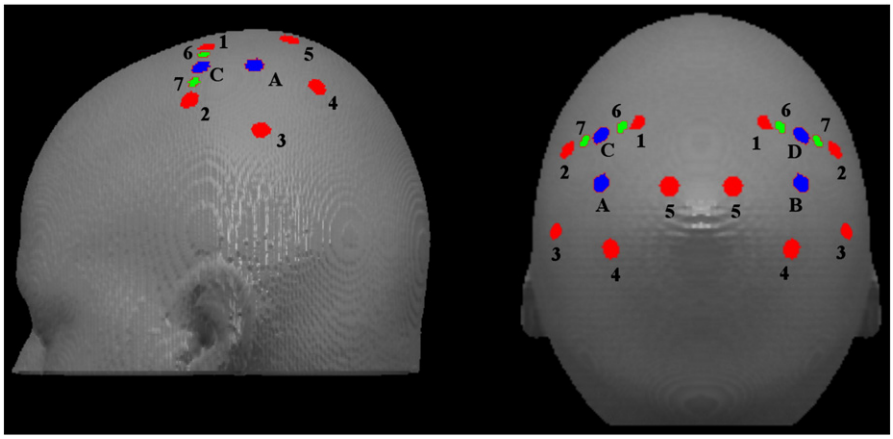

Fig 1. Probe placement: sources (red circles) and detectors (blue circles) overlaid on the head surface of the ICBM152 template. Green circles represent the sources of the reference-channels.

Synthetic data

Simulated datasets were generated so as to resemble as close as possible the characteristics of real data. For each of the 30 simulated subjects, 10 standard-channels for HbO and the corresponding 10 channels for HbR were generated, in addition to 2 reference-channels for each hemisphere, for both HbO and HbR. Thus, there were a total of 14 channels for each type of hemoglobin.

Samples $y_{sim}(t)$ of each simulated channel were generated as:

$$y_{sim}(t) = k \cdot u_{true}(t) + \phi_{sim}(t) + \eta(t) + r(t) \tag{1}$$

where $u_{true}(t)$ was the true HR and was multiplied by the constant k (which can be equal to 0, 0.5 or 1 depending on the channel; i.e., k is 0 in the reference-channels and halved in some standard-channel, described below), $\varphi_{sim}(t)$ was the physiological noise term, $\eta(t)$ was the random noise, and $r(t)$ was a noise term due to possible motion artifacts.

The HR evoked by a single stimulus ($t=0$ corresponds to the presentation of the stimulus), was modeled by a linear combination of two gamma-variant functions Γ_n (Abdelnour and Huppert, 2009; Lindquist and Wager, 2007), time dependent, with a total of 6 variable parameters:

$$u_{true}(t) = \alpha \times [\Gamma_n(t, \tau_1, \rho_1) - \beta \times \Gamma_n(t, \tau_2, \rho_2)] \tag{2}$$

with:

$$\Gamma_n\left(t, \tau_j, \rho_j\right) = \frac{1}{p!\tau_j}\left(\frac{t-\rho_j}{\tau_j}\right)^p e^{-(t-\rho_j)/\tau_j}\delta\left(t-\rho_j\right), \quad \delta\left(t-\rho_j\right) = \begin{cases} 1 \text{ if } t-\rho_j \geq 0 \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

where α tuned the amplitude, τ_j and ρ_j tuned the response width and the onset time respectively, and β determined the ratio of the response to undershoot. The coefficient p was set to 5 following the nominative values reported in Glover (1999). In order to simulate the HR due to two different stimuli and with shapes, amplitudes and latencies in agreement with previous findings regarding finger tapping tasks, two u_{true} profiles were generated by properly tuning the parameters in Eq. (2), allowing small variations in peak amplitude and latency between a trial and another. For HbO, this led to a first HR profile with a peak amplitude of 420 ± 20 nM and a peak latency equal to $5.0 \pm .2$ s, while the second HR profile had a peak amplitude of 360 ± 20 nM and a peak latency equal to $5.5 \pm .2$ s. The first profile was associated with Condition 1, the second with Condition 2. Note that HR amplitude was lower than that of physiological components ($\approx$400 nM vs $\approx$2000 nM). 40 stimuli for each condition were simulated, with an ISI ranging between 12 and 15 s. The simulated HR u_{true} in Eq. (2) was added ($k=1$) in channels 1, 4, 5 for Condition 1, and in channels 6, 9, 10 for Condition 2. In channels 2, 3, 7, 8 the amplitudes of the added HRs were halved ($k=.5$). In the other channels no HR

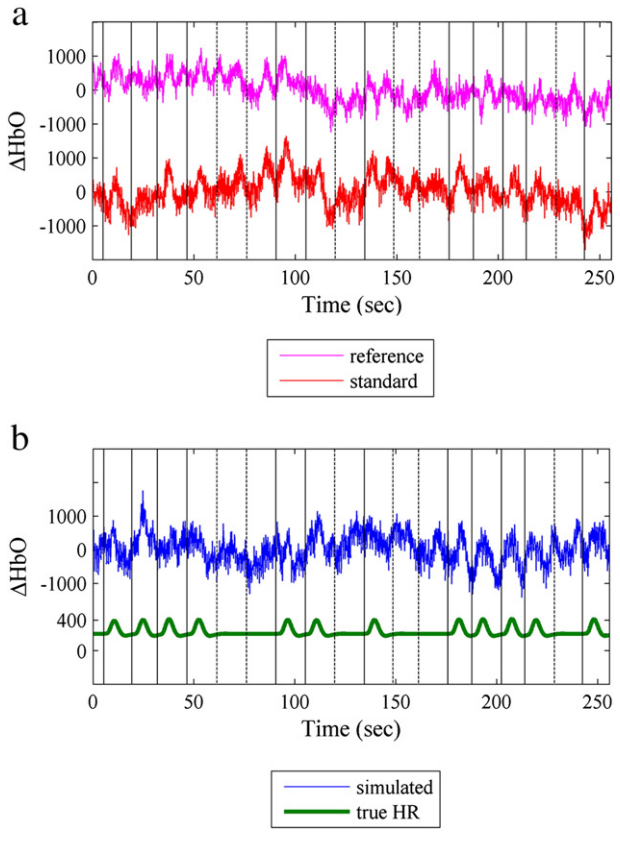

Fig 2. a) Representative examples of signal acquired by reference-channel (magenta) and by standard-channel (red). b) Synthetically generated time series (blue line) and true HR (green line) in the synthetic dataset. The onset time is represented by a solid black vertical line for Condition 1, and by a dashed vertical line for Condition 2: it is the same for all represented signals in a) and b). The real signals reported in figure are relative to channels located in the left hemisphere, so in the reported simulated signal the HR was added only in correspondence of a stimulus relative to Condition 1.

was added ($k=0$). Thus, there were 3 channels with full HR and 2 channels with halved HR in the left hemisphere for Condition 1 and the same in the right hemisphere for Condition 2.

Physiological noise $\varphi_{sim}(t)$ in Eq. (1) can be considered to be additive and modeled as a linear combination of sinusoids (Abdelnour and Huppert, 2009):

$$\phi_{sim}(t) = \sum_{j=1}^{5}\left[A_j \sin\left(2\pi f_j t\right)\right]. \quad (4)$$

Frequency f_j and amplitude A_j of each sinusoid varied in the channel, and were different between participants. They are reported in Table 1.

Random noise $\eta(t)$ in Eq. (1) was modeled as a normal white noise process with standard deviation tuned to be close to that of real data. The measurement noise was different between participants and between channels (400 nM ± 180 nM). In order to simulate artifacts (e.g., due to movements of the participant or shifts of a source or a detector) short non-cyclic abrupt drifts were added in 6 out of 30 participants (it is the typical fraction noticed in our experiment), at a random temporal position and with random amplitude, and is represented by the noise term $r(t)$, different channel by channel.

HbR channels were generated in the same way. According to our real data, for each component in Eq. (1), the sign was changed, a delay of 1 s was added, and all the amplitudes, except that of random noise, were reduced to 25%.

A representative example of the simulated signals is reported in Fig. 2b, which also shows the true HR (u_{true}) which has been added in the synthetic signal. Note that real (Fig. 2a) and simulated (Fig. 2b) signals are dominated by physiological components and that their amplitude is greater than that of the true HR.

Reference-channel modeling corrected Bayesian approach (ReMCoBA)

The proposed method consists of two main steps: (in the section "Step 1: reference-channel modeling and fNIRS data correction") a model of the physiological noise is derived by the reference-channel and used to correct raw fNIRS data of the other channels; (in the section "Step 2: single trial Bayesian filtering and HR estimation") the resulting data are then filtered, on a single-trial basis, with a nonparametric Bayesian approach to reduce residual random noise.

A prototype of the whole algorithm described in this section was implemented in Matlab© (version R2010a, The Mathworks, Natick, Massachusetts, USA) and run on a personal computer.

Raw data time-series of standard and reference-channels were first band-pass filtered (Butterworth, 4th order, band-pass .01–3 Hz) to remove any very slow drift and noise with frequency far from that of the HR. The signal was segmented in 12 second trials and underwent the two step procedure.

Step 1: reference-channel modeling and fNIRS data correction

The fNIRS signal $y(t)$ acquired by standard-channels after a stimulus given at $t=0$ contained as useful component $u(t)$ the event-related HR, a physiological noise component $\varphi(t)$ and random noise $\upsilon(t)$:

$$y(t) = u(t) + \phi(t) + \upsilon(t). \quad (5)$$

An fNIRS reference-channel signal $y_{ref}(t)$ contained, by hypothesis, the same physiological noise $\varphi(t)$ of standard-channels scaled by a coefficient s (constant in time) which took into account the different paths crossed by photons and random noise $\varepsilon(t)$:

$$y_{ref}(t) = \phi(t)/s + \varepsilon(t). \quad (6)$$

Thus, the idea was to exploit the signal $y_{ref}(t)$ to remove, or at least reduce, the physiological noise $\varphi(t)$, which was present also in the standard-channels $y(t)$. This was feasible only if $y_{ref}(t)$ and $y(t)$ had a good correlation, otherwise there was the risk of adding more noise instead of reducing it. For this reason, Step 1 was performed only if the total Pearson's correlation coefficient between the two channels (y and y_{ref}) was above .6; otherwise, Step 1 was skipped and the algorithm went directly to Step 2 (in the section "Step 2: single trial Bayesian filtering and HR estimation"). In order to maximize the usefulness of the reference-channel, for each standard-channel, the reference-channel chosen between the available ones was the one with the highest Pearson's correlation coefficient.

