Journal of Biomedical Optics

BiomedicalOptics.SPIEDigitalLibrary.org

Comparison of motion correction techniques applied to functional near-infrared spectroscopy data from children

Xiao-Su Hu
Maria M. Arredondo
Megan Gomba
Nicole Confer
Alexandre F. DaSilva
Timothy D. Johnson
Mark Shalinsky
Ioulia Kovelman

SPIE.

Comparison of motion correction techniques applied to functional near-infrared spectroscopy data from children

Xiao-Su Hu,[a] Maria M. Arredondo,[b] Megan Gomba,[a] Nicole Confer,[a] Alexandre F. DaSilva,[a,c] Timothy D. Johnson,[d] Mark Shalinsky,[a] and Ioulia Kovelman[a,b,*]
[a]University of Michigan, Center for Human Growth and Development, 300 North Ingalls Street, Ann Arbor, MI 48104
[b]University of Michigan, Department of Psychology, 530 Church Street, Ann Arbor, Michigan 48109
[c]University of Michigan, Biologic & Materials Sciences Department, School of Dentistry, Headache & Orofacial Pain Effort Lab, 1011 North University Avenue, Ann Arbor, Michigan 48109
[d]University of Michigan, Department of Biostatistics, School of Public Health, 1415 Washington Heights, Ann Arbor, Michigan 48109

Abstract. Motion artifacts are the most significant sources of noise in the context of pediatric brain imaging designs and data analyses, especially in applications of functional near-infrared spectroscopy (fNIRS), in which it can completely affect the quality of the data acquired. Different methods have been developed to correct motion artifacts in fNIRS data, but the relative effectiveness of these methods for data from child and infant subjects (which is often found to be significantly noisier than adult data) remains largely unexplored. The issue is further complicated by the heterogeneity of fNIRS data artifacts. We compared the efficacy of the six most prevalent motion artifact correction techniques with fNIRS data acquired from children participating in a language acquisition task, including wavelet, spline interpolation, principal component analysis, moving average (MA), correlation-based signal improvement, and combination of wavelet and MA. The evaluation of five predefined metrics suggests that the MA and wavelet methods yield the best outcomes. These findings elucidate the varied nature of fNIRS data artifacts and the efficacy of artifact correction methods with pediatric populations, as well as help inform both the theory and practice of optical brain imaging analysis. *© 2015 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JBO.20.12.126003]

Keywords: functional near-infrared spectroscopy; motion artifact correction; child brain imaging.

Paper 150717R received Oct. 26, 2015; accepted for publication Nov. 5, 2015; published online Dec. 11, 2015; corrected Apr. 28, 2020.

1 Introduction

Functional near-infrared spectroscopy (fNIRS) is an emerging optical brain imaging technique that is becoming an increasingly popular method for pediatric brain imaging.[1] fNIRS measures the hemodynamic changes that effectively reflect brain activities occurring while people perform a wide range of mental tasks;[2–6] it can provide both topographic[3,4,7] and tomographic brain images.[6,8] Specifically, fNIRS monitors concentration variations of oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) in the blood of the brain, through the skull, with absorption changes of near-infrared light at wavelengths between 700 and 1000 nm.[9]

Similar to electroencephalography (EEG), fNIRS allows for noninvasive observations of awake infants and children, but unlike EEG, fNIRS provides complimentary hemodynamic response data. Compared with functional magnetic resonance imaging, fNIRS devices are usually portable, silent, and use near-infrared light to detect brain activity.[2] These aspects of fNIRS technology make it a useful tool for researchers interested in child[10–20] and infant[21–27] development. Therefore, the use of this method with pediatric populations has grown threefold in the past decade.

Motion artifacts caused by participants' head, jaw, eyebrow, or body movement can distort the data analysis.[28,29] A major impediment of interpreting fNIRS brain imaging data of child participants is that children's data typically have more motion artifacts than the fNIRS data of adults. The field has put forth a variety of methods for addressing the fNIRS motion artifact problem in general.[28–30] The most direct method is trial rejection (for event-related design) or block rejection (for block design), which excludes all the data sections that contain motion artifacts from further analysis.[31] Selb et al. compared several motion correction techniques on synthetic data and found trial rejection was the best method for preserving the original shape of the data. However, in the context of pediatric brain imaging, the major problem with this method is that the entire data samples tend to be relatively small, as children and infants have a short attention span and experimental designs can be as short as a few minutes. Thus, researchers have attempted methods that allow them to retain more of the collected data. These methods can be divided into two categories: the first category includes hardware-based methods (which need additional sensor or special optode setup), including short separation channel-based independent component analysis,[32] using specially designed collodion-fixed fibers,[31] accelerometer-based regression, or rescaling.[33,34] The difficulty with these methods is that additional hardware may further complicate and increase participant setup time, which is always a problem with young participants. The second category includes software-based methods, such as principal

*Address all correspondence to: Ioulia Kovelman, E-mail: kovelman@umich.edu

1083-3668/2015/$25.00 © 2015 SPIE

component analysis (PCA),[35] spline interpolation,[36] wavelet,[37] correlation-based signal improvement (CBSI),[38] and moving average (MA), plus Kalman filter and recursive least-square based methods for real-time preprocessing.[34,39]

There have been a few prior studies that have attempted to explore the relative efficacy of different artifact correction approaches using simulated adult fNIRS data. These studies concluded that the wavelet method was best suited for the simulated adult fNIRS data.[28,29] Yet, it remains largely unknown whether the same is true of real pediatric fNIRS data from children subjects. Thus, in the present study, we aimed to improve the field's methodology for child fNIRS brain imaging by directly comparing the efficacy of six software-based offline motion correction methods on real fNIRS data acquired from children (ages 6 to 12) completing a language task.

2 Methods

2.1 *Participants*

Twelve children (eight females, age range: 6.8 to 12.6 years, mean age = 9.9 years, SD = 1.75) took part in the study. Participants were excluded from the imaging analysis if they did not complete all the necessary tasks or did not fit this study's inclusion criteria (see fNIRS analysis section below for more details). Since this paper discusses signal qualities, we also revealed the hair and skin color of all participants (Table 1, for more info on participants see Arredondo et al.[40]) The University of Michigan Institutional Medical Review Boards reviewed and approved the study. Parents provided consent and children provided assent to participate in the study.

