Removal of eye blink artifacts in wireless EEG sensor networks using reduced-bandwidth canonical correlation analysis

Ben Somers[1] and Alexander Bertrand[2]

[1]KU Leuven, Department of Neurosciences, Research Group Experimental Oto-Rhino-Laryngology (ben.somers@med.kuleuven.be)

[2]KU Leuven, Department of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics (alexander.bertrand@esat.kuleuven.be)

***Abstract*— *Objective*: Chronic, 24/7 EEG monitoring requires the use of highly miniaturized EEG modules, which only measure a few EEG channels over a small area. For improved spatial coverage, a wireless EEG sensor network (WESN) can be deployed, consisting of multiple EEG modules, which interact through short-distance wireless communication. In this paper, we aim to remove eye blink artifacts in each EEG channel of a WESN by optimally exploiting the correlation between EEG signals from different modules, under stringent communication bandwidth constraints. *Approach*: We apply a distributed canonical correlation analysis (CCA-)based algorithm, in which each module only transmits an optimal linear combination of its local EEG channels to the other modules. The method is validated on both synthetic and real EEG data sets, with emulated wireless transmissions. *Main Results*: While strongly reducing the amount of data that is shared between nodes, we demonstrate that the algorithm achieves the same eye blink artifact removal performance as the equivalent centralized CCA algorithm, which is at least as good as other state-of-the-art multi-channel algorithms that require a transmission of all channels. *Significance*: Due to their potential for extreme miniaturization, WESNs are viewed as an enabling technology for chronic EEG monitoring. However, multi-channel analysis is hampered in WESNs due to the high energy cost for wireless communication. This paper shows that multi-channel eye blink artifact removal is possible with a significantly reduced wireless communication between EEG modules.**

I. Introduction

Current clinical EEG devices often make use of an electrode head cap with wired connections to a computer and a power source. This makes them impractical for everyday use in the patient's natural environment, and limits the possibilities of comfortable measurements for extended durations. A first step towards achieving mobile EEG recordings is the development of wireless EEG headsets [1], [2]. Several wireless headsets have appeared on the market, but these are still too heavy and too bulky to wear them continuously in everyday life, and offer only limited autonomy since the wireless transmission of all EEG channels to a computer or smartphone consumes a substantial amount of energy. However, advancements in miniaturization of biomedical sensors, on-body or off-body wireless transmitters and power efficient design give prospects of miniature EEG modules capable of processing, logging and transmitting relevant neurological data 24/7 over a long period of time [3], [4].

For example, to design so-called neuro-steered hearing prostheses [5], [6], a small EEG module can be integrated into a hearing prosthesis for in-the-ear [7] or around-the-ear [8] EEG recordings, and even the implanted electrodes of a cochlear implant can be used to record EEG [9]. Such small EEG modules are discreet, albeit limited in number of channels and spatial coverage. If necessary, a second wireless EEG module can be deployed at the other ear to obtain a better spatial coverage, and to increase the number of available channels [8]. This is similar to the existing concept of binaural hearing aids, where a wireless link between hearing aids at both ears allows communication for improved sound quality and localization [10]. Similarly, two EEG modules at both ears can exchange EEG data, which can be viewed as a two-node wireless EEG sensor network (WESN) [11]–[14].

We can imagine other future applications which can rely on discreet wearable EEG modules, such as sleep monitoring, epilepsy monitoring, stress monitoring, and brain-computer interfaces. There exist many examples of the research being done to realize these small EEG modules, such as the development of soft, flexible arrays of electrodes that can be mounted on the auricle [15], and the integration of EEG electrodes into common headgear or silicone earpieces [16]. Other examples include electrodes that latch on to the scalp despite the presence of hair [17], [18], or small electronics that are implantable underneath the skin [19]. By letting a multitude of such modules exchange EEG data with each other in a WESN-like architecture, the spatio-temporal correlation between their EEG signals can be exploited to solve a certain signal processing task.

A big challenge that still needs to be overcome to make such WESNs viable is to reduce the energy consumption, as battery lifetime is a major limiting factor. The energy consumption in such a system is dominated by wireless transmission. On-chip processing can provide a solution here, but this is only possible for local (single-channel) processing at each node individually [3], [20]. Multi-channel EEG processing algorithms, e.g., for artifact removal, spatial filtering or source localization

implicitly require data centralization, which is highly energy-inefficient. Distributed algorithms provide a key solution here, as they can exploit the spatio-temporal correlation structure across the different modules under the severe bandwidth and energy constraints in a WESN [11]. This correlation structure is exploited implicitly without centralizing the data, i.e., the modules only share fused or compressed data with each other, without information loss for the signal processing task at hand. In this paper, we focus on eye blink artifact removal as the signal processing task that is to be solved cooperatively by the modules within the WESN. From here on, we will refer to the different EEG modules in the WESN as the *nodes* of the network.

Artifact removal is an important pre-processing step before interpreting EEG, as the artifacts obscure underlying neural activity, which is often an order of magnitude smaller in amplitude. As eye blink artifacts are a common problem in most EEG recordings [21], we aim to remove them using a spatial filtering technique using all the channels of the WESN, while taking the severe bandwidth constraints of a WESN into account. Our goal is to remove the eye blink artifacts in every channel of the WESN, by exploiting the spatio-temporal correlation structure across *all* nodes in the network.

Although Independent Component Analysis (ICA) is currently the most popular method to remove eye blink artifacts [22], it is computationally demanding and not amenable to a distributed realization in WESNs. Previous work on distributed eye blink artifact removal in WESNs takes a Multi-channel Wiener Filtering (MWF) approach [11], [23], which is a semi-supervised method, as the data needs to be split into clean and corrupted EEG segments. Segregation methods such as thresholding are easy and cheap to implement, however they lead to a high amount of false positives and negatives due to misclassification errors. Furthermore, this threshold is a parameter that is very subject-dependent, and even time-dependent due to changes in the electrode-skin contact impedance, which cause signal drift. The segregation problem is further complicated in a WESN, because each node needs a different threshold tuning for optimal results, or the segmentation should be performed by a single node (with highest artifact amplitude), and then the results should be broadcast to all other nodes, which creates additional communication load.

The above problems can be overcome by Canonical Correlation Analysis (CCA). We will show that the most important parameter to be tuned is a time lag, which can be chosen equal across the nodes in the WESN and even across subjects. Because of the Blind Source Separating (BSS) capabilities of CCA, there is no need for an artifact detection step. In [24], CCA is used to remove ocular artifacts from EEG measurements, by including information from an electro-oculogram (EOG) measured with an electrode placed near the eyes. Similar results are obtained in [25] by making use of only two EEG measurement channels and four EOG electrodes. CCA has also been shown to be an effective method to remove muscle artifacts from EEG data [26]. To our knowledge CCA has not been applied yet as a method for eye blink artifact removal in EEG recordings without an EOG reference.

In this paper, we will use the BSS capabilities of CCA to remove eye blink artifacts without EOG references, and then extend the algorithm towards a distributed realization, which allows to substantially reduce the transmission cost in a WESN, while still achieving a provable equivalent performance as the centralized CCA algorithm. We validated the DCCA algorithm's performance using both synthetic and real EEG data. Both these datasets are emulated as a WESN, meaning that the electrodes are clustered and re-referenced into galvanically separated nodes. For the synthetic EEG data, a known artifact-free ground truth is available to verify our results.

This paper is structured as follows: in Section II, we explain the problem statement and how (D)CCA can be applied to remove eye blink artifacts. In Section III, the experimental procedures are given. In Section IV, we present the results of the experiments, which are further discussed in Section V. Finally, we draw conclusions in Section VI.