Physiological noise was modeled as a sum of M sinusoidal waves (similarly to Prince et al., 2003), on a trial-by-trial basis:

$$\phi(t) = \sum_{i=1}^{M}[a_i \sin(2\pi\omega_i t) + b_i \cos(2\pi\omega_i t)] + c + w(t) \quad (7)$$

where $w(t)$ was the model error. For each trial, the number M and the value ω_i of the dominant low frequencies (<.18 Hz) were individuated from the power spectrum. In the procedure of Step 1 the maximum value allowed to M was 3, since we had seen that this was the maximum value of dominant low frequencies that could be detected in real data (corresponding to respiratory frequency, Mayer's and very low-frequency oscillations). Unlike Prince et al. (2003) where the number of sinusoidal waves was fixed at 3, M was kept variable among the trials, given that oscillatory components were not always detectable with a sufficient accuracy during the entire recording session. This yielded a reduction of the number of parameters (e.g. in data in the "Real data" section, M will result to be 1 in 19% of trials, 2 in 59%, and 3 in 22%). Even if 5 components have been used in the simulation of physiological noise, we were interested in modeling only the 3 low frequency components, which lie in the same frequency band of HR. The initial value of the frequencies ω_i obtained from the spectrum was then optimized through a grid search method (Handel, 2000). In Eq. (7), the parameters a_i, b_i and c were estimated by the least-squares method. Having an estimate of φ, the scaling coefficient s in Eq. (6) was determined by minimizing the squared difference between the reference-signal $y_{ref}(t)$ and the standard-channel signal $y(t)$ observed in 30 s of resting state data acquired before each experimental session (so as to compensate for possible differences in the magnitude of the path-length factor between reference and standard-channels). Having provided the estimates of all unknown parameters (denoted by the hat symbol) in Eqs. (6) and (7), the corresponding trial of the standard-channels was so corrected:

$$y_c(t) = y(t) - \hat{s}\cdot\sum_{i=1}^{\hat{M}}\left[\hat{a}_i \sin(2\pi\hat{\omega}_i t) + \hat{b}_i \cos(2\pi\hat{\omega}_i t)\right] + \hat{c}. \quad (8)$$

The resulting corrected signal $y_c(t)$ was then submitted to Step 2.

The procedure of the present subsection produced a result which may resemble that of a notch filter. However, this is not completely equivalent to what a notch filter would do on a data-stream. In fact,

Table 1
Physiological components.

	Frequency (Hz)	Amplitude (nM)
Very low freq.	$f_1 = .002 \pm .0001$	$A_1 = 700 \pm 100$
Low freq.	$f_2 = .01 \pm .001$	$A_2 = 700 \pm 100$
Vasomotor	$f_3 = .07 \pm .04$	$A_3 = 400 \pm 10$
Respiratory	$f_4 = .2 \pm .03$	$A_4 = 200 \pm 10$
Cardiac	$f_5 = 1.1 \pm .1$	$A_5 = 400 \pm 10$

Mean and standard deviation of frequency and amplitude of each physiological component.

a notch filter would unavoidably introduce a distortion on the HR because of the superimposition of the spectral content of the HR with that of the physiological noise components.

Step 2: single trial Bayesian filtering and HR estimation

The signal was filtered with a Bayesian approach developed and tested on fNIRS data in Scarpa et al. (2010), to which we refer the reader for details. Briefly, data $y_c(t)$ of each standard-channel were described as:

$$y_c(t) = u(t) + v(t) \tag{9}$$

where $u(t)$ was the useful component and $v(t)$ was noise. Trial by trial, the signal $y_c(t)$ was filtered within a Bayesian embedding by exploiting a priori expectation on both HR and noise. Re-expressing Eq. (9) in vector form, we got:

$$y_c = u + v \tag{10}$$

where **u** and **v** were vectors containing the n samples of the trial. A stationary autoregressive (AR) model was employed as a priori description of **v** (different for each of the N available trials of each condition). In order to keep the number of parameters as low as possible, model order was determined by using the Akaike information criterion (AIC). On 80% of the trials, the most parsimonious model order was 4. This was the value that provided the best estimation error using the data of our previous work (Scarpa et al., 2010). Hence, model order was set to 4 and the a priori covariance matrix of **v** was:

$$\Sigma_v = \sigma^2 \left(A^T A\right)^{-1} \tag{11}$$

where A was a square (n-size) Toeplitz matrix the first column of which was $[1, a_1, a_2, a_3, a_4, 0, \ldots 0]^T$, $\{a_k\}_{k=1,\ldots,4}$ being the coefficients of the AR model, and σ^2 was the variance of the noise process which drives the AR model. For each of the available trials, $\{a_k\}_{k=1,\ldots,4}$ and σ^2 were identified from data measured in an interval lasting 4 s and starting from 1.5 s before the stimulus (when HR was not present). Therefore, A and σ^2 in Eq. (11) were determined on a single-trial basis. As far as the a priori information on **u**, the strategy was to model its expected smoothness as the realization of a stochastic process obtained by the cascade of 2 integrators driven by a zero-mean white noise process $\{\varepsilon_k\}$ with variance λ^2. Therefore, the covariance matrix of **u** was:

$$\Sigma_u = \lambda^2 \left(F^T F\right)^{-1} \tag{12}$$

where $F = \Delta^2$, with Δ^2 being the square n-dimensional lower-triangular Toeplitz matrix the first column of which was $[1, -2, 1, \ldots, 0]^T$. While matrix F in Eq. (12) was fixed throughout the N trials, the scalar λ^2 varied on a single-trial basis and was estimated, independently trial-by-trial, by the so-called "discrepancy" criterion (Twomey, 1965). The optimally filtered trial in the sense of minimum variance estimation was thus:

$$\hat{u} = \left(A^T A + \gamma F^T F\right)^{-1} A^T A y_c \tag{13}$$

where $\gamma = \sigma^2/\lambda^2$.

In order to remove trials irremediably affected by motion artifacts, for each vector $\hat{\mathbf{u}}_\mathbf{j}$ the difference d_j between its maximum and minimum was considered:

$$d_j = \max\left(\hat{u}_j\right) - \min\left(\hat{u}_j\right) \tag{14}$$

Vector characterized by a difference d_j greater than the average of the d_js ± 2.5 SDs had been discarded (≈10%). The estimated HR $\bar{\mathbf{u}}$ was obtained from the average of the N_1 trials filtered as in Eq. (13), which remain after the artifact removal procedure above, belonging to the same condition:

$$\bar{u} = \left(\sum_{j=1,\ldots,N_1} \hat{u}_j\right)/N_1 \tag{15}$$

To further reduce high frequency residual oscillations and facilitate later peak amplitude and latency determination, $\bar{\mathbf{u}}$ was then smoothed with a Savitzky-Golay filter with 3rd polynomial order and frame size equal to 25 time-points (3 s). Finally, $\bar{\mathbf{u}}$ was baseline-corrected by subtracting its mean intensity in the 0–500 ms interval from the onset.

Methods used for comparison

The proposed methodology (ReMCoBA) will be applied in the "Results" section to simulated and real data, and assessed against two widely used literature methods briefly described below.

Reference-channel based fNIRS data correction followed by conventional averaging (rCA)

It is convenient to start from the presentation of a two step method which first exploits the reference-channel signal to correct fNIRS data and then resorts to conventional averaging (rCA) (Saager et al., 2011). This method has been chosen for comparison because it is the most general and simple method for HR estimation based on the reference-channel and, like the proposed method, does not require a model of the unknown stimulus-evoked HR. As in the section "Reference-channel modeling corrected Bayesian approach (ReMCoBA)", data were first pre-processed by band-pass filtering (Butterworth, pass band: from .01 Hz to 1.25 Hz, Zhang et al., 2009). The upper cutoff frequency was reduced from 3 to 1.25 Hz: such a choice allowed to take into account the cardiac component facilitating the correlation analysis made below, while removing higher frequencies which would not be removed by the Bayesian filtering approach used in ReMCoBA. Then, similarly to what was done in the section "Step 1: reference-channel modeling and fNIRS data correction", for each standard-channel, the reference-channel signal with the highest correlation was scaled to fit the former signal in a least-squares sense (Saager and Berger, 2005). If the correlation between the reference- and standard-channel signals was greater than .6, the scaled reference-channel signal was subtracted from the standard-channel signal, otherwise no subtraction was done, for doing so brings with it the potential of increasing noise consequently (Zhang et al., 2009). The obtained time-series was then segmented into 12 s trials and those belonging to the same condition were grouped. Trials with artifacts were discarded following the same criteria in the section "Step 2: single trial Bayesian filtering and HR estimation". After CA of the trials corresponding to the same condition, residual high frequency noise was reduced as in the section "Step 2: single trial Bayesian filtering and HR estimation" by a Savitzky-Golay's filter with polynomial order equal to 3 and frame-size equal to 25 time-points (corresponding to 3 s). The HR estimate is finally obtained after a baseline-correction done with the same strategy in the section "Step 2: single trial Bayesian filtering and HR estimation".

Conventional averaging (CA)

The second method entailed the same operations as that in the section "Reference-channel based fNIRS data correction followed by conventional averaging (rCA)", but reference-channels were not used. For simplicity, it will be referenced to as conventional averaging (CA). Since differences in the results will be due to the use of the reference-channel only, the comparison CA vs. rCA can be exploited to quantitatively demonstrate the benefits of the use of the reference-channel in HR estimation.

