Hammerstein–Wiener Motion Artifact Correction for Functional Near-Infrared Spectroscopy: A Novel Inertial Measurement Unit-Based Technique

Hayder R. Al-Omairi [1,2], Arkan AL-Zubaidi [1,3], Sebastian Fudickar [4,5], Andreas Hein [4] and Jochem W. Rieger [1,3,*]

1 Applied Neurocognitive Psychology Lab, Carl von Ossietzky Universität Oldenburg, 26129 Oldenburg, Germany; hayder.r.alomairi@uotechnology.edu.iq (H.R.A.-O.); arkan.al-zubaidi@uni-oldenburg.de (A.A.-Z.)

2 Department of Biomedical Engineering, University of Technology—Iraq, Baghdad 10066, Iraq

3 Cluster of Excellence Hearing4all, Carl von Ossietzky Universität Oldenburg, 26129 Oldenburg, Germany

4 Assistance Systems and Medical Device Technology, Carl von Ossietzky University Oldenburg, 26129 Oldenburg, Germany; sebastian.fudickar@uni-oldenburg.de (S.F.); andreas.hein@uni-oldenburg.de (A.H.)

5 Institute for Medical Informatics, University of Lübeck, 23562 Lübeck, Germany

* Correspondence: jochem.rieger@uni-oldenburg.de

Abstract: Participant movement is a major source of artifacts in functional near-infrared spectroscopy (fNIRS) experiments. Mitigating the impact of motion artifacts (MAs) is crucial to estimate brain activity robustly. Here, we suggest and evaluate a novel application of the nonlinear Hammerstein–Wiener model to estimate and mitigate MAs in fNIRS signals from direct-movement recordings through IMU sensors mounted on the participant's head (head-IMU) and the fNIRS probe (probe-IMU). To this end, we analyzed the hemodynamic responses of single-channel oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) signals from 17 participants who performed a hand tapping task with different levels of concurrent head movement. Additionally, the tapping task was performed without head movements to estimate the ground-truth brain activation. We compared the performance of our novel approach with the probe-IMU and head-IMU to eight established methods (PCA, tPCA, spline, spline Savitzky–Golay, wavelet, CBSI, RLOESS, and WCBSI) on four quality metrics: SNR, $\triangle$AUC, RMSE, and R. Our proposed nonlinear Hammerstein–Wiener method achieved the best SNR increase ($p < 0.001$) among all methods. Visual inspection revealed that our approach mitigated MA contaminations that other techniques could not remove effectively. MA correction quality was comparable with head- and probe-IMUs.

Keywords: functional near-infrared spectroscopy fNIRS; motion artifact; multi-channel IMU; accelerometer; gyroscope; motion correction; NIRS signal improvement

1. Introduction

Functional near-infrared spectroscopy (fNIRS) noninvasively measures changes in oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) levels in the cortex with near-infrared light (e.g., 760 and 850 nm) [1]. To this end, light sources and detectors are attached to the head. The intensity of light arriving at a detector situated several centimeters away from the source is measured and then converted to concentration changes of HbO and HbR with the modified Beer–Lambert law. Ideally, the concentration changes reflect neuronal activity in the superficial cortical layers [2]. Portable fNIRS systems have been developed to facilitate brain activation measurements in realistic settings with relatively freely moving subjects, e.g., during driving [3]. However, motion artifacts (MAs) induced in less restricted settings, e.g., by the movement of the probes, are an essential challenge to overcome [4].

MAs come in various forms, including shape and magnitude changes, as well as oscillations. MAs alter the form of the observed shape of the HbO and HbR signals, resulting in erroneous estimates of brain activity [5]. Generally, sudden optode movements generate spike-like artifacts [6], whereas optode position changes generate baseline shifts [7], steps [8], oscillations [9], and slow drifts [10]. MAs are easier to detect when they have large magnitudes and contain high-frequency oscillations. However, small-magnitude and slow, low-frequency MAs can be hard to detect. This can limit the effectiveness of MA correction algorithms that include an MA detection step.

The most consequent approach to treating MAs once they are detected is to reject the contaminated data segment. However, this can lead to unacceptable data loss. In order to remove MAs without data loss, researchers have recently developed several MA correction approaches. Currently, a gold standard for effectively removing the MA effects is still lacking [11]. Popular methods make strong assumptions about the nature of the hemodynamic response and the MA. Some require additional MA detection algorithms, e.g., spline interpolation and Targeted Principal Component Analyses (tPCAs). Others, such as principal component analyses (PCAs), correlation-based signal improvement (CBSI) [12], and wavelet filtering [13] can be run without MA detection. In addition, smoothing methods have been used to filter out MAs, e.g., Savitzky–Golay (SG) filtering [14] and robust locally estimated scatterplot smoothing (RLOESS) [15]. However, hybrid or combined methods, such as spline interpolation with a Savitzky–Golay filter [10], wavelet filtering with spline interpolation [16], and wavelet filtering with CBSI (WCBSI) [17], have proven to be more effective in MA correction than each method alone. It should be noted that all these methods can either leave residual MAs to various degrees or introduce new artifacts if their assumptions are not met or when the wrong parameters are chosen [17].

Some recent MA correction methods linearly combine direct movement measurements from accelerometers and gyroscopes with fNIRS to estimate and/or remove MAs from fNIRS data. Examples include an auto-regressive model with an exogenous input (ARX) [18], accelerometer data with canonical correlation analysis [11], and accelerometer-based motion artifact removal (ABAMAR), which uses the standard deviation of the accelerometer data to detect MAs in fNIRS and corrects baseline shifts [19]. Linear methods assume a linear relationship between the IMU measurements and the fNIRS artifact. In this study, we investigated the capabilities of the linear and nonlinear Hammerstein–Wiener model (HWM) to improve MA correction. Our HWM approach integrates fNIRS data with concurrent measurements from a three-axis accelerometer and gyroscope to estimate MAs in the fNIRS signal and uses the estimates to correct them. In addition, we applied a bandpass filter (BPF) to refine the data further. This filter enforces constraints on the signal bandwidth imposed by the slow hemodynamic response function (i.e., the system's impulse response) and eliminates high-frequency noise as well as physiological artifacts from noise sources other than MAs [2,20].

We evaluated the performance of our suggested nonlinear MA correction method on empirical data and compared it to eight popular methods. In our experiment, we recorded brain activity in the motor cortex with fNIRS while participants performed a hand tapping task in three experimental conditions. In the first condition, participants kept their heads motionless while performing the task. This provided a ground truth signal of the expected brain activation. The second and third conditions required participants to perform small and large head movements while performing the task. We used this dataset to thoroughly assess the quantitative and qualitative aspects of our suggested approach and to validate its performance by comparing it with the most popular MA correction algorithms. For evaluation, we used four metrics that capture different aspects of algorithm performance: area under the curve (AUC), Root Mean Square Error (RMSE), Pearson correlation coefficient (R), and signal-to-noise ratio (SNR). A paired *t*-test was performed on all metrics to follow up on the significant differences in the performance of each tested algorithm. The dataset was made open access for future development and evaluations at https://www.doi.org/10.17605/OSF.IO/U3F89 (accessed on 12 May 2024).

2. Materials and Methods

2.1. Characteristics of the fNIRS Signal

The basic features of the interaction of near-infrared light (NIR) with human tissues relevant for functional neuroimaging are the following: (1) Human tissues are relatively transparent to light in the near-infrared spectral range (650–1000 nm) [21]. (2) NIR light is either absorbed by pigmented compounds (chromophores) or scattered in tissue [21]. (3) The dominant transport factor of NIR light in tissue is scattering, which is typically about 100 times more likely than absorption [22]. (4) Hemoglobin, which is located in vessels of the microcirculation, such as capillaries, arterioles, and venular beds, is a main contributor to the absorption of NIR light in the brain. The arterial blood volume fraction in the human brain is approximately 30% [23]. However, in fNIRS light, sources and detectors are attached to the scalp, and therefore the raw fNIRS signal is caused by a complex mixture of contributions by the blood supply and oxygen consumption in the brain, absorption in the tissues between sources and detectors (the scalp, bones, and the brain), and autonomous and heartbeat-related blood pressure changes, to name some [24]. Some of these non-brain influences can be mitigated by filtering techniques. Importantly, our study was designed to keep such factors comparable across experimental conditions.

2.2. Modified Beer–Lambert Law in the fNIRS Measurement

The fNIRS technique measures relative changes in light attenuation and uses a modified Beer–Lambert law to quantify changes in hemoglobin concentration as given by the following equation [25,26]:

$$\Delta OD_\lambda = \frac{\Delta c}{\varepsilon_\lambda \cdot L \cdot DPF}$$

where ΔOD_λ is the optical density change, ε_λ is the chromophore's extinction coefficient, Δc is the concentration change, L is the distance between the light entry and exit points, λ is the wavelength, and DPF is the differential pathlength factor. Most of the fNIRS devices use two wavelengths between 650 and 1000 nm to capture the HbO and HbR concentration changes [26,27].

2.3. Hammerstein–Wiener Model Configuration

The Hammerstein–Wiener model has its application in nonlinear system identification and integrates elements from both the Hammerstein and Wiener models. The Hammerstein model was first presented by Narendra and Gallman (1966) and the Wiener model by Norbert Wiener (1942). The Hammerstein–Wiener model applies a static nonlinearity to its input followed by a linear block. Conversely, the Wiener model passes the input through a linear block followed by a static nonlinearity. The Hammerstein–Wiener model combines both, resulting in a nonlinear–linear–nonlinear model structure as shown below in Figure 1. The Hammerstein–Wiener model represents a nonlinear system that is easier to implement and estimate than other linear or nonlinear models [28].

As depicted in Figure 1, we employed the Hammerstein–Wiener model with multiple inputs (IMU signals) and a single output (fNIRS signal). We used MATLAB's System Identification Toolbox (MATLAB version R2020a) to estimate the nonlinear and linear parameters. The estimation report provides details of the model configuration, e.g., the names of regressors used in both (first and last) nonlinear blocks, along with suggestions for parameters. Based on this report, we selected a sigmoid network with ten units for the nonlinear block and a state space model for the linear block, which is configured as a transfer function with three poles and two zeros with a short input delay of one.
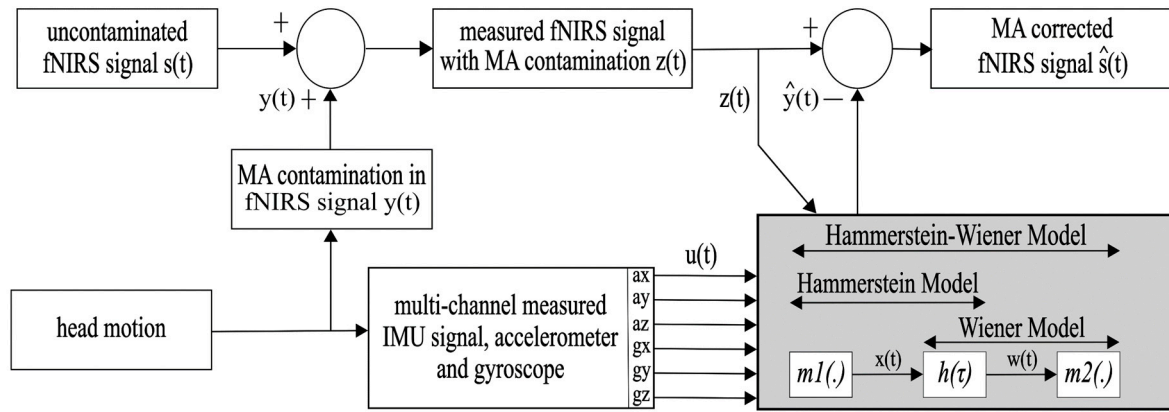
Figure 1. The block diagram of the HWM MA correction pipeline. Here, $s(t)$ is the MA-uncontaminated fNIRS signal, and $y(t)$ is the MA contamination by the head motion in the fNIRS signal. The measured fNIRS signal with the MA contamination is $z(t)$, and $u(t)$ holds the three-axis accelerometer and gyroscope signals. Both $z(t)$ and $u(t)$ are the inputs of the model. The HWM model (grey box) consists of two nonlinear blocks (m1 and m2) and one linear block $h(\tau)$. The model output $\hat{y}(t)$ represents the estimated MA signal, which is subtracted from the measured fNIRS signal with MA contamination $z(t)$ to obtain the $\hat{s}(t)$, the MA-corrected fNIRS signal.