II. CCA-Based Artifact Removal in a WESN

A. Problem statement

Consider a WESN with K nodes. Node $k \in \{1, \ldots, K\}$ has access to M_k local EEG channels, stacked in a M_k-channel signal $\mathbf{x}_k[t]$, which is treated as a M_k-dimensional stochastic vector of which realizations are observed at different sample times $t \in \mathbb{N}$. The total number of EEG channels in the network is $\sum_{k=1}^{K} M_k = M$. We model the M_k-channel EEG signal in node k as

$$\mathbf{x}_k[t] = \mathbf{d}_k[t] + \mathbf{n}_k[t], \tag{1}$$

consisting of an eye blink artifact signal $\mathbf{d}_k[t]$, and the clean EEG signal $\mathbf{n}_k[t]$. For the sake of conciseness, we will omit the sample index t in the sequel, where we treat $\mathbf{x}_k[t]$ as a single observation of a multi-variate stochastic variable denoted by $\mathbf{x}_k$. In the estimation literature, the symbol d usually represents the signal to be estimated ('desired'), and the symbol n represents the undesired components ('noise'). Although it may appear strange to denote the artifact component as 'desired' and the clean EEG component as 'noise', this fits with the rationale of our algorithm, which first estimates the eye blink artifact signal by removing the EEG, and then subtracts a scaled version of the estimated artifact signal from the raw EEG data. From this point of view, the clean background EEG can be viewed as 'noise' in the first (artifact estimation) step of the algorithm. In the EEG-BSS literature, it is common practice to write the EEG signal $\mathbf{x}_k$ as an instantaneous mixture of S point-source signals $\mathbf{s} = [s_1 \ldots s_S]^T$ [22], i.e.,

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{s} + \mathbf{r}_k, \tag{2}$$

where $\mathbf{r}_k$ is a residu, and $\mathbf{A}_k$ is an $M_k \times S$ mixing matrix with element $A_{k,ij}$ denoting the contribution of source j in channel i, for $i = 1, \ldots, M_k$ and $j = 1, \ldots, S$. We assume without loss of generality that s_1 is the eye blink source signal generated at the eyes, and hence $\mathbf{d}_k$ can be modelled as [11], [22]

$$\mathbf{d}_k = \mathbf{a}_{k,1} s_1 \tag{3}$$

where $\mathbf{a}_{k,1}$ is the first column of $\mathbf{A}_k$ denoting the contribution of the eye blink signal at each channel of node k. The clean

EEG signal is then modelled as

$$\mathbf{n}_k = \mathbf{A}_{k,-1}\mathbf{s}_{-1} + \mathbf{r}_k \tag{4}$$

where $\mathbf{A}_{k,-1}$ denotes $\mathbf{A}_k$ with the first column removed, and $\mathbf{s}_{-1}$ is $\mathbf{s}$ with s_1 removed. Note that the source signal $\mathbf{s}$ is not node-specific, while the mixing matrix $\mathbf{A}_k$, which determines how each source is mixed in the local channels, is node-specific.

Our goal is not to estimate or unmix all source signals in $\mathbf{s}$ (as is often the case in traditional BSS problems). Instead, we only aim to estimate $\mathbf{d}_k$ at each node, and hence each node will produce a local estimate of the eye blink artifact signal $\hat{\mathbf{d}}_k$ and subtract it from its measured EEG channels to find an estimate of the clean EEG data $\hat{\mathbf{n}}_k$. To do this, the node needs a local estimate of the eye blink source signal s_1, and an estimate of the local eye blink mixing vector $\mathbf{a}_{k,1}$. In the next Subsection II-B, we will explain how CCA can be used to estimate the eye blink source signal, after which $\mathbf{a}_{k,1}$ can be estimated with a simple least squares fit. We first consider the case of centralized CCA, where all nodes centralize their EEG channels in a local processing unit. While the network-wide correlation structure is more straightforward to exploit when all channels are available, such data centralization is very bandwidth- and energy-consuming, and is preferably avoided in a WESN. In Subsection II-C, we explain how DCCA can be used to avoid the expensive data centralization in a WESN by exploiting the low dimensionality of the underlying artifact sources.

B. Centralized CCA-based eye blink artifact removal

CCA is a statistical tool used to find common patterns between two different multi-dimensional data sets based on correlations in the data [27]. As we will show later on, it is capable of separating eye blink artifacts from the neural responses in EEG, because the former show specific temporal and spatial correlation patterns.

Consider two sets of multi-channel stochastic time signals $\mathbf{x}$ and $\mathbf{y}$, their respective numbers of channels being M and N. CCA then finds the M-dimensional vector $\hat{\mathbf{v}}_j$ and the N-dimensional vector $\hat{\mathbf{w}}_j$ such that the single-channel signals $\hat{\mathbf{v}}_j^T\mathbf{x}$ and $\hat{\mathbf{w}}_j^T\mathbf{y}$ show maximal cross-correlation. These vectors can be found for values of j up to $\min(M,N)$, under the constraints that the j^{th} direction is perpendicular to all previous $j-1$ directions. Defining the data covariance matrices $\mathbf{R}_{xx}$, $\mathbf{R}_{xy}$ and $\mathbf{R}_{yy}$ as respectively $E\{\mathbf{x}\mathbf{x}^T\}$, $E\{\mathbf{x}\mathbf{y}^T\}$ and $E\{\mathbf{y}\mathbf{y}^T\}$, the maximization problem to find the j^{th} direction can be formulated as

$$(\hat{\mathbf{v}}_j, \hat{\mathbf{w}}_j) = \arg\max_{\mathbf{v}_j,\mathbf{w}_j} \frac{\mathbf{v}_j^T\mathbf{R}_{xy}\mathbf{w}_j}{\sqrt{\mathbf{v}_j^T\mathbf{R}_{xx}\mathbf{v}_j \cdot \mathbf{w}_j^T\mathbf{R}_{yy}\mathbf{w}_j}} \tag{5}$$

under the constraints that

$$\begin{aligned} \mathbf{v}_i^T\mathbf{R}_{xx}\mathbf{v}_j &= 0 \qquad \forall\, i \in \{1,2,\ldots,j-1\} \\ \mathbf{w}_i^T\mathbf{R}_{yy}\mathbf{w}_j &= 0 \qquad \forall\, i \in \{1,2,\ldots,j-1\}. \end{aligned} \tag{6}$$

The vectors $\hat{\mathbf{v}}_j$ and $\hat{\mathbf{w}}_j$ are referred to as the j^{th} principal CCA direction between $\mathbf{x}$ and $\mathbf{y}$. The single-channel signals $\hat{\mathbf{v}}_j^T\mathbf{x}$ and $\hat{\mathbf{w}}_j^T\mathbf{y}$ are called the j^{th} principal CCA compononents of $\mathbf{x}$ and $\mathbf{y}$. The j^{th} correlation value ρ_j, maximized in (5), is referred to as the j^{th} canonical correlation coefficient.

To summarize, CCA finds orthogonal (i.e., uncorrelated) directions in the M- and N-dimensional spaces in which $\mathbf{x}$ and $\mathbf{y}$ are maximally correlated [27].

We can apply CCA to our eye blink artifact estimation problem, based on the source separation properties of CCA [25], [26], [28]. Imagine all the EEG channels $\mathbf{x}_k, \forall k \in \{1,\ldots,K\}$ measured by the nodes in the WESN are transmitted to a central processing unit and stacked in one M-dimensional signal $\mathbf{x}$. As the second multi-channel input signal for CCA (i.e., $\mathbf{y}$ in the CCA description above), we use the time-delayed version of $\mathbf{x}$, denoted as $\mathbf{x}_\tau$. It is defined as the EEG channels in $\mathbf{x}$ delayed with a time lag τ, i.e.

$$\mathbf{x}_\tau[t] \triangleq \mathbf{x}[t-\tau]. \tag{7}$$

Then, after performing a CCA on $\mathbf{x}$ and $\mathbf{x}_\tau$, a demixing matrix $\mathbf{V}$ can be constructed by putting all principal CCA directions $\hat{\mathbf{v}}_j$ in the columns of $\mathbf{V}$. It can be shown that $\mathbf{V}^T\mathbf{x}$ approximates the source signals in $\mathbf{s}$ up to a scaling factor, assuming that the sources have a different autocorrelation function at the chosen time lag [26], [28], [29]. Ideally, the autocorrelation function of the eye blink artifact signal at time lag τ is significantly larger than the autocorrelation of the other (EEG) sources at time lag τ. If this is the case, applying CCA to signals $\mathbf{x}$ and $\mathbf{x}_\tau$ will extract the eye blink artifacts in the first principal CCA component and hence (5) only has to be solved for $j=1$. In Section III-D, we will further elaborate on the choice of τ. Note that this method of using CCA for BSS assumes a good a priori choice of τ, so that the eye blink artifacts show a sufficiently high autocorrelation at that time lag. For BSS problems where the sources to be extracted are completely unknown, CCA can be adapted to predict the time lag (or a linear combination of multiple time lags) for each unknown source [30]. However, for the case of eye blink artifact removal, one single time lag can be chosen so that the eye blink source is consistently extracted across subjects, as will be shown in Section IV-B. This allows to reduce the computational complexity of the algorithm.