Results

In this Section, the new method (ReMCoBA) and the two literature methods (rCA and CA) are compared by using simulated and real data, and defining some objective indexes to quantify accuracy of HR estimates. For the sake of completeness, in Appendix we report results, obtained by HR peak amplitudes analyses, on the different activation in each hemisphere induced by each condition, in line with previous findings of the literature regarding finger-tapping tasks.

Results on synthetic data

Fig. 3a (left) shows a representative example of synthetically generated HR (black line) estimated by ReMCoBA (green line), rCA (blue line) and CA (red line). Qualitatively, the ReMCoBA profile seems closer to the true profile than the rCA and the CA ones. Also, of note is that the rCA profile seems more accurate than CA. Similar results are obtained for all the other synthetic subjects.

In order to quantitatively assess the performance of the three methods and the impact of the reference-channel, four quantitative accuracy indexes, E_{HR}, E_A, E_L and E_D were defined as follows. E_{HR} was a percentage estimation error index, defined as:

$$E_{HR} = 100^{*}\|\{u_{true}\}-\bar{u}\|^2/\|u_{true}\|^2 \quad (16)$$

where $\mathbf{u_{true}}$ and $\tilde{\mathbf{u}}$ were the true and the estimated HR, respectively. Indexes E_A and E_L were used to describe the error in estimating crucial HR parameters such as peak amplitude and latency and were defined as:

$$E_A = 100^{*}|A_{true}-A|/|A_{true}| \quad (17)$$

$$E_L = 100^{*}|L_{true}-L|/|L_{true}| \quad (18)$$

where A_{true} and L_{true} were the peak amplitude and latency of the true HR, while A and L were those of the estimated HR. Finally, index E_D, defined as:

$$E_D = 100^{*}\sqrt{E_A^2+E_L^2} \quad (19)$$

was used to assess the accuracy in estimating both amplitude and latency at the same time. To quantitatively compare ReMCoBA, rCA and CA, these four parameters have been computed for each of the 30 simulated participants. The mean value of the 30 participants and its standard deviation are reported in Figs. 4a, b, c and d. Channels with full HR (peak amplitude equal to 400 nM) and channel with halved HR (peak amplitude equal to 200 nM) were kept separately. Notably, the best E_{HR} was obtained with ReMCoBA. In the case of entire HR (peak amplitude equal to 400 nM), ReMCoBA reduced the estimation error by 16.14% and by 50.54% for HbO, and 10.92% and 17.10% for HbR relative to rCA and CA, respectively. In the case of halved HR (peak amplitude equal to 200 nM), ReMCoBA reduced the estimation error of 13.20% and 49.44% for HbO and 19.57% and 24.48% for HbR with respect to rCA and CA respectively.

Similarly, ReMCoBA reduced E_A, E_L, as well as E_D. The higher values of the error in the HbR case were due to its smaller amplitude compared to HbO (it is reduced to 25%); thus, the HR in the acquired signal had a smaller signal-to-noise ratio (SNR), and its estimation was consequently more difficult compared with the HbO case. For the same reason, error indexes had greater values in the case of halved HR. The best values, for all indexes, obtained by rCA with respect to CA, underline the benefits provided by the use of the reference-channel in HR estimation. Values of each index for both HbO and HbR were submitted to separate repeated measures ANOVAs with method (ReMCoBA, rCA and CA) and amplitude (full vs. halved) as within-subject factors. All ANOVAs revealed a main effect of the method (for details, see Table 2). The performance of the different methods as measured with the different indexes has been compared with a series of paired *t*-tests: statistically significant differences are indicated by gray ($p<.05$) and black ($p<.01$) horizontal lines over the corresponding bars in Figs. 4a, b, c and d. Remarkably, the ReMCoBA obtained the best performance for all indexes and conditions, with a significant difference in most cases.

Additional analyses for the simulated data were performed to quantitatively compare the performance of ReMCoBA with the other two methods (CA and rCA). The estimation error (E_{HR}), the most important index among those used in our work since it considers the difference in the whole profile between true and estimated HR, was used to objectively compare the methods on a single channel basis.

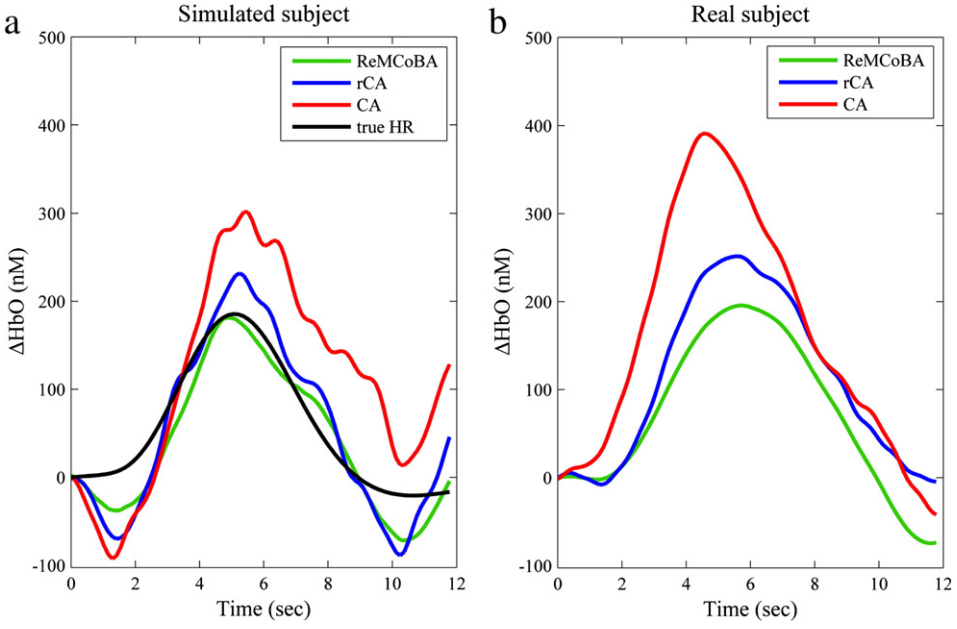

Fig 3. Representative examples of the estimated HRs by the proposed method (ReMCoBA) (green), by rCA (blue) and by CA (red), on simulated data (a) and real data (b). On simulated data (a) the true HR (black line) is also shown.

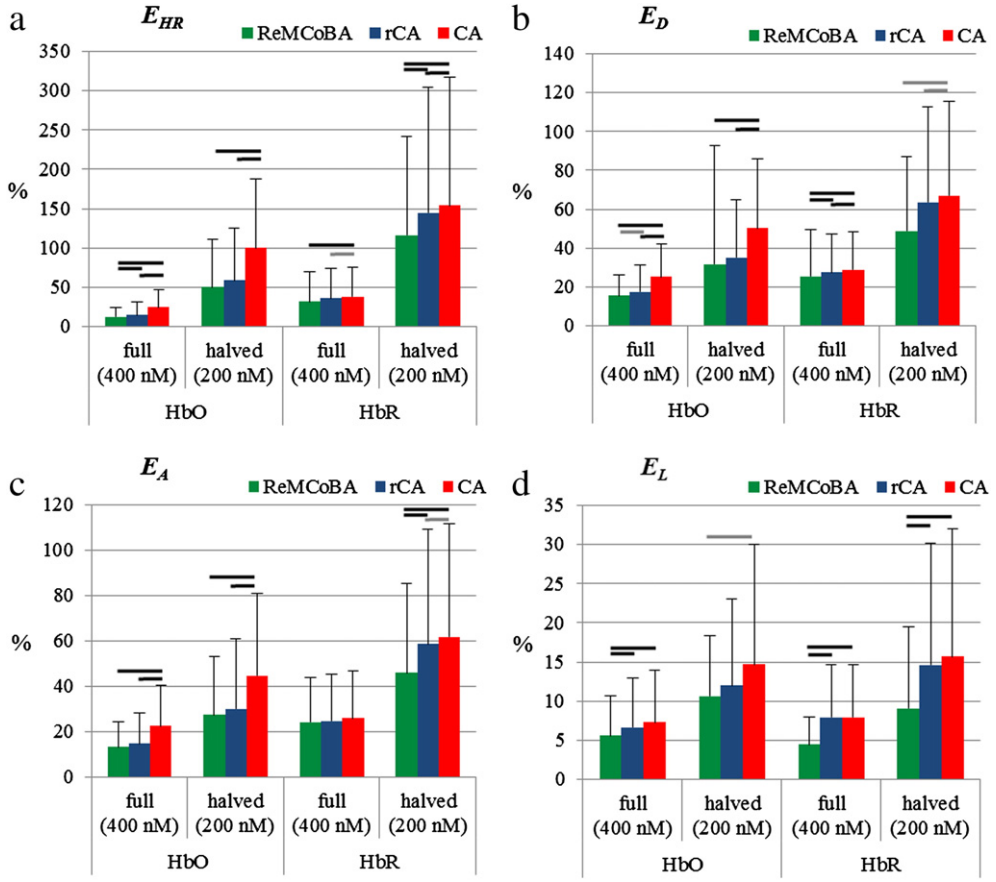

Fig 4. Estimation error (E_{HR}) (a), Euclidean distance between estimated and true peak (E_D) (b), error on peak amplitude (E_A) (c) and peak latency (E_L) (d). Means and standard deviations computed across all simulated participants, Statistical differences between the three methods are indicated by a gray ($p<.05$) or black ($p<.01$) horizontal line over the corresponding bars. Evoked hemodynamic response with peak amplitude equal to 400 nM (full) and 200 nM (halved) are analyzed separately.

With respect to CA, ReMCoBA improved E_{HR} in 84% of the simulated channels for HbO, and 70% for HbR. With respect to rCA, ReMCoBA improved E_{HR} in 58% of the simulated channels for HbO, and 61% for HbR. Results are reported in Figs. 5 and 6: 300 points should be displayed (30 subjects each for 10 channels) but for the sake of clarity each subject is represented by one dot, where this dot is the mean E_{HR} of all his channels. While the performances of ReMCoBA and CA are markedly different, with a significant reduction of E_{HR} obtained by ReMCoBA, the improvement obtained by ReMCoBA with respect to rCA is clear but less apparent, in line with results shown in Fig. 4a.