2.4. Hammerstein–Wiener Model (HWM) for Artifact Removal

The whole MA-estimation-and-removal process with the HWM is outlined in Figure 1. We assume that the motion-contaminated fNIRS signal $z(t)$ measured by the fNIRS optodes is the sum of the uncontaminated ground truth fNIRS signal $s(t)$ and the MAs produced by head motions in the fNIRS signal $y(t)$. This can be written as the following equation:

$$z(t) = s(t) + y(t) \tag{1}$$

Here, we used the HWM to estimate the MA in the fNIRS-measured signal. HW modelling assumes that the input–output relationship of a system can be decomposed into two or more interconnected elements when their output is nonlinear. More precisely, the HWM configuration consists of a static nonlinear block $m1(.)$ cascading into a linear dynamic block $h(\tau)$, followed by another static nonlinear block $m2(.)$ [29].

The first nonlinear block $m1(.)$ transforms the input IMU data $u(t)$ as the following:

$$x(t) = c_1(u(t)) = \sum_{i=1}^{m} \sum_{q1=0}^{Q1} c_i^{q1} u_i^{q1}(t - \tau), \tag{2}$$

where $u(t)$ is the input with six dimensions (ax, ay, az, gx, gy, and gz) and $x(t)$ is the one-dimensional output of the first nonlinear block. i is the counter and m is the number of IMU data channels, τ is the time delay, and t is the time interval. Cascading (2) into the linear stage, we have the following:

$$w(t) = \sum_{\tau=0}^{T-1} h(\tau) \left\{ \sum_{i=1}^{m} \sum_{q1=0}^{Q1} c_i^{q1} u_i^{q1}(t - \tau) \right\} \tag{3}$$

Here, T is the memory length. Cascading the output from the linear stage into the second nonlinearity, and rearranging, we obtain the output equation for the Hammerstein–Wiener model MA estimation:

$$\hat{y}(t) = \sum_{i=1}^{m} \left\{ \sum_{q2=0}^{Q2} c_i^{q2} \left(\sum_{q1=0}^{Q1} c_i^{q1} \sum_{\tau=0}^{T-1} h(\tau) u_i^{q1}(t - \tau) \right)^{q2} \right\} \tag{4}$$

where $Q1$ is the order of the first nonlinear kernel in Equation (2) and $Q2$ is the order of the second nonlinear kernel in Equation (4).

In order to calculate the corrected fNIRS signal $\hat{s}(t)$, we subtract the HWM MA-estimated signal $\hat{y}(t)$ from the motion contaminated fNIRS signal $z(t)$, as described by the following equation:

$$\hat{s}(t) = z(t) - \hat{y}(t) \tag{5}$$

2.5. Motion Artifact Correction Algorithms

Here, we briefly review the popular MA correction methods employed in this study for comparison with the HW method. We refer the reader to [30] for details of the methods.

Principal component analysis (PCA) finds an orthogonal transformation (principal components) to transform the original time series of N fNIRS channels into N uncorrelated time series of the principal components. If the magnitude of the MAs is larger than the brain signal and is correlated across sensors, then PCA may be able to separate MAs from brain signals [30]. Unfortunately, PCA applied to the entire measured signal often has difficulty distinguishing between brain and MA signals and may, therefore, partly remove the signal as well [31]. An extension of traditional PCA called Targeted Principal Component Analysis (tPCA) aims to improve MA and brain signal separation. This algorithm first detects MAs, then concatenates the MA-contaminated segments to estimate the PCA, and finally removes MAs only from those segments. The method has some complexity, as the user must set several parameters [32].

Correlation-based signal improvement (CBSI) is based on the hypothesis that HbO and HbR are always negatively correlated. MAs change this correlation because they have similar effects on both signals. Therefore, HbO and HbR are positively correlated in MA segments [12]. The MA correction is performed by estimating the parameters of a simple linear function that predicts, e.g., HbR from HbO data. However, a consequence of CBSI correction is that HbR and HbO become redundant, as one is predicted from the other. This may not be true in the actually acquired data [30].

Wavelet correction (WC) is a time–frequency method. It is separately applied to the time series of every fNIRS channel and decomposes the signal in a set of frequency bands with varying temporal resolutions. WC produces detail and approximation coefficients for every frequency resolution. The outliers of the distribution of the detail coefficients are associated with MAs. Therefore, MAs can be eliminated by setting these outlier coefficients to zero before performing the inverse wavelet transform to reconstruct the MA-corrected signal time series [13]. The other approaches employed here perform MA correction by local function fitting: The spline interpolation (SI) technique first applies an MA detection algorithm and fits a cubic spline to the MA intervals. Then, the interpolation is removed from the original signal to obtain the MA-corrected signal [33]. Savitzky–Golay (SG) filtering is related, as it locally fits a polynomial function to the MA interval and replaces the observed time series with the estimate from the fitted polynomial [14]. Moreover, robust locally estimated scatterplot smoothing (RLOESS) first fits a locally weighted regression, which is then iteratively adjusted depending on the magnitude of the residuals [15].

WCBSI combines the wavelet with the CBSI approach for MA correction. In a previous study [17], WCBSI was proven to be the most effective MA correction method among the methods reviewed in this section across all quality measures reported in the current study.

2.6. Participants

We originally recorded data from twenty participants, but we lost the IMU data from three of them, so the analysis is based on seventeen healthy participants with an average age of 29 ± 4 years, including twelve females. Ten had various shades of brown hair, six had black hair, and one had red hair. Individual participants' age, gender, and hair color are listed in Table S2 in the Supplementary Materials. Participants had no neurological or psychological disorders. We obtained their written informed consent prior to their participation in the study. The IRB of Carl von Ossietzky Universität Oldenburg approved the study protocol under the code (Drs. EK/2020/021-01).

2.7. Experimental Procedure and Study Design

We measured the fNIRS, accelerometer, and gyroscope data while participants performed a hand tapping task and concurrent head movements with different amplitudes. A condition without head movements allowed us to empirically determine the ground truth fNIRS signal elicited by hand tapping. This approach allowed us to have significant control over the experiment. The experimental procedure was the same as in [17,34].

The experiment was performed in a dark and soundproof room. The participants were seated comfortably in front of a 24-inch LCD computer monitor at an approximately 75 cm distance. We used Psychtoolbox (v3.0.10), a MATLAB toolbox [35], to implement three tasks. Each trial involved 10 s of activity followed by a 20 s rest period for each task (Figure 2). This cycle was repeated 25 times, with a fixed task sequence, resulting in a total of 75 trials with a 41.5 min total data acquisition time per participant [17]. The experiment started and ended with a 120 s long baseline measurement without any overt activity from the participants. Psychtoolbox sent two signals at the beginning of each task to synchronize the IMU and the fNIRS recordings. (1) The event time for each task was sent to the fNIRS recording software (Oxysoft v3.0.103.3) through the lab streaming layer. (2) Simultaneously, a beep sound was sent to the Arduino board through the auxiliary port and was recorded in an extra channel in parallel with the IMU recordings.
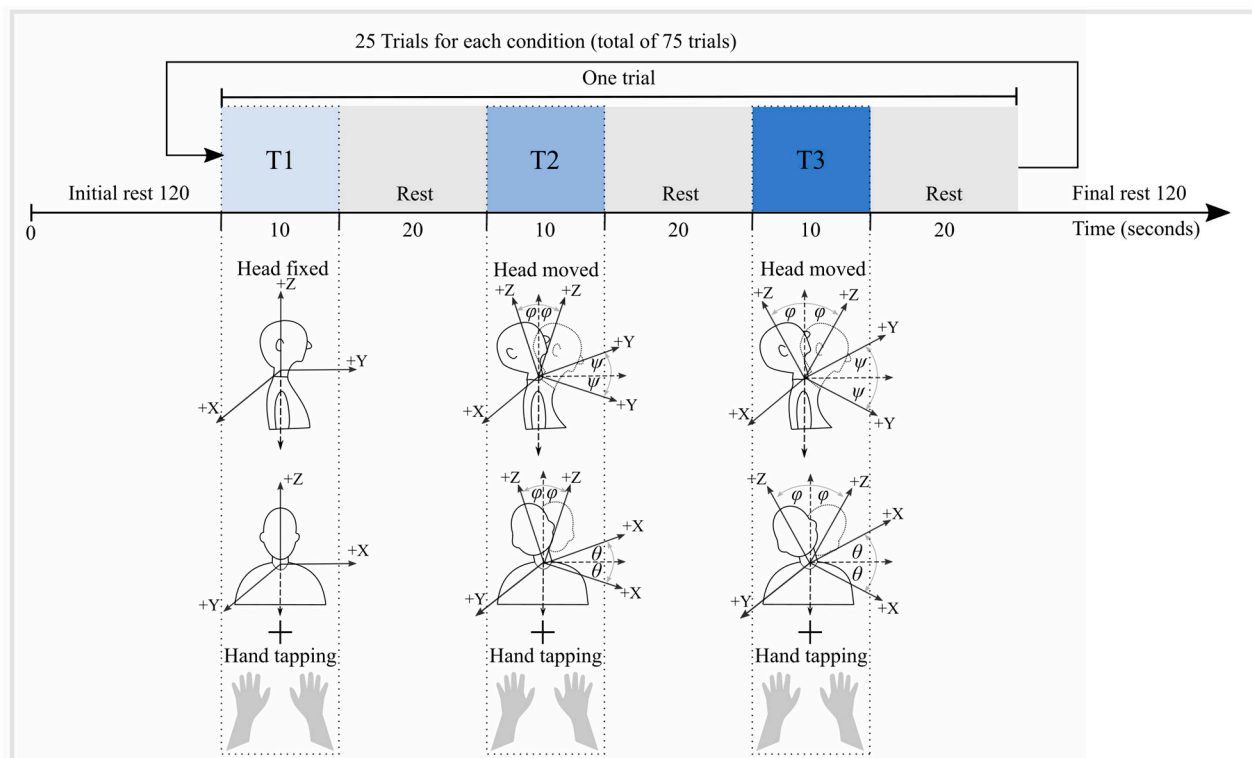

Figure 2. The experimental procedure and study design. The experiment begins with a 120 s initial rest period, followed by 10 s of task T1 (light blue box), involving hand tapping without head movements. After a 20 s rest period, 10 s of T2 (blue box) follows, involving hand tapping with small head movements (left–right–forward–backward). After another 20 s rest period, 10 s of task T3 follows (dark blue box), involving hand tapping and larger head movements. This sequence is repeated 25 times, resulting in 75 movement blocks (25 blocks for each task). The experiment concludes with a final 120 s rest period.

The participants were instructed to initially place their hands on the table and to avoid head movements. Upon the appearance of the word 'tapping' on the screen, they started tapping their hand on the table, keeping their head steady in the first task (T1). This task served to record the ground truth signal. In the second task (T2), participants were asked

to move their heads slightly while tapping their hands. The third task (T3) was similar to the second but with more substantial head movements. The head movements included flexion–extension (forward–backward) and left–right movements performed in random order. The hand tapping was performed at approximately 1 ± 0.5 Hz [17,36]. Participants practiced the tasks for 5–10 min before the experiment started to ensure that they followed the procedure correctly.

