Distributed Multiple Constraints Generalized Sidelobe Canceler for Fully Connected Wireless Acoustic Sensor Networks

Shmulik Markovich-Golan[1], *Student Member, IEEE,* Sharon Gannot[1], *Senior Member, IEEE* and Israel Cohen[2], *Senior Member, IEEE*

Abstract

This paper proposes a distributed multiple constraints generalized sidelobe canceler (GSC) for speech enhancement in an N-node fully connected wireless acoustic sensor network (WASN) comprising $\bar{M}$ microphones. Our algorithm is designed to operate in reverberant environments with P constrained speakers (including both desired and competing speakers). Rather than broadcasting $\bar{M}$ microphone signals, a significant communication bandwidth reduction is obtained by performing *local* beamforming at the nodes, and utilizing only $N+P$ transmission channels. Each node processes its own microphone signals together with the transmitted signals. The GSC-form implementation, by separating the constraints and the minimization, enables the adaptation of the BF during speech-absent time segments, and relaxes the requirement of other distributed LCMV based algorithms to re-estimate the sources RTFs after each iteration. We provide a full convergence proof of the proposed structure to the centralized GSC-beamformer (BF). An extensive experimental study of both narrowband and (wideband) speech signals verifies the theoretical analysis.

[1] Shmulik Markovich-Golan and Sharon Gannot are with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan 56000, Israel, e-mail: shmulik.markovich@gmail.com; sharon.gannot@biu.ac.il

[2] Israel Cohen is with the Department of Electrical Engineering, Technion, Technion City, Haifa 32000, Israel, e-mail: icohen@ee.technion.ac.il

I. INTRODUCTION

Recent advances in the fields of nano-technology and communications encourage the development of low-cost, low-power and miniaturized modules which can be incorporated in wireless sensor network (WSN) applications. A WSN comprises several nodes (WSN modules) interconnected in some manner via a wireless medium. Each node consists of one or more sensors, a processing unit and a wireless communication module allowing them to exchange data. The goal of the system is to perceive some physical phenomenon, to process it, and to yield a required result. In classical array processing systems, the sensing and the processing of the acquired data are concentrated in a single location denoted a *fusion center*. A phenomenon originating in the enclosure, results in a disturbance that propagates in space. The closer the sensors are to the origin of the phenomenon, the higher is the signal to noise ratio (SNR) of the acquired signal, resulting in lower estimation errors and better quality at the output of the signal processing procedure. The concept of the WSN is to distribute the system resources (sensors, processing units and actuators) and to provide a scalable, easy to deploy, and robust structure. The wireless interface allows for the extension of the sensing range beyond the limits of the wired fusion center systems. The distribution of the sensors in a larger volume enables a better coverage with higher SNR. For a survey on the topic of WSN please refer to [1], [2], [3], [4]. Limited power and communication bandwidth resources set bounds on the amount of data shared between nodes and necessitate developing distributed algorithms. In recent years, many contributions to the field of WASN have been introduced, circumventing the severe network constraints [5], [6], [7], [8], [9], [10], [11]. A trivial solution is obtained by utilizing only microphones local to the node without any communication link. However this solution fails to utilize the entire information from the network and hence is sub-optimal. A common scheme for distributed signal processing algorithms in WASNs comprises the following steps. First, local processing of microphone signals results in intermediate signals or estimates at each node, requiring less communication-bandwidth. Second, the results of the first step are broadcast in the WASN. Finally, a global estimate or an enhanced signal is obtained by merging all intermediate signals or estimates. Since the data available at each node is incomplete, an iterative (or time-recursive) solution becomes necessary.

Several contributions have considered using a WASN system for speech processing applications. Two main criteria are common in speech beamforming applications: the minimum mean squared error (MMSE) and the minimum variance distortionless response (MVDR). The mean squared error (MSE) between the output signal and the desired signal comprises two components, namely the distortion and the residual noise. The multi-channel Wiener filter (MWF)-BF [12], [13], [14] minimizes the MSE between the desired signal and the output signal, while the MVDR, first introduced by Capon [15], minimizes the noise power at the output signal while maintaining a distortionless response towards the desired signal, i.e., resulting in zero distortion. Er and Cantoni [16] generalized the single distortionless response to a set of linear constraints, and denoted the BF as linearly constrained minimum variance (LCMV)-BF. The speech distortion weighted (SDW)-MWF-BF, proposed by Doclo et al. [17], generalizes both BF criteria. It introduces a trade-off factor between noise reduction and distortion. It can be shown that two special cases of the SDW-MWF are the MWF-BF and, in case of a single desired speaker, the MVDR-BF.

Signals and parameters at a node which are obtained by processing its own microphone signals are denoted "local" to the node. Other signals and parameters which are obtained by processing data received from other nodes in the WASN are denoted "global". Doclo et al. [7] addressed the problem of enhancing a single desired speaker contaminated by a stationary noise. They adopted the SDW-MWF criterion and used a binaural hearing aid system comprising two apparatuses with multiple microphones in each ear.

Bertrand and Moonen [8] considered the more general case of an N node WASN and P desired sources. They allowed each node to define individual desired signals by using different weighting of the spatial components of the speech. They proposed a distributed adaptive node-specific signal estimation (DANSE)-P algorithm which necessitates transmission of P channels from each node and proved the convergence of the algorithm to the global SDW-MWF-BF. In complicated scenarios where multiple speakers exist and more control over the beampattern is desired, the LCMV-BF is a more suitable option. The linear constraints set can be designed to maintain undistorted desired speakers while mitigating competing speakers.

Adaptive formulation of the MVDR-BF was proposed by Frost [18]. Frost developed a constrained least mean squares (LMS) algorithm for the adaptation of the BF coefficients. Griffiths and Jim [19] showed that the MVDR criterion, can be equivalently described in a two branch structure, denoted GSC. This structure conveniently separates the constraining and the minimization operations. Breed and Strauss [20] further proved the equivalence between the closed-form LCMV and the GSC-form in the case of multiple constraints.