A ROC analysis was performed to test the performance of the methods considered in the present context following the systematic variation of SNR values (Machado et al., 2011). HRs from all simulated subjects were split in order to obtain a subset of active HRs with a full HR amplitude (≈400 nM) and a subset of non-active HRs with HR true amplitude equal to 0. For HbO, HRs derived by channels 1, 4, 5 in Condition 1 and channels 6, 9, 10 in Condition 2 were treated as active HRs, whereas HRs derived by channels 6, 9, 10 in Condition 1 and channels 1, 4, 5 in Condition 2 were treated as non-active HRs. An analogous subdivision was applied on channels with halved HR (≈200 nM). In this latter case, for HbO, HRs derived by channels 2 and 3 for Condition 1 and channels 7 and 8 for Condition 2 were treated as active HRs, and HRs derived by channels 7 and 8 in Condition 1 and channels 2 and 3 in Condition 2 were treated as non-active HRs. HbR underwent the same procedure. Full and halved HRs led to two different SNRs, whose values were −17 dB (full HR) and −23 dB (halved HR) for HbO, and −20 dB (full HR) and −26 dB (halved HR) for HbR. In order to provide parameters reflecting a substantial portion of the HR function, and not only of the peak amplitude, an area subtended by the HR curve in a 4 s interval centered at the peak's latency was considered. A threshold was set on the value of the 4 s area to enable the separation of true from false positive

Table 2
ANOVAs results on simulated data.

	HbO						HbR					
	Method		Amplitude		Method∗amplitude		Method		Amplitude		Method∗amplitude	
	F(2,58)	P	F(1,29)	P	F(2,58)	P	F(2,58)	p	F(1,29)	P	F(2,58)	F(2,58)
E_{HR}	54.3	<.001	63.8	<.001	23.7	<.001	15.5	<.001	74.1	<.001	11.5	<.001
E_D	36.9	<.001	69.3	<.001	7.8	=.001	19.5	<.001	64.3	<.001	12.8	<.001
E_A	32.1	<.001	50.2	<.001	5	=.01	11.4	<.001	49.7	<.001	10.8	<.001
E_L	6.4	=.003	36	<.001	1.9	=0.159	26.4	<.001	49.2	<.001	3.1	=.053

Results obtained by repeated measures ANOVAs, considering method (ReMCoBA, rCA and CA) and amplitude (full vs. halved) as within-subject factors, for estimation error (E_{HR}), Euclidean distance between estimated and true peak (E_D), error on peak amplitude (E_A), error on peak latency (E_L).

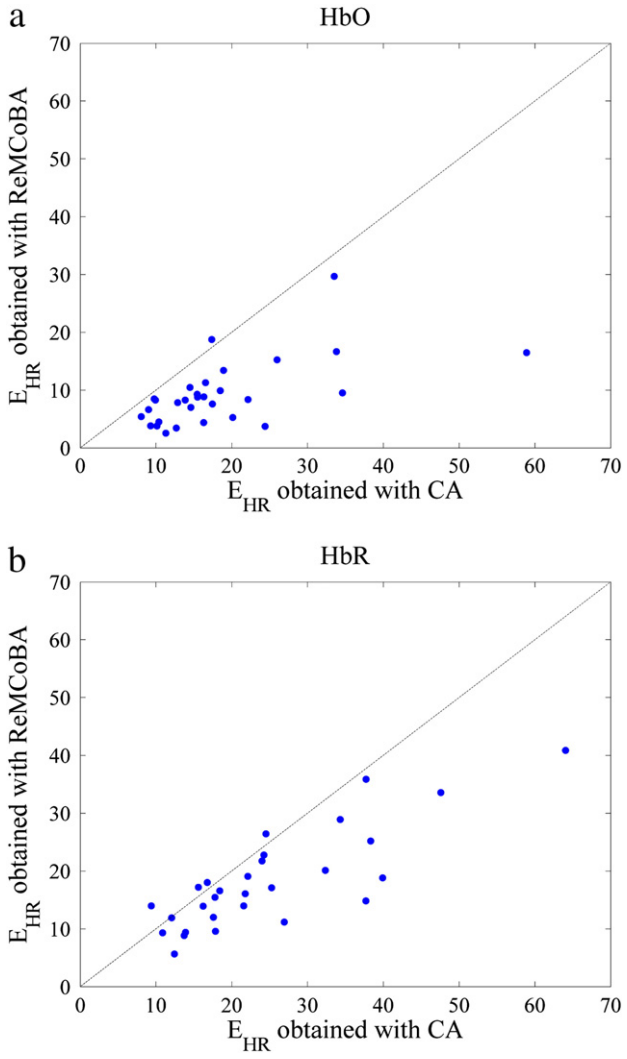

Fig 5. Estimation error (E_{HR}) obtained with CA (horizontal axis) and ReMCoBA (vertical axis), for HbO (a) and HbR (b). Each dot is the mean E_{HR} obtained from channels relative to the same simulated subject.

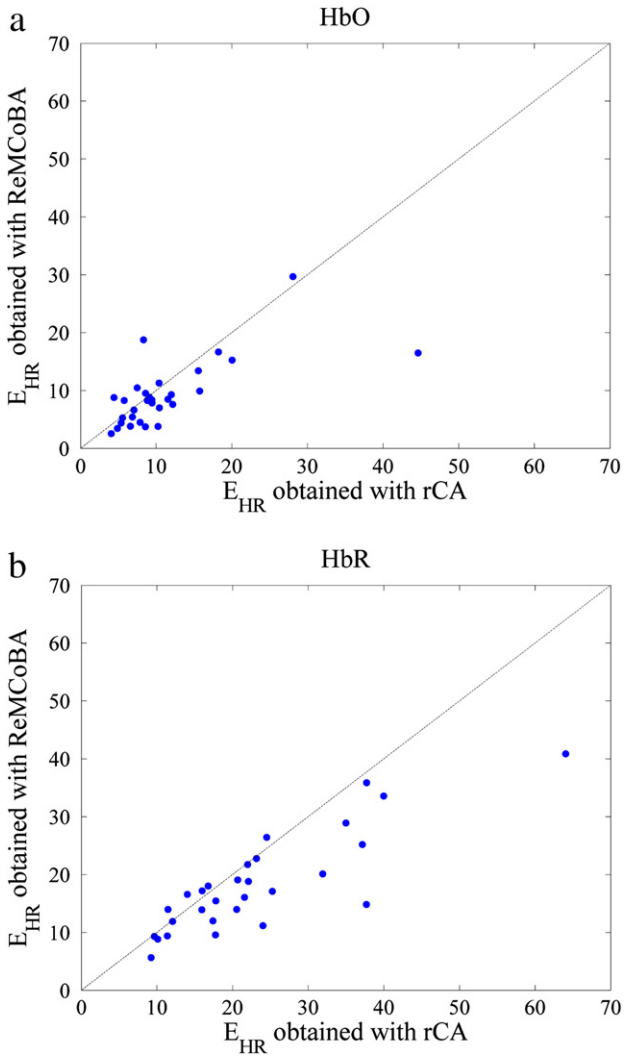

Fig 6. Estimation error (E_{HR}) obtained with rCA (horizontal axis) and ReMCoBA (vertical axis), for HbO (a) and HbR (b). Each dot is the mean E_{HR} obtained from channels relative to the same simulated subject.

rates. The threshold value was then varied from the greatest area, where neither true nor false positives were found, to the lowest area, where the concentration of true and false positives was assumed to be maximal. A total of four cases were generated for the analyses, each contributing 200 points. Table 3 reports the values of the area (*AUC*) under the curve estimated using the present ROC approach. The results of the ROC analyses are illustrated in Figs. 7a, b, c and d. The results from the different methods were compared via a series of paired two-tailed *t*-tests on the individual *AUC* values. Significant differences were found in HbO (full HR) between ReMCoBA and CA ($t(29)=2.041$, $p=.05$), in HbO (halved HR) between rCA and CA ($t(29)=2.371$, $p<.05$), and between both ReMCoBA and rCA ($t(29)=2.948$, $p<.05$) and ReMCoBA and CA ($t(29)=2.916$, $p<.05$) in HbR (full HR). No other significant difference was detected in the comparison. In line with the results described in the foregoing sections, the ROC analyses corroborated the remarkable superior performance of ReMCoBA relative to CA. Albeit with slightly less power, the results of the ROC analyses – and the particularly clear pattern visible in Fig. 7 – allows us to argue that ReMCoBA performed better than rCA (and particularly so in the case of HbR, full HR), given the highest *AUC* value obtained by ReMCoBA vs. rCA in three out of the four tests.

Additional analyses were performed on the precision of the HR estimates provided by ReMCoBA. The number of parameters used by our method ranges between 10 and 16 (depending on *M*), if the reference-channel is used, while it is 5 otherwise. This number is of the same order of that of previously published methods, e.g. at least 6 parameters in Prince et al. (2003), 8 parameters in Abdelnour and Huppert, (2009), 16 parameters in Gagnon et al., (2011). A lower number of parameters would not be able to describe the complexity of the fNIRS signal shown by the data, resulting in a biased description of noise and in suboptimal HR estimates. Indeed, our procedure tries to minimize model complexity by letting the data indicate the most suited model order. For instance, *M* in Eq. (8) can be 1, 2 or 3, depending on the data, while in other methods (e.g. Prince et al., 2003) *M* is identically equal to 3 for all the trials (with consequent risk of overfitting). Similarly, the order 4 for the auto-regressive model was determined by AIC. In general, an aspect of strength of

Table 3
Area under curve (AUC) of ROC curves.

	HbO		HbR	
	Full HR	Halved HR	Full HR	Halved HR
ReMCoBA	.9777	.8535	.9272	.7780
rCA	.9679	.8621	.9052	.7689
CA	.9390	.7847	.8934	.7556

Area under curve (AUC) obtained by each method (ReMCoBA, rCA and CA) in the ROC analysis in case of full HR (≈400 nM) and halved HR (≈200 nM), for HbO and HbR.

