

# MULTI-TASK WIRELESS ACOUSTIC SENSOR NETWORK FOR NODE-SPECIFIC SPEECH ENHANCEMENT AND DOA ESTIMATION

*Amin Hassani, Jorge Plata-Chaves, Alexander Bertrand, Marc Moonen*

Department of Electrical Engineering-ESAT, STADIUS, KU Leuven, B-3001 Leuven, Belgium  
E-mails: {amin.hassani, jplata, alexander.bertrand, marc.moonen}@esat.kuleuven.be

## ABSTRACT

We consider the design of a distributed algorithm that is suitable for a wireless acoustic sensor network formed by nodes solving multiple tasks (MDMT). In the network, some of the nodes aim at estimating the node-specific direction-of-arrival of some desired sources. Additionally, there are other nodes that aim at implementing either a multi-channel Wiener filter or a minimum variance distortionless response beamformer in order to estimate node-specific desired signals as they impinge on their microphones. By using compressive filter-and-sum operations that incorporate a low-rank approximation of the sensor signal correlation matrix, the proposed MDMT algorithm let the nodes cooperate to achieve the network-wide centralized solution of their node-specific estimation problems without any knowledge about the tasks of other nodes. Finally, the effectiveness of the algorithm is shown through computer simulations.

**Index Terms**— Distributed node-specific signal estimation, subspace estimation, wireless acoustic sensor networks.

## 1. INTRODUCTION

Recently, there has been an increasing interest in distributed algorithms that can be implemented over so-called wireless acoustic sensor networks (WASNs). A WASN consists of a multitude of wireless nodes which are equipped with a microphone or an array of microphones. Traditionally, the design of distributed algorithms has focused on networks where the nodes observe the same phenomenon and/or are interested in the same network-wide signal processing task [1]- [3]. However, motivated by the heterogeneity of today's digital networks, recent advances in distributed adaptive signal processing and communication networking are currently enabling a novel paradigm where the networks are formed by Multiple Devices cooperating in Multiple Tasks (MDMT) [4], [5].

Unlike distributed parameter estimation algorithms for single-task networks, (e.g. [6]- [10]), under the MDMT paradigm the design of distributed parameter estimation algorithms assumes that the nodes are interested in estimating different but coupled parameters. Toward this goal, all the resulting distributed parameter estimation algorithms rely on novel node-specific implementations of a particular adaptive filtering technique such as least mean squares

(LMS), recursive least squares (RLS) or the affine projection algorithm (APA). For instance, in the context of cognitive radio networks and smart grids, there are node-specific incremental [11], [12] and diffusion [13], [14] algorithms that solve a distributed parameter estimation problem where the nodes are interested in different but overlapping vectors of parameters. Similarly, several diffusion-based algorithms were also derived to facilitate the cooperation among subsets of nodes with similar estimation interests [15]- [18]. The aforementioned works are all focused on regression-based parameter estimation problems. Other works have focused on other parameter estimation problems such as node-specific direction of arrival (DOA) estimation [19]- [20]. In this setting, each node has a different orientation and hence a different DOA with respect to the target source [19].

Another class of distributed MDMT algorithms focuses on node-specific signal estimation (NSSE) problems, which rely on a network-wide spatial filtering. Most of these algorithms rely on compressive filter-and-sum operations on the sensor signals in order to let the nodes solve their NSSE problems with the same performance as a central unit that collects and processes all the sensor signals. These compressive filter-and-sum operations were used to solve NSSE problems under different beamformer criteria where the nodes are interested in estimating samples of desired signals that share a common latent signal subspace, e.g., for speech enhancement in WASNs. Based on the multi-channel Wiener filter (MWF), a distributed algorithm was derived to obtain the centralized linear minimum mean square error estimates of the node-specific desired signals in binaural hearing aids [21] or in wireless networks with a fully-connected topology [22], a tree topology [23] and combinations thereof [24]. To run all these algorithms over networks that operate under non-stationary and low-SNR conditions, in [25] the estimation of each node-specific desired signal is performed by a MWF in which a low-rank approximation based on a generalized eigenvalue decomposition (GEVD) is incorporated. Moreover, the authors in [26] derived a distributed algorithm under which the estimation of the node-specific signals is undertaken through two different but coupled blind minimum variance distortionless response (MVDR) beamformers. Unlike the MWF-based node-specific signal estimation algorithm, under the blind MVDR criterion the goal of each node is to minimize the output power of a filter subject to linear constraints to avoid distortion of a desired (e.g., speech) signal.

All the aforementioned distributed node-specific estimation algorithms consider a setting where all nodes cooperate to obtain different but coupled solutions of the *same* signal processing (SP) task (e.g., signal enhancement, spectrum estimation, DOA estimation etc.). Moreover, when obtaining these node-specific estimates, all the existing works assume that all nodes apply the *same* SP technique, e.g., a particular adaptive filter (e.g., LMS, RLS or APA), beamformer (e.g., MWF or MVDR) or subspace-based DOA esti-

---

This work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven Research Council CoE PFV/10/002 (OPTEC) and BOF/STG-14-005, the Interuniversity Attractive Poles Programme initiated by the Belgian Science Policy Office IUAP P7/23 'Belgian network on stochastic modeling analysis design and optimization of communication systems' (BESTCOM) 2012-2017, Research Project 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 EU/FP7 project 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.

mation method. However, in heterogeneous multi-task networks, the nodes may be interested in solving *different* but interrelated SP tasks. Furthermore, each node may apply *different* SP techniques (filters or beamformers) in order to fulfill the particular performance requirement of its application layer. Motivated by these facts, we first define in this paper an instance of an MDMT network where some nodes aim at obtaining estimates of the node-specific DOAs of some desired sources, while others are interested in enhancing node-specific desired sources by using different beamformer criteria, i.e., a GEVD-based MWF or blind MVDR. Next, after describing the corresponding centralized problems of the different SP tasks, we rely on compressive linear estimation techniques to design a distributed MDMT-based algorithm that lets the nodes cooperate while solving their node-specific SP tasks as if they had access to all the sensor signals available in the network. In this particular setting, the nodes do not even have to be aware of the tasks that other nodes are solving, i.e., MWF, MVDR or DOA estimation. Finally, we provide simulation results for such a multi-task WASN to illustrate its effectiveness.

