Journal of Biomedical Optics

SPIEDigitalLibrary.org/jbo

Accelerometer-based method for correcting signal baseline changes caused by motion artifacts in medical near-infrared spectroscopy

Jaakko Virtanen
Tommi Noponen
Kalle Kotilahti
Juha Virtanen
Risto J. Ilmoniemi

SPIE

Accelerometer-based method for correcting signal baseline changes caused by motion artifacts in medical near-infrared spectroscopy

Jaakko Virtanen,[a,b] Tommi Noponen,[c] Kalle Kotilahti,[a,b] Juha Virtanen,[d] and Risto J. Ilmoniemi[a,b]
[a]Aalto University School of Science, Department of Biomedical Engineering and Computational Science, P.O. Box 12200, FI-00076 Aalto, Finland
[b]Helsinki University Central Hospital, BioMag Laboratory, HUSLAB, P.O. Box 340, FI-00029 HUS, Finland
[c]Turku University Hospital, Department of Nuclear Medicine and Turku PET Centre, P.O. Box 52, FI-20521 Turku, Finland
[d]GE Healthcare Finland, P.O. Box 300, FI-00031 GE, Finland

Abstract. In medical near-infrared spectroscopy (NIRS), movements of the subject often cause large step changes in the baselines of the measured light attenuation signals. This prevents comparison of hemoglobin concentration levels before and after movement. We present an accelerometer-based motion artifact removal (ABAMAR) algorithm for correcting such baseline motion artifacts (BMAs). ABAMAR can be easily adapted to various long-term monitoring applications of NIRS. We applied ABAMAR to NIRS data collected in 23 all-night sleep measurements and containing BMAs from involuntary movements during sleep. For reference, three NIRS researchers independently identified BMAs from the data. To determine whether the use of an accelerometer improves BMA detection accuracy, we compared ABAMAR to motion detection based on peaks in the moving standard deviation (SD) of NIRS data. The number of BMAs identified by ABAMAR was similar to the number detected by the humans, and 79% of the artifacts identified by ABAMAR were confirmed by at least two humans. While the moving SD of NIRS data could also be used for motion detection, on average 2 out of the 10 largest SD peaks in NIRS data each night occurred without the presence of movement. Thus, using an accelerometer improves BMA detection accuracy in NIRS. *© 2011 Society of Photo-Optical Instrumentation Engineers (SPIE).* [DOI: 10.1117/1.3606576]

Keywords: near-infrared spectroscopy; motion artifact; accelerometer; sleep.

Paper 11168R received Apr. 6, 2011; revised manuscript received May 19, 2011; accepted for publication Jun. 13, 2011; published online Aug. 1, 2011.

1 Introduction

Medical near-infrared spectroscopy (NIRS) is a noninvasive technique for estimating hemodynamic variables from attenuation changes of 650 to 950-nm light in tissue, including concentration changes of oxygenated ($\Delta[HbO_2]$) and deoxygenated ($\Delta[HbR]$) hemoglobin.[1,2] Whereas positron emission tomography exposes tissue to ionising radiation, and functional magnetic resonance imaging requires the subject to maintain the same position throughout the study, NIRS has no risks involved with long-term use, and tolerates small movements of the subject. Combined with high temporal resolution, portability, and relatively low cost of equipment, these features make NIRS the only functional imaging modality currently suitable for long-term monitoring of cerebral oxygenation and hemodynamics.[3]

Although not directly disturbed by movement, the NIRS measurement is sensitive to the quality of the tissue–optode contact.[4] Large or sudden movements often disturb the contact, so that either the position of the optode relative to skin changes or hair enters/exits the space between optode and tissue. This may result in either transient noise in the NIRS signal, such as spiking, or if the new optode position persists, a signal baseline change. Transient noise mainly causes problems in studying stimulus-evoked responses, whereas baseline discontinuities typically prevent tracking $\Delta[HbO_2]$ and $\Delta[HbR]$ over time periods of tens of minutes or longer.

Human detection of motion artifacts in NIRS data is time consuming and does not allow for correcting the errors. Moreover, discarding artifact-affected data is wasteful, and may even be unfeasible if the study itself requires the subject to move. Analysis of NIRS data would therefore greatly benefit from an automated method for identifying and correcting motion artifacts in the data while leaving artifact-free data untouched. Such methods have been introduced before for removing transient motion artifacts from cerebral activation measurements,[5] continuous sinusoidal motion artifacts resulting from breathing or nodding,[6,7] and any physiological signal changes and artifacts that affect all measurement channels.[8,9] One novel approach, the motion artifact reduction algorithm (MARA) by Scholkmann et al., uses the moving standard deviation (SD) of NIRS signals to detect various types of motion artifacts.[10] However, most of these methods are founded on identifying motion artifacts solely based on their influence on NIRS data, and concentrate on transient artifacts instead of baseline changes.

In this study, we describe a motion artifact identification and correction method that is based on directly measuring the movement of the subject. Our algorithm for baseline motion artifact removal (ABAMAR) uses combined NIRS and accelerometer

Address all correspondence to: Jaakko Virtanen, Aalto University School of Science, Department of Biomedical Engineering and Computational Science, P.O. Box 12200, Espoo, FI-00076 Aalto, Finland; Tel: +358 50 344 3461; Fax: +358 9 470 23182; E-mail: jaakko.virtanen@aalto.fi.

1083-3668/2011/16(8)/087005/9/$25.00 © 2011 SPIE

data to identify motion artifacts that change the baseline of the NIRS amplitude signal. We apply the method to all-night NIRS measurements of sleeping subjects, and demonstrate that it allows tracking hemodynamic changes over sleep stage transitions that are often associated with movement.

Although no gold standard exists for motion artifact identification and removal from NIRS signals, visual interpretation is a fundamental, if laborious, approach to data classification. Therefore, in order to validate ABAMAR's performance, we compare its results to motion artifacts detected independently in the same data by three NIRS researchers. This is, to our knowledge, the first NIRS study comparing extensively the results of manual and automatic classification of motion artifacts in long-term monitoring.