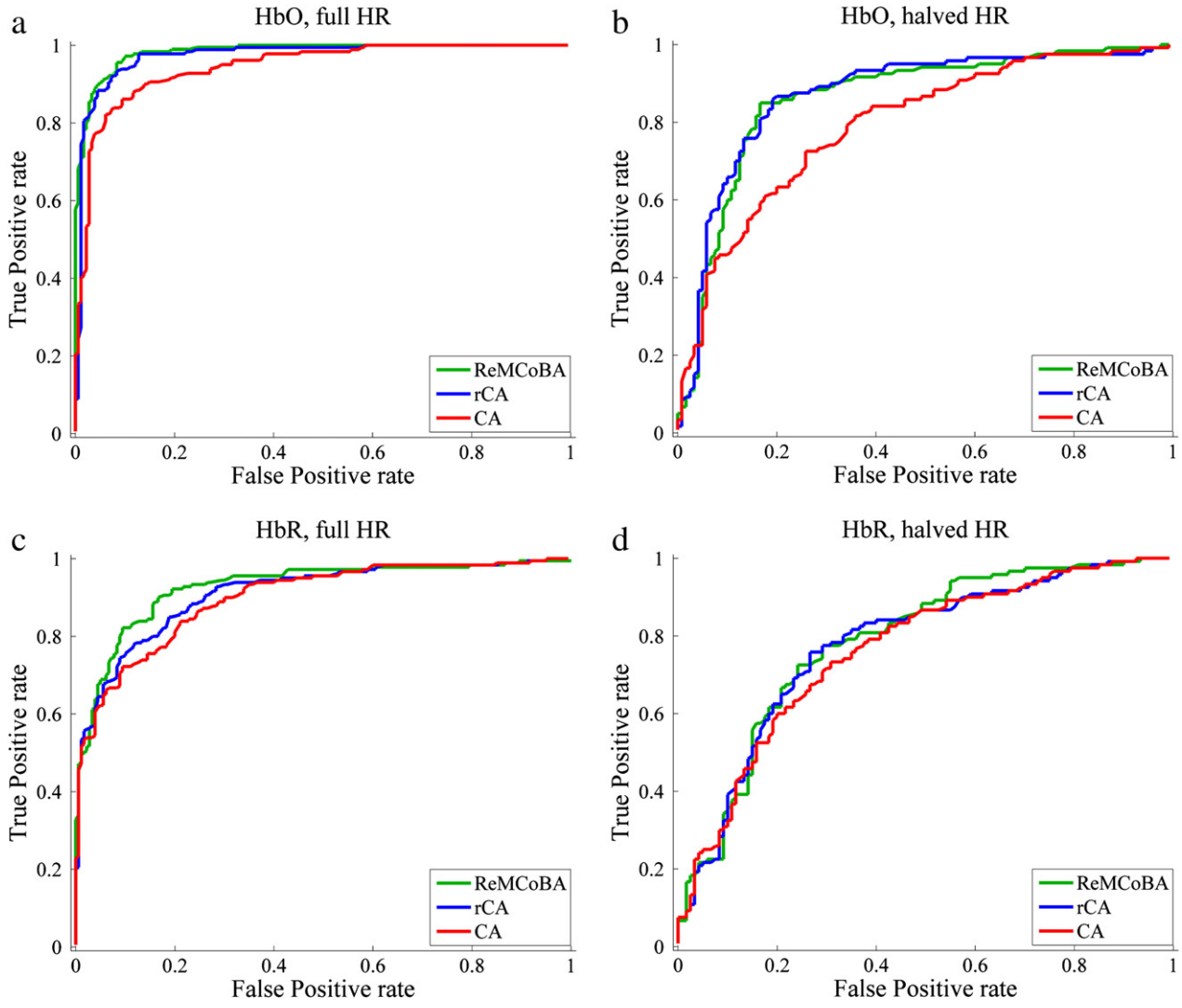

Fig 7. ROC curves obtained by ReMCoBA (green), rCA (blue) and CA (red) in the case of (a) HbO, full HR, (b) HbO, halved HR, (c) HbR, full HR and (d) HbR, halved HR. Full HR (≈400 nM) and halved HR (≈200 nM) denote two different SNR, about −17 dB and −23 dB respectively for HbO, and −20 dB and −26 dB for HbR.

ReMCoBA is that the values of all the employed parameters are computed by objective and reproducible (even if, sometimes, empirical to some extent) procedures, the aim being to reduce possible user's subjectivity as much as possible. In order to have a quantification of the precision of the HR estimates obtained by ReMCoBA, we performed a Monte Carlo simulation. For each simulated subject, 200 realizations of the same channel were created, each with a different random noise sequence ($\eta(t)$ in Eq. (1)). From each of these 200 realizations, the HR was estimated. From the sample distribution of the estimates, the average and the 5th and 95th percentile variability bands were finally computed. As shown in Figs. 8a,b, the bias was very small. The fact that the variability band was narrow indicates that the procedure was not overly sensitive to noise, as it would happen if models were overfitting the data. In particular, the mean variance of E_{HR} was equal to 2.3% for HbO and 5.1% for HbR, denoting the good precision of the estimates obtained by ReMCoBA.

Results on real data

A representative example of the HR obtained in a real subject is reported in Fig. 3b (right). On this particular dataset, all methods achieved a reasonable HR estimation. However, differences in peak amplitude and latency were apparent. Obviously, in the real data case a quantitative/objective assessment of the three HR estimation algorithms was more difficult to be done than in the simulated data case. The following strategy has been considered. First, samples of each estimated HR were fitted by the canonic mathematical model of the HR, Eq. (2), by a nonlinear least squares algorithm (initial values of the parameters, i.e., $\alpha=1202$, $\varphi_1=.5$, $\varphi_2=3.5$, $\tau_1=\tau_2=.8$ and $\beta=.6$, were derived by an overall average of all real participants and channels). The fit was then computed on each estimated HR with lower and upper constraints on the model's parameters ($0\leq\alpha<\text{inf}$, $-1.5<\varphi_1<1.5$, $2.5<\varphi_2<5.5$, $.5<\tau_i<.9$ and $0<\beta<.7$). The constraints on amplitude (α) and time delay (φ_i) were set to take into account the possible variability between participants and channels, and, in order to obtain a good fit, only if the estimated HR had a reasonable physiological shape (Figs. 9a,b). Stringent constraints were set, instead, to the parameters (τ_i) which tuned the shape of the HR and to the amplitude of the undershoot (β). The coefficients of the fit (R^2, *rmse*) were used to have a grasp of the "likelihood" of the estimated HR.

For each index, a mean value was obtained by each of the 10 real participants. For each method, the mean value of the 10 participants and its standard deviation are reported in Figs. 10a,b. All methods achieved good results on the Pearson's correlation coefficient (R^2), ≈.9 and ≈.8 for HbO and HbR respectively. Values were submitted to a repeated measures ANOVA, considering method as within-subject

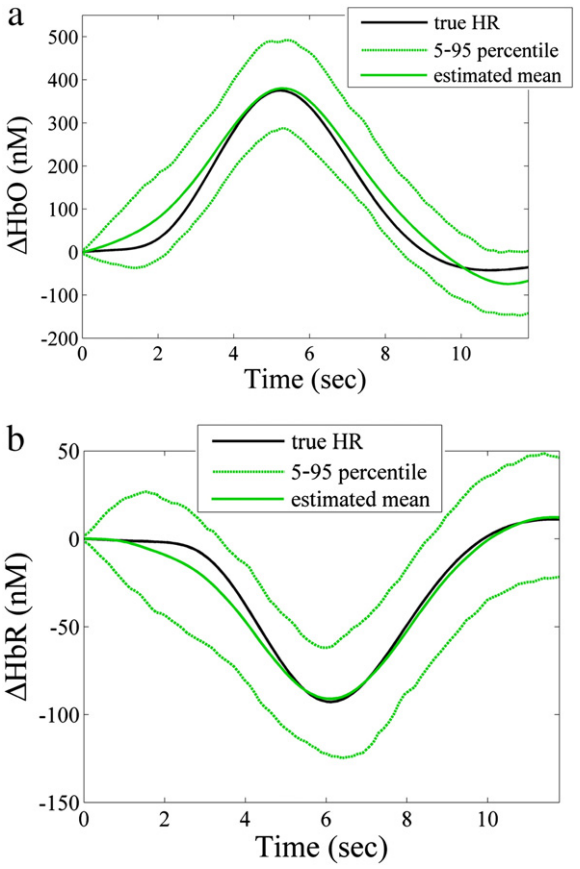

Fig 8. Monte Carlo estimate: true HR (black line), estimated mean (green line) obtained by 200 realizations of a channel with different random noise, 5th and 95th percentiles (dotted green line), for HbO (a) and for HbR (b).

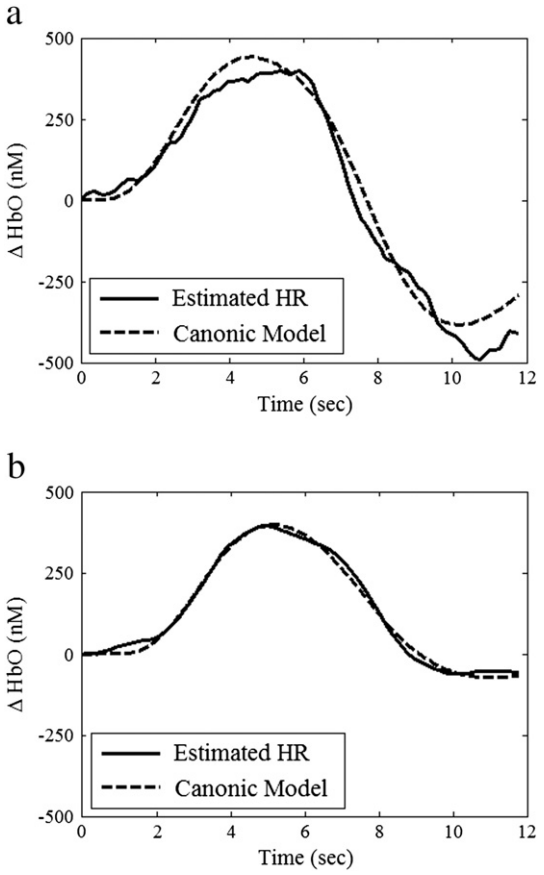

Fig 9. Representative examples of canonic mathematical model of the hemodynamic response (dashed line) fitted to estimated HR (solid line). On the left (a), the model cannot properly fit the estimated HR because of the presence of an unrealistic undershoot, whose amplitude is greater to that of the peak, probably due to the presence of physiological noise not yet removed by the signal processing method ($R^2 = .96$, $rmse = 44$). On the right (b), the estimated HR has a reasonable physiological shape and it is well described by the canonic model ($R^2 = .99$, $rmse = 15$).

factor, for both HbO and HbR. ANOVAs revealed a significant effect of the method (Table 4). Hence, a paired *t*-test was conducted between methods. Statistical differences between the three methods are indicated by gray ($p<.05$) and black ($p<.01$) horizontal lines over the corresponding bars in Figs. 10a,b. The proposed method achieved the best R^2, with a significant difference with respect to CA for HbO, and to rCA and CA for HbR. The same analyses were conducted on the root mean square error (*rmse*) and the new methodology achieved the lowest *rmse*, with a significant difference ($p<.05$) with respect to rCA and CA, for both HbO and HbR. Again, the best values, for both R^2 and *rmse*, obtained by rCA with respect to that obtained by CA, underline the benefits provided by the use of the reference-channel in HR estimation.