Gannot et al. [21] considered the single desired source case and suggested to implement the MVDR-BF in its GSC-form in the short time Fourier transform (STFT) domain. They also proposed to use the relative transfer function (RTF) rather than the acoustic transfer function (ATF) of the desired speaker, and proposed an applicable estimation procedure based on the non-stationarity of the speech. Markovich-Golan et al. [22] considered the multiple speakers case and proposed to use an LCMV-BF in a GSC-form. They constructed a constraints set from an estimate of the RTFs of the desired speakers and an estimate of the basis spanning the ATFs of the competing speakers and the stationary noise. In [9] the authors adopted the MVDR criterion and proposed an iterative distributed MVDR-BF for a binaural hearing aid system. Bertrand and Moonen [23] proposed a distributed LCMV algorithm, denoted linearly constrained DANSE (LC-DANSE). They considered the case of P speakers and noise picked up by microphones of an N node WASN. Assuming that each node may define the set of desired speakers differently, they proposed that the constraints matrix will be common to all nodes, whereas the desired response will be node-specific. Their proposed algorithm constructs P node-specific constraints LCMV-BFs that require each node to transmit P audio channels. A total of $N \times P$ transmission channels (the output signals of all local BFs) are required. At each iteration, each node has to re-estimate two sets of basis vectors spanning the ATFs of the desired and the interfering speakers.

Ahmed and Vorobyov [24] presented a novel technique for controlling the sidelobe level in collaborative beamforming for WSNs where nodes comprise both sensors and actuators. They considered the problem of transmitting multiple data streams from different clusters of nodes to some remote target nodes. Each cluster forms a beam pattern by properly setting the phases and amplitudes at the transmission such that the signals received at the designated target node are with equal phases and amplitudes. An efficient algorithm for controlling the inter-channel interference is based on repeatedly and randomly selecting the nodes which participate in the beamforming, and using low communication-bandwidth feedback channels from the target nodes which report the interference level that they experience.

In the current contribution we consider the case where the nodes *agree* on the classification of desired and competing speakers and share a common constraints set as well as desired responses. A distributed time-recursive version of the centralized GSC, denoted distributed GSC (DGSC), is proposed. We prove that the proposed algorithm converges to the centralized GSC. The proposed algorithm requires the transmission of only $N + P$ audio channels. In static scenarios, the RTFs of the sources need to be estimated only once at the initialization stage. The estimation procedure of the RTFs may require non-overlapping activity patterns of the speakers.

The structure of the paper is as follows. In Sec. II, the problem is formulated. In Sec. III, a closed-form and a GSC structure of the centralized LCMV-BF are presented. We show that, under certain conditions, an LCMV which operates on a transformation of the inputs is equivalent to the regular BF. In Sec. IV, we derive the DGSC algorithm. The latter is based on a specific transformation which allows to reformulate the centralized BF as a sum of local GSC-BFs. The proposed algorithm makes use of shared signals, one for each source, which are broadcast in the WASN. We give an analytical proof of the equivalence between the DGSC and the centralized GSC-BF. In Sec. V, we propose a scheme for constructing the shared signals. We compare the proposed DGSC and the LC-DANSE in Sec. VI. An extensive experimental study, which verifies the equivalence of the DGSC and the centralized GSC, is presented in Sec. VII. Conclusions are drawn in Sec. VIII.

II. Problem formulation

Consider a WASN of microphones comprised of N nodes. Denote the number of microphones in the nth node by $\bar{M}_n$. The total number of microphones is denoted $\bar{M}$ and equals

$$\bar{M} \triangleq \sum_{n=1}^{N} \bar{M}_n. \tag{1}$$

The problem is formulated in the STFT domain where k denotes the frequency index and ℓ denotes the time-frame index. The vector of signals received by the microphones of all nodes is $\bar{\mathbf{z}}(\ell, k)$. It is composed by concatenating the microphone signals of all nodes:

$$\bar{\mathbf{z}}(\ell, k) \triangleq \begin{bmatrix} \bar{\mathbf{z}}_1^T(\ell, k) & \cdots & \bar{\mathbf{z}}_N^T(\ell, k) \end{bmatrix}^T \tag{2}$$

where $\bar{\mathbf{z}}_n(\ell, k)$ is an $\bar{M}_n \times 1$ vector consisting of locally received signals at the nth node. The vector of all received signals is given by:

$$\bar{\mathbf{z}}(\ell, k) \triangleq \bar{\mathbf{H}}(\ell, k)\mathbf{s}(\ell, k) + \bar{\mathbf{v}}(\ell, k) \tag{3}$$

where

$$\mathbf{s}(\ell, k) \triangleq \begin{bmatrix} s^1(\ell, k) & \cdots & s^P(\ell, k) \end{bmatrix}^T \tag{4}$$

is a $P \times 1$ vector comprised of the speech sources, and

$$\bar{\mathbf{H}}(\ell, k) \triangleq \begin{bmatrix} \bar{\mathbf{h}}^1(\ell, k) & \cdots & \bar{\mathbf{h}}^P(\ell, k) \end{bmatrix} \tag{5}$$

is an $\bar{M} \times P$ matrix which columns are the ATFs relating the P speakers and the $\bar{M}$ microphones. The vector $\bar{\mathbf{v}}(\ell, k)$ is a vector of interfering signals picked up by the microphones. Assuming that the P speakers' signals and the noise sources are uncorrelated, the $\bar{M} \times \bar{M}$ dimensional covariance matrix of the received signals may be written as:

$$\bar{\mathbf{\Phi}}_{zz}(\ell, k) \triangleq \bar{\mathbf{H}}(\ell, k)\mathbf{\Gamma}(\ell, k)\bar{\mathbf{H}}^\dagger(\ell, k) + \bar{\mathbf{\Phi}}_{vv}(k, \ell). \tag{6}$$