2 Materials and Methods

2.1 *Measurements*

The data in this study were collected in a series of 30 all-night NIRS measurements on 13 healthy volunteers (9 males, 4 females, mean age 26 years, range 21 to 32). The study plan was approved by the ethical committee of the Helsinki University Central Hospital. The subjects were asked to sleep the whole night (approximately from 11 p.m. to 6:30 a.m.) as normally as possible until they were awoken by the researcher monitoring the measurement. Each subject was measured on at least two nights to reduce the influence of the unfamiliar surroundings on sleep quality; however, no such influence was observed when comparing sleep time to total measurement time on the first and second measurement night.

The measurement setup included NIRS, an accelerometer attached to the NIRS probe, polysomnography consisting of electroencephalography (EEG), electrooculography, and electromyography, and a fingertip pulse oximeter for recording arterial oxygen saturation and pulse (Fig. 1). Based on the polysomnography recording, a neurophysiologist scored the sleep into stages according to the rules by Rechtschaffen and Kales, so that the recording was divided into 30 s epochs designated as S1, S2, S3, S4, rapid eye movement (REM), wake, or movement time (MT).[11] The MT designation is used when motion artifacts in the polysomnography signals prevent sleep stage scoring. Since movements of the subjects were not restricted, also the NIRS data were expected to contain motion artifacts from movements during sleep.

Seven out of the 30 measurements were discarded due to problems in NIRS or EEG data acquisition, leaving 172 h of NIRS data in the remaining 23 measurements for analysis. The measurement duration was 7.5 ± 0.9 h (mean ± SD) per night.

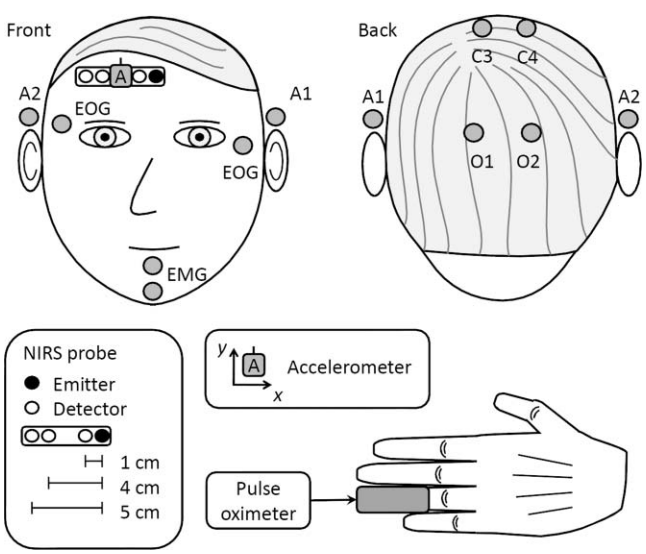

Fig. 1 Positioning of the NIRS probe and accelerometer on the forehead, and the accelerometer measurement axes. Placements of the polysomnography electrodes and pulse oximeter are also illustrated.

2.2 *NIRS Instrumentation*

A frequency domain NIRS device was used for the measurements.[12,13] The NIRS probe had one light source fiber and three detector fiber bundles at source-detector distances of 1, 4, and 5 cm, and was placed on the right side of the forehead just below the hairline (Fig. 1). During measurement, light pulses at 685 and 830 nm were time-multiplexed through the source fiber so that the effective sampling rate of $\Delta[HbO_2]$ and $\Delta[HbR]$ at the three detector locations was approximately 10 Hz.

The NIRS probe was secured to the head with an elastic strap. The end of each fiber was coupled to a reflecting prism terminal, so that the fibers could be placed parallel to the surface of the head and rest on the pillow (Fig. 2). Each fiber was attached to the scalp with adhesive tape to further reduce the sensitivity of the skin–optode contact to motion.

$\Delta[HbO_2]$ and $\Delta[HbR]$ were estimated using the modified Beer–Lambert law:[14]

$$\begin{pmatrix} \Delta[HbO_2] \\ \Delta[HbR] \end{pmatrix} = \frac{(\alpha^T\alpha)^{-1}\alpha^T}{d} \begin{pmatrix} \Delta A_{685nm}/DPF_{685nm} \\ \Delta A_{830nm}/DPF_{830nm} \end{pmatrix}, \quad (1)$$

where the 2×2 matrix α contains the specific extinction coefficients of HbO_2 and HbR at 685 and 830 nm, d is the geometrical distance between light source and detector, ΔA_i is the logarithmic light attenuation change at wavelength i, and DPF_i is the differential pathlength factor characterizing the average pathlength of photons in tissue. Attenuation changes due to light scattering and other chromophores than HbO_2 and HbR are typically considered negligible in medical NIRS.[14]

Although the frequency domain technique allows estimating the photon pathlength individually for each subject, technical difficulties in obtaining the necessary phase data prevented this in 10 measurements. Thus, we estimated average DPF values (6.16 for 685 nm and 5.84 for 830 nm) from the available phase

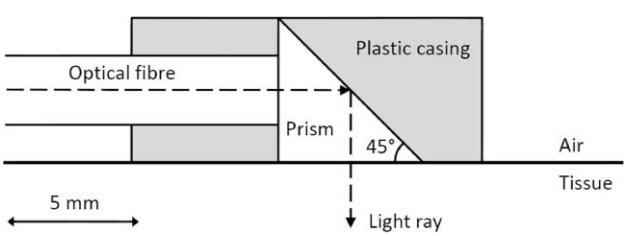

Fig. 2 Design of the prism terminal at the end of each optical fiber and fibre bundle.

data and used these values for all subjects for consistency. Values for α were obtained from the literature.[15]

2.3 *Accelerometer*

The accelerometer was manufactured by casting two SCA600-C28H1G MEMS acceleration sensors (VTI Technologies, Finland) inside a 20×30×8 mm resin block. The sensors were oriented to measure acceleration in two orthogonal directions, x and y, parallel to the head surface (Fig. 1). The accelerometer was designed to interface with the invasive blood pressure port of an E-PRESTN module of the S/5 patient monitor (GE Healthcare, Finland). The blood pressure waveform can be collected from the patient monitor at a 100-Hz sampling frequency and 0.01-mmHg resolution. The accelerometer receives a 5-V supply voltage from the module, and the output voltage from each sensor is scaled so that the measured acceleration is given by

$$a_i = \frac{P_i - 150\,\text{mmHg}}{150\,\text{mmHg}} \cdot g \qquad (2)$$

where a_i is the acceleration along axis $i = x, y$, P_i is the recorded blood pressure value, and g is 9.81 m/s^2.