2.8. fNIRS and IMU Data Collection and Processing

We employed a wearable fNIRS Artinis medical OctaMon system to measure experimentally induced HbO and HbR changes in the motor cortex. The system consisted of eight sources emitting light at 850 and 760 nm wavelengths and two detectors sampled at 10 Hz. Sources and detectors were integrated into a specialized cap, with holder positions following the international transcranial positioning pattern 10/20 [37]. Detectors were placed in positions C3/C4 and sources in positions C1/C2, C5/C6, FC3/FC4, and CP3/CP4 (Figure 3). This created four fNIRS channels approximately placed over the hand area of the primary somatomotor cortex of each hemisphere. Before data collection, we checked the signal quality using the Oxysoft v3.0.103.3 recording program. This assessment included checking the light intensity and potential ambient light interference (<1% of the total signal) and ensuring that the light wavelengths emitted from each source fell within the specified range of 760 ± 5 nm and 850 ± 6 nm.
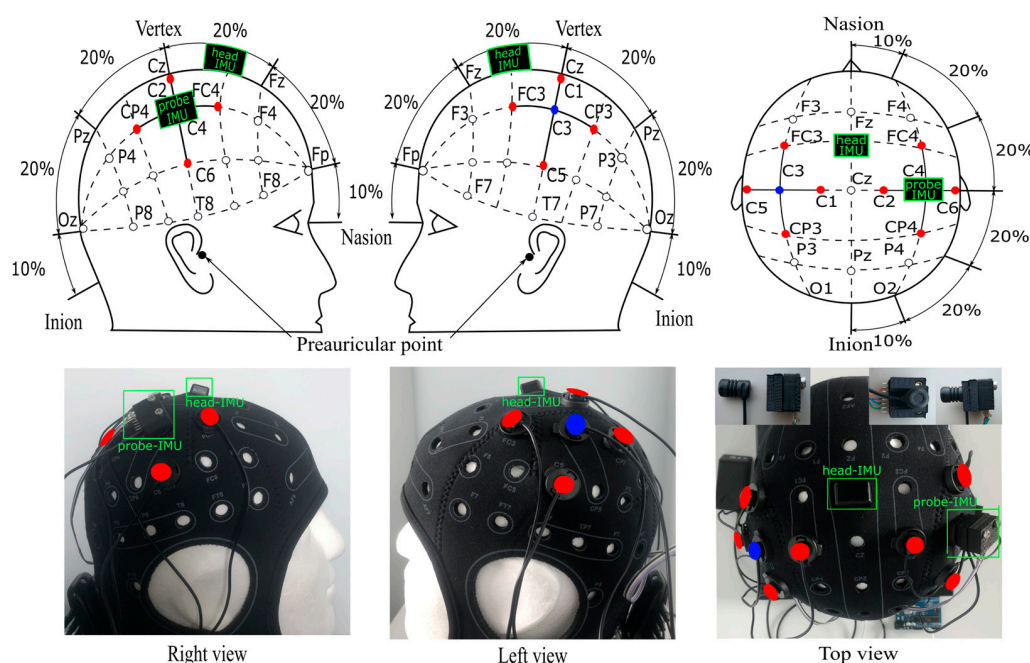

Figure 3. The positioning of optodes and IMUs. Positions follow the international transcranial positioning pattern 10/20 [37]. The fNIRS source and detector pairs are separated by 3 cm. Sources are shown in red and detectors in blue. The two IMU sensors are outlined in green. The head-IMU is at FCZ and the probe-IMU is on the top of detector C4.

We used two MPU6050 IMUs (each housing a 3-axis accelerometer and a gyroscope; AZ Delivery Vertriebes GmbH, Deggendorf, Germany) to measure linear and rotational movement components. Both were firmly fixed with custom-made holders for precise measurements of the detector and head movements. One IMU was fixed centrally at the position FCZ on the cap (head-IMU). The other IMU was fixed to the top of the fNIRS detector used in the analysis, which was at position C4 (probe-IMU) (see Figure 3). Both IMU sensors were recorded at a sampling frequency of 10 Hz, the same as the fNIRS sampling frequency, through an Arduino UNO board (for the code, see Supplementary Materials), and the data were stored on a PC.

2.9. Data Processing

Before the data recording, we checked the signal quality of all tested channels according to the user guide of the fNIRS (Artinis medical system) using the manufacturer's software (Oxysoft v3.0.103.3). The data processing was performed using the Homer3 v1.31.2 NIRS processing MATLAB toolbox [2], which implements the popular MA correction methods employed here. For HWM-based MA correction, we used the System Identification ToolboxTM [38] in MATLAB version R2020a. The block diagram in Figure 4 depicts the processing steps for all MA correction methods.
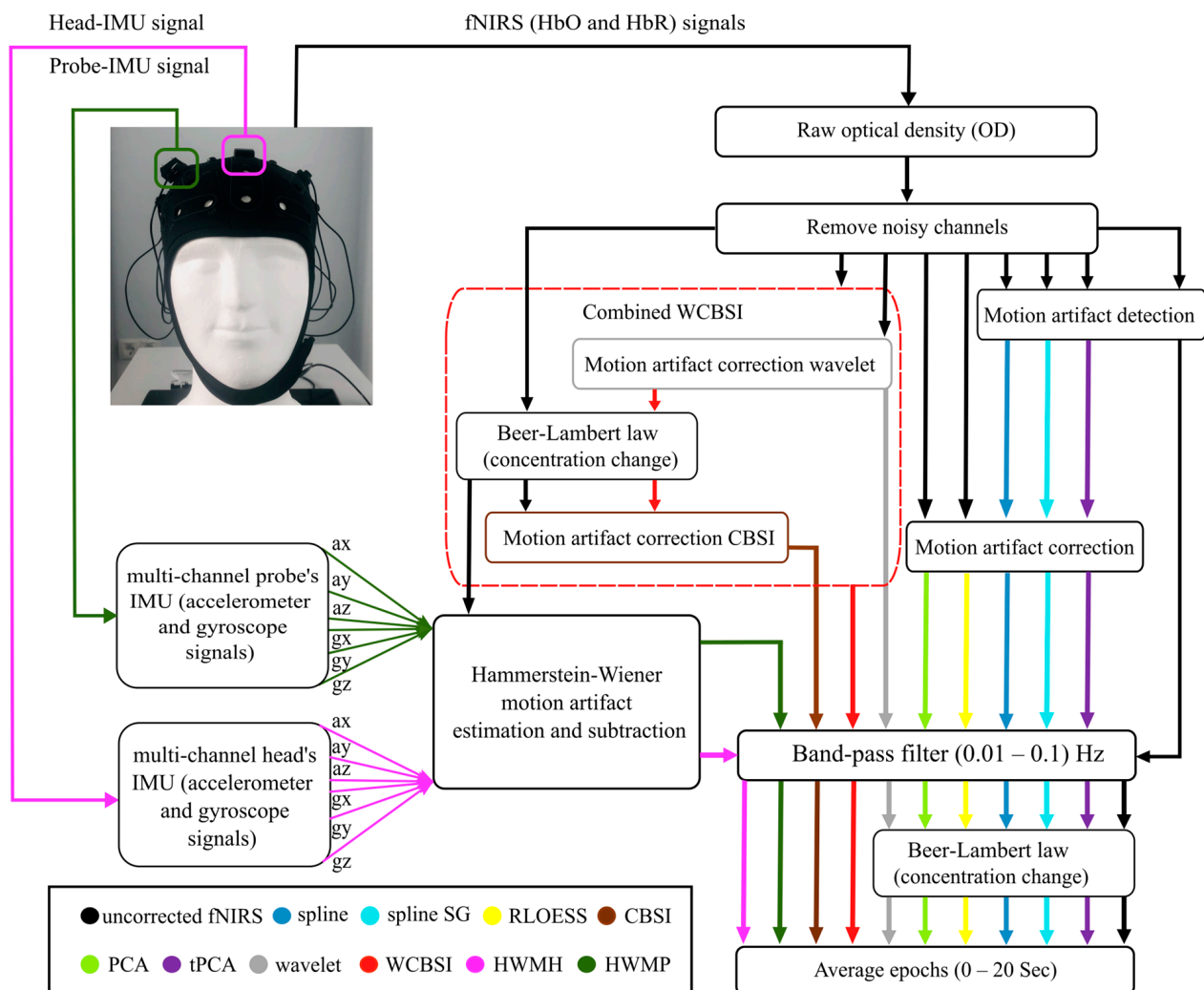

Figure 4. The processing streams for each MA correction technique are indicated by colored arrows. The processing of the uncorrected fNIRS signal is in black. The pipelines of the different correction approaches are shown in pink for HWMH, olive for HWMP, green for PCA, yellow for RLOESS, blue for spline, cyan for splineSG, and purple for tPCA, brown for CBSI, red for WCBSI, and grey for the wavelet.

The first step is applying the modified Beer–Lambert law to convert the raw intensity fNIRS measurements to the optical density (OD) data. The value of the differential pathlength factor depends on the age of the participant, and was determined by the method of Scholkmann and Wolf [39], which is integrated into the fNIRS manufacturer's software (Oxysoft v3.0.103.3), and this factor averaged to 6 ± 0.5 for 760 nm and 850 nm wavelengths [40].

The HWM estimation is applied separately with the head-IMU and the probe-IMU data. The accelerometer and gyroscope data are then run through the estimated HWMs to estimate the MA, which is subtracted from the fNIRS signal, resulting in the HWM-corrected fNIRS signals. HWMH refers to head-IMU correction and HWMP to probe-IMU correction.

A 3rd-order Butterworth band-pass filter with a cut-off frequency of 0.01–0.1 Hz is applied to the corrected data to reduce very slow drifts and high-frequency noise [2,20]. Finally, to estimate the mean HbO and HbR response elicited by the hand tapping, all epochs in each experimental task (T1, T2, and T3) were averaged over a 20 s long interval, starting with the hand tapping. This results in three mean HbO and three mean HbR responses (a ground truth hand tapping response signal, a hand tapping with a small-head-movement response signal, and a hand tapping with a large-head-movement response signal).

For the correction techniques that require defined artifact intervals (spline, splineSG, and tPCA), we used amplitude and standard deviation thresholds to mark the data samples with MAs. All parameters for motion detection and the MA correction algorithms were the same as in [2].

2.10. Metrics of Comparison

In order to quantify the performance of the MA correction approaches, we compared the averaged 20 s epochs of the $\widehat{HRF}$ motion-corrected signal to the average HRF of the ground truth signal obtained without head movements individually for the HbO and HbR signals at different levels of motion artifacts (T2 and T1), and for each technique. We used four metrics for this comparison:

The signal-to-noise ratio (SNR), measured in decibels, relates the ground truth-signal power to the power of the residual noise in the motion-corrected signal. It is calculated as in [5], where the ground truth signal is $HRF(i)$ with $i = 1, \ldots N$, the sample index. $\widehat{HRF}(i)$ is the motion-corrected signal. By subtracting both, we can calculate the residual noise signal $E(i)$.

$$E(i) = \widehat{HRF}(i) - HRF(i), \tag{6}$$

and with that, the SNR

$$\text{SNR} = 10 * log_{10} \left(\frac{\sum_{i=1}^{N} (HRF(i))^2}{\sum_{i=1}^{N} (E(i))^2} \right) \tag{7}$$

The Area under the curve difference (ΔAUC) is a global measure that compares the overall deviation of two curves (ground truth and motion-corrected signal) from the baseline. We calculated $AUC_{HRF(i)}$ and $AUC_{\widehat{HRF}(i)}$ using a MATLAB function for numerical integration (trapz). The difference is calculated using the following formula:

$$\Delta AUC = AUC_{HRF(i)} - AUC_{\widehat{HRF}(i)} \tag{8}$$

The Root Mean Square Error (RMSE) measures the average deviation between the ground truth signal and the motion-corrected signal using the following equation:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(HRF(i) - \widehat{HRF}(i) \right)^2}{N}} \tag{9}$$

Here, N is the number of samples in an epoch and i is the sample count. Note that RMSE is sensitive to shape and scaling differences.

Pearson's correlation coefficient (R) measures the similarity of the shape of the ground truth $HRF(i)$ and the motion-corrected signals $\widehat{HRF}(i)$. The Pearson correlation ranges between -1 and 1 and is insensitive to scaling differences between the two signals.