where $(\bullet)^\dagger$ denotes the conjugate-transpose operator, $\mathbf{\Gamma}(\ell, k) = \operatorname{diag}\left\{\lambda^1(\ell, k), \ldots, \lambda^P(\ell, k)\right\}$ is the $P \times P$ dimensional covariance matrix of the P speech signals and $\bar{\mathbf{\Phi}}_{vv}(\ell, k)$ is the covariance matrix of the noise. Note that multiple speakers and noise sources may be simultaneously active at each frequency bin. We assume that the network is fully connected, hence any transmitted signal is available to all nodes. In cases that the network is not fully connected a hierarchial algorithm, for example based on a spanning tree of the network, can be sought. However, this is beyond the scope of the current contribution. As an example for a distributed algorithm in a partially connected WASN please refer to [25]. The locations of the speakers are assumed static, therefore their corresponding ATFs are time-invariant, and hence the frame index is omitted in $\bar{\mathbf{H}}(k)$. The algorithm is applied to each frequency bin independently. For brevity, the index k is hereafter omitted. The noise statistics is assumed to vary significantly slower than the convergence-time of the algorithm. For brevity, the index ℓ is also omitted from $\bar{\mathbf{\Phi}}_{vv}$ hereafter.

Denote the set of microphone indexes at the nth node by $\bar{\mathcal{M}}_n \triangleq \left\{m_n(1), \ldots, m_n(\bar{M}_n)\right\}$, where $\bar{M}_n \triangleq |\bar{\mathcal{M}}_n|$ and $|\bullet|$ denotes the number of elements in a set. The vector of the received signals at the nth node is given by

$$\bar{\mathbf{z}}_n(\ell) = \mathbf{T}_n^{\dagger}\bar{\mathbf{z}}(\ell) \tag{7}$$

where $\mathbf{T}_n$ is an $\bar{M} \times \bar{M}_n$ selection matrix which extracts the $\bar{M}_n$ entries that correspond to the microphone indexes of the nth node:

$$\mathbf{T}_n = \left[\begin{array}{c|c|c} \mathbf{0}_{\bar{M}_n \times \left(\sum_{n'=1}^{n-1} \bar{M}_{n'}\right)} & \mathbf{I}_{\bar{M}_n} & \mathbf{0}_{\bar{M}_n \times \left(\sum_{n'=n+1}^{N} \bar{M}_{n'}\right)} \end{array} \right]^T \tag{8}$$

and $\mathbf{I}_m$ is an $m \times m$ identity matrix.

III. An equivalent centralized LCMV-BF

In the following, the centralized LCMV-BF is formulated. We show that under certain conditions, an LCMV-BF which operates on a transformation of the inputs is equivalent to the LCMV-BF which directly processes the microphone signals. A common design relaxation of using the RTFs rather than the ATFs is formulated, and the GSC-form implementation is defined. The distributed algorithm, derived in Sec. IV, will be based on a specific transformation matrix, that will conveniently split the centralized BF into a sum of N BFs. Each of the BFs utilizes only local microphones and P shared signals, generated as a linear combination of the local microphone signals in some remote nodes. Together with the transmission of the N local BF outputs, a total of $N + P$ transmission channels is required.

The centralized LCMV-BF, denoted $\bar{\mathbf{w}}_{\text{LCMV}}$, is given by:

$$\bar{\mathbf{w}}_{\text{LCMV}} \triangleq \underset{\left\{\mathbf{w};\bar{\mathbf{H}}^{\dagger}\mathbf{w}=\mathbf{g}\right\}}{\operatorname{argmin}} \left\{\mathbf{w}^{\dagger}\bar{\mathbf{\Phi}}_{vv}\mathbf{w}\right\} \tag{9}$$

where the global constraints set is

$$\bar{\mathbf{H}}^{\dagger}\bar{\mathbf{w}} = \mathbf{g} \tag{10}$$

and $\mathbf{g}$ is a $P \times 1$ desired response vector. Typically, the desired response vector $\mathbf{g}$ is comprised of values of zeros and ones, where a value of 1 is associated with a desired speaker and a value of 0 is associated with an interfering speaker. In this case the BF is required to yield a combination of all the desired speakers while mitigating the interfering speakers and the noise. Generally, $\mathbf{g}$ can be any arbitrary $P \times 1$ vector. We assume that the ATFs are linearly independent, i.e., the column rank of the constraints matrix $\bar{\mathbf{H}}$ is P. In practice, when $\bar{M} \gg P$ the latter assumption usually holds, however, of course it is not guaranteed. In cases for which the ATFs are linearly dependent, the constraints set might consist of contradicting requirements. Hence, no solution that satisfy all constraints can be obtained. When contradicting constraints exist, the system designer has to compromise and alleviate the contradiction by reducing the number of constraints. The closed-form solution of (9) is given by Van Veen and Buckley in [12]:

$$\bar{\mathbf{w}}_{\text{LCMV}} = \bar{\mathbf{\Phi}}_{vv}^{-1}\bar{\mathbf{H}}\left(\bar{\mathbf{H}}^{\dagger}\bar{\mathbf{\Phi}}_{vv}^{-1}\bar{\mathbf{H}}\right)^{-1}\mathbf{g} \tag{11}$$

where we assume that $\bar{\mathbf{\Phi}}_{vv}$ is invertible since one of its components is a spatially white sensor noise.

The output of the LCMV-BF is given by:

$$\bar{y}_{\text{LCMV}}(\ell) = \sum_{p=1}^{P}(g^p)^* s^p(\ell) + \bar{\mathbf{w}}_{\text{LCMV}}^{\dagger}\bar{\mathbf{v}}(\ell) \tag{12}$$

where $\mathbf{g} = \begin{bmatrix} g^1 & \cdots & g^P \end{bmatrix}^T$. Note that the output comprises of the sum of the constrained sources weighted by their corresponding desired responses and a residual noise component.

