DISTRIBUTED DISTORTIONLESS SIGNAL ESTIMATION IN WIRELESS ACOUSTIC SENSOR NETWORKS

Alexander Bertrand[∗,†], *Joseph Szurley*[∗,†] *and Marc Moonen*[∗,†]

∗ KU Leuven, Dept. of Electrical Engineering-ESAT, SCD-SISTA \ † IBBT Future Health Department
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
E-mail: alexander.bertrand@esat.kuleuven.be
joseph.szurley@esat.kuleuven.be
marc.moonen@esat.kuleuven.be

ABSTRACT

Wireless microphone networks or so-called wireless acoustic sensor networks (WASNs) consist of physically distributed microphone nodes that exchange data over wireless links. In this paper, we propose a novel distributed distortionless signal estimation algorithm for noise reduction in WASNs. The most important feature of the proposed algorithm is that the nodes broadcast only single-channel signals while still obtaining optimal estimation performance, even in a scenario with multiple desired sources or speakers (in existing distributed methods, this is achieved only in scenarios with a single desired source). The idea is to create a one-dimensional desired signal subspace by using the same reference microphone at all the nodes. Since the theory is based on a distortionless signal estimation technique, namely linearly constrained minimum variance (LCMV) beamforming, we will show that this reference microphone does not need to be transmitted over the wireless link. We provide simulations to demonstrate the performance of the algorithm.

1. INTRODUCTION

Traditional microphone arrays often have strict space and power constraints, limiting the number of microphones and the physical size of the array, especially in miniature and portable devices (e.g., hearing aids or cell phones). Although such microphone arrays exploit spatial properties of the acoustic scenario, they only sample the sound field locally, i.e., in a small area. This limitation can be overcome by distributing many microphone nodes over a large area, where each node contains one or more microphones and facilities for wireless communication and signal processing. These nodes can then exchange microphone signals over wireless communication links with nearby nodes or a central processing unit. This yields a wireless microphone network, often referred to as a wireless acoustic sensor network (WASN), which is viewed as a next-generation technology for audio signal acquisition and audio signal processing [1]. However, since WASNs consist of physically distributed microphone nodes, they usually require dedicated audio processing algorithms, preferably allowing for distributed computing.

Microphone arrays are often used for multi-channel noise reduction or beamforming [2, 3]. In this paper, we focus on a noise reduction technique for WASNs, which is based on the so-called linearly constrained minimum variance (LCMV) beamformer. It provides a distortionless estimate of the desired signal components in an arbitrarily chosen reference microphone signal [3]. We envisage a distributed approach, i.e., the noise reduction needs to be computed in the network itself without gathering all the microphone signals in a central processing unit. For the sake of an easy exposition, we only address fully-connected networks where a signal broadcast by one node is received by all other nodes, but the results can be relatively easily modified for partially connected networks, based on similar techniques as in [4].

Distributed noise reduction in WASNs has been addressed in earlier work [5–8]. In particular, the so-called 'distributed adaptive node-specific signal estimation' (DANSE) algorithm [9] is able to achieve an optimal noise reduction in a distributed fashion (see, e.g., [5]). The same holds for the linearly constrained DANSE (LC-DANSE) algorithm, which is an LCMV-based modification of DANSE. An important feature of the (LC-)DANSE algorithm is that each node optimally estimates a desired signal component in its own reference microphone signal, rather than a joint network-wide signal, which explains the 'node-specific' aspect. However, it is shown that DANSE (and all of its extensions) can only achieve optimal noise reduction if the nodes transmit N-channel signals, where N is equal to the dimension of the signal subspace containing the desired signals of all the nodes. In a scenario with S desired sources, each node-specific reference microphone signal contains a different mixture of these source, hence $N = S$.

The idea is now to transform this S-dimensional desired signal subspace to a one-dimensional signal subspace by removing this node-specific aspect in DANSE. Indeed, if each node in DANSE would use the same reference microphone signal, then $N = 1$ (even if $S > 1$) and so single-channel broadcast signals are sufficient to achieve optimal noise reduction. However, this would require that each node is provided with this common reference microphone signal to locally compute its noise reduction filters, which significantly increases the communication bandwidth[1]. In this paper, we

Acknowledgements: The work of A. Bertrand was supported by a Postdoctoral Fellowship of the Research Foundation - Flanders (FWO). This work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE EF/05/006 'Optimization in Engineering' (OPTEC) and PFV/10/002 (OPTEC), Concerted Research Action GOA-MaNet, the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization', 2007-2011), Research Project IBBT, and Research Project FWO nr. G.0600.08 ('Signal processing and network design for wireless acoustic sensor networks'). The scientific responsibility is assumed by its authors.

