Get rid of the beat in mobile EEG applications: A framework towards automated cardiogenic artifact detection and removal in single-channel EEG

Journal Article

Author(s):

Chiu, Neng-Tai; [Huwiler, Stephanie](#) ; Ferster, Maria L.; [Karlen, Walter](#) ; Wu, Hau-Tieng; Lustenberger, Caroline

Publication date:

2022-02

Permanent link:

<https://doi.org/10.3929/ethz-b-000517346>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Originally published in:

Biomedical Signal Processing and Control 72, <https://doi.org/10.1016/j.bspc.2021.103220>

Get rid of the beat in mobile EEG applications: A framework towards automated cardiogenic artifact detection and removal in single-channel EEG

Neng-Tai Chiu^a, Stephanie Huwiler^b, M. Laura Ferster^c, Walter Karlen^c, Hau-Tieng Wu^{d,e,f,1,*}, Caroline Lustenberger^{b,g,h,1,2,*}

^a School of Medicine, National Yang Ming Chiao Tung University, Taiwan

^b Neural Control of Movement Lab, Institute of Human Movement Sciences and Sport, Department of Health Sciences and Technology, ETH Zurich, 8092 Zurich, Switzerland

^c Mobile Health Systems Lab, Institute of Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zurich, 8092 Zurich, Switzerland

^d Department of Mathematics, Duke University, Durham, NC 27708, USA

^e Department of Statistical Science, Duke University, Durham, NC 27708, USA

^f Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan

^g Neuroscience Center Zurich, ETH Zurich and University of Zurich, Zurich, Switzerland

^h Sleep & Health Zurich, University Center of Competence, University of Zurich, Zurich, Switzerland

ARTICLE INFO

Keywords:

Electroencephalogram
Cardiogenic artifact
Automated artifact removal
Mobile technology
Machine learning

ABSTRACT

Brain activity recordings outside clinical or laboratory settings using mobile EEG systems have gained popular interest allowing for realistic long-term monitoring and eventually leading to identification of possible biomarkers for diseases. The less obtrusive, minimized systems (e.g., single-channel EEG, no ECG reference) have the drawback of artifact contamination with varying intensity that are particularly difficult to identify and remove. We developed brMEGA, the first open-source algorithm for automated detection and removal of cardiogenic artifacts using non-linear time-frequency analysis and machine learning to (1) detect whether and where cardiogenic artifacts exist, and (2) remove those artifacts. We compare our algorithm against visual artifact identification and a previously established approach and validate it in one real and semi-real datasets. We demonstrated that brMEGA successfully identifies and substantially removes cardiogenic artifacts in single-channel EEG recordings. Moreover, recovery of cardiogenic artifacts, if present, gives the opportunity for future extraction of heart rate features without ECG measurement.

1. Introduction

Wearable and portable EEG technology emerged in the last decades allowing for brain activity monitoring in a natural setting. Moving from clinical and laboratory settings into in-home settings opens up a vast

amount of opportunities for real-life, long-term and large-scale monitoring of brain activity during different vigilance states and conditions, such as during sleep or while performing specific everyday tasks. Ultimately, this technology has the potential to serve as a digital biomarker for specific disorders, to detect in real-time aberrant brain activity (e.g.,

Abbreviations: AR, artifact residue; brMEGA, beat removal mobile EEG algorithm; CGMH, Chang Gang Memorial Hospital; CWT, continuous wavelet transform; EAS, ensemble average subtraction; EEMD-ICA, ensemble empirical mode decomposition ICA; ICA, independent component analysis; IF, instantaneous frequency; IHR, instantaneous heart rate; mSC, modified spectral concentration; NMSE, normalized mean square error; NLEM, nonlocal Euclidean median; OS, optimal shrinkage; OT, optimal transport; PSNR, peak-signal-to-noise ratio; SANAR, subtraction or shape adaptive nonlocal artifact removal; SC, spectral concentration; scICA, single-channel ICA; SNR, signal-to-noise ratio; SST, synchrosqueezing transform; SVM, support vector machine; TF, time-frequency; TFR, time-frequency representation; TIDIS, Taiwan Integrated Database for Intelligent Sleep; TS, template subtraction.

* Corresponding authors at: Department of Mathematics, Duke University, Durham, NC 27708, USA (H.-T. Wu); Department of Health Sciences and Technology, ETH Zurich, 8092 Zurich, Switzerland (C. Lustenberger).

E-mail addresses: hauwu@math.toronto.edu (H.-T. Wu), caroline.lustenberger@hest.ethz.ch (C. Lustenberger).

¹ Shared correspondence, contributed equally to work.

² Lead contact.

<https://doi.org/10.1016/j.bspc.2021.103220>

Received 20 May 2021; Received in revised form 27 August 2021; Accepted 30 September 2021

Available online 1 November 2021

1746-8094/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

epileptic seizures), to better understand the functional role of brain oscillations in behavior and restoration, and to enable brain-computer interfaces in-home settings (e.g., rehabilitation, neurofeedback [1–4]).

In order to integrate this technology in real-life settings, the developed systems thrive for reduced obtrusiveness. This is achieved with miniaturization, reduction of EEG electrodes to a minimal number (sometimes only containing one EEG derivation), and localization of the electrodes to easy reachable places on the head (e.g., electrodes on forehead, mastoid, around-ear, and in-ear [5–7]). While reduction of electrodes increases comfort to the user it also comes with different drawbacks. One of them is the increased difficulty to separate real brain signal from noise and artifacts. EEG can be strongly contaminated by different artifacts, either from physiological (e.g., cardiogenic, sweating) and non-physiological sources (e.g., power-line noise) [8]. Especially when the artifacts have components in the brain activity frequency ranges of interest, they need to be removed for an accurate estimation of underlying brain activity. Cardiogenic artifacts and other artifacts such as eyeblinks provide redundant information that is present in multiple channels of the EEG data recorded with a multi-channel system. Having this information from several channels simplifies the separation of information from different sources and consequently the separation of artifact and EEG information. Removing artifact from single-channel EEG data is therefore more challenging as compared to multi-channel EEG.

The cardiogenic artifact in EEG can occur when cardiac electrical fields affect the surface potentials on the scalp [9]. It can arise in all vigilance states, e.g., also during sleep that is otherwise less affected by movement artifacts or eye blinks. The cardiogenic artifact has frequency components in the delta-theta range but due to its non-sinusoidal shape, the artifact is also represented in higher harmonics of the initial band leading to artifacts across the whole power spectrum. The affected frequency bands are of fundamental interest in neuroscientific research (e.g., slow-frequency activity in sleep, theta activity in wakefulness). Furthermore, cardiogenic artifacts are typically observable as a QRS shaped form, yet they can strongly vary in size and form. This adds an additional layer of difficulty in successfully detecting and removing them [9].

Removal of cardiogenic artifacts necessitates (a) determination of existence of cardiogenic contamination in the EEG; (b) when cardiogenic artifacts exist, determine their locations; (c) removal of the artifact from the EEG (including correctly identifying cardiogenic artifact locations and separating the artifact from the real brain activity). For (a), to our knowledge, there is very limited work aiming to automatically determine whether cardiogenic artifacts exist when only a single-channel EEG signal is available. Tamburro et al. (2019) and Shao et al. (2009) proposed a framework for multi-channel cardiac artifact detection and removal using independent component analysis (ICA) [10,11]. Other approaches include hybrid blind-source separation and machine learning algorithm [12], polynomial neural network and decision tree [13], and feature selection-based approach [14]. For all these listed approaches multiple EEG channels are also required and are therefore not suitable for single-channel EEG applications. For (b), there have been some works with only single-channel EEG applications [15,16]. When the cardiogenic artifacts are *known to exist*, different approaches have been used to define the presence of the QRS complex without ECG reference including the energy interval histogram [15], a modified S-transform approach [16], and the continuous wavelet transform algorithm (CWT) [17]. Yet, all the approaches did not provide an application to tackle (a) and only well-controlled recording segments with known ECG artifact presence were included. We emphasize that while (a) seems to ask the same problem as (b), they are essentially different. For (c), several methods have been proposed to remove the cardiogenic artifact in EEG specifically when there is more than one EEG channel, or when at least an extra ECG reference channel is available [18–21]. To date, there is limited work when only a single EEG channel was recorded. The proposed methods to remove cardiogenic artifacts in

single-channel EEG could be classified into two categories. The first category includes algorithms related to the template subtraction (TS), or ensemble average subtraction [15,16,22] when the locations of cardiogenic artifacts are known; for example, the modified template subtraction by weighting [15] or shape adaptive nonlocal artifact removal (SANAR) [23]. The second category includes algorithms generalizing ICA to the single-channel setup; for example, the single-channel ICA (scICA) [24], the wavelet ICA [25], and the ensemble empirical mode decomposition (EEMD) ICA is combined in Mijovic et al. (2010) [26]. While these ICA based algorithms might work, they are limited by the need of visual inspection to select channels for the cardiogenic artifact removal and sometimes machine learning tools are considered to automatize the mission of artifact component selection. Also, it is not easy to avoid the mode mixing issue when the wavelet or EEMD decomposition is applied. Moreover, the lack of theoretical justification of EEMD further limits its application for scientific research.