Suppose that rather than $\bar{\mathbf{z}}(\ell)$, a linear transformation of the inputs is available:

$$\mathbf{z}(\ell) \triangleq \mathbf{U}^{\dagger}\bar{\mathbf{z}}(\ell) \tag{13}$$

where $\mathbf{U}^{\dagger}$ is an $M \times \bar{M}$ matrix and $M > \bar{M}$. Assuming that the column-subspace of $\mathbf{U}^{\dagger}$ is full rank, i.e., its rank is $\bar{M}$, we will show that the LCMV-BFs in the original and in the transformed domains are equivalent. Denote the following terms in the transformed domain:

$$\tilde{\mathbf{H}} \triangleq \mathbf{U}^{\dagger}\bar{\mathbf{H}} \tag{14a}$$

$$\mathbf{\Phi}_{vv} \triangleq \mathbf{U}^{\dagger}\bar{\mathbf{\Phi}}_{vv}\mathbf{U}. \tag{14b}$$

Consider the following constraints set in the transformed domain:

$$\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{w}} = \mathbf{g}. \tag{15}$$

According to the fundamental theorem of linear algebra, any $M \times 1$ BF, $\tilde{\mathbf{w}}$, in the transformed domain can be expressed as the sum of two components:

$$\tilde{\mathbf{w}} = \tilde{\mathbf{w}}^u + \tilde{\mathbf{w}}^{uc} \tag{16}$$

where $\tilde{\mathbf{w}}^u$ and $\tilde{\mathbf{w}}^{uc}$ lie in the column-subspace of $\mathbf{U}^\dagger$ and its complementary subspace, respectively. Similarly to (9), the LCMV criterion in the transformed domain is:

$$\tilde{\mathbf{w}}_{\text{LCMV}} \triangleq \underset{\left\{\tilde{\mathbf{w}};\tilde{\mathbf{H}}^\dagger\tilde{\mathbf{w}}=\mathbf{g}\right\}}{\operatorname{argmin}} \left\{\tilde{\mathbf{w}}^\dagger \mathbf{\Phi}_{vv}\tilde{\mathbf{w}}\right\}. \tag{17}$$

Note that from the definition of $\tilde{\mathbf{H}}$ and $\mathbf{\Phi}_{vv}$ in (14a) and (14b), their columns lie in the column-subspace of $\mathbf{U}^\dagger$. Hence, substituting (16) in the transformed constraint set (15) and in the minimization of the transformed LCMV-BF (17) yields:

$$\tilde{\mathbf{H}}^\dagger \tilde{\mathbf{w}}^u = \mathbf{g} \tag{18a}$$

$$\tilde{\mathbf{w}}^u_{\text{LCMV}} = \underset{\left\{\tilde{\mathbf{w}}^u;\tilde{\mathbf{H}}^\dagger\tilde{\mathbf{w}}^u=\mathbf{g}\right\}}{\operatorname{argmin}} \left\{(\tilde{\mathbf{w}}^u)^\dagger \mathbf{\Phi}_{vv}\tilde{\mathbf{w}}^u\right\} \tag{18b}$$

$$\tilde{\mathbf{w}}_{\text{LCMV}} = \tilde{\mathbf{w}}^{uc} + \tilde{\mathbf{w}}^u_{\text{LCMV}} \tag{18c}$$

where the orthogonal component $\tilde{\mathbf{w}}^{uc}$ can be chosen arbitrarily, since it affects neither the noise power at the output nor the satisfaction of the constraints set. Any $\tilde{\mathbf{w}}^u$ can be expressed as a linear combination of the columns of $\mathbf{U}^\dagger$:

$$\tilde{\mathbf{w}}^u \triangleq \mathbf{U}^\dagger \boldsymbol{\omega} \tag{19}$$

where $\boldsymbol{\omega}$ is an $\bar{M} \times 1$ vector.

Substituting (19) in (18a),(18b),(18c), $\tilde{\mathbf{w}}_{\text{LCMV}}$ becomes:

$$\begin{aligned}\tilde{\mathbf{w}}_{\text{LCMV}} &= \tilde{\mathbf{w}}^{uc} \\ &+ \mathbf{U}^\dagger \underset{\left\{\boldsymbol{\omega};\left(\mathbf{U}\tilde{\mathbf{H}}\right)^\dagger\boldsymbol{\omega}=\mathbf{g}\right\}}{\operatorname{argmin}} \left\{\boldsymbol{\omega}^\dagger \mathbf{U}\mathbf{\Phi}_{vv}\mathbf{U}^\dagger\boldsymbol{\omega}\right\}.\end{aligned} \tag{20}$$

Note that

$$\mathbf{U}\mathbf{\Phi}_{vv}\mathbf{U}^\dagger = \mathbf{U}\mathbf{U}^\dagger\bar{\mathbf{\Phi}}_{vv}\mathbf{U}\mathbf{U}^\dagger \tag{21}$$

is a full-rank $\bar{M} \times \bar{M}$ matrix since both $\mathbf{U}\mathbf{U}^\dagger$ and $\bar{\mathbf{\Phi}}_{vv}$ are $\bar{M} \times \bar{M}$ dimensional rank-$\bar{M}$ matrices. Hence, similarly to (11), the closed-form LCMV-BF of (20) in the transformed domain equals:

$$\begin{aligned}\tilde{\mathbf{w}}_{\text{LCMV}} =& \tilde{\mathbf{w}}^{uc} + \mathbf{U}^\dagger \left(\mathbf{U}\mathbf{\Phi}_{vv}\mathbf{U}^\dagger \right)^{-1} \\ & \times \mathbf{U}\tilde{\mathbf{H}} \left(\left(\mathbf{U}\tilde{\mathbf{H}} \right)^\dagger \left(\mathbf{U}\mathbf{\Phi}_{vv}\mathbf{U}^\dagger \right)^{-1} \mathbf{U}\tilde{\mathbf{H}} \right)^{-1} \mathbf{g}. \end{aligned} \tag{22}$$