[1]This is especially true in partially connected networks where it is not possible to directly broadcast a signal to all the nodes.

will indeed use a common reference microphone signal for all the nodes, but without explicitely broadcasting this reference signal. Instead, we use a distortionless noise reduction framework which allows each node to generate a virtual reference signal that has exactly the same desired signal component as the common reference microphone signal.

The new algorithm, referred to as 'single-reference distributed distortionless signal estimation' (1Ref-DDSE), has several interesting advantages compared to DANSE. As already mentioned, it obtains an optimal noise reduction at each node with only single-channel broadcast signals, even in e.g. speech scenarios with multiple desired speakers. Furthermore, one can choose a high-SNR microphone as the common reference microphone, which improves robustness against ripple of estimation errors. Indeed, DANSE has some robustness issues in the sense that errors in the estimation of signal correlation matrices at low-SNR nodes ripple through the network, significantly affecting the noise reduction performance at all other nodes[2] [5]. However, there are also some minor drawbacks in using 1Ref-DDSE instead of DANSE, i.e., the node-specific aspect is lost, and the constraints that impose the distortionless response remove some degrees of freedom that could have otherwise been used for extra noise reduction. Furthermore, it is based on LCMV beamforming, which requires robust subspace estimation methods if the source-microphone transfer functions are not known [3, 10].

2. DATA MODEL

Consider a WASN with a set of nodes $\mathcal{K} = \{1, \ldots, K\}$. Node k has access to M_k microphones, and the total number of microphones in the WASN is denoted by $M = \sum_{k \in \mathcal{K}} M_k$. Each microphone signal m of node k can be described in the frequency domain as

$$y_{km}(\omega) = x_{km}(\omega) + n_{km}(\omega), \;\; m = 1, ..., M_k \qquad (1)$$

where ω denotes the frequency-domain variable, $x_{km}(\omega)$ is the desired signal component (e.g. a speech signal or a mixture of multiple speech signals) and $n_{km}(\omega)$ is an undesired noise component. All subsequent algorithms will be implemented in the short-time Fourier transform (STFT) domain, where (1) is approximated based on finite-length time-to-frequency domain transformations. For conciseness, the frequency-domain variable ω will be omitted. All signals y_{km} of node k are stacked in an M_k-dimensional vector $\mathbf{y}_k$, and all vectors $\mathbf{y}_k$ are stacked in an M-dimensional vector $\mathbf{y}$. The vectors $\mathbf{x}_k$, $\mathbf{n}_k$ and $\mathbf{x}$, $\mathbf{n}$ are similarly constructed. The network-wide data model can then be written as $\mathbf{y} = \mathbf{x} + \mathbf{n}$.

The desired signal components $\mathbf{x}$ are modeled as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \qquad (2)$$

where $\mathbf{s}$ is an S-channel source signal, and $\mathbf{A}$ contains the transfer functions from each source to each microphone. The columns of $\mathbf{A}$ are referred to as the steering vectors, and the subspace spanned by these steering vectors is referred to as the steering subspace. We assume that $\mathbf{x}$, $\mathbf{A}$, and $\mathbf{s}$ are all unknown, i.e., we envisage a blind approach. Therefore, we will choose an arbitrary reference microphone, and try to estimate the desired signal component in this microphone signal[3], rather than the source signals in $\mathbf{s}$. This reference microphone is preferably a high-SNR microphone where all desired sources have a strong component. This is to avoid an ill-conditioned subspace estimation problem (see further). Without loss of generality (w.l.o.g.), we choose the first microphone of node 1 as the reference microphone, hence we estimate x_{11}.

[2]In the DANSE framework, this is avoided by using the so-called 'robust-DANSE' (R-DANSE) algorithm [5].

[3]It is noted that we do not aim to demix the S source signals in $\mathbf{x}$.