Correct removal of artifacts initially requires identification of participants and epochs that are contaminated by ECG. Even though similar EEG montages might be used, not all participants will show artifacts. Furthermore, specifically for long-lasting recordings (e.g., nocturnal sleep period), these artifacts might only temporarily be present and will presumably vary in their magnitude and shape across the recording period. Specifically, for large-scale and long-term recordings that are collected with mobile applications, visual inspection (which is typically involved) is not feasible.

Collectively, there is still lack for a framework that involves automated identification and removal of cardiogenic artifacts from single-channel EEG that have been collected over longer recording durations and without an ECG reference signal. Yet, such a framework is urgently needed, given that mobile, non-obtrusive EEG applications have gained tremendous interest in research. In order to fill this gap, we combine here for the first time nonlinear-type time–frequency analysis tools and machine learning to automatically determine whether a cardiogenic artifact exists and to remove it from single-channel EEG without the need of an ECG reference. We first train a machine learning algorithm based on novel features extracted from nonlinear-type time–frequency analysis tools to determine whether cardiogenic artifacts exist. Second, we apply a manifold learning algorithm similar to SANAR [23] to remove cardiogenic artifacts. The algorithm is modified to handle the specific characteristics of cardiogenic artifact. Unlike the stimulation artifact removal mission, the locations of cardiogenic artifacts are in general unknown. We coin our algorithm *beat removal mobile EEG algorithm* (brMEGA). We apply brMEGA to long-term, nocturnal EEG recordings that were obtained from a portable device in a participant administered in-home study and validate our framework in datasets for which ECG reference channels exist. We demonstrate that contaminated EEG can be successfully identified and substantially restored with our proposed framework.

2. Methods

2.1. Experimental model and subject details

2.1.1. Single-channel EEG recordings from portable device

Thirteen EEG recordings of nocturnal sleep periods from different participants have been selected for this publication. Participants (all healthy, 63–74 years of age, 4 female) were part of an ongoing clinical trial (NCT03420677) that has been approved by the provincial ethics committee and the Swiss Agency for Therapeutic Products Swissmedic. All participants gave written informed consent prior to participation. The EEG data reported here was collected with the MHSL-Sleepband, a mobile and configurable system for high-quality biosignal recordings at home [6]. Specifically, participants used this system independently at home and therefore applied single-use auto-adhesive electrodes (Neuroline 720, Ambu A/S, DK) to different electrode positions on the scalp and face to capture EEG, EOG and EMG activity. For the purpose of this

article, we will only focus on the EEG, which was placed at Fpz with a reference on the right Mastoid (M2) and the ground on the left Mastoid (M1). The signal was sampled at 250 Hz.

Our neuroscientist expert visually examined the EEG signal (filtered at 0.5 Hz high-pass and 40 Hz low-pass for visualization purpose only) and checked for cardiogenic artifacts over non-overlapping 4 s. For 13 whole night polysomnograms, the neuroscientist labeled 100 randomly selected 4 s epochs between the first and second hour of each recording. An epoch was labeled as 0 if no clear cardiogenic artifacts could be visually identified. An epoch was labeled as 1 when a cardiogenic artifact was visible.

2.1.2. Control datasets with ECG reference

The control dataset consists of 2 clinical subjects (one 32-year-old male and one 63-year-old female) being evaluated for sleep apnea by the whole night polysomnogram but diagnosed to be free of sleep apnea. The recordings were collected from the sleep center at Chang Gang Memorial Hospital (CGMH), Linkou, Taiwan. At the time of these recordings no informed consent was obtained as they were coming for a medical screening of sleep apnea. The institutional review board of the CGMH approved the study protocol and the re-use of this data retrospectively without the obtained consent. All signals were acquired by a biosignal amplifier system from Philips Alice. These recordings are part of the Taiwan Integrated Database for Intelligent Sleep (TIDIS, <https://tidis.org/>) project, so we refer to this database as TIDIS. The simultaneously recorded EEG and ECG signals were sampled at 200 Hz. For the validation purpose, we select F4M1 channel for one case and O1A2 channel for another case that have visually identifiable cardiogenic artifacts.

2.1.3. Semi-real database with EEG and ECG ground truth

The semi-real dataset consists of 7 simulated over-night recordings. These recordings are from the TIDIS database, referring to recordings from the sleep center at CGMH, Linkou, Taiwan (see details in paragraph 2.1.2). We randomly select 7 EEG recordings without cardiogenic artifacts and generate the semi-real recording in the following way. For the EEG, there are in total 7,146 epochs, and 28 epochs are removed by the following criterion. If there exists an interval longer than 4 s with the EEG magnitude less than or equal to 1.5 μV , that epoch is removed. We preprocess the raw ECG signal by resample it to 250 Hz and then applying a median filter with the window length 0.2 s followed by a low pass filter with the cutoff at 80 Hz. Then, select 20 s intervals with bSQI [27] greater than or equal to 0.95, and we also remove the 20 s intervals if the maximum value of the difference between the background power of that interval and the median height of all peaks in this interval is greater than 200 μV or if the maximum of the difference between any peak and the median height of all peaks in this interval is greater than 120 μV . We concatenate them to form the clean ECG signal. If two consecutive beats have RR interval less than 0.4 s, the first one is removed. Next, we generate the cardiogenic artifact by randomly scaling the peaks in the following way. Let R_i be the height of the i th R peak in the clean ECG, and QS_i be the interval of the QRS interval of the i th heartbeat, and SQ_i be the interval from the S wave of the i th QRS to the Q wave of the $(i + 1)$ th QRS, which includes the i th T wave. Those parameters are illustrated in Fig. 1. We scale the ECG over QS_i by $60h_i/R_i$, where h_i is independently and identically sampled uniformly from [0.8, 1], and scale the ECG over SQ_i by $12h_i/R_i$. We consider this scaling since the T wave is not always there and not strong even if it exists in the cardiogenic artifact contaminated EEG signal. Finally, we sum the simulated cardiogenic artifact and the cardiogenic artifact free EEG signal to form the final semi-real EEG signal.

2.2. Cardiogenic artifact detection and removal algorithm

The algorithm is composed of two main steps and illustrated in Fig. A1 in the Appendix. The first step is determining if an epoch is

Fig. 1. Illustration of ECG parameters used to generate the artificial cardiogenic artifacts. R_i : height of the i th R peak in the clean ECG. QS_i : interval of the QRS interval of the i th heartbeat. SQ_i : interval from the S wave of the i th QRS to the Q wave of the $(i + 1)$ th QRS, which includes the i th T wave.

contaminated by cardiogenic artifacts. The second step is detecting the location of each cardiogenic artifact and removing the cardiogenic artifact. The algorithm we consider is a modification of the SANAR designed for removing the stimulation artifacts during direct cortical stimulation [23]. The main difference is that the periods between two consecutive stimulation artifacts are usually fixed and their times are known, while the periods between consecutive cardiogenic artifacts are time-varying and their times are unknown. Below we detail the algorithm step by step. Denote the digitalized EEG signal as $x_0 \in \mathbb{R}^n$, where we pad the recorded signal by 0 so that the overall length of the EEG signal is of the multiple of 4 s. We assume that the signal is uniformly sampled every $\Delta_t > 0$ second. Without loss of generality, we assume that $1/\Delta_t$ is an integer to simplify the notation and suppose there are $L > 0$, 4 s non-overlapping epochs; that is $n = L \times 4/\Delta_t$.

Step 0: Preprocessing of EEG data for algorithm.

The mean of the EEG signal is removed and an IIR notch filter applied at 50 Hz (or 60 Hz, depending on where the signal is recorded) to suppress the power line interference. Then, we remove the wandering baseline by applying the median filter with a window length of 100 ms, followed by a window smoothing of length 150 ms. Fix $\Delta_t > 0$. If the sampling rate is different from $1/\Delta_t$ Hz, we resample the signal to $1/\Delta_t$ Hz to balance between the computational time and a better recovery of the cardiogenic artifact information [28]. In practice, if the sampling rate is higher than $1/\Delta_t$ Hz, the signal is downsampled to $1/\Delta_t$ Hz, and the reconstructed cardiogenic artifact is upsampled back to the original sampling rate and subtracted from the EEG signal. We comment that if computational time is not an issue, the following algorithm can be applied to the original signal with parameters scaled accordingly. We denote the resulting signal as $x_1 \in \mathbb{R}^n$.

Step 1: Automatic detection of cardiogenic artifact.

In the first step, we automatically determine if a given epoch is contaminated by cardiogenic artifacts or not.

Step 1-1: We first correct the polarity of the signal. To do this, we need to locate peaks in x_1 . We rectify the signal and take the square root of it. Denote the resulting signal as y_1 and detect peaks in y_1 . The peak locations of x_1 are determined by the peak locations of y_1 . If the mean value of peak heights in x_1 is negative, we correct the polarity by flipping the sign of x_1 , so that the peaks in x_1 are positive. After the correction, to strengthen the power of peaks relative to the background brainwave, we set the negative part of x_1 to be zero and keep the positive part of x_1 . Denote the resulting signal x .

Step 1-2: Then we run the nonlinear-type time-frequency (TF) analysis tools, synchrosqueezing transform (SST), on x . We focus on the frequency band $[0, U)$, where $U \leq 1/2\Delta_t$. Note that $1/2\Delta_t$ is the Nyquist rate. Suppose the frequency axis is equally spaced into $m > 0$ bins. Set $\Delta_\xi := U/m$. The resulting time-frequency representation (TFR) by SST is denoted as $S_x \in \mathbb{C}^{n \times m}$ respectively. We then divide S_x into L submatrices associated with L non-overlapping epochs, denoted as $S_{x,1} \dots S_{x,L} \in \mathbb{C}^{(4/\Delta_t) \times m}$.