2.2 *Materials and Procedure*

The children completed an auditory grammatical judgment language task during the fNIRS data collection. The task was based on the Test of Early Grammatical Impairment,[41] and it included three conditions: (1) control sentences with developmentally atypical errors (e.g., "Yesterday he bake a cake"), (2) experimental sentences with developmentally typical (optional infinitive) errors (e.g., "He am tallest in class"), and (3) correct sentences with no grammatical errors (e.g., "She is the prettiest cat."). A total of 60 trials were presented; each condition had 20 trials, and no trials were repeated. Children were instructed to respond by pressing buttons in a button-box. If a sentence did not contain any mistakes, the children were instructed to use their right hand to press the right button, and if the sentence contained mistakes, children were instructed to use their left hand to press the left button.

The task was designed to be rapid event-related, and it was randomized using OptSeq2.[42] Jittered rest time in between trials varied, and it totaled 90 s. Each sentence played for ~4 s, which was followed by a 2-s display of a question mark in the center of the screen that allowed time for the participant to respond. The children could respond any time between the start of the sentence, which was then followed by an additional response time period that presented a question mark in the middle of the screen. A fixation cross was displayed in the center of the screen between jittered rest periods. The approximate total time of the task was 456 s (7.6 min). The task was presented using E-Prime 2 (Psychology Software Tools) on a 23-in. Philips 230E wide LCD screen connected to a Dell Optiplex 780 desktop computer. Sound was played via a Creative Inspire T12 2.0 multimedia speaker system.

Table 1 Participants' information (hair and skin colors).

Participant	Hair color	Skin tone
1	4	Medium
2	4	Light
3	1	Light
4	1	Light
5	4	Medium
6	2	Light
7	6	Light/medium
8	2	Light
9	1	Light
10	2	Light
11	6	Medium
12	4	Light

Hair color: 0, light blond; 1, dirty blond; 2, light brown; 3, medium brown; 4, dark brown; 5, light black; and 6, dark black.

2.3 *Functional Near-Infrared Spectroscopy Recordings*

We used a TechEN-CW6 system with 690 and 830 nm wavelengths. The setup included one emitter and three detectors spaced 2.7 cm apart, yielding three data channels sampled at 10 Hz (Fig. 1). We examined brain activation in the left inferior frontal gyri (a language-related area). The probe localization was established and applied consistently for each participant using the international 10–10 transcranial system positioning;[43] Fz, Cz, and preauricular were measured for each participant and the emitter was anchored at F7 (see Fig. 1). The probes were mounted on the participants' head using costume made caps (foam). Two caps were built to fit the different head sizes of the age range. Each cap was equipped with optode holders into

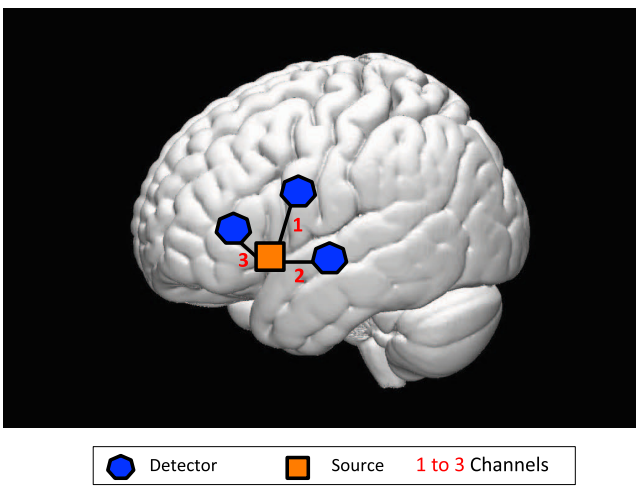

Fig. 1 Measurement location and channel distribution.

which the optodes were plugged in. An additional wrapping band was used to secure the optodes in place.

2.4 *Data Processing*

Data processing was completed using Homer2 fNIRS processing package[44] and an fNIRS data analysis tool (fNIRSDAT, a homemade software for general linear model (GLM) regression-based individual and group-level analysis) based on MATLAB® (MathWorks, Natick, Massachusetts). In order to understand the performance of different methods, we categorized the motion artifacts in our data into four types: type A—defined as a spike with a standard deviation of 50 that was over the mean within a second; type B—a peak with a standard deviation of 100 from the mean during a time portion ranging from 1 to 5 s; type C—a gentle slope between 5 and 30 s with a standard deviation of 300 from the mean; type D—a slow baseline shifting longer than 30 s with a standard deviation of 500 from the mean.

The raw data were first converted into optical density change (delta OD) using the optical density change calculation function (hmrIntensity2OD in Homer2 software package) in Homer2. We then slightly modified the motion artifact identification function (hmrMotionArtifactbyChannel in Homer2 software package) and applied it to the delta OD data to detect the number of the type A motion artifact. The parameters were selected as SDThresh = 50, tMotion = 1 (SDThresh, standard deviation threshold; tMotion, motion artifact time period); these parameters guaranteed the identification of the type A motion artifact (50 standard deviations to the mean within 1 s). However, the motion artifact identification function in Homer2 is only designed for identification of type A motion artifact (sudden move generated motion artifact). In this study, we also wanted to study the effect of type B, C, and D correction. Therefore, we used a home-made trend detection and detrending function using an MA algorithm to track the trend in fNIRS signal at different frequencies to identify type B, C, and D motion artifacts. This process also located the motion artifacts in the time series, thus providing the prior information for PCA and spline techniques.

We then applied five different motion correction techniques (PCA, spline, MA, wavelet, as well as MA and wavelet) to the delta OD series after the first round of the motion artifact detection process (except CBSI method; this method needs HbO and HbR data for motion correction; therefore, a different processing stream was applied for the CBSI method). Then, the motion artifact detection process was applied again in order to count the number of motion artifacts of different types. The motion corrected delta OD data were further filtered with a third-order Butterworth low-pass filter of ~0.5 Hz to prevent the physiological noise interference (e.g., cardiac noises) and high-frequency measurement noises; the filter coefficients are automatically designed by the MakeFilter function in MATLAB® (coefficients: A: [1 −2.37 1.93 −0.53] B: [0.0029 0.0087 0.0087 0.0029]). Then the filtered OD data were converted to HbO and HbR concentration data by the modified Beer-Lambert law (MBLL) (differential path length factor is selected to be 6 for both 690 and 830 nm wavelengths). Finally, the HbO and HbR data derived from previous steps went through a block averaging process and were regressed with a GLM.[45] A flowchart of the processing steps can be found in Fig. 2.

