

Temporal Derivative Distribution Repair (TDDR): A motion correction method for fNIRS



Frank A. Fishburn^{a,*}, Ruth S. Ludlum^b, Chandan J. Vaidya^{b,c}, Andrei V. Medvedev^d

^a Interdisciplinary Program in Neuroscience, Georgetown University, Washington, DC, USA

^b Department of Psychology, Georgetown University, Washington, DC, USA

^c Children's Research Institute, Children's National Health System, Washington, DC, USA

^d Department of Neurology, Georgetown University, Washington, DC, USA

ARTICLE INFO

Keywords:

Functional near-infrared spectroscopy

NIRS

Head motion

Artifact

Denoising

Children

ABSTRACT

Functional near-infrared spectroscopy (fNIRS) is an optical neuroimaging technique of growing interest as a tool for investigation of cortical activity. Due to the on-head placement of optodes, artifacts arising from head motion are relatively less severe than for functional magnetic resonance imaging (fMRI). However, it is still necessary to remove motion artifacts. We present a novel motion correction procedure based on robust regression, which effectively removes baseline shift and spike artifacts without the need for any user-supplied parameters. Our simulations show that this method yields better activation detection performance than 5 other current motion correction methods. In our empirical validation on a working memory task in a sample of children 7–15 years, our method produced stronger and more extensive activation than any of the other methods tested. The new motion correction method enhances the viability of fNIRS as a functional neuroimaging modality for use in populations not amenable to fMRI.

1. Introduction

Functional near-infrared spectroscopy (fNIRS) is an optical neuroimaging technique that has gained interest in recent years as a tool for investigation of cortical activity. Since fNIRS optodes are placed on the scalp of the participant, fNIRS is less sensitive to artifacts arising from head motion than stationary-sensor methods like functional magnetic resonance imaging (fMRI). However, motion artifacts may still be problematic, particularly in populations with more numerous and extreme head movements, such as young children and infants. This is especially true of studies testing for atypical patterns of neural activity in groups that have elevated levels of head motion, such as autism spectrum disorder, attention-deficit/hyperactivity disorder, and epilepsy (Yerys et al., 2009). If left unchecked, artifacts arising from head motion could produce spurious differences, or mask true differences, in patterns of neural activity. Therefore, despite the relative insensitivity of fNIRS to head motion, it is still necessary to apply preprocessing strategies that eliminate artifacts resulting from excessive head movement.

While several motion-correction algorithms have been developed for fNIRS data, they each suffer from some drawback. For instance, the

wavelet-based methods (Chiarelli et al., 2015; Molavi and Dumont, 2012) deals well with motion spikes but exacerbates baseline shift artifacts. Other methods such as the Movement Artifact Reduction Algorithm (MARA) (Scholkmann et al., 2010) and Targeted Principal Components Analysis (tPCA) (Yücel et al., 2013) rely on detection of artifacts, which requires several parameters that must be supplied by the user. These parameters are not necessarily the same across instruments, participants, or brain regions, imposing a large burden on the user. The tPCA and correlation-based signal improvement (CBSI) (Cui et al., 2010) methods make assumptions that don't always hold, regarding the principal component magnitudes and the relationships between oxygenated and deoxygenated hemoglobin, respectively. With these limitations of the existing methods in mind, our objective was to develop a method that efficiently removes both types of motion artifacts, has no tuning parameters, and makes minimal assumptions about the data being corrected.

Here we introduce a novel artifact correction method, Temporal Derivative Distribution Repair (TDDR), that satisfies these criteria of good performance on both types of artifacts, no tuning parameters, and minimal assumptions about the data. The relative performance of the

* Corresponding author.

E-mail address: fishburnf@upmc.edu (F.A. Fishburn).

¹ Present Address: Department of Psychiatry University of Pittsburgh Loeffler Building 121 Meyran Ave, Room 139 Pittsburgh, PA 15213.

proposed method was assessed using simulations in which a task activation pipeline was carried out on data that contained either 1) no motion artifacts, 2) uncorrected motion artifacts, or 3) motion artifacts corrected using either the proposed method or one of 5 comparison methods. This allowed for quantification of the activation performance in reference to a theoretical optimum. To demonstrate the viability of this method on real data, an empirical study was carried out in a pediatric sample, which typically has higher motion than adults. A sample of children 7–15 years old were administered a working memory task, as numerous fMRI and fNIRS studies have consistently shown that working memory is a reliable activator of the prefrontal cortex in children, making it an ideal cognitive process on which to validate our algorithm. Task activation was then evaluated with and without our proposed method, and compared against 5 existing motion correction methods.

2. Material and methods

2.1. Theory

The measured fNIRS signal (x) can be represented as a function of time (t), where the signal is the temporal integral of all prior moment-to-moment fluctuations (y),

$$x_t = \sum_{i=1}^t y_i. \quad (1)$$

These fluctuations contain contributions from the hemodynamic response reflecting cognitive and/or motor processing (the main signal of interest, (h), hemodynamic activity related to systemic (physiological) regulation (e.g., cardiac and respiratory oscillations and Mayer waves; in particular, those signals contribute to the signal recorded from the superficial layers of tissue such as scalp and skull) (p), motion artifact (m), and instrument noise (ϵ),

$$y_t = h_t + p_t + m_t + \epsilon_t. \quad (2)$$

Because the instrument noise is random and broadband, it is hard to remove it in preprocessing by any analytical method, therefore this noise has to be dealt with at the hardware level. Therefore, we assume that ϵ_t is negligible. The systemic physiological signals are relatively narrow-band and quasi-stationary therefore they can be removed by many existing corrective methods, for example, the Independent Component Analysis. Also, because these contributions dominate the signal from the superficial layers, they can be recorded by additional detectors placed in close proximity to the source and subsequently regressed out of the signal of interest (Saager and Berger, 2005). The proposed algorithm aims to specifically reduce motion artifacts (m_t). The approach outlined here relies on three assumptions:

- 1) Non-motion fluctuations (i.e., the derivative of non-motion-related activity) are approximately normally distributed.
- 2) Most fluctuations contain no contribution from motion artifact.
- 3) When present, the fluctuations related to motion artifacts (i.e., their derivative) have much greater magnitude than non-motion fluctuations.