Discussion

The fNIRS is an emerging neuroimaging technique which can be usefully employed to provide, in a non-invasive way and with reasonable laboratory costs, useful information for the study of cerebral activity. Unfortunately, in the event-related design used in the study of cognitive processes, the estimation of HR from fNIRS signals is especially challenging due to its small amplitude with respect to that of the other ongoing physiological components. In the present paper, a methodology to improve the estimation of HR from fNIRS signals has been presented. The new reference-channel modeling corrected Bayesian approach (ReMCoBA) is a two-step method with no assumption on HR shape, duration, amplitude and latency, nor on the kind of experiment and stimulus. In the first step, ReMCoBA takes advantage from the availability of noise-only data, the so-called reference-channel signal, which allows a massive reduction of physiological noise (due to respiration, vasomotor waves, etc.) disturbing HR estimation from standard-channels. In Step 1, in fact, the reference-channel signal is used to estimate a parametric model of physiological noise, which is in turn used to correct the signal recorded from standard-channels. The model derived from the reference-signal describes low frequency physiological components, so as to remove from standard-signal only these components, without the risk of propagating noise from reference-channel to standard-channel. To perform Step 1, a good correlation between reference and standard-channel is necessary, otherwise physiological noise may be amplified instead of reduced. Indeed, in the present work only 32% of standard-channels have a Pearson's correlation coefficient with at least one of the reference-channels greater than .6: the majority of the channels skips data correction and goes directly to Step 2. The setting of a threshold to a high value of the Pearson's correlation coefficient is crucial, because if the physiological model subtracted from a standard-channel is derived by a not sufficiently correlated reference-channel, the obtained HR estimate gets worse instead of improving. This can be seen in Figs. 11a,b, which show the percentage difference of the estimation error obtained with and without the use of the reference-channel in relation to the Pearson's correlation coefficient during the resting period. In agreement with Zhang et al. (2009), data in Figs. 11a,b suggest a correlation coefficient threshold of approximately 0.6: when the correlation coefficient is lower than 0.6, the obtained estimation error increases in 50% of channels and in some cases it drastically increases (three and even four times) with respect to the estimation error obtained without the use of the reference-channel.

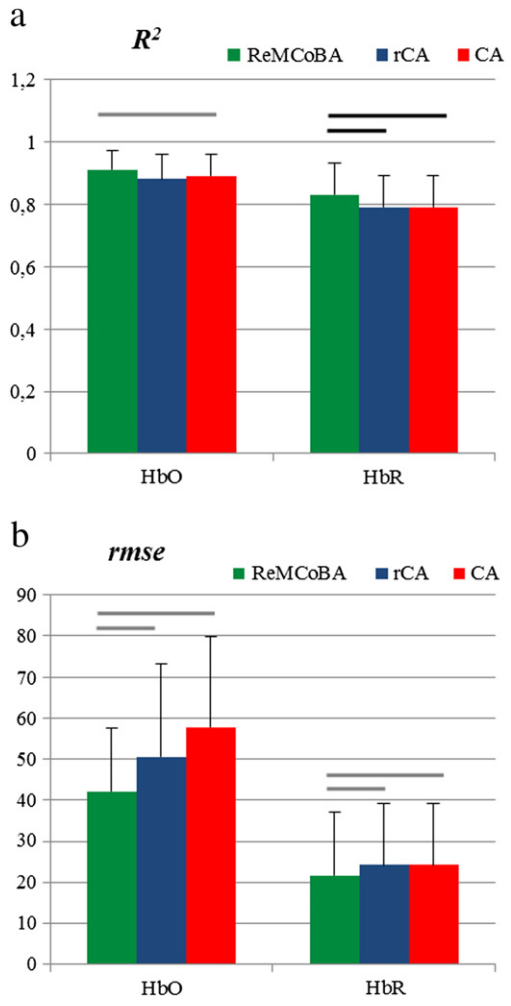

Fig 10. a,b. Pearson's correlation coefficients (R^2) and root mean square error (*rmse*) between estimated hemodynamic responses and the canonic model of the hemodynamic response. Means and standard deviations are computed across all real participants. Statistical differences between the three methods are indicated by a gray ($p<.05$) or black ($p<.01$) horizontal line over the corresponding bars.

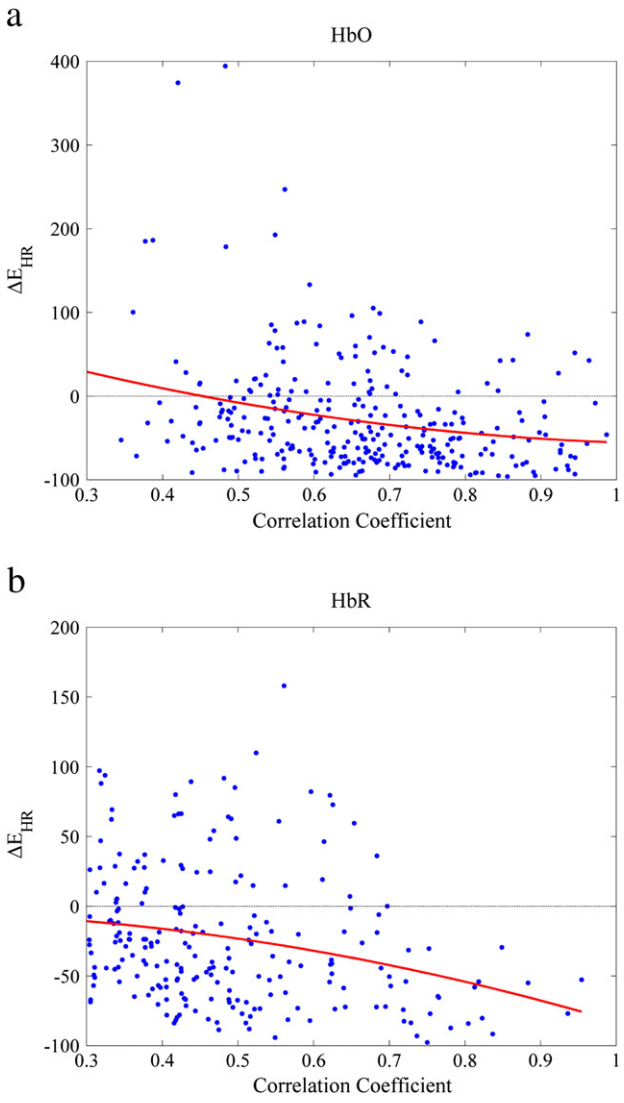

Fig 11. Percentage difference of the estimation error obtained with and without the use of the reference-channel ($\Delta E_{HR} = 100 * (E_{HRwith} - E_{HRwithout}) / E_{HRwithout}$) in relation to the Pearson's correlation coefficient during the resting period (from 0.3 to 1), for HbO (a) and for HbR (b). Red lines represent the second order polynomial fit.

Step 2 embeds a non-parametric (i.e., no assumptions on shape, amplitude and latency) Bayesian approach which is able to individuate, on a trial-by-trial basis, a suitable compromise between measured noisy data and a priori expectations on HR smoothness. The procedure in Step 2 is in large part that already presented in Scarpa et al. (2010). Step 1 and 2 take into account the complexity of fNIRS signal, letting the data indicate the most suited model order and parameter values.

The new ReMCoBA method was assessed on both simulated and real data against two other literature methods. For simulated data, several quantitative accuracy indexes could be obtained. The comparison of ReMCoBA and rCA, a method which subtracts a scaled version of the reference-signal from standard-channels, reveals a significant improvement in HR estimation (with a statistically significant difference for all accuracy indexes detectable in simulated data). A CA methodology that does not use the reference-channel was also used as a comparative method, in order to further validate the use of the reference-channel. The fact that CA leads to the worst HR estimates supports the importance of the reference-channel. The superior performance of ReMCoBA over the other two methods was confirmed by the results of ROC analyses considering full and halved HR in each cell of the present design and for each subject estimated based on both HbO and HbR concentration indices. Notably, as regards real data, we used a finger tapping task instead of using a more sophisticated paradigm, because we wanted to obtain a known HR. In this way, we provide an ideal platform to compare the performance of the three methods straightforwardly.

Given that the total number of channels available in NIRS devices is typically limited, the use of only two reference-channels for hemisphere is an important advantage of the proposed method with respect to other reference-channel based methods (Gagnon et al., 2011, Zhang et al., 2009). There is ample evidence in the literature that low frequency physiological fluctuations are highly correlated across space, with some relatively constant time delay across time (Cooper et al., 2011; Frederick et al., 2012; Tong & Frederick, 2010; Tong et al., 2011). Even if the Pearson's correlation coefficient between reference and standard-channel is generally greater if they are close to each other (<2 cm) (Gagnon et al., 2012), we found good correlation (>.6) also between

Table 4
ANOVAs results on real data.