## 2. DATA MODEL AND PROBLEM STATEMENT

We consider a fully-connected WASN with  $K$  nodes observing a single target speech source<sup>1</sup>. Each node  $k \in \mathcal{K} = \{1, \dots, K\}$  is equipped with a microphone array consisting of  $M_k$  microphones, where its  $M_k$ -channel microphone signal is denoted as  $\mathbf{y}_k$ . We consider frequency domain processing, where the microphone signal  $\mathbf{y}_k$  can be modeled as ( $\omega$  is the frequency variable)

$$\mathbf{y}_k(\omega) = \mathbf{a}_k(\theta_k, \omega)s(\omega) + \mathbf{n}_k(\omega) \quad (1)$$

where  $s$  is the signal of the target speech source,  $\mathbf{a}_k$  is a node-specific  $M_k$ -dimensional steering vector (acoustic transfer function from the source to the microphones of node  $k$ ),  $\theta_k$  is the DOA at node  $k$ , and  $\mathbf{n}_k$  is additive noise which includes both spatially correlated (e.g., due to localized noise sources) and uncorrelated (e.g., microphone self-noise) noise contributions. We assume that each node's microphone array has a different orientation, and hence the DOA  $\theta_k$  is different for each node. In the sequel, we always omit  $\omega$  for the sake of brevity. We also define  $\mathbf{s}_k \triangleq \mathbf{a}_k s$  as the  $M_k$ -channel speech component of the noisy microphone signals at node  $k$ . By stacking all  $\mathbf{y}_k$ ,  $\mathbf{n}_k$  and  $\mathbf{s}_k$ , we obtain the network-wide  $M$ -channel signals  $\mathbf{y}$ ,  $\mathbf{s}$  and  $\mathbf{n}$ , respectively, where  $M = \sum_{k=1}^K M_k$ . With this we can write  $\mathbf{y} = \mathbf{s} + \mathbf{n} = \mathbf{a}\mathbf{s} + \mathbf{n}$ , where  $\mathbf{a}$  denotes the network-wide  $M$ -dimensional steering vector.

Each node is then tasked to attain a node-specific goal from the following three cases, i.e.,  $\mathcal{K} \triangleq \{\mathcal{K}^{\text{MWF}} \cup \mathcal{K}^{\text{MVDR}} \cup \mathcal{K}^{\text{DOA}}\}$ . Each node  $k \in \mathcal{K}^{\text{MWF}}$  estimates its node-specific desired speech signal  $d_k$  from the available noisy microphone signals using MWF [27]. Each node  $k \in \mathcal{K}^{\text{MVDR}}$  minimizes the output power of its beamformer under a single linear constraint that steers the beam towards the location of the target speech source such that the target speech signal (as received at its *reference* microphone) is processed without distortion [26]. Finally, each node  $k \in \mathcal{K}^{\text{DOA}}$  estimates its node-specific DOA  $\theta_k$  from the target speech source [19]. We assume that the local microphone array geometry of the nodes  $k \in \mathcal{K}^{\text{DOA}}$  is known, but the position of these nodes as well as the relative geometry between them and the other nodes are unknown.

## 3. CENTRALIZED ESTIMATION

We first consider the centralized estimation problems where we assume that each node  $k$  transmits its unprocessed  $M_k$ -channel micro-

<sup>1</sup>These assumptions are mainly for the sake of easy exposition, as all results can be extended to networks with nearest-neighbor topology or scenarios where multiple target speech sources are present [22]-[24].

phone signal  $\mathbf{y}_k$  to all other nodes. Therefore the objective for node  $k$  is to carry out its node-specific task based on the network-wide  $M$ -channel microphone signal  $\mathbf{y}$ .

### 3.1. MWF

Under the MWF criterion, the goal for each node  $k \in \mathcal{K}^{\text{MWF}}$  is to process the network-wide noisy microphone signal  $\mathbf{y}$  in order to obtain the linear minimum mean square error (LMMSE) estimate of its node-specific desired signal  $d_k = a_{k,r}s$ , where  $a_{k,r}$  is the  $r$ -th element of  $\mathbf{a}_k$  corresponding to the reference microphone. Toward this goal, each node  $k \in \mathcal{K}^{\text{MWF}}$  then applies an  $M$ -dimensional linear estimator  $\hat{\mathbf{w}}_k$  to estimate  $d_k$  as  $\hat{d}_k = \hat{\mathbf{w}}_k^H \mathbf{y}$ , with

$$\hat{\mathbf{w}}_k = \min_{\mathbf{w}_k} E\{|d_k - \mathbf{w}_k^H \mathbf{y}|^2\} \quad (2)$$