If these assumptions are met, contributions from motion will be large and infrequent, therefore residing in the distant tails of the otherwise normal distribution of signal fluctuations. One possible approach to resolving this issue is to reduce the weights of abnormally large fluctuations,

$$y'_t = w_t y_t, \quad (3)$$

where w is a weighting function with range of $[0,1]$. An obvious choice is to use a fixed threshold based on Chauvenet's criterion,

$$w_t = \begin{cases} 0 & \left| \frac{y_t - \mu}{\sigma} \right| > \alpha, \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

where signal fluctuations with magnitude greater than α standard deviations (σ) from the mean (μ) are set to zero, where α is often some value around 3–5. While it would be possible to set a cutoff for classification of individual fluctuations as containing outliers in this manner, this introduces the burden of threshold selection which may vary between populations, individuals, or brain regions. Rather, we propose an iterative reweighting scheme like that used in robust regression.

The most common type of robust estimator is the M-estimator. These estimators are defined as minimizing the expression,

$$\sum_{i=1}^N \rho(d_i), \quad (5)$$

where d_i are the scaled deviations between the predicted and observed values, and ρ is a weighting function. M-estimators are a generalization of the maximum likelihood estimator, with the least-squares solution being a specific non-robust case: $\rho(d_i) = d_i^2$. In the case of robust M-estimators, the weighting function places lower weight on larger deviations. A common robust estimator is Tukey's biweight function:

$$w_t = \begin{cases} (1 - d_t^2)^2 & |d_t| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where d_t is the scaled residual of each observation. This function places lower weight on observations that are far from the mean, with extreme deviations reduced to zero. For this estimator and many others, it is necessary to employ an iterative reweighting scheme in which the parameters, residuals, and observation weights are all iteratively recomputed until the parameter estimate converges on a value. Using this iterative reweighting approach, it is possible to produce a set of weights which shrinks the excessively large fluctuations (e.g., $|d_t| \geq 1$) most commonly associated with head motion.

2.2. Description of the algorithm

2.2.1. Compute the temporal derivative of the measured signal

$$y_t = x_t - x_{t-1} \quad (7)$$

2.2.2. Initialize vector of observation weights (w)

$$w_t = 1 \quad (8)$$

2.2.3. Iteratively estimate the robust observation weights

- a Estimate the weighted mean of the fluctuations.

$$\mu = \frac{1}{\sum(w)} \sum(w_t y_t) \quad (9)$$

- b Compute the absolute residuals of the estimate

$$r_t = |y_t - \mu| \quad (10)$$

- c Compute the robust estimate of the standard deviation of the residuals. This is achieved using the median absolute residual scaled by the appropriate constant for the normal distribution.

$$\sigma = 1.4826 \cdot \text{median}(r) \quad (11)$$

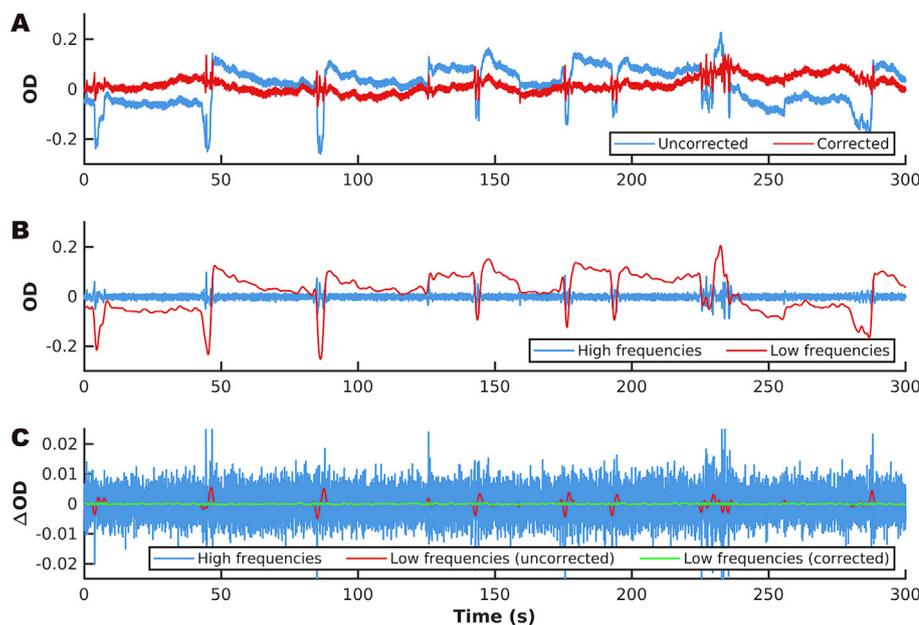


Fig. 1. Demonstration of TDDR on real data.

Demonstration of the effect of TDDR on a fNIRS signal collected from the prefrontal cortex of a child. The signal is shown before and after application of TDDR (A). The high and low frequency components of the uncorrected signal are shown in (B). The temporal derivative is shown for the high and low frequency components of the uncorrected signal, as well as the corrected low frequency derivative (C).

- d Compute the scaled deviations of each observation using the standard deviation of the residuals and the tuning constant that achieves 95% efficiency on normally distributed data.

$$d_i = \frac{r_i}{4.685\sigma} \quad (12)$$

- e Compute new observation weights using Tukey's biweight function.

$$w_i = \begin{cases} (1 - d_i^2)^2 & d_i < 1 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

- f Repeat until the value of μ converges.