For CBSI method, we first applied a low-pass filter of (~0.5 Hz) followed by the MBLL conversion to the delta OD data after the first round of motion detection process. The derived HbO and HbR data were then passed through the CBSI motion correction, and the processed HbO and HbR data were sent to the block averaging process and GLM regression, as well as converted back to delta OD data for the motion artifacts counting process (see Fig. 2).

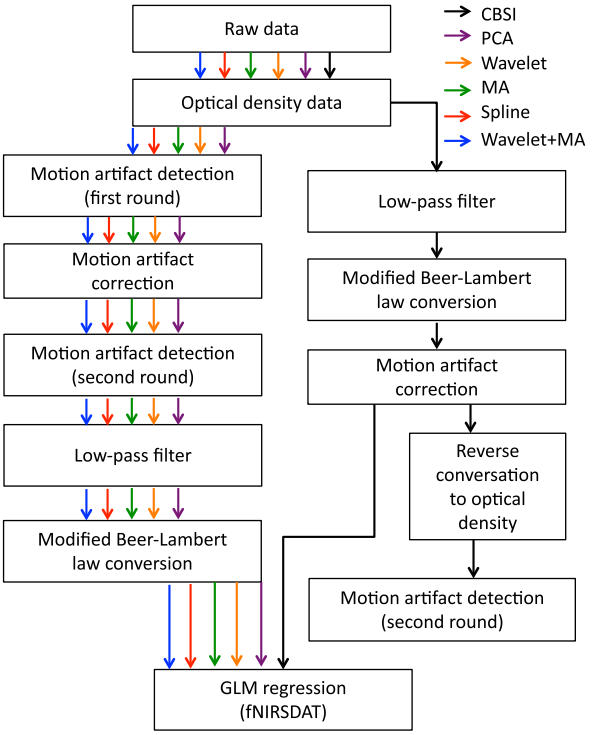

Fig. 2 Data processing flowchart.

2.5 *Motion Artifacts Correction Methods*

The approach for evaluating the performance of motion correction methods with pediatric fNIRS data in this work was modeled on Brigadoi et al. (2014), where they previously examined adult fNIRS data. Below is a brief description of the motion artifacts correction methods.

2.5.1 *Spline*

We applied a channel-by-channel spline interpolation method that is integrated with the Homer2 toolbox[44] and developed by Scholkmann and colleagues.[36] The method uses a cubic smoothing spline curve modeling motion artifacts. The reduction of motion artifacts is then achieved by subtracting the modeled spline interpolation curve. The curve is created using the csaps function in MATLAB software. To guide the curve's shape, this function used the smoothing parameter $p \in [0,1]$, which p increased from 0 to 1, thereby modifying the curve's shape from a straight line to a cubic spline. We used the integrated version of the spline interpolation method in Homer 2 toolbox,[44] and guided by prior work,[28,29,36] we set the p to 0.99.

2.5.2 *Principal component analysis*

PCA converts a set of signals into a set of uncorrelated time series retaining the variation. In the fNIRS context, PCA can transform the observed time series to linearly uncorrelated

principal components. Each component is accounting for a certain portion of the variance in the data, ranging from large to small. This process can be expressed as

$$X_{pc} = w \cdot X,$$

where X is the data matrix with dimension $N \times T$, w is the weight matrix with dimension $N \times N$, X_{pc} is the transformed $N \times T$ uncorrelated matrix, N is the number of fNIRS channels, and T is the number of time points in the collected signal. Given that the variance of motion artifacts is often much larger compared with the more meaningful hemodynamic responses, we usually remove a predefined portion of the variance in the transformed signal matrix X_{pc}, then restored the pruned principle components matrix to the signal using the weight matrix w, to achieve motion correction. We used the integrated version of the PCA method in Homer2 toolbox,[44] by setting the parameter nSV to be 0.9.

2.5.3 *Wavelet*

Wavelet analysis decomposes signals at various frequencies along time. This method was first applied to the fNIRS signal for motion artifact correction purposes by Molavi and Dumont.[37] Briefly, this method applies discrete wavelet transforms to decompose the fNIRS signal collected from each channel at multiple time-frequency locations. The time-frequency details of the signal are estimated as wavelet coefficients. The wavelet method assumes that the fNIRS signal linearly consists of task-evoked hemodynamic responses and motion artifacts. Therefore, the estimated wavelet coefficients are assumed to follow a normal distribution, where motion artifact relevant coefficients lie far from the center of the distribution. Thus, the motion artifacts can be diminished by removing these outlier coefficients then restoring the signal from the rest of the coefficients. In this study, we used the wavelet filtering function in Homer2 toolbox with Daubechies 5 wavelet,[44] with the tuning parameter iqr set to 1.5, following previous studies.[28,29,37]

2.5.4 *Correlation-based signal improvement*

Cui and colleagues originally developed the correlation-based signal improvement (CBSI) algorithm for fNIRS signal motion correction purposes.[38] This method presumes that the observed signal is a linear combination of the true hemodynamic responses, motion artifacts, and other measurement noises. It also assumes that the ratio between motion artifact amplitudes presented in HbO and HbR equals to the ratio between the amplitudes of the true HbO and HbR signals. Based on the assumptions made above, the CBSI algorithm can estimate the true physiological HbO and HbR signals according to the observed signal. In the current study, we applied CBSI algorithm by using its integrated version in Homer2 toolbox,[44] which requires no input parameters.[29]

2.5.5 *Moving average*

MA is a signal smoothing technique that captures the trend of the signal. It is a type of low pass finite impulse response filter. In this study, we applied the simple moving average (SMA) algorithm to the fNIRS data collected from each channel. The process can be described as

$$\mathrm{SMA} = \frac{P_{(t-n)} + \cdots + P_{t-1} + P_t + P_{t+1} + \cdots + P_{t+n}}{2n+1},$$

where t is the time and n is the number of points included in the MA. In this study, we chose n of 25 (5 s data). We first extracted the low-frequency trend from the signal using the moving average algorithm, then subtracted the trend from the signal.

2.6 *Metrics of Comparison*

We defined five metrics in order to quantitatively evaluate the performance of the different motion correction methods.