where the hat ( $\hat{\cdot}$ ) refers to the fact that the centralized estimation is considered and where  $E\{\cdot\}$  and the superscript  $H$  denote the expected value operator and the conjugate transpose operator, respectively. Assuming that  $\mathbf{R}_{yy} = E\{\mathbf{y}\mathbf{y}^H\}$  is a full rank matrix, the unique solution of (2) is [27]

$$\hat{\mathbf{w}}_k = \mathbf{R}_{yy}^{-1} \mathbf{R}_{ss} \mathbf{e}_k \quad \text{with} \quad \mathbf{R}_{ss} \triangleq E\{\mathbf{s}\mathbf{s}^H\} = P_s \mathbf{a}\mathbf{a}^H \quad (3)$$

where  $\mathbf{e}_k$  selects the column of  $\mathbf{R}_{ss}$  corresponding to the reference microphone of node  $k$ , and where  $P_s = E\{|s|^2\}$  is the power of the target speech source signal  $s$ .

Note that  $\mathbf{R}_{yy}$  can be estimated using sample averaging during 'speech-and-noise' segments. In addition, the noise correlation matrix, defined as  $\mathbf{R}_{nn} = E\{\mathbf{n}\mathbf{n}^H\}$ , is assumed to be either known or to be estimated in the 'noise-only' segments when the target speech source is silent. To distinguish between such segments, a Voice Activity Detection (VAD) is required (as explained in [19, 27]). In the sequel, we use an overline to denote a correlation matrix that is estimated from the data, i.e.,  $\bar{\mathbf{R}}$ .

When  $\mathbf{s}$  and  $\mathbf{n}$  are uncorrelated, we have  $\mathbf{R}_{ss} = \mathbf{R}_{yy} - \mathbf{R}_{nn}$ , where  $\mathbf{R}_{ss}$  is a rank-1 matrix (see (3)). In practice, however, the rank of  $\bar{\mathbf{R}}_{ss} = \bar{\mathbf{R}}_{yy} - \bar{\mathbf{R}}_{nn}$  is often greater than one, which is due to the finite window size in the short-time Fourier transform (STFT) analysis and/or nonstationarity of the noise. Moreover, in low-SNR conditions,  $\bar{\mathbf{R}}_{ss}$  may even lose its positive semi-definiteness, leading to suboptimal or even unstable filters [28]. A GEVD-based rank-1 approximation of  $\bar{\mathbf{R}}_{ss}$  can be alternatively incorporated in the MWF solution (3) to increase the estimation performance in such cases (more discussion in [25], [28]).

A GEVD of the matrix pair  $(\bar{\mathbf{R}}_{yy}, \bar{\mathbf{R}}_{nn})$  is defined as [29]

$$\bar{\mathbf{R}}_{yy} \hat{\mathbf{X}} = \bar{\mathbf{R}}_{nn} \hat{\mathbf{X}} \hat{\Lambda} \quad \text{s.t.} \quad \hat{\mathbf{X}}^H \bar{\mathbf{R}}_{nn} \hat{\mathbf{X}} = \mathbf{I}_M \quad (4)$$

where  $\hat{\Lambda}$  and  $\hat{\mathbf{X}}$  contain the Generalized EigenVaLues (GEVLs) and their corresponding Generalized EigenVeCtors (GEVCs), respectively. It is assumed w.l.o.g. that the GEVLs are sorted in descending order and that the GEVCs are scaled such that their  $\bar{\mathbf{R}}_{nn}$ -weighted norm is 1 (as expressed in (4)). Assuming that  $\bar{\mathbf{R}}_{nn}$  is invertible, the GEVD problem (4) is equivalent to a joint diagonalization of  $\bar{\mathbf{R}}_{yy}$  and  $\bar{\mathbf{R}}_{nn}$ , i.e., it can be verified from (4) that

$$\bar{\mathbf{R}}_{yy} = \hat{\mathbf{Q}} \hat{\Lambda} \hat{\mathbf{Q}}^H, \quad \bar{\mathbf{R}}_{nn} = \hat{\mathbf{Q}} \hat{\mathbf{Q}}^H \quad (5)$$

where  $\hat{\mathbf{Q}} = \hat{\mathbf{X}}^{-H}$ , with  $\hat{\mathbf{Q}}$  a full-rank  $M \times M$  matrix (not necessarily orthogonal). We can then write  $\bar{\mathbf{R}}_{ss} = \bar{\mathbf{R}}_{yy} - \bar{\mathbf{R}}_{nn} = \hat{\mathbf{Q}}(\hat{\Lambda} - \mathbf{I}_M)\hat{\mathbf{Q}}^H$ . Comparing this with (3), the GEVD-based rank-1 approximation of  $\bar{\mathbf{R}}_{ss}$  can be computed as

$$\bar{\mathbf{R}}_{ss} = (\hat{\lambda} - 1)\hat{\mathbf{q}}\hat{\mathbf{q}}^H \quad (6)$$

where  $\hat{\lambda}$  is the first (i.e., largest) GEVL of  $(\bar{\mathbf{R}}_{yy}, \bar{\mathbf{R}}_{nn})$ , and where  $\hat{\mathbf{q}}$  denotes the first column of  $\hat{\mathbf{Q}}$ . By plugging (6) into (3), the GEVD-based estimate of the node-specific desired signal at node  $k \in \mathcal{K}^{\text{MWF}}$  is  $\hat{d}_k = \hat{\mathbf{w}}_k^H \mathbf{y}$  with

$$\hat{\mathbf{w}}_k = \bar{\mathbf{R}}_{yy}^{-1} \hat{\mathbf{q}} (\hat{\lambda} - 1) \hat{q}_{k,r}^* \quad (7)$$