2.2.4. Apply the resulting robust weights to the centered temporal derivative to produce the corrected derivative

$$y'_i = w_i(y_i - \mu) \quad (14)$$

2.2.5. Integrate the corrected temporal derivative to yield the corrected signal

$$x'_i = \sum_{t=1}^N y'_t \quad (15)$$

2.3. Practical considerations

A demonstration of the effect of this method on real data is shown in Fig. 1A. As TDDR makes use of the temporal derivative of the signal, it is fundamentally dependent on the sampling rate of that signal. This is because the sampling period is directly related to the magnitude of the values in the temporal derivative. In addition, this algorithm relies on estimation of the variance of the temporal derivative to compute robust weights. For this reason, the presence of high frequency (HF) components (whether due to instrument/measurement noise or of physiological origin) inflates the estimated variance, thus reducing its efficacy. To address these issues, our implementation first checks whether the sampling rate is high enough that excessive HF components would be present (>1 Hz). If so, the signal is split into low and high frequency components by applying a low-pass (LP) filter with a cutoff of 0.5 Hz and retaining the filter residual (Fig. 1B). The TDDR algorithm is then applied to the low frequency component of the signal, which is then added back to the

uncorrected HF component. As can be seen in Fig. 1C, the variance of the temporal derivative of HF components is far greater than that of even large motion artifacts. As we show in the simulation described in section 2.4.2, this LP filtering procedure effectively buffers the core algorithm from issues of excessively high sampling rate and the presence of high frequency components.

We also note that while this algorithm is iterative in nature, it is still extremely fast. This is because each iteration involves only computing the weighted mean of the signal derivative, then computing the new set of weights. All of the operations involved are relatively computationally straightforward. To illustrate this point, a simulated 10-min scan with 32 channels sampled at 20 Hz takes around 0.13 s to apply TDDR on our 2.40 GHz workstation in single-threaded mode. In our simulation and experimental data, the weights generally converge in less than 20 iterations. Source code implementing the TDDR algorithm can be accessed at <https://github.com/frankfishburn/TDDR>.

2.4. Simulation validation

2.4.1. Dataset generation

Simulated fNIRS signals were produced using a randomly-generated autoregressive model with an order of 10 samples. Signals were generated at 20 Hz for 32 channels and a spatial covariance of 0.33 was imposed. The task structure consisted of a 10-min scan subdivided into blocks of a fixed duration randomly chosen from 1 to 60 s, separated by inter-block intervals of 10–60 s. The simulated task response was generated by convolving the task boxcar with the canonical hemodynamic response function. The task response was then added to the autoregressive noise of half of the channels, with an amplitude of 0.5–25 times the standard deviation of the noise. Systemic physiological oscillations were added to the resulting hemoglobin signals by generating sine waves having frequencies drawn from a normal distribution for cardiac ($\mu = 1$, $\sigma = 0.1$), respiration ($\mu = 0.25$, $\sigma = 0.025$), and Mayer wave ($\mu = 0.1$, $\sigma = 0.01$) noises. Cumulative phase drifts were added to each of these signals using a normal distribution with standard deviation of 0.1 rad/s. These hemoglobin concentration signals were then converted back to raw intensity. Spike and baseline shift motion artifacts were then generated with a frequency of 2 spikes and 2 baseline shifts per minute. Spikes were modeled as exponential growth curves ($x(t) = b^{t/\tau}$), with a positive growth factor (b) randomly selected from a normal distribution with standard deviation of 25 times the

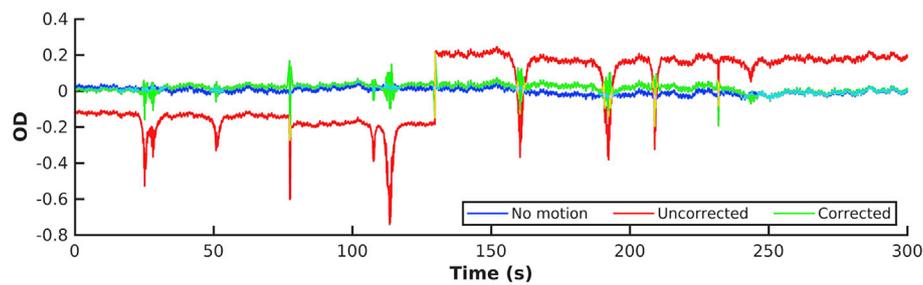


Fig. 2. Demonstration of TDDR on simulated data.

The optical density is shown for a motion-free signal (blue), the same signal with uncorrected artifacts (red), and the same signal after artifacts are corrected using TDDR (green). Note that the corrected signal is almost identical to the motion-free signal.

standard deviation of the input signal, and time constant (τ) selected to yield a final spike duration of 0.1–10.0 s. Baseline shifts were modeled as random positive or negative shifts with amplitude selected the same way as for spikes. Ten-thousand files were generated in this manner. Simulations were carried out using the NIRS Brain AnalyzIR toolbox (Santosa et al., 2018). A sample simulated time course without artifacts, with uncorrected artifacts, and with artifacts corrected by TDDR is shown in Fig. 2.

2.4.2. Analysis #1: the effect of sampling rate and filtering

2.4.2.1. Preprocessing. Raw intensity signals were converted to optical density. The signals were then resampled to 1–20 Hz. The resampled optical density signals were then either 1) submitted to direct application of TDDR, or 2) had only the low-frequency portion of the signal corrected using TDDR, as described in section 2.3. The corrected optical density signals were then converted to hemoglobin concentration using the modified Beer-Lambert relationship (Cope and Delpy, 1988). To avoid sampling rate directly altering the performance of the activation detection, all signals were then resampled to 1 Hz.

2.4.2.2. Activation. Activation was assessed using a GLM. To account for low frequency drift in the signals, discrete cosine transform matrix terms were included in the model, with a maximum frequency of 1/128 Hz. Performance of activation detection was assessed using receiver operating characteristic (ROC) curves. To generate bounds on the ROC curves, the results of the 10,000 iterations (32,000 channels total) were partitioned into 100 bins and the means and standard deviations calculated on the curve. For a global metric of performance, the area under the curve (AUC) was computed by trapezoidal integration. To validate our choice of the number of iterations, the results of each test (e.g., each sample rate and prefiltering) were randomly separated in half 100 times and the mean-squared error (MSE) in AUC calculated between the two halves $MSE = \frac{1}{100} \sum_{i=1}^{100} (|AUC_1 - AUC_2|^2)$. The MSE of the AUC for each test was less than 10^{-5} .

2.4.3. Analysis #2: method comparison

2.4.3.1. Preprocessing. Simulated fNIRS signals were corrected using TDDR and 5 different comparison methods detailed below, as implemented in Homer2 (Huppert et al., 2009). Signals were then converted to hemoglobin concentration according to the modified Beer-Lambert law and resampled to 1 Hz.