The first metric is defined as the amount variation of different motion artifacts (the number of motion artifacts identified before subtracted by the number of motion artifacts identified after). The numbers of four kinds of motion artifacts, identified before and after motion correction methods application, were calculated and compared for data on both 690- and 830-nm wavelengths.

The second and third metrics are the R values and t values generated during the GLM-based regression. The R value is the residual of the GLM-based regression. It indicates the similarity between the predesigned model and the recorded data. Therefore, a smaller R value indicates a better performance of a motion correction technique. A t value is the ratio between the regression coefficient beta and the residual. An increasing t value can be caused by either an increase of the beta value or a decrease of the residual, thus also indicating a promising performance of a motion correction method. In the current study, we expected the experimental conditions to have the highest responses in the language-related cortical area. Therefore, we only focused on the experimental condition relevant t values.

The fourth and fifth metrics are the area under the averaged hemodynamic response curve between 0 and 1.5 s, and the ratio of the area under the curve between 1.5 and 3 s to the area under the curve between 0 and 1.5 s. These two metrics are important parameters of the block-averaged response. These are concepts similar to those adopted in Ref. 29. However, the participants in this study are children, not adults. Therefore, we assumed that the time period it takes for a hemodynamic curve to reach its peak is shorter than that of an adult; thus, we used 1.5 s instead of 2 s.

3 Results

The quantity variations of different motion artifacts are displayed in Fig. 3.

For type A motion artifacts, the motion correction methods we applied, which produced the best results for both 690 and 830 nm wavelengths, were wavelet, CBSI, and the wavelet and MA combination, while the MA and PCA methods produced poor results for both wavelengths. For type B motion artifacts, the wavelet method alone, as well as the wavelet and MA combination, attained the best results for both wavelengths. For type C motion artifacts, the spline and wavelet and MA combination each performed better than other methods for both wavelengths. For type D motion artifacts, spline, MA, and the wavelet and MA combination outperformed the other methods for the 690 nm data, while all methods achieved the same performance compared to each other in 830 nm data, yet wavelet and CBSI performed better.

The GLM-based regression results are presented in Fig. 4. The t values (averaged across three channels) of the HbO signal

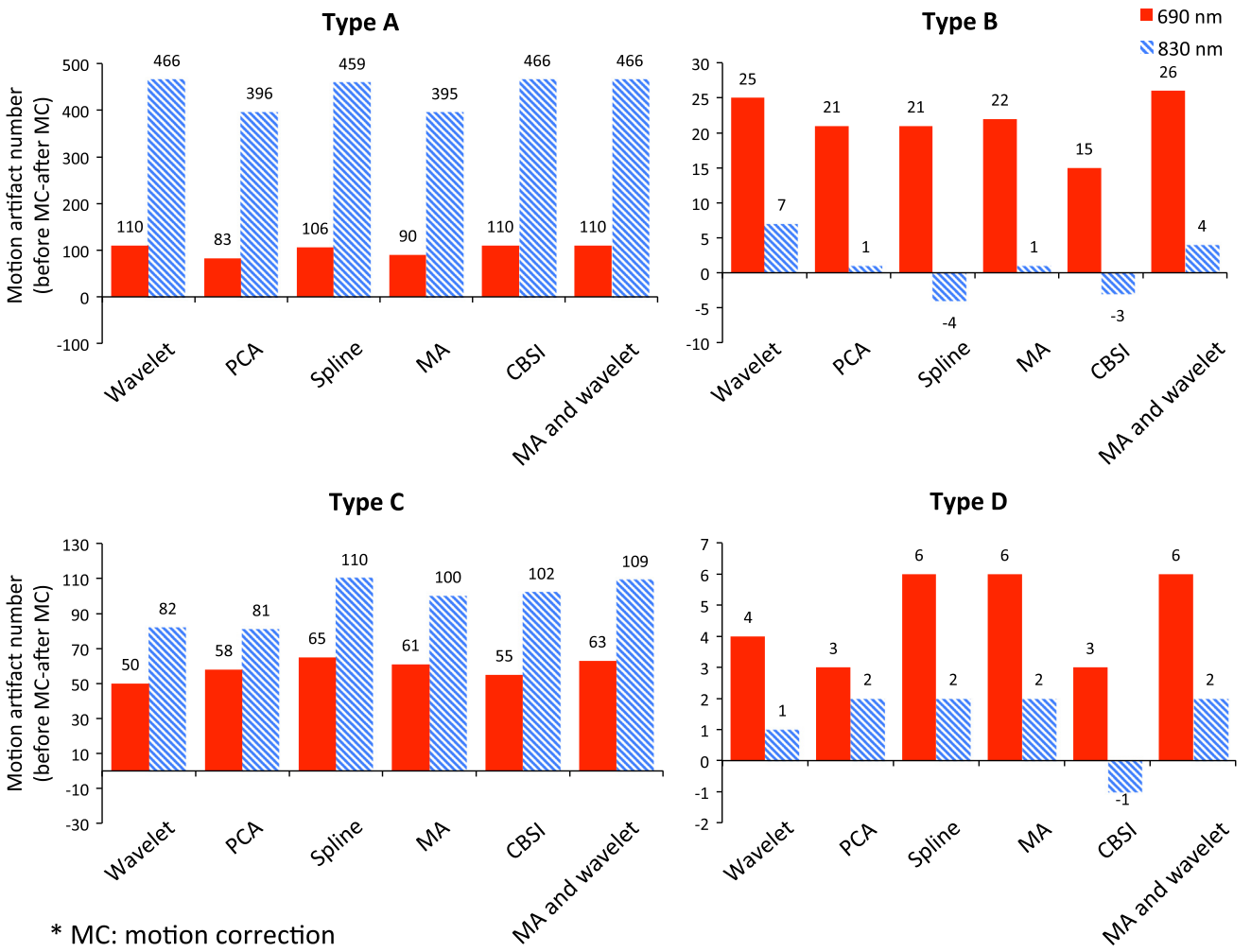

Fig. 3 The amount of variation of four types of motion artifacts after applying motion correction techniques.

from all channels increased after we applied CBSI, wavelet, MA, and the wavelet and MA combination, but decreased after we applied PCA and spline. The *t* values of the HbR signal changed in accordance with the HbO signal, but with a weaker magnitude. The *R* values of both HbO and HbR decreased after we applied wavelet, MA, CBSI, and the wavelet and MA combination.