Instrumental noise of the sensor component at the measurement band is 0.4 mVrms at output sensitivity of 1.2 V/g. This corresponds to 3.3×10^{-4} g in acceleration. Typical physiological noise from breathing movements and pulse was measured to correspond to 1.6×10^{-3} g.

The accelerometer signals were recorded with a laptop computer connected to the S/5 monitor. A separate computer recorded the NIRS data. The two computers were synchronized manually, with an error margin of approximately 0.2 to 0.3 s. Since there was no practical advantage from having a much higher sampling frequency for the accelerometer than for NIRS data, the accelerometer signals were downsampled by a factor of 10 to reduce the computational load of processing several nights of data.

2.4 *ABAMAR*

2.4.1 *Detecting motion*

When the subject is at rest, a_x and a_y consist of a baseline determined by the alignment of the sensor in Earth's gravitational field, signal fluctuation primarily from breathing, and random noise [Figs. 3(a) and 3(b)]. If the signal change between two consecutive samples exceeds a level equivalent to 1.3 g/s on either accelerometer channel, ABAMAR flags the corresponding time interval T_m as a motion event [Figs. 3(c) and 3(d)]. Motion detected within 20 s of a preceding motion event is considered part of the preceding event, extending T_m accordingly. If the duration of T_m is less than 1 s, the event is not counted as motion. The minimum thresholds for signal change and event duration were chosen based on visual inspection of the accelerometer signals so that breathing and very small movements are eliminated.

2.4.2 *Identifying baseline shifts*

For each motion event and NIRS channel, the NIRS amplitude baseline $\overline{A}$ is examined before and after T_m. The baseline is defined as the mean amplitude over 15 s, so that $\overline{A}_{\text{before}}$ and $\overline{A}_{\text{after}}$ are calculated for the time periods T_{before} (20 to 5 s before T_m) and T_{after} (5 to 20 s after T_m), respectively. The 5 s margin is left around T_m to ensure that the baseline is not affected by the motion. If $|\overline{A}_{\text{before}} - \overline{A}_{\text{after}}| > 2.6 \times s_{\text{before}}$, where s_{before} denotes the SD of amplitude during T_{before}, a baseline shift is identified. The empirically chosen threshold of $2.6 \times s_{\text{before}}$ corresponds to the 99th percentile around the mean for normally distributed data.

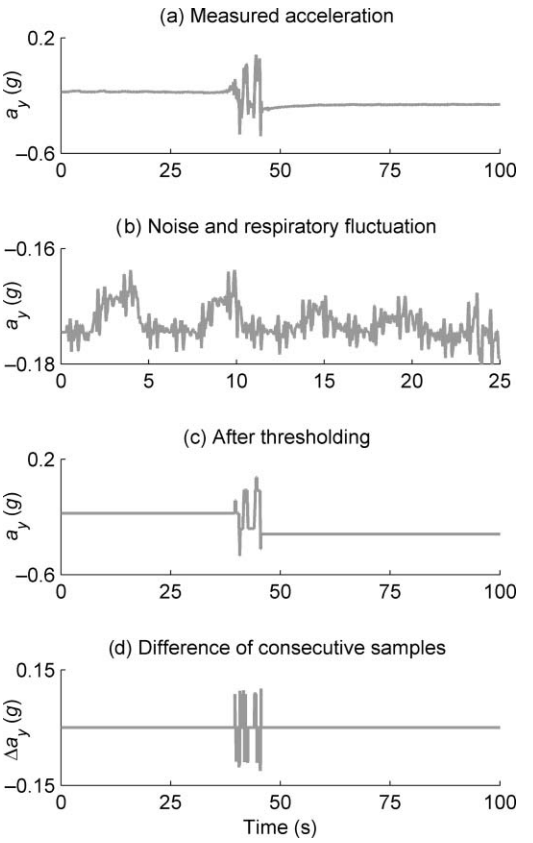

Fig. 3 (a) A typical 100-s segment of a_y during sleep, showing movement at approximately 40 to 50 s; (b) close-up of the first 25 s of a_y, demonstrating noise and respiratory fluctuation; (c) the signal after noise thresholding; and (d) difference of consecutive samples of the thresholded signal, indicating movement.

2.4.3 *Removing baseline shifts*

If a baseline shift is identified during the motion event on at least two NIRS detector channels, or at both light wavelengths on any one channel, it is considered to result from a motion artifact. To correct the artifact, the amplitude baseline preceding the event is imposed on all channels and wavelengths by multiplying amplitude values after T_m by $\overline{A}_{\text{before}}/\overline{A}_{\text{after}}$. Amplitude values between T_{before} and T_{after} are set to $\overline{A}_{\text{before}}$ to remove any transient motion artifact.

2.5 *Validation of ABAMAR*

2.5.1 *Validation against human classifiers*

To validate motion artifact identification with ABAMAR, three NIRS researchers independently viewed the amplitude and accelerometer data in 15 min blocks, and marked the beginning and

end of what they considered to be motion-artifact-related baseline changes. Each researcher had a background in biomedical engineering and signal processing, and at least 5 yr of experience in collecting and analyzing NIRS data in monitoring and evoked response studies. Due to the lack of a gold standard in NIRS motion artifact detection, no strict criteria for motion artifacts were defined to avoid biasing the classification toward the criteria used by ABAMAR. The final interpretation of what constitutes as a baseline change by motion artifact (as opposed to transient artifacts and physiological changes) was left to the intuition of each human classifier (HC).

Student's paired t-test was used for comparing the average number of artifacts per night identified by different classifiers, and student's unpaired t-test was used for comparing the average duration of all artifacts identified. The level of statistical significance was chosen as $p < 0.01$. If an artifact identified by ABAMAR overlapped in time with artifacts identified by at least two HCs, they were registered as true positives. Any other artifacts identified by ABAMAR were registered as false positives.