Step 1-3: According to the well-established theory, the cardiogenic artifact in EEG will be represented as a dominant horizontal curve in the

TFR [29]. See Fig. 2, right panel, for an example of the TFR of an EEG segment contaminated by cardiogenic artifacts by SST. It is clear that when the cardiogenic artifact is strong, we can easily identify a dominant curve. Thus, since the cardiogenic artifact reflects the heart beats, that curve represents the instantaneous heart rate (IHR), and we obtain the IHR of the subject if we can extract that curve. Since we focus on subjects without severe arrhythmia like ventricular fibrillation, we follow the normal physiological constraint and assume that the heart beats at the frequency less than 4 Hz and higher than $1/3$ Hz; that is, the consecutive two beats are separated by at least 250 ms and no longer than 3 s [30]. Under this constraint, in this work, we apply the following curve extraction algorithm to determine the IHR.

$$c^* = \max_{c \in Z^n} \sum_{l=1}^n \log \left[\frac{|S_x(c(l), l)|}{\sum_{i=1}^m \sum_{j=1}^N |S_x(j, i)|} \right] - \lambda \sum_{l=2}^n |c(l) - c(l-1)|^2, \quad (1)$$

where $\lambda > 0$ is the penalty term chosen by the user, $Z = \{ \lfloor 1/3\Delta_\xi \rfloor + 1, \lfloor 1/3\Delta_\xi \rfloor + 2, \dots, 4/\Delta_\xi \}$, $\lfloor 1/3\Delta_\xi \rfloor$ means the largest integer smaller than $1/3\Delta_\xi$, and c^* is a n dim vector representing a curve on the TFR. Z is the set indexing the ticks in the frequency axis. The first term in the function in (1) essentially says that we fit a curve c of length n that is parametrized by the time into the TFR so that it captures the frequency with the highest intensity at each time. The second term imposes a smoothness constraint on the extracted curve; that is, while we allow c to vary from time to time, we expect that the variation is constrained and penalized by λ . As a result, when cardiogenic artifacts exist, the IHR is

captured, and at time $i\Delta_t$ th second the IHR is $c^*(i)\Delta_\xi$ Hz. When there is no cardiogenic artifact, the extracted curve represents the instantaneous frequency associated with the spectral distribution of the EEG as a random process.

With the estimated c^* , we apply the peak tracing algorithm [31] to determine the location and heights of peaks from x . Specifically, we consider the following optimization:

$$\hat{B}_m = \underset{B_m}{\operatorname{argmax}} \left[\sum_{i=1}^{M_n} x(b_i) + \gamma \sum_{i=2}^{M_n} P(b_i, b_{i-1}) \right], \quad (2)$$

where $B_m = \{b_1, \dots, b_{M_n}\}$ is the set of estimated peak locations, $\gamma > 0$ is the penalty term, and

$$P(b_i, b_{i-1}) := - \left(\log_2 \left(\frac{(b_i - b_{i-1})\Delta_t}{1/c_{b_i}^*} \right) \right)$$

Note that $P(b_i, b_{i-1})$ is the term constraining the estimated peak locations so that they follow the estimated curve c_i^* . Indeed, $(b_i - b_{i-1})\Delta_t$ is the period between the estimated $(i-1)$ th and i th peaks, and $1/c_{b_i}^*$ is the estimated instantaneous frequency at the estimated i th peak. Note that when cardiogenic artifacts exist, $\hat{B}_m$ contains the locations of cardiogenic artifacts. Otherwise, we obtain random peaks associated with the EEG signal. Since we do not consider patients with severe arrhythmia, if two detected peaks are separated less than 250 ms, we keep the beat with the larger amplitude and remove the other one.

Fig. 2. The flowchart of the first step of brMEGA. The four designed features are marked from 1 to 4. The first feature is the peak height in the time domain, the second and third features are the features in the time-frequency (TF) domain that are determined by the “optimal transport distance”, and the fourth feature is the total energy around the peak. The features in the time domain depend on the peak tracking algorithm with the extracted instantaneous frequency from the TF representation signals. x_1 : preprocessed raw signal (50 Hz IIR notch filter with removed wandering baseline). y_1 : rectified and square root of x_1 . S_x : Time-frequency representation after the synchrosqueezing transform of y_1 .

Step 1–4 and 1–5: Motivated by the physiological meaning of c^* , we design features and construct an automatic cardiogenic artifact detection algorithm from x and its SST. Overall, we consider the following four features associated with the cardiogenic artifact in each 4-s epoch, and then train a support vector machine (SVM) model [32].

Since the cardiogenic artifact represents itself as a spike with the spike height generally higher than the EEG signal, the first feature we consider is the *peak height* of cardiogenic artifact. See Fig. 2 for an example of the cardiogenic artifacts. For the l th 4-s epoch, we take the median height of all detected peaks in that epoch to be the feature associated with the epoch. We denote this first feature as $X_l(1)$. This feature is obvious from the time domain.

Next, for the l th epoch, we consider the optimal transport (OT) distance, or the earth mover distance to capture the strength of the cardiogenic artifacts from the TF domain as the second feature. To this end, we first prepare some quantities associated with the cardiogenic artifact in the TF domain and take the extracted curve $c^{(1)}$. As is mentioned above, it is associated with the fundamental frequency of the cardiogenic artifact. We apply (1) on $|S_x|$, after masking the extracted curve $c^{(1)}$ by 0, to extract the instantaneous frequency of the first harmonic of the cardiogenic artifact. Denote it as $c^{(2)} \in \mathbb{R}^n$. By iterating this process, we obtain the second and third harmonics of the cardiogenic artifact, denoted as $c^{(3)} \in \mathbb{R}^n$ and $c^{(4)} \in \mathbb{R}^n$. It is critical to note that the instantaneous frequencies of “harmonics” might not exactly be integer multiples of the fundamental frequency. A critical observation of cardiogenic artifact is that while cardiogenic artifacts look similar, their patterns might change from one to the other. This time-varying pattern has been extensively discussed in the literature [33]. With this preparation, we restrict $c^{(1)} \dots c^{(4)}$ on the l th epoch, and denote them as $c_1^{(1)} \dots c_l^{(4)} \in \mathbb{R}^{4/\Delta_t}$. For the k th sampling point of the l th epoch, define the *ideal* time-varying spectrum as a m -dim vector ν_k , so that $\nu_k(c^{(1)}(k)) = |S_{x,l}(k, c^{(1)}(k))| / \sum_{i=1}^4 |S_{x,l}(k, c^{(i)}(k))|$, $\dots$, $\nu_k(c^{(4)}(k)) = |S_{x,l}(k, c^{(4)}(k))| / \sum_{i=1}^4 |S_{x,l}(k, c^{(i)}(k))|$ and other entries 0. We also define the estimated time-varying spectrum associated with the k th sampling point as $\mu_k := |S_{x,l}(k, :)| / \sum_{i=1}^m |S_{x,l}(k, i)|$. Clearly, ν_k and μ_k are both probability measures. Note that when the cardiogenic artifact is strong, the energy will be concentrated around the instantaneous frequencies of the fundamental and harmonic components. Thus, the estimated time-varying spectrum will be close to the ideal time-varying spectrum. Otherwise, they will be different. To quantify this difference at the k th sampling point, we apply the OT distance [34], which is defined as $d_k :=$

$$d_{OT}(\mu_k, \nu_k) := \int_{-\infty}^{\infty} |f_{\mu_k}(x) - f_{\nu_k}(x)| dx \text{ where } f_{\nu_k} \text{ and } f_{\mu_k} \text{ are the cumulative}$$

functions of ν_k and μ_k . Thus, the first feature in the TF domain for the l th epoch, $X_l(2)$, is defined as the median of $\{d_k\}_{k=1}^{4/\Delta_t}$. The second feature in the TF domain for the l th epoch is defined similarly, but we only focus on the fundamental component; that is, for the k th sampling point of the l th epoch, consider a m -dim vector $\check{\nu}_k$, so that $\check{\nu}_k(c^{(1)}(k)) = 1$ and other entries 0, and define the second feature in the TF domain for the l th epoch, denoted as $X_l(3)$, as the median of $\left\{ \check{d}_k \right\}_{k=1}^{4/\Delta_t}$, where $\check{d}_k := d_{OT}(\mu_k, \check{\nu}_k)$. The fourth feature is the total energy over the l th epoch, which is denoted as $X_l(4)$.

With the above four features for the l th epoch, denoted as a 4-dim vector X_l , we apply the kernel SVM with a radial basis function kernel to train an automatic cardiogenic artifact annotation system. The features are standardized by the z-score. For the reproducibility purpose, we use the SVM implementation commonly applied in practice [35].

Step 2: Recover EEG without cardiogenic artifact.

Once we confirm that the EEG signal is contaminated by the cardiogenic artifact, we apply the nonlocal Euclidean median (NLEM)

algorithm to remove the cardiogenic artifact. The NLEM approach belongs to the category of TS, which is particularly useful when we have only single or few channels and the adaptive filter technical and blind source separation tools cannot be applied. It is designed by a critical observation that while the cardiogenic artifacts follow a similar pattern, it does vary from beat to beat. This observation motivated us to introduce the *wave-shape manifold* model [33], which parametrizes all possible patterns of cardiogenic artifacts. By finding all cardiogenic artifacts with a similar pattern, the NLEM helps recover the cardiogenic artifact pattern more efficiently compared with the traditional TS algorithms.