2.4.3.2. Comparison methods. Correlation-Based Signal Improvement (CBSI) – CBSI (Cui et al., 2010) relies on the assumption that oxygenated and deoxygenated hemoglobin species are strongly negatively correlated, while motion artifacts are positively correlated between these measurements. Essentially, the deoxygenated signal is scaled to the same variance

of the oxygenated signal, then subtracted from the oxygenated signal. The rationale being that positively-correlated features of the signal (i.e., motion) will be negated by the subtraction, while negatively-correlated features (i.e., hemodynamics) will sum. Notably, this method requires no user-specified thresholds. The function *hmrMotionCorrectCbsi* was used to carry this out.

Movement Artifact Reduction Algorithm (MARA) – MARA (Scholkmann et al., 2010) is a correction method in which the artifacts are detected by evaluating short time windows for changes in amplitude or standard deviation exceeding a user-specified threshold. The time windows that are found to contain artifacts are widened to include a short time preceding and following, then the artifact is approximated with a smoothing spline, which is then subtracted out to give the motion-corrected signal. The function *hmrMotionArtifactByChannel* was used to detect motion time periods, while *hmrMotionCorrectSpline* was used to apply the correction. The parameters from Cooper et al. (2012) were used: $tMotion = 0.5$, $tMask = 2$, $STDEVthresh = 20$, $AMPthresh = 0.5$, $pSpline = 0.99$.

Targeted Principal Component Analysis (tPCA) – The first widely-used method for removing artifacts from fNIRS signals was PCA (Zhang et al., 2005). In this method, the set of fNIRS channels are submitted to PCA (Pearson, 1901), a decomposition in which the set of input signals are transformed into an orthogonal set of principal components. Those components accounting for the greatest variance are then discarded and the signal is reconstructed from the presumed non-artifactual components. An extension of this method known as targeted principal components analysis was later created (Yücel et al., 2013) to apply the original PCA method only on short segments where artifacts have been detected. This was carried out using the *hmrMotionCorrectPCArecuse* function. The motion detection parameters used were the same as those used in the MARA method, with the additional parameters of $nSV = 0.97$ and $maxIter = 3$ as used in the original paper.

Kurtosis Wavelet (kWavelet) – Wavelet filtering (Molavi and Dumont, 2012) is a correction method wherein a discrete wavelet transform is taken on the input signal, and the wavelet coefficients with magnitude exceeding a threshold are rejected before reconstructing the signal. A later extension of this work uses the kurtosis of the wavelet coefficients to automatically set the threshold (Chiarelli et al., 2015). The function *hmrMotionCorrectKurtosisWavelet* was used with the kurtosis threshold parameter set to 3.3 as recommended in the original paper.

Spline Savitzky-Golay (Spline-SG) – The Spline-SG method is a multi-step correction procedure that aims to extend MARA by reducing the number of parameters supplied by the user and adding a separate process for removing fast spikes (Jahani et al., 2018). In this method, a Sobel filter is first used to extract the gradient of the signal, from which time periods containing motion are identified as those exceeding 1.5 times the inter-quartile interval. The method then attempts to identify baseline shifts and slow spikes and correct those time periods using the spline approach employed by MARA. This is followed up with a Savitzky-Golay smoothing filter to remove fast spikes. The function *hmrMotionCorrectSplineSG* was used with parameters: $p = .99$ and $FrameSize_sec = 10$.

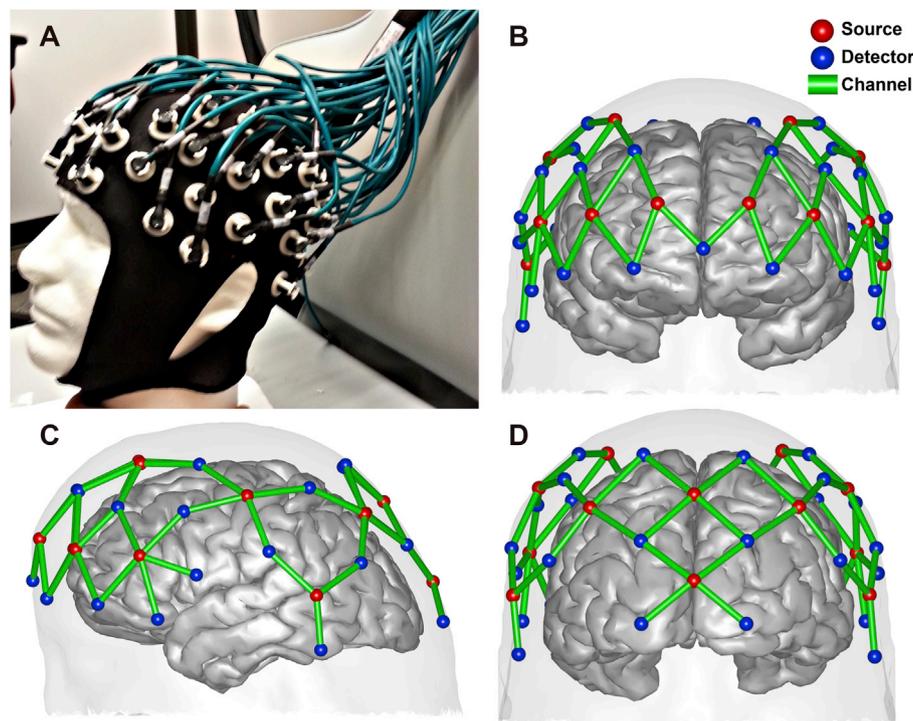


Fig. 3. Probe configuration.

The optode layout is shown on a head phantom (A) and the anterior (B), left hemisphere (C), and posterior (D) views of the probe registered to the Colin27 brain template.

2.4.3.3. Activation. The activation performance of each method was then assessed in the same way as described in section 2.4.2.2. The MSE of the AUC for each test was less than 10^{-5} .