where  $\hat{q}_{k,r}$  denotes the entry of  $\hat{\mathbf{q}}$  corresponding to the  $r$ -th microphone of node  $k$ , i.e.,  $\hat{q}_{k,r}^* = \hat{\mathbf{q}}^H \mathbf{e}_k$ , where superscript  $*$  is the complex conjugate operator. Note that by plugging (5) in (7), it can be easily shown that  $\hat{\mathbf{w}}_k$  in (7) is equal to  $\hat{\mathbf{x}}$  up to a scaling, where  $\hat{\mathbf{x}}$  is the first column of  $\hat{\mathbf{X}}$ . The node-specific desired signal at node  $k \in \mathcal{K}^{\text{MWF}}$  is then estimated as  $\hat{d}_k = \hat{\mathbf{w}}_k^H \mathbf{y}$ .

### 3.2. MVDR

The objective for each node  $k \in \mathcal{K}^{\text{MVDR}}$  is to design an  $M$ -dimensional beamformer  $\hat{\mathbf{w}}_k$  that minimizes the output noise power, subject to a unity gain constraint in the target speech source direction, i.e.,

$$\hat{\mathbf{w}}_k = \min_{\mathbf{w}_k} E\{|\mathbf{w}_k^H \mathbf{n}|^2\} \quad \text{s.t.} \quad \mathbf{a}^H \mathbf{w}_k = 1. \quad (8)$$

In practice, due to the fact that often the network-wide steering vector  $\mathbf{a}$  is either unknown or difficult to estimate, the constraint is replaced by  $\hat{\mathbf{q}}^H \mathbf{w}_k = \hat{q}_{k,r}^*$  [30]. This replacement is essentially motivated by the fact that  $\hat{\mathbf{q}}$  is an estimate for the steering vector  $\mathbf{a}$  (up to an unknown scaling), as can be verified by comparing (6) with (3). With the latter constraint, the optimization problem (8) alternatively preserves the target speech signal as it impinges on the reference microphone of node  $k$ . The closed-form solution of (8) is then given by [30]

$$\hat{\mathbf{w}}_k = \frac{\mathbf{R}_{nn}^{-1} \hat{\mathbf{q}}}{\hat{\mathbf{q}}^H \mathbf{R}_{nn}^{-1} \hat{\mathbf{q}}} \hat{q}_{k,r}^* \quad (9)$$

in which by plugging (5), it can be easily shown that  $\hat{\mathbf{w}}_k$  is equal to  $\hat{\mathbf{x}}$  up to a scaling. The node-specific desired signal at node  $k \in \mathcal{K}^{\text{MVDR}}$  is then estimated as  $\hat{d}_k = \hat{\mathbf{w}}_k^H \mathbf{y}$ .

### 3.3. DOA estimation

The objective for each node  $k \in \mathcal{K}^{\text{DOA}}$  is to estimate its node-specific DOA  $\theta_k$  from the network-wide  $M$ -channel microphone signal  $\mathbf{y}$ . To achieve this, an estimate of the local steering vector (up to a complex scalar) can be fed into a subspace-based DOA estimation method, such as MUSIC [31], or ESPRIT [32]. We first partition  $\hat{\mathbf{q}}$  as  $\hat{\mathbf{q}} = [\hat{\mathbf{q}}_1^T \dots \hat{\mathbf{q}}_K^T]^T$ , where  $\hat{\mathbf{q}}_k^T$  contains the  $M_k$  entries of  $\hat{\mathbf{q}}$  corresponding to the local array of node  $k$ . Note that the node-specific DOA estimation at each node  $k \in \mathcal{K}^{\text{DOA}}$  is then carried out only using  $\hat{\mathbf{q}}_k$ , which is due to the fact that the relative geometry between the nodes is unknown. Although this means that we only partially exploit the information, cooperation between nodes has led to an improved GEVD-based estimate of  $\hat{\mathbf{q}}$  (and hence its sub-vectors  $\hat{\mathbf{q}}_k$ 's) [19]. The obtained DOA estimate is denoted as  $\hat{\theta}_k$ .

## 4. DISTRIBUTED MDMT-BASED ALGORITHM

In the proposed distributed algorithm, each node  $k \in \mathcal{K}$  first fuses its  $M_k$ -channel microphone signal  $\mathbf{y}_k$  into a single-channel signal  $z_k = \mathbf{f}_k^H \mathbf{y}_k$  with an  $M_k$ -dimensional linear compressor  $\mathbf{f}_k$  (which will be defined later, see (14)), and then broadcasts  $z_k$  to all other nodes. As a result, the required per-node communication bandwidth is reduced by a factor of  $M_k$ , compared to the centralized approach.

Considering a  $K$ -channel signal  $\mathbf{z} = [z_1 \dots z_K]^T$ ,  $\mathbf{z}_{-k}$  denotes the vector  $\mathbf{z}$  with  $z_k$  excluded. Assuming a fully-connected WASN, each node  $k$  then has access to a  $P_k$ -channel signal  $\tilde{\mathbf{y}}_k$  which is defined as  $\tilde{\mathbf{y}}_k = [\mathbf{y}_k^T \mathbf{z}_{-k}^T]^T$ , with  $P_k = M_k + (K-1)$ . In the sequel, we use the  $\tilde{\cdot}$  notation for quantities that are computed based on the

extended signal  $\tilde{\mathbf{y}}_k = \tilde{\mathbf{s}}_k + \tilde{\mathbf{n}}_k$ . Moreover, the corresponding  $P_k$ -dimensional correlation matrix estimates at each node  $k$  are denoted as  $\tilde{\mathbf{R}}_{\tilde{\mathbf{y}}_k \tilde{\mathbf{y}}_k}$ ,  $\tilde{\mathbf{R}}_{\tilde{\mathbf{s}}_k \tilde{\mathbf{s}}_k}$  and  $\tilde{\mathbf{R}}_{\tilde{\mathbf{n}}_k \tilde{\mathbf{n}}_k}$ .