Substituting (14a),(14b) and (11) in (22) yields

$$\tilde{\mathbf{w}}_{\text{LCMV}} = \mathbf{U}^\dagger \left(\mathbf{U}\mathbf{U}^\dagger \right)^{-1} \bar{\mathbf{w}}_{\text{LCMV}} + \tilde{\mathbf{w}}^{uc}, \tag{23}$$

where we also used the invertibility of $\mathbf{U}\mathbf{U}^\dagger$. It can be easily deduced that the BFs in the original and transformed domains are equivalent as their outputs coincide:

$$\tilde{\mathbf{w}}_{\text{LCMV}}^\dagger \mathbf{z}(\ell) = \bar{\mathbf{w}}_{\text{LCMV}}^\dagger \bar{\mathbf{z}}(\ell). \tag{24}$$

In practice the ATFs of the speakers are unknown, and difficult to estimate. A practical solution can be obtained by replacing the sources in (12) with filtered versions thereof [21], [22], [26], [27]. Let $h_{\text{ref}}^p;\ p = 1, \ldots, P$ be such filters. The RTF of the pth source in the transformed domain is defined as:

$$\mathbf{h}^p \triangleq \frac{\tilde{\mathbf{h}}^p}{h_{\text{ref}}^p}. \tag{25}$$

The filters $h_{\text{ref}}^p;\ p = 1, \ldots, P$ will be determined in Sec. IV. Note that these procedures may require non-overlapping activity patterns of the speakers.

Define the transformed ATF and RTF matrices of dimensions $M \times P$, respectively:

$$\tilde{\mathbf{H}} \triangleq \begin{bmatrix} \tilde{\mathbf{h}}^1 & \cdots & \tilde{\mathbf{h}}^P \end{bmatrix} \tag{26a}$$

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{h}^1 & \cdots & \mathbf{h}^P \end{bmatrix}. \tag{26b}$$

The modified constraints set is finally given by substituting $\tilde{\mathbf{H}}$ by $\mathbf{H}$ in (10):

$$\mathbf{H}^\dagger \mathbf{w}_{\text{LCMV}} = \mathbf{g}. \tag{27}$$

The modified centralized LCMV-BF (in the transformed domain), which satisfies the modified constraints set in (27), is denoted by $\mathbf{w}_{\text{LCMV}}$ and is given in closed-form, similarly to (22):

$$\begin{aligned}\mathbf{w}_{\text{LCMV}} =& \mathbf{w}^{uc} + \mathbf{U}^\dagger \left(\mathbf{U}\mathbf{\Phi}_{vv}\mathbf{U}^\dagger \right)^{-1} \\ & \times \mathbf{U}\mathbf{H} \left((\mathbf{U}\mathbf{H})^\dagger \left(\mathbf{U}\mathbf{\Phi}_{vv}\mathbf{U}^\dagger \right)^{-1} \mathbf{U}\mathbf{H} \right)^{-1} \mathbf{g}. \end{aligned} \tag{28}$$

where $\mathbf{w}^{uc}$ is an arbitrary vector lying in the null-subspace of the column-subspace of $\mathbf{U}^\dagger$. Similarly to (18b),(18c) we identify that the component of $\mathbf{w}_{\text{LCMV}}$ which lies in the column-subspace of $\mathbf{U}^\dagger$ is:

$$\mathbf{w}^u_{\text{LCMV}} = \mathbf{U}^\dagger \left(\mathbf{U}\boldsymbol{\Phi}_{vv}\mathbf{U}^\dagger\right)^{-1} \times \mathbf{U}\mathbf{H}\left((\mathbf{U}\mathbf{H})^\dagger \left(\mathbf{U}\boldsymbol{\Phi}_{vv}\mathbf{U}^\dagger\right)^{-1}\mathbf{U}\mathbf{H}\right)^{-1}\mathbf{g}. \tag{29}$$

The GSC-form implementation of (29), denoted centralized GSC-BF [19], [21], is obtained by splitting $\mathbf{w}^u$ into two components:

$$\mathbf{w}^u_{\text{LCMV}} \triangleq \mathbf{q}_{\text{GSC}} - \mathbf{B}_{\text{GSC}}\mathbf{f}_{\text{GSC}}. \tag{30}$$

Both $\mathbf{q}_{\text{GSC}}$ and the columns of $\mathbf{B}_{\text{GSC}}$ lie in the column-subspace of $\mathbf{U}^\dagger$. The vector $\mathbf{q}_{\text{GSC}}$, denoted fixed beamformer (FBF), lies in the column-subspace of $\mathbf{H}$. $\mathbf{q}_{\text{GSC}}$ is responsible for maintaining the modified constraints set (27), and equals:

$$\mathbf{q}_{\text{GSC}} = \mathbf{H}\left(\mathbf{H}^\dagger\mathbf{H}\right)^{-1}\mathbf{g}. \tag{31}$$

The blocking matrix (BM) matrix $\mathbf{B}_{\text{GSC}}$ blocks the RTFs of the constrained speakers. Explicitly,

$$\mathbf{B}_{\text{GSC}}{}^\dagger\mathbf{H} = \mathbf{0}. \tag{32}$$

Since the ranks of $\mathbf{U}^\dagger$ and $\mathbf{H}$ are $\bar{M}$ and P, respectively, the rank of $\mathbf{B}_{\text{GSC}}$ is $\bar{M} - P$ and its dimensions are $M \times \left(\bar{M} - P\right)$. The BM is not unique and can be obtained in several ways, for example, as suggested in [12], [28], by applying the singular value decomposition (SVD). To construct the BM, the SVD is applied to the $\bar{M} \times P$ matrix $\mathbf{U}\mathbf{H}$, rather than $\mathbf{H}$, and then projected to the transformed domain. Using this procedure a $M \times (\bar{M} - P)$ BM is obtained. Denote the noise canceler (NC) by an $(\bar{M} - P) \times 1$ vector $\mathbf{f}_{\text{GSC}}$. According to [12] it equals:

$$\mathbf{f}_{\text{GSC}} = \left(\mathbf{B}_{\text{GSC}}{}^\dagger\boldsymbol{\Phi}_{vv}\mathbf{B}_{\text{GSC}}\right)^{-1}\mathbf{B}_{\text{GSC}}{}^\dagger\boldsymbol{\Phi}_{vv}\mathbf{q}_{\text{GSC}}. \tag{33}$$

Note that the invertibility of $\mathbf{B}_{\text{GSC}}{}^\dagger\boldsymbol{\Phi}_{vv}\mathbf{B}_{\text{GSC}}$ is guaranteed by the definition (14b) and by the BM construction procedure above.