3. CENTRALIZED LCMV BEAMFORMING

We first assume that all microphone signals stacked in $\mathbf{y}$ are available in a central processing unit (we will later extend this to the distributed case). We will apply a multi-channel spatial filter or beamformer $\mathbf{w}$ to $\mathbf{y}$ such that the output signal $d = \mathbf{w}^H\mathbf{y}$ is a good estimate of x_{11} (superscript H denotes a conjugate transpose operator). We want to minimize the residual noise variance $\mathbf{w}^H\mathbf{n}$, while preserving an undistorted version of the desired speech component, i.e., $\mathbf{w}^H\mathbf{x} = x_{11}$. In [3], this is achieved by solving the following LCMV problem [11]

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{nn} \mathbf{w} \qquad (3)$$

$$\text{s.t. } \mathbf{Q}^H\mathbf{w} = \mathbf{Q}^H\mathbf{e}_1 \qquad (4)$$

where $\mathbf{R}_{nn} = E\{\mathbf{n}\mathbf{n}^H\}$ (with $E\{.\}$ the expected value operator), $\mathbf{e}_1 = [1 \; 0 \; \ldots \; 0]^T$ (selecting the column of $\mathbf{Q}^H$ corresponding to the reference microphone), and $\mathbf{Q}$ is an $M \times S$ matrix with columns spanning the steering subspace, i.e., $\mathbf{Q} = \mathbf{A}\mathbf{T}$ with $\mathbf{T}$ a non-singular $S \times S$ transformation matrix. The solution of (3)-(4) is given by

$$\hat{\mathbf{w}} = \mathbf{R}_{nn}^{-1}\mathbf{Q}\left(\mathbf{Q}^H\mathbf{R}_{nn}^{-1}\mathbf{Q}\right)^{-1}\mathbf{Q}^H\mathbf{e}_1 \qquad (5)$$

where $\mathbf{R}_{nn}$ can be estimated during noise-only frames, requiring a so-called voice-activity detection algorithm when applied in speech applications. In [3], it is proven that $\hat{\mathbf{w}}^H\mathbf{x} = x_{11}$, and hence the beamformer output yields an undistorted version of the desired signal in the reference microphone.

It is noted that, even though it may be hard to obtain the individual steering vectors in $\mathbf{A}$, it may be relatively easy to find an orthogonal basis for the steering subspace [3]. For example, the eigenvectors corresponding to the S non-zero eigenvalues of $\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^H\} = E\{\mathbf{y}\mathbf{y}^H\} - E\{\mathbf{n}\mathbf{n}^H\}$ indeed span the steering subspace defined by $\mathbf{A}$, and can therefore be used to construct $\mathbf{Q}$ in (5). More advanced subspace tracking algorithms can be found in [10, 12]. In the sequel, we make abstraction of this subspace tracking algorithm, i.e., we assume that for any set of input signals an orthogonal basis for the corresponding steering subspace can be computed, purely based on an analysis of these input signals.

4. DISTRIBUTED SIGNAL ESTIMATION WITH SINGLE-CHANNEL BROADCASTS

4.1 Problem statement and notation

In this section, we aim to compute (5) in a distributed fashion. In particular, we aim to have the LCMV output signal $\hat{d} = \hat{\mathbf{w}}^H\mathbf{y}$ available at each node in the network, without letting each node broadcast the full signal $\mathbf{y}_k, \forall\, k \in \mathcal{K}$. Instead, we only allow each node to broadcast a single-channel signal.

The single-channel signal that is broadcast by node k is defined as $z_k = \mathbf{r}_k^H \mathbf{y}_k$ where $\mathbf{r}_k$ is a (for the time being) undefined compression vector. All the z_k's are stacked in the K-channel signal $\mathbf{z}$ and we define $\mathbf{z}_{-k}$ as the vector $\mathbf{z}$ with z_k removed. Assuming full connectivity, node k has access to $\mathbf{y}_k$ and $\mathbf{z}_{-k}$, yielding an $(M_k + K - 1)$-channel input signal for node k (see Fig. 1):

$$\widetilde{\mathbf{y}}_k = \begin{bmatrix} \mathbf{y}_k \\ \mathbf{z}_{-k} \end{bmatrix} . \tag{6}$$

The $\widetilde{\mathbf{x}}_k$ and $\widetilde{\mathbf{n}}_k$ are constructed similarly, and the corresponding correlation matrices are denoted as $\widetilde{\mathbf{R}}_{xx,k}$ and $\widetilde{\mathbf{R}}_{nn,k}$ respectively. Furthermore, a basis for the corresponding (compressed) steering subspace is given by the columns of $\widetilde{\mathbf{Q}}_k = \widetilde{\mathbf{A}}_k \mathbf{T}_k$ with $\widetilde{\mathbf{A}}_k$ the compressed steering matrix such that $\widetilde{\mathbf{x}}_k = \widetilde{\mathbf{A}}_k \mathbf{s}$ and where $\mathbf{T}_k$ denotes a non-singular $S \times S$ matrix. The matrix $\widetilde{\mathbf{Q}}_k$ can be estimated, e.g., from the S dominant eigenvectors of $\widetilde{\mathbf{R}}_{xx,k}$, as explained earlier.