2.5. Experimental validation

2.5.1. Participants

Twenty-three typically-developing participants aged 7–15 years (Mean = 11.72, SD = 2.13), participated in the study after complying with consenting guidelines of the Georgetown University Institutional Review Board. Participants were recruited from the Washington, DC area through advertisements at public venues and pediatrician offices.

Exclusion criteria included: (1) Full-scale IQ below 80 as measured by the Wechsler Intelligence Scale for Children (WISC-IV) or Wechsler Abbreviated Scale of Intelligence (WASI) (2) Neurological diagnosis (e.g., epilepsy); based on parent report; (3) Psychiatric diagnosis based on Child and Adolescent Symptom Inventory-4R (Gadow and Sprafkin, 2005).

2.5.2. Task procedure

Experimental sessions consisted of a 7-min n-back task. Participants were presented with a series of single consonant letters and instructed to press a button with their dominant hand when the presented letter was the same as the one presented n letters ago. Subjects were tested on three blocks of each of the three load conditions defined by $n=0, 1,$ and 2 trials, 0-back, 1-back, and 2-back blocks, respectively. The 0-back condition required the participant to only press whenever they see an 'X', thus not engaging working memory beyond the minimum required to maintain task instructions. The load condition order was pseudorandomized using a modified Latin square. Each block consisted of 9 trials, each lasting 3000 ms, with the letter exposed for 500 ms followed by a lag of 2500 ms. Each 27-s block was followed by a 14-s interval of fixation to allow the hemodynamic response to return to baseline, followed by a 5-s warning that the next block was about to begin. Subjects practiced two sample blocks of each task condition prior to the scanning session.

2.5.3. Imaging procedure

Optical signals were recorded on a two-wavelength (690 and 830 nm) continuous-wave CW6 (TechEn, Inc., Milford, MA) imaging system. Data were collected from detectors at a sampling rate of 50 Hz. The 56 optical channels were made up of 16 sources and 29 detectors that were mounted to elastic fabric caps (EASYCAP, Brain Products GmbH, Germany). The caps were placed on a model head phantom alongside an EEG system to determine optode coordinates in standard 10–20 space. The AtlasViewer program (Aasted et al., 2015) was then used to register optode and channel locations to the Colin27 brain template (Fig. 3).

2.5.4. Analysis

2.5.4.1. Preprocessing. Preprocessing and activation analysis were carried out using the NIRS Brain AnalyzIR toolbox (Santosa et al., 2018). Each motion correction procedure described in section 2.4.3.2 was applied independently. Signals were then resampled to 5 Hz.

2.5.4.2. Activation. Task activation was quantified by creating a working memory load-dependent regressor. This was done as described previously (Fishburn et al., 2014), by creating a weighted task boxcar function in which each condition (0-, 1-, and 2-back) had an amplitude corresponding to their relative load (1, 2, and 3, respectively). The weighted boxcar was then convolved with the canonical hemodynamic response function (HRF) and submitted to a general linear model. In order to account for variations in the hemodynamic response function, temporal and dispersion derivatives were included in the model. Low-frequency drifts were accounted for by including discrete cosine transform terms in the model, with maximum frequency of $1/128$ Hz. Group-level analysis was carried out by performing a regularized surface-based image reconstruction described by Abdelnour and Huppert (2011) using the sensitivity matrix generated by AtlasViewer during probe registration.

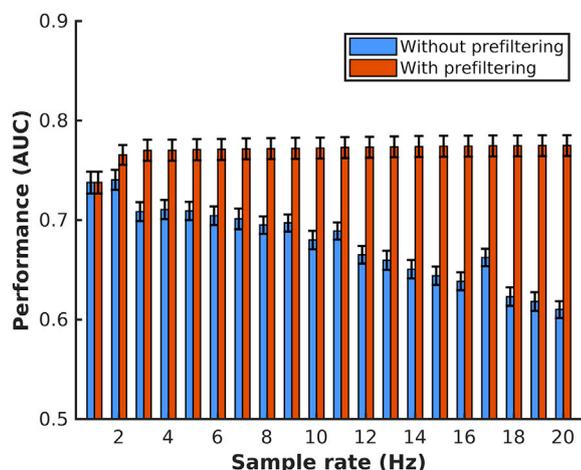


Fig. 4. Effect of sample rate and low-pass filtering on TDDR performance. Simulation results showing TDDR performance (AUC) as a function of sampling rate of the input data, both without prefiltering (blue) and with prefiltering (orange). Higher sampling rates were detrimental to performance, but only when prefiltering was not performed. Error bars reflect the standard deviation.

3. Results

3.1. Simulation #1: the effect of sampling rate and low-pass filtering

The simulation assessing the effects of sample rate and LP filtering on TDDR performance found that when using the core TDDR algorithm without prefiltering, performance was highly dependent on the sampling rate of the input data, with 2 Hz yielding the best results (Fig. 4). However, the addition of the low-pass prefiltering step fully recovered the performance for all sample rates higher than 1 Hz.

3.2. Simulation #2: method comparison

The ROC curves assessing activation detection performance are shown in Fig. 5A. The mean AUC value closest to that of motion-free data (0.869) was achieved with TDDR (0.775), which was greater than CBSI (0.733), MARA (0.563), tPCA (0.591), kWavelet (0.513), spline-SG (0.652), and uncorrected (0.516). The mean and standard deviation of AUC values for each method are shown in Fig. 5B.

3.3. Experiment

Activation maps produced by each method on the oxygenated hemoglobin data are shown in Fig. 6. Of the methods tested, TDDR produced the greatest maximum activation t-statistic (4.88), as compared to CBSI (3.02), MARA (2.96), tPCA (3.79), kWavelet (3.96), spline-SG (3.68), and uncorrected (3.67). In addition, TDDR produced the greatest number of mesh vertices with positive significant ($p < .05$) values (2,399), compared to CBSI (903), MARA (924), tPCA (891), kWavelet (935), spline-SG (1,153), and uncorrected (1,560). Results using a channel-based approach to activation analysis are provided in Supplementary Fig. 1. Analysis of deoxygenated hemoglobin using both the image reconstruction approach and a channel-based approach can be found in Supplementary Fig. 2 and Supplementary Fig. 3, respectively.