At iteration  $i$ , node  $q$  is the only updating node, which uses a block of  $L$  samples to locally estimate the required correlation matrices. In the next iteration the updating node  $q$  is changed, and a new block of  $L$  samples (over a different time window) is used, which means that the iterations are spread out over time in a block-wise fashion. Similar to (4)-(5) node  $q$  then computes a local GEVD on the matrix pencil  $(\tilde{\mathbf{R}}_{\tilde{\mathbf{y}}_k \tilde{\mathbf{y}}_k}^i, \tilde{\mathbf{R}}_{\tilde{\mathbf{n}}_k \tilde{\mathbf{n}}_k}^i)$ , leading to  $P_q$ -dimensional matrices containing the local GEVCs and GEVLs denoted as  $\tilde{\mathbf{X}}_q^i$ ,  $\tilde{\mathbf{\Lambda}}_q$  (with ordered as GEVLs in (4)) and  $\tilde{\mathbf{Q}}_q^i$ , respectively, where  $\tilde{\mathbf{Q}}_q^i = (\tilde{\mathbf{X}}_q^i)^{-H}$ . The iteration index  $i$  will be dropped in the sequel for conciseness. We also define  $\tilde{\mathbf{q}}_q$  as the first column of  $\tilde{\mathbf{Q}}_q$ . Regarding the update procedure at the updating node  $q$ , we first consider the following three single-SP task cases:

- If all nodes were MWF nodes, i.e., if  $\mathcal{K} = \mathcal{K}^{\text{MWF}}$ , one could run the GEVD-based distributed adaptive node-specific signal estimation (DANSE) algorithm [25], in which all nodes sequentially perform the following operations (compare to (2),(7)).

$$\tilde{\mathbf{w}}_q = \min_{\tilde{\mathbf{w}}_q} E\{|d_q - \tilde{\mathbf{w}}_q^H \tilde{\mathbf{y}}_q|^2\} \quad (10)$$

$$\tilde{\mathbf{w}}_q = \tilde{\mathbf{R}}_{\tilde{\mathbf{y}}_q \tilde{\mathbf{y}}_q}^{-1} \tilde{\mathbf{q}}_q (\tilde{\lambda}_q - 1) \tilde{q}_{q,r}^* \quad (11)$$

with  $\tilde{q}_{q,r}$  denoting the  $r$ -th entry of  $\tilde{\mathbf{q}}_q$  and  $\tilde{\lambda}_q$  equal to the largest GEVL in  $\tilde{\mathbf{\Lambda}}_q$ .

- If all nodes were MVDR nodes, i.e., if  $\mathcal{K} = \mathcal{K}^{\text{MVDR}}$ , one could run the linearly constrained (LC-) DANSE algorithm [33], in which all nodes sequentially perform

$$\tilde{\mathbf{w}}_q = \min_{\tilde{\mathbf{w}}_q} E\{|\tilde{\mathbf{w}}_q^H \tilde{\mathbf{n}}_q|^2\} \quad \text{s.t.} \quad \tilde{\mathbf{q}}_q^H \tilde{\mathbf{w}}_q = \tilde{q}_{q,r}^* \quad (12)$$

$$\tilde{\mathbf{w}}_q = \frac{\tilde{\mathbf{R}}_{\tilde{\mathbf{n}}_q \tilde{\mathbf{n}}_q}^{-1} \tilde{\mathbf{q}}_q}{\tilde{\mathbf{q}}_q^H \tilde{\mathbf{R}}_{\tilde{\mathbf{n}}_q \tilde{\mathbf{n}}_q}^{-1} \tilde{\mathbf{q}}_q} \tilde{q}_{q,r}^* \quad (13)$$

- If all nodes were DOA nodes, i.e., if  $\mathcal{K} = \mathcal{K}^{\text{DOA}}$ , one could run the algorithm in [20] to estimate the first  $M_q$  entries of  $\tilde{\mathbf{q}}_q$  corresponding to the local array at node  $q$ , defined as  $\mathbf{q}_q$ . As an estimate of the steering vector  $\mathbf{a}_k$  (up to an unknown scaling),  $\mathbf{q}_q$  is then fed into a subspace-based DOA estimation algorithm such as MUSIC or ESPRIT. The resulting DOA estimate is denoted as  $\tilde{\theta}_q$ . Note that in this case node  $q$  updates an auxiliary vector  $\tilde{\mathbf{w}}_q$  via  $\tilde{\mathbf{w}}_q = \tilde{\mathbf{x}}_q$ , where  $\tilde{\mathbf{x}}_q$  is the first column of  $\tilde{\mathbf{X}}_q$ , i.e., it is the largest principal GEVC of  $(\tilde{\mathbf{R}}_{\tilde{\mathbf{y}}_k \tilde{\mathbf{y}}_k}^i, \tilde{\mathbf{R}}_{\tilde{\mathbf{n}}_k \tilde{\mathbf{n}}_k}^i)$ . This auxiliary parameter will define the linear compressor  $\mathbf{f}_q$  (see (14)). The reason behind this choice is explained in [20], where it is shown that using (part of) the local principal GEVC as a compressor, eventually results in a local steering vector estimate  $\mathbf{q}_q$  which is a subset of the network-wide steering vector  $\hat{\mathbf{q}}$  (details omitted).