2.11. Statistical Analysis

In order to demonstrate that the HWM can reduce MAs, we compared the quality metrics (SNR, ΔAUC, RMSE, and R) between the HWM motion-corrected signal and the uncorrected signal with paired *t*-tests. For the comparison with the other MA corrections, we built on prior knowledge from our previous study [17], in which we demonstrated that WCBSI has the best MA correction performance among all comparison methods and quality measures employed in the present study. Therefore, in order to avoid an excessive number of statistical tests, we only statistically compared HWM to WCBSI, the "best in class" in the previous study. For each quality metric, we added a Benjamini–Hochberg correction with a false discovery rate of 5%. In addition, we calculated Cohen's d as a measure of the effect sizes. All statistics were performed using IBM SPSS (v29.0.0.0).

2.12. Measuring Participants' Head Orientation during Experimental Tasks

We measured participants' head motion in the experimental tasks (T1, T2, and T3) by calculating the rotational angles of the head-IMU accelerometer data (X, Y, and Z). Figure 2 shows pitch (θ) for the *X*-axis, roll (ψ) for the *Y*-axis, and yaw (φ) for the *Z*-axis. These angles are expressed according to the below formula in degrees relative to the Earth's gravity g = 9.81 m/s2 [41].

$$\theta = arctan\left[\frac{a_x}{\sqrt{a_y^2 + a_z^2}}\right] \times \frac{180}{\pi}, \text{ is the rotation angle of the } x\text{-axis.}$$

$$\psi = arctan\left[\frac{a_y}{\sqrt{a_x^2 + a_z^2}}\right] \times \frac{180}{\pi}, \text{ is the rotation angle of the } y\text{-axis.}$$

$$\varphi = arctan\left[\frac{a_z}{\sqrt{a_x^2 + a_y^2}}\right] \times \frac{180}{\pi}, \text{ is the rotation angle of the } z\text{-axis.}$$

where a_x, a_y, and a_z are the acceleration output on the (*x*-, *y*-, and *z*-) axis.

2.13. Head- and Probe-IMU Sensors' Relationship Based on Canonical Correlation Analysis (CCA)

Canonical correlation analysis (CCA) was first introduced by Hotelling (1936) with the goal of fitting projections of two datasets into a common space in which they are maximally correlated. Here, we use CCA to align the six-dimensional recording of the two IMU sensors (head-IMU and probe-IMU) in a common space to analyze their relationship.

3. Results

3.1. Qualitative Comparison of Motion Artifact Correction Methods

In this section, we provide a brief qualitative comparison of how different MA correction methods cope with different types of MAs. Figure 5 provides an illustrative example of single-trial HbO and HbR signals with various MA types marked in red. These include baseline shifts, drifts, and spikes. It can be seen in these examples that for both small (T2) and large (T3) head movements, the nonlinear HWM approach excels in eliminating the high-amplitude spikes, rectifying the up- and down-baseline shifts, and successfully mitigating slow drifts. All methods seem to correct both up and down spikes, but this may be at least partly due to the low-pass filters applied as the last step after MA correction. Of the other MA correction methods tested, only HWMH, HWMP, WCBSI, and RLOESS corrected the downshift and slow drift in T2. Furthermore, in T3, our suggested HWMH, HWMP, and WCBSI succeeded in remedying the up-baseline shift, but none of the other techniques did. Notably, HWMH and HWMP perform best in resetting the offset amplitude of the signal in all three tasks. Moreover, only HWMH and HWMP maintain consistent fNIRS response amplitudes across all three tasks (T1, T2, and T3). Note how well the HWM-corrected HbO and HbR time courses follow the expected time courses (positive

for HbO and negative for HbR), with comparable peak amplitudes toward the end of the movement interval despite large head movements. Also, note the similarity between the HWM corrections obtained with the head- and the probe-IMUs. In the following, we will provide a quantitative evaluation of the performance of all MA correction algorithms.
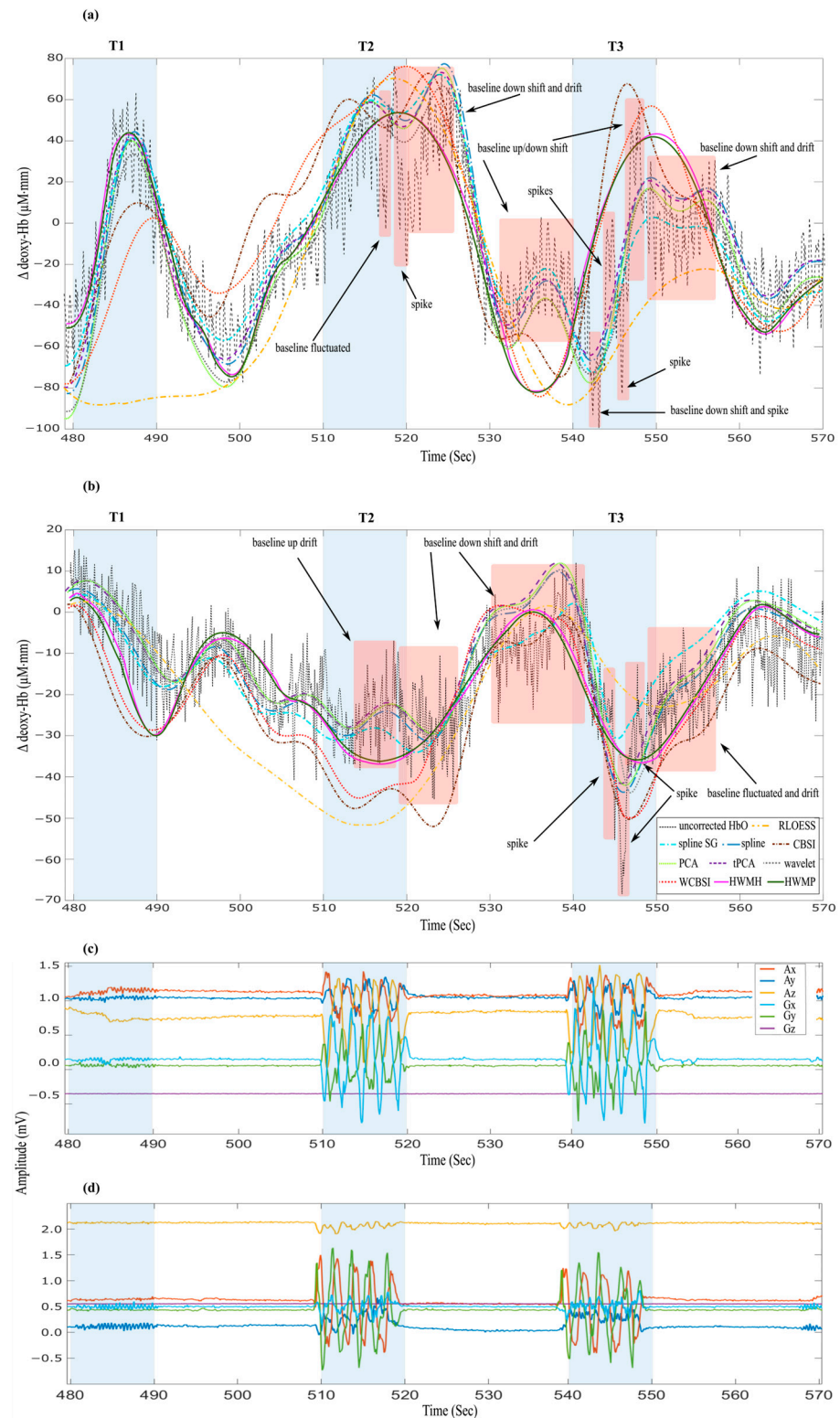

Figure 5. Effects of the different MA correction methods on various types of MAs are marked in red. The intervals of the three tasks (T1, T2, and T3) are marked in blue. Single-trial time series of (**a**) HbO,

(**b**) HbR, (**c**) the probe-IMU, and (**d**) the head-IMU are shown. The uncorrected signal is shown as a black dashed line. Compared to all other methods, HWMH and HWMP correction (pink and green solid lines) exhibit remarkable efficacy in mitigating all types and magnitudes of MA for both hemoglobin signals during the T2 and T3 intervals. The corrected signals generated by the other MA corrections are shown as dashed lines with different colors: spline (blue), spline SG (cyan), RLOESS (orange), tPCA (purple), CBSI (brown), wavelet (grey), PCA (green), and WCBSI (red).

3.2. Quantitative Comparison of Motion Correction Methods

All measures of MA correction quality are referenced to the average tapping-related fNIRS activations measured in the T1 epochs without head movement. Consequently, no MA correction was applied to obtain T1 averages. We consider them to be the empirical ground truth to which we compare the MA-contaminated and the MA-corrected fNIRS responses measured in the T2 and T3 epochs with small and large head movements, respectively. As depicted in Figure 4, averages for comparison with the ground truth signal were taken after MA correction. To quantify the results, we calculated four distinct quality metrics, capturing different aspects of signal quality (see the Methods Section): the SNR capturing the residual noise level, ΔAUC, the RMSE capturing the magnitude and shape reconstruction combined, and the Pearson correlation coefficient R capturing shape reconstruction alone. The epochs to calculate the metrics were 20 s long and started with the beginning of the respective tasks.

Our main interest was in the comparison of the performance of our novel nonlinear HWM MA correction approach with the uncorrected average epochs (i.e., averaging as the only MA treatment) and the linear WCBSI correction that we recently suggested [17] and proved to be superior to all alternative MA correction methods tested in the previous study (spline, spline SG, RLOESS, CBSI, PCA, tPCA, and wavelet) and here. Therefore, in order to reduce the number of statistical comparisons among MA correction methods, we focused on testing differences between the HWM- and WCBSI-corrected signals. Furthermore, our results indicated that MA correction success is virtually identical with the head-IMU (HWMH) and the probe-IMU (HWMP). Therefore, we averaged the quality measures over the two corrections into a single value for the HWM. We used this average value for the statistical comparisons.

All tests were conducted as two-tailed paired *t*-tests. In the plots below, we use the symbol (*) to indicate significance levels $p < 0.05$ (**) for $p < 0.01$ and (***) for $p < 0.001$ [11]. We added Benjamini–Hochberg-corrected *p*-values for a false discovery rate (FDR) of 5% with $n = 8$ comparisons (two signals (HbO and HbR), two movement levels (T2 and T3), and two methods (WCBSI and HWM) for each quality measure and calculated Cohen's d as a measure of the effect size.

Signal-to-noise ratio (SNR) relates the power of residual noise after MA correction to the power in the ground truth signals. A higher SNR value indicates a better MA correction. Figure 6 summarizes the averaged SNR (17 participants) for the HbO and HbR signals after MA correction in tasks T2 and T3 for each correction approach. The SNRs differ very little between HWM types (all two-tailed paired *t*-tests with df = 16: HbO T2: $p > 0.05$, Cohen's d = 0.451; HbO T3: $p > 0.05$, Cohen's d = 0.287, HbR T2: $p > 0.05$, Cohen's d = 0.448; HbR T3: $p > 0.05$, Cohen's d = 0.262). Therefore, we averaged the HWMP and HWMH SNRs for HWM correction in the following analysis.