Figure 5 shows the $AUC_{0-1.5}$ results of the HbO and HbR data. All of the methods we applied achieved similar

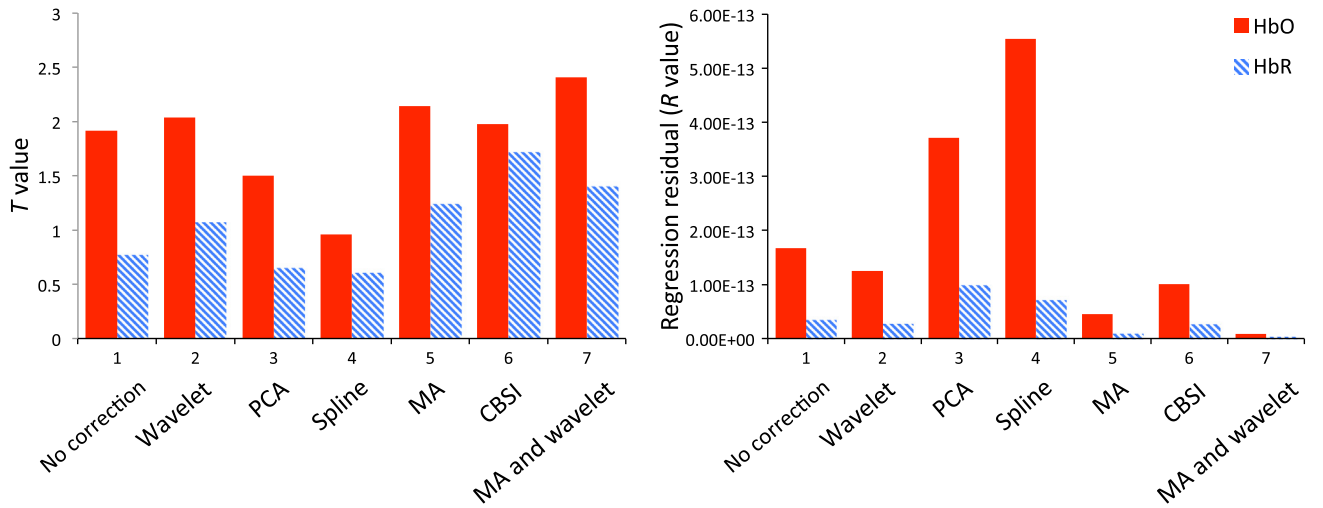

Fig. 4 General linear model regression derived *t* value and *R* value of three channels (three channels regressed as one) before and after applying different motion correction techniques.

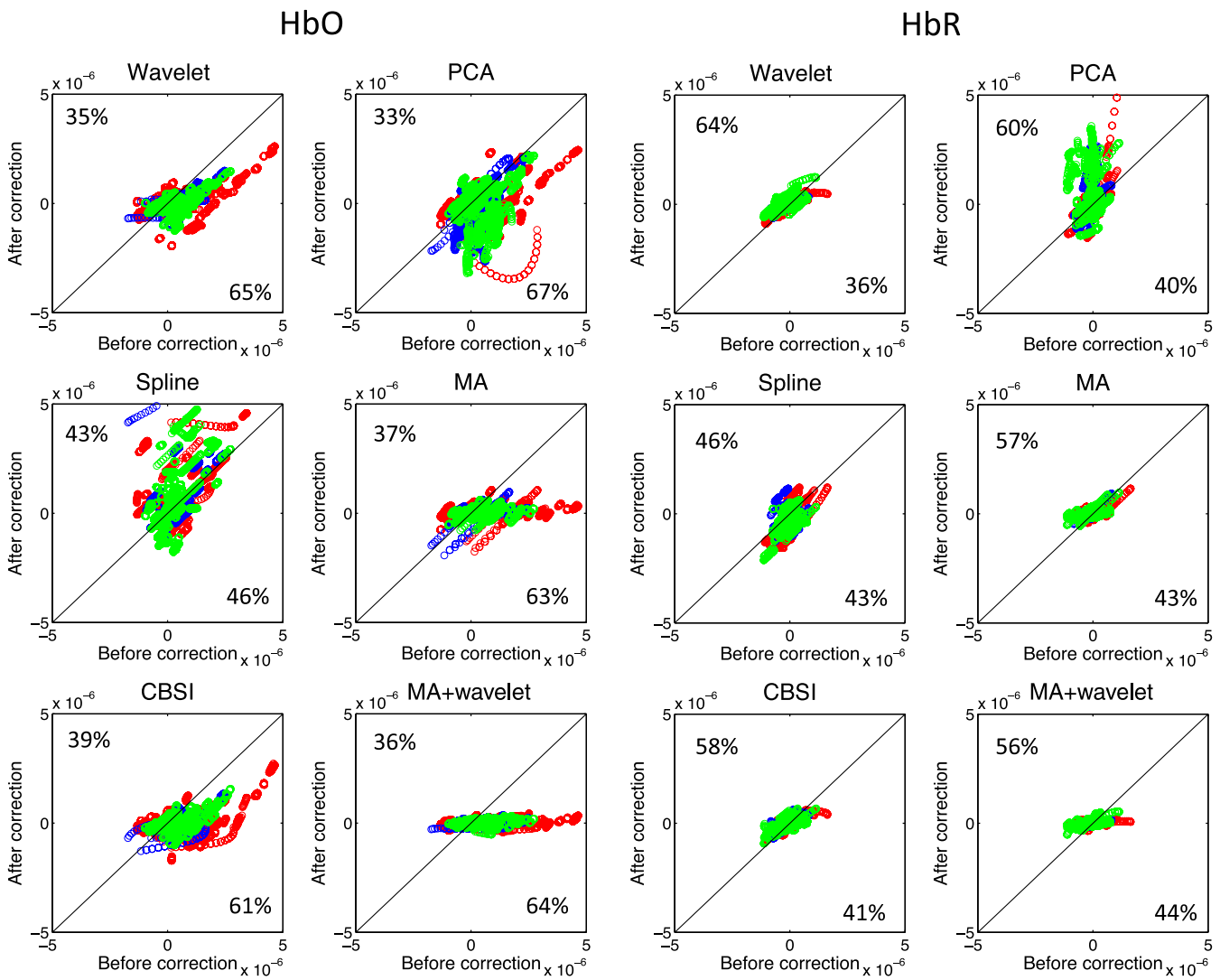

Fig. 5 $AUC_{0-1.5}$ variation after applying different motion correction techniques.

performances, except the spline and wavelet methods. Wavelet performed better than the other methods with a 65% decrease in HbO data and 64% increase in HbR data. Spline presented a less promising performance, showing only a 46% decrease in HbO data and a 46% increase in HbR data. The AUC_{ratio} results are shown in Fig. 6. The MA method achieved the best performance for both HbO and HbR data, while the spline and CBSI methods performed worse than other methods for both HbO and HbR data.

