Signal Processing 107 (2015) 68–81

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Cooperative integrated noise reduction and node-specific direction-of-arrival estimation in a fully connected wireless acoustic sensor network

Amin Hassani*, Alexander Bertrand, Marc Moonen

KU Leuven, Department of Electrical Engineering (ESAT), Stadius Center for Dynamical Systems, Signal Processing and Data Analytics, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 18 February 2014
Received in revised form
9 July 2014
Accepted 1 September 2014
Available online 6 September 2014

Keywords:
Wireless acoustic sensor network
Distributed noise reduction
Node-specific DOA estimation
Generalized eigenvalue decomposition

ABSTRACT

In this paper, we consider cooperative node-specific direction-of-arrival (DOA) estimation in a fully connected wireless acoustic sensor network (WASN). We consider a scenario where each node is equipped with a local microphone array with a known geometry, but where the position of the nodes, as well as their relative geometry and hence the between-nodes signal coherence model is unknown. The local array geometry in each node defines node-specific DOAs with respect to a set of target speech sources and the aim is to estimate these in each node. We assume a noisy environment with localized and/or diffuse noise sources, i.e., the noise can be correlated over the different microphones. A distributed noise reduction algorithm can then be applied as a preprocessing step to denoise all the microphone signals of the WASN, based on the distributed adaptive node-specific signal estimation (DANSE) algorithm. The denoised local microphone signals can then be used in each node to estimate the node-specific DOAs by using a subspace-based DOA estimation, involving a (generalized) eigenvalue decomposition of the local microphone signal correlation matrices. It is seen that the fused microphone signals that are exchanged between the nodes in the DANSE algorithm can also be included in these correlation matrices to obtain improved DOA estimates, leading to a cooperative integrated noise reduction and DOA estimation scheme, where the noise reduction can actually be shortcut. The improved performance achieved by this cooperative DOA estimation is demonstrated by means of numerical simulations for two different subspace-based DOA estimation methods (MUSIC and ESPRIT).

1. Introduction

Microphone arrays facilitate spatiotemporal processing in acoustic applications and allow to exploit the spatial characteristics of the acoustic scenario to estimate a parameter or signal of interest. For example, they allow us to estimate the direction from which target sound signals originate, and/or to perform spatial filtering to suppress undesired sound signals coming from other directions. Microphone arrays have been widely used in hearing aids, teleconferencing systems, automatic speech recognition, hands-free telephony, etc. [1,2]. In general, the estimation performance improves when more microphones are used, and often also when the spacing between the microphones is increased. However, due to limitations in terms of space, power and processing capabilities of devices with embedded microphone arrays, it is not always

* Corresponding author. Tel.: +32 16 321927; fax: +32 16 321970.
E-mail addresses: amin.hassani@esat.kuleuven.be (A. Hassani), alexander.bertrand@esat.kuleuven.be (A. Bertrand), marc.moonen@esat.kuleuven.be (M. Moonen).

http://dx.doi.org/10.1016/j.sigpro.2014.09.001
0165-1684/© 2014 Elsevier B.V. All rights reserved.

possible to have an array with these desired characteristics [3].

One remedy could be to use a so-called wireless acoustic sensor network (WASN) [3]. A WASN consists of spatially distributed nodes, which are each equipped with a microphone array, a digital signal processing (DSP) unit and with wireless communication facilities to exchange data with other nodes in the WASN. As a result, the nodes can cooperate to solve certain acoustic signal processing tasks by exchanging relevant information amongst each other.

Direction-of-arrival (DOA) estimation of a target sound signal with respect to a given microphone array plays a crucial role in many applications. For example, based on the estimated DOA, one can control the look direction of a camera, or design an adaptive spatial filter to steer a beam towards the actual location of the target sound source and steer nulls towards the location of noise sources [4,5]. In this paper, we consider subspace-based DOA estimation techniques, which rely on the estimation of a so-called signal and noise subspace from the (generalized) eigenvalue decomposition of the microphone signal correlation matrices. For example, in the case of a narrowband signal and a fully calibrated microphone array, e.g., multiple signal classification (MUSIC) [6], maximum likelihood methods (MODEs) [7] or weighted subspace fitting (WSF) [8] can be used for DOA estimation. In this category, perhaps MUSIC is the most popular super resolution algorithm which can be applied to an array with an arbitrary but known geometry. However, MUSIC has a relatively high computational complexity. To reduce the computational cost, the so-called estimation of signal parameters via the rotational invariance technique (ESPRIT) [9] may be used as an alternative, which is also more robust with respect to array imperfections compared to MUSIC [9]. However, ESPRIT can only be applied to arrays with specific geometries [9].

When a wideband signal (such as a speech signal) is considered, a wideband extension of a narrowband DOA estimation should be utilized. Methods such as steered covariance matrix (STCM) [10] and spatial smoothing or array manifold interpolation (AMI) [11] are based on coherent focusing, i.e., they perform a narrowband method on a single frequency-steered coherent covariance matrix [10,12]. The class of so-called incoherent methods applies a narrowband method on each frequency bin independently and averages them all in the end [13,14]. It has been demonstrated in [15] that for sources with non-flat spectra (such as a speech signal), incoherent wideband MUSIC (IWM) leads to more accurate results compared to results of the coherently steered MUSIC. Therefore, but without loss of generality (w.l.o.g.), we consider incoherently averaged wideband methods in the sequel.

Although we will focus on MUSIC and ESPRIT in this paper, it is noted that there are several other subspace-based DOA estimation techniques for (partially) calibrated microphone arrays, e.g., rank reduction (RARE) [16], multiple invariance ESPRIT [17] and multiple invariances MUSIC and MODE [18].

In this paper, we consider cooperative node-specific direction-of-arrival (DOA) estimation in a fully connected wireless acoustic sensor network (WASN). We consider a noisy scenario where the position of the nodes as well as the relative geometry between them is unknown, but where each node is equipped with a local microphone array, with a known local geometry. The local geometry defines node-specific DOAs with respect to a set of target speech sources and the aim is to estimate these in each node.[1] This means that, unlike e.g., [19,20], the aim is to take benefit from the correlation between the microphone signals of the different arrays without modeling the unknown coherence structure between them. In practice, this is of great importance since even theoretical modeling of the spatial coherence cannot perfectly describe the environmental impacts (turbulence) that disturb the natural spherical propagation of wavefronts, especially when nodes are placed far apart [21,19].

We assume a noisy environment where the noise can be spatially correlated, i.e., due to localized and/or diffuse noise sources, which may deteriorate the performance of the DOA estimation. Therefore a multi-channel noise reduction algorithm can be applied as a preprocessing step to denoise all the microphone signals of the WASN. However, it is important that this noise reduction does not remove the spatial information associated with the target speech signal in the individual microphone signals. Furthermore, due to the unknown node and source positions, the noise reduction must rely on a blind beamforming technique, e.g., the multi-channel Wiener filter (MWF) [22]. In essence, MWF adopts a minimum mean square error (MMSE) criterion to estimate the desired target speech signal as it is observed in the microphones and therefore allows us to preserve the spatial characteristics of the target speech signal in the individual microphones such that DOA estimation can be performed on the denoised signals.

In order to apply a network-wide MWF, all the microphone signals of the WASN must be centralized and processed in a fusion center which may however demand a large communication bandwidth and computational power. An alternative could be a decentralized processing which is inherently scalable in terms of the communication bandwidth and computational complexity. The distributed adaptive node-specific signal estimation (DANSE) algorithm [23,24] is an iterative algorithm that distributes the processing task of the centralized MWF amongst the nodes. In the case of DANSE, the nodes broadcast fused microphone signals which, assuming a fully connected topology, can be captured by all other nodes in the network. Under mild conditions, DANSE converges to the centralized MWF solution as if all microphone signal signals were available in each node [23,24], allowing each node to optimally denoise all of its local microphone signals.

Because of the node-specific nature of DANSE, it is well suited to be applied in conjunction with the node-specific DOA estimation. The denoised local microphone signals can then be used in each node to estimate the node-specific DOAs by using a subspace-based DOA estimation, involving a (generalized) eigenvalue decomposition of the local microphone signal correlation matrices. It will be

[1] One application could be a video conferencing in which on top of the noise reduction for speech enhancement, we are also interested in steering each node's built-in camera towards the location of a certain speaker.

demonstrated that the fused microphone signals that are exchanged between the nodes in the DANSE algorithm can also be included in these correlation matrices to obtain improved DOA estimates, leading to a cooperative integrated noise reduction and DOA estimation scheme, where the DANSE final filtering stage can actually be shortcut.

The paper is organized as follows. The data model and the problem statement are presented in Section 2. Section 3 briefly reviews three subspace-based DOA estimation methods. Section 4 first describes the MWF algorithm for noise reduction and then outlines its distributed implementation based on the DANSE algorithm. Section 5 presents DANSE-based node-specific DOA estimation. Section 6 first addresses some evaluation aspects and then presents the simulation results. Finally, conclusions are drawn in Section 7.

2. Data model, problem statement and preview

We consider a WASN with K nodes in which each node $k \in \{1,\ldots,K\}$ is equipped with M_k microphones forming a uniform linear array (ULA), a set of M_k collinear microphones with equal spacing. The ULA geometry is selected here for the sake of an easy exposition, but w.l.o.g., i.e., other geometries may be considered as well as long as a DOA estimation procedure is used that can handle general geometries (e.g., MUSIC). The topology of the network is assumed to be fully connected which means that data broadcast by one node can be received by all other nodes in the network. The signal of microphone m at node k (frequency domain representations) can be decomposed as

$$y_{km}(\omega) = s_{km}(\omega) + n_{km}(\omega) \tag{1}$$

where $s_{km}(\omega)$ and $n_{km}(\omega)$ are the target speech component and the undesired noise component, respectively, and ω is the discrete frequency domain variable, where the resolution is defined by the discrete Fourier transform (DFT) of size L. In the sequel, whenever it is possible, we omit ω for the sake of brevity. By stacking (1) for $m = 1,\ldots,M_k$, we obtain $\mathbf{y}_k = [y_{k1} \;\; \ldots \;\; y_{kM_k}]^T = \mathbf{s}_k + \mathbf{n}_k$. All the $\mathbf{y}_k$'s are stacked in the full M-dimensional signal vector $\mathbf{y} = [\mathbf{y}_1^T \ldots \mathbf{y}_K^T]^T$ in which $M = \sum_{k=1}^{K} M_k$. Considering $\check{\mathbf{s}}$ as the signal generated by S target speech sources, we have $\mathbf{s}_k = \mathbf{A}_k(\boldsymbol{\theta}_k)\check{\mathbf{s}}$ in which the steering matrix $\mathbf{A}_k(\boldsymbol{\theta}_k)$ is defined as