We define $d = \sum_{k \in \mathcal{K}} z_k$. If $\hat{\mathbf{w}}$ as defined in (5) would be known, then the signal d can be set equal to the network-wide LCMV output $\hat{d} = \hat{\mathbf{w}}^H \mathbf{y}$ by setting $\mathbf{r}_k = \hat{\mathbf{w}}_k$, where $\hat{\mathbf{w}}_k$ is the part of $\hat{\mathbf{w}}$ that is applied to $\mathbf{y}_k$. However, since none of the nodes have access to the full signal $\mathbf{y}$, the matrices $\mathbf{R}_{nn}$ and $\mathbf{Q}$ cannot be computed and hence (5) cannot be used to compute $\hat{\mathbf{w}}$. However the 1Ref-DDSE algorithm described in Subsection 4.3 will be able to iteratively find this solution, i.e., the $\mathbf{r}_k$'s at the different nodes are sequentially updated to converge towards their corresponding $\hat{\mathbf{w}}_k$'s. Therefore we will add an iteration index i as a superscript in the sequel, e.g., $z_k^i = \mathbf{r}_k^{i\,H} \mathbf{y}_k$, etc. It is important to remark that this iterative nature of our approach does not imply that previous samples of z_k^i are recompressed and retransmitted after each update of the $\mathbf{r}_k^i$ that generates this signal. This is similar to the output signal of adaptive (recursive) filters, i.e., previously filtered/compressed/transmitted data is not refiltered/recompressed/retransmitted when the filter is updated.

4.2 Relationship with distributed LCMV beamforming

Let us initialize all compression vectors $\mathbf{r}_k^0, \forall\, k \in \mathcal{K}$, with random entries, such that z_k^0 contains random linear combinations of the microphone signals in $\mathbf{y}_k$. Consider the following distributed algorithm that updates the $\mathbf{r}_k$'s, which we refer to as Algorithm A:

1. Initialize $q \leftarrow 1$ and $i \leftarrow 0$.
2. Node q observes the $(M_q + K - 1)$-channel input signal $\widetilde{\mathbf{y}}_q^i$ and computes the LCMV beamformer $\widetilde{\mathbf{w}}_q$ with respect to these inputs (similar to (5), but with $\widetilde{\mathbf{R}}_{nn,k}^i$ and $\widetilde{\mathbf{Q}}_k^i$, rather than $\mathbf{R}_{nn}$ and $\mathbf{Q}$). It chooses one of its microphone signals $y_q^{\text{ref}} \in \mathbf{y}_q$ as the reference microphone signal. We partition this local LCMV beamformer in two parts:

$$\widetilde{\mathbf{w}}_q = \begin{bmatrix} \mathbf{b}_q \\ \mathbf{g}_q \end{bmatrix} \tag{7}$$

   where $\mathbf{b}_q$ is the part that is applied to the signal $\mathbf{y}_q$, and $\mathbf{g}_q$ is the part that is applied to $\mathbf{z}_{-q}^i$.

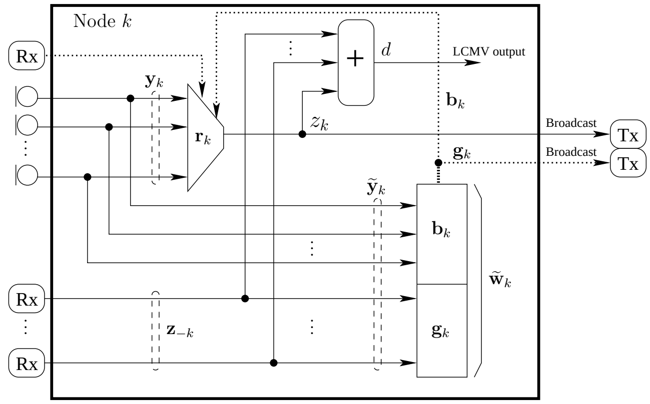

Figure 1: Illustration of the signal flow within node k in the 1Ref-DDSE algorithm. Full lines show audio signal flows, and dotted lines show the exchange of control parameters.