When the eye blink source signal has been estimated, i.e. as the first principal CCA component $\hat{s}_1 = \hat{\mathbf{v}}_1^T\mathbf{x}$, it needs to be rescaled to fit the artifact in each of the channels, and then subtracted to find the artifact-free EEG data. In order to estimate the appropriate scaling factors, first define

$$\hat{S}_1 \triangleq \Big[\hat{s}_1[t-L+1],\ \ldots\ ,\hat{s}_1[t]\Big] \tag{8}$$

as the single-channel time signal containing observations of the estimated eye blink source $\hat{s}_1$ in a window of length L, and

$$\mathbf{X} \triangleq \Big[\mathbf{x}[t-L+1],\ \ldots\ ,\mathbf{x}[t]\Big] \tag{9}$$

as the M-channel time signal containing observations of the EEG channels in a window of length L. We find the scaling factors as the solution of the Least Squares (LS) problem

$$\boldsymbol{\alpha} = \arg\min_{\boldsymbol{\alpha}} \left\| \left(\boldsymbol{\alpha}\hat{S}_1 - \mathbf{X}\right) \right\|_F^2, \tag{10}$$

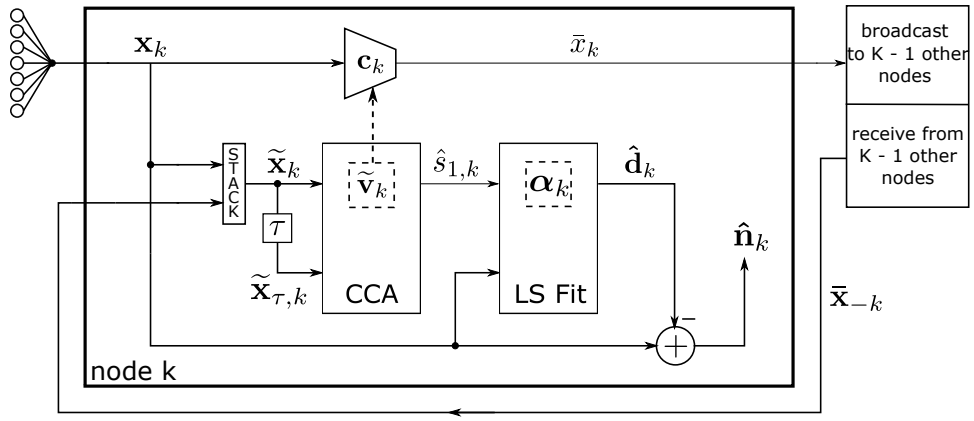


Fig. 1. Schematic illustration of the signal flow in the DCCA-based eye blink artifact removal algorithm at the k-th node of a WESN.

where $\|\cdot\|_F$ denotes the Frobenius norm and where $\boldsymbol{\alpha}$ is the vector containing the scaling factors, defined as

$$\boldsymbol{\alpha} = [\alpha_1, \ \dots \ , \alpha_M]^T, \tag{11}$$

in which α_i is the scaling factor that fits the estimated eye blink source $\hat{s}_1$ to EEG channel i. Solving the optimization problem in (10) can be done by computing a LS fit of $\hat{S}_1$ to $\mathbf{X}$. The LS solution for $\boldsymbol{\alpha}$ can be found as

$$\boldsymbol{\alpha} = (\hat{S}_1 \hat{S}_1^T)^{-1} \mathbf{X} \hat{S}_1^T. \tag{12}$$

Note that $\boldsymbol{\alpha}$ is a least squares estimation of the network-wide eye blink source mixing vector $\mathbf{a}_1$, analogous to the node-specific source mixing vector in (3). We now obtain the clean EEG data $\hat{\mathbf{n}}$ by subtracting the scaled estimated artifacts as

$$\hat{\mathbf{n}} = \mathbf{x} - \hat{\mathbf{d}} = \mathbf{x} - \boldsymbol{\alpha} \hat{s}_1, \tag{13}$$

where $\hat{\mathbf{n}}$ and $\hat{\mathbf{d}}$ denote the estimates of the $\mathbf{n}_k$'s and $\mathbf{d}_k$'s, respectively, stacked over all nodes.

C. Distributed CCA-based eye blink artifact removal

Keeping in mind the need to avoid energy-inefficient data centralization in WESNs, we are interested in the capabilities of CCA to remove eye blink artifacts in a distributed fashion. An algorithm for Distributed CCA (DCCA) is derived in [28]. In the sequel, we assume that the WESN is fully connected, i.e., a signal broadcast by a node is received by all other nodes. However, this is mainly for the sake of an easy exposition, as DCCA can also be formulated for partially connected networks with short-distance communication between nearby nodes [28].

Assuming we aim to extract the Q first principal CCA components, then the DCCA algorithm lets each node perform a linear compression of its M_k signals into a Q-channel signal, which is then transmitted to the other nodes. As experiments in [11] have demonstrated, eye blink artifacts can be assumed to be one-dimensional (i.e., they appear as scaled versions of a single-channel signal throughout all EEG channels). Furthermore, we will demonstrate in Section IV that the eye blink artifacts appear in the first principal CCA component if the time lag τ is properly chosen. This means that in our case $Q = 1$ and each node k linearly compresses its M_k EEG channels into a single-channel signal, denoted as

$$\bar{x}_k = \mathbf{c}_k^T \mathbf{x}_k, \tag{14}$$

where $\mathbf{c}_k$ is a $M_k \times 1$ compression vector which will be defined later on. The DCCA algorithm will automatically learn an optimal compression vector $\mathbf{c}_k$, such that *all* the nodes are able to compute the *centralized* principal CCA component, allowing to remove the eye blink artifacts as if each node would have access to all the channels in the WESN. This means that, although (14) is a lossy compressor in general, it converges to a compressor that is lossless with respect to the information that is needed to solve the network-wide CCA problem.

The $(K-1)$-channel signal containing the compressed signals from the other nodes that node k receives is denoted as $\bar{\mathbf{x}}_{-k}$. The subscript $-k$ indicates that the compressed signal from node k itself is not included in $\bar{\mathbf{x}}_{-k}$, i.e.,

$$\bar{\mathbf{x}}_{-k} = [\bar{x}_1, \ \dots \ , \bar{x}_{k-1}, \bar{x}_{k+1}, \ \dots \ , \bar{x}_K]^T. \tag{15}$$

In the DCCA algorithm with $Q = 1$, each node keeps track of its local compression vector $\mathbf{v}_k$, which is initialized randomly. The nodes take turns to update their compression vector until its entries have converged to stable values. The signal flow and updating step of a node k within the DCCA algorithm is depicted in Fig. 1 and goes as follows: the node collects its own M_k EEG channels, $\mathbf{x}_k$, and stacks them together with the $K-1$ compressed signals from the other nodes, $\bar{\mathbf{x}}_{-k}$, into an $(M_k + K - 1)$-channel signal:

$$\widetilde{\mathbf{x}}_k = \begin{bmatrix} \mathbf{x}_k \\ \bar{\mathbf{x}}_{-k} \end{bmatrix}. \tag{16}$$

The same stacking operation is performed on the time-delayed signals to produce $\widetilde{\mathbf{x}}_{\tau,k}$, where subscript τ indicates the time-delayed version, i.e.

$$\widetilde{\mathbf{x}}_{\tau,k}[t] \triangleq \widetilde{\mathbf{x}}_k[t - \tau]. \tag{17}$$

These two stacked signals are used as the inputs for a local CCA operation (only to extract the first principal CCA component), which allows to construct the principal CCA direction $\widetilde{\mathbf{v}}_k$ of dimension $(M_k + K - 1)$ as described in Subsection

II-B. From [28], it is then found that the linear compressor $\mathbf{c}_k$ should be updated by replacing it with the first M_k elements of $\widetilde{\mathbf{v}}_k$, i.e., the part of $\widetilde{\mathbf{v}}_k$ that is applied to the local channels $\mathbf{x}_k$. Hence, the first M_k elements of $\widetilde{\mathbf{v}}_k$ are copied into $\mathbf{c}_k$ as indicated by the dashed line in Fig. 1. Once the node compression vector $\mathbf{c}_k$ is found, it is applied to $\mathbf{x}_k$. Note that the role of vector $\mathbf{c}_k$ in each node k is twofold: it compresses the local EEG channels into a single-channel signal that is broadcast to other nodes, but the compressor is also part of the estimator $\widetilde{\mathbf{v}}_k$ that estimates the eye blink source signal $\hat{s}_{1,k} = \widetilde{\mathbf{v}}_k^T \widetilde{\mathbf{x}}_k$ at node k.

If each node computes its local reduced-dimension CCA in a sequential round-robin fashion, each time adapting its local compression vector $\mathbf{c}_k$ accordingly (while the other nodes keep their $\mathbf{c}_k$ fixed), then it can be shown that all the compressors $\mathbf{c}_k, \forall\, k \in \{1, \ldots, K\}$ will converge, after which all the local estimates $\hat{s}_{1,k} = \widetilde{\mathbf{v}}_k^T \widetilde{\mathbf{x}}_k$ will be the same as the signal $\hat{s}_1$ computed by the centralized CCA [28], i.e., as if each node would have access to all channels. The iterations can each time be re-computed over the same signal segment, resulting in an increased communication bandwidth. However, since the spatial time-varying pattern $\mathbf{a}_1$ of the eye blink signal is typically not or only slowly time-varying, the iterations of DCCA can be spread out over different signal segments in a time-recursive fashion, similar to an adaptive filter.