In all of the three aforementioned algorithms (i.e., [25], [33], [20]), the fusion rule is updated in a similar fashion, i.e., the updating node  $q$  updates  $\mathbf{f}_q$  by replacing it with the first  $M_q$  rows of  $\tilde{\mathbf{w}}_q$ , i.e.,

$$\mathbf{f}_q = [\mathbf{I}_{M_q} \mathbf{0}] \tilde{\mathbf{w}}_q \quad (14)$$

where  $\mathbf{I}_{M_q}$  is the  $M_q$ -dimensional identity matrix and  $\mathbf{0}$  is an all-zero matrix with proper dimension. Nevertheless, note that the fusion update in (14) is different for each of these three single-SP cases, as the respective  $\tilde{\mathbf{w}}_q$ 's are different. Using these ingredients, we now define the distributed MDMT algorithm case where we let the nodes

**Table 1.** Distributed MDMT-based algorithm

1. Set  $i \leftarrow 0$ ,  $q \leftarrow 1$ , and initialize all  $\mathbf{f}_k^0$  and  $\tilde{\mathbf{w}}_k^0$ ,  $\forall k \in \mathcal{K}$ , with random entries.
2. Each node  $k \in \mathcal{K}$  broadcasts  $L$  new observations of its single-channel compressed signal  $z_k^i$ :

$$z_k^i[iL + j] = \mathbf{f}_k^i H \mathbf{y}_k^i[iL + j], \quad j = 1 \dots L \quad (15)$$

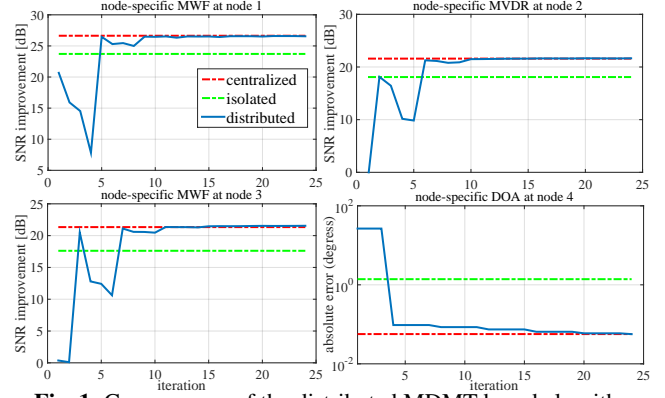
where the notation  $[\cdot]$  denotes a time (STFT-frame) index.

3. **Updating node  $q$ :** first update  $\tilde{\mathbf{R}}_{\tilde{\mathbf{y}}_q}^i$  and  $\tilde{\mathbf{R}}_{\tilde{\mathbf{n}}_q}^i$  via sample averaging and then computes  $\tilde{\mathbf{X}}_q^i$  and  $\tilde{\mathbf{A}}_q^i$  from the GEVD of  $(\tilde{\mathbf{R}}_{\tilde{\mathbf{y}}_q}^i, \tilde{\mathbf{R}}_{\tilde{\mathbf{n}}_q}^i)$  from which  $\tilde{\mathbf{q}}_q^i$  is estimated.
  - if  $q \in \mathcal{K}^{\text{MWF}}$ : compute the node-specific MWF  $\tilde{\mathbf{w}}_q^{i+1}$  as in (11).
  - if  $q \in \mathcal{K}^{\text{MVDR}}$ : compute the node-specific MVDR  $\tilde{\mathbf{w}}_q^{i+1}$  as in (13).
  - if  $q \in \mathcal{K}^{\text{DOA}}$ : use  $\tilde{\mathbf{q}}_q^i$  and estimate the node-specific DOA  $\tilde{\theta}_q$ , e.g., via ESPRIT or MUSIC and update  $\tilde{\mathbf{w}}_q^{i+1} = \tilde{\mathbf{x}}_q^{i+1}$ .
4. **Updating node  $q$ :** updates  $\mathbf{f}_q^{i+1} = [\mathbf{I}_{M_q} \mathbf{0}] \tilde{\mathbf{w}}_q^{i+1}$ .
5. The other nodes  $k \in \mathcal{K} \setminus q$  update their parameters as  $\tilde{\mathbf{w}}_k^{i+1} = \tilde{\mathbf{w}}_k^i$  and  $\mathbf{f}_k^{i+1} = \mathbf{f}_k^i$ .
6. Each node  $k \in \{\mathcal{K}^{\text{MWF}} \cup \mathcal{K}^{\text{MVDR}}\}$  estimates the next  $L$  samples of its single-channel signal  $d_k$ , as  $\tilde{\mathbf{d}}_k = (\tilde{\mathbf{w}}_k^{i+1})^H \tilde{\mathbf{y}}_k$ . Each node  $k \in \mathcal{K}^{\text{DOA}} \setminus q$  keeps its latest node-specific DOA  $\tilde{\theta}_k$ .
7.  $i \leftarrow i + 1$  and  $q \leftarrow (q \bmod K) + 1$  and return to step 2.

use the fused signals of the other nodes, independent of how they locally update their  $\tilde{\mathbf{w}}_q$  with the fusion rule  $\mathbf{f}_q$  based on GEVD-based DANSE, LC-DANSE or DOA estimation. The resulting distributed MDMT-based algorithm is described in Table 1. Note that nodes  $k \in \{\mathcal{K}^{\text{MWF}} \cup \mathcal{K}^{\text{MVDR}}\}$  estimate their node-specific desired signal as  $\tilde{\mathbf{d}}_k = \tilde{\mathbf{w}}_k^H \tilde{\mathbf{y}}_k$  in each iteration. It is noted that the nodes are not aware of each others tasks, and hence perform the same operations as they would perform in a hypothetical homogeneous network where all the (other) nodes perform the same network-wide distributed algorithm (DANSE, LC-DANSE or DOA estimation). Remarkably, despite the fact that each node solves a different local task, it can be shown that all their local estimates converge to the corresponding centralized solution as if all nodes would have access to the microphone signals from all other nodes.