3. Update $\mathbf{r}_q^{i+1} = \mathbf{b}_q$ and $\mathbf{r}_k^{i+1} = g_q(k)\mathbf{r}_k^i, \forall\, k \in \mathcal{K}\backslash\{q\}$, where $g_q(k)$ denotes the entry of $\mathbf{g}_q$ corresponding to z_k^i. Notice that this update changes all broadcast signals from $\mathbf{z}^i$ to $\mathbf{z}^{i+1}$.
4. Update $i \leftarrow i + 1$ and $q \leftarrow (q \mod K) + 1$.
5. Go back to step 2.

In each iteration, one node q solves an LCMV beamforming problem based on its local inputs, and all $\mathbf{r}_k^i$'s, $\forall\, k \in \mathcal{K}$, are updated based on this LCMV solution (the $\mathbf{r}_k^i$'s at nodes $k \neq q$ are only scaled with some factor chosen by node q). To perform this scaling, node q must broadcast the vector $\mathbf{g}_q$ to the other nodes. However, this is merely a transmission of control parameters that happens every now and then, which is negligible compared to the continuous transmission of the signals in $\mathbf{z}^i$.

It is important to note that, after step 3, the speech component in the signal $d^{i+1} = \sum_{k \in \mathcal{K}} z_k^{i+1}$ will be equal to x_q^{ref}, i.e., the speech component in the reference microphone of node q. This is because $d^{i+1} = \widetilde{\mathbf{w}}_q^H \widetilde{\mathbf{y}}_q^i$, and $\widetilde{\mathbf{w}}_q$ is the LCMV beamformer based on reference microphone signal y_q^{ref}. Notice that in Algorithm A, each node uses its own local reference microphone to define its local LCMV problem, which will hinder convergence to a common solution. In the 1Ref-DDSE algorithm however, a single reference microphone for all the nodes is chosen, i.e., y_{11} (w.l.o.g.), and yet we wish to avoid that node 1 needs to broadcast this signal (in addition to z_1^i) to the other nodes. Before explaining how this can be achieved, we first state the following convergence theorem concerning algorithm A in a hypothetical scenario:

Theorem 4.1. *If the reference microphone signal of each node has the same desired signal component as the network-wide reference microphone signal y_{11}, i.e.,*

$$y_k^{ref} = x_{11} + n_k^{ref} , \ \forall\, k \in \mathcal{K} . \tag{8}$$

and if some technical conditions are satisfied (details omitted), then the $\mathbf{r}_k$'s of algorithm A converge to the corresponding $\hat{\mathbf{w}}_k$'s. This means that $\lim_{i \to \infty} d^i = \hat{d} = \hat{\mathbf{w}}^H \mathbf{y}$, hence the network-wide LCMV beamformer output $\hat{d}$ can be computed at each node.

The 'technical conditions' mentioned in the theorem are due to a similarity between algorithm A and the so-called distributed LCMV (D-LCMV) algorithm described in [13], which requires two sufficient conditions to guarantee convergence. Basically, they require that $\mathbf{R}_{nn}$ is full rank, and that $\mathcal{R}^{i\,H}\mathbf{Q}$ has at least S linearly independent rows for all i, where $\mathcal{R}^i$ denotes the block-diagonal compression matrix $\mathcal{R}^i = \mathit{Blockdiag}(\mathbf{r}_1^i, \ldots, \mathbf{r}_K^i)$ (such that $\mathbf{z}^i = \mathcal{R}^{i\,H}\mathbf{y}$). In practice, this second condition is usually satisfied if the number of nodes K is much larger than S (a safe choice is $K > 2S$ [13]).

Due to space constraints, we only give an outline of the proof of Theorem 4.1. Denote $\mathbf{r}^i$ as the stacked vector of all the $\mathbf{r}_k^i$'s. With this notation, it can be shown that Algorithm A is equivalent to computing $\mathbf{r}^{i+1}$ from $\mathbf{r}^i$ as the solution of

$$\mathbf{r}^{i+1} = \arg\min_{\mathbf{r}} \mathbf{r}^H \mathbf{R}_{nn} \mathbf{r} \tag{9}$$

$$\text{s.t.} \quad \mathbf{T}^i \mathbf{Q}^H \mathbf{r} = \mathbf{T}^i \mathbf{Q}^H \mathbf{e}_1 \tag{10}$$