2.5.2 *Validation against MARA*

To determine whether directly measuring movement of the subject improves motion artifact identification, accelerometer-based motion detection with ABAMAR was compared to NIRS-data–based motion detection with MARA.[10] The motion detection algorithm for MARA was implemented with a window size of 51 samples, or approximately 5 s for calculating the moving SD of the 4-cm $\Delta[HbO_2]$ signal. This corresponds to the approach Scholkmann et al. used for baseline shift artifacts.[10]

In MARA, a user-determined threshold for the moving SD of NIRS data is required for identifying motion artifacts. Particularly in long measurements, the baseline level of the SD of $\Delta[HbO_2]$ changes over time due to physiological variation and changes in the signal-to-noise ratio during the measurement. Since manually adjusting the threshold for SD baseline changes is impractical for large datasets and may introduce observer bias, the 10 highest SD peaks from each night were analyzed to determine whether they coincided with motion detected by ABAMAR, regardless of whether the peaks were associated with NIRS amplitude baseline change. This approach was based on the assumption that NIRS data would be affected by motion at least 10 times per each night, since people regularly change their position while asleep. SD peaks within 5 s of each other were counted as a single peak.

2.5.3 *Stationarity of the physiological baseline during motion*

ABAMAR is based on the assumption that any physiological changes in the signal baseline between T_{before} and T_{after} are negligible. To quantify the possible error resulting from this assumption, we determined for each incident of movement identified by ABAMAR the change in amplitude relative to the SD of amplitude before movement, $|\Delta A_{rel}| = |\overline{A}_{after} - \overline{A}_{before}|/s_{before}$. The median value of $|\Delta A_{rel}|$ over both measurement wavelengths at the 4-cm channel was compared between situations where ABAMAR detected a motion artifact and where movement was detected but no motion artifact.

Table 1 Frequency and duration of motion artifacts. Data in the second and third rows are given as mean ± SD.

	HC1	ABAMAR	HC2	HC3
Detected artifacts	324	384	456	736
Artifacts per night	14 ± 7	17 ± 9	20 ± 9	32 ± 17
Duration of artifact	20 ± 13 s	17 ± 20 s	46 ± 23 s	26 ± 14 s

3 Results

3.1 *Identifying Motion Artifacts*

The frequency and duration of motion artifacts as detected by ABAMAR and the HCs are given in Table 1. The number of motion artifacts per night detected by ABAMAR did not significantly differ from HC1 or HC2 ($p < 0.01$). However, differences between HC1, HC2, and HC3 were statistically significant, as well as the difference between ABAMAR and HC3. The average duration of artifacts was different between all classifiers except ABAMAR and HC1 ($p < 0.01$).

Of the 384 artifacts detected by ABAMAR, 304 (79%) were registered as true positive and 80 (21%) as false positive. For comparison, 78% of artifacts identified by HC1 were identified by both other HCs. This proportion naturally declined as the total number of detected artifacts increased, being 53% for HC2 and 34% for HC3. On the other hand, the proportion of artifacts not verified by at least one HC was 8% for ABAMAR, and 6, 11, and 39% for the three humans. Figure 4 shows motion artifact identification agreement between all four classifiers.

In addition to the data in Table 1, ABAMAR also detected on average 14 ± 16 instances of motion per night that were not associated with amplitude baseline change. The total number of ABAMAR-detected artifacts and nonartifact motion ranged from 2 to 70 per night, and exceeded 10 on all but one night. However, of the 10 highest SD peaks for each night detected by MARA, 47% did not coincide with motion detected by ABAMAR. When allowing for a margin for coincidence so that SD peaks within 5, 10, and 15 s of ABAMAR-detected motion were

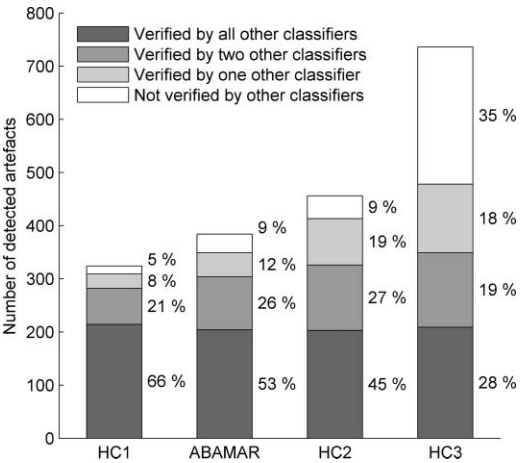

Fig. 4 The number and proportions of each classifier's artifacts that were verified by other classifiers.

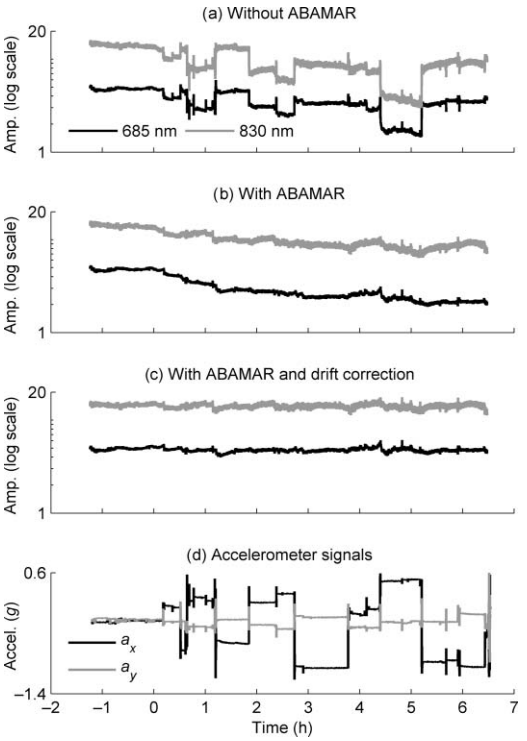

Fig. 5 (a) NIRS amplitude signals during a typical night; (b) the same signals after removing motion artifacts with ABAMAR; (c) after removing both motion artifacts and baseline drift; and (d) a_x and a_y. The amplitude signals have been scaled so that they do not overlap.