After estimating the eye blink source signal $\hat{s}_{1,k}$, each node will compute a set of scaling factors $\boldsymbol{\alpha}_k$ in an analogous way as in (10). Again, this can be done in practice by performing a LS fit of $\hat{s}_{1,k}$ to each of the node's M_k local EEG channels in a least squares sense, by solving

$$\boldsymbol{\alpha}_k = [\alpha_{k,1},\ \ldots\ , \alpha_{k,M_k}]^T = (\hat{S}_{1,k}\hat{S}_{1,k}^T)^{-1}\mathbf{X}_k\hat{S}_{1,k}^T, \quad (18)$$

where $\hat{S}_{1,k}$ and $\mathbf{X}_k$ are defined as the node-specific time signals containing observations of respectively $\hat{s}_{1,k}$ and $\mathbf{x}_k$ in a window of length L, analogous to (8) and (9), and where $\alpha_{k,i}$ is the scaling factor that fits the eye blink estimate $\hat{s}_{1,k}$ to EEG channel i of node k. The artifacts in node k can then be subtracted to find the clean EEG channels as

$$\hat{\mathbf{n}}_k = \mathbf{x}_k - \hat{\mathbf{d}}_k = \mathbf{x}_k - \boldsymbol{\alpha}_k \cdot \hat{s}_{1,k}. \quad (19)$$

As mentioned earlier, it can be shown that the first principal CCA component $\hat{s}_{1,k}$ estimated in every node with the DCCA scheme converges to the same first component estimated with a centralized CCA over all channels [28]. This result is remarkable, since the problem statement of CCA in (5) requires the network-wide covariance matrices to be computed, which is not done anywhere during the DCCA algorithm as it would be too bandwidth- and energy-intensive in a WESN. Each node only uses local correlations between its own channels and compressed channels from other nodes without losing performance compared to centralized CCA. It should be noted that such equivalence with the centralized CCA does not hold for the other CCA components, since information about them is completely lost in the compression to a single-channel signal. However, for our purpose, where the eye blink artifacts are separable in the first component, this is not a problem. Nevertheless, the other components can be estimated if an increase in bandwidth is allowed (see [28]). In conclusion, with DCCA, a substantial reduction in communication bandwidth is achieved while maintaining the same artifact removal accuracy as in the centralized case, because DCCA exploits the underlying one-dimensional character of the eye blink artifacts.

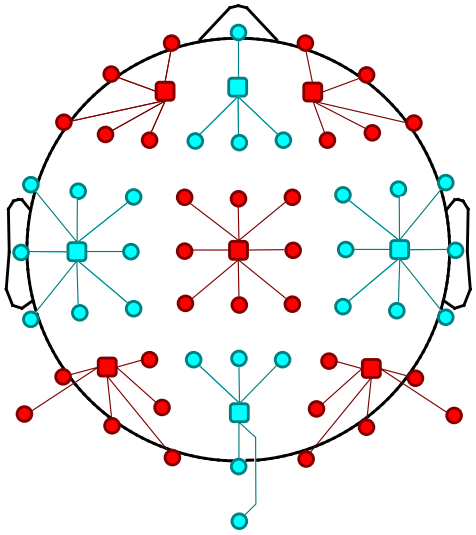


Fig. 2. Clustering of the 64 EEG electrodes into 9 nodes. The electrodes are positioned according to the 10-20 system. Each node consists of several recording electrodes (circles) and one fusion center with processor and radiotransmitter (squares). The recording electrodes are referenced locally to an electrode in the fusion center.

III. Experiment

A. EEG recording sessions

EEG measurements are performed with a 64-electrode EEG system (Biosemi) with international 10-20 electrode placement configuration. Several short EEG recordings of 9 subject are obtained. The subjects blink the eyes every 5 seconds in order to purposefully induce artifacts in the recordings at a controlled rate. The subjects do not perform any other physical or mental tasks during the recording. The data is recorded at a sample rate of 8192 Hz and later downsampled to 200 Hz in a referential montage where electrode Cz is the reference electrode. A mean subtraction on the EEG data is performed.

B. Emulating the WESN

In the envisioned WESN, the different sensor nodes are galvanically isolated from each other. This means that no common reference channel for the entire EEG electrode set can be chosen. In each node, one channel needs to be selected as the local reference channel for that node.

We emulate a WESN by partitioning the complete set of electrodes into 9 nodes, as is shown in Fig. 2. Within each node, one electrode is selected as the local reference. The choice of a local reference electrode within a node does not matter significantly, however electrodes near the sides of the head and neck are more prone to pick up muscle and electrode displacement artifacts, and should be avoided. The local reference electrodes are chosen to be located centrally

within their node, as indicated by the squares in Fig. 2. All other channels of that node are re-referenced with respect to the chosen reference electrode. With r the index of the channel corresponding to the selected reference electrode in node k when referenced to Cz, the re-referencing operation can be represented as a linear transformation as

$$\mathbf{x}_{k,ref} = \begin{bmatrix} \mathbf{I}_{(r-1)} & & \mathbf{O}_{(r-1)\times(M_k-r)} \\ -\mathbf{1}_{M_k-1} & & \\ \mathbf{O}_{(M_k-r)\times(r-1)} & & \mathbf{I}_{(M_k-r)} \end{bmatrix} \cdot \mathbf{x}_k \tag{20}$$

where $\mathbf{I}_n$ denotes an $n \times n$ identity matrix, $\mathbf{O}_{m\times n}$ denotes an $m \times n$ all-zero matrix, and $-\mathbf{1}_n$ denotes a vector of length n containing only -1, and where $\mathbf{x}_k$ denotes the vector with all channels that correspond to the electrodes of node k. The re-referencing matrix is an identity matrix of which the column corresponding to the reference channel in node k only contains -1, and the row corresponding to the reference channel in node k is discarded. Note that the re-referencing matrix is $(M_k - 1 \times M_k)$-dimensional.

From a mathematical point of view, the information in the re-referenced data sets do not differ much from the original ones, since they are obtained by a full rank linear transformation. This will not impact the linear mixing model which is assumed in source separation algorithms such as CCA and ICA. There is, however, a reduction by 1 in the number of channels of each node, since the transformation matrix is non-square and hence the original data can not be perfectly reconstructed any more. This implies a loss of information due to the galvanic decoupling between the nodes. It is noted that the channels of each node float with respect to other nodes, so that network-wide mixing vectors cannot be interpreted across nodes any more.

In the emulated network, all wireless transmission channels are assumed to be ideal. The processing in all further experiments are performed in batch mode, i.e. all correlation matrices are computed over the full signal length.

C. Synthetic EEG data with known ground truth

In order to verify the artifact removal algorithms, a set of synthetic EEG data is constructed in addition to the real artifact data collected in the experiment described in Subsection III-A. The EEG recordings of all subjects were cleared of major artifacts with an ICA decomposition and manual component rejection. In one subject, some heavily corrupted data segments are cut out. The resulting data set is considered as clean EEG data being the ground truth, i.e. $\mathbf{n}_k$ in (1).

An eye blink template is created by cutting out an eye blink segment from one subject and low-pass filtering it. This template is repeated at random times with an average occurrence of 1 blink every 5 seconds, resulting in the single-channel eye blink artifact source s_1. A mixing vector $\mathbf{a}_1$ is constructed by estimating eye blink artifact amplitudes in real EEG data. The synthetic data is obtained by mixing the artifact source over all channels as $\mathbf{d} = \mathbf{a}_1 s_1$, and adding the resulting multi-channel artifact signal $\mathbf{d}$ to the ground truth. This constructs a synthetic EEG data set conform to the artifact model in (3), with realistic eye blink amplitudes that are higher in the channels near the eyes.

D. Effect of time lag

The choice of the time lag τ in (7) is an important parameter for the performance of CCA. As indicated in Subsection II-B, for a good choice of τ, the autocorrelation of the eye blink artifacts should be sufficiently larger than the autocorrelation of the other EEG sources at that same time lag. If this is the case, the eye blink artifact source can be extracted from the EEG data in the first principal CCA component.

At the given sample rate of $200\,\mathrm{Hz}$, the time lag can be tuned with a resolution of $5\,\mathrm{ms}$. We will investigate whether the same choice of time lag for all subjects suffices to remove eye blink artifacts. For the purpose of comparison, the optimal value will be sought for each individual subject.