$$\forall\, k \in \mathcal{K}\backslash\{q\},\ \exists\, \gamma_k \in \mathbb{C}:\ \mathbf{r}_k = \mathbf{r}_k^i \gamma_k\ . \tag{11}$$

where q is sequentially updated according to $q \leftarrow (q \mod K) + 1$. Here, $\mathbf{T}^i$ is a nonsingular $S \times S$ transformation matrix that models the fact that in Algorithm A, each node's local (compressed) subspace estimation corresponds to a different[4] basis for the (uncompressed) steering subspace. Note that, due to (8), the first row in the matrix $\mathbf{Q}$ and the rows corresponding to the different y_k^{ref}'s have the same entries, and therefore it is allowed to use the same selection vector $\mathbf{e}_1$ in every iteration.

Note that removing the $\mathbf{T}^i$'s in (10), $\forall\, i \in \mathbb{N}$, does not change the solution of this optimization problem. The resulting updating procedure without the $\mathbf{T}^i$'s has fixed linear constraints (independent of i), and is then equivalent to the so-called D-LCMV algorithm, for which convergence to the network-wide LCMV solution (under the above mentioned technical conditions[5]) is proven in [13].

4.3 The 1Ref-DDSE algorithm

We will now convert Algorithm A to a practical algorithm, such that (8) is not required, while still relying on the convergence result described in Theorem 4.1. Notice that the first updating node in Algorithm A is node 1. This means that, for $i = 1$, it holds that

$$d^i = \sum_{k\in\mathcal{K}} z_k^i = x_{11} + \sum_{k\in\mathcal{K}} \mathbf{r}_k^{i\,H} \mathbf{n}_k \tag{12}$$

i.e., the summation of the z_k^i signals yields a signal d^i that has exactly the same desired signal component as the reference microphone y_{11}. Furthermore, since each node has access to all the z_k^i's, each node can generate d^i. The main trick to derive the 1Ref-DDSE algorithm, is to use this signal d^i as the reference microphone signal in all nodes (except for node 1, where the actual reference microphone signal y_{11} is used, see also Remark I). Define the vector

$$\mathbf{v}_k^i = \begin{bmatrix} \mathbf{r}_k^i \\ \mathbf{1}_{K-1} \end{bmatrix} \tag{13}$$

[4]This is due to the fact that each node estimates this basis based on a differently compressed version of $\mathbf{R}_{xx}$.

[5]The D-LCMV algorithm can be modified to operate in simply connected networks, in which case the conditions for convergence are slightly different (see [13] for more details).

Table 1: Description of the 1Ref-DDSE algorithm.

1Ref-DDSE algorithm

1. Initialize $\mathbf{r}_k^0, \forall k \in \mathcal{K}$, with random non-zero entries and set $q \leftarrow 1$, $i \leftarrow 0$.
2. Node q observes the $(M_q + K - 1)$-channel input signal $\widetilde{\mathbf{y}}_q^i$ (yielding a new estimate of $\widetilde{\mathbf{R}}_{nn,q}^i$ and $\widetilde{\mathbf{Q}}_q^i$) and it computes the local LCMV beamformer $\widetilde{\mathbf{w}}_q$ according to (16). If $q = 1$, the same formula (16) is used, but $\mathbf{v}_k^i$ is replaced with $\mathbf{e}_1$. We define the partition $\widetilde{\mathbf{w}}_q = \begin{bmatrix}\mathbf{b}_q^T & \mathbf{g}_q^T\end{bmatrix}^T$, similar to (7).
3. Update $\mathbf{r}_q^{i+1} = \mathbf{b}_q$ and $\mathbf{r}_k^{i+1} = g_q(k)\mathbf{r}_k^i$, $\forall\, k \in \mathcal{K}\backslash\{q\}$, where $g_q(k)$ denotes the entry of $\mathbf{g}_q$ corresponding to z_k^i.
4. Update $i \leftarrow i + 1$ and $q \leftarrow (q \mod K) + 1$.
5. Go back to step 2.