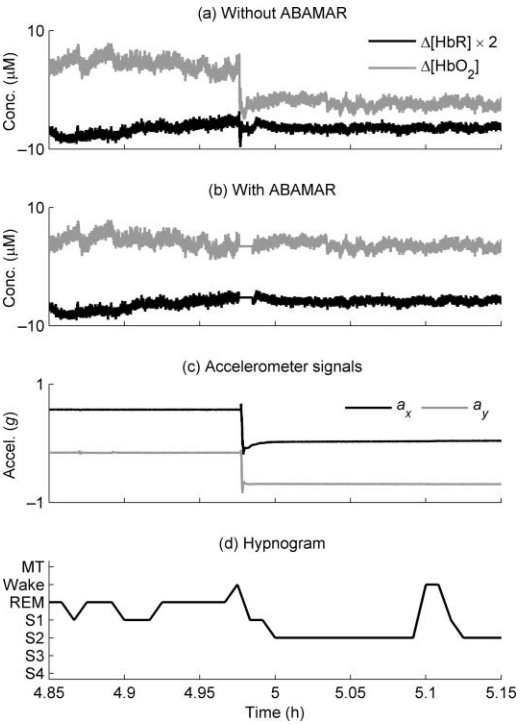

Fig. 6 Hemoglobin concentration changes (a) without and (b) with ABAMAR; (c) a_x and a_y from the same period; and (d) hypnogram showing transition from REM to S2 sleep. The zero levels of $\Delta[HbO_2]$ and $\Delta[HbR]$ have been chosen to avoid overlapping of the two curves, and $\Delta[HbR]$ has been multiplied by two for graphical clarity. Time is relative to midnight. ABAMAR produces constant values for $\Delta[HbO_2]$ and $\Delta[HbR]$ for approximately 30 s around the motion artifact.

attributed to movement, the percentage was 24, 21, and 20%, respectively.

3.2 *Removing Baseline Change Artifacts with ABAMAR*

Figure 5 shows typical results from applying ABAMAR to NIRS amplitude data. These data, as well as all NIRS data in subsequent figures, were acquired using the 4-cm detector. Prominent step discontinuities in amplitude signals at both wavelengths [Fig. 5(a)] that coincide with large changes in the accelerometer signals [Fig. 5(d)] are corrected by ABAMAR [Fig. 5(b)]. The downward slope of the remaining amplitude baseline is most likely caused by instrumental drift. In Fig. 5(c), an elliptic low-pass filter with passband at 0.2 mHz and stopband at 0.4 mHz is used for removing this drift.

Figure 6 shows how a baseline shift caused by movement during sleep stage transition affects $\Delta[HbO_2]$ and $\Delta[HbR]$. The subject is initially in REM sleep, but awakens for a short period before entering S2 sleep. The short awakening is accompanied by movement, causing a step discontinuity in $\Delta[HbO_2]$ and $\Delta[HbR]$. The original data thus imply that the sleep stage transition leads to a large decrease in $\Delta[HbO_2]$ and $\Delta[HbR]$, whereas data corrected with ABAMAR suggest that both $\Delta[HbO_2]$ and $\Delta[HbR]$ stay relatively constant after the termination of REM sleep.

In some cases, relatively rapid changes in $\Delta[HbO_2]$ and $\Delta[HbR]$ baseline during movement may be attributed to physiological changes. Figure 7 shows a case where step increases in $\Delta[HbO_2]$ coincide with motion during short arousals. The increases are comparable in magnitude to physiological fluctuation and are followed by a gradual return to the preceding baseline within approximately 5 min. However, they can also represent motion artifacts where the optode–tissue coupling is temporarily disturbed but returns to the previous state over time.

Both ABAMAR and the HCs typically counted as artifacts those events where the step discontinuity was large and rapid compared to surrounding physiological changes, and the new concentration level was sustained for at least several minutes. Such changes were sometimes followed by another motion artifact that returned the concentration baseline to the original level. Figure 8(a) shows such a case where the new concentration baseline after motion artifact persists for a period of approximately 20 min, and then returns to the preceding level following another instant of movement. The accelerometer signal baselines [Fig. 8(c)] show that the artifacts in question most likely result from the subject first changing his head position and later returning to the original position. ABAMAR successfully removes both motion artifacts [Fig. 8(b)].

3.3 *Stationarity of the Physiological Baseline During Motion*

ABAMAR detected, in total, 1274 incidents of movement, 384 of which were labeled as artifacts. The median value of $|\Delta A_{rel}|$ for the 4-cm channel was 5.5 for motion artifacts, 0.74 for nonartifact movement, and 1.4 when both types of incidents were counted. For comparison, the median values of $|\Delta[HbO_2]|$

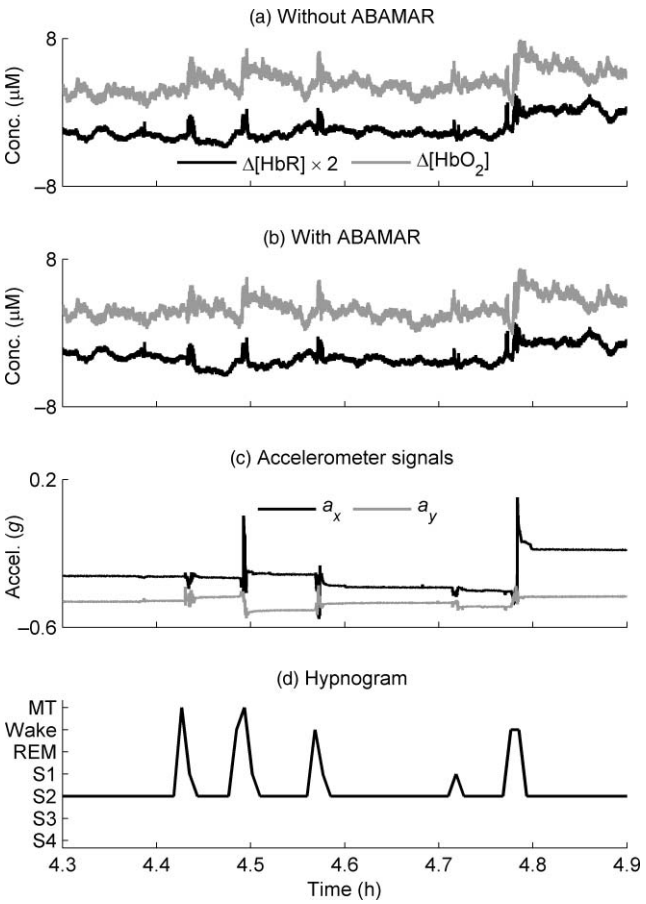

Fig. 7 An example illustrating the difficulty of determining whether step changes in hemoglobin concentration are caused by motion artifacts or physiological changes. Within the time window shown, ABAMAR identified no motion artifacts, while HC1, HC2, and HC3 identified 1, 3, and 6 artifacts, respectively.