	HbO		HbR	
	F(2,18)	P	F(2,18)	p
R^2	4.0	=.036	14.5	<.001
rmse	5.9	=.011	6.6	=.007

Results obtained by repeated measures ANOVAs, considering method (ReMCoBA, rCA and CA) as within-subject factor, for the Pearson's correlation coefficients (R^2), the root mean square error (*rmse*) between estimated hemodynamic responses and the canonic model of the hemodynamic response.

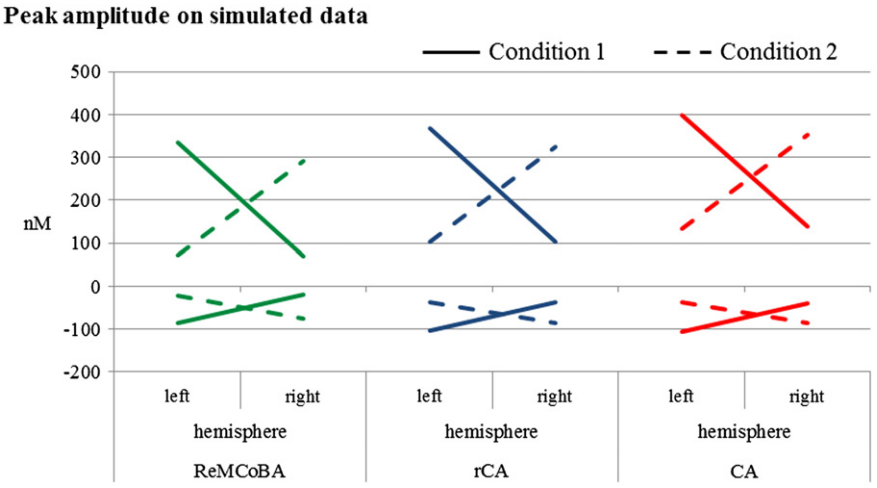

Fig 12. Mean values of the peak amplitude obtained for each condition (movement of the right or left forefinger for Condition 1 and Condition 2 respectively) and hemisphere (left or right) by the three methods (Condition 1 is represented by a solid line, Condition 2 is represented by a dashed line). Mean values are computed across all simulated participants.

distant channels, even from those located in different hemispheres: the model of physiological components is derived by the reference-channel with the greatest correlation coefficient with the considered standard-channel, and 45% of the selected reference-channel is located in the opposite hemisphere. However, just because their number is limited, it is important to maximize the SNR of the signal acquired by reference-channels, so special attention must be paid on their location (in order to minimize the distance from standard-channels) and on a perfect contact between optic fibers and the participant's skin.

Furthermore, the distance between source and detector of each reference-channel must be less than 1 cm, in order to measure the signal relative to scalp and skull only. If the distance is greater than 1 cm, part of the acquired signal is relative to the cerebral cortex and may contain the stimulus-evoked hemodynamic response. The probe arrangement used for the placement of our sources and detectors provides, besides the 10 standard-channels and the 4 reference-channels so far mentioned, 4 additional channels (sources 1 and 2 with detector C on the left hemisphere and sources 1 and 2 with detector D on the right hemisphere) (Fig. 1) with a source-detector distance equal to 1.5 cm. The HRs estimated by these channels are comparable with those of standard-channels (peak amplitude $\approx$200 nM). Thus, they probably contain also some cerebral signal. Since they cannot be used as reference-channels nor as normal-channels, they have been discarded from analysis.

Simulated and real data on which the new method has been tested on comprised two conditions, each consisting in the repetition of an identical stimulus. In order to find out the minimum number of repetitions (trials) needed to obtain a good HR, we estimated HR using 30, 40 (so far discussed), 50 trials and we evaluated the corresponding estimation error on simulated data. No significant difference was found in the estimation of the event-related hemodynamic response between 40 and 50 trials (for HbO, E_{HR} was equal to 12.2% and 10.2% respectively; for HbR, E_{HR} was equal to 31.9% and 21.7% respectively), but a significantly greater estimation error was found in the 30 trial case (E_{HR} was equal to 18.6% for HbO and 43.8% for HbR). Thus, when the signal is dominated by physiological components, a minimum number of 40 trials seems to be mandatory for a correct estimation of HR in an event-related design.

In conclusion, the proposed method based on a model derived by the reference-channel provides a valuable estimation of the stimulus-evoked hemodynamic response, without a priori information about its shape, duration, amplitude, latency and no model of the unknown hemodynamic response is required. These features make the proposed method a general and flexible way to correctly estimate evoked hemodynamic response.

Acknowledgments

This study was supported by the University of Padova project "Quantitative understanding of the human brain functioning through advanced EEG and NIRS signal processing" and by a grant from the European Research Council (210922-GENMOD).

Appendix A

Investigation of differences between hemispheres and conditions

In order to demonstrate the effectiveness of the used methodologies, hemodynamic responses estimated by the proposed method (ReMCoBA) and by the methods used for comparison (rCA and CA) were analyzed to confirm with each method previous findings known by literature regarding finger tapping tasks. The standard investigation of fNIRS data consists in the analysis of HR's peak amplitudes. Note that, on simulated data, the whole HRs and in particular their peak amplitudes and peak latencies were created just according to information known by literature. Since we were not interested in the HR obtained by each single channel but in differences between conditions and hemispheres, all channels of the same hemisphere were grouped, while the two conditions have been kept separate.

Table 5
ANOVAs results on simulated data.

	HbO						HbR					
	Hemisphere		Condition		Hemisphere∗condition		Hemisphere		Condition		Hemisphere∗condition	
	F(1,29)	P	F(1,29)	P	F(1,29)	p	F(1,29)	p	F(1,29)	p	F(1,29)	p
ReMCoBA	16.3	<.001	2.6	=.115	3155.6	<.001	10.2	=.003	.712	.406	841.8	<.001
rCA	15.2	=.001	3.7	=.065	1914.6	<.001	11.2	=.002	4.1	.053	892.0	<.001
CA	9.7	=.004	1.5	=.234	1039.3	<.001	10.7	=.003	4.8	.036	818.8	<.001

Results obtained by repeated measures ANOVAs, with condition (right- vs. left-hand tapping) and hemisphere (left vs. right) as within-subjects factors, for ReMCoBA, rCA and CA. The mean peak amplitude of every participant obtained grouping channels by hemisphere and condition was considered.

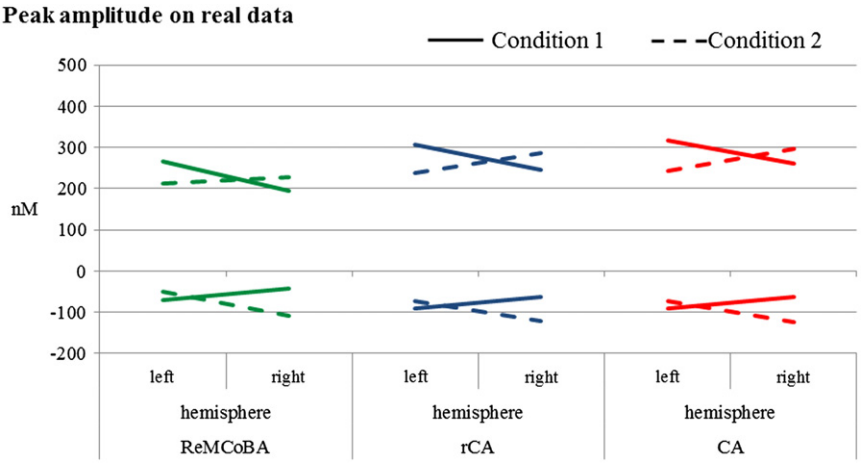

Fig 13. Mean values of the peak amplitude obtained for each condition (movement of the right or left forefinger for Condition 1 and Condition 2 respectively) and hemisphere (left or right) by the three methods (Condition 1 is represented by a solid line, Condition 2 is represented by a dashed line). Mean values are computed across all real participants.

Then, peak amplitudes of each channel in both conditions have been analyzed in order to find significant differences between hemispheres and conditions. The mean peak amplitude of every participant obtained grouping channels by hemisphere and condition (e.g. mean value of the channels of the right hemisphere in Condition 1) was considered. The values obtained with the three methods (reported in Fig. 12) have been separately submitted to a repeated measures ANOVA, with condition (right- vs. left-hand tapping) and hemisphere (left vs. right) as within-subject factors. Although all the ANOVAs revealed a significant interaction between hemisphere and condition for all methods (Table 5), the HRs estimated with the proposed method were the closest to the real values (336 nM for Condition 1 in the left hemisphere and 288 nM for Condition 2 in the right hemisphere for HbO, and −88 nM for Condition 1 in the left hemisphere and −72 nM for Condition 2 in the right hemisphere for HbR).

In order to find significant differences between hemispheres and conditions, the same analyses conducted on simulated data have been replicated on real data, analyzing peak amplitudes of each channel and condition. Similarly to simulated data, the mean values for each condition and hemisphere obtained with the three methods are reported in Fig. 13. As in simulated data, they were submitted to repeated measures ANOVAs, with hemisphere and condition as within-subject factors. For each method, ANOVAs revealed a significant (Table 6) interaction between hemisphere and condition. In real data, hemodynamic activity was observed both contralaterally and ipsilaterally to the moved forefinger. Anyway, as expected, the amplitude of the HRs was greater in the contralateral hemisphere, without reaching significance, though. Instead, a significant difference between hemispheres was found on simulated data, but (unlike real data) HR was added contralaterally only. Equivalent outcomes are found for HbR. All the results obtained in real data are in line with previous findings concerning finger tapping protocols (Franceschini et al., 2006; Holper et al., 2009; Leff et al., 2011; Lutz et al., 2005; Sato et al., 2007).