4. Discussion

4.1. General discussion

We have developed a novel algorithm for the removal of artifacts resulting from head motion in fNIRS data. By leveraging an iterative-reweighting approach mirroring that used in robust regression, this method efficiently removes artifacts without the need for any user-supplied parameters. Our simulations show that this method yields activation-detection performance better than other methods and nearly as good as motion-free data. In addition, our testing on a dataset from a working memory task collected from a wide age range of typically-developing children showed that TDDR produced stronger and more extensive activation than any other method tested.

The CBSI method strongly outperformed all methods aside from TDDR in the simulation, though performance was poor in the experimental data. One possibility is that the CBSI assumption of strong anti-correlation between oxygenated and deoxygenated species is maintained more strongly in the simulation than in real data. This interpretation is supported by previous studies generally finding CBSI performance similar to or less than other methods in real data (Brigadoi et al., 2014; Hu et al., 2015; Jahani et al., 2018). Future work should attempt to better characterize the departure of real data from the assumptions of CBSI to improve the validity of simulated datasets and potentially extend the CBSI method itself.

While the MARA method performed better than no correction on the simulated data, it did not excel in either the simulation nor experimental data. While some previous studies have shown excellent performance of MARA (Cooper et al., 2012), most have found its performance to be

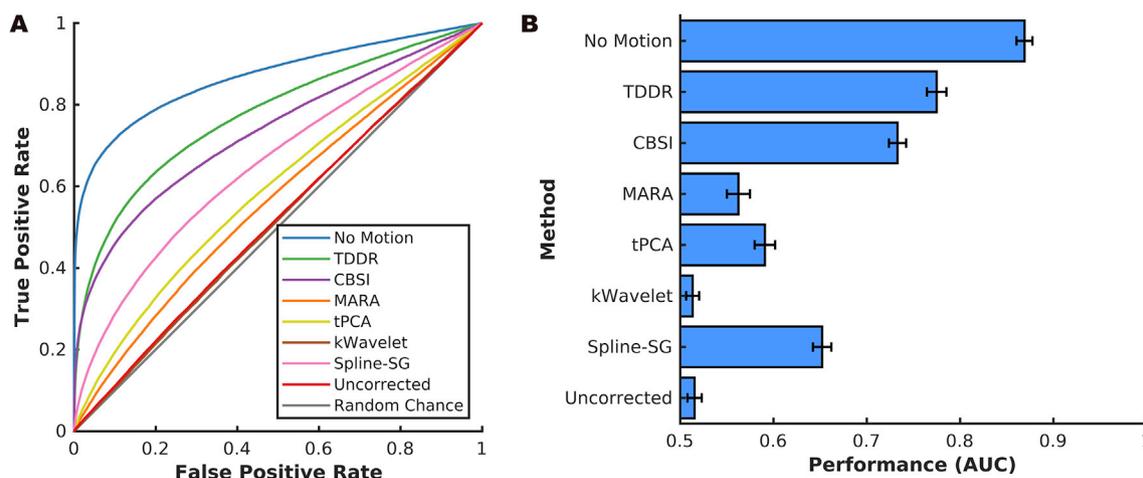


Fig. 5. Comparison of TDDR performance with other correction methods.

The performance of activation detection is shown via ROC curves for simulation data with no motion, as well as motion-contaminated data corrected with each method (A). The area under each of the ROC curves is shown in (B). Error bars reflect the standard deviation.