To enable the construction of the DGSC in Sec. IV, an extended GSC-structure is proposed:

$$\mathbf{w} \triangleq \mathbf{q} - \mathbf{B}\mathbf{f} \tag{34}$$

where the regular GSC components, $\mathbf{q}_{\text{GSC}}$, $\mathbf{B}_{\text{GSC}}$ and $\mathbf{f}_{\text{GSC}}$, are replaced by:

$$\mathbf{q} = \mathbf{q}_{\text{GSC}} + \mathbf{B}\mathbf{a} + \mathbf{q}^{uc} \tag{35a}$$

$$\mathbf{B} = \mathbf{B}_{\text{GSC}} + \mathbf{B}^{uc} \tag{35b}$$

$$\mathbf{f} = \left(\mathbf{B}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{B}\right)^{-1}\mathbf{B}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{q}. \tag{35c}$$

Here, the regular FBF $\mathbf{q}_{\text{GSC}}$ is extended by the vectors $\mathbf{B}\mathbf{a}$ and $\mathbf{q}^{uc}$, and the regular BM $\mathbf{B}_{\text{GSC}}$ is extended by the matrix $\mathbf{B}^{uc}$. The extensions $\mathbf{q}^{uc}$ and $\mathbf{B}^{uc}$ lie in the columns null-subspace of the matrix $\mathbf{U}^{\dagger}$, and $\mathbf{B}\mathbf{a}$ lies in null-subspace of $\mathbf{H}$. For any choice of $\mathbf{a}$, $\mathbf{q}^{uc}$, $\mathbf{B}^{uc}$ the modified constraints set (27) is maintained. Note that the regular GSC can be obtained as a special case of (35a), (35b) and (35c) by setting $\mathbf{a} = \mathbf{0}$, $\mathbf{q}^{uc} = \mathbf{0}$ and $\mathbf{B}^{uc} = \mathbf{0}$. As will be seen in the sequel, the introduction of $\mathbf{a} \neq \mathbf{0}$, $\mathbf{q}^{uc} \neq \mathbf{0}$ and $\mathbf{B}^{uc} \neq \mathbf{0}$ will enable us to derive a distributed version of the GSC.

Now, we show that:

$$\mathbf{w}^{\dagger}\mathbf{z}(\ell) = \mathbf{w}_{\text{LCMV}}^{\dagger}\mathbf{z}(\ell) \tag{36}$$

i.e., that $\mathbf{w}$ and $\mathbf{w}_{\text{LCMV}}$ are equivalent. By substituting (35a), (35b), (35c) in (34), it is evident that:

$$\begin{aligned}
\mathbf{w} &= \mathbf{q}_{\text{GSC}} + \mathbf{B}\mathbf{a} + \mathbf{q}^{uc} - \mathbf{B}\left(\mathbf{B}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{B}\right)^{-1}\mathbf{B}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{q} \\
&\overset{(33)}{=} \mathbf{q}_{\text{GSC}} - \mathbf{B}_{\text{GSC}}\mathbf{f}_{\text{GSC}} \\
&\quad + \mathbf{q}^{uc} - \mathbf{B}^{uc}\left(\mathbf{B}_{\text{GSC}}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{B}_{\text{GSC}}\right)^{-1}\mathbf{B}_{\text{GSC}}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{q}_{\text{GSC}} \\
&\overset{(29)}{=} \mathbf{w}_{\text{LCMV}}^{u} + \mathbf{w}^{uc}
\end{aligned} \tag{37}$$

where $\mathbf{w}^{uc}$ is identified as:

$$\mathbf{w}^{uc} = \mathbf{q}^{uc} - \mathbf{B}^{uc}\left(\mathbf{B}_{\text{GSC}}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{B}_{\text{GSC}}\right)^{-1}\mathbf{B}_{\text{GSC}}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{q}_{\text{GSC}}. \tag{38}$$

This concludes the proof of the equivalence between the extended and the regular GSC-structures.

The output signal of the proposed GSC-structure is given by:

$$\begin{aligned}
y(\ell) &\triangleq \mathbf{w}^{\dagger}\mathbf{z}(\ell) \\
&= y^{\text{FBF}}(\ell) - y^{\text{NC}}(\ell)
\end{aligned} \tag{39}$$

where $y^{\text{FBF}}(\ell)$ and $y^{\text{NC}}(\ell)$ are the outputs of the upper and lower branches of the GSC, respectively:

$$y^{\text{FBF}}(\ell) \triangleq \mathbf{q}^{\dagger}\mathbf{z}(\ell) \tag{40a}$$

$$y^{\text{NC}}(\ell) \triangleq \mathbf{f}^{\dagger}\mathbf{u}(\ell) \tag{40b}$$

$$\mathbf{u}(\ell) \triangleq \mathbf{B}^{\dagger}\mathbf{z}(\ell) \tag{40c}$$

and $\mathbf{u}(\ell)$ are the noise reference signals at the output of the BM. Substituting the constraints set of (27) in (39) yields:

$$y\left(\ell\right)=\sum_{p=1}^{P}g_{p}^{*}h_{\text{ref}}^{p}s^{p}\left(\ell\right)+\mathbf{w}^{\dagger}\mathbf{v}\left(\ell\right). \tag{41}$$

Note that the output of the GSC in the transformed domain and, by equivalence, the LCMV in the original domain, is comprised of a summation of filtered versions of the P sources and a residual noise component. It is interesting to compare the different combinations of the constrained sources at the output of the regular LCMV-BF (12) and the extended GSC-BF (41).