E. Effect of input Signal-to-Noise Ratio

It is informative to evaluate the performance of CCA as a function of the artifact amplitude, relative to the background EEG level. We can only do this for the case of synthetic EEG data, where the ground truth is known. The synthetic multi-channel artifact signal $\mathbf{d}$ is scaled with a scaling factor γ in order to obtain a range of Signal-to-Noise Ratio (SNR) values. We define the input SNR as the power of the pure artifacts d_1 scaled with γ, divided by the power of the clean EEG data n_1, converted to decibel, i.e.

$$SNR = 10\log_{10}\frac{E\{(\gamma d_1)^2\}}{E\{(n_1)^2\}}. \tag{21}$$

Note that we selected channel 1 as the reference for defining the SNR, which corresponds with electrode Fp1 near the eyes. This is just an arbitrary choice in order to define an SNR measure, where preferably a channel with large eye blink amplitude (as in Fp1) is used. The artifact scaling factor γ can be chosen in order to obtain a desired SNR, e.g. we can create synthetic data with an SNR of 3 dB by choosing $\gamma = 2$.

Because the SNR varies between subjects, it should be normalized to allow for inter-subject comparisons. A subject's SNR values are normalized by subtracting the SNR when $\gamma = 1$, i.e. the unscaled synthetic EEG data as constructed in Subsection III-C. This defines the normalized SNR values as

$$SNR_{\text{norm}} = SNR - SNR|_{\gamma=1}\,. \tag{22}$$

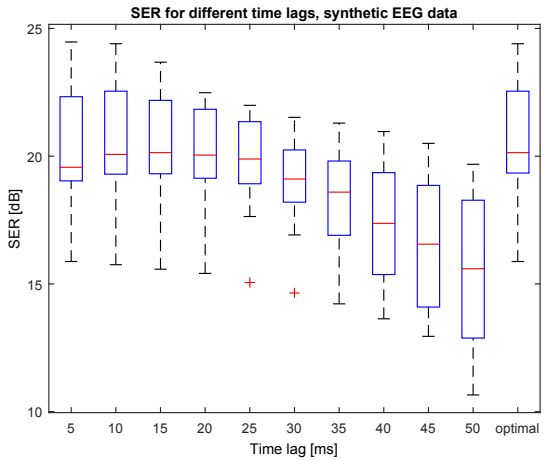
This means that, for the case where $\gamma = 1$, SNR_{norm} is 0 dB for all subjects. From here onwards, we will always refer to normalized SNR values.

IV. Results

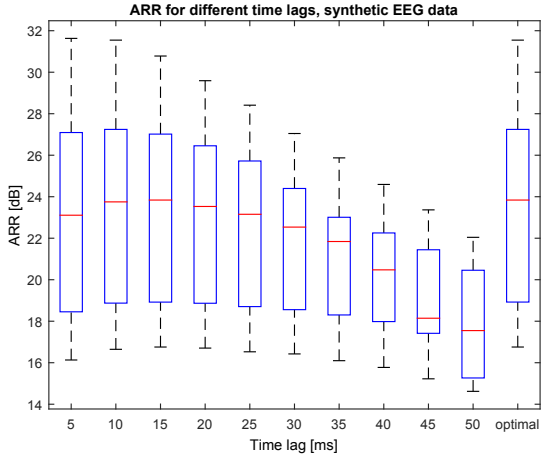
A. Performance measures

We define a first measure that indicates to what degree the actual EEG component (i.e. $\mathbf{n}_k$ in (1)) is altered by applying the artifact removal algorithm. Since the estimated artifacts are subtracted from the EEG data, the estimated artifacts $\hat{\mathbf{d}}_k$ should be small in the segments where no artifacts are present in order to not add noise or remove actual EEG data in (13) and (19). We propose the following Signal-to-Error Ratio (SER):

$$SER = \frac{1}{M}\sum_{i=1}^{M} p_i \cdot 10\log_{10}\frac{E\{(x_i)^2\}}{E\{(\hat{d}_i)^2\}}\bigg|_{\text{clean segments}} \tag{23}$$

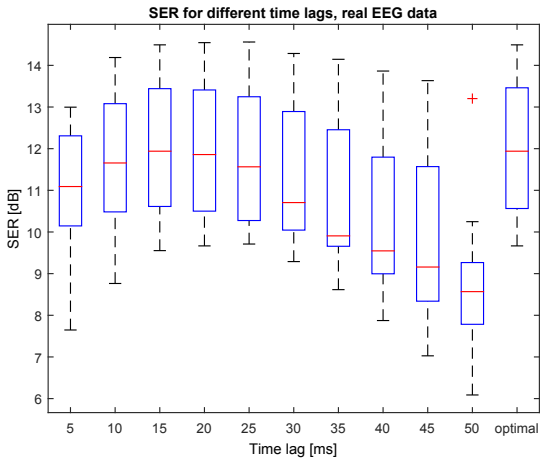


(a) SER

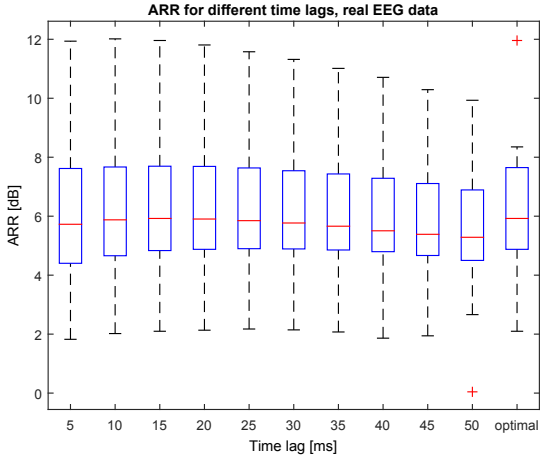


(b) ARR

Fig. 3. Effect of time lag τ on performance measures for synthetic EEG data.



(a) SER



(b) ARR

Fig. 4. Effect of time lag τ on performance measures for real measured EEG data.

where i denotes the index of the channel, and both numerator and denominator are evaluated in the artifact-free (clean) segments. The individual SER for each channel is averaged over all channels with weights p_i to obtain a single number as performance measure. The channel-specific weights p_i are defined as the estimated artifact power in channel i, obtained by subtracting the measured signal power in the clean segments from the signal power in the segments where artifacts are present:

$$p_i = E\{(x_i)^2\}\big|_{\text{corrupted segments}} - E\{(x_i)^2\}\big|_{\text{clean segments}}. \quad (24)$$

In the chosen weighting approach, artifact removal in channels that are more heavily corrupted with artifacts has a greater contribution to the overall performance measure.

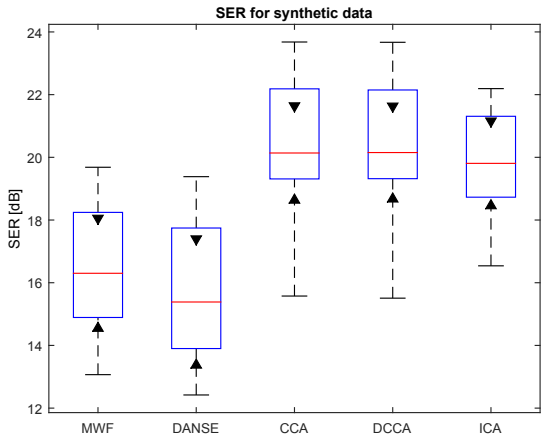
The second performance measure indicates whether the estimated artifacts resemble the actual artifact shapes closely. This means that the residue in every channel, obtained as $d_i - \hat{d}_i$, should be as small as possible. This leads to the following Artifact-to-Residue Ratio (ARR):

$$ARR = \frac{1}{M}\sum_{i=1}^{M} p_i \cdot 10\log_{10}\left.\frac{E\{(d_i)^2\}}{E\{(d_i-\hat{d}_i)^2\}}\right|_{\text{corrupted segments}} \quad (25)$$

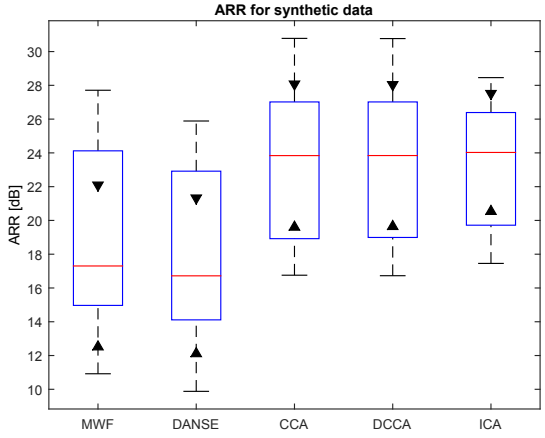
where both numerator and denominator are evaluated only in the segments that contain artifacts. However, in the case of real EEG data, the artifact signal $\mathbf{d}$ is unknown. The ARR can only be used as a performance measure in synthetic data where a ground truth is known. An estimate of the ARR for real data can be obtained by approximating the pure artifact signal $\mathbf{d}$ with the measured signal $\mathbf{x}$ in the corrupted segments, based on the fact that the amplitude of the artifact is significantly higher than the EEG amplitude. This leads to the following ARR measure for real data:

$$ARR = \frac{1}{M}\sum_{i=1}^{M} p_i \cdot 10\log_{10}\left.\frac{E\{(x_i)^2\}}{E\{(x_i-\hat{d}_i)^2\}}\right|_{\text{corrupted segments}}. \quad (26)$$