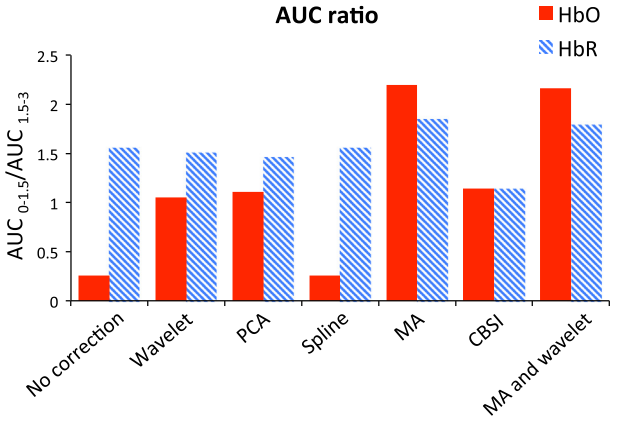

Fig. 6 AUC_{ratio} before and after applying different motion correction techniques.

4 Discussion

The goal of the study was to examine the efficiency of standard motion correction methods toward developmental fNIRS data. Specifically, we compared the performances of six motion correction techniques on fNIRS data derived from children 6 to 12 years old. The motion correction techniques we selected for comparison did not require extra equipment or a real-time data acquisition environment. Importantly, we did not consider the trial rejection method. The comparison metrics we selected aimed to qualitatively and quantitatively compare the performances of the motion correction abilities of these different techniques.

Previous studies reported that spline interpolation can achieve promising[28,36,47] and less promising results.[29] In theory, the spline technique can reduce all four types of motion artifacts equally. However, in our study, using this technique yielded

minimal improvement in the data, in comparison to the data before motion correction techniques were applied in all metrics. The reason for this is that the spline technique uses the preidentified motion artifacts to subtract a generated cubic function. However, the motion artifact shapes vary in fNIRS data, thus causing unstable results. Vinette et al.[47] tested the spline interpolation method on real long-term fNIRS data measured from patients with epilepsy. Their results showed spline interpolation method is good at reducing baseline shift motion artifacts. However, they also pointed out the limitations of the spline interpolation method, requiring different parameters for different participants and lower performance when brain response signal and motion artifact overlapped. Thus, in our study, the use of same parameters for all participants and possible overlapping between task evoked responses and motion artifacts may lead to the less promising motion correction results.

Cooper et al. showed that using PCA produced better results than using no motion correction technique,[28] and Brigadoi et al. claimed the performance of PCA is variable depending on the data set.[29] In the current study, we examined a data set acquired from children. The performance of the PCA technique in our study was not promising. Our data show that PCA successfully reduced type B, C, and D motion artifacts, but did not reduce type A motion artifacts. Although the PCA technique performed fairly well in the area under the curve (AUC)-related metrics, the GLM-based regression-related metrics decreased after applying PCA for both HbO and HbR. The reason for this may be that the PCA technique is a multichannel approach; thus, it requires that an artifact be presented in multiple channels (to be identified as a principal component), which is not always the case. The basic assumption of the PCA technique, that the components need to be independent, was violated. This may have additionally contributed to the poor performance we observed.

The CBSI technique performed well with our data. It decreased the number of all motion artifacts, except type B. The $AUC_{0-1.5}$ and AUC_{ratio} metrics jointly showed an improvement in both averaged HbO and HbR data. The GLM-related metrics increased in t value, and also slightly decreased in residual for both HbO and HbR data. However, the results of CBSI can be unstable, especially because an assumption of the CBSI method is that the HbO and HbR data are always negatively correlated, which may not always be the case. According to our data, in most cases, when the HbO signal increases, the HbR signal decreases, but sometimes HbR increases or remains unchanged. Furthermore, this is a reason why fNIRS researchers have reported that the HbO signal is more reliable than the HbR signal in fNIRS studies.[48]

In our study, the wavelet technique achieved a promising result, as it greatly reduced type A and type B motion artifacts, and barely reduced type C and type D motion artifacts. It also achieved very good results in two AUC-related metrics, and slightly increased the t value and R value for both HbO and HbR data. An advantage of the wavelet technique is that it does not require any prior motion artifact detection algorithm. Wavelet decomposes the signal into components on different time resolutions and then detects and fixes motion artifacts, thus eliminating the need for detection algorithms. Finally, the corrected component signals are restored to form the motion corrected signal. Brigadoi et al.[29] concluded that the wavelet technique achieved the best results in their study. However, according to our data, wavelet showed a weaker ability in reducing type C and type D motion artifacts. Therefore, we introduced the MA technique to work jointly with the wavelet technique to enhance the motion correction results.

Our data showed that using MA accomplished decent performance on reducing type C and type D motion artifacts. It also achieved good results for all other four metrics. The problem with the MA method is that it has a weak ability to reduce type A and type B motion artifacts. Therefore, the MA method may not work independently when type A and type B motion artifacts are significantly dominant in the data.

The wavelet and MA methods independently yielded some of the best results in reducing type A/B and type C/D motion artifacts. Due to this, we applied wavelet and MA combined methods to our data to observe the cumulative results. As expected, the wavelet and MA combination method attained the best performance of reducing motion artifacts in our data. The method greatly reduced all four kinds of motion artifacts and acquired very good AUC-related metrics. The GLM regression t value for HbR signal we obtained was not as high as CBSI method. The R value we obtained was the smallest among all methods, indicating that this motion corrected data had the

Table 2 Methods' performance summarization.