References

Abdelnour, a F., Huppert, T., 2009. Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model. NeuroImage 46 (1), 133–143.
Brigadoi, S., Cutini, S., Scarpa, F., Scatturin, P., Dell'Acqua, R., 2012. Exploring the role of primary and supplementary motor areas in simple motor tasks with fNIRS. Cogn. Process. 13 (1), 97–101.
Boas, D.A., Franceschini, M.A., Dunn, A.K., Strangman, G., 2002. Noninvasive imaging of cerebral activation with diffuse optical tomography. In: Frostig, R.D. (Ed.), In-Vivo Optical Imaging of Brain Function. CRC Press, pp. 193–221 (Chap 8).
Bunce, S.C., Izzetoglu, M., Izzetoglu, K., Onaral, B., Pourrezaei, K., 2006. Functional near-infrared spectroscopy. IEEE Eng. Med. Biol. Mag. 25 (4), 54–62.
Cooper, R.J., Gagnon, L., Goldenholz, D.M., Boas, D.A., Greve, D.N., 2011. The utility of near-infrared spectroscopy in the regression of low-frequency physiological noise from functional magnetic resonance imaging data. NeuroImage 59 (4), 3128–3138.
Cutini, S., Scatturin, P., Menon, E., Bisiacchi, P.S., Gamberini, L., Zorzi, M., Dell'Acqua, R., 2008. Selective activation of the superior frontal gyrus in task-switching: an event-related fNIRS study. NeuroImage 42 (2), 945–955.
Cutini, S., Scarpa, F., Scatturin, P., Jolicœur, P., Pluchino, P., Zorzi, M., Dell'Acqua, R., 2011a. A hemodynamic correlate of lateralized visual short-term memories. Neuropsychologia 49 (6), 1611–1621.
Cutini, S., Scatturin, P., Zorzi, M., 2011b. A new method based on ICBM152 head surface for probe placement in multichannel fNIRS. NeuroImage 54 (2), 919–927.
Cutini, S., Basso, S., Bisconti, S., 2012. Functional near infrared optical imaging in cognitive neuroscience: an introductory review. J. Near Infrared Spectrosc. 20 (1), 75–92.
Cutini, S., Scarpa, F., Scatturin, P., Dell'Acqua, R., Zorzi, M., in press. Number–space interactions in the human parietal cortex: enlightening the snarc effect with functional near-infrared spectroscopy. Cereb. Cortex. http://dx.doi.org/10.1093/cercor/bhs321.
Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8 (2–3), 109–114.
Ferrari, M., Quaresima, V., 2012. A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. NeuroImage 63 (2), 921–935.
Franceschini, M.A., Joseph, D.K., Huppert, T.J., Diamond, S.G., Boas, D.A., 2006. Diffuse optical imaging of the whole head. J. Biomed. Opt. 11 (5), 054007.

Table 6
ANOVAs results on real data.

	HbO						HbR					
	Hemisphere		Condition		Hemisphere∗condition		Hemisphere		Condition		Hemisphere∗condition	
	F(1,29)	p	F(1,29)	P	F(1,29)	p	F(1,29)	P	F(1,29)	p	F(1,29)	p
ReMCoBA	.804	=.393	.194	=.67	6.3	=.033	2.6	=.142	7.2	=.025	12.0	=.007
rCA	.033	=.86	.289	=.604	16.5	=.003	1.4	=.269	6.2	=.035	14.0	=.005
CA	.003	=.955	.584	=.464	16.8	=.003	2.0	=.194	6.9	=.028	13.7	=.005

Results obtained by repeated measures ANOVAs, with condition (right- vs. left-hand tapping) and hemisphere (left vs. right) as within-subjects factors, for the proposed method (ReMCoBA), rCA and CA. The mean peak amplitude of every participant obtained grouping channels by hemisphere and condition was considered.

Frederick, B.D., Nickerson, L.D., Tong, Y., 2012. Physiological denoising of BOLD fMRI data using Regressor Interpolation at Progressive Time Delays (RIPTiDe) processing of concurrent fMRI and near-infrared spectroscopy (NIRS). NeuroImage 60 (3), 1913–1923.

Gagnon, L., Perdue, K., Greve, D.N., Goldenholz, D., Kaskhedikar, G., Boas, D.A., 2011. Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state–space modeling. NeuroImage 56 (3), 1362–1371.

Gagnon, L., Cooper, R.J., Yücel, M. a, Perdue, K.L., Greve, D.N., Boas, D. a, 2012. Short separation channel location impacts the performance of short channel regression in NIRS. NeuroImage 59 (3), 2518–2528.

Gervain, J., Mehler, J., Werker, J.F., Nelson, C.A., Csibra, G., Lloyd-Fox, S., Shukla, M., Aslin, R.N., 2011. Near-infrared spectroscopy: a report from the McDonnell infant methodology consortium. Dev. Cogn. Neurosci. 1 (1), 22–46.

Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9 (4), 416–429.

Handel, P., 2000. Evaluation of a standardized sine wave fit algorithm. Proceedings of the IEEE Nordic Signal Processing Symposium, Sweden.

Holper, L., Biallas, M., Wolf, M., 2009. Task complexity relates to activation of cortical motor areas during uni- and bimanual performance: a functional NIRS study. NeuroImage 46 (4), 1105–1113.

Huppert, T.J., Diamond, S.G., Franceschini, M.A., Boas, D.A., 2009. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Appl. Opt. 48 (10), D280–D298.

Jasdzewski, G., Strangman, G., Wagner, J., Kwong, K., Poldrack, R., Boas, D., 2003. Differences in the hemodynamic response to event-related motor and visual paradigms as measured by near-infrared spectroscopy. NeuroImage 20 (1), 479–488.

Jöbsis, F.F., 1977. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 198, 1264–1267.

Leff, D.R., Orihuela-Espina, F., Elwell, C.E., Athanasiou, T., Delpy, D.T., Darzi, A.W., Yang, G.-Z., 2011. Assessment of the cerebral cortex during motor task behaviours in adults: a systematic review of functional near infrared spectroscopy (fNIRS) studies. NeuroImage 54 (4), 2922–2936.

Lindquist, M.A., Wager, T.D., 2007. Validity and power in hemodynamic response modeling: a comparison study and a new approach. Hum. Brain Mapp. 28 (8), 764–784.

Lutz, K., Koeneke, S., Wüstenberg, T., Jäncke, L., 2005. Asymmetry of cortical activation during maximum and convenient tapping speed. Neurosci. Lett. 373 (1), 61–66.

Machado, A., Lina, J.M., Tremblay, J., Lassonde, M., Nguyen, D.K., Lesage, F., Grova, C., 2011. Detection of hemodynamic responses to epileptic activity using simultaneous electro encephalography (EEG)/near infra red spectroscopy (NIRS) acquisitions. NeuroImage 56 (1), 114–125.

Näsi, T., Kotilahti, K., Noponen, T., Nissilä, I., Lipiäinen, L., Meriläinen, P., 2010. Correlation of visual-evoked hemodynamic responses and potentials in human brain. Exp. Brain Res. 202 (3), 561–570.

Prince, S., Kolehmainen, V., Kaipio, J.P., Franceschini, M.G., Boas, D.A., Arridge, S.R., 2003. Time-series estimation of biological factors in optical diffusion tomography. Phys. Med. Biol. 48, 1491–1504.

Rosen, B.R., Buckner, R.L., Dale, a M., 1998. Event-related functional MRI: past, present, and future. Proc. Natl. Acad. Sci. U. S. A. 95 (3), 773–780.

Saager, R.B., Berger, A.J., 2005. Direct characterization and removal of interfering absorption trends in two-layer turbid media. J. Opt. Soc. Am. A Opt. Image Sci. Vis. 22 (9), 1874–1882.

Saager, R.B., Telleri, N.L., Berger, A.J., 2011. Two-detector corrected near infrared spectroscopy (C-NIRS) detects hemodynamic activation responses more robustly than single-detector NIRS. NeuroImage 55 (4), 1679–1685.

Sassaroli, A., Fantini, S., 2004. Comment on the modified Beer–Lambert law for scattering media. Phys. Med. Biol. 49 (14), N255–N257.

Sato, T., Ito, M., Suto, T., Kameyama, M., Suda, M., Yamagishi, Y., Ohshima, A., Uehara, T., Fukuda, M., Mikuni, M., 2007. Time courses of brain activation and their implications for function: a multichannel near-infrared spectroscopy study during finger tapping. Neurosci. Res. 58 (3), 297–304.

Scarpa, F., Cutini, S., Scatturin, P., Dell'Acqua, R., Sparacino, G., 2010. Bayesian filtering of human brain hemodynamic activity elicited by visual short-term maintenance recorded through functional near-infrared spectroscopy (fNIRS). Opt. Express 18 (25), 26550–26568.

Taga, G., Watanabe, H., Homae, F., 2011. Spatiotemporal properties of cortical haemodynamic response to auditory stimuli in sleeping infants revealed by multichannel near-infrared spectroscopy. Philos. Transact. A Math. Phys. Eng. Sci. 369 (1955), 4495–4511.

Tong, Y., Frederick, B.D., 2010. Time lag dependent multimodal processing of concurrent fMRI and near-infrared spectroscopy (NIRS) data suggests a global circulatory origin for low-frequency oscillation signals in human brain. NeuroImage 53 (4), 2047–2057.

Tong, Y., Bergethon, P.R., Frederick, B.D., 2011. An improved method for mapping cerebrovascular reserve using concurrent fMRI and near-infrared spectroscopy with Regressor Interpolation at Progressive Time Delays (RIPTiDe). NeuroImage 56 (2), 553–564.

Twomey, S., 1965. The application of numerical filtering of the solution of integral equation encountered in indirect sensing measurements. J. Franklin Inst. 279, 95–109.

Ye, J.C., Tak, S., Jang, K.E., Jung, J., Jang, J., 2009. NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. NeuroImage 44 (2), 428–447.

Zhang, Q., Brown, E.N., Strangman, G.E., 2007. Adaptive filtering for global interference cancellation and real-time recovery of evoked brain activity: a Monte Carlo simulation study. J. Biomed. Opt. 12 (4), 044014.

Zhang, Q., Strangman, G.E., Ganis, G., 2009. Adaptive filtering to reduce global interference in non-invasive NIRS measures of brain activation: how well and when does it work? NeuroImage 45 (3), 788–794.