$$\mathbf{A}_k(\boldsymbol{\theta}_k) = [\mathbf{a}_{k1}(\theta_{k1}) \ldots \mathbf{a}_{kS}(\theta_{kS})] \tag{2}$$

where $\mathbf{a}_{ks}(\theta_{ks})$ is the node-specific M_k-dimensional steering vector which is composed of the acoustic transfer functions (including room acoustics and microphone characteristics) from the s-th target speech source to the microphones of node k, and where $\boldsymbol{\theta}_k = [\theta_{k1} \ldots \theta_{kS}]^T$ is the set of corresponding node-specific DOAs with respect to the ULA of node k. For a ULA, the so-called array steering (response) vector $\mathbf{g}_k(\omega,\theta)$, which expresses the relative phase shifts of the target speech signal s in all microphones at node k with respect to the first microphone of its local ULA for a given DOA θ, can be generally modeled as [4]

$$\mathbf{g}_k(\omega,\theta) = \begin{bmatrix} 1 \\ e^{-j\omega d \;\cos(\theta) f_s/c} \\ \vdots \\ e^{-j\omega(M_k-1)\, d \;\cos(\theta) f_s/c} \end{bmatrix} \tag{3}$$

where f_s is the sampling frequency, c is the speed of sound, and d is the inter-microphone distance of the ULA of node k. Note that (3) assumes that all microphones have the same ideal omni-directional directivity response, that the relative attenuation factors are neglectable, and that far-field conditions are satisfied. These are common assumptions in DOA estimation algorithms [4], and they are a reasonable approximation when the inter-microphone distances are small. This is indeed the case for the local ULAs that are embedded in a sensor node, as envisaged in this paper. It is noted that we only impose these assumptions locally on a per-node basis, but not with respect to the network-wide array.

It is reiterated that each node $k \in \{1,\ldots,K\}$ observes node-specific DOAs θ_k originating from the same set of target speech sources, and that the goal for each node is to estimate its node-specific DOAs θ_k in a noisy acoustic environment. To this end, the nodes can first cooperate to denoise their local microphone signals, where each node broadcasts observations of fused microphone signals, as defined by the DANSE algorithm. In a second step, each node can use its M_k denoised microphone signals, as well as the denoised fused signals that are exchanged between the nodes in the noise reduction step, as inputs to a subspace estimation and a subspace-based DOA estimation algorithm. A schematic diagram of this approach with cascaded noise reduction and subspace estimation is depicted in Fig. 1 where the different blocks will be explained later in more detail. It will be demonstrated, however, that the DANSE algorithm allows us to integrate the subspace estimation into the noise reduction and that

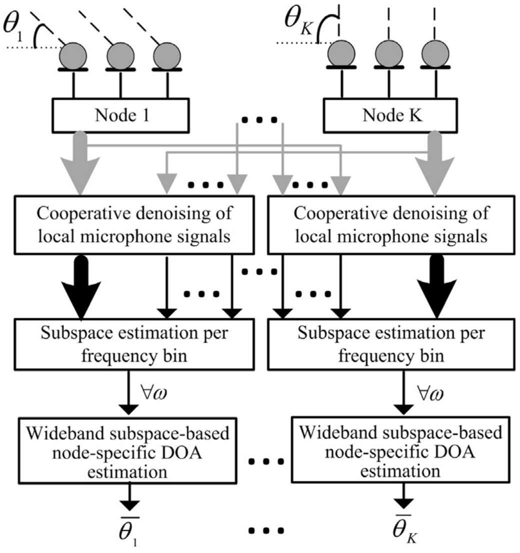
θ_1
θ_K
Node 1
Node K
Cooperative denoising of local microphone signals
Cooperative denoising of local microphone signals
Subspace estimation per frequency bin
Subspace estimation per frequency bin
$\forall\omega$
$\forall\omega$
Wideband subspace-based node-specific DOA estimation
Wideband subspace-based node-specific DOA estimation
$\hat{\theta}_1$
$\hat{\theta}_K$

Fig. 1. Node-specific DOA estimation scheme with cascaded noise reduction and subspace estimation.

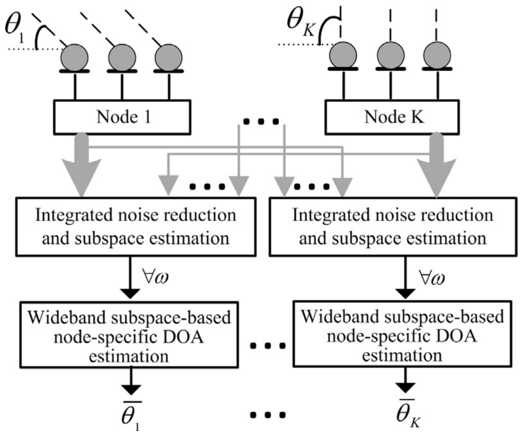
θ_1
θ_K
Node 1
Node K
Integrated noise reduction and subspace estimation
Integrated noise reduction and subspace estimation
$\forall\omega$
$\forall\omega$
Wideband subspace-based node-specific DOA estimation
Wideband subspace-based node-specific DOA estimation
$\bar{\theta}_1$
$\bar{\theta}_K$

Fig. 2. Node-specific DOA estimation scheme with integrated noise reduction and subspace estimation.

the DANSE final filtering stage can then be shortcut. A schematic diagram of this approach with integrated noise reduction and subspace estimation is depicted in Fig. 2.

For the sake of brevity, throughout Sections 3–5 we only consider the special case of a single target speech source, i.e., $S=1$. The multi-source case can be derived straightforwardly, where all vector-variables can be replaced by their matrix equivalent. A multi-source scenario is considered in Section 6 to further show the effectiveness of the proposed cooperative method in the general case.

3. Subspace-based DOA estimation

Subspace-based DOA estimation algorithms essentially extract the so-called signal and noise subspace from the microphone signal correlation matrices and estimate the DOAs based on them. In this section, we briefly review the MUSIC and ESPRIT algorithms and their incoherent wideband extensions. It is noted that other subspace-based DOA estimation algorithms can be used as well. We first consider the case where the noise at each microphone is independent and identically distributed (i.i.d.), and we will later consider the more general case. The microphone signal correlation matrix at node k is then equal to

$$\mathbf{R}_{\mathbf{y}_k\mathbf{y}_k} = E\{\mathbf{y}_k\mathbf{y}_k^H\} = \mathbf{R}_{\mathbf{s}_k\mathbf{s}_k} + \sigma_{n_k}^2 \mathbf{I}_{M_k} \tag{4}$$

where $\mathbf{R}_{\mathbf{s}_k\mathbf{s}_k} = E\{\mathbf{s}_k\mathbf{s}_k^H\}$, $\sigma_{n_k}^2$ is the noise power on each microphone of node k, $E\{\cdots\}$ denotes the expected value operator, the superscript H indicates the conjugate transpose operator, and $\mathbf{I}_{M_k}$ is the $M_k \times M_k$ identity matrix. It is noted that in (4), $\mathbf{R}_{\mathbf{y}_k\mathbf{y}_k}$ and $\mathbf{R}_{\mathbf{s}_k\mathbf{s}_k}$ have the same eigenvectors, which is due to the i.i.d. assumption on the noise. In case of a single target speech source, the correlation matrix of the target speech signal component of the microphone signals, $\mathbf{s}_k$, can be written as

$$\mathbf{R}_{\mathbf{s}_k\mathbf{s}_k} = \sigma_s^2 \mathbf{a}_k \mathbf{a}_k^H \tag{5}$$

where $\sigma_s^2 = E\{|\check{s}|^2\}$ is the power of the target speech source signal.

3.1. MUSIC

In this section, we provide a very brief outline of the MUSIC algorithm, and we refer to [6] for further details. Basically, MUSIC decomposes the correlation matrix $\mathbf{R}_{\mathbf{y}_k\mathbf{y}_k}$ at each frequency ω, into a signal and noise subspace which are orthogonal to each other, e.g., by means of an eigenvalue decomposition (EVD). In the case of a single target speech source, the signal subspace is defined by the eigenvector corresponding to the largest eigenvalue of $\mathbf{R}_{\mathbf{y}_k\mathbf{y}_k}$, and the noise subspace is spanned by the remaining (M_k-1) eigenvectors. The matrices containing the basis vectors for the signal and noise subspace are then denoted as

$$\mathbf{E}_{\mathbf{s}_k} = [\mathbf{q}_{k1}] \tag{6}$$

$$\mathbf{E}_{\mathbf{n}_k} = [\mathbf{q}_{k2}|\ldots|\mathbf{q}_{kM_k}] \tag{7}$$

where $\mathbf{q}_{k1}$ is the eigenvector corresponding to the largest eigenvalue, and $\mathbf{E}_{\mathbf{n}_k}^H \mathbf{q}_{k1} = \mathbf{0}$. Note that these subspaces are different at each frequency ω. For a narrowband signal with a central frequency ω, we define the so-called pseudospectrum as

$$\frac{1}{|\mathbf{g}_k^H \mathbf{E}_{\mathbf{n}_k} \mathbf{E}_{\mathbf{n}_k}^H \mathbf{g}_k|} \tag{8}$$

where $\mathbf{g}_k(\omega,\theta)$ is defined in (3). It is noted that the denominator will be close to zero if θ equals the true DOA, since then $\mathbf{g}_k \approx \mathbf{a}_{k1} \approx q_{k1}$. Therefore, the θ_k for which the wideband pseudospectrum[2] is maximized will be the estimated DOA, i.e. (we use an overline (bar) to denote an estimate)

$$\bar{\theta}_k = \arg\max_{\theta_k} \frac{1}{\sum_\omega |\mathbf{g}_k^H \mathbf{E}_{\mathbf{n}_k} \mathbf{E}_{\mathbf{n}_k}^H \mathbf{g}_k|} \tag{9}$$

where an exhaustive search over all possible θ_k is performed. Note that although we are considering a ULA, MUSIC also works with other array topologies as long as the array geometry (and hence the array steering vector $\mathbf{g}_k(\omega,\theta)$) is known and fully calibrated.

3.2. ESPRIT

ESPRIT [9] is an alternative subspace-based DOA estimation algorithm, which does not require an exhaustive search over all possible DOAs, leading to a computational complexity that is typically lower compared to MUSIC. ESPRIT essentially operates on a doublet structure which means that it decomposes the array into several two-element sub-arrays (doublets) with a known identical displacement vector, i.e. all doublets are identically oriented with the same local inter-microphone distance. Fig. 3 shows how ESPRIT splits a three-microphone ULA into two overlapping doublets.

In the rest of this section we briefly describe the ESPRIT algorithm for a ULA and the reader is referred to [9] for further details.

[2] Note that there are also other alternatives to combine the different frequencies (incoherent averaging) in a wideband pseudospectrum [25].