Suppose there are M epochs that are labeled with cardiogenic artifact contamination. Assume the cardiogenic artifacts happen at $t_1 < t_2 \dots < t_N$ determined by the peak tracing algorithm [31]. According to the assumption and the peak tracking algorithm, $t_i - t_{i-1}$ is sufficiently large so that two cardiogenic artifacts are separated by 250 ms. Before applying the OS algorithm, we apply the 5th order butterworth highpass filter with the cutoff at 5 Hz to the EEG signal to suppress the background EEG signal. Denote the resulting EEG signal $\check{x}$. For each t_i , determine a segment that contains the i th cardiogenic artifact by $\check{X}_i$ to be $\check{x}$ over the period $[t_i - L_L, t_i + L_R]$, where $L_L = \lceil 0.2/\Delta_t \rceil$ and $L_R = \lceil 0.5/\Delta_t \rceil$. The parameters L_L and L_R are chosen based on the knowledge of electrophysiological activity of a heart cycle. Since the peak of the cardiac artifact is associated with the ventricular depolarization, L_L is chosen so that the segment includes atrial depolarization and L_R is chosen so that the segment includes ventricular repolarization. Then, form a p by N data matrix $\check{\mathcal{X}}$, where $p = L_R + L_L + 1$ is the length of each segment and the i th column is the i th artifact cycle.

Step 2–1: Suppose we have $\check{\mathcal{X}} = \check{\mathcal{A}} + \check{\mathcal{E}}$, where the i th column of $\check{\mathcal{A}}$ and $\check{\mathcal{E}}$ is $\check{A}_i$ and $\check{E}_i$ respectively, where $\check{A}_i$ is the cardiogenic artifact and $\check{E}_i$ is the EEG signal. Note that after the 5 Hz highpass filter, the EEG signal $\check{E}_i$ has been suppressed while $\check{A}_i$ is kept less perturbed. Then, perform the OS algorithm with the Frobenius norm [36] on $\check{\mathcal{X}}$. The OS result of $\check{\mathcal{X}}$, denoted as $\tilde{\mathcal{X}}$, is the optimal approximation of $\check{\mathcal{A}}$ in the sense of the Frobenius norm. Thus, we can view the i th column of $\tilde{\mathcal{X}}$ as a good estimator of the i th cardiogenic artifact. We can then define the following “metric” to find those EEG segments that have the same or similar cardiogenic artifacts: $d(X_i, X_j)$ is defined as the diffusion distance [37] between $\tilde{X}_i$ and $\tilde{X}_j$, where $\tilde{X}_i$ is the i th column of $\tilde{\mathcal{X}}$.

Step 2–2: Finally, we can apply the NLEM. For the i th EEG segment, denoted as X_i , find all EEG segments with similar cardiogenic artifacts by finding the K EEG segments with the smallest $d(X_i, X_j)$. Denote those K EEG segments as $\{X_{ij}\}_{j=1}^K$.

Step 2–3: Then, find the median of $\{X_{ij}\}_{j=1}^K$, denoted as $\tilde{A}_i$, as the final recovered cardiogenic artifact. Finally, we can recover the EEG signal by removing $\tilde{A}_i$ from the original EEG signal, with the common stitching policy [23].

The key step for the NLEM is finding all EEG segments that have similar cardiogenic artifacts. However, it is not an easy job due to the existence of the EEG signal. To appreciate this, for the EEG signal x over the period $[t_i - L_L, t_i + L_R]$ denoted as X_i , we write it as the sum of two parts, that is, $X_i = A_i + E_i$, where A_i is deterministic that models the cardiogenic artifact and E_i is stochastic with mean 0 that models the EEG signal with possibly the inevitable noise. The usual L^2 distance between two EEG segments, X_i and X_j , is $\|X_i - X_j\|_{\mathbb{R}^p}^2 = \|A_i - A_j\|_{\mathbb{R}^p}^2 + 2\langle A_i - A_j | E_i - E_j \rangle + \|E_i - E_j\|_{\mathbb{R}^p}^2$. Clearly, if two cardiogenic artifacts are close, $\|A_i - A_j\|_{\mathbb{R}^p}^2$ will be close to 0 and $\langle A_i - A_j | E_i - E_j \rangle$ will be a random variable with mean 0 and small variance. However, the noise term $\|E_i - E_j\|_{\mathbb{R}^p}^2$ will be nonzero and might deviate the neighboring infor-

mation determined by $\|X_i - X_j\|_{\text{pp}}^2$. This domination might render two EEG segments with similar cardiogenic artifact patterns to have a huge distance. As a result, the neighbors determined by the usual L^2 distance might not have similar cardiogenic artifact patterns. Thus, we need a new method to find those EEG segments with the same or similar cardiogenic artifacts. To this end, we propose to apply the currently developed OS algorithm [36] to suppress E_i from X_i . OS can be viewed as a nonlinear matrix denoising algorithm that removes the noise in an efficient way. Note that in the proposed algorithm, we also applied a 5 Hz highpass filter before applying OS to help suppress the impact of EEG signals.

2.3. Quantification indices

The artifact residue (AR) index for the i th cardiogenic artifact, denoted as AR_i , is defined as

$$AR_i := \left| \log \left(\frac{1}{2} \left[\frac{\text{med}|\tilde{E}_i - \text{med}(\tilde{E}_i)|}{\text{med}|E_i - \text{med}(E_i)|} + \frac{\text{med}|E_i - \text{med}(E_i)|}{\text{med}|\tilde{E}_i - \text{med}(\tilde{E}_i)|} \right] \times \frac{1}{2} \left[\frac{Q95|\tilde{E}_i - \text{med}(\tilde{E}_i)|}{\max|E_i - \text{med}(E_i)|} + \frac{\max|E_i - \text{med}(E_i)|}{Q95|\tilde{E}_i - \text{med}(\tilde{E}_i)|} \right] \right) \right|,$$

where $\tilde{E}_i$ is the estimated EEG signal recovered over the i th cardiogenic artifact, E_i is the concatenation of true EEG signal over the interval *without* any cardiogenic artifact from $i-5, \dots, i+5$ cardiogenic artifacts, *med* means median, *Q95* means the 95% quartile, and *max* means the maximal value. As is indicated above, the cardiogenic artifact might have a “T” wave, thus, depending on the heart rate, sometimes the true EEG signal might not exist over an artifact to artifact interval. The AR index of a good reconstruction algorithm should be close to 0.

The spectral concentration (SC) index measures the performance of the algorithm in the frequency domain. The original SC index for the EEG signal is defined as the ratio of the power change, relative to raw signal, over the fundamental frequency and the harmonics of the stimulation frequency to the energy over the rest of the frequencies in the band (1–100 Hz). While this definition works well for stimulation artifacts removal problem, however, it is not suitable for cardiogenic artifact, due to spectral spreading caused by the existence of heart rate variability. To avoid this issue, we propose the following modified SC (mSC) index. First, the EEG signal is warped in the time domain by the estimated IHR so that the cardiogenic artifacts have a constant frequency that is equivalent to the mean of the estimated IHR. Second, the SC index is evaluated from the warped EEG signal. The EEG signal and warped EEG signal are both sampled at 250 Hz. The mSC index is defined as the ratio of the power change, relative to raw signal, over the fundamental frequency and the harmonics of the stimulation frequency to the energy over the rest of the frequencies in the band (1–100 Hz). The band 1–100 Hz is chosen since it is sufficiently wide to cover the spectrum of interest, including delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz) and gamma (30–50 Hz), even if the spectrum of warped EEG signal is clearly different from that of the original EEG. The SC index should be small if the reconstruction quality is good.

To apply the AR and SC indices, we adhere to the following protocol. For each 20-s epoch, we first get 5 prediction results for all non-overlapping 4-s sub-epochs. In each 20-s epoch, if more than and equal to 3 sub-epochs are determined to have cardiogenic artifacts, then we apply the cardiogenic removal algorithm to that 20-s epoch and evaluate the AR and SC indices. For the AR index, it is evaluated for each peak in the 20-s epoch and take the median of them to represent the AR index for that 20-s epoch. For the SC index, we calculate the index over the 20-s epoch.

The above AR and SC indices are designed to evaluate the algorithm

performance when the ground truth EEG signal is not available. When the true EEG signal is available, we apply the following four metrics [38] to quantify how well the EEG is recovered from the cardiogenic artifact contaminated EEG signal. Denote the true EEG as $a \in \mathbb{R}^N$ and the estimated EEG as $\tilde{a} \in \mathbb{R}^N$. Here, the estimated EEG can be the one after applying brMEGA or ensembled average subtraction (EAS), or without applying any algorithm. The first index is the normalized mean square error (NMSE), which is defined as

$$NMSE = \frac{\sum_{i=1}^N (a(i) - \tilde{a}(i))^2}{\sum_{i=1}^N a(i)^2}.$$

Clearly, the smaller NMSE is, the better the EEG is estimated by $\tilde{a}$. The second index is the cross correlation between a and $\tilde{a}$, denote as ρ . The better $\tilde{a}$ estimates a , the closer ρ is to 1. The third and fourth indices are *signal-to-noise ratio* (SNR) and *peak-signal-to-noise ratio* (PSNR),

which are defined by

$$SNR = 20 \log_{10} \left(\frac{\text{std}(a)}{\text{rmse}(a, \tilde{a})} \right), \quad PSNR = 20 \log_{10} \left(\frac{\max_{1 \leq i \leq N} \{|a(i)|\}}{\text{rmse}(a, \tilde{a})} \right),$$

where $\text{rmse}(a, \tilde{a})$ is the root mean square error between a and $\tilde{a}$. Clearly, the higher the SNR and PSNR, the better the recovery is.