In conclusion, applying a transformation $\mathbf{U}^{\dagger}$ that preserves the rank-$\bar{M}$ signal subspace, guarantees the equivalence between the LCMV-BFs in the original and the transformed domains. Furthermore, an equivalent extended GSC structure exists in the transformed domain. Its optimality can be guaranteed by designing a FBF (35a) which satisfies the transformed constraints set (27), and by designing a BM (35b) with $\bar{M}-P$ linearly independent noise references.

In the following section we propose a specific transformation $\mathbf{U}$ which enables the construction of a distributed version of the extended GSC-BF.

IV. DGSC

A recursive distributed version of the GSC-BF is now proposed. We present a specific transformation matrix $\mathbf{U}$ which conveniently splits the centralized GSC into a sum of N GSC-BFs, denoted $\mathbf{w}_n$ for $n\in\{1,\ldots,N\}$, operating in each of the WASN nodes. The proposed transformation matrix consists of N sub-matrices:

$$\mathbf{U}\triangleq\begin{bmatrix}\mathbf{U}_1 & \cdots & \mathbf{U}_N\end{bmatrix} \tag{42}$$

where the *transformed* inputs of the nth node are constructed by

$$\mathbf{z}_n(\ell)\triangleq\mathbf{U}_n^{\dagger}\bar{\mathbf{z}}(\ell) \tag{43}$$

and the concatenation of all transformed inputs yields:

$$\mathbf{z}(\ell)\triangleq\begin{bmatrix}\mathbf{z}_1^T(\ell) & \cdots & \mathbf{z}_N^T(\ell)\end{bmatrix}^T. \tag{44}$$

Note that $\mathbf{U}_n$ is an $\bar{M}\times M_n$ matrix and the corresponding transformed input $\mathbf{z}_n(\ell)$ is an $M_n\times 1$ vector for $n=1,\ldots,N$. The sub-matrices $\mathbf{U}_n;\ n=1,\ldots,N$ will be later defined.

The transformed inputs of each node will comprise all of its local microphone signals and a subset of the P shared signals. With the proposed transformation each node has at least P input signals, allowing for P constraints to be maintained locally, without unnecessary sacrificing degrees of freedom, as will be shown in the following sub-sections. In this section, the selection of the P shared signals is arbitrary, and should only satisfy linear independence. We will elaborate on this matter in Sec. IV-C. A specific and simple selection of the P shared signals is given in Sec. V. The N outputs of the GSC-BFs, denoted $y_n(\ell)$ for $n = 1, \ldots, N$, and the P shared signals are transmitted in the WASN, where:

$$y_n(\ell) \triangleq \mathbf{w}_n^\dagger \mathbf{z}_n(\ell) \tag{45}$$

and $\mathbf{w}_n$ is the GSC-BF at the nth node. Hence, a total of $N + P$ transmission channels are required by the algorithm. These channels effectively extend the number of available microphones at each node and should be continuously broadcast (also after the algorithm has converged). Note that for a node n that comprises a single microphone, i.e., $\bar{M}_n = 1$, no communication-bandwidth reduction is obtained, since the single microphone signal is transmitted. The global GSC-BF is given by augmenting the N nodes' BFs:

$$\mathbf{w} \triangleq \begin{bmatrix} \mathbf{w}_1^\dagger & \cdots & \mathbf{w}_N^\dagger \end{bmatrix}^\dagger. \tag{46}$$

The final output of the algorithm is obtained by substituting (44), (45) and (46) in (39):

$$\begin{aligned} y(\ell) &= \sum_{n=1}^{N} \mathbf{w}_n^\dagger \mathbf{z}_n(\ell) \\ &= \sum_{n=1}^{N} y_n(\ell). \end{aligned} \tag{47}$$

The GSC-BF at the nth node is given by:

$$\mathbf{w}_n \triangleq \mathbf{q}_n - \mathbf{B}_n \mathbf{f}_n \tag{48}$$

where $\mathbf{q}_n$, $\mathbf{B}_n$ and $\mathbf{f}_n$ are the FBF, BM and NC at each node. Substituting (48) in (47), the output of the algorithm can be restated as:

$$y(\ell) = \sum_{n=1}^{N} \left(\mathbf{q}_n - \mathbf{B}_n \mathbf{f}_n\right)^\dagger \mathbf{z}_n(\ell). \tag{49}$$

Considering (49), we identify the global components of the GSC-BF (34) as a concatenation of $\mathbf{q}_n$ and $\mathbf{f}_n$ for $n \in \{1, \ldots, N\}$, respectively:

$$\mathbf{q} \triangleq \begin{bmatrix} \mathbf{q}_1^T & \cdots & \mathbf{q}_N^T \end{bmatrix}^T \tag{50a}$$

$$\mathbf{f} \triangleq \begin{bmatrix} \mathbf{f}_1^T & \cdots & \mathbf{f}_N^T \end{bmatrix}^T. \tag{50b}$$

The global BM, $\mathbf{B}$, is constructed as a block-diagonal matrix with N blocks:

$$\mathbf{B} \triangleq \text{blkdiag}\left\{ \begin{array}{ccc} \mathbf{B}_1, & \cdots, & \mathbf{B}_N \end{array} \right\}. \tag{51}$$

Similarly to the notation in (39), (40a), (40b), (40c) for the global GSC, the outputs of the upper and lower branches, and the noise references at the nth node, are defined as:

$$y_n(\ell) \triangleq y_n^{\text{FBF}}(\ell) - y_n^{\text{NC}}(\ell) \tag{52a}$$

$$y_n^{\text{FBF}}(\ell) \triangleq \mathbf{q}_n^\dagger \mathbf{z}_n(\ell) \tag{52b}$$

$$y_n^{\text{NC}}(\ell) \triangleq \mathbf{f}_n^\dagger \mathbf{u}_n(\ell) \tag{52c}$$

$$\mathbf{u}_n(\ell) \triangleq \mathbf{B}_n^\dagger \mathbf{z}_n(\ell). \tag{52d}$$

The global noise references vector is given by augmenting the noise reference signals of all nodes:

$$\mathbf{u}(\ell) \triangleq \left[\begin{array}{ccc} \mathbf{u}_1^T(\ell) & \cdots & \mathbf{u}_N^T(\ell) \end{array} \right]^T. \tag{53}$$

A proper selection of P shared signals ensures that the number of noise references at the output of the global BM is $\bar{M}-P$, and hence satisfies the requirement that $\mathbf{B}^\dagger \mathbf{\Phi}_{vv} \mathbf{B}$ is a full-rank $(\bar{M}-P)\times(\bar{M}-P)$ matrix.