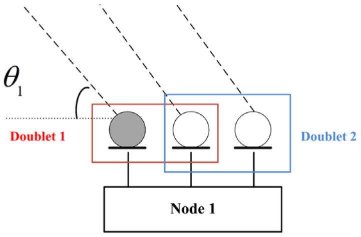

Fig. 3. Two sub-arrays (doublets) which are used by ESPRIT for a three-element array.

Given the signal subspace $\mathbf{E}_{\mathbf{s}_k}$ in (6) which is extracted from the correlation matrix $\mathbf{R}_{\mathbf{y}_k\mathbf{y}_k}$ by means of an EVD, ESPRIT defines two subvectors $\mathbf{v}_{1_k}$ and $\mathbf{v}_{2_k}$ with the first M_k-1 and the last M_k-1 entries of $\mathbf{E}_{\mathbf{s}_k}$. The i-th component in $\mathbf{v}_{1_k}$ and $\mathbf{v}_{2_k}$ then corresponds to the i-th doublet in the array. Whenever $\mathbf{E}_{\mathbf{s}_k}$ exactly matches the array steering vector expressed in (3) we can write

$$\mathbf{v}_{2_k} = \mathbf{v}_{1_k}\psi \tag{10}$$

where for an estimated $\mathbf{E}_{\mathbf{s}_k}$, the ψ is estimated by using a least squares solver, i.e.,

$$\overline{\psi} = \left(\mathbf{v}_{1_k}^H \mathbf{v}_{1_k}\right)^{-1} \mathbf{v}_{1_k}^H \mathbf{v}_{2_k}. \tag{11}$$

Since both $\mathbf{v}_{1_k}$ and $\mathbf{v}_{2_k}$ are often noisy, a total least squares (TLS) solver can be used alternatively which is based on the singular value decomposition (SVD) of the matrix $[\mathbf{v}_{1_k}|\mathbf{v}_{2_k}]$. Considering the array manifold vector expressed in (3), we can then write

$$\overline{\theta}_k = \cos^{-1}\left(\frac{\angle\overline{\psi}}{\omega d f_s/c}\right). \tag{12}$$

For the case of a wideband signal, the DOA estimation is then obtained from the average over the DOA estimates for different frequencies (incoherent averaging).

4. Noise reduction preprocessing

In the previous section we have considered the problem of DOA estimation in an acoustic environment with i.i.d. microphone noise, i.e. the noise signals are spatially uncorrelated and identically distributed. In the sequel, we consider the general case where the noise signals in the different microphones may be correlated, e.g., due to the presence of localized acoustic noise sources. In such an environment and in order to estimate the DOAs using one of the subspace-based algorithms explained so far, we can perform a noise reduction as a preprocessing step to denoise the local microphone signals in each node. The target speech correlation matrix $\mathbf{R}_{\mathbf{s}_k\mathbf{s}_k}$ (see (5)) is then estimated based on the denoised local microphone signals. For the noise reduction we consider multichannel Wiener filtering (MWF) [22] which in contrast to standard beamforming does not require prior information on the microphone or source positions. In Section 4.1, we briefly review the MWF, and in Section 4.2, we explain how the nodes can cooperate to improve the overall noise reduction performance.

4.1. Multi-channel Wiener filter

The goal of MWF is to estimate the target speech signal s_{km} as it is observed in the m-th microphone of node k. MWF performs a filter-and-sum operation in which the filter coefficients $\mathbf{w}_{km}$ are selected such that the following mean square error (MSE) cost function is minimized:

$$\min_{\mathbf{w}_{km}} E\{|\mathbf{e}_m^H\mathbf{s}_k - \mathbf{w}_{km}^H\mathbf{y}_k|^2\} \tag{13}$$

where $\mathbf{e}_m = [0\ldots0\;1\;0\ldots0]^T$, where 1 is the m-th coefficient. The solution to this minimum MSE (MMSE) problem, assuming independence between $\mathbf{s}_k$ and $\mathbf{n}_k$, is given as [22]

$$\mathbf{w}_{km} = \left(\mathbf{R}_{\mathbf{y}_k\mathbf{y}_k}\right)^{-1}\mathbf{R}_{\mathbf{s}_k\mathbf{s}_k}\mathbf{e}_m. \tag{14}$$

Again assuming independence between $\mathbf{s}_k$ and $\mathbf{n}_k$, we can write

$$\mathbf{R}_{\mathbf{s}_k\mathbf{s}_k} = \mathbf{R}_{\mathbf{y}_k\mathbf{y}_k} - \mathbf{R}_{\mathbf{n}_k\mathbf{n}_k} \tag{15}$$

where $\mathbf{R}_{\mathbf{n}_k\mathbf{n}_k} = E\{\mathbf{n}_k\mathbf{n}_k^H\}$. Estimation of the covariance matrices ($\mathbf{R}$ matrices) can be done by time averaging in the short-time-Fourier-transform (STFT) domain. $\mathbf{R}_{\mathbf{y}_k\mathbf{y}_k}$ can be estimated during “speech-and-noise” signal segments and $\mathbf{R}_{\mathbf{n}_k\mathbf{n}_k}$ can be estimated during “noise-only” signal segments. To distinguish between “noise-only” and “speech-and-noise” signal segments, a voice activity detection (VAD) mechanism must be applied [22].

In the case of a single speech source, $\mathbf{R}_{\mathbf{s}_k\mathbf{s}_k}$ is given by (5) and hence is a rank-1 matrix. In practice, however, due to (a) the finite DFT size in the STFT analysis, (b) the non-stationarity of the noise and (c) the finite observation set (which leads to estimation errors), the rank of the estimated $\overline{\mathbf{R}}_{\mathbf{s}_k\mathbf{s}_k}$ will be greater than one. Moreover, in low input signal to noise ratio (iSNR) conditions, we have that $\overline{\mathbf{R}}_{\mathbf{y}_k\mathbf{y}_k} \approx \overline{\mathbf{R}}_{\mathbf{n}_k\mathbf{n}_k}$, such that the estimation of $\mathbf{R}_{\mathbf{s}_k\mathbf{s}_k}$ via subtraction in (15) may result in a covariance matrix $\overline{\mathbf{R}}_{\mathbf{s}_k\mathbf{s}_k}$ which is not positive (semi-)definite and this may result in an unstable noise reduction performance [26]. A remedy for this problem is to choose a rank-1 approximation based on either the EVD of $\overline{\mathbf{R}}_{\mathbf{s}_k\mathbf{s}_k}$ or the generalized EVD (GEVD) of $\overline{\mathbf{R}}_{\mathbf{y}_k\mathbf{y}_k}$ and $\overline{\mathbf{R}}_{\mathbf{n}_k\mathbf{n}_k}$ [27]. GEVD-based rank-1 approximation has been shown to deliver the best performance, as it effectively selects the “mode” corresponding to the highest SNR [27]. Therefore the GEVD based rank-1 approximation is utilized in the sequel.[3]

Given the matrices $\overline{\mathbf{R}}_{\mathbf{y}_k\mathbf{y}_k}$ and $\overline{\mathbf{R}}_{\mathbf{n}_k\mathbf{n}_k}$, their joint diagonalization can be written as

$$\begin{aligned}\overline{\mathbf{R}}_{\mathbf{y}_k\mathbf{y}_k} &= \overline{\mathbf{V}}_k\overline{\boldsymbol{\Sigma}}_{\mathbf{y}_k}\overline{\mathbf{V}}_k^H\\ \overline{\mathbf{R}}_{\mathbf{n}_k\mathbf{n}_k} &= \overline{\mathbf{V}}_k\overline{\boldsymbol{\Lambda}}_{\mathbf{y}_k}\overline{\mathbf{V}}_k^H\end{aligned} \tag{16}$$

so that

$$\overline{\mathbf{R}}_{\mathbf{n}_k\mathbf{n}_k}^{-1}\overline{\mathbf{R}}_{\mathbf{y}_k\mathbf{y}_k} = \overline{\mathbf{V}}_k^{-H}(\overline{\boldsymbol{\Lambda}}_{\mathbf{n}_k}^{-1}\overline{\boldsymbol{\Sigma}}_{\mathbf{y}_k})\overline{\mathbf{V}}_k^H = \overline{\mathbf{V}}_k^{-H}\overline{\boldsymbol{\Sigma}}_k\overline{\mathbf{V}}_k^H \tag{17}$$

where $\overline{\mathbf{V}}_k$ is an invertible matrix (not necessarily orthogonal) and the columns of $\overline{\mathbf{V}}_k^{-H}$ are the generalized eigenvectors, $\overline{\boldsymbol{\Sigma}}_{\mathbf{y}_k} = diag\{\overline{\sigma}_1\cdots\overline{\sigma}_{M_k}\}, \overline{\boldsymbol{\Lambda}}_{n_k} = diag\{\overline{\lambda}_1\cdots\overline{\lambda}_{M_k}\}$ and the

[3] This is w.l.o.g. since EVD-based rank-1 approximation can also be utilized for the proposed cooperative DOA method.

real-valued generalized eigenvalues are defined by the diagonal matrix $\overline{\mathbf{\Sigma}}_k = diag\{\overline{\sigma}_1/\overline{\lambda}_1 \cdots \overline{\sigma}_{M_k}/\overline{\lambda}_{M_k}\}$ [28,22]. The GEVD-based rank-1 approximation of $\mathbf{R}_{\mathbf{s}_k\mathbf{s}_k}$ is then given by

$$\overline{\mathbf{R}}_{\mathbf{s}_k\mathbf{s}_k} = \overline{\mathbf{v}}_k(\overline{\sigma_1} - \overline{\lambda}_1)\overline{\mathbf{v}}_k^H \tag{18}$$

where $\overline{\mathbf{v}}_k$ is the first column of $\overline{\mathbf{V}}_k$. The MWF formula then becomes (compare with (14))

$$\overline{\mathbf{w}}_{km} = \left(\overline{\mathbf{R}}_{\mathbf{y}_k\mathbf{y}_k}\right)^{-1}\overline{\mathbf{v}}_k\overline{v}_{km}^*(\overline{\sigma}_1 - \overline{\lambda_1}) \tag{19}$$

where $\overline{v}_{km}$ is the m-th component of $\overline{\mathbf{v}}_k$. Finally, the denoised version of the m-th microphone of node k is computed as

$$\overline{d}_{km} = \overline{\mathbf{w}}_{km}^H\mathbf{y}_k. \tag{20}$$

After denoising all the microphone signals in each node k, the resulting denoised microphone signals $\overline{\mathbf{d}}_k = [\overline{d}_{k1} \cdots \overline{d}_{kM_k}]$ can be fed to the DOA estimation algorithm.