where $\mathbf{1}_X$ is the X-dimensional vector containing 1 in each entry. Consider the following optimization problem corresponding to node k at iteration i

$$\min_{\mathbf{w}} \mathbf{w}^H \widetilde{\mathbf{R}}_{nn,k}^i \mathbf{w} \tag{14}$$

$$\text{s.t.}\ \widetilde{\mathbf{Q}}_k^{i\,H} \mathbf{w} = \widetilde{\mathbf{Q}}_k^{i\,H} \mathbf{v}_k^i \tag{15}$$

with solution

$$\widetilde{\mathbf{w}}_k = \left(\widetilde{\mathbf{R}}_{nn,k}^i\right)^{-1} \widetilde{\mathbf{Q}}_k^i \left(\widetilde{\mathbf{Q}}_k^{i\,H} \left(\widetilde{\mathbf{R}}_{nn,k}^i\right)^{-1} \widetilde{\mathbf{Q}}_k^i\right)^{-1} \widetilde{\mathbf{Q}}_k^{i\,H} \mathbf{v}_k^i\ . \tag{16}$$

Since the desired signal component in d^i is equal to $\mathbf{v}_k^{i\,H}\widetilde{\mathbf{x}}_k^i$, it can be seen from the righthand side of (15) that the signal d^i is actually chosen as a (virtual) reference microphone signal, rather than one of the actual input signals in $\widetilde{\mathbf{y}}_k^i$.

Let us now consider Algorithm A, where $\widetilde{\mathbf{w}}_k$ is computed as in (16) (except in node 1), which results in the 1Ref-DDSE algorithm, which is described in Table 1 and the corresponding signal flow at node k is schematically depicted in Fig. 1. It is important to note that, since each node (except for node 1) uses d^i as a (virtual) reference microphone signal to compute a distortionless LCMV beamformer, and since d^{i+1} is equal to the output of this beamformer, (12) will hold for any iteration $i \in \mathbb{N}$. Therefore, the desired signal component of the signal d^i will always be equal to x_{11}. Furthermore, this also means that condition (8) in Theorem 4.1 is now satisfied for each iteration and in every node, since d^i (or y_{11} in node 1) is used as a reference microphone. Therefore, convergence and optimality of the 1Ref-DDSE algorithm follows immediately from Theorem 4.1.

Remark I: As explained above, the desired signal component of the signal d^i will always be equal to x_{11}. In practice, however, estimation errors in $\widetilde{\mathbf{R}}_{nn,k}^i$ and $\widetilde{\mathbf{Q}}_k^i$ will add some distortion on this speech component, which ripples through all subsequent iterations. That is why it is important that node 1 uses its own reference microphone (rather than d^i) to stop this ripple, allowing the algorithm to correct itself.

Remark II: The entries of $\mathbf{g}_q$ in step 3 of the 1Ref-DDSE algorithm can also be used by all the nodes to scale the corresponding rows and columns in the estimates of their local covariance matrices (i.e., the $\widetilde{\mathbf{R}}^i_{nn,k}$'s, and possibly the $\widetilde{\mathbf{R}}^i_{xx,k}$'s). This may be useful since these covariance matrices need to be continuously tracked.

Remark III: It is noted that the 1Ref-DDSE algorithm has a complexity of $O\left((M_k + K - 1)^3\right)$ at node k. Therefore, if $M \gg K$, the power consumption in the 1Ref-DDSE algorithm is significantly smaller than in a centralized LCMV beamformer, which has complexity $O(M^3)$.

5. SIMULATION RESULTS

In this section, we provide simulation results for the 1Ref-DDSE algorithm in a scenario with two desired speakers and two babble noise sources, and some uncorrelated sensor noise. We have 5 nodes, each with 4 microphones, in a 5m by 5m reverberant room. Full details and an illustration of the simulated scenario are omitted here for brevity, but can be found in [7], which describes the same scenario[6]. We aim to estimate the mixture of the two desired speaker signals as they impinge on the reference microphone (at node 1). This reference microphone is in the middle of the room, hence none of the observed speech signals heavily dominates the other, which is important for the subspace estimation. The SNR at this reference microphone is -0.8dB. For the subspace estimation at node k, we used the locally observed clean speech correlation matrix (based on the speech components in $\widetilde{\mathbf{y}}_k$), hence isolating subspace estimation errors. We use an STFT with block size 1024.