	Metrics					
	Motion artifact reduction					
Methods	Short term	Long term	General linear model (GLM) t value	GLM R value	$AUC_{0-1.5}$	$AUC_{1.5-3}/AUC_{0-1.5}$
Wavelet	√					
Principal component analysis					√	
Spline		√				
Moving average (MA)		√				√
Correlation-based signal improvement						
MA and wavelet	√	√	√	√	√	√

Note: √ indicates the good performance in certain category.

smallest deviation from the modeled data. As stated above, the reason for this is that the wavelet method and the MA method can, respectively, handle high-frequency and low-frequency noises. However, the disadvantage of using these combined methods is that the motion correction process must be applied two times, which may cause an over-reduction of the magnitude of the signal and can lead to low regression coefficients and low t values.

In general, the spline technique was fairly successful at reducing all types of motion artifacts, but leads to unsatisfied metrics. This was possibly due to improper correction of motion artifacts. The PCA method reduced different motion artifacts less successfully than other methods in this study, thus leading to less promising results in all metrics except $AUC_{0-1.5}$. The reason for this could be the absence of motion artifacts in the channels in our study. If so, PCA could not catch the motion artifacts as a principal component. CBSI demonstrated good results for reducing type A and type C motion artifacts, wavelet reduced type A and type B motion artifacts better than other methods, MA corrected type C and type D motion artifacts better than other methods, and the combination of wavelet and MA effectively reduced all types of motion artifacts. These latter four techniques achieved similar results in $AUC_{0-1.5}$, indicating they reduced the curve from 0 to 1.5 s. MA and wavelet combination and MA performed better than other methods for the AUC_{ratio} metric. Interestingly, both MA and wavelet combination and MA achieved promising results for t value and R value metrics. This indicates that type C and type D motion artifacts affect GLM metrics more than type A and type B motion artifacts. Table 2 provides a performance summarization for all the methods applied in the current study.

5 Conclusion

In this study, we applied six different techniques on fNIRS data recorded from children completing a language task and compared the performance of these techniques. We compared the differences in amount of motion artifacts, $AUC_{0-1.5}$, AUC_{ratio}, GLM regression-based t values and R values before and after motion correction application. Our data revealed that the MA and wavelet combination produced the largest improvement in the computed metrics. We also found that type C and type D motion artifacts affect GLM-related metrics more than type A and type B motion artifacts.

Acknowledgments

The authors thank the University of Michigan Center for Human Growth and Development and the University of Michigan Office of Research. We also thank participating families and members of the Language & Literacy Laboratory for their help in data collection, analyses, and manuscript preparation.

References

1. M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *NeuroImage* **63**(2), 921–935 (2012).
2. H. Obrig and A. Villringer, "Beyond the visible—imaging the human brain with light," *J. Cereb. Blood Flow Metab.* **23**(1), 1–18 (2003).
3. H. Koizumi et al., "Optical topography: practical problems and new applications," *Appl. Opt.* **42**(16), 3054–3062 (2003).
4. X. S. Hu et al., "Kalman estimator- and general linear model-based online brain activation mapping by near-infrared spectroscopy," *Biomed. Eng. Online* **9**, 82 (2010).
5. X. S. Hu, K. S. Hong, and S. S. Ge, "Recognition of stimulus-evoked neuronal optical response by identifying chaos levels of near-infrared spectroscopy time series," *Neurosci. Lett.* **504**(2), 115–120 (2011).
6. D. A. Boas, A. M. Dale, and M. A. Franceschini, "Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy," *NeuroImage* **23**, S275–S288 (2004).
7. V. Y. Toronov, X. F. Zhang, and A. G. Webb, "A spatial and temporal comparison of hemodynamic signals measured using optical and functional magnetic resonance imaging during activation in the human primary visual cortex," *NeuroImage* **34**(3), 1136–1148 (2007).
8. R. L. Barbour et al., "Optical tomographic imaging of dynamic features of dense-scattering media," *J. Opt. Soc. Am. A* **18**(12), 3018–3036 (2001).
9. M. Ferrari, L. Mottola, and V. Quaresima, "Principles, techniques, and limitations of near infrared spectroscopy," *Can. J. Appl. Physiol.* **29**(4), 463–487 (2004).
10. H. Zhu et al., "Reduced interhemispheric functional connectivity of children with autism spectrum disorder: evidence from functional near infrared spectroscopy studies," *Biomed. Opt. Express* **5**(4), 1262–1274 (2014).
11. F. H. Tian et al., "Quantification of functional near infrared spectroscopy to assess cortical reorganization in children with cerebral palsy," *Opt. Express* **18**(25), 25973–25986 (2010).
12. L. Sugiura et al., "Sound to language: different cortical processing for first and second languages in elementary school children as revealed by a large-scale study using fNIRS," *Cereb. Cortex* **21**(10), 2374–2393 (2011).
13. M. H. Shalinsky et al., "Exploring cognitive functions in babies, children & adults with near infrared spectroscopy," *J. Vis. Exp* (29), e1268 (2009).
14. Y. Monden et al., "Clinically-oriented monitoring of acute effects of methylphenidate on cerebral hemodynamics in ADHD children using fNIRS," *Clin. Neurophysiol.* **123**(6), 1147–1157 (2012).
15. M. J. Kurz, T. W. Wilson, and D. J. Arpin, "An fNIRS exploratory investigation of the cortical activity during gait in children with spastic diplegic cerebral palsy," *Brain Dev.* **36**(10), 870–877 (2014).
16. I. Kovelman et al., "At the rhythm of language: brain bases of language-related frequency perception in children," *NeuroImage* **60**(1), 673–682 (2012).
17. B. Khan et al., "Identification of abnormal motor cortex activation patterns in children with cerebral palsy by functional near-infrared spectroscopy," *J. Biomed. Opt.* **15**(3), 036008 (2010).
18. K. K. Jasinska and L. A. Petitto, "How age of bilingual exposure can change the neural systems for language in the developing brain: a functional near infrared spectroscopy investigation of syntactic processing in monolingual and bilingual children," *Dev. Cogn. Neurosci.* **6**, 87–101 (2013).
19. A. Gallagher, R. Beland, and M. Lassonde, "The contribution of functional near-infrared spectroscopy (fNIRS) to the presurgical assessment of language function in children," *Brain Lang.* **121**(2), 124–129 (2012).
20. X. P. Ding, G. Fu, and K. Lee, "Neural correlates of own and other-race face recognition in children: a functional near-infrared spectroscopy study," *NeuroImage* **85**(Pt 1), 335–344 (2014).
21. K. L. Perdue et al., "Extraction of heart rate from functional near-infrared spectroscopy in infants," *J. Biomed. Opt.* **19**(6), 067010 (2014).
22. M. Matsui et al., "Referential framework for transcranial anatomical correspondence for fNIRS based on manually traced sulci and gyri of an infant brain," *Neurosci. Res.* **80**, 55–68 (2014).
23. M. Kobayashi et al., "Size-invariant representation of face in infant brain: an fNIRS-adaptation study," *NeuroReport* **23**(17), 984–988 (2012).
24. D. C. Hyde et al., "Near-infrared spectroscopy shows right parietal specialization for number in pre-verbal infants," *NeuroImage* **53**(2), 647–652 (2010).
25. T. Grossmann, S. Lloyd-Fox, and M. H. Johnson, "Brain responses reveal young infants' sensitivity to when a social partner follows their gaze," *Dev. Cogn. Neurosci.* **6**, 155–161 (2013).
26. A. Cristia et al., "Neural correlates of infant accent discrimination: an fNIRS study," *Dev. Sci.* **17**(4), 628–635 (2014).
27. A. Cristia et al., "An online database of infant functional near infrared spectroscopy studies: a community-augmented systematic review," *PLoS One* **8**(3), e58906 (2013).
28. R. J. Cooper et al., "A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy," *Front. Neurosci.* **6**, 147 (2012).