4.2. DANSE-based cooperative noise reduction

In Section 4.1 we have assumed that each node operates on its own in order to denoise its local microphone signals. If a node k would also have access to the microphone signals of all the other nodes, i.e., the entire M-dimensional signal vector $\mathbf{y}$, it could compute the network-wide MWF to obtain a substantially better noise reduction. However, this would require a large communication bandwidth and computational power in each node. An alternative could be a decentralized processing which is inherently scalable in terms of the communication bandwidth and computational complexity. This is achieved by the DANSE algorithm [23,24], which can be viewed as a distributed implementation of the network-wide MWF. The computational cost is then shared between the different nodes, and each node only broadcasts one fused signal to the other nodes, rather than its full M_k-dimensional signal vector $\mathbf{y}_k$. Consequently and compared to the centralized network-wide MWF (based on the full M-dimensional signal vector $\mathbf{y}$), the algorithm reduces the required per-node communication bandwidth by a factor M_k as well as the number of input channels in each node which results in a significant computational complexity reduction. It has been shown in [23] that the DANSE algorithm is able to denoise the microphone signals in each node as if each node would have access to all the WASN microphone signals, despite the fact that only one signal per node is broadcast. In Section 5, we will explain that the fused microphone signals that are exchanged between the nodes in the DANSE algorithm can also be exploited to improve the subspace estimation in each node and hence the DOA estimation. In the rest of this section we briefly review the DANSE algorithm for a single target speech source in a fully connected WASN. It is noted that this is only a very concise review to give an idea of the underlying principles. For more details, as well as extensions to multiple speakers and other network topologies, we refer to [23,24,29].

In DANSE (for a single source), each node k creates one fused microphone signal z_k by means of a filter-and-sum operation on its own microphone signals and then broadcasts it to all other nodes. The signal z_k at node k is computed as

$$z_k = \mathbf{f}_k^H\mathbf{y}_k \tag{21}$$

where the fusion vector $\mathbf{f}_k$ will be defined later. We define $\mathbf{z} = [z_1 \ \ldots \ z_K]^T$ and we write $\mathbf{z}_{-k}$ to denote the vector $\mathbf{z}$ in which z_k is excluded.

Node k's own microphone signals together with the z_k-signals received from the other nodes are stacked in a vector

$$\tilde{\mathbf{y}}_k = [\mathbf{y}_k^T\mathbf{z}_{-k}^T]^T \tag{22}$$

For the sake of an easy exposition, we first assume that each node only estimates the target speech signal in its first microphone, and we later extend this for the other microphones. Each node k then computes the local MWF (compare with (19)) as

$$\tilde{\mathbf{w}}_{k1} = \mathbf{R}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}^{-1}\tilde{\mathbf{v}}_k\tilde{v}_{k1}^*(\tilde{\sigma}_1 - \tilde{\lambda}_1) \tag{23}$$

where the $\tilde{\cdot}$ notation is used for quantities that are computed based on the extended signal $\tilde{\mathbf{y}}_k$ rather than $\mathbf{y}_k$, and we also replace $\tilde{\bar{\cdot}}$ with $\tilde{\cdot}$ in the sequel for the sake of conciseness. The $\tilde{\mathbf{w}}_{k1}$ is then partitioned into two parts, one applied to $\mathbf{y}_k$ and one applied to $\mathbf{z}_{-k}$, i.e.,

$$\tilde{\mathbf{w}}_{k1} = \begin{bmatrix} \mathbf{h}_{k1} \\ \mathbf{g}_{k1} \end{bmatrix} \tag{24}$$

and the denoised signal of the first microphone at node k can then be written as

$$\tilde{d}_{k1} = \tilde{\mathbf{w}}_{k1}^H\tilde{\mathbf{y}}_k = \mathbf{h}_{k1}^H\mathbf{y}_k + \mathbf{g}_{k1}^H\mathbf{z}_{-k}. \tag{25}$$

In DANSE, the $\mathbf{f}_k$ in (21) is then set to $\mathbf{h}_{k1}$, i.e.,

$$\forall k \in \{1, \ldots, K\} : \mathbf{f}_k = \mathbf{h}_{k1} \tag{26}$$

Note that the fusion vector $\mathbf{f}_k$ is not only a compressor in each node k to generate the z_k signal from the local microphone signals, but also is a part of the MWF for the first microphone signal in (25). However, this is a chicken-and-egg problem since to obtain $\mathbf{f}_k$ we have to compute (23)–(26) first, which in turn require the $\mathbf{z}_{-k}$ from the other nodes. Starting with random entries for the $\mathbf{f}_k$s, $\forall k \in \{1, \ldots, K\}$, DANSE lets each node k iteratively update first its $\mathbf{R}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}$ and $\mathbf{R}_{\tilde{\mathbf{n}}_k\tilde{\mathbf{n}}_k}$ and then $\tilde{\mathbf{w}}_{k1}$ and $\mathbf{f}_k$ (using (23)–(26)) based on the most recent microphone signals of $\tilde{\mathbf{y}}_k$. The updating procedure can be done in a sequential round-robin fashion [23], or all the nodes can update simultaneously (requiring some minor modifications) [24]. In [23] it is demonstrated that DANSE converges to the network-wide MWF, as if all microphone signals were available in each node.

So far we have only denoised the first microphone of each node $k \in \{1, \ldots, K\}$ with the DANSE algorithm. It can be shown that the other microphone signals can also be optimally denoised based on the same z_k-signals, even though the fusion vectors that generate these z_k-signals are based on the MWF problems corresponding to the first microphones in each node. For example to denoise the second microphone signal of node k, $\tilde{\mathbf{w}}_{k2}$ is computed with (23), where $\tilde{v}_{k1}$ is merely replaced by $\tilde{v}_{k2}$.

5. Cooperative integrated noise reduction and DOA estimation

5.1. Cooperative DOA estimation

In the previous section, we have introduced the DANSE algorithm to denoise the local microphone signals in each node where the nodes cooperate with each other by exchanging fused microphone signals. This preprocessing step allows us to reduce the effect of noise in the DOA estimation in each node. As depicted in Fig. 1, the next step is to estimate the node-specific DOA at each node k, based on all available denoised signals. Now the objective is to re-use the fused microphone signals that are broadcast in the DANSE algorithm to further improve the node-specific DOA estimation performance, leading to a cooperative integrated noise reduction and DOA estimation scheme. To achieve this, the $\mathbf{z}_{-k}$ signals should first also be denoised by the local MWF, i.e., all the signals in $\tilde{\mathbf{y}}_k$ are first denoised and then fed to the DOA estimation. By stacking the M_k denoised microphone signals together with $K-1$ denoised z_k signals, we can define $\tilde{\mathbf{d}}_k$ as

$$\tilde{\mathbf{d}}_k = [\tilde{d}_{k1} \cdots \tilde{d}_{kM_k}, \cdots \tilde{d}_{k(M_k+K-1)}] \tag{27}$$

and its corresponding correlation matrix as

$$\mathbf{R}_{\tilde{\mathbf{d}}_k\tilde{\mathbf{d}}_k} = E\{\tilde{\mathbf{d}}_k \tilde{\mathbf{d}}_k^H\} \tag{28}$$

which will be used in the sequel for the node-specific DOA estimation. In order to extract the local signal and noise subspace at node k, an EVD of $\mathbf{R}_{\tilde{\mathbf{d}}_k\tilde{\mathbf{d}}_k}$ is performed. If $\overline{\mathbf{u}}_{k,max}$ is the eigenvector corresponding to the largest eigenvalue, then it is noted that since the relative geometry between the nodes is unknown, only the first M_k entries of $\overline{\mathbf{u}}_{k,max}$, defined as $\overline{\mathbf{u}}_k$, can be used for the DOA estimation. Although this means that we throw away information, there is still implicit cooperation between the nodes as the EVD indeed also relies on the fused microphone signals from other nodes, allowing us to exploit more correlation structure in the subspace estimation.[4] Fig. 4 visualizes the dimension of the correlation matrix $\mathbf{R}_{\tilde{\mathbf{d}}_k\tilde{\mathbf{d}}_k}$ in the proposed cooperative approach, compared with two other approaches. The first is a “centralized” approach where each node has access to all M microphone signals throughout the entire network, i.e. $\mathbf{y}_k = \mathbf{y}$ in (13). In this case we can estimate the full M-dimensional correlation matrix which indeed leads to a better node-specific subspace estimation and hence DOA estimation, but which has a high communication cost. Secondly we will consider an “isolated” approach where each node has only access to its own microphone signals and where there is no cooperation, i.e. the input of each local MWF is merely the $\mathbf{y}_k$ already introduced in Section 2. As can be seen in Fig. 4, the DANSE-based node-specific DOA estimation uses more data than the isolated approach, but less data than the centralized approach. This figure also shows that each node k estimates its node-specific DOA only based on the M_k local entries corresponding to its local array.

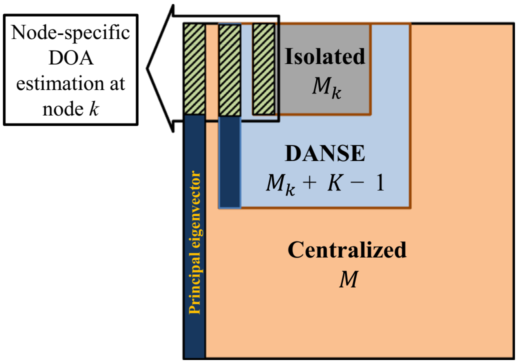

Fig. 4. Dimension of the correlation matrix $\mathbf{R}_{\tilde{\mathbf{d}}_k\tilde{\mathbf{d}}_k}$ when extra signals are included in the centralized and DANSE approaches, compared to the dimension of the isolated approach.

Intuitively, there are three effects which explain why the proposed cooperative approach results in a better node-specific DOA estimation at each node:

1. The distributed noise reduction allows the DOA estimation to perform better for a given input SNR and number of microphones.
2. An enhanced subspace estimation is obtained by exploiting more structure due to extending the covariance matrix (see also Section 5.2). Our proposed cooperative approach exploits coherence between nodes, but without using a model for this coherence based on relative geometry etc., because such models with large inter-microphone distances are typically inaccurate in the first place due to the environmental impacts (turbulence) that disturb the natural spherical propagation of wavefronts.
3. The proposed approach uses the broadcasting signals of DANSE to extend the correlation matrix, which typically have better SNR than when the nodes would merely transmit raw microphone signals (see also Section 6).

According to (5), $\overline{\mathbf{u}}_k$ can be treated as a normalized estimate of the steering vector, i.e. $\overline{\mathbf{a}}_k \approx \beta \overline{\mathbf{u}}_k$, where β is an unknown complex number.