Compared to uncorrected signals, ΔAUC decreased significantly with HWM correction for both HbO and HbR in T2 and T3 (all two-tailed paired *t*-tests with df = 16: HbO T2: $p < 0.001$, Cohen's d = 1.011; HbO T3: $p < 0.001$, Cohen's d = 1.195; HbR T2: $p < 0.01$, Cohen's d = 0.730; HbR T3: $p < 0.001$, Cohen's d = 2.180). All effects remained statistically significant after Benjamini–Hochberg correction for n = 8 comparisons at an FDR of $p < 0.05$ (corrected *p*-values: HbO T2: $p < 0.05$, HbR T2: $p < 0.05$, HbO T3: $p < 0.001$, HbR T3: $p < 0.01$).
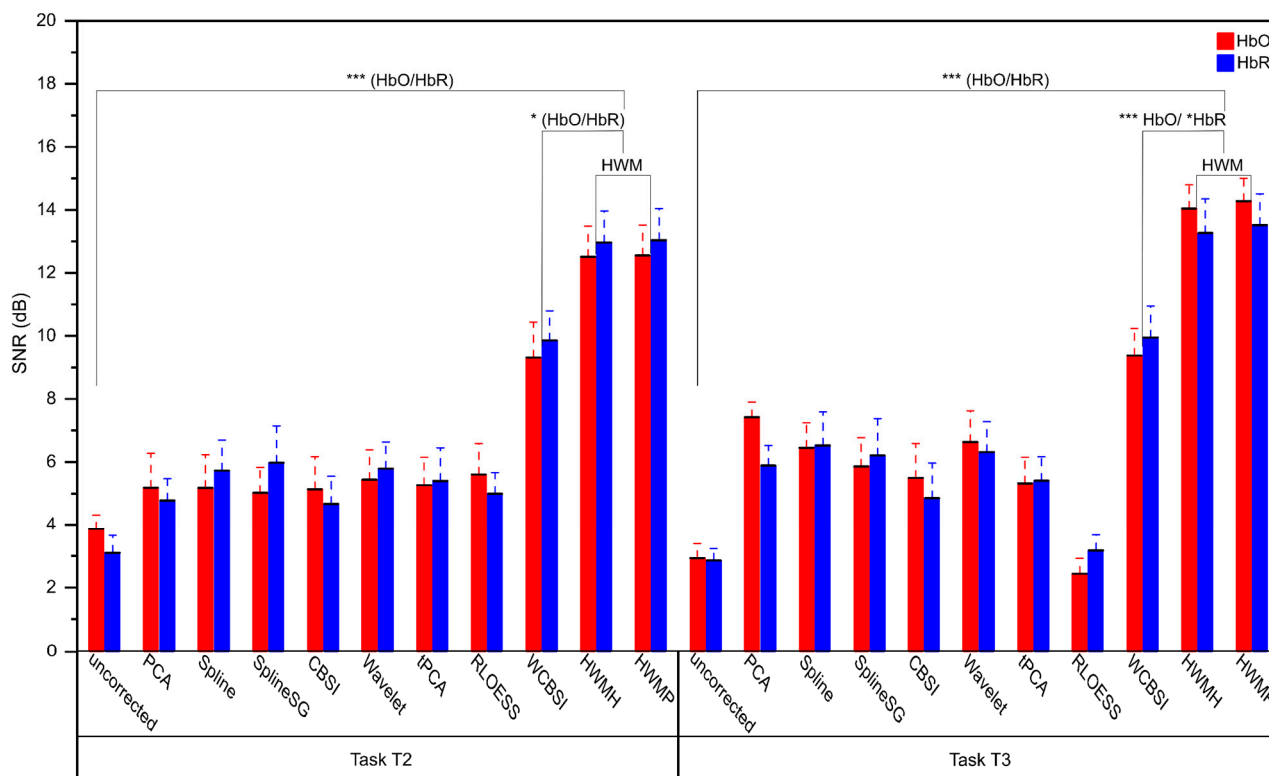
Figure 6. Mean signal-to-noise ratio (SNR) with standard errors after MA correction for the different correction methods plus the SNR of the uncorrected data. SNRs for HbO and HbR signals are shown in red and blue, respectively. The left panel shows SNRs for experimental tasks T2 (small head movements) and the right panel for T3 (large head movements), with a significance level of $p < 0.001$. Note that HWMH and HWMP have superior SNRs compared to all other methods tested. Three stars (***) indicate a significance level of $p < 0.001$, and one star (*) $p < 0.05$.

Compared to uncorrected signals, the SNR increased significantly with HWM correction for both HbO and HbR in T2 and T3 (all two-tailed paired *t*-tests with df = 16: HbO T2: $p < 0.001$, Cohen's d = 1.918; HbO T3: $p < 0.001$, Cohen's d = 3.620; HbR T2: $p < 0.001$, Cohen's d = 2.012; HbR T3: $p < 0.001$, Cohen's d = 2.180). All the comparisons are still significant after Benjamini–Hochberg correction for n = 8 comparisons at an FDR of $p < 0.05$ (corrected *p*-values: HbO T2: $p < 0.001$; HbO T3: $p < 0.001$; HbR T2: $p < 0.001$; HbR T3: $p < 0.001$).

Compared to WCBSI, the SNR increased numerically with HWM correction for both HbO and HbR in T2 and T3 (see Figure 6). This difference reached significance in the uncorrected paired *t*-tests (all two-tailed paired *t*-tests with df = 16: HbO T2: $p < 0.05$, Cohen's d = 0.530; HbO T3: $p < 0.001$, Cohen's d = 1.46; HbR T2: $p < 0.05$, Cohen's d = 0.692; HbR T3: $p < 0.05$, Cohen's d = 0.700). Also, after Benjamini–Hochberg correction for n = 8 comparisons at an FDR of $p < 0.05$, the measures were significantly different (corrected *p*-values: HbO T2: $p < 0.05$; HbO T3: $p < 0.001$; HbR T2: $p < 0.05$; HbR T3: $p < 0.05$).

In sum, we obtained for both head movement levels the highest SNRs with the HWM correction. The SNR reached up to 13–14 dB for both HWMH and HWMP, which amounts to a ratio of ca 20:1 to 25:1.

The area under the curve difference (ΔAUC) measures the difference between the integrals of the ground truth signal and the MA-corrected signals. This is a global measure that can, for example, detect signal rescaling, and smaller values indicate better MA correction. Reported alone, it can be hard to interpret. Figure 7 shows the ΔAUCs for the different correction approaches for HbO and HbR separately. Note that the HbR signals were multiplied by −1 to compensate for the sign flip. The ΔAUC did not show significant differences between HWMP and HWMH correction (all two-tailed paired *t*-tests with

df = 16: HbO T2: $p > 0.05$, Cohen's d = 0.316; HbO T3: $p > 0.05$, Cohen's d = 0.041; HbR T2: $p > 0.05$, Cohen's d = 0.139; HbR T3: $p > 0.05$, Cohen's d = 0.141). Therefore, we averaged the HWMP and HWMH ΔAUCs for HWM correction in the following analysis.
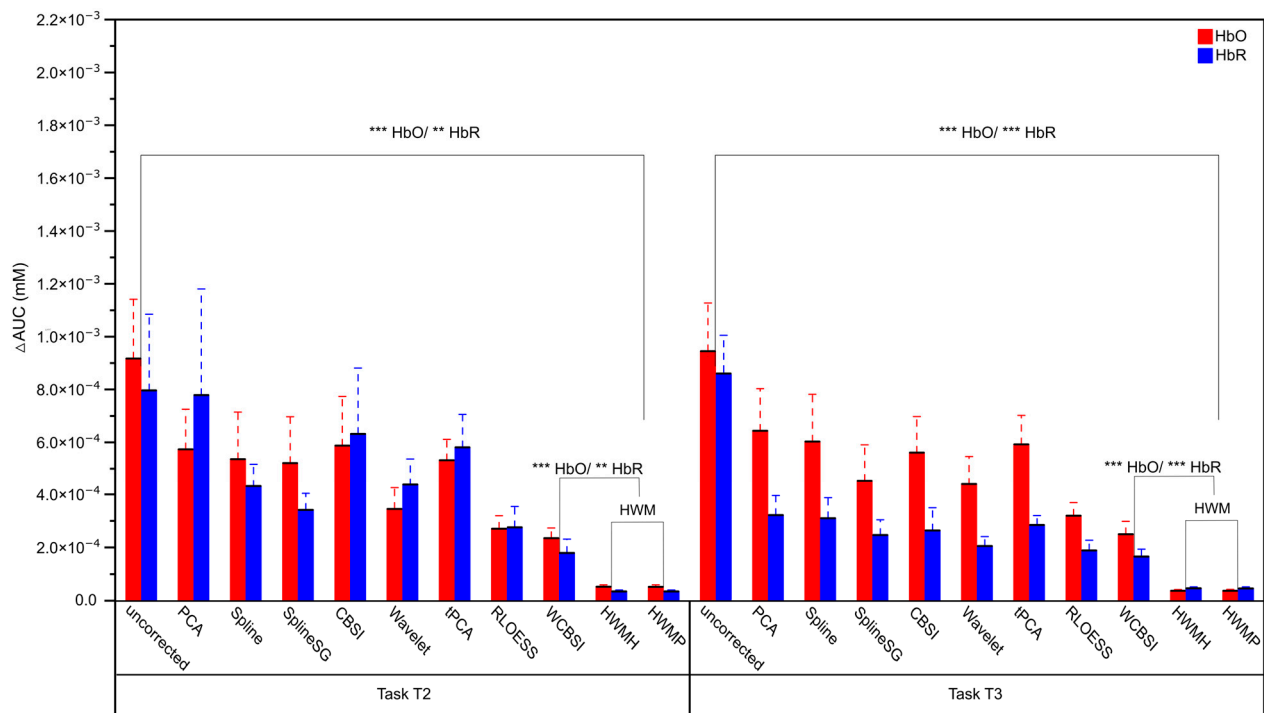

Figure 7. Mean area under the curve difference (ΔAUC) with standard errors after MA correction for the different correction methods plus the uncorrected data. ΔAUCs for HbO and HbR signals are shown in red and blue, respectively. The left panel shows results for experimental tasks T2 (small head movements) and the right panel for T3 (large head movements). Three stars (***) indicate a significance level of $p < 0.001$, and two stars (**) $p < 0.01$. Note that HWMH and HWMP corrections produce fNIRS signals with the smallest AUC deviation from the ground truth signal.

Compared to WCBSI, ΔAUC decreased numerically with HWM correction for both HbO and HbR in T2 and T3 (see Figure 7). This difference reached significance in the uncorrected paired *t*-tests (all two-tailed paired *t*-tests with df = 16: HbO T2: $p < 0.001$, Cohen's d = 1.120; HbO T3: $p < 0.001$, Cohen's d = 1.081; HbR T2: $p < 0.01$, Cohen's d = 0.708; HbR T3: $p < 0.001$, Cohen's d = 0.964), After Benjamini–Hochberg correction for n = 8 comparisons at an FDR of $p < 0.05$, all effects remained significant (corrected *p*-values: HbO T2: $p < 0.05$; HbO T3: $p < 0.05$; HbR T3: $p < 0.05$; HbR T2: $p < 0.05$).

In sum, similar to SNR, the HWM produced for ΔAUC the best MA correction results over the range of head movements investigated (T2 ΔAUC: 5.2×10^{-5} for HbO and 3.5×10^{-5} for HbR; T3 ΔAUC: 3.7×10^{-5} for HbO and 4.6×10^{-5} for HbR).