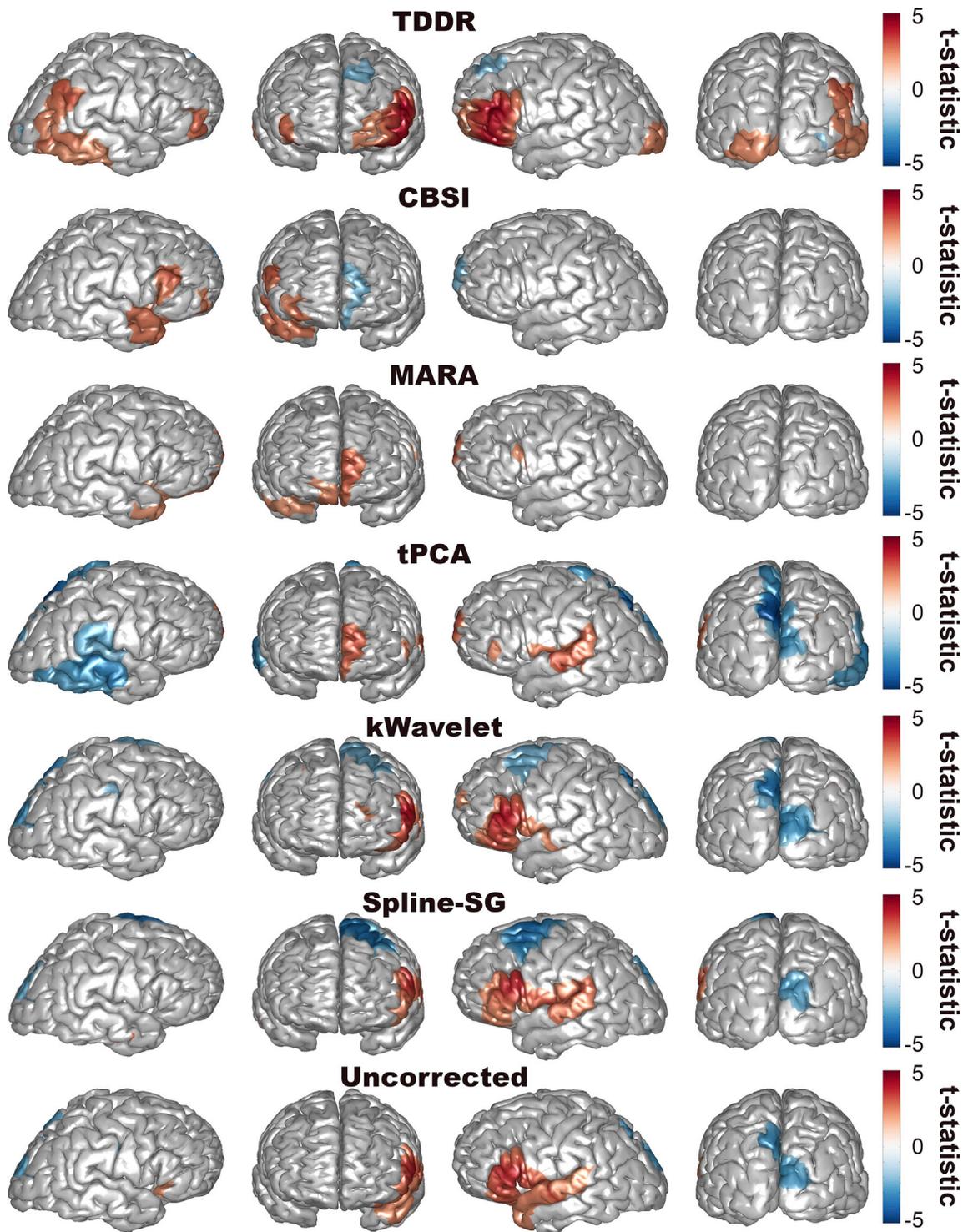


Fig. 6. Comparison of activation results using each method.

Results of the activation analysis on the oxygenated hemoglobin data are shown with each motion correction method. Images are thresholded at $p < .05$.

sub-par compared to other methods (Brigadoi et al., 2014; Chiarelli et al., 2015; Jahani et al., 2018; Yücel et al., 2013). One known challenge facing MARA (and tPCA) is the need to set several parameters for detecting periods of motion contamination. This introduces a large burden on the user to select these values through trial-and-error and makes replication of studies much more difficult. Upon visual inspection, the parameters used in this study yielded reasonable motion-detection performance, however that doesn't necessarily mean that these parameters were at their optimal value in terms of activation performance.

Another possibility is that the issue is more central to the algorithm. MARA uses the spline method of correcting motion, which essentially models a smoothed spline of the data during the affected time period and subtracts that from the signal. From this description, it would seem that hemodynamic fluctuations during this period would likely be modeled as part of the spline, as they are also slow. Thus, if motion is detected at the start of a task block, the task-related increase of hemoglobin concentration would potentially be removed along with the motion artifact.

The tPCA method achieved results slightly better than MARA in both

the simulation and experimental data. This is in line with previous work showing somewhat better performance of tPCA than MARA (Chiarelli et al., 2015; Jahani et al., 2018; Yücel et al., 2013), though the absolute performance of both methods was better in those studies. The poor performance of both tPCA and MARA may very well be due to their shared dependence on detection of motion using a set of user-selected parameters. Another potential issue with this method is that it blindly removes the largest principal components accounting for a fixed proportion of the variance in the data. This again leaves the possibility that some of the hemodynamic response is inadvertently being removed during correction.

The kWavelet method suffered from abysmal performance in the simulation but reasonably good performance on the experimental data. A known limitation of wavelet-based methods is their poor performance on low frequency motion artifacts, such as baseline shifts and slow spikes. The result of using wavelet-based methods on these types of artifacts is generally a temporal smearing or smoothing of the artifact, which is detrimental to later analyses. It could be that the particular dataset used for the experimental validation had fewer or smaller baseline shifts than those in the simulated data.

The spline-SG method yielded strong results in the simulation and in the pattern and extent of activation in the experimental data. However, it failed to achieve peak activation as strong as other methods. One possibility is that the data are being over-corrected by this method, which could potentially be resolved by a modification of thresholds. However, this method is subject to the same previous criticism of spline methods in general, namely that subtraction of the smoothing spline may also remove some of the hemodynamic response.

4.2. Limitations

While the proposed TDDR method is effective at removing motion-related artifacts such as spikes and baseline shifts, it comes with some limitations. First, the presence of high-frequency components in the signal dramatically reduces the effectiveness of the procedure. High-frequency oscillations inflate the variance of the temporal derivative, biasing the distribution estimate such that artifacts are much closer to the mean and therefore subject to smaller correction. Fortunately, this issue is easily remedied by applying a low-pass filter to the data prior to application of TDDR. If desired, the high-frequency information can then be reintroduced to the resulting motion-corrected low-frequency signal. One downside of this approach is that the high frequency component of motion artifacts is not corrected. However, it is also true that these motion remnants can be removed via low-pass filtering.

Another limitation is that this procedure is most effective for artifacts that have a large temporal derivative (e.g., high amplitude and/or short duration). This makes it very effective for typical spikes and baseline shifts, however it is possible that in some cases low amplitude or very slow artifacts could remain. It is also true that this is the case for any motion correction method, since smaller and slower artifacts are more difficult to isolate from the hemodynamic signal in general.

4.3. Future directions

The growing number of motion correction methods for fNIRS have different strengths and weaknesses. For this reason, future studies should investigate whether performance can be improved by combining multiple methods. Previous research has shown that high performance can be achieved using both moving average and wavelet motion correction, as compared to using either alone (Hu et al., 2015). Future work should investigate whether TDDR performance can be enhanced by pre- or post-processing with another correction method, and whether this performance is greater than the possible combinations of other correction methods. Since TDDR only corrects the low-frequency portion of the signal, the development of a post-processing step for the high-frequency component could be especially beneficial for applications where simply

removing the high frequencies is undesirable. Additionally, there are strategies for dealing with motion that are not strictly motion correction methods (i.e., methods that do not strictly convert a motion-contaminated signal into an uncontaminated signal). These include affixing optodes to the scalp using clinical adhesive (Yücel et al., 2014), as well as motion-robust algorithms for estimating activation (Barker et al., 2013) and functional connectivity (Santosa et al., 2017). Hopefully, the fNIRS field will converge upon a standard set of validated and generalizable strategies for dealing with motion in the coming years.

5. Conclusions

We have devised a novel motion correction method for fNIRS. As the method is based on robust estimators, it has no tuning parameters and is effective at removing both spikes and baseline shifts. Validation using simulations and an empirical dataset show performance superior to that of other current correction methods. This method increases the viability of application of fNIRS for functional neural imaging in populations such as infants and young children and those with intellectual disability in whom restricting head motion is difficult due to compliance or communication limitations.