To assess the performance of the cooperative DOA estimation in conjunction with the DANSE algorithm, we can apply any of the subspace-based DOA estimation algorithms explained in Section 3. In MUSIC, referring to (6), we will have $\mathbf{q}_{k1} = \overline{\mathbf{u}}_k$ and then $\mathbf{E}_{\mathbf{n}_k}$ in (7) is computed as the (M_k-1)-dimensional subspace orthogonal to $\overline{\mathbf{u}}_k$. Likewise, in ESPRIT, $\mathbf{E}_{\mathbf{s}_k} = \mathbf{q}_{k1} = \overline{\mathbf{u}}_k$.

Remark: In terms of the computational complexity of the proposed cooperative approach, the following items should be considered:

1. When the noise reduction part is taken into account, we have the inversion (see (23)) of K times (one per node) an $(M_k+K-1) \times (M_k+K-1)$ matrix, versus one $M \times M$ matrix (in the centralized case).

[4] We will later provide some more motivation for this claim.

2. In the subspace-based DOA estimation part, we compute an EVD (see (28)) of K times an $(M_k+K-1)\times(M_k+K-1)$ matrix, versus one $M\times M$ matrix.

Therefore in both cases, there is a significant benefit since both inversion and EVD are $O(N^3)$ procedures, where N is the dimension of the matrix. However it is known that in practice the communication unit of a WASN node (often battery-powered) consumes much more energy than its DSP unit. Therefore, even more important than the computational gain, as mentioned in Section 4.2, there is a reduction in communication cost with a factor M_k per node.

5.2. Theoretical motivation

In this section we provide a brief theoretical motivation that explains why the proposed cooperative method improves the performance of the node-specific DOA estimation method. For the sake of an easy exposition, we consider a single-node WASN with M microphones in which all the microphones receive the signal of a target source at the same time (corresponds to a DOA of 90° in far-field conditions). Moreover, for the sake of mathematical tractability, we consider i.i.d. noise components. Therefore, in this case we can write the normalized steering vector as $\hat{\mathbf{a}}=(1/\sqrt{M})\mathbf{1}_M$, where $\mathbf{1}_M$ is a M-dimensional vector with all entries equal to 1. Similar to (4) and (5), we can here write

$$\mathbf{R}_{\mathbf{yy}}=E\{\mathbf{y}\mathbf{y}^H\}=\mathbf{a}\sigma_s^2\mathbf{a}^H+\sigma_n^2\mathbf{I}_M. \tag{29}$$

In practice, $\mathbf{R}_{\mathbf{yy}}$ is estimated via time averaging. By defining the $M\times N$ matrix $\mathbf{Y}$ in which each column corresponds to an observation of $\mathbf{y}$ at a certain time instant, we can approximate $\mathbf{R}_{\mathbf{yy}}$ as

$$\mathbf{R}_{\mathbf{yy}}\approx\overline{\mathbf{R}}_{\mathbf{yy}}=\frac{1}{N}\mathbf{Y}\mathbf{Y}^H. \tag{30}$$

Based on an EVD we have

$$\mathbf{R}_{\mathbf{yy}}\hat{\mathbf{a}}=\lambda_{max}\hat{\mathbf{a}} \tag{31}$$

where $\lambda_{max}=\sigma_s^2M+\sigma_n^2$. The other eigenvalues $\lambda_m, m=2,\dots,M$ are equal to σ_n^2 and correspond to the $(M-1)$-dimensional noise subspace. Let $\overline{\mathbf{a}}$ denote the normalized steering vector estimate computed from the sample covariance matrix $\overline{\mathbf{R}}_{\mathbf{yy}}$. Define the estimation error then as $\Delta\mathbf{a}=\overline{\mathbf{a}}-\hat{\mathbf{a}}$. The second order statistic of $\Delta\mathbf{a}$ can then be described as (see, e.g., formula (4) in [30])

$$E\{\Delta\mathbf{a}\Delta\mathbf{a}^H\}=\frac{\lambda_{max}}{N}\sum_{m=2}^{M}\frac{\lambda_m}{(\lambda_{max}-\lambda_m)^2}\hat{\mathbf{a}}_m\hat{\mathbf{a}}_m^H. \tag{32}$$

By plugging $\sum_{m=2}^{M}\hat{\mathbf{a}}_m\hat{\mathbf{a}}_m^H=\mathbf{I}_M-\hat{\mathbf{a}}\hat{\mathbf{a}}^H$ in (32), and setting $\lambda_{max}=\sigma_s^2M+\sigma_n^2$ and $\lambda_m=\sigma_n^2, m=2,\dots,M$, we can write

$$E\{\Delta\mathbf{a}\Delta\mathbf{a}^H\}=\frac{M\sigma_s^2\sigma_n^2+\sigma_n^4}{M^2N\sigma_s^4}\left(\mathbf{I}_M-\hat{\mathbf{a}}\hat{\mathbf{a}}^H\right). \tag{33}$$

Now the objective is to determine how adding more signals (increasing M) affects the steering vector estimation performance. To this end, we examine the MSE of the estimation error $\Delta\mathbf{a}$. Since $\hat{\mathbf{a}}=(1/\sqrt{M})\mathbf{1}_M$, and with some straightforward simplifications, we find that

$$\begin{aligned}E\{\|\Delta\mathbf{a}\|^2\}&=E\{\mathrm{Tr}\{\Delta\mathbf{a}\Delta\mathbf{a}^H\}\}=\mathrm{Tr}\{E\{\Delta\mathbf{a}\Delta\mathbf{a}^H\}\}\\&=\frac{M\sigma_s^2\sigma_n^2+\sigma_n^4}{M^2N\sigma_s^4}(M-1).\end{aligned} \tag{34}$$

Finally we define MSE(M) as the MSE per entry of $\Delta\mathbf{a}$ (hence independent of the length of M), i.e.,

$$\mathrm{MSE}(M)=\frac{E\{\|\Delta\mathbf{a}\|^2\}}{M}=\frac{1}{MN}\left[\left(\frac{M-1}{M}\right)\frac{\sigma_n^2}{\sigma_s^2}+\left(\frac{M-1}{M^2}\right)\frac{\sigma_n^4}{\sigma_s^4}\right]. \tag{35}$$

As can be seen, $\lim_{M\to\infty}\mathrm{MSE}(M)=0$, i.e., the performance of the steering vector estimation is improved as the dimension of the sample covariance matrix increases, i.e., if extra signals are added. Indeed, this also leads to a better DOA estimation when a subspace-based DOA estimation is considered. Fig. 5 is provided to better clarify the relationship between the dimension of the sample covariance matrix, i.e., M, and the steering vector estimation performance. This simulation is carried out with $\sigma_s=2$, $\sigma_n=1.3$, N=200 over different values of M, and averaged over 200 Monte Carlo runs. The data for the stochastic matrix $\mathbf{Y}$ is drawn from a zero-mean normal distribution based on the covariance matrix described by (29). The figure clearly shows how the MSE of the entries of $\Delta\mathbf{a}$ in this case is

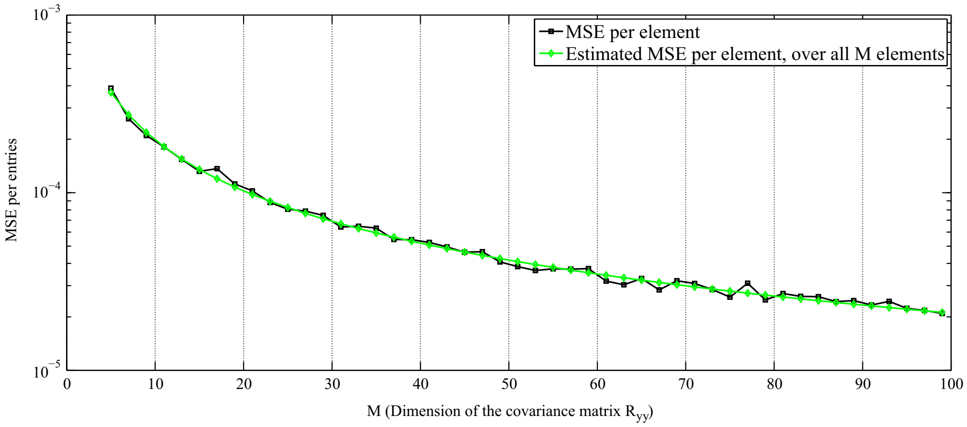

Fig. 5. MSE of the entries of $\Delta\mathbf{a}$.

reduced, as M increases. The results of the Monte Carlo simulations are compared to the theoretical results by plotting (35) as a function of M. It is observed that the theoretical prediction is very close to the simulated values. This verifies that increasing the dimension of the sample covariance matrix decreases the per-entry MSE of the steering vector estimation.

Remark: Note that for the case where also correlated noise components exist, it can be shown again that $\lim_{M\to\infty} \text{MSE}(M) = 0$.

5.3. Shortcutting the noise reduction

Until now, a cascaded scheme has been proposed in which the first step is to denoise the microphone signals by DANSE and the second step is to estimate the node-specific DOAs based on the denoised signals (Fig. 1). However it will be shown in this section that exactly the same DOA estimates can be obtained, without explicitly computing the signals in $\tilde{\mathbf{d}}_k$ and the EVD of its resulting correlation matrix (see (28)), which effectively leads to a cooperative integrated noise reduction and DOA estimation scheme, where the noise reduction is shortcut (Fig. 2).

From (23) we can define $\tilde{\mathbf{W}}_k$ as

$$\tilde{\mathbf{W}}_k = \overline{\mathbf{R}}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}^{-1}\tilde{\mathbf{v}}_k(\tilde{\sigma}_1 - \tilde{\lambda}_1)\tilde{\mathbf{v}}_k^H \tag{36}$$

where the m-th column of the $\tilde{\mathbf{W}}_k$ corresponds to the MWF to estimate the target speech signal in the m-th component of $\tilde{\mathbf{y}}_k$ at node k. Note that $\tilde{\mathbf{d}}_k$ in (27) can then be written as $\tilde{\mathbf{W}}_k^H\tilde{\mathbf{y}}_k$. Considering (28) and (36) we can write

$$\begin{aligned}\overline{\mathbf{R}}_{\tilde{\mathbf{d}}_k\tilde{\mathbf{d}}_k} &= \tilde{\mathbf{W}}_k^H\overline{\mathbf{R}}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}\tilde{\mathbf{W}}_k \\ &= \tilde{\mathbf{v}}_k(\tilde{\sigma}_1 - \tilde{\lambda}_1)^*\tilde{\mathbf{v}}_k^H\overline{\mathbf{R}}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}^{-H}\overline{\mathbf{R}}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}\overline{\mathbf{R}}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}^{-1}\tilde{\mathbf{v}}_k(\tilde{\sigma}_1 - \tilde{\lambda}_1)\tilde{\mathbf{v}}_k^H\end{aligned} \tag{37}$$

and by taking $\rho = |(\tilde{\sigma}_1 - \tilde{\lambda}_1)|^2\tilde{\mathbf{v}}_k^H\overline{\mathbf{R}}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}^{-H}\tilde{\mathbf{v}}_k$, we have

$$\overline{\mathbf{R}}_{\tilde{\mathbf{d}}_k\tilde{\mathbf{d}}_k} = \rho\tilde{\mathbf{v}}_k\tilde{\mathbf{v}}_k^H \tag{38}$$

which means that $\overline{\mathbf{R}}_{\tilde{\mathbf{d}}_k\tilde{\mathbf{d}}_k}$ is immediately a rank-1 matrix and hence the eigenvector corresponding to the largest (non-zero) eigenvalue is equal to $\tilde{\mathbf{v}}_k$, which is already available from the GEVD of $\overline{\mathbf{R}}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}$ and $\overline{\mathbf{R}}_{\tilde{\mathbf{n}}_k\tilde{\mathbf{n}}_k}$. As a result, we can shortcut the final filtering stage of the DANSE algorithm that computes the denoised signals of $\tilde{\mathbf{d}}_k$ which clearly leads to a substantial reduction in computational complexity. Note that this shortcut only holds if an EVD- or a GEVD-based rank-1 approximation of $\overline{\mathbf{R}}_{\tilde{\mathbf{s}}_k\tilde{\mathbf{s}}_k}$ is used in the local MWFs of the DANSE algorithm.