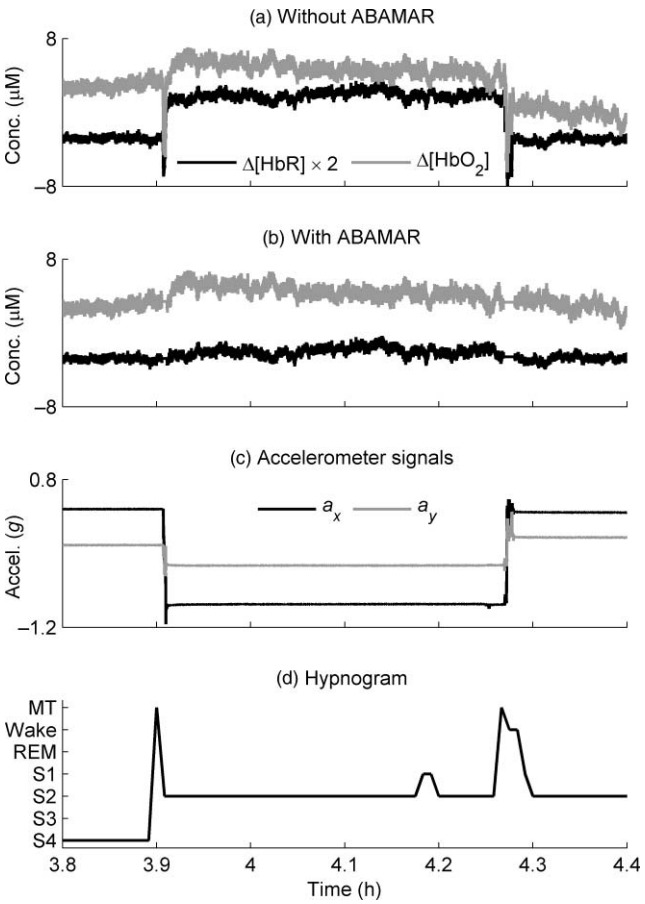

Fig. 8 A typical example of the correlation between head position and concentration baseline. Note that the first motion artifact is followed by a physiological increase in $\Delta[HbO_2]$, which returns to the previous level within the 20 min period. This physiological change is most likely related to a sleep stage transition as demonstrated by the hypnogram.

and $|\Delta[HbR]|$ for nonartifact movement were 0.50 μM and 0.15 μM, respectively.

4 Discussion

4.1 *Motion Artifact Correction in Literature*

Motion-related amplitude baseline changes are present in almost all NIRS measurements, and they complicate monitoring hemoglobin concentration changes over long time periods. The primary method for NIRS motion artifact reduction is careful experiment design, e.g., minimizing movement of the subject and ensuring that the optodes are attached firmly to the skin.[16,17] However, motion artifacts cannot be completely eliminated from sleep, critical care, operation room, or several other long-term physiological measurements because of the involuntary nature of movement in these situations. Although our sleep measurement probe was designed to be as rigid as possible without producing excessive discomfort, small movements of the optodes could not be prevented, and sometimes the whole probe came off during the night.

Most approaches to automatic motion artifact identification and/or correction in NIRS have concentrated on removing transient motion artifacts from brain activation measurements. For example, Cui et al. observed that while $\Delta[HbO_2]$ and $\Delta[HbR]$ are strongly negatively correlated during neuronal activation, motion artifacts tend to cause a strong positive correlation between the two.[5] Under this condition, they developed a correlation-based signal improvement method for NIRS motion artifact reduction. However, the assumption of strong negative correlation appears to be primarily valid in neuronal activation studies and for short time periods, where background cerebral hemodynamic activity can be considered constant. For example, an increase in total cerebral blood volume could increase $\Delta[HbO_2]$ without decreasing $\Delta[HbR]$.

Several studies have used adaptive filtering to remove both motion artifacts and physiological interference from NIRS data.[18–20] These are typically evoked response studies, where a short-distance source–detector channel is used for modeling the interference in a long-distance channel, and the performance of the filtering method is quantified by, e.g., comparing contrast-to-noise ratios of the cerebral activation signal before and after filtering. Again, while this approach is appropriate for cerebral activation studies where the hemodynamic response is short and independent from systemic hemodynamics, there are many monitoring applications where cerebral and systemic hemodynamics changes are tracked over long time periods and may exhibit parallel behavior. For example, hypercapnia often leads

to an increase in $\Delta[HbO_2]$ in both scalp and the brain, and sleep stage transitions are associated with changes in both systemic and cerebral hemodynamics.[21–23] In these cases, adaptive filtering may lose valuable data on cerebral hemodynamics. Also, if the tissue–optode contact is disturbed only in a single long-distance detector, adaptive filtering may miss the artifact.

While adaptive filtering methods are intended for removing physiological interference as well as both transient and baseline motion artifacts over the whole time period recorded, ABAMAR identifies and removes only baseline motion artifacts and leaves the signal otherwise untouched. The different domains of application and the lack of a known, quantifiable cerebral response in sleep data make comparison of ABAMAR with adaptive filtering methods in motion artifact reduction impossible. However, ABAMAR can conceivably be used as a preprocessing stage before adaptive filtering, or for tomographic reconstruction methods that assume that the signal baseline is continuous. Assessing the potential benefit of ABAMAR to separating systemic and cerebral hemodynamics with other methods warrants a separate study.