SER and ARR should be evaluated simultaneously as they measure different quality aspects of artifact removal. Good

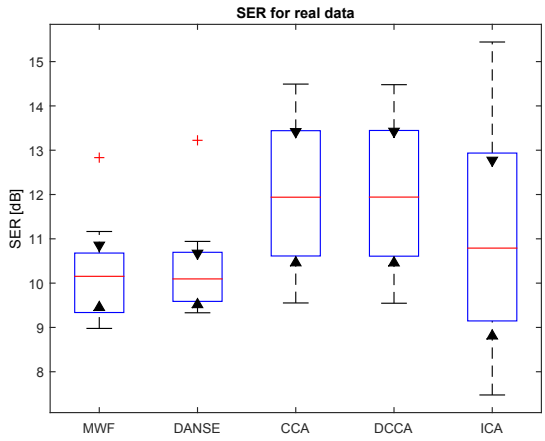


(a) SER

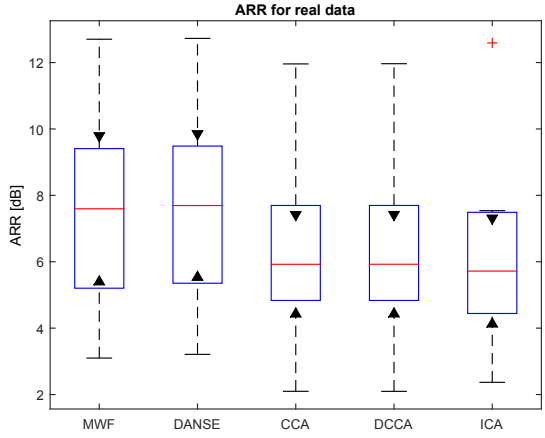


(b) ARR

Fig. 5. Results of applying different artifact removal methods on the synthetic EEG data.



(a) SER



(b) ARR

Fig. 6. Results of applying different artifact removal methods on the real measured EEG data.

performance is indicated by both a high SER and a high ARR.

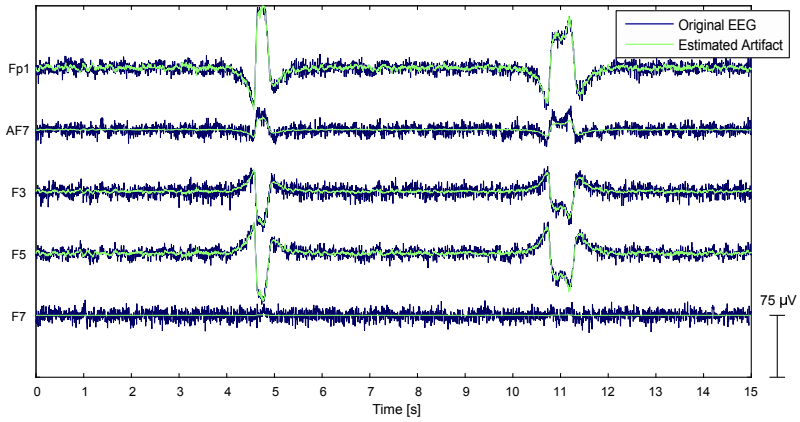
B. Effect of time lag

We perform CCA for different time lags τ on both the synthetic and real data. Fig. 3 shows boxplots of the SER and ARR for the synthetic data of all 9 subjects, with time lags varying from 5 to $50\,\mathrm{ms}$. The rightmost boxplot represents the performance measures in case the best time lag is chosen for each subject individually. Outliers are shown as '+' symbols.

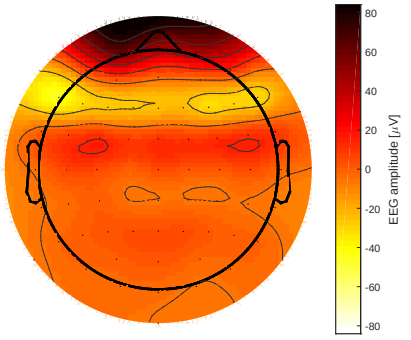
Fig. 4 shows the same, but this time with the real data.

C. Results on synthetic data

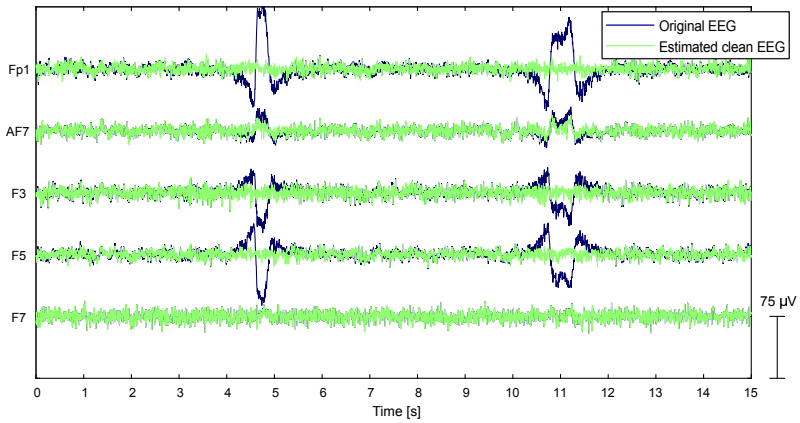
To our knowledge (D)CCA has not been applied yet as a method for eye blink artifact removal in EEG recordings without an EOG reference. Therefore, it is instructive to compare it with other algorithms applied to the same set of EEG recordings, while using the same performance measures. The results for the synthetic data are shown in Fig. 5 for 5 different competing algorithms for eye blink artifact removal: MWF [11], [23], DANSE [11], [23], [31], ICA [22], CCA and DCCA. The MWF method relies on separating the data in segments with and without artifacts to compute the filter coefficients. This is achieved for eye blink artifacts by using a threshold. DANSE is a distributed realization of a MWF, which makes it interesting to compare with DCCA. For the results with ICA, the data is processed with the ICA infomax implementation of EEGLAB [22] to separate the data in independent components. The components corresponding to eye blinks are selected manually by observing their time domain representation and spectrum, and are removed from the data. For 7 out of 9 subjects, the eye blink components are captured in one independent component, for the others multiple components needed to be removed. For both CCA and DCCA, a time lag of $15\,\mathrm{ms}$ is used for all subjects, which was chosen based on the results in Subsection IV-B (see also the discussion on time lags in Subsection V-A). It is noted that only CCA and DCCA did not rely on a manual intervention: for MWF/DANSE, the eye blink segments had to be identified based on a subject-dependent threshold and a manual validation. For ICA, the eye blink components were manually selected from the full set of independent



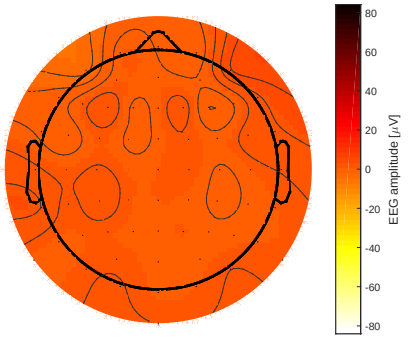
(a)



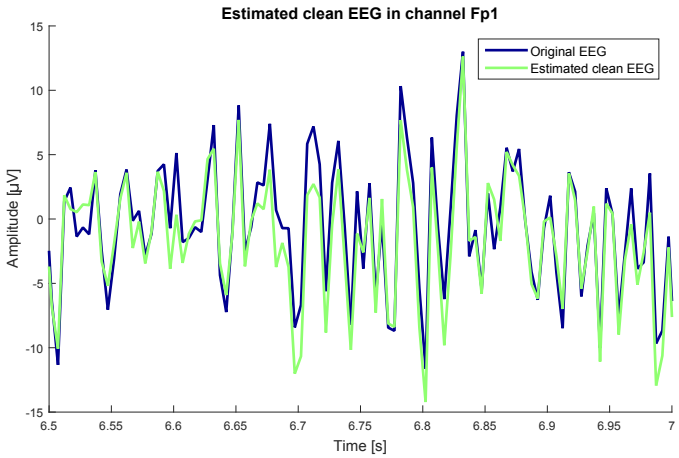
(b)



(c)



(d)



(e)

Fig. 7. Illustration of eye blink artifact removal using DCCA in an emulated WESN. (a) Eye blink artifacts plotted on top of the original EEG signals as measured by WESN node 1 (channels re-referenced to electrode AF3). (b) Topographic plot of the original EEG amplitude across the scalp during an eye blink. (c) EEG data after eye blink artifact removal using DCCA, plotted on top of the original EEG signals as measured by WESN node 1. (d) Topographic plot of the EEG amplitude across the scalp during an eye blink, after eye blink artifact removal. (e) Close-up view of the estimated clean EEG data in channel Fp1, plotted on top of the original EEG.

components.