Remark: As mentioned in Section 4, the use of a GEVD for the rank-1 approximation of $\overline{\mathbf{R}}_{\mathbf{s}_k\mathbf{s}_k}$ in the MWF often improves the noise reduction performance compared to the use of an EVD [26,27]. However, in view of the DOA estimation, there is an additional benefit in using a GEVD rather than an EVD. If a random scaling is applied to one of the z_k-signals in $\tilde{\mathbf{y}}_k$, this results in a similar scaling of the corresponding row and column of the correlation matrices $\overline{\mathbf{R}}_{\tilde{\mathbf{y}}_k\tilde{\mathbf{y}}_k}$, $\overline{\mathbf{R}}_{\tilde{\mathbf{s}}_k\tilde{\mathbf{s}}_k}$ and $\overline{\mathbf{R}}_{\tilde{\mathbf{n}}_k\tilde{\mathbf{n}}_k}$. This scaling then actually changes the eigenvectors of $\overline{\mathbf{R}}_{\tilde{\mathbf{s}}_k\tilde{\mathbf{s}}_k}$ and therefore also a steering vector estimate based on the EVD. This is undesired, i.e., a simple scaling of the fused microphone signals in one node should not have any effect on the steering vector estimate (and the resulting DOA estimate) in other nodes. It can be shown that the GEVD does not have this effect, i.e., the scaling of a z_k-signal only affects the component in the generalized eigenvectors corresponding to the scaled signal, while the other components remain the same up to a common scaling. As a result, the local steering vector estimate is never affected, as the scaled component is not part of it (remember that the steering vector only consists of the components in the generalized eigenvector that correspond to the microphone signals, and not to the z_k-signals).

6. Simulation results

6.1. Evaluation aspects

To demonstrate the effectiveness of the proposed cooperative node-specific DOA estimation, it will be compared with the centralized case and the case where each node performs a local noise-reduction and DOA estimation on its own (the ‘isolated’ case). Moreover, we consider an approach where each node k merely broadcasts one of its raw microphone signals to the other nodes (instead of the z_k-signal defined in the DANSE) and where these signals are then directly used as additional inputs to the local MWFs, followed by the subspace-based DOA estimation in each node. This is similar to the DANSE-based node-specific DOA estimation, but it relies on a suboptimal cooperative noise reduction scheme instead of the (optimal) DANSE algorithm. This approach does not only result in a reduced noise reduction performance,[5] but it also results in a slightly worse DOA estimation performance, as will be demonstrated with the simulations.

All experiments are performed in a simulated cubic room with dimensions $5^m \times 5^m \times 5^m$ and with wall reflection coefficients $\beta = 0.2$ using the image method [31]. First we consider a WASN with four nodes ($K=4$), each having a ULA with 3 microphones ($M_k=3$ for each node k and $M=12$), with a single target speech source placed at the center of the room (The acoustic scenario is depicted in Fig. 6). In Section 6.4, we will also consider the multi-source case. We perform Monte-Carlo simulations (using different speech signals in each run), in a room. Four localized multi-talker noise sources are placed in the room with equal noise power, which is varied to manipulate the input SNR. We use a sampling frequency of $f_s = 16$ kHz, a Hann-windowed DFT with size $L=512$ and with 50% Hann-window overlaps. An ideal VAD is used to exclude the effect of VAD errors. The target speech source produces short English sentences with a silence period between each two consecutive sentences. Sensor noise and all other spatially uncorrelated noise sources are modeled as uncorrelated white Gaussian noise with 20% of the power of the target speech signal as observed at the microphones. We simulate DANSE in batch mode which means that the required correlation matrices are estimated over the full

[5] This follows from the fact that the DANSE algorithm always results in an optimal noise reduction [23].

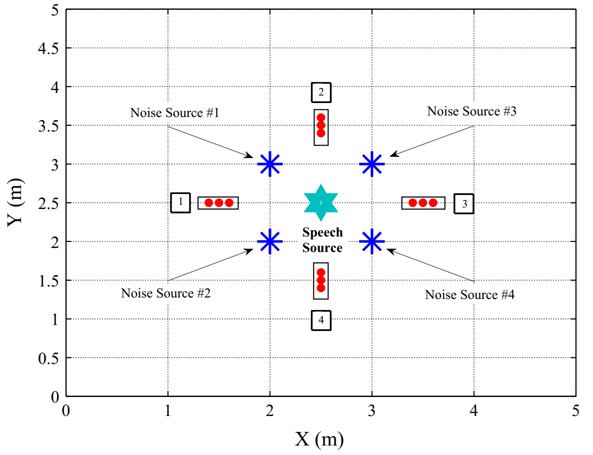

Fig. 6. Acoustic scenario.

signal length in each iteration and the DOA estimation is performed after convergence of the DANSE algorithm.

All results in Sections 6.2 and 6.3 are averaged over 56 independent Monte Carlo runs and over all nodes. All the figures are plotted as a function of the input SNR at a reference node k which is defined in the time-domain as the power ratio of the speech and noise component in the first microphone signal of node k, i.e.,

$$iSNR_k = \frac{E\{|s_{k1}|^2\}}{E\{|n_{k1}|^2\}} = \frac{E\{|s_{k1}|^2\}}{E\{|n^m_{k1}+n^p_{k1}|^2\}} \tag{39}$$

where n^m_{k1} and n^p_{k1} are assumed to be the signal components corresponding to the uncorrelated sensor noise and the localized noise sources, respectively.

Since the actual subspace estimation performance plays an import role in all the subspace-based DOA estimation algorithms, a proper assessment of the subspace estimation can give a better insight into the merits of our proposed technique. Since only the relative phase differences between the microphones are important, we should define a measure for the subspace estimation performance that is independent of phase or sign ambiguities. To achieve this goal, we again consider the overlapping doublets structure for a ULA as explained in Section 3.2, and we compute the difference between the phase of the two doublets' estimated steering vectors and of the true array manifold vector at each frequency bin. For node k, this yields (see (11) and Fig. 3)[6]

$$e_k = \frac{1}{\Omega}\sum_{\omega} |\angle\overline{\Psi} - \angle\psi| \tag{40}$$

where $\Omega = L/2+1$ (50% Hann-window overlaps) is the number of the DFT bins.

The performance of the node-specific DOA estimation by using the DANSE algorithm is evaluated with the subspace-based DOA estimation algorithms outlined earlier in Section 3, i.e., MUSIC and ESPRIT.

6.2. Scenario 1

We first assume a symmetric scenario in which the true value of the DOA in each node is chosen to be 0° (this corresponds to so-called end-fire arrays). Due to this symmetry, iSNR is identical at each node and hence all the nodes are equally important. To change the iSNRs, we change the power of the four localized noise sources uniformly and identically, while keeping the uncorrelated noise level on the microphones unchanged. Fig. 7 compares the subspace estimation performance based on the measure (40) as a function of iSNR when averaged over all frequency bins and all MC runs. As can be seen, DANSE achieves a better subspace estimation than the isolated approach and the approach where nodes merely broadcast their first microphone signal to the other nodes (note that the plot corresponding to the proposed cooperative distributed DOA estimation method based on DANSE almost fully overlaps with the plot for the centralized method).

Fig. 8 shows the averaged absolute values of the DOA estimation errors using ESPRIT. Moreover, the results for DOA estimation with MUSIC are illustrated in Fig. 9. As can be seen in these figures, there is a clear benefit in terms of DOA estimation when there is cooperation between the nodes, compared to the isolated approach. If this cooperation is based on the z_k-signals of the DANSE algorithm, the performance of the DOA estimation is closer to the centralized performance compared to the approach where the nodes merely broadcast one microphone signal. The results also show that MUSIC is more robust than ESPRIT. This comes with a significantly higher computational complexity due to the exhaustive searches, which might be impractical in WASNs with limited power supply.

It is noted that the obtained results are better than those in a preliminary study [32]. This is partly due to the fact that the present simulations use a GEVD, rather than an EVD based rank-1 approximation (as used in [32]). As explained in the remark in Section 5.3, such a GEVD-based approach is more robust and less dependent on the differences in signal power between the fused microphone signals that are exchanged between the nodes.

6.3. Scenario 2

In order to further investigate the effectiveness of the proposed cooperative node-specific DOA estimation, we now rotate the microphone array in each node independently with a random angle in each MC run. Fig. 10 compares the subspace estimation performance based on the measure in (40) as a function of iSNR when averaged over all the frequency bins and all the MC runs with different true DOAs. Moreover, Figs. 11 and 12 show the averaged absolute values of the DOA estimation errors in degrees for ESPRIT and MUSIC, respectively.

Again, we observe that cooperation between nodes results in a better subspace estimation and hence a better DOA estimation.

[6] It could be necessary to add or subtract multiples of 2π to ensure that the absolute phase e_k is in the interval $[0, \pi]$.

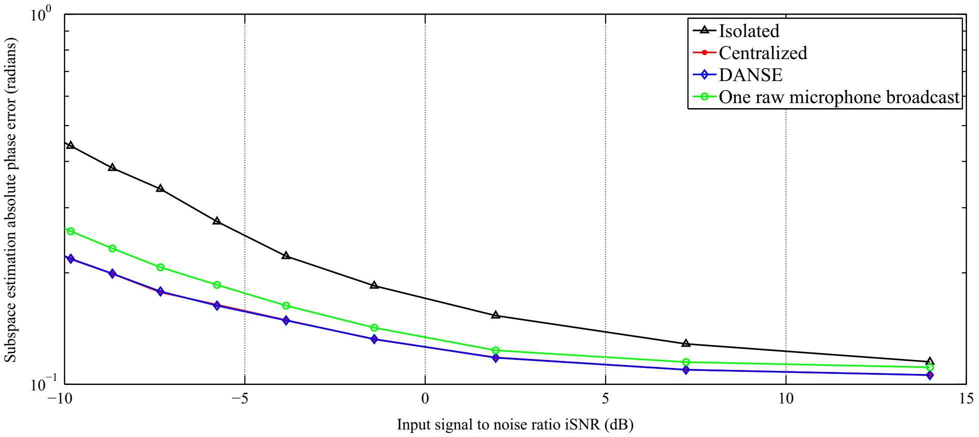

Fig. 7. Subspace estimation performance.