The Root Mean Square Error (RMSE) quantifies the unscaled averaged absolute discrepancy between the ground truth signal and the MA correction signals. Lower RMSE scores signify more effective MA correction. Figure 8 depicts the RMSE measures of MA correction success in the experimental conditions T2 and T3. There are only minor RMSE differences between HWMP and HWMH (all two-tailed paired *t*-tests with df = 16: HbO T2: $p > 0.05$, Cohen's d = 0.361; HbO T3: $p > 0.05$, Cohen's d = 0.267; HbR T2: $p > 0.05$, Cohen's d = 0.538; HbR T3: $p > 0.05$, Cohen's d = 0.423). Therefore, we calculated the following analysis using the average HWM-correction RMSE score.
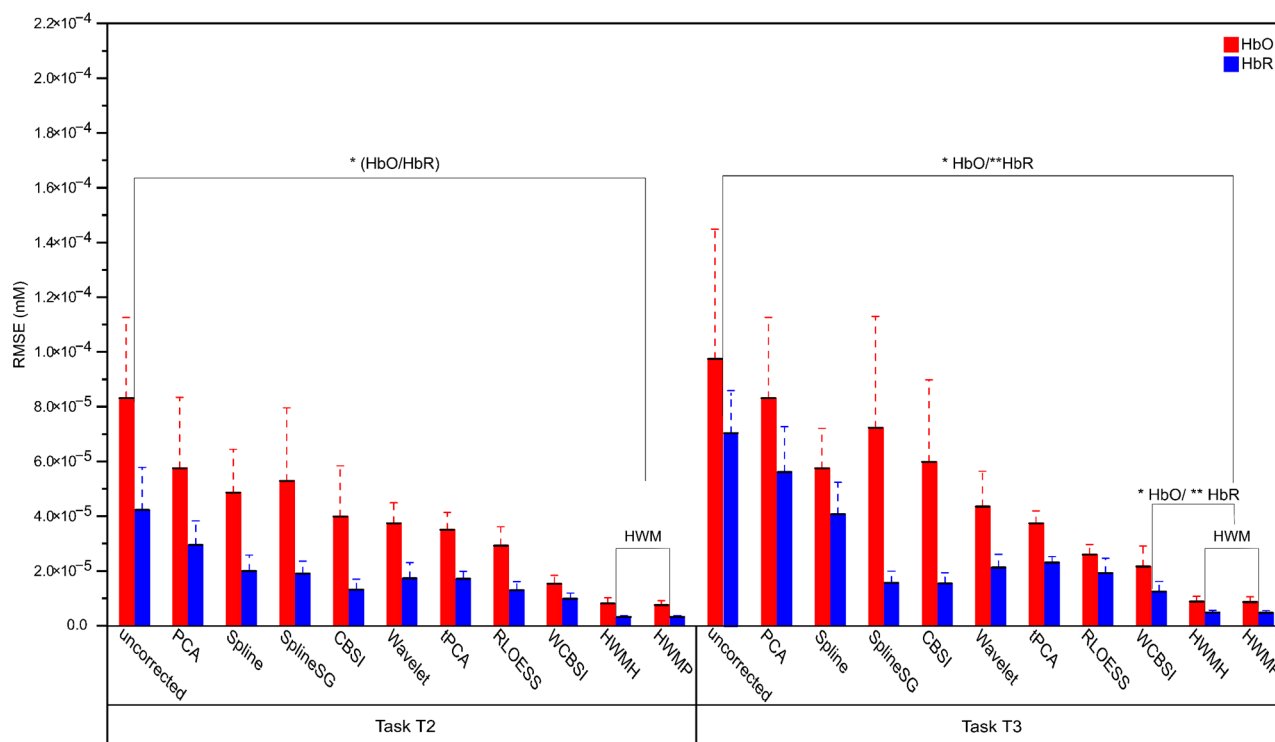
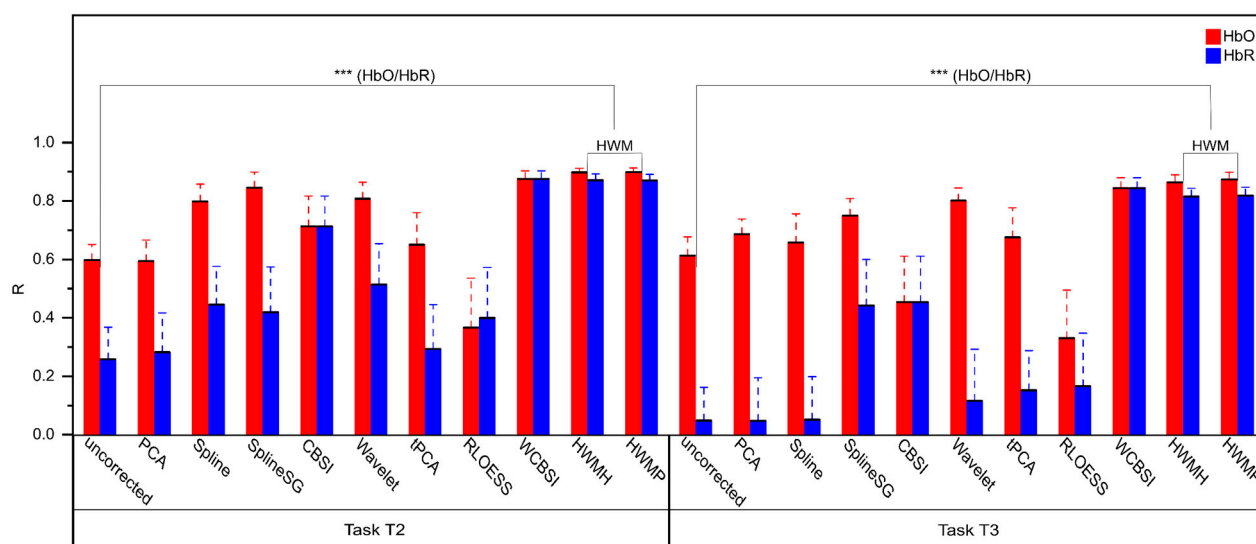**Figure 8.** Mean RMSE with standard errors after MA correction for the different correction methods plus the uncorrected data. RMSEs for HbO and HbR signals are shown in red and blue, respectively. The left panel shows results for experimental tasks T2 (small head movements) and the right panel for T3 (large head movements). HWMH and HWMP corrections produce fNIRS signals, with the smallest RMSE from the ground truth signal. Two stars (**) indicate $p < 0.01$, and one star (*) $p < 0.05$.

Compared to uncorrected signals, the RMSE decreased for both HbO and HbR in T2 and T3 (all two-tailed paired *t*-tests with df = 16: HbO T2: $p < 0.05$, Cohen's d = 0.494; HbO T3: $p < 0.05$, Cohen's d = 0.624; HbR T2: $p < 0.05$, Cohen's d = 0.611; HbR T3: $p < 0.01$, Cohen's d = 0.903). Most of the tests remained significant after Benjamini–Hochberg correction for n = 8 comparisons at an FDR of $p < 0.05$; the comparison measures were significant (corrected *p*-values: T2; HbO T2: $p < 0.05$, HbR T2: $p < 0.05$, and HbR T3: $p < 0.01$). Only the comparison of the HbO RMSE scores for the large head movements (T3) did not pass correction (HbO T3: $p > 0.05$). We consider the HWM-correction effect to be statistically marginal but note that the mean numerical RMSE was smaller for all HWM-corrected than for uncorrected signals (see Figure 8).

Compared to WCBSI, the RMSE decreased numerically with the HWM correction for both HbO and HbR in T2 and T3 (see Figure 8). This difference reached significance in the uncorrected paired *t*-tests only in condition T2, the smaller head movement (all two-tailed paired *t*-tests with df = 16: HbO T2: $p < 0.05$, Cohen's d = 0.593; HbO T3: $p > 0.05$, Cohen's d = 0.421; HbR T2: $p < 0.01$, Cohen's d = 0.870; HbR T3: $p > 0.05$, Cohen's d = 0.470). The same pattern emerges after Benjamini–Hochberg correction for n = 8 comparisons at an FDR of $p < 0.05$: only the differences for T2 remained significant (corrected *p*-values: HbO T2: $p < 0.05$; HbO T3: $p > 0.05$; HbR T2: $p < 0.01$; HbR T3: $p > 0.05$).

In sum, we obtained consistently the best numerical RMSE scores with the HWM correction. However, the differences between the uncorrected and WBCSI-corrected RMSEs tended to be marginal in the larger-head-movement condition T3 and did not survive correction for multiple comparisons.

Pearson's correlation coefficient (R) captures the relative shape similarity between the ground truth signal and the MA-corrected signals. R is a normalized value, where a value of one indicates that the shapes of the two curves are perfect reproductions of each other, and zero indicates that they are unrelated. Note that R is insensitive to scaling differences.

Figure 9 shows the correlation coefficients for the different correction approaches for HbO and HbR separately. The highest correlations are obtained with the proposed HWMH and HWMP approaches. Again, the correlation coefficients obtained with HWMP and HWMH are very similar and do not differ significantly (all two-tailed paired *t*-tests with df = 16: HbO T2: $p > 0.05$, Cohen's d = 0.113; HbO T3: $p > 0.05$, Cohen's d = 0.402; HbR T2: $p > 0.05$, Cohen's d = 0.085; HbR T3: $p > 0.05$, Cohen's d = 0.329). Therefore, we calculated the following analysis with the averaged HWM Pearson correlation coefficient in the following analysis.

Figure 9. Mean Pearson correlation coefficient with standard errors after MA correction for the different correction methods plus the uncorrected data. Rs for HbO and HbR signals are shown in red and blue, respectively. The left panel shows results for experimental tasks T2 (small head movements) and the right panel for T3 (large head movements). Three stars (***) indicate a significance level of $p < 0.001$. Note that HWMH, HWMP, and WCBSI produce MA-corrected fNIRS signals with the highest correlation with the ground truth signal.

Compared to uncorrected signals, R increased significantly for both HbO and HbR in T2 and T3 (all two-tailed paired *t*-tests with df = 16: HbO T2: $p < 0.001$, Cohen's d = 1.498; HbO T3: $p < 0.001$, Cohen's d = 0.941; HbR T2: $p < 0.001$, Cohen's d = 1.329; HbR T3: $p < 0.001$, Cohen's d = 1.696). Also, after Benjamini–Hochberg correction for n = 8 comparisons at an FDR of $p < 0.05$, all comparisons remained significant (corrected *p*-values: HbO T2: $p < 0.001$; HbO T3: $p < 0.01$; HbR T2: $p < 0.001$; HbR T3: $p < 0.001$).

HWM correction achieves similarly accurate shape reconstructions to WCBSI (see Figure 9). The mean R overall conditions are for HWM 0.86 (STD = 0.03) and for WCBSI 0.86 (STD = 0.02). Consequently, there were no significant differences between any conditions (all two-tailed paired *t*-tests with df = 16: HbO T2: $p > 0.05$, Cohen's d = 0.226; HbO T3: $p > 0.05$, Cohen's d = 0.190; HbR T2: $p > 0.05$, Cohen's d = 0.054; HbR T3: $p > 0.05$, Cohen's d = 0.271). Also, after Benjamini–Hochberg correction for n = 8 comparisons at an FDR of $p < 0.05$, no comparison was significant (corrected *p*-values: HbO T2: $p > 0.05$; HbO T3: $p > 0.05$; HbR T2: $p > 0.05$; HbR T3: $p > 0.05$). In addition, Pearson correlations were remarkably similar across head movement levels (T2 HbO: R = 0.89 vs. T3 HbO R = 0.86 and T2 HbR: R = 0.87 vs. T3 HbR R = 0.81).

Given the observation that RMSE values were somewhat lower for the HWM than for WCBSI, and the fact that RMSE is sensitive to scaling and shape whereas Pearson correlation is only sensitive to shape, this suggests that the difference between WCBSI and HWM may be found mostly in scaling differences, at least at the smaller-head-movement levels in T2. Importantly, reliable response-amplitude estimates are key in many analysis approaches that operate on MA-corrected data.