To report NMSE, we divide the signal into non-overlapping 20-s epochs, and we take the median value of all NMSE s in each epoch, which represents the algorithm performance over that epoch. The same procedure is applied to ρ , SNR and PSNR in the same way. In the end we have four indices to quantify the cardiogenic artifact, and hence EEG, recovery quality over each epoch.

2.4. Implementation details, parameters and statistical analysis

The proposed brMEGA algorithm is implemented in MATLAB, and we adhere to the following parameters in this work. We set $1/\Delta_t = 250$. In the nonlinear-type TF analyses, including SST, the window is Gaussian and the full width at half maximum is set to be 10 s, and the maximal frequency is set to be $U = 25$ Hz. We fix the frequency resolution at $\Delta_\xi := U/m = 0.02$ Hz. Set $\lambda = 1$ in the curve extraction algorithm (1), we set the parameter $\gamma = 20$ in the peak tracing algorithm in (2). To train the kernel SVM classifier, we use all epochs in the training dataset in the leave-one-subject-out cross validation. The kernel is the radial basis function with $\gamma = 2$, the cost is set to $C = 1$, and the weight of label 1 v.s. label 0 is set to 3:1, where label 0 is the cardiogenic artifact free epochs and 1 for those epochs with cardiogenic artifacts. We do not carry out any grid optimization search for the parameters to avoid overfitting. In the NLEM, we take $K = 100$. Note that the batch size is the whole training dataset and the number of epochs is 1. In other words, we simply follow the standard application of libsvm in MATLAB without any additional tuning.

To avoid the possible interaction between the EEG signal and the sleep dynamics and to speed up the computation, we suggest to randomly permute the cardiogenic artifacts and then perform OS on every non-overlapping 500 cardiogenic artifacts. To further speed up the computation, the NLEM step is suggested to be carried out in the following way. We first divide the EEG segments $\{X_i\}_{i=1}^n$ into non-overlapping

groups containing 250 consecutive EEG segments. Suppose X_i is in the l th group. Find K EEG segments with similar cardiogenic artifacts in the sense of the smallest diffusion distance between the l th and the $(l+1)$ th groups.

To evaluate the proposed automatic determination of existence of cardiogenic artifacts, the leave-one-subject-out cross validation was carried out. In the leave-one-subject-out cross validation, the validation set and training set are determined on the subject level; that is, the training set and the validation set contain different subjects. In other words, each subject is viewed as the testing dataset when all the remaining subjects are viewed as the training dataset; that is, we run thus training-testing mission for 13 times including all epochs of the respective subjects. The summary result is shown in Table 1 and performance for each subject separately is illustrated in supplementary Table A1. Compared with the 5-fold cross validation shown in Supplementary Table 1, the result is compatible and encouraging. Note that while the leave-one-subject-out cross validation scheme is challenged by the inter-individual variability, this scheme is close to the real-world scenario, particularly when there is no available information for a newly arrival subject.

The power spectral density is calculated using Welch's method, featuring 5000 discrete Fourier transform points, Hamming windows of 5000 samples, and 50% overlapping. The 5000 point Fourier transform is chosen to reflect a 5-s window for a frequency resolution of 0.2 Hz. Spectrogram plots were derived using a multitaper approach to optimally illustrate the nocturnal neurophysiological EEG dynamics as well as the artifact over the whole recording period [39]. Specifically, we used the *mtspecgramc* function from the Chronux open source library (<http://www.chronux.org/>). As previously optimized for whole night spectral representation, we chose the following parameters which were previously optimized: 30 s windows spaced at 5 s intervals, 1 Hz spectral resolution, the time-half-bandwidth product of 15, and the number of tapers of 29 [39].

For the algorithm performance evaluation, the mean and standard deviation of the AR and SC indices for different algorithms were reported. For each subject, the null-hypothesis that one algorithm does not perform better than the other one in the sense of AR or SC index is tested by the paired-sample *t*-test. We view p value < 0.05 as statistical significance. To compare recovery of the EEG in the semi-real dataset, we first evaluated whether EAS and brMEGA improved EEG signal quality compared to the uncleaned EEG. Therefore, we performed a linear mixed model using method as a fixed factor (levels: uncleaned, brMEGA, EAS) and condition and subject as a random slope factor. To verify whether the cleaning with brMEGA was different to EAS we took the difference of each epoch relative to the uncleaned signal and performed a linear mixed model of these differences with the fixed factor cleaning method (levels: brMEGA, EAS) condition and subject as a random slope factor.

3. Results

Here we included thirteen nocturnal single-channel EEG recordings that showed various degrees of cardiogenic artifact contamination, both across and within recordings and that were obtained from a mobile EEG device, the MHSL-Sleepband [6]. In addition, two additional datasets

with and without cardiogenic contamination from a sleep lab were included that also had synchronous ECG channels from the whole night polysomnogram.

3.1. Non-linear time-frequency analysis and machine learning enable automatic determination of existence of cardiogenic artifacts in EEG (Step 1)

In the first step of brMEGA, we aimed at automatically determining if cardiogenic artifacts exist. Particularly, without the ECG signal or cardiac rhythmic information, we need to estimate where the cardiogenic artifacts are. However, due to the variability of the heart rate, the traditional time-domain and spectral approaches are limited since the cardiac rhythm is not fixed. To handle these challenges, we take a well-known fact into account. When cardiogenic artifact exists as a periodic component, the time-frequency representation (TFR) of the EEG signal with cardiogenic artifact contains dominant curves associated with the cardiogenic artifacts; otherwise, there is no dominant curve. This specific characteristic in the TFR is converted into features so that we can learn if cardiogenic artifacts exist. To do so, the signal is composed with the rectified linear unit activation function and then smoothed (step 1-1). Then we apply the recently established non-linear time-frequency methods of synchrosqueezing transform (SST) [29,30,40] (step 1-2) to extract dominant spectral features from the TFR, particularly the instantaneous frequency (IF). When the cardiogenic artifacts exist, the IF is the instantaneous heart rate (IHR), so we obtain the IHR from SST at the same time; otherwise, we obtain the IF associated with the EEG spectrum. With the IF, the peak tracking algorithm is applied to obtain the peak locations (step 1-3). When cardiogenic artifacts exist, we obtain the cardiogenic artifact locations; otherwise, we obtain random EEG spikes. From steps 1-2 and 1-3, we form four features associated with cardiogenic artifacts, two from the time domain and two from the TFR (step 1-4). We apply the well-established support vector machine algorithm (step 1-5) to learn whether a 4 s epoch is contaminated by cardiogenic artifact or not. The flowchart of this part of algorithm is shown in Fig. 2.

Our neuroscientist expert visually examined the EEG signal and labeled non-overlapping 4-s epochs using a quality index that describes the strength of cardiogenic artifacts. An epoch was labeled as 0 if no clear cardiogenic artifacts could be visually identified, and 1 if a cardiogenic artifact was clearly visible. For 13 whole night polysomnograms (all from different participants), the neuroscientist labeled randomly selected 100 4-s epochs between the first and second hour of each recording.

There were in total 1300 labeled 4-s epochs from 13 subjects. The leave-one-subject-out cross validation result is shown in Table 1, where the sensitivity is 0.95, the specificity 0.70, and the F1 score is 0.84 and the overall accuracy is 0.81. The performance metrics for each individual subject are shown in Table A1 in the Appendix. The 5-fold cross validation carried out over all epochs from all subjects revealed similar performance values (Appendix Table A2).

3.2. Template subtraction (TS) using the wave-shape manifold model can accurately remove cardiogenic artifacts of different sizes and forms (Step 2)

In the second step of brMEGA, we aimed at removing the cardiogenic artifacts. Once an epoch is determined to be contaminated by cardiogenic artifacts in step 1, we apply the beat tracking algorithm combined with the estimated IHR by SST to obtain the location of cardiogenic artifacts (step 1-3). With these locations, we move forward to remove the cardiogenic artifacts. A first glance at the cardiogenic artifact contaminated EEG signal suggests that the cardiogenic artifacts share a common spike-like pattern. This common pattern motivated the TS algorithm. The traditional TS approach, however, is limited when the spike-like patterns are not fixed from one to another, which is the case in

Table 1
The overall confusion matrix of the leave-one-subject-out cross validation.

	Visual: artifact (1) (576)	Visual: no artifact (0) (724)	
Prediction: artifact (1)	539	220	Precision = 0.71
Prediction: no artifact (0)	37	504	False predictive value = 0.93
	Sensitivity = 0.93	Specificity = 0.69	Accuracy = 0.80 F1 score = 0.84

the cardiogenic artifacts in EEG. To fix this issue, the manifold learning algorithm was introduced. As a generalization of linear model, a manifold model is capable of describing the nonlinear relationship among cardiogenic artifacts [33]. Based on this model, we apply the optimal shrinkage (OS) to find similar cardiogenic patterns (step 2-1), so that the nonlocal Euclidean median (step 2-2) leads to a less biased cardiogenic artifact estimates. Unlike the traditional TS approach that mainly counts on the previous few cardiogenic artifacts, this step allows us a better reconstruction of spike-like cardiogenic artifacts (step 2-3) and hence a more accurate EEG recovery (Step 2-4). The flowchart of this part of algorithm is shown in Fig. 3.