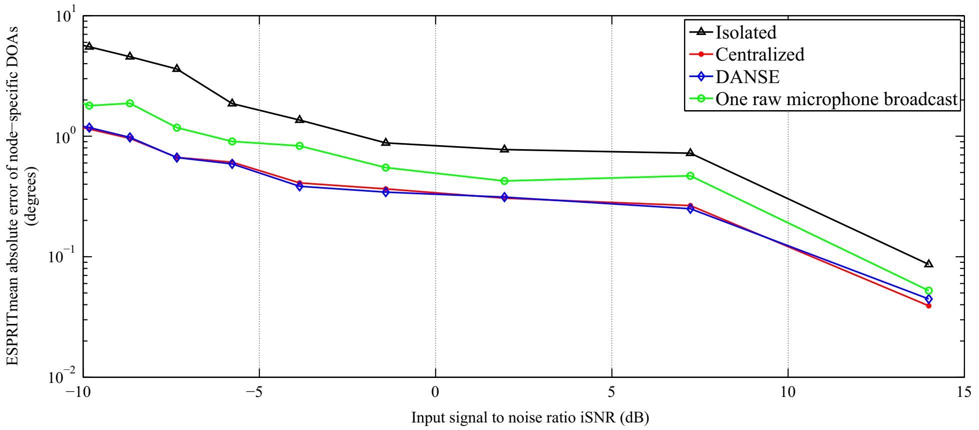

Fig. 8. Absolute DOA estimation errors based on wideband ESPRIT.

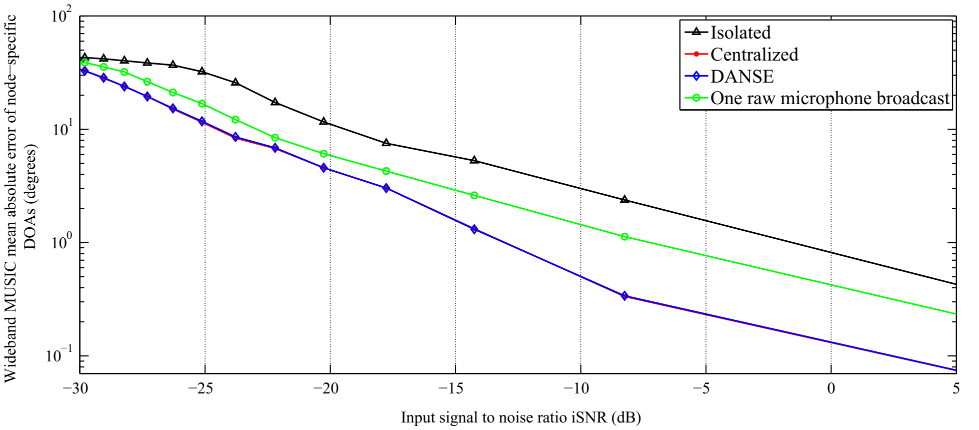

Fig. 9. Absolute DOA estimation errors based on wideband MUSIC.

6.4. *Scenario 3: multi-source case*

In this section we consider a multi-source scenario with two target speech sources, while three localized multi-talker noise sources contaminate the captured target speech signals. The multi-source acoustic arrangement is depicted in Fig. 13. To change the iSNRs, we again increase the power of the three localized noise sources uniformly and identically, while keeping the uncorrelated noise level on the microphones unchanged. Moreover, to have a noise subspace with a higher dimension, here we consider $M_k=5$ for each node k, hence $M=20$. While each source consists of a different speech signal, there are some silent intervals for both sources to let nodes estimate the noise statistics. The true DOAs of

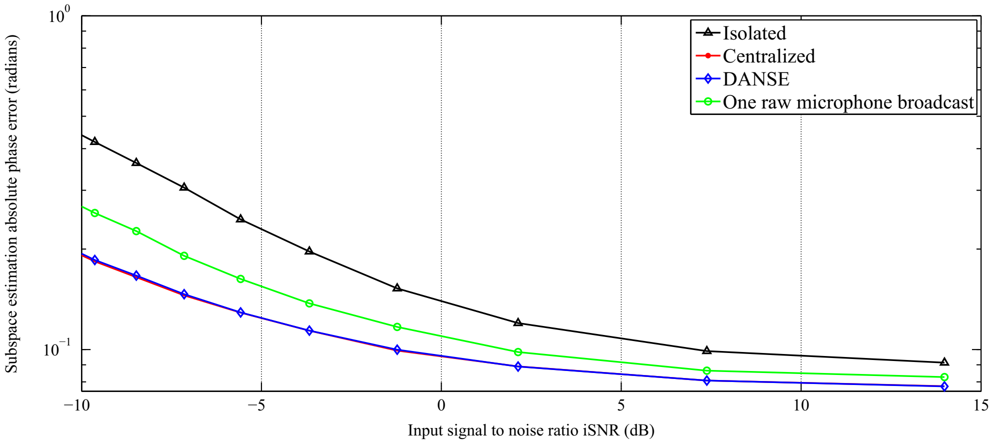

Fig. 10. Subspace estimation performance with random rotation of the arrays.

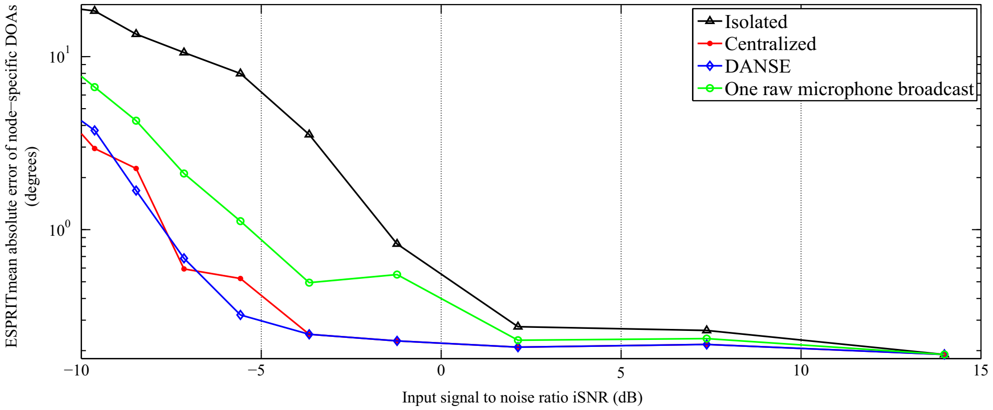

Fig. 11. Absolute DOA estimation errors based on wideband ESPRIT with random rotation of the arrays.

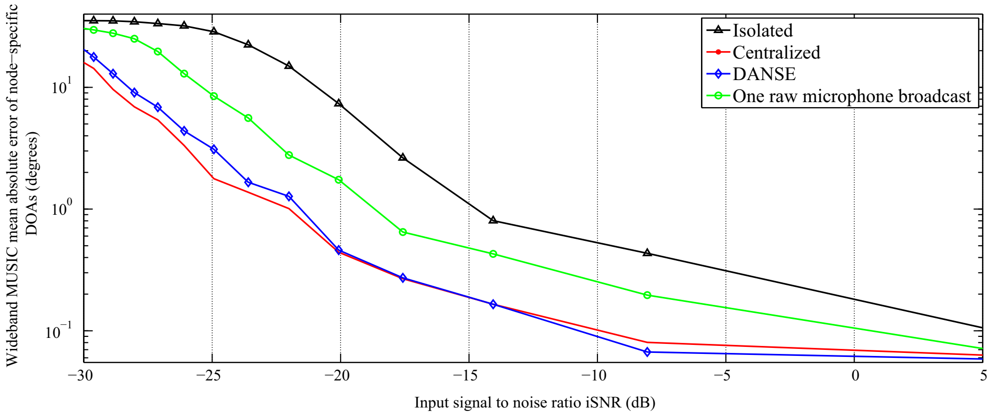

Fig. 12. Absolute DOA estimation errors based on wideband MUSIC with random rotation of the arrays.

$K=4$ nodes with respect to the first and the second (see Fig. 13) target speech sources are [45° 71° 71° 45°] and [108° 135° 135° 108°], respectively. The simulations are performed by averaging first over absolute estimation errors of 28 Monte Carlo runs, and then over the two estimated DOAs at each node k, and finally over all $K=4$ nodes. It has been shown in [23] that for multi-source cases, DANSE converges to the centralized MWF performance when each node k broadcasts $\min\{S, M_k\}$ fused signal to the other nodes (resulting in a per-node compression factor of $\max\{M_k/S, 1\}$ for the data to be sent). Therefore, and since here $S=2$, DANSE compresses the 5-channel microphone signal of each node k into a 2-channel signal that is broadcast to the other nodes. Figs. 14 and 15 show the resulting DOA estimation error when ESPRIT and wideband MUSIC are used, respectively. Although the performance plots are now slightly different, the general trend remains the same, i.e., these figures verify that in the general case of estimating DOAs for multiple target speech sources, cooperation between nodes again leads to significantly better DOA estimation. As a reference, the case where nodes merely exchange two of their raw microphone signals is also considered. As can be clearly seen, the performance of the DANSE case is again substantially closer to the centralized case.

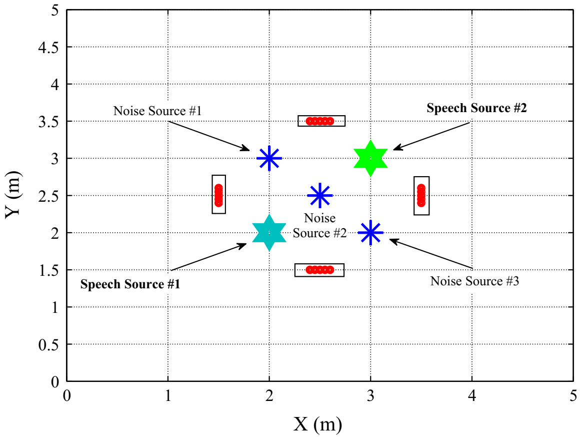

Fig. 13. Multi-source acoustic scenario.