**Theorem I:** *If  $\tilde{\mathbf{R}}_{\tilde{\mathbf{n}}_n}$  is full rank, then the estimates obtained from the proposed distributed MDMT-based algorithm converge for any initialization of the fusion rules  $\mathbf{f}_q$  to the corresponding estimates obtained from the centralized solutions, i.e., when  $i \rightarrow \infty$ ,  $\forall k \in \{\mathcal{K}^{\text{MWF}} \cup \mathcal{K}^{\text{MVDR}}\}$ ,  $\tilde{\mathbf{d}}_k = \mathbf{d}_k$ , and  $\forall k \in \{\mathcal{K}^{\text{DOA}}\}$ ,  $\tilde{\theta}_k = \hat{\theta}_k$ .*

We do not provide a rigorous proof here due to space constraints. The proof relies on the fact that, even though the  $\tilde{\mathbf{w}}_q$ 's are estimated in a different manner at each node, there is an inherent compatibility, i.e., it can be shown that all the  $\tilde{\mathbf{w}}_q$ 's represent a scaled version of the local GEVC  $\tilde{\mathbf{x}}_q$  (only for the DOA node, this link is explicit as we define  $\tilde{\mathbf{w}}_q$  as  $\tilde{\mathbf{x}}_q$ ). This makes the algorithm akin to Distributed Adaptive Covariance Generalized Eigenvector Estimation (DACGEE) [34], of which convergence and centralized optimality can be proven. The proof of Theorem I then follows from the convergence of the latter and the fact that the proposed algorithm is immune to the different scaling applied by each node's fusion rule  $\mathbf{f}_k$ .



**Fig. 1.** Convergence of the distributed MDMT-based algorithm

## 5. NUMERICAL SIMULATIONS

To investigate both the convergence and the performance of the proposed distributed MDMT-based algorithm, an acoustic scenario is simulated using the image method [35]. The room is rectangular ( $5m \times 5m \times 5m$ ), with reflection coefficients 0.2 for all surfaces. A WASN with 4 nodes ( $K = 4$ ) is considered, where each node is equipped with a uniform linear array with 3 microphones ( $M_k = 3$ ,  $\forall k \in \mathcal{K}$ ) and where the inter-microphone distance is  $10cm$ . The target speech source produces seven speech sentences, with one second of silence between each two consecutive sentences. Four localized multi-talker noise sources (mutually uncorrelated) are placed in the room at the broadside direction of the nodes, with equal noise power. We use a sampling frequency of 16kHz, a Hann-windowed DFT with window size 256 and with 50% overlaps. We assume a perfect VAD to exclude the effect of VAD errors. An uncorrelated white Gaussian noise is also added to each microphone signal to model the microphone's self-noise and other possible isotropic noise contributions. The simulations are carried out in batch mode, which means that the signal statistics are estimated over the full signal length in each iteration. Nodes 1 and 3 are tasked with MWF, node 2 with MVDR and node 4 with DOA estimation (via wideband ESPRIT). As a performance measure, at MWF and MVDR nodes we utilize the SNR improvement (in dB), i.e., the difference between the input and the output SNR, and at the DOA node we use the absolute error of the estimates (in degrees). Figure 1 illustrates both the convergence and the performance of the proposed distributed MDMT-based algorithm at all nodes. Cases where nodes estimate their node-specific tasks on their own, called as 'isolated', are added to also show the effectiveness of the algorithm. Results show that the estimates obtained with the distributed MDMT-based algorithm converges to the corresponding centralized estimates obtained in the centralized case.

## 6. CONCLUSION

We have studied a distributed multi-task problem in a WASN formed by three different groups of nodes. One of the groups is composed of nodes that aim at estimating the node-specific DOA of a desired speech source. The second and the third group are formed by nodes that aim at solving different node-specific speech enhancement problems by implementing either a MWF or a MDVR beamformer, respectively. We have derived a distributed algorithm that let the nodes cooperate to attain the network-wide centralized solution of their estimation problems without any knowledge on the tasks solved by the other nodes. To do so, the proposed algorithm employs compressive filter-and-sum operations and a low-rank approximation of the sensor correlation matrix based on the GEVD. Finally, simulations have shown the efficiency of the proposed algorithm.