Fig. 4 illustrates the TF representation of one representative nocturnal recording before and after cardiogenic artifact removal. All other 12 nights can be found in Supplementary Figures A2-A13. Application of the brMEGA recovers the clean EEG by substantial removal of the artifacts in the delta-theta and high frequency region (including higher harmonics). In the temporal domain, this finding is confirmed by an accurate recovery of the signal (see Fig. 5 for an exemplary temporal trace).

3.3. ECG QRS location and structure match well with algorithm identified artifact

In order to validate the performance of our brMEGA to accurately define ECG peak locations we included two in-lab measured subjects from one study site that also included ECG as reference recordings. In Fig. 5, a representative segment before and after cardiogenic artifact removal is shown, and the simultaneously recorded electrocardiogram is illustrated for a visual comparison. Visually identified cardiogenic artifacts coincide with QRS complexes in the ECG signal. To quantify the performance [41], the R-peaks of the ECG are viewed as the ground truth. To avoid the impact when the ECG signal is of low quality, the ECG signal quality is determined by bSQI [27]. We keep all epochs when the associated ECG signal is of bSQI > 0.95 in the final evaluation, 4.4% epochs are removed. As is illustrated in Table 2, the detected cardiogenic

artifact peaks in two cases by brMEGA correspond with 96% precision and 96% sensitivity to R-peaks defined in the ECG with the grace period of 50 ms. The mean absolute error is 1.24 ms. We therefore conclude that brMEGA can detect the existence of cardiogenic artifact and remove cardiogenic artifacts from the single-channel EEG signal.

In Fig. 6, an additional comparison of the recovered cardiogenic artifacts and ECG signal is demonstrated. It is noteworthy that in addition to the dominant spike associated with the R peak in the ECG signal, there is a bump after the dominant spike in the reconstructed cardiogenic artifact (indicated by red arrows). Based on the temporal characteristics of a classical ECG curve as derived from our averaged ECG signal, the second bump observed in the cardiogenic artifact is likely associated with the T wave in the ECG signal.

3.4. Manifold TS outperforms previously published TS approach in removing the cardiogenic artifacts from contaminated EEG

In addition to our proposed brMEGA algorithm, we also show the result by a previously published TS approach [15], which is called ensemble average subtraction (EAS). For a fair comparison, step 2 of the proposed brMEGA is replaced by EAS.

In order to quantify and compare the TS performance, we consider the following two indices: the AR index, which was considered in Malik et al. [38] for the f-wave extraction from ECG signals, and a modification of the spectral concentration (mSC) index proposed in Castells et al. [42]. The quantitative results shown in Fig. 7 indicates that overall, the proposed brMEGA algorithm outperforms the existing EAS algorithm as demonstrated by a significant reduction in both SC and artifact residue (AR; values get closer to 0) across all participants. Note that 2 out of 13 cases have less than 30 epochs contaminated by cardiogenic artifacts (S029: 24/1562 and S053: 2/1669), so they are not shown and summarized below. Overall, the mean and standard deviation of the AR index for brMEGA is 0.04 ± 0.015 , while those for EAS is 0.054 ± 0.028 . For each subject, the null-hypothesis that brMEGA does not perform better than EAS in the sense of AR index is rejected by the paired-sample

Fig. 3. The flowchart of the second step of brMEGA. Signal x_1 : preprocessed raw signal (50 Hz IIR notch filter with removed wandering baseline). $\tilde{x}$: x_1 after 5th order butterworth highpass filter with the cutoff at 5 Hz (to suppress the background EEG signal). $\tilde{H}$: Signal after applying optimal shrinkage. X_i : representation of the i th artifact. $\hat{A}_i$: Signal representing the i th estimated artifact.

Fig. 4. TF representation of the original EEG signal (all-night nocturnal recording, left) and after removing the cardiogenic artifact (right).

Fig. 5. Illustrations of the cardiogenic-artifact contaminated EEG signal (second panel) and the simultaneously recorded ECG signal (top panel). Red $\times$ marks the respective peak detection for both ECG (R peak) and EEG (cardiogenic artifact peak) separately. The R peak and artifact peaks are correctly identified and overlap. The recovered EEG signal by brMEGA is shown in the third panel.

t -test with the p value $< 10^{-10}$. Regarding the SC index, the mean and standard deviation for brMEGA is 0.013 ± 0.003 , while those for EAS is 0.015 ± 0.003 . For each subject, the null-hypothesis that brMEGA does not perform better than EAS in the sense of SC is rejected by the paired-sample t -test with the p value $< 10^{-5}$.

3.5. Manifold TS outperforms previously published TS approach in recovering the EEG signal from ECG artifact

To further confirm that the EEG is properly recovered, we apply brMEGA and EAS to the semi-real database. In this dataset, we combine cardiogenic artifact-free EEG with varying cardiogenic artifacts that are characterized by the range and variability of the cardiogenic artifacts reported in the previous paragraphs (7 nocturnal EEG recordings from

Table 2

The confusion matrix of cardiogenic artifact detection. A detected cardiogenic artifact is viewed as a false positive when there does not exist a R peak within the 50 ms grace period. A false negative cardiogenic artifact is defined when there is a R peak, but we do not detect a cardiogenic artifact within the 50 ms grace period.

	ECG R peak (number = 39,942)		
Detected EEG cardiogenic artefact (number = 39,730)	True positive number = 38,262	False positive number = 1468	PPV = 0.96
	False negative number = 1680		F1 = 0.96
	Sensitivity = 0.96		MAE = 1.24

different participants). This approach allows us to compare the EEG with contamination and after applying artifact removal (brMEGA and EAS) to the ground truth EEG without ECG artifacts. To do so, we report here four indices that quantify EEG recovery. First, normalized mean square error (NMSE) is substantially lower after artifact removal with brMEGA (0.151 ± 0.113), and EAS (0.184 ± 0.129) compared to the contaminated EEG (1.241 ± 0.941). Second, the cross-correlation (ρ) between the ground truth EEG and the brMEGA cleaned EEG (0.926 ± 0.058) and EAS cleaned EEG (0.916 ± 0.06) strongly increases compared to the correlation with the cardiogenic contaminated EEG (0.69 ± 0.118). Third, the signal-to-noise ratio (SNR) substantially increases after artifact removal with brMEGA (8.919 ± 2.361) and EAS (8.130 ± 2.536) compared to the contaminated EEG (-0.025 ± 2.739). The same holds true for the fourth index, peak-SNR (PSNR), with brMEGA (17.806 ± 2.245) and EAS (17.041 ± 2.416) cleaned EEG being substantially higher than the uncleaned EEG (8.888 ± 2.651). Overall, both brMEGA and EAS substantially recover the EEG as indicated by all four indices being pronouncedly improved compared to the contaminated EEG. For each index, brMEGA-cleaned EEG significantly differs from the contaminated EEG in all four indices when performing a linear mixed effects model with p -values < 0.001 . Furthermore, comparing the differences to the uncleaned signal, brMEGA outperforms the EAS approach for all four indices with p -values significantly for SNR, and pSNR p -values < 0.02 , and trend-level for NMSE, and ρ , p -values = 0.065 and 0.057, respectively.

Fig. 6. An illustration of the cardiogenic artifacts and the “T wave” bumps (marked by red arrows). If the window is short, the T-wave-like artifact cannot be recovered. Red $\times$ indicate cardiogenic artifact peak or R peak occurrence.

4. Discussion

Here we propose the automated framework brMEGA for determination, peak location identification, and removal of cardiogenic artifacts from single-channel EEG. We demonstrate that our algorithm can successfully detect and substantially recover EEG periods with cardiogenic artifacts without the need of an additional ECG derivation. BrMEGA is the first algorithm for automated cardiogenic artifact determination and removal that can be applied to long duration recordings with minimal EEG channel settings. Therefore, it provides a highly attractive artifact removal solution for mobile EEG applications that have gained tremendous impetus in the last decade [1–4].

4.1. Automatic determination of existence of the cardiogenic artifact

In the first step, brMEGA determines a possible presence of a cardiogenic artifact in the EEG. The cardiogenic artifact varies in amplitude from being completely absent to strongly contaminating the EEG, both across and within different recordings. Classical spectral analysis is generally limited when the signals are with time-varying frequency and amplitude that need identification with high temporal and frequency resolution (e.g., IHR) such as is the case for cardiogenic artifacts. Therefore, a method is needed that overcomes the limitations of classical spectral analysis. We have recently established that SST can successfully isolate and characterize this kind of signal, even when the signal oscillates in a non-sinusoidal way (e.g., non-sinusoidal signals such as ECG) [29]. Accordingly, the brMEGA detection algorithm included a combination of features extracted from the raw signal and the

Fig. 7. The artifact residue (AR) and spectral concentration (SC) indices of 11 cases in which visually ECG artifact have been identified in the course of the nocturnal recording. The number of 20-s epochs of each case identified as been contaminated by cardiogenic artifacts, and the total number of 20-s epochs are listed below the case ID as the numerator and denominator respectively.