7. Conclusion

In this paper, we have studied a cooperative node-specific DOA estimation algorithm in a fully connected WASN in a noisy environment where the position of the nodes as well as the relative geometry or coherence models between them is unknown. The DANSE algorithm is employed as a preprocessing step to first denoise all the WASN microphone signals in a distributed fashion where in each node GEVD-based MWFs are applied for the filtering process. In addition to achieving an optimal noise reduction, the fused microphone signals that are exchanged between the nodes in DANSE are also exploited to improve the node-specific subspace estimation in the DOA estimation algorithm, resulting in a cooperative integrated noise reduction and DOA estimation where the computational cost can be reduced by shortcutting the DANSE final filtering stage. An incoherent wideband version of MUSIC and ESPRIT has been employed to show

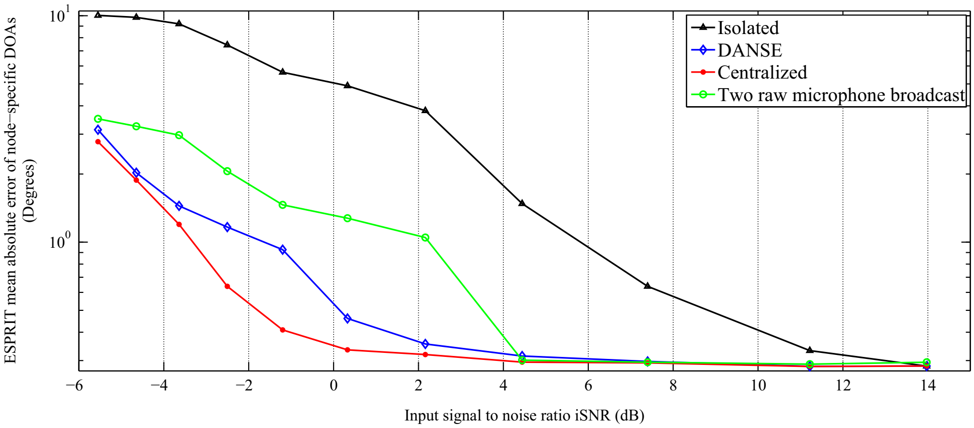

Fig. 14. Absolute Multi-source DOA estimation errors based on wideband ESPRIT with random rotation of the arrays.

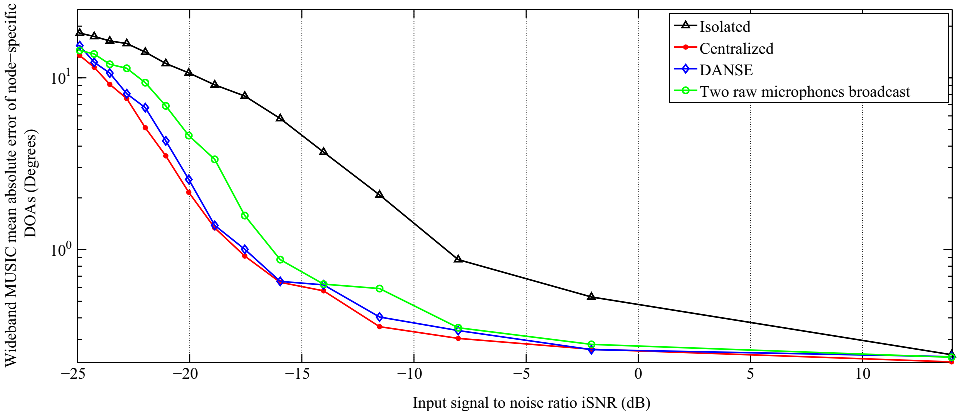

Fig. 15. Absolute Multi-source DOA estimation errors based on wideband MUSIC with random rotation of the arrays.

the effectiveness of the proposed cooperative node-specific DOA estimation. Monte-Carlo simulations have demonstrated that the cooperation between the nodes indeed improves the subspace estimation, and therefore also the multi-source DOA estimation in each node.

Acknowledgments

A conference precursor of this paper has been published in the proceedings of EUSIPCO-2013 [32]. 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), Concerted Research Action GOA-MaNet, 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.0763.12 'Wireless Acoustic Sensor Networks for Extended Auditory Communication', and 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. The work of A. Bertrand was supported by a Postdoctoral Fellowship of the Research Foundation - Flanders (FWO). The scientific responsibility is assumed by its authors.

References

[1] M. Brandstein, D. Ward, Microphone Arrays: Signal Processing Techniques and Applications, Springer-Verlag, Berlin Heidelberg, 2001.

[2] I. Tashev, Sound Capture and Processing: Practical Approaches, Wiley, West Sussex, UK, 2009.

[3] A. Bertrand, Applications and trends in wireless acoustic sensor networks: a signal processing perspective, in: Proceedings of the IEEE Symposium on Communications and Vehicular Technology (SCVT), Ghent, Belgium, 2011.

[4] H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach, Signal Process. Mag. IEEE 13 (4) (1996) 67–94, http://dx.doi.org/10.1109/79.526899.

[5] S. Pillai, C. Burrus, Array signal processing, Signal Processing and Digital Filtering, Springer-Verlag, New York, 1989.

[6] R. Schmidt, Multiple emitter location and signal parameter estimation, in: IEEE Transactions on Antennas and Propagation, vol. 34, 1986, pp. 276–280.

[7] P. Stoica, K. Sharman, Maximum likelihood methods for direction-of-arrival estimation, IEEE Trans. Acoust. Speech Signal Process. 38 (7) (1990) 1132–1143, http://dx.doi.org/10.1109/29.57542.

[8] M. Viberg, B. Ottersten, T. Kailath, Detection and estimation in sensor arrays using weighted subspace fitting, IEEE Trans. Signal Process. 39 (11) (1991) 2436–2449, http://dx.doi.org/10.1109/78.97999.

[9] R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust. Speech Signal Process. 37 (7) (1989) 984–995, http://dx.doi.org/10.1109/29.32276.

[10] H. Wang, M. Kaveh, Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources, IEEE Trans. Acoust. Speech Signal Process. 33 (4) (1985) 823–831, http://dx.doi.org/10.1109/TASSP.1985.1164667.

[11] J. Evans, D. Sun, J. Johnson, Application of Advanced Signal Processing Techniques to Angle of Arrival Estimation in ATC Navigation and Surveillance Systems, Technical report, DTIC Document, 1982.

[12] F. Sellone, Robust auto-focusing wideband DOA estimation, Signal Process. 86 (1) (2006) 17–37, http://dx.doi.org/10.1016/j.sigpro.2005.04.009.

[13] S. Chandran, M. Ibrahim, DOA estimation of wide-band signals based on time-frequency analysis, IEEE J. Ocean. Eng. 24 (1) (1999) 116–121, http://dx.doi.org/10.1109/48.740160.

[14] M. Wax, T.-J. Shan, T. Kailath, Spatio-temporal spectral analysis by eigenstructure methods, IEEE Trans. Acoust. Speech Signal Process. 32 (4) (1984) 817–827, http://dx.doi.org/10.1109/TASSP.1984.1164400.

[15] T. Pham, B. Sadler, Wideband Array Processing Algorithms for Acoustic Tracking of Ground Vehicles, US Army Research Laboratory, Report. Available at ⟨http://www.arl.army.mil/sedd/acoustics/reports.htm⟩.

[16] M. Pesavento, A. Gershman, K.M. Wong, Direction finding in partly calibrated sensor arrays composed of multiple subarrays, IEEE Trans. Signal Process. 50 (9) (2002) 2103–2115, http://dx.doi.org/10.1109/TSP.2002.801929.

[17] A. Swindlehurst, B. Ottersten, R. Roy, T. Kailath, Multiple invariance ESPRIT, IEEE Trans. Signal Process. 40 (4) (1992) 867–881, http://dx.doi.org/10.1109/78.127959.

[18] A. Swindlehurst, P. Stoica, M. Jansson, Exploiting arrays with multiple invariances using MUSIC and MODE, IEEE Trans. Signal Process. 49 (11) (2001) 2511–2521, http://dx.doi.org/10.1109/78.960398.

[19] J. Chen, K. Yao, R. Hudson, Source localization and beamforming, Signal Process. Mag. IEEE 19 (2) (2002) 30–39.

[20] R. Kozick, B.M. Sadler, Near-field localization of acoustic sources with imperfect spatial coherence, distributed processing, and low communication bandwidth, in: Aerospace/Defense Sensing, Simulation, and Controls, International Society for Optics and Photonics, 2001, pp. 52–63.

[21] G. Müller, M. Möser, Handbook of Engineering Acoustics, Springer-Verlag, Berlin Heidelberg, 2013.

[22] S. Doclo, M. Moonen, GSVD-based optimal filtering for single and multimicrophone speech enhancement, IEEE Trans. Signal Process. 50 (2002) 2230–2244.

[23] A. Bertrand, M. Moonen, Distributed adaptive node-specific signal estimation in fully connected sensor networks part I: sequential node updating, IEEE Trans. Signal Process. 58 (2010) 5277–5291.

[24] A. Bertrand, M. Moonen, Distributed adaptive node-specific signal estimation in fully connected sensor networks part II: simultaneous and asynchronous node updating, IEEE Trans. Signal Process. 58 (2010) 5292–5306.

[25] M. Azimi-Sadjadi, A. Pezeshki, N. Roseveare, Wideband DOA estimation algorithms for multiple moving sources using unattended acoustic sensors, IEEE Trans. Aerosp. Electron. Syst. 44 (4) (2008) 1585–1599.

[26] R. Serizel, M. Moonen, B. Van Dijk, J. Wouters, Rank-1 approximation based multichannel Wiener filtering algorithms for noise reduction in cochlear implants, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013.

[27] R. Serizel, M. Moonen, B. Van Dijk, J. Wouters, Low-rank approximation based multichannel Wiener filtering algorithms for noise reduction in cochlear implants, IEEE Transactions on Audio, Speech and Language Processing 22 (4) (2014) 785–799.

[28] C.F.V.L. Gene, H. Golub, Matrix Computations, 3rd ed. John Hopkins Univ. Press, Baltimore, MD, 1996.

[29] A. Bertrand, M. Moonen, Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology, IEEE Trans. Signal Process. 59 (5) (2011) 2196–2210, http://dx.doi.org/10.1109/TSP.2011.2108290.

[30] B. Friedlander, A.J. Weiss, On the second-order statistics of the eigenvectors of sample covariance matrices, IEEE Trans. Signal Process. 46 (11) (1998) 3136–3139.

[31] J.B. Allen, D.A. Berkley, Image method for efficiently simulating small-room acoustics, Journal of the Acoustical Society of America 65 (4) (1979) 943–950.

[32] A. Hassani, A. Bertrand, M. Moonen, Distributed node-specific direction-of-arrival estimation in wireless acoustic sensor networks, in: Proceedings of the European Signal Processing Conference (EUSIPCO), 2013.