3.3. IMU Data Analysis

3.3.1. Head Orientation Analysis across Experimental Tasks

In this section, we analyze the head motion characteristics in the three experimental conditions. Figure 10 shows the mean pitch (θ), roll (ψ), and yaw (φ) head rotations for the three tasks, derived from the accelerometer data of the head-IMU. As instructed, head orientation remained stable during the T1 task, from which we derived the ground truth fNIRS responses. T2 and T3 show considerable head rotations, and those increased from T2 to T3 (T2: $\theta \sim 10°$, $\psi \sim 7°$, and $\varphi \sim 4°$, T3: $\theta \sim 12°$, $\psi \sim 8°$, and $\varphi \sim 7°$).
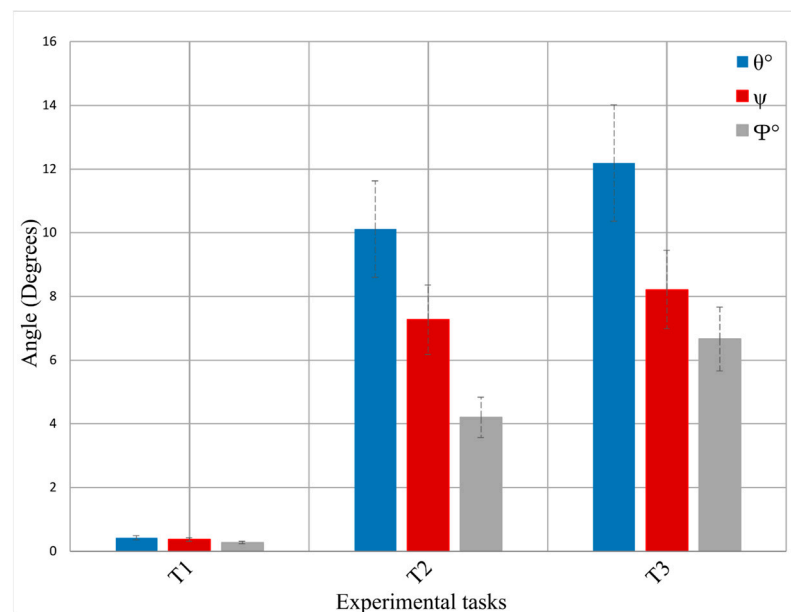
Figure 10. Mean head rotation angles in the three experimental conditions: pitch (θ, blue) is the rotation angle around the *x*-axis, roll (ψ, red) is the rotation angle around the *y*-axis, and yaw (φ, grey) is the rotation angle around the *z*-axis. Error bars indicate the standard error of the mean.

3.3.2. Canonical Correlation Analysis (CCA) of the Head- and Probe-IMU Motion Signals

We originally assumed that the head-IMU and the probe-IMU might measure at least partly independent motion signals, as only the probe-IMU could measure the probe in addition to the head movements. In practice, it turned out that both IMUs combined with the HWM provide indistinguishable motion correction performance. This suggests that the simpler head mount was sufficient for the HWM correction in our recording setup. However, an equal level of MA correction does not necessarily mean that the two IMUs measured comparable motion acceleration signals, as the HWM may adapt to different inputs. Here, we wanted to see if the actual movement signals measured were comparable between the two IMU positions. As the IMUs are attached with different orientations, their sensor channels are not directly comparable. To align them, we used CCA, which finds a common space in which linear combinations of the original sensor channels (the CC factors) are maximally correlated. If both IMUs measured comparable motion signals, CCA should find a transformation for each sensor such that the CCA time series of the head- and probe-IMUs are highly correlated. Figure 11 depicts the mean CCA correlations for the two sensors. The three CCA channels for the accelerometers indicate a very high correlation, indicating that the IMU accelerometers measure virtually identical signals. For the gyroscopes, one CCA channel has a very high correlation and the second a somewhat lower one. The fact that the correlation in the third CCA channel is low is likely due to the nature of the head movements, which rotate around only two axes. Therefore, CCA can only find two correlated rotation axes. In sum, this analysis implies that head- and probe-IMUs measured very similar signals in our fNIRS measurement setup.
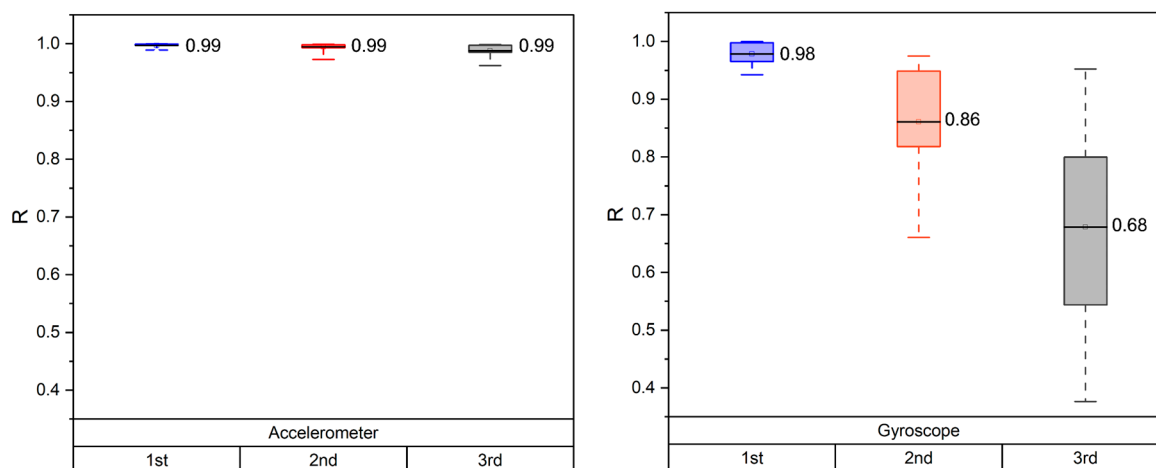
Figure 11. The correlation coefficients of the first three CCA variables are denoted by the colors blue, red, and grey, respectively, for the accelerometer and gyroscope measurements. These coefficients represent the relationships between the head-IMU and probe-IMU in three degrees of freedom.

4. Discussion

The main goal of this study was to develop and test a novel application of the linear–nonlinear Hammerstein–Wiener model for optimizing the MA correction of fNIRS measurements in a realistic experiment and to compare this approach with the most popular MA correction techniques. Notably, our approach provides the best, or at least on par with the "best in class", MA correction, WCBSI, of the other investigated methods in several MA correction quality metrics, for both hemoglobin signals (HbO/HbR) and among a range of movement levels (small and large movement tasks). For both hemoglobin signals and movement tasks, the HWM correction method produces the best SNR of the MA corrected signals among all MA correction methods investigated. The HWM-corrected fNIRS also had the lowest values for differences in AUC and RMSE, indicating that it best recovers the shape and amplitude of the ground truth signal. This exceptional performance, reflected by the lowest value of (ΔAUC) among all other techniques, underscores its pronounced capability to effectively mitigate the high-amplitude spikes and high up–down-baseline shifts by successfully resetting the signal's baseline offset. The lowest RMSE values for HWM-corrected fNIRS indicate minimal errors between the ground truth signal and MA-corrected signals compared to the other tested algorithms. Additionally, the HWM correction significantly improves the Pearson correlation between the MA corrected signal and the empirically measured ground truth signal compared to the uncorrected signal. Interestingly, the WCBSI method, which we previously developed [17], is on par with the HWM correction regarding the Pearson correlation. This suggests that WCBSI recovers the shape of the fNIRS signal with comparable quality to the HWM. However, the lower performance of WCBSI in the other quality measures indicates that it does not recover the amplitude of the signal as well as the HWM. Finally, visual inspection indicates optimal correction of segments with very low-amplitude and slow MAs by HWM, a task beyond the capability of other techniques.

There are two important differences between the HWM MA correction method developed here and the other popular approaches we tested. First, the HWM method uses IMU-based measurements of the head movements and integrates them into the MA correction process. This is different from the other MA correction approaches, which do not use movement information but try to correct MAs based on statistical signal features, employing signal processing techniques, or making assumptions about the relation between HbO and HbR signals. When these assumptions are violated, these techniques may fail to various degrees. Second, HWM correction is a nonlinear technique, while most other techniques are linear. The superior performance of the nonlinear HWM correction method indicates that the linearity of the MA is a strong assumption that is likely violated in practice. Different types of movement artifacts may not only have different dynamics,

but also different amplitude scaling behavior, which is not possible to capture with linear methods. This may be the reason why the linear methods are good at correcting some types of MA but not others, as indicated by the examples in Figure 5.

The HWM method was insensitive to differences in the placement of the IMUs used for correction in our fNIRS measurement setup, in which we used a standard cap with optode holders. This suggests that IMU placement can be solved in a relatively simple way by attaching it to the optode holder cap. An open question is whether the use and integration of multiple IMUs can further improve the HWM correction. In principle, the approach can handle inputs from multiple IMUs. However, the CCA analysis indicates that multiple IMUs may provide redundant information. In this case, no further improvement can be expected.

The advantage of the HWM method lies in its ability to estimate the MA based on IMU data by using the linear–nonlinear Hammerstein–Wiener model's integration of the participants' head movement data and the fNIRS data. This allows for precise quantification and removal across a wide range of movement levels. Another advantage of the HWM method is that it is a fully automated correction technique that does not require user-defined parameter entry for each movement level. Other MA correction methods rely on the manual selection of appropriate threshold parameters that must be adjusted based on the level of MA contamination before MA correction. Importantly, opposed to the "best in class" WCBSI, the HWM method corrects the HbR signal independently of the HbO signal. This way, it provides independent HbR and HbO estimates. Conversely, (W)CBSI combines the HbR and HbO signals in one synthetic signal that can be expressed as HbO or as HbR, but both synthetic signals are fully linearly dependent. Another important advantage is that the HWM method performs comparably over the movement levels employed here. This robustness against MA-level variations is consistent with [18,42].

A limitation of this study is that all processing was conducted offline. Although most fNIRS studies implement offline analyses in some use cases, online MA correction may be desirable. However, processing time may become an issue with larger sensor arrays because the HWM correction method is a channel-by-channel processing technique, and processing time scales approximately linearly with the number of fNIRS channels. Table S1 (Supplementary Materials) shows that the processing time for a single fNIRS channel consisting of HbO and HbR signals amounts to approximately 30 s for 40 min of data. This is comparable to most other techniques (16–690 s). However, for whole-head recordings with more than 100 fNIRS channels, the processing time may become considerable. Notably, the HWM correction was faster than that of wavelet, WCBSI, and RLOESS, which took 62, 65, and 690 s, respectively. However, future investigations should test the capability to perform online corrections. Another limitation of our study pertains to the performance of HWM correction in real-world recordings where head movements, e.g., during walking, may be even stronger or MAs have different dynamics. This requires further investigations into new datasets recorded under diverse conditions. However, the observation that the HWM can correct MAs caused by different levels of head movements and a range of different types of MAs without a loss in accuracy opens the possibility that it could also produce favorable results in less controlled fNIRS recording settings. Finally, the HWM correction requires additional head movement recordings, for example, from an IMU. These can be obtained with a commercial movement-tracking system or, as was done in this study, with a relatively cheap custom-made system. The MPU6050 three-axis accelerometer and gyroscope IMU used here are relatively cheap consumer-grade devices that were read at the low sampling rate of the fNIRS device with a cheap standard Arduino Uno microcontroller board. The code for the Arduino board is provided in the Supplementary Materials. With that, it should be possible to reproduce our head motion measurement setup with minimal soldering skills.

5. Conclusions

In the current study, we developed and tested a novel application of the nonlinear Hammerstein–Wiener model for combining the IMU-based head movement recordings with fNIRS recordings for the correction of fNIRS motion artifacts. According to four commonly used quality metrics (SNR, ΔAUC, RMSE, and R), the IMU-based HWM approach significantly improves MA removal compared to several popular MA correction techniques. Furthermore, we found that the placement of the IMU sensor on the participant's head or on the top of the fNIRS probe did not significantly affect the performance of the HWM approach. Based on our results, we conclude that combining HWM with IMU-based motion data is the most favorable approach to improve the quality and reliability of the MA correction of fNIRS signals.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/s24103173/s1, Arduino codes; Table S1: The processing time for each tested technique; and Table S2: The participant's age, gender, and hair colour.