29. S. Brigadoi et al., "Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data," *NeuroImage* **85**(Pt 1), 181–191 (2014).
30. S. Tak and J. C. Ye, "Statistical analysis of fNIRS data: a comprehensive review," *NeuroImage* **85**(Pt 1), 72–91 (2014).
31. J. Selb et al., "Effect of motion artifacts and their correction on near-infrared spectroscopy oscillation data: a study in healthy subjects and stroke patients," *J. Biomed. Opt.* **20**(5), 056011 (2015).
32. F. C. Robertson, T. S. Douglas, and E. M. Meintjes, "Motion artifact removal for functional near infrared spectroscopy: a comparison of methods," *IEEE Trans. Biomed. Eng.* **57**(6), 1377–1387 (2010).
33. J. Virtanen et al., "Accelerometer-based method for correcting signal baseline changes caused by motion artifacts in medical near-infrared spectroscopy," *J. Biomed. Opt.* **16**(8), 087005 (2011).
34. A. Blasi et al., "Automatic detection of motion artifacts in infant functional optical topography studies," *Adv. Exp. Med. Biol.* **662**, 279–284 (2010).
35. Y. H. Zhang et al., "Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging," *J. Biomed. Opt.* **10**(1), 011014 (2005).
36. F. Scholkmann et al., "How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation," *Physiol. Meas.* **31**(5), 649–662 (2010).
37. B. Molavi and G. A. Dumont, "Wavelet-based motion artifact removal for functional near-infrared spectroscopy," *Physiol. Meas.* **33**(2), 259–270 (2012).
38. X. Cui, S. Bray, and A. L. Reiss, "Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics," *NeuroImage* **49**(4), 3039–3046 (2010).
39. M. Izzetoglu et al., "Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering," *Biomed. Eng. Online* **9**, 16 (2010).
40. M. Arredondo et al., "Bilingualism alters children's frontal lobe functioning for attentional control" *Dev. Sci.* (in press)
41. M. L. Rice, K. Wexler, and S. M. Redmond, "Grammaticality judgements of an extended optional infinitive grammar: evidence from English-speaking children with specific language impairment," *J. Speech Lang. Hear. Res.* **42**(4), 943–961 (1999).
42. A. M. Dale, "Optimal experimental design for event-related fMRI," *Hum. Brain Mapp.* **8**(2–3), 109–114 (1999).
43. V. Jurcak, D. Tsuzuki, and I. Dan, "10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems," *NeuroImage* **34**(4), 1600–1611 (2007).
44. T. J. Huppert et al., "HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain," *Appl. Opt.* **48**(10), D280–D298 (2009).
45. K. Friston, P. Jezzard, and R. Turner, "Analysis of functional MRI time-series," *Hum. Brain Mapp.* **1**, 153–171 (1994).
46. F. Scholkmann et al., "How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation," *Physiol. Meas.* **31**(5), 649–662 (2010).
47. S. A. Vinette et al., "Artifact reduction in long-term monitoring of cerebral hemodynamics using near-infrared spectroscopy," *Neurophotonics* **2**(2), 025004 (2015).
48. Y. Hoshi, "Functional near-infrared spectroscopy: current status and future prospects," *J. Biomed. Opt.* **12**(6), 062106 (2007).

Xiao-Su Hu is a research investigator at the fNIRS laboratory at the Center for Human Growth and Development, University of Michigan. His research interests are methodology development for neuroimaging techniques (fNIRS & EEG) and their multidisciplinary applications.

Maria M. Arredondo is a PhD student in developmental psychology at the University of Michigan, and is a recipient of an NSF Graduate Research Fellowship. Her research interests include understanding bilingual children's cognitive mechanisms as applied during language acquisition and academic achievement; her methods include using experimental and neuroimaging techniques (fMRI, fNIRS).

Alexandre F. DaSilva is the director of H.O.P.E. (Headache & Orofacial Pain Effort), which is a multidisciplinary collaborative team to investigate the brain as a research and therapeutic target for chronic trigeminal pain disorders. He is an assistant professor at the University of Michigan School of Dentistry. He has collaborated with his colleagues on novel neuroimaging techniques (PET and MRI-based), non-invasive brain stimulation (HD)-tDCS, mobile technology, and advanced 3D-neuronavigation projects, all related to trigeminal pain.

Timothy D. Johnson is professor of biostatistics at the University of Michigan, School of Public Health. His research interests are in Bayesian modeling, computational statistics, the interface of Bayesian methods and big data, and statistical image analysis. He received his PhD in biostatistics at UCLA in 1997. He is a fellow of the American Statistical Association.

Ioulia Kovelman is an assistant professor at the University of Michigan. She is the director of the Language & Literacy Laboratory in the Department of Psychology and a co-directory of the fNIRS laboratory at the Center for Human Growth and Development. She studies bilingual language and reading acquisition in typically developing children and those with learning impairments. She also works to improve fNIRS neuroimaging method for the study of child development.