SST to capture typical cardiogenic signatures. With these features we have trained a support vector machine (SVM) to identify cardiogenic artifacts in EEG based on visual expert labelling. Compared to a neuroscience expert, the algorithm achieved a very high specificity (at ~95%) in identifying an artifact that was also identified visually. The sensitivity was slightly lower at around 70%, thus the algorithm identified some epochs as including cardiogenic artifacts in the EEG that were not labelled by the expert. However, this might be biased to the visual capacity of the expert. Artifacts could be less pronouncedly visible, specifically if strong EEG activity is present, even though they exist in the EEG. We took a closer look at some epochs that were only identified by the algorithm and could also see indices of cardiogenic artifacts that were labelled by the algorithm indication that at least in some epochs, the algorithm might outperform the scorer (see Appendix Figure A14). A limitation of the datasets derived from the mobile EEG device is that we did not have the ground truth ECG to evaluate whether the located cardiogenic peaks with our algorithm matched with real R-

peaks in the ECG. To overcome this limitation, we include additional in-lab datasets that provided simultaneous EEG and ECG recordings. Applying our brMEGA algorithm to these recordings demonstrated highly precise peak overlap between the cardiogenic artifact and the ECG with 96% precision and 96% sensitivity. This result is comparable with the accuracy of previously published approaches, for which accuracies around 97–99% were reported [10,16]. Here, we have to consider the achieved performance is slightly lower than in previous publications because we include a whole night recording with some time periods that might not have had an artifact. In other words, we include step 1 in this metric and likely have some inaccuracy in the determination of 4-s epochs with artifact contamination whereas the previous publications only report accuracy from segments where a clear visual artifact was present [10], or for which cardiogenic artifacts were added to clean EEG [16].

To date, only few previous studies introduced a framework to automatically determine the existence of and remove cardiogenic artifacts in

recorded data with unknown/different levels of contamination and without ECG reference [10–12]. They applied independent component analysis (ICA) in high-density EEG and developed an algorithm that could define IC with ECG artifact. However, their approach cannot be applied in single-channel EEG settings that are common in mobile applications. ICA is computationally very intensive and not ideal for long-term recordings, (a) because very long recordings would take up too much computational power, and (b) ICA requires a certain level of signal stationarity that is not provided for hour-long sleep and wake recordings therefore requiring sophisticated windowing approaches. Our presented framework here overcomes these limitations and enables accurate definition of contaminated segments and peak location independent of recording length also for single-channel EEG.

4.2. Removal and recovery of the cardiogenic artifact

Upon detection of cardiogenic artifacts, we applied the wave-shape manifold model template subtraction to remove the cardiogenic artifact. Our approach can substantially recover the EEG as illustrated in the time-domain but also in the frequency domain, for which we should removal in the primary frequency bands but also the affected higher harmonics. This is also quantified in four indices that depict EEG recovery in contrast to the ground truth EEG in our semi-real dataset. All indices show pronounced improvement after brMEGA application compared to the contaminated EEG. The proposed brMEGA is suitable to handle non-stationary time series, particularly when the signal is recorded for hours or longer. It is well known that the EEG signal is non-stationary. The cardiogenic artifact, on the other hand, is also non-stationary, due to the inevitable heart rate variability and the fact that the cardiogenic artifact strengths might vary from one to another. Furthermore, changes in head positions can affect the extent of the artifact, which is often happening during sleep or recordings that involve head movements [9]. Unlike the traditional template subtraction (TS) algorithms, like EAS we compared in this work [15] and its variations [16,22], we leverage this non-stationarity to more efficiently remove cardiogenic artifacts. Such non-stationarity is captured by the wave-shape manifold, which quantifies how cardiogenic artifacts change from one cycle to another. Specifically, for each given cardiogenic artifact, we design a specific filter, the optimal shrinkage, to “clean up the EEG signal” so that the median is taken only over those cardiogenic artifacts that are similar to the given one. Note that unlike the traditional TS algorithms, those chosen cardiogenic artifacts for the median operator might be temporally far away. When there are multiple channels, the proposed brMEGA can be generalized. A further exploration of such generalization and its comparison with ICA will be reported in our future work.

Our modified TS outperformed the EAS reported in Park et al. as quantified by the AR and mSC index [15]. The first term of the AR index measures how accurate the EEG signal is recovered and we need to avoid over-smoothing. The second term of the AR index, on the other hand, measures the residue of the cardiogenic artifact. The product captures simultaneously how well the cardiogenic artifact is removed and how well the EEG signal is recovered. Compared with its application to iEEG stimulation artifact removal [23] and f-wave extraction [38], its application to our problem is natural without too much change. The SC index, however, is much more challenging due to the time-varying frequency and time-varying strength of the cardiogenic artifact, and a modification of the SC index is introduced. By warping the EEG signal, the spectrum associated with cardiogenic artifacts becomes more concentrated around the fundamental frequency and its multiples. While the EEG spectrum is also deformed due to the warping, the spectral spreading issue caused by the inevitable heart rate variability is mitigated, and hence the EEG recovery can be better evaluated. Furthermore, our algorithm also outperformed in metrics for EEG signal recovery when we compared recovered signal in the semi-real dataset to the ground truth EEG.

We further identified that the cardiogenic artifact does not only entail the strong QRS complex but also the less pronounced T-wave, which has previously been reported [9]. Thus, we highlight the importance of capturing a long enough window around the QRS complex to remove the T-wave of the cardiogenic artifact.

4.3. Beyond artifact removal: Opportunity to capture heart related activity without ECG

Besides using brMEGA for cardiogenic artifact removal, it can also be leveraged to extract heart-rate related activity (e.g., heart rate variability measures, IHR) without the need of additional ECG. In some applications, like telemedicine and mobile health, where only limited channels are available or ECG channels around the chest might be disturbing (e.g., sport science), this extra information, when exists, might provide more health information and benefits. Even in the monitor-intensive environment like an intensive care unit or operation room, this information might also be useful when other first line heart rate information is of bad quality. Of note, capturing heart-rate related activity requires existence of cardiogenic artifacts and as already illustrated in this data set, this is even with similar montages not always the case. Therefore, to optimize capturing the artifacts one should carefully consider the location of the recording electrodes.

4.4. Limitations of the study

Our brMEGA approach offers several advantages in removing cardiogenic artifacts from single-channel mobile EEG data, that is a completely automatic framework to identify cardiogenic artifacts, peak detection and removal without the need of an additional ECG reference. Yet, this algorithm has clear potential to be further improved in future studies.

First, the sensitivity of the algorithm could be enhanced and validated against different approaches. Using visual identification as the ground truth of artifact existence is the step that can be improved. One problem is that strong EEG background activity might mask its presence so that the ground truth is not correct, which will mislead the SVM model and downgrade the overall performance. Therefore, while the algorithm could in theory outperform the visual capacity of the experimenter, this ultimate goal cannot be achieved in this work. Even with a better annotation, yet, this would need to be tested in different frameworks, such as with many experts and definition of a consensus rule among the independently rated expert opinions.

Second, while the brMEGA removes the artifact substantially and outperforms a classical EAS approach, there is still some remaining artifact that are hard to visualize by naked eyes. To further improve the algorithm, however, will be a combined effort of optimizing all three steps of the brMEGA, since all of them contribute to the artifact removal performance of long-term recordings.

Finally, it remains to be established whether brMEGA could be used to clean EEG signals that are contaminated by other strong artifact sources that resemble spike-like patterns, e.g., epileptic seizures.

5. Conclusion

EEG recordings with a minimal number of channels have gained popular interest because they allow for large-scale and long-term brain activity recordings. However, those systems come at the cost of making it more difficult to clean the EEG from artifacts and noise. Here, we propose brMEGA as an open-source algorithm for (1) automated cardiogenic artifact detection and (2) removal in single-channel EEG settings, using non-linear time-frequency analysis and machine learning. We validated brMEGA in a real and semi-real dataset containing long-term recordings and illustrate that our algorithm can substantially recover EEG from cardiogenic artifacts and outperforms already existing algorithms to remove detected cardiogenic artifacts.

Yet, it remains to be established whether our algorithm can also be trained to detect and remove additional artifacts that are difficult to identify by visual scorings.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was conducted as part of the SleepLoop Flagship of Hochschulmedizin Zürich and funded in part by the Schweizerische Hirn Stiftung and the Swiss National Science Foundation (P3P3PA 171525 and PZ00P3_179795 to CL). The authors thank all the participants for taking part in the study and R. Büchi, N. Demarmels, G. Hoppeler, J. Kurz, and E. Silberschmidt for performing the recruitment and data collection. SleepLoop consortium members provided uncountable discussions and feedback. This work and the Taiwan Integrated Database for Intelligent Sleep (TIDIS) project are partially supported by Ministry of Science and Technology 109-2119-M-002-014-, NCTS Taiwan. Neng-Tai Chiu is partially supported by the Undergraduate Summer Research Program hosted by National Center for Theoretical Sciences, Mathematics Division, Taiwan.

Author contributions

N.T.C.: Software, validation, formal analysis, visualisation, writing - review & editing. S.H.: Investigation, data curation, writing - original draft, review & editing. M.L.F.: Investigation, methodology, writing - review & editing. W.K.: Conceptualization, investigation, writing - review & editing. H.T.W.: Conceptualization, software, validation, formal analysis, supervision, visualization, writing - original draft, review & editing. C.L.: Conceptualization, investigation, data curation, validation, formal analysis, supervision, visualization, writing - original draft, review & editing. H.T.W. and C.L. contributed equally to this work, they are shared corresponding authors.