The performance of the 1Ref-DDSE algorithm is shown in Fig. 2, and the performance of the corresponding centralized LCMV beamformer is also shown as a reference. The upper plot shows the output SNR as a function of the number of iterations. It is observed that the 1Ref-DDSE algorithm converges and achieves the same output SNR as the centralized approach. The middle plot shows the signal-to-distortion ratio (SDR) defined as

$$\text{SDR}^i = 10\log_{10}\frac{E\{x_{11}[t]^2\}}{E\{(x_{11}[t]-d^i_k[t])^2\}} \qquad (17)$$

where $x_{11}[t]$ and $d^i_k[t]$ are now defined in the time domain. In theory, the SDR should be infinitely large in each iteration because we envisage a distortionless estimate. This is of course not the case in practice due to estimation errors in the correlation matrices and due to finite length DFTs. However, a very high SDR is indeed immediately obtained in the first iteration. The SDR slightly drops each time a node $k \neq 1$ updates. This is because each iteration will introduce some small distortion on the desired signal component in the beamformer output d^i for the same reasons as mentioned earlier. Since the next update uses the previous d^i as a reference, there is a slight ripple of distortion errors over multiple iterations, until node 1 updates again (which does not use d^i as a reference). The third plot shows the mean squared error (MSE) between the centralized LCMV filters in $\hat{\mathbf{w}}$ and the corresponding filter entries in $\mathbf{r}^i = [\mathbf{r}^{i\,T}_1 \ \ldots \ \mathbf{r}^{i\,T}_K]^T$ obtained by the 1Ref-DDSE algorithm.

6. CONCLUSIONS

We have proposed a novel distributed noise reduction algorithm, referred to as the 1Ref-DDSE algorithm, for distortionless signal estimation in WASNs. Even though nodes broadcast only single-channel signals, it is proven that the 1Ref-DDSE algorithm obtains the optimal (centralized) performance as if all nodes have access to all microphone signals. This also holds in scenarios with multiple desired sources, which is not the case for other existing methods, where multi-channel broadcasts are required to obtain optimal performance in such scenarios. This is due to the fact that the 1Ref-DDSE algorithm is based on a single (virtual) reference microphone that is the same for all the nodes, reducing the desired signal subspace dimension to one. Simulation results have been provided to demonstrate the performance of the algorithm.

[6] Except for an extra node placed in the middle of the room.

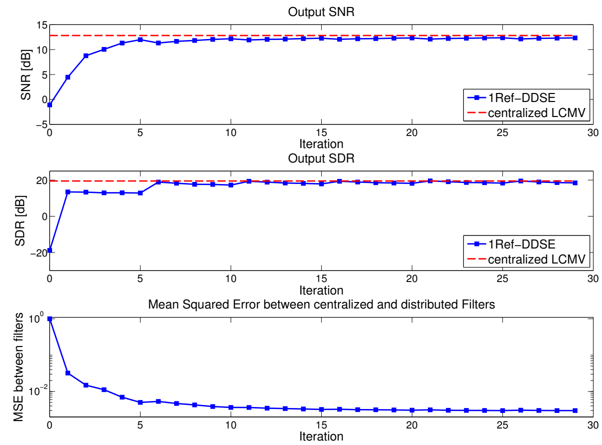

Figure 2: Performance of the 1Ref-DDSE algorithm, compared with the centralized LCMV beamformer.

REFERENCES

[1] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: a signal processing perspective," in *Proc. IEEE Symposium on Communications and Vehicular Technology (SCVT)*, (Ghent, Belgium), November 2011.

[2] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Berlin Heidelberg New York: Springer-Verlag.

[3] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, pp. 1071–1086, August 2009.

[4] A. Bertrand and M. Moonen, "Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology," *IEEE Trans. Signal Processing*, vol. 59, pp. 2196–2210, May 2011.

[5] A. Bertrand and M. Moonen, "Robust distributed noise reduction in hearing aids with external acoustic sensor nodes," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 530435, 14 pages, 2009.

[6] I. Himawan, I. McCowan, and S. Sridharan, "Clustered blind beamforming from ad-hoc microphone arrays," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, pp. 661 –676, may 2011.

[7] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, pp. 233–246, January 2012.

[8] Y. Jia, Y. Luo, Y. Lin, and I. Kozintsev, "Distributed microphone arrays for digital home and office," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1065–1068, May 2006.

[9] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part I: sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, pp. 5277–5291, 2010.

[10] S. Markovich Golan, S. Gannot, and I. Cohen, "Subspace tracking of multiple sources and its application to speakers extraction," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, (Dallas, Texas USA), pp. 201 –204, Mar. 2010.

[11] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, pp. 4 –24, apr. 1988.

[12] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, pp. 2230 – 2244, Sept. 2002.

[13] A. Bertrand and M. Moonen, "Distributed LCMV beamforming in a wireless sensor network with single-channel per-node signal transmission," *Internal Report KU Leuven, ESAT/SCD-SISTA (submitted for publication)*, 2012.