On Fig. 5, the black triangles indicate the edges of the 95% confidence intervals for the median. If confidence intervals of different boxplots do not overlap, it can be concluded with 95% confidence that the medians are significantly different.

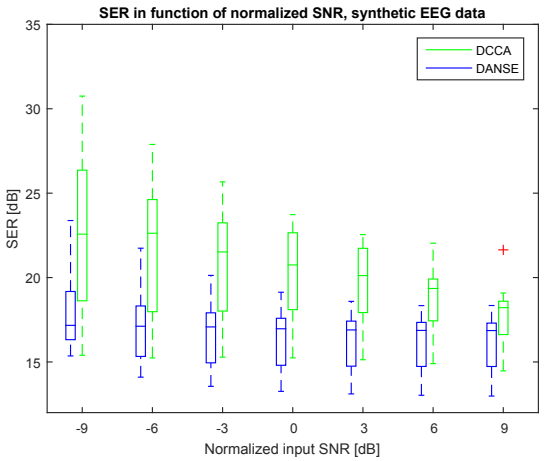
D. Results on real EEG data

The results of applying different eye blink artifact removal algorithms on the real data are shown in Fig. 6 in an analogous way as for the synthetic data. For CCA and DCCA, again a time lag of 15 ms is used.

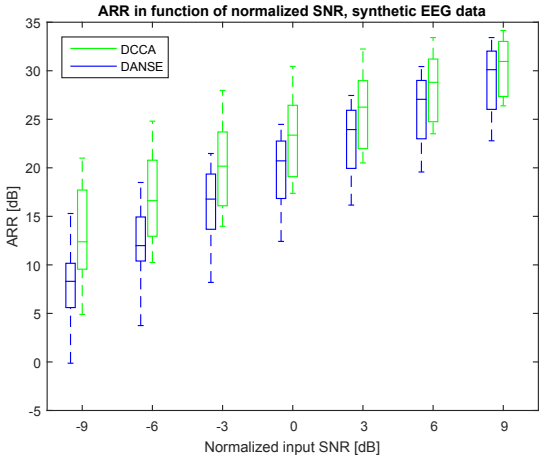
For illustrative purposes, the result of applying distributed CCA on a real EEG measurement, where again a time lag of 15 ms is used, is shown in Fig. 7. In Fig. 7(a), the channels of node 1 of the WESN are shown, as well as the eye blink artifact that is estimated by DCCA in each of those channels. The same channels after removing the estimated artifacts are shown in Fig. 7(c). Note that channel F7 shows no clear eye blink artifacts, despite being relatively close to the eyes. This can be explained from the fact that this node is referenced with respect to electrode AF3. In the emulated WESN, channels with similar eye blink artifact amplitude as their local node reference will have their artifact amplitude strongly reduced. Also, channels in more posterior positions than the local reference channel show an inversion of the eye blink artifact. Fig. 7(e) shows a close-up view of a signal segment between two artifacts in channel Fp1. The original EEG is barely altered in signal segments in between the artifacts. To give an illustration of the spatial pattern of the eye blink artifacts, Fig. 7(b) shows a topographic scalp plot of the original EEG amplitude during an eye blink, as captured by the electrodes of the emulated WESN. As can be expected, the largest amplitudes are found near the eyes. However, due to the introduction of local references by the WESN, the artifact amplitudes don't decrease monotonically from frontal to posterior channels. Fig. 7(d) shows the EEG amplitude at the same point in time, but after artifact removal using DCCA. The performance measures for this example of artifact removal are a SER of 10.54 dB and an ARR of 11.96 dB.

E. Effect of input SNR

As explained in Subsection III-E, the normalized SNR of the synthetic EEG data can be altered to study the effect of artifact amplitude relative to the background EEG on performance. The study is carried out for both distributed algorithms DCCA and DANSE. The results are shown on Fig. 8. The normalized SNR of 0 dB corresponds to the synthetic EEG data with realistic artifact amplitudes derived from real EEG data. A normalized SNR greater than 0 dB corresponds to synthetic EEG where the artifact amplitudes are increased compared to the original synthetic data set of Subsection III-C. An SNR smaller than 0 dB corresponds to synthetic EEG where the artifacts are downscaled. Every step of 3 dB corresponds to an artificial doubling or halving of the eye blink artifact power. For normalized SNRs lower than -6 dB, the artifact amplitudes are smaller than the background EEG.



(a) SER



(b) ARR

Fig. 8. Performance measures as a function of normalized input SNR.

V. Discussion

A. Choice of time lag

From the results of the time lag study depicted in Fig. 3 (synthetic EEG data) and Fig. 4 (real EEG data), we can infer how the time lag should be chosen. For increasing time lags, a downward trend is observed in the boxplots. This can be explained from the fact that the autocorrelation of the sources in $\mathbf{s}$ typically becomes smaller if the time lag increases, and hence also the canonical correlation coefficients of the different CCA components become smaller, with smaller differences between them. This leads to worse source separation by CCA and thus worse artifact removal. For time lags over 50 ms, the eye blink artifacts are no longer contained in the first CCA component for all subjects and hence manual selection of the eye blink component is needed. The artifacts not being present in the first component has an additional disadvantage for the distributed computation, since more than one CCA component then has to be computed. As explained in Subsection II-C, the number of channels that needs to be transmitted between the nodes is equal to the number of principal CCA components that need to be computed.

For very small time lags, slowly varying neural activity also shows high temporal correlation with its time-delayed waveform, resulting in high canonical correlations corresponding to the components that capture these sources. Choosing the time lag too small thus also leads to poorly separated CCA components.

These lag-dependent effects are observable in Fig. 3 and Fig. 4, however the ARR is less susceptible to variations in time lag, especially for real EEG data. The rightmost boxplot in each of the figures represents the results where for each of the 9 subjects, the time lag was chosen in a subject-dependent fashion to produce the highest performance measures. In practice these optimal time lags were all found within the 5 - $20\,\mathrm{ms}$ range. There is barely any observable difference between this "optimal" subject-dependant choice of τ versus choosing the same time lag for all subjects, as long as it is in the 10 to $20\,\mathrm{ms}$ range.

Therefore, we propose, as a rule-of-thumb for separating the eye blink artifact component with CCA-based methods, to choose the time lag to be $15\,\mathrm{ms}$. In both synthetic as in real EEG data, this parameter choice leads to near-optimal algorithm performance for all subjects, and it makes sure the eye blinks are present in the first CCA component without leakage to other components. This results in a parameter setting that can be generally applied to all subjects without significant performance loss, and removes the need for manual artifact component selection.

B. Performance in varying SNR conditions

Varying the input SNR influences the SER and ARR differently, as shown on Fig. 8. The ARR, which measures accurateness of the artifact estimate, shows a clear upward trend with increasing SNR. This is also intuitive, as a higher input SNR of the artifacts allows for a more accurate estimation. However, this seems to result in a corresponding decrease in SER, which measures undesired distortion of the EEG after applying artifact removal. This exposes a trade-off: at low SNRs the artifacts are less accurately estimated, but the distortion of the clean EEG is smaller. A similar trade-off can be seen for DANSE, which is consistently outperformed by DCCA.

Of course, for real EEG measurements, the SNR of artifacts relative to the background EEG cannot be chosen, but are fixed at a certain SNR depending on the subject. However, this study shows that DCCA consistently captures the eye blink artifact source in the principal CCA component over a wide range of input SNRs.

C. Comparison of eye blink artifact removal algorithms

From the results it is clear that performance of CCA-based methods is on par with ICA, the current state-of-the-art method for EEG eye blink artifact rejection, for both of our performance parameters. Besides this, the CCA-based methods have the additional advantages that they are more energy-efficient due to a lower complexity, there is no need for a manual component selection, and they facilitate a distributed realization.

When comparing CCA with the centralized MWF, and its adaptive distributed implementation DANSE [11], [23], the results indicate a superior performance of the CCA-based methods for the SER (i.e. the degradation of artifact-free segments) parameter, and an equal performance for the ARR parameter (i.e. the ability to estimate the artifact shape accurately). An advantage of ICA and CCA-based methods over the MWF-based methods is the property of blind source separation. For the MWF and DANSE, a separation of the signals in clean and corrupted segments is required based on a subject-specific threshold, which is prone to misclassification errors.

DCCA achieves a high bandwidth reduction compared to the centralized algorithms: each node compresses its local EEG channels to a single-channel signal that is sent to all other nodes. For K nodes with each M_k channels, the compression factor is M_k. Also the computational complexity per node is reduced: an M-channel CCA is a $\mathcal{O}(M^3)$ procedure since it can be solved as a generalized eigenvalue decomposition [32]. In the emulated WESN, each node only performs a local $(M_k + K - 1)$-channel CCA.