Data availability

All datasets analyzed during the current study are available from the corresponding author on reasonable request. The TIDIS dataset is available on <https://tidis.org/>. In the future, the dataset will be made available in a repository.

Code availability

The code of brMEGA is available on: <https://github.com/NengtaiChiu/brMEGA>.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bspc.2021.103220>.

References

- [1] M. De Vos, S. Debener, Mobile EEG: towards brain activity monitoring during natural action and cognition, *Int. J. Psychophysiol.* 91 (1) (2014) 1–2.
- [2] A. Lau-Zhu, M.P.H. Lau, G. McLoughlin, Mobile EEG in research on neurodevelopmental disorders: opportunities and challenges, *Dev. Cogn. Neurosci.* 36 (2019), 100635.
- [3] J. Askamp, M.J.A.M. van Putten, Mobile EEG in epilepsy, *Int. J. Psychophysiol.* 91 (1) (2014) 30–35.
- [4] C. Kranczioch, C. Zich, I. Schierholz, A. Sterr, Mobile EEG and its potential to promote the theory and application of imagery-based motor rehabilitation, *Int. J. Psychophysiol.* 91 (1) (2014) 10–15.
- [5] S. Debener, R. Emkes, M. De Vos, M. Blechner, Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear, *Sci. Rep.* 5 (2015) 16743.
- [6] M.L. Ferster, C. Lustenberger, W. Karlen, Configurable mobile system for autonomous high-quality sleep monitoring and closed-loop acoustic stimulation, *IEEE Sensors Lett.* 3 (5) (2019) 1–4, <https://doi.org/10.1109/LSNS.778263410.1109/LSNS.2019.2914425>.
- [7] K.B. Mikkelsen, S. L. Kappel, D. P. Mandic, P. Kidmose, EEG recorded from the ear: characterizing the Ear-EEG Method, *Front. Neurosci.*, 9, (NOV), 2015, 10.3389/fnins.2015.00438.
- [8] X. Jiang, G.-B. Bian, Z. Tian, Removal of artifacts from EEG signals: a review, *Sensors* 19 (5) (2019) 987.
- [9] G. Dirlich, L. Vogl, M. Plaschke, F. Strian, Cardiac field effects on the EEG, *Electroencephalogr. Clin. Neurophysiol.* 102 (4) (1997) 307–315.
- [10] G. Tamburro, D.B. Stone, S. Comani, Automatic removal of cardiac interference (ARCI): a new approach for EEG data, *Front. Neurosci.* 13 (2019) 441.
- [11] S.-Y. Shao, K.-Q. Shen, C.J. Ong, E.P.V. Wilder-Smith, X.-P. Li, Automatic EEG artifact removal: a weighted support vector machine approach with error correction, *IEEE Trans. Biomed. Eng.* 56 (2) (2008) 336–344.
- [12] L. Shoker, S. Sanei, J. Chambers, Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm, *IEEE Signal Process. Lett.* 12 (10) (2005) 721–724.
- [13] V. Schetinin, J. Schult, The combined technique for detection of artifacts in clinical electroencephalograms of sleeping newborns, *IEEE Trans. Inf. Technol. Biomed.* 8 (1) (2004) 28–35.
- [14] S. Bhattacharyya, et al., Detection of artifacts from high energy bursts in neonatal EEG, *Comput. Biol. Med.* 43 (11) (2013) 1804–1814.
- [15] H.-J. Park, D.-U. Jeong, K.-S. Park, Automated detection and elimination of periodic ECG artifacts in EEG using the energy interval histogram method, *IEEE Trans. Biomed. Eng.* 49 (12) (2002) 1526–1533.
- [16] C. Dora, P.K. Biswal, Efficient detection and correction of variable strength ECG artifact from single channel EEG, *Biomed. Signal Process. Control* 50 (2019) 168–177.
- [17] J.-A. Jiang, C.-F. Chao, M.-J. Chiu, R.-G. Lee, C.-L. Tseng, R. Lin, An automatic analysis method for detecting and eliminating ECG artifacts in EEG, *Comput. Biol. Med.* 37 (11) (2007) 1660–1671.
- [18] S.P. Cho, M.H. Song, Y.C. Park, H.S. Choi, K.J. Lee, Adaptive noise canceling of electrocardiogram artifacts in single channel electroencephalogram, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 3278–3281.
- [19] S. Devuyt, T. Dutoit, P. Stenuit, M. Kerkhofs, E. Stanus, Removal of ECG artifacts from EEG using a modified independent component analysis approach, *Conf Proc IEEE Eng Med Biol Soc* 2008 (2008) 5204–5207, <https://doi.org/10.1109/IEMBS.2008.4650387>.
- [20] A. Jafarifarmand, M.A. Badamchizadeh, Artifacts removal in EEG signal using a new neural network enhanced adaptive filter, *Neurocomputing* 103 (2013) 222–231.
- [21] A.G. Correa, E. Laciari, H.D. Patiño, M.E. Valentinuzzi, Artifact removal from EEG signals using adaptive filters in cascade, *J. Phys. Conf. Ser.* 90 (1) (2007) 12081.
- [22] M. Nakamura, H. Shibasaki, Elimination of EKG artifacts from EEG records: a new method of non-cephalic referential EEG recording, *Electroencephalogr. Clin. Neurophysiol.* 66 (1) (1987) 89–92.
- [23] S. Alagapan, H.W. Shin, F. Fröhlich, H. Wu, Diffusion geometry approach to efficiently remove electrical stimulation artifacts in intracranial electroencephalography, *J. Neural Eng.* 16 (3) (2019) 36010.
- [24] M.E. Davies, C.J. James, Source separation using single channel ICA, *Signal Process.* 87 (8) (2007) 1819–1832.
- [25] J. Lin, A. Zhang, Fault feature separation using wavelet-ICA filter, *NDT e Int.* 38 (6) (2005) 421–427.
- [26] B. Mijović, M. De Vos, I. Gligorijević, J. Taelman, S. Van Huffel, Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis, *IEEE Trans. Biomed. Eng.* 57 (9) (2010) 2188–2196.
- [27] A.E.W. Johnson, J. Behar, F. Andreotti, G.D. Clifford, J. Oster, Multimodal heart beat detection using signal quality indices, *Physiol. Meas.* 36 (8) (2015) 1665–1677.
- [28] P. Laguna, L. Sörnmo, Sampling rate and the estimation of ensemble variability for repetitive signals, *Med. Biol. Eng. Comput.* 38 (5) (2000) 540–546.
- [29] I. Daubechies, J. Lu, H.-T. Wu, Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool, *Appl. Comput. Harmon. Anal.* 30 (2) (2011) 243–261.
- [30] C.L. Herry, M. Frasch, A.J.E. Seely, H.-T. Wu, Heart beat classification from single-lead ECG using the synchrosqueezing transform, *Physiol. Meas.* 38 (2) (2017) 171–187.
- [31] D.P.W. Ellis, Beat tracking by dynamic programming, *J. New Music Res.* 36 (1) (2007) 51–60.
- [32] B. Scholkopf, A.J. Smola, Learning with kernels: support vector machines, regularization, optimization, and beyond, *Adapt. Comput. Machine Learn. Series* (2018).
- [33] Y.-T. Lin, J. Malik, H.-T. Wu, Wave-shape oscillatory model for biomedical time series with applications, *arXiv Prepr. arXiv1907.00502*, 2019.
- [34] G. Peyré, M. Cuturi, Computational optimal transport. *Foundations and Trends® Machine Learn.* 11 (5–6) (2019) 355–607.
- [35] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol.* 2 (3) (2011) 1–27.

- [36] M. Gavish, D.L. Donoho, Optimal shrinkage of singular values, *IEEE Trans. Inf. Theory* 63 (4) (2017) 2137–2152.
- [37] R.R. Coifman, S. Lafon, Diffusion maps, *Appl. Comput. Harmon. Anal.* 21 (1) (2006) 5–30.
- [38] J. Malik, N. Reed, C.-L. Wang, H.-T. Wu, Single-lead f-wave extraction using diffusion geometry, *Physiol. Meas.* 38 (7) (2017) 1310–1334.
- [39] M. J. Prerau, R. E. Brown, M. T. Bianchi, J. M. Ellenbogen, P. L. Purdon, Sleep neurophysiological dynamics through the lens of multitaper spectral analysis, vol. 32, no. 1. 2017, pp. 60–92.
- [40] F. Auger, P. Flandrin, Y.-T. Lin, S. McLaughlin, S. Meignen, T. Oberlin, H.-T. Wu, Time-frequency reassignment and synchrosqueezing: an overview, *IEEE Signal Process. Mag.* 30 (6) (2013) 32–41.
- [41] A.-A. EC57 and A. for the A. of M. Instrumentation, Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, *Assoc. Adv. Med. Instrumentation, Arlington, VA*, 1998.
- [42] F. Castells, J.J. Rieta, J. Millet, V. Zarzoso, Spatiotemporal blind source separation approach to atrial activity estimation in atrial tachyarrhythmias, *IEEE Trans. Biomed. Eng.* 52 (2) (2005) 258–267.