Unlike adaptive filtering algorithms, the MARA algorithm is explicitly intended for identifying motion artifacts based on the moving SDs of $\Delta[HbO_2]$ and $\Delta[HbR]$.[10] MARA also corrects baseline change artifacts somewhat similarly to ABAMAR. These features make it the most appropriate NIRS-data–based motion artifact correction algorithm for comparison with ABAMAR.

MARA can also be used for removing transient motion artifacts while preserving any physiological changes. This is in practice accomplished by making *a priori* assumptions about the frequency content of the motion artifact and approximating it with a spline function. However, this method is not capable of separating artifacts from physiological changes that occur on the same time scale, so the concentration estimates may be, to some extent, unreliable even after the correction. Another practical issue is that the use of several signal- and artifact-dependent parameters make algorithms such as MARA cumbersome for processing large amounts of data. This is particularly true when the algorithm and end result are not controlled by a NIRS signal processing expert. In monitoring applications, artifact identification similar to ABAMAR, i.e., by a single data-based threshold for the SD of the amplitude or concentration signals would be a more suitable approach.

4.2 *Benefit from Using an Accelerometer to Detect Motion*

Since nearly all existing methods for motion artifact removal rely solely on NIRS data, it is important to determine whether the accuracy of artifact detection could be improved by directly measuring motion. We compared motion artifacts detected with MARA's moving SD criterion against motion detected by the accelerometer. The results in Table 1 show that NIRS data can be expected to be affected by movement at least 10 times per night, so if motion causes peaks in the moving SD that are much larger than those from physiological changes, the 10 highest SD peaks should be associated with movement. However, comparison of the SD peaks against accelerometer data indicates that, on average, 2 of the 10 highest peaks for each night were most likely not associated with motion. Thus, if motion artifact detection relies solely on the properties of NIRS signals, distinguishing motion artifacts from other signal changes is prone to error. Of course, NIRS-data–based methods may be able to also detect artifacts caused by disturbances in the tissue–optode coupling that are not registered by the accelerometer. However, determining the frequency of such incidents in the data and thus their significance to data interpretation is difficult.

4.3 *Accuracy of Artifact Detection*

We compared artifact identification by ABAMAR to the performance of three humans to provide an independent measure of detection accuracy. Manual inspection has been regarded superior to automated algorithms in distinguishing between physiological signals and artifacts in NIRS.[3] Moreover, since any automated algorithm inevitably reflects the programmer's interpretation of data, we believe that human verification is the best available measure for motion artifact detection accuracy. Although there was considerable variance in the number of artifacts detected by the three humans, it appears that this can be largely attributed to sensitivity to the size of the artifact. In other words, the results can be interpreted so that HC1 marked only large or otherwise unambiguous artifacts, while HC3 marked even relatively ambiguous ones. Differences in artifact duration may be explained by a tendency to classify successive baseline changes as a single artifact, or to leave a large margin between the artifact and unaffected data.

The agreement between the HCs can be compared to inter- and intraclassifier reliability in sleep scoring, where visual inspection of electrophysiological signals is used for assigning sleep into stages based on formal criteria. A meta-analysis has shown that both the inter- and intraclassifier agreement can range approximately from 20 to 90% depending on the sleep stage and study, with approximately 70 to 90% being the norm.[24] Although there are differences between motion artifact identification and sleep stage scoring, it can be concluded that even visual inspection should not be expected to lead to near-perfect agreement between classifiers. Therefore, we used a majority vote criterion, so that artifacts detected by ABAMAR were accepted only if at least two of the three HCs identified the event as artifact. Under this criterion, ABAMAR agreed with the humans on a core set of approximately 300 artifacts, representing 79% of all artifacts detected by ABAMAR and comparable to interclassifier agreement between the HCs. Thus, ABAMAR appears to be generally capable of similar accuracy in baseline motion artifact detection as a human classifier.

We also investigated the effects of adjusting the amplitude threshold and the required number of channel–wavelength combinations on motion artifact identification with ABAMAR (data not shown). Increasing either of these parameters decreased the number of artifacts detected, but also slightly increased the ratio of true positives to false positives. The combination of parameters presented in this study should be applicable in most long-term monitoring applications independent of the measurement setup.

4.4 *Features of the Algorithm*

The amplitude baseline change caused by a motion artifact results from a change in light attenuation at the tissue–optode

coupling. As long as the amplitude level stays well above the noise level of the instrumentation, the artifact mathematically corresponds to multiplicative scaling of the amplitude signal. Therefore, the baseline correction made by ABAMAR (scaling the amplitude back to the level preceding the artifact) does not affect the magnitude or contrast of physiological changes. Both the artifact and the correction correspond to addition of a negative or positive constant term to $\Delta[HbO_2]$ and $\Delta[HbR]$.

ABAMAR uses a 15-s time window to estimate the amplitude baseline before and after movement. This time window should, in our opinion, be sufficient to minimize the effect of pulse and respiratory oscillations on the baseline, while using a longer time window would have increased the risk of slower hemodynamic changes affecting the baseline. In addition, a 5-s margin was set around movement detected by the accelerometer to ensure that the estimated amplitude baseline is free of motion artifacts. The baseline correction itself is based on the assumption that any physiological changes that may occur in the signal baseline during the movement are negligible. Our analysis of periods containing nonartifact movement supports this assumption, showing that such physiological baseline changes are typically smaller than the SD of the signal, and ABAMAR considerably reduces baseline tracking errors caused by motion artifacts.

ABAMAR ignores any motion artifacts when there is no amplitude baseline change, and does not recover any physiological concentration changes that may occur during the movement. These limitations can be addressed by adding a filtering algorithm (e.g., wavelet or spline interpolation based[8,10]) to ABAMAR to preserve physiological signal changes during movement. Our main reasons for not using such filtering are the uncertainty in distinguishing motion-related changes from physiological changes discussed earlier, and our primary interest in hemodynamic changes that occur over several minutes, e.g., in connection with sleep stage transitions.

Even after applying ABAMAR, signal drift originating from the measurement equipment or setup may prevent analysis of hemoglobin concentration changes over time periods of several hours. This baseline drift can be removed, e.g., by scaling the amplitude signal by a low-pass filtered version of itself, but such filtering also loses any physiological changes that have a time scale of approximately 1 h or more. Tracking such long-term changes requires improving the measurement setup or quantifying the drift, e.g., by simultaneously measuring a reference channel.