From this comparison we conclude that DCCA is a well-suited method to apply to distributed artifact removal in a WESN with bandwith and energy constraints.

VI. Conclusions

In this paper, we have successfully used distributed CCA to remove eye blink artifacts in EEG recordings from 9 subjects, using both synthetic and real EEG data. In order to demonstrate the algorithm performance, we emulated a distributed wireless EEG network where each node in the network uses a local reference. We found that the only important parameter to be tuned, the time lag, can be chosen to be $15\,\mathrm{ms}$ for all nodes in the network and across subjects. For this setting, the eye blinks always occur in the first CCA component without significant leakage to other components. This means that the artifact rejection using DCCA can happen in an unsupervised way.

Comparing DCCA with DANSE, which is another distributed algorithm that is applicable to distributed eye blink artifact removal, we have found that DCCA achieves a better performance while achieving the same bandwidth reduction. Furthermore, for the envisioned application in WESNs, DCCA has the advantage of unsupervised artifact removal. Comparing DCCA to ICA, which is the current state-of-the-art for EEG artifact removal, shows no significant performance differences. However, ICA is not amenable to a distributed realization and has a higher complexity.

VII. Acknowledgments

The authors declare no conflict of interest. This work was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Fund KU Leuven BOF/STG-14-005, CoE PFV/10/002 (OPTEC), projects of the Research Foundation Flanders (FWO) nr. G.0931.14 'Design of distributed signal processing algorithms and scalable hardware platforms for energy-vs-performance adaptive wireless acoustic sensor

networks', and nr. G0D7516N 'Distributed signal processing algorithms for spike sorting in next-generation high-density neuroprobes', and HANDiCAMS. The project HANDiCAMS acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 323944. The scientific responsibility is assumed by its authors.

REFERENCES

[1] Brown, L., Van de Molengraft, J., Yazicioglu, R. F., Torfs, T., Penders, J., and Van Hoof, C. (2010) A low-power, wireless, 8-channel EEG monitoring headset. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE* IEEE pp. 4197–4200.

[2] Chi, Y. M., Wang, Y., Wang, Y.-T., Jung, T.-P., Kerth, T., and Cao, Y. (2013) A practical mobile dry EEG system for human computer interfaces. In *Foundations of Augmented Cognition* pp. 649–655 Springer.

[3] Casson, A. J., Yates, D. C., Duncan, J. S., Rodriguez-Villegas, E., et al. (2010) Wearable electroencephalography. *Engineering in Medicine and Biology Magazine, IEEE,* **29**(3), 44–56.

[4] Gyselinckx, B., Van Hoof, C., Ryckaert, J., Yazicioglu, R. F., Fiorini, P., and Leonov, V. (2005) Human++: autonomous wireless sensors for body area networks. In *Custom Integrated Circuits Conference, 2005. Proceedings of the IEEE 2005* IEEE pp. 13–19.

[5] Biesmans, W., Das, N., Francart, T., and Bertrand, A. (2016) Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario. Accepted for publication in IEEE Trans. Neural Systems and Rehabilitation Engineering.

[6] Van Eyndhoven, S., Francart, T., and Bertrand, A. (2016) EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses. *IEEE Transactions on Biomedical Engineering,*.

[7] Looney, D., Kidmose, P., Park, C., Ungstrup, M., Rank, M. L., Rosenkranz, K., and Mandic, D. P. (2012) The in-the-ear recording concept: User-centered and wearable brain monitoring. *Pulse, IEEE,* **3**(6), 32–42.

[8] Debener, S., Emkes, R., De Vos, M., and Bleichner, M. (2015) Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Scientific reports,* **5**.

[9] Mc Laughlin, M., Lu, T., Dimitrijevic, A., and Zeng, F.-G. (2012) Towards a closed-loop cochlear implant system: Application of embedded monitoring of peripheral and central neural activity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* **20**(4), 443–454.

[10] Sockalingam, R., Holmberg, M., Eneroth, K., and Shulte, M. (2009) Binaural hearing aid communication shown to improve sound quality and localization. *The Hearing Journal,* **62**(10), 46–47.

[11] Bertrand, A. (2015) Distributed signal processing for wireless EEG sensor networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* **23**(6), 923–935.

[12] Wu, M.-F. and Wen, C.-Y. (2012) Distributed cooperative sensing scheme for wireless sleep EEG measurement. *Sensors Journal, IEEE,* **12**(6), 2035–2047.

[13] Chi, Y. M., Deiss, S. R., and Cauwenberghs, G. (2009) Non-contact low power EEG/ECG electrode for high density wearable biopotential sensor networks. In *Wearable and Implantable Body Sensor Networks, 2009. BSN 2009. Sixth International Workshop on* IEEE pp. 246–250.

[14] Wu, M.-F. and Wen, C.-Y. (2009) The Design of Wireless Sleep EEG Measurement System With Asynchronous Pervasive Sensing.. In *SMC* pp. 714–721.

[15] Norton, J. J., Lee, D. S., Lee, J. W., Lee, W., Kwon, O., Won, P., Jung, S.-Y., Cheng, H., Jeong, J.-W., Akce, A., et al. (2015) Soft, curved electrode systems capable of integration on the auricle as a persistent brain–computer interface. *Proceedings of the National Academy of Sciences,* **112**(13), 3920–3925.

[16] Bleichner, M. G., Lundbeck, M., Selisky, M., Minow, F., Jäger, M., Emkes, R., Debener, S., and De Vos, M. (2015) Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see?. *Physiological reports,* **3**(4), e12362.

[17] Sun, M., Jia, W., Liang, W., and Sclabassi, R. J. (2012) A low-impedance, skin-grabbing, and gel-free EEG electrode. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* IEEE pp. 1992–1995.

[18] Lee, S. M., Kim, J. H., Park, C., Hwang, J.-Y., Hong, J. S., Lee, K. H., and Lee, S. H. (2016) Self-Adhesive and Capacitive Carbon Nanotube-Based Electrode to Record Electroencephalograph Signals From the Hairy Scalp. *IEEE Transactions on Biomedical Engineering,* **63**(1), 138–147.

[19] Kim, D.-H., Lu, N., Ma, R., Kim, Y.-S., Kim, R.-H., Wang, S., Wu, J., Won, S. M., Tao, H., Islam, A., et al. (2011) Epidermal electronics. *science,* **333**(6044), 838–843.

[20] Yates, D. C. and Rodriguez-Villegas, E. (2007) A key power trade-off in wireless EEG headset design. In *Neural Engineering, 2007. CNE'07. 3rd International IEEE/EMBS Conference on* IEEE pp. 453–456.

[21] Croft, R. and Barry, R. (2000) Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique/Clinical Neurophysiology,* **30**(1), 5–19.

[22] Delorme, A. and Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods,* **134**(1), 9–21.

[23] Bertrand, A. and Moonen, M. (May, 2014) Distributed eye blink artifact removal in a wireless EEG sensor network. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal processing (ICASSP)*.

[24] Zhang, L., Wang, Y., and He, C. (2012) Online removal of eye blink artifact from scalp EEG using canonical correlation analysis based method. *Journal of Mechanics in Medicine and Biology,* **12**(05), 1250091.

[25] Xie, J., Qiu, T., and Wen-hong, L. (2013) An ocular artifacts removal method based on canonical correlation analysis and two-channel EEG recordings. In *World Congress on Medical Physics and Biomedical Engineering May 26-31, 2012, Beijing, China* Springer pp. 465–468.

[26] De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., and Van Huffel, S. (2006) Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Transactions on Biomedical Engineering,* **53**(12), 2583–2587.

[27] Hotelling, H. (1936) Relations between two sets of variates. *Biometrika,* pp. 321–377.

[28] Bertrand, A. and Moonen, M. (2015) Distributed canonical correlation analysis in wireless sensor networks with application to distributed blind source separation. *IEEE Transactions on Signal Processing,* **63**(18), 4800–4813.

[29] Borga, M. and Knutsson, H. (2001) A canonical correlation approach to blind source separation. *Report LiU-IMT-EX-0062 Department of Biomedical Engineering, Linkoping University,*.

[30] Liu, W., Mandic, D. P., and Cichocki, A. (2007) Analysis and online realization of the CCA approach for blind source separation. *IEEE Transactions on Neural Networks,* **18**(5), 1505–1510.

[31] Bertrand, A. and Moonen, M. (2010) Distributed adaptive node-specific signal estimation in fully connected sensor networks – Part I: Sequential node updating. *IEEE Transactions on Signal Processing,* **58**(10), 5277–5291.

[32] Golub, G. H. and Van Loan, C. F. (2012) Matrix computations, JHU Press, 4th edition.