Acknowledgments

This work was supported by National Institute of Mental Health (R01 MH084961 to Chandan Vaidya), National Institute of General Medical Sciences (R21 GM103526 to Andrei Medvedev), Autism Speaks (Dennis Weatherstone Predoctoral Fellowship to Frank Fishburn), and the National Institute of Mental Health (F31 Predoctoral National Research Service Award to Frank Fishburn).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.neuroimage.2018.09.025>.

References

- Aasted, C.M., Yücel, M.A., Cooper, R.J., Dubb, J., Tsuzuki, D., Becerra, L., Petkov, M.P., Borsook, D., Dan, I., Boas, D.A., 2015. Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. *NeuroPhotonics* 2, 020801. <https://doi.org/10.1117/1.NPh.2.2.020801>.
- Abdelnour, F., Huppert, T., 2011. A random-effects model for group-level analysis of diffuse optical brain imaging. *Biomed. Optic Express* 2, 1–25. <https://doi.org/10.1364/BOE.2.000001>.
- Barker, J.W., Aarabi, A., Huppert, T.J., 2013. Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomed. Optic Express* 4, 1366–1379. <https://doi.org/10.1364/BOE.4.001366>.
- Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., Gagnon, L., Boas, D.A., Cooper, R.J., 2014. Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data. In: *NeuroImage, Celebrating 20 Years of Functional Near Infrared Spectroscopy (fNIRS)*, 85, pp. 181–191. Part 1. <https://doi.org/10.1016/j.neuroimage.2013.04.082>.
- Chiarelli, A.M., Maclin, E.L., Fabiani, M., Gratton, G., 2015. A kurtosis-based wavelet algorithm for motion artifact correction of fNIRS data. *NeuroImage* 112, 128–137. <https://doi.org/10.1016/j.neuroimage.2015.02.057>.
- Cooper, R., Selb, J., Gagnon, L., Phillip, D., Schytz, H.W., Iversen, H.K., Ashina, M., Boas, D.A., 2012. A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. In: *Brain Imaging Methods*, 6, p. 147. <https://doi.org/10.3389/fnins.2012.00147>.
- Cope, M., Delpy, D., 1988. System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Med. Biol. Eng. Comput.* 26, 289–294. <https://doi.org/10.1007/BF02447083>.
- Cui, X., Bray, S., Reiss, A.L., 2010. Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics. *NeuroImage* 49, 3039–3046. <https://doi.org/10.1016/j.neuroimage.2009.11.050>.
- Fishburn, F.A., Norr, M.E., Medvedev, A.V., Vaidya, C.J., 2014. Sensitivity of fNIRS to cognitive state and load. *Front. Hum. Neurosci.* 8. <https://doi.org/10.3389/fnhum.2014.00076>.
- Gadow, K.D., Sprafkin, J., 2005. *Child and Adolescent Symptom Inventory-4R*. Checkmate Plus, Stony Brook, NY.

- Hu, X.-S., Arredondo, M.M., Gomba, M., Confer, N., DaSilva, A.F., Johnson, T.D., Shalinsky, M., Kovelman, I., 2015. Comparison of motion correction techniques applied to functional near-infrared spectroscopy data from children. *J. Biomed. Optic.* 20, 126003.
- 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. Optic.* 48, D280–D298.
- Jahani, S., Setarehdan, S.K., Boas, D.A., Yücel, M.A., 2018. Motion artifact detection and correction in functional near-infrared spectroscopy: a new hybrid method based on spline interpolation method and Savitzky-Golay filtering. *Neurophotonics* 5, 015003–015003.
- Molavi, B., Dumont, G.A., 2012. Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiol. Meas.* 33, 259. <https://doi.org/10.1088/0967-3334/33/2/259>.
- Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philos. Mag. A* 2, 559–572. <https://doi.org/10.1080/14786440109462720>.
- Saager, R.B., Berger, A.J., 2005. Direct characterization and removal of interfering absorption trends in two-layer turbid media. *JOSA A* 22, 1874–1882. <https://doi.org/10.1364/JOSAA.22.001874>.
- Santosa, H., Aarabi, A., Perlman, S.B., Huppert, T., 2017. Characterization and correction of the false-discovery rates in resting state connectivity using functional near-infrared spectroscopy. *J. Biomed. Optic.* 22, 055002. <https://doi.org/10.1117/1.JBO.22.5.055002>.
- Santosa, H., Zhai, X., Fishburn, F., Huppert, T., 2018. The NIRS brain AnalyzIR toolbox. *Algorithms* 11, 73. <https://doi.org/10.3390/a11050073>.
- Scholkmann, F., Spichtig, S., Muehleemann, T., Wolf, M., 2010. How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation. *Physiol. Meas.* 31, 649. <https://doi.org/10.1088/0967-3334/31/5/004>.
- Yerys, B.E., Jankowski, K.F., Shook, D., Rosenberger, L.R., Barnes, K.A., Berl, M.M., Ritzl, E.K., VanMeter, J., Vaidya, C.J., Gaillard, W.D., 2009. The fMRI success rate of children and adolescents: typical development, epilepsy, attention deficit/hyperactivity disorder, and autism spectrum disorders. *Hum. Brain Mapp.* 30, 3426–3435.
- Yücel, M.A., Selb, J., Boas, D.A., Cash, S.S., Cooper, R.J., 2014. Reducing motion artifacts for long-term clinical NIRS monitoring using collodion-fixed prism-based optical fibers. *NeuroImage, Celebrating 20 Years of Functional Near Infrared Spectroscopy (fNIRS)* 85, 192–201. <https://doi.org/10.1016/j.neuroimage.2013.06.054>.
- Yücel, M.A., Selb, J., Cooper, R.J., Boas, D.A., 2013. Targeted principle component analysis: a new motion artifact correction approach for near-infrared spectroscopy. *J. Innov. Opt. Health Sci.* 07, 1350066. <https://doi.org/10.1142/S1793545813500661>.
- Zhang, Y., Brooks, D.H., Franceschini, M.A., Boas, D.A., 2005. Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *J. Biomed. Optic.* 10, 011014–01101411. <https://doi.org/10.1117/1.1852552>.