Abbreviations

ABAMAR	Accelerometer-based motion artifact removal
ARX	Exogenous auto-regressive model
AUC	Area under the curve
BPF	Bandpass filter
CBSI	Correlation-based signal improvement
CCA	Canonical correlation analysis
DPF	Differential pathlength factor
FDR	False discovery rate
fNIRS	Functional near-infrared spectroscopy
HbO	Oxyhemoglobin
HbR	Deoxyhemoglobin
HRF	Hemodynamic response function
HWM	Hammerstein–Wiener model
HWMH	Hammerstein–Wiener model of the head sensor
HWMP	Hammerstein–Wiener model of the probe sensor
Hz	Hertz
IMU	Inertial measurement unit
IRB	Institutional Review Board

LCD	Liquid-crystal display
MA	Motion artifact
NIR	Near infrared
nm	nanometers
OD	Optical density
PCA	Principal component analyses
R	Pearson correlation coefficient
RLOESS	Robust locally estimated scatterplot smoothing
RMSE	Root Mean Square Error
SG	Savitzky–Golay
SI	Spline interpolation
SNR	Signal-to-noise ratio
tPCA	Targeted principal component analysis
WC	Wavelet correction
WCBSI	Wavelet with correlation-based signal improvement

References

1. Cui, X.; Bray, S.; Bryant, D.M.; Glover, G.H.; Reiss, A.L. A Quantitative Comparison of NIRS and FMRI across Multiple Cognitive Tasks. *Neuroimage* **2011**, *54*, 2808–2821. [CrossRef]
2. Huppert, T.J.; Diamond, S.G.; Franceschini, M.A.; Boas, D.A. HomER: A Review of Time-Series Analysis Methods for near-Infrared Spectroscopy of the Brain. *Appl. Opt.* **2009**, *48*, D280–D298. [CrossRef]
3. Trende, A.; Unni, A.; Jablonski, M.; Biebl, B.; Lüdtke, A.; Fränzle, M.; Rieger, J.W. Driver's Turning Intent Recognition Model Based on Brain Activation and Contextual Information. *Front. Neuroergon.* **2022**, *3*, 956863. [CrossRef] [PubMed]
4. Villringer, A.; Planck, J.; Hock, C.; Schleinkofer, L.; Dirnagl, U. Near Infrared Spectroscopy (NIRS): A New Tool to Study Hemodynamic Changes during Activation of Brain Function in Human Adults. *Neurosci. Lett.* **1993**, *154*, 101–104. [CrossRef]
5. Gao, L.; Wei, Y.; Wang, Y.; Wang, G.; Zhang, Q.; Zhang, J.; Chen, X.; Yan, X. Hybrid Motion Artifact Detection and Correction Approach for Functional Near-Infrared Spectroscopy Measurements. *J. Biomed. Opt.* **2022**, 27. [CrossRef]
6. Janani, A.; Sasikala, M. Investigation of Different Approaches for Noise Reduction in Functional Near-Infrared Spectroscopy Signals for Brain–Computer Interface Applications. *Neural Comput. Appl.* **2017**, *28*, 2889–2903. [CrossRef]
7. Noponen, T.E.J.; Kotilahti, K.; Nissilä, I.; Kajava, T.; Meriläinen, P.T. Effects of Improper Source Coupling in Frequency-Domain near-Infrared Spectroscopy. *Phys. Med. Biol.* **2010**, *55*, 2941–2960. [CrossRef] [PubMed]
8. Huang, R.; Qing, K.; Yang, D.; Hong, K.S. Real-Time Motion Artifact Removal Using a Dual-Stage Median Filter. *Biomed. Signal Process. Control* **2022**, *72*, 103301. [CrossRef]
9. Izzetoglu, M.; Chitrapu, P.; Bunce, S.; Onaral, B. Motion Artifact Cancellation in NIR Spectroscopy Using Discrete Kalman Filtering. *Biomed. Eng. Online* **2010**, *9*, 934–938. [CrossRef]
10. Jahani, S.; Setarehdan, S.K.; Boas, D.A.; Yücel, M.A. Motion Artifact Detection and Correction in Functional Near-Infrared Spectroscopy: A New Hybrid Method Based on Spline Interpolation Method and Savitzky–Golay Filtering. *Neurophotonics* **2018**, *5*, 1. [CrossRef]
11. von Lühmann, A.; Boukouvalas, Z.; Müller, K.-R.; Adalı, T. A New Blind Source Separation Framework for Signal Analysis and Artifact Rejection in Functional Near-Infrared Spectroscopy. *Neuroimage* **2019**, *200*, 72–88. [CrossRef]
12. Cui, X.; Bray, S.; Reiss, A.L. Functional near Infrared Spectroscopy (NIRS) Signal Improvement Based on Negative Correlation between Oxygenated and Deoxygenated Hemoglobin Dynamics. *Neuroimage* **2010**, *49*, 3039–3046. [CrossRef]
13. Molavi, B.; Dumont, G.A. Wavelet-Based Motion Artifact Removal for Functional near-Infrared Spectroscopy. *Physiol. Meas.* **2012**, *33*, 259–270. [CrossRef]
14. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
15. Cleveland, W.S. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836. [CrossRef]
16. Di Lorenzo, R.; Pirazzoli, L.; Blasi, A.; Bulgarelli, C.; Hakuno, Y.; Minagawa, Y.; Brigadoi, S. Recommendations for Motion Correction of Infant FNIRS Data Applicable to Multiple Data Sets and Acquisition Systems. *Neuroimage* **2019**, *200*, 511–527. [CrossRef]
17. Al-Omairi, H.R.; Fudickar, S.; Hein, A.; Rieger, J.W. Improved Motion Artifact Correction in FNIRS Data by Combining Wavelet and Correlation-Based Signal Improvement. *Sensors* **2023**, *23*, 3979. [CrossRef]
18. Siddiquee, M.R.; Marquez, J.S.; Atri, R.; Ramon, R.; Perry Mayrand, R.; Bai, O. Movement Artefact Removal from NIRS Signal Using Multi-Channel IMU Data. *Biomed. Eng. Online* **2018**, *17*, 120. [CrossRef]
19. Virtanen, J.; Noponen, T.; Kotilahti, K.; Virtanen, J.; Ilmoniemi, R.J. Accelerometer-Based Method for Correcting Signal Baseline Changes Caused by Motion Artifacts in Medical near-Infrared Spectroscopy. *J. Biomed. Opt.* **2011**, *16*, 087005. [CrossRef]
20. Pinti, P.; Scholkmann, F.; Hamilton, A.; Burgess, P.; Tachtsidis, I. Current Status and Issues Regarding Pre-Processing of FNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Front. Hum. Neurosci.* **2019**, *12*, 505. [CrossRef]

21. Ferrari, M.; Quaresima, V. A Brief Review on the History of Human Functional Near-Infrared Spectroscopy (FNIRS) Development and Fields of Application. *Neuroimage* **2012**, *63*, 921–935. [CrossRef]

22. Delpy, D.T.; Cope, M.; Van Der Zee, P.; Arridge, S.; Wray, S.; Wyatt, J. Estimation of Optical Pathlength through Tissue from Direct Time of Flight Measurement. *Phys. Med. Biol.* **1988**, *33*, 1433–1442. [CrossRef]

23. Ito, H.; Kanno, I.; Fukuda, H. Human Cerebral Circulation: Positron Emission Tomography Studies. *Ann. Nucl. Med.* **2005**, *19*, 65–74. [CrossRef]

24. Reddy, P.; Izzetoglu, M.; Shewokis, P.A.; Sangobowale, M.; Diaz-Arrastia, R.; Izzetoglu, K. Evaluation of FNIRS Signal Components Elicited by Cognitive and Hypercapnic Stimuli. *Sci. Rep.* **2021**, *11*, 23457. [CrossRef]

25. Yücel, M.A.; Lühmann, A.V.; Scholkmann, F.; Gervain, J.; Dan, I.; Ayaz, H.; Boas, D.; Cooper, R.J.; Culver, J.; Elwell, C.E.; et al. Best Practices for FNIRS Publications. *Neurophotonics* **2021**, *8*, 12101. [CrossRef]

26. Kocsis, L.; Herman, P.; Eke, A. The Modified Beer–Lambert Law Revisited. *Phys. Med. Biol.* **2006**, *51*, N91. [CrossRef]

27. Chen, W.-L.; Wagner, J.; Heugel, N.; Conant, L.; Quirk, B.; Beardsley, S.A.; Whelan, H.T. Functional Near-Infrared Spectroscopy and Its Clinical Application in the Field of Neuroscience: Advances and Future Directions. *Front. Neurosci.* **2020**, *14*, 540710. [CrossRef]

28. Copaci, D.; Moreno, L.; Blanco, D. Two-Stage Shape Memory Alloy Identification Based on the Hammerstein--Wiener Model. *Front. Robot. AI* **2019**, *6*, 83. [CrossRef]

29. Westwick, D.T.; Kearney, R.E. Models of Nonlinear Systems. In *Identification of Nonlinear Physiological Systems*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2003; pp. 57–101, ISBN 9780471722960.

30. Brigadoi, S.; Ceccherini, L.; Cutini, S.; Scarpa, F.; Scatturin, P.; Selb, J.; Gagnon, L.; Boas, D.A.; Cooper, R.J. Motion Artifacts in Functional Near-Infrared Spectroscopy: A Comparison of Motion Correction Techniques Applied to Real Cognitive Data. *Neuroimage* **2014**, *85*, 181–191. [CrossRef]

31. Zhang, Y.; Brooks, D.H.; Franceschini, M.A.; Boas, D.A. Eigenvector-Based Spatial Filtering for Reduction of Physiological Interference in Diffuse Optical Imaging. *J. Biomed. Opt.* **2005**, *10*, 011014. [CrossRef]

32. Yücel, M.A.; Selb, J.; Cooper, R.J.; Boas, D.A. Targeted Principle Component Analysis: A New Motion Artifact Correction Approach for near-Infrared Spectroscopy. *J. Innov. Opt. Health Sci.* **2014**, *7*, 1350066. [CrossRef]

33. Scholkmann, F.; Spichtig, S.; Muehlemann, T.; Wolf, M. How to Detect and Reduce Movement Artifacts in Near-Infrared Imaging Using Moving Standard Deviation and Spline Interpolation. *Physiol. Meas.* **2010**, *31*, 649–662. [CrossRef]

34. Boas, D.A.; Dale, A.M.; Franceschini, M.A. Diffuse Optical Imaging of Brain Activation: Approaches to Optimizing Image Sensitivity, Resolution, and Accuracy. *Neuroimage* **2004**, *23*, S275–S288. [CrossRef]

35. Brainard, D.H. The Psychophysics Toolbox. *Spat. Vis.* **1997**, *10*, 433–436. [CrossRef]

36. Pfeifer, M.D.; Scholkmann, F.; Labruyère, R. Signal Processing in Functional Near-Infrared Spectroscopy (FNIRS): Methodological Differences Lead to Different Statistical Results. *Front. Hum. Neurosci.* **2018**, *11*, 641. [CrossRef]

37. Jurcak, V.; Tsuzuki, D.; Dan, I. 10/20, 10/10, and 10/5 Systems Revisited: Their Validity as Relative Head-Surface-Based Positioning Systems. *Neuroimage* **2007**, *34*, 1600–1611. [CrossRef]

38. Ljung, L. *System Identification*; Prentice Hall: Upper Saddle River, NJ, USA, 1999.

39. Scholkmann, F.; Wolf, M. General Equation for the Differential Pathlength Factor of the Frontal Human Head Depending on Wavelength and Age. *J. Biomed. Opt.* **2013**, *18*, 105004. [CrossRef]

40. Yücel, M.A.; Selb, J.; Aasted, C.M.; Lin, P.-Y.; Borsook, D.; Becerra, L.; Boas, D.A. Mayer Waves Reduce the Accuracy of Estimated Hemodynamic Response Functions in Functional Near-Infrared Spectroscopy. *Biomed. Opt. Express* **2016**, *7*, 3078–3088. [CrossRef]

41. Tuck, K. *Tilt Sensing Using Linear Accelerometers*; Freescale Semiconductor Application Note AN 3107; Freescale Semiconductor, Inc.: Austin, TX, USA, 2007.

42. Izzetoglu, M.; Devaraj, A.; Bunce, S.; Onaral, B. Motion Artifact Cancellation in NIR Spectroscopy Using Wiener Filtering. *IEEE Trans. Biomed. Eng.* **2005**, *52*, 934–938. [CrossRef]