5 Conclusions

We have presented a novel accelerometer-based method for automatic identification and removal of motion artifacts that change the amplitude baseline of NIRS signals. Our results show that the use of accelerometer data improves motion artifact detection compared to purely NIRS-data–based methods, and the ABAMAR algorithm identifies baseline motion artifacts with similar accuracy to a human operator. The algorithm is based on simple rules, requiring only the addition of an inexpensive accelerometer to the measurement setup. It can be applied in nearly real time, with only a 20-s delay, or in post-processing. The algorithm allows tracking slow hemodynamic changes in NIRS monitoring applications over periods of tens of minutes or even hours.

Acknowledgments

The authors would like to thank Tiina Näsi for her help on motion artifact identification and Jussi Toppila for sleep scoring. This work was financially supported by the Finnish Cultural Foundation.

1. F. F. Jöbsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science* **198**, 1264–1267 (1977).
2. D. A. Boas, D. H. Brooks, E. L. Miller, C. A. DiMarzio, M. Kilmer, R. J. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," *IEEE Signal Process. Mag.* **18**, 57–75 (2001).
3. Y. Hoshi, "Functional near-infrared spectroscopy: Current status and future prospects," *J. Biomed. Opt.* **12**, 062106 (2007).
4. T. Noponen, K. Kotilahti, I. Nissilä, T. Kajava, and P. Meriläinen, "Effects of improper source coupling in frequency-domain near-infrared spectroscopy," *Phys. Med. Biol.* **55**, 2941–2960 (2010).
5. 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**, 3039–3046 (2010).
6. S. Fantini, D. Hueber, M. A. Franceschini, E. Gratton, W. Rosenfeld, P. G. Stubblefield, D. Maulik, and M. R. Stankovic, "Non-invasive optical monitoring of the newborn piglet brain using continuous-wave and frequency-domain spectroscopy," *Phys. Med. Biol.* **44**, 1543–1563 (1999).
7. M. Izzetoglu, P. Chitrapu, S. Bunce, and B. Onaral, "Motion artifact cancellation in NIR spectroscopy using Kalman filtering," *BioMed. Eng. OnLine* **9**, 16 (2010).
8. K. E. Jang, S. Tak, J. Jung, and J. Jang, "Wavelet minimum description length detrending for near-infrared spectroscopy," *J. Biomed. Opt.* **14**, 034004 (2009).
9. Q. Zhang, G. E. Strangman, and G. Ganis, "Adaptive filtering to reduce global interference in non-invasive NIRS measures of brain activation: How well and when does it work?," *NeuroImage* **45**, 788–794 (2009).
10. F. Scholkmann, S. Spichtig, T. Muehlemann, and M. Wolf, "How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation," *Physiol. Meas.* **31**, 649–662 (2010).
11. A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects," Public Health Service, NIH Publication No. 204, U.S. Government Printing Office, Washington, D.C. (1968).
12. I. Nissilä, K. Kotilahti, K. Fallström, and T. Katila, "Instrumentation for the accurate measurement of phase and amplitude in optical tomography," *Rev. Sci. Instrum.* **73**, 3306–3312 (2002).
13. I. Nissilä, T. Noponen, K. Kotilahti, T. Katila, L. Lipiäinen, T. Tarvainen, M. Schweiger, and S. Arridge, "Instrumentation and calibration methods for the multichannel measurement of phase and amplitude in optical tomography," *Rev. Sci. Instrum.* **76**, 044302 (2005).
14. I. Nissilä, T. Noponen, J. Heino, T. Kajava, and T. Katila, "Diffuse optical imaging," in *Advances in Electromagnetic Fields in Living Systems*, J. Lin, Ed., Vol. 4, 77–130, Springer Science, New York (2005).
15. M. Cope, "The application of near infrared spectroscopy to non invasive monitoring of cerebral oxygenation in the newborn infant," Ph.D. thesis, University College of London, Department of Medical Physics and Bioengineering (1991).
16. I. Miyai, H. C. Tanabe, I. Sase, H. Eda, I. Oda, I. Konishi, Y. Tsunazawa, T. Suzuki, T. Yanagida, and K. Kubota, "Cortical mapping of gait in humans: A near-infrared spectroscopic topography study," *NeuroImage* **14**, 1186–1192 (2001).
17. T. J. Huppert, S. G. Diamond, M. A. Franceschini, and D. A. Boas, "HomER: a review of time-series analysis methods for near-

infrared spectroscopy of the brain," *Appl. Opt.* **48**, D280–D298 (2009).
18. Q. Zhang, E. N. Brown, and G. E. Strangman, "Adaptive filtering for global interference cancellation and real-time recovery of evoked brain activity: a Monte Carlo simulation study," *J. Biomed. Opt.* **12**, 044014 (2007).
19. 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**, 1377–1387 (2010).
20. L. Gagnon, K. Perdue, D. N. Greve, D. Goldenholz, G. Kaskhedikar, and D. A. Boas, "Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state-space modeling," *NeuroImage* **56**, 1362–1371 (2011).
21. J. Virtanen, T. Noponen, and R. J. Ilmoniemi, "Properties of end-expiratory breath hold responses measured with near-infrared spectroscopy," *Proc. SPIE* **7896**, 78960D (2011).
22. P. Bušek, J. Vaňková, J. Opavský, J. Salinger, and S. Nevšímalová, "Spectral analysis of heart rate variability in sleep," *Physiol. Res.* **54**, 369–376 (2005).
23. Y. Hoshi, S. Mizukami, and M. Tamura, "Dynamic features of hemodynamic and metabolic changes in the human brain during all-night sleep as revealed by near-infrared spectroscopy," *Brain Res.* **652**, 257–262 (1994).
24. M. H. Silber, S. Ancoli-Israel, M. H. Bonnet, S. Chokroverty, M. M. Grigg-Damberger, M. Hirshkowitz, S. Kapen, S. A. Keenan, M. H. Kryger, T. Penzel, M. R. Pressman, and C. Iber, "The visual scoring of sleep in adults," *J. Clin. Sleep Med.* **3**, 121–131 (2007).