## 7. REFERENCES

- [1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2001. ICASSP 2001*, 2001, vol. 4, pp. 2033–2036.
- [2] C. David, E. Deborah, and S. Mani, "Overview of sensor networks," *Computer*, vol. 32, pp. 41–50, 2004.
- [3] S. Haykin and K. J. R. Liu, *Handbook on array processing and sensor networks*, vol. 63, John Wiley & Sons, 2010.
- [4] S. Chouvardas, M. Muma, K. Hamaidi, S. Theodoridis, and A. M. Zoubir, "Distributed robust labeling of audio sources in heterogeneous wireless sensor networks," in *IEEE 40th International Conference on Acoustics, Speech and Signal Processing, 2015. ICASSP 2015*, 2015, pp. 5783–5787.
- [5] F. K. Teklehaymanot, M. Muma, B. Béjar, P. Binder, A. M. Zoubir, and M. Vetterli, "Robust diffusion-based unsupervised object labelling in distributed camera networks," in *AFRICON, 2015*, 2015, pp. 1–6.
- [6] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4583–4588, 2009.
- [7] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.
- [8] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, 2007.
- [9] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2010.
- [10] S. Chouvardas, K. Slavakis, and S. Theodoridis, "Adaptive robust distributed learning in diffusion sensor networks," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4692–4707, 2011.
- [11] J. Plata-Chaves, N. Bogdanovic, and K. Berberidis, "Distributed incremental-based RLS for node-specific parameter estimation over adaptive networks," in *IEEE 21st European Signal Conference, 2013. EUSIPCO 2013*, 2013.
- [12] N. Bogdanovic, J. Plata-Chaves, and K. Berberidis, "Distributed incremental-based LMS for node-specific adaptive parameter estimation," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5382–5397, 2014.
- [13] J. Plata-Chaves, N. Bogdanovic, and K. Berberidis, "Distributed diffusion-based LMS for node-specific parameter estimation over adaptive networks," *IEEE Transactions on Signal Processing*, vol. 13, no. 63, pp. 3448–3460, 2015.
- [14] J. Plata-Chaves, M. H. Bahari, M. Moonen, and A. Bertrand, "Unsupervised diffusion-based LMS for node-specific parameter estimation over wireless sensor networks," in *IEEE 41th International Conference on Acoustics, Speech and Signal Processing, 2016. ICASSP 2016*, 2016.
- [15] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4129–4144, 2014.
- [16] V. C. Gogineni and M. Chakraborty, "Diffusion adaptation over clustered multitask networks based on the affine projection algorithm," 2015 [Online]. Available: <http://arxiv.org/abs/1507.08566>.
- [17] V. C. Gogineni and M. Chakraborty, "Distributed multi-task APA over adaptive networks based on partial diffusion," 2015 [Online]. Available: <http://arxiv.org/abs/1509.09157>.
- [18] R. Nassif, C. Richard, A. Ferrari, and A. H. Sayed, "Multitask diffusion adaptation over asynchronous networks," 2014 [Online]. Available: <http://arxiv.org/abs/1412.1798>.
- [19] A. Hassani, A. Bertrand, and M. Moonen, "Cooperative integrated noise reduction and node-specific direction-of-arrival estimation in a fully connected wireless acoustic sensor network," *Signal Processing*, vol. 107, pp. 68–81, 2015.
- [20] A. Hassani, A. Bertrand, and M. Moonen, "Distributed signal subspace estimation based on local generalized eigenvector matrix inversion," in *IEEE 23rd European Signal Conference, 2015. EUSIPCO 2015*, 2015, pp. 1386–1390.
- [21] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *Audio, Speech and Language Processing, IEEE Transactions on*, vol. 17, no. 1, pp. 38–51, 2009.
- [22] 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, no. 10, pp. 5277–5291, 2010.
- [23] A. Bertrand and M. Moonen, "Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2196–2210, 2011.
- [24] J. Szurley, A. Bertrand, and M. Moonen, "Distributed adaptive node-specific signal estimation in heterogeneous and mixed-topology wireless sensor networks," *Signal Processing*, vol. 117, no. 12, pp. 44–60, 2015.
- [25] A. Hassani, A. Bertrand, and M. Moonen, "GEVD-based low-rank approximation for distributed adaptive node-specific signal estimation in wireless sensor networks," *To appear in IEEE Transactions on Signal Processing*, 2016 [Online]. Available: [ftp://ftp.esat.kuleuven.be/pub/SISTA/abertran/papers\\_website/AHassaniGEVDDANSE.pdf](ftp://ftp.esat.kuleuven.be/pub/SISTA/abertran/papers_website/AHassaniGEVDDANSE.pdf).
- [26] S. M. Golan, S. Gannot, and I. Cohen, "A reduced bandwidth binaural MVDR beamformer," in *9th International Workshop on Acoustic Echo and Noise Control, 2010. IWAENC 2010*, 2010.
- [27] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," in *IEEE Trans. Signal Processing*, 2002, vol. 50, pp. 2230–2244.
- [28] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filtering algorithms for noise reduction in cochlear implants," *Audio, Speech and Language Processing, IEEE Transactions on*, vol. 22, no. 4, pp. 785–799, 2014.
- [29] C. F. Van Loan G. H. Golub, *Matrix Computations, 3rd ed.*, Baltimore, MD: John Hopkins Univ. Press, 1996.
- [30] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 6, pp. 1071–1086, Aug 2009.
- [31] R. Schmidt, "Multiple emitter location and signal parameter estimation," in *IEEE Trans. on Antennas and Propagation*, 1986, vol. 34, pp. 276–280.
- [32] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984–995, 1989.
- [33] S. Markovich-Golan, A. Bertrand, M. Moonen, and S. Gannot, "Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks," *Signal Processing*, vol. 107, pp. 4–20, Feb. 2015.
- [34] A. Bertrand and M. Moonen, "Distributed adaptive generalized eigenvector estimation of a sensor signal covariance matrix pair in a fully-connected sensor network," *Signal Processing*, vol. 106, pp. 209–214, Jan. 2015.
- [35] J. Allen and D. Berkley, "Image method for efficiently simulating smallroom acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, 1979.