In the following, we prove analytically that the proposed DGSC converges to the centralized GSC. In Sec. IV-A we propose a proper transformation matrix $\mathbf{U}$, that will allow us to split the BF into the structure defined by (49). We show that the proposed transformation matrix preserves the rank-$\bar{M}$ signals subspace, as required for the equivalence shown in Sec. III. The design of the FBF, the BM, and the NC of the DGSC is presented in Secs. IV-B,IV-C,IV-D. This structure is shown to satisfy the requirements of Sec. III.

A. The transformation matrix

In the following, we define some notations for formulating the DGSC. The node that transmits the shared signal of the pth speaker is denoted as the "owner" of the pth source. In Sec. V we describe the procedure for selecting the owners of each of the P signals[1], and for generating the shared signals. Denote by $\chi(p)$ the index of the node which is the owner of the pth source. The shared signals are denoted by $r^p(\ell);\ p = 1, \ldots, P$. Consider the pth shared signal, corresponding to the pth source. Assume that the

[1]A node can be the owner of several sources.

pth source is owned by the nth node, i.e., $\chi(p) = n$. We suggest to construct the shared signal as:

$$\begin{aligned} r^p(\ell) &\triangleq (\mathbf{d}_n^p)^\dagger \, \bar{\mathbf{z}}_n(\ell) \\ &= (\mathbf{d}_n^p)^\dagger \, \mathbf{T}_n^\dagger \left(\sum_{p=1}^{P} \bar{\mathbf{h}}^p s^p(\ell) + \bar{\mathbf{v}}(\ell) \right) \end{aligned} \tag{54}$$

where $\mathbf{d}_n^p$ is an $\bar{M}_n \times 1$ "local" BF that processes only the microphone signals of the nth node. A specific choice of the BFs $\mathbf{d}_n^p$ for $p = 1, \ldots, P$ and $n = \chi(p)$ will be defined in Sec. V.

Denote by $\mathcal{P}_n \triangleq \left\{ \begin{array}{ccc} p_n(1), & \cdots, & p_n(P_n) \end{array} \right\}$ the set of sources owned by the nth node, and by $P_n \triangleq |\mathcal{P}_n|$ the number of sources owned by the nth node. The shared signals generated by the nth node, are defined in a vector notation by the $P_n \times 1$ vector:

$$\mathbf{r}_n(\ell) \triangleq \left[\begin{array}{ccc} r^{p_n(1)}(\ell) & \cdots & r^{p_n(P_n)}(\ell) \end{array} \right]^T \tag{55}$$

$$= \mathbf{D}_n^\dagger \bar{\mathbf{z}}_n(\ell) \tag{56}$$

where

$$\mathbf{D}_n \triangleq \left[\begin{array}{ccc} \mathbf{d}_n^{p_n(1)} & \cdots & \mathbf{d}_n^{p_n(P_n)} \end{array} \right]. \tag{57}$$

The $\bar{M}_n \times P_n$ dimensional matrix $\mathbf{D}_n$ should be properly constructed to have a rank P_n. As each source is exclusively owned by a single node

$$P = \sum_{n=1}^{N} P_n. \tag{58}$$

The $P \times 1$ vector of all shared signals is constructed by augmenting the contributions of all nodes:

$$\mathbf{r}(\ell) \triangleq \left[\begin{array}{ccc} \mathbf{r}_1^T(\ell) & \cdots & \mathbf{r}_N^T(\ell) \end{array} \right]^T. \tag{59}$$

Note, that some of the nodes may own no sources. For instance, suppose that the n'th node does not own any source. In that case, $P_{n'} = 0$ and the corresponding vector of shared signals $\mathbf{r}_{n'}(\ell)$ will be empty.

The set of indexes of the P speakers is denoted by $\mathcal{P} \triangleq \{1, \ldots, P\}$. Denote the set of shared signals that the nth node receives as $\dot{\mathcal{P}}_n$. It comprises the indexes of all sources except the self-owned sources:

$$\begin{aligned} \dot{\mathcal{P}}_n &\triangleq \mathcal{P} \backslash \mathcal{P}_n \\ &= \left\{ \begin{array}{ccc} \dot{p}_n(1) & \cdots & \dot{p}_n(\dot{P}_n) \end{array} \right\} \end{aligned} \tag{60}$$

where \ denotes the set subtraction operation and $\dot{P}_n = |\dot{\mathcal{P}}_n|$. The $\dot{P}_n \times 1$ vector of shared signals received by the nth node is denoted by:

$$\dot{\mathbf{r}}_n^T(\ell) \triangleq \\ \left[\begin{array}{cccccc} \mathbf{r}_1^T(\ell) & \cdots & \mathbf{r}_{n-1}^T(\ell) & \mathbf{r}_{n+1}^T(\ell) & \cdots & \mathbf{r}_N^T(\ell) \end{array} \right]. \tag{61}$$

As previously defined in (43), the signals available for processing at the nth node are denoted by $\mathbf{z}_n(\ell)$, an $M_n \times 1$ vector:

$$\mathbf{z}_n(\ell) \triangleq \mathbf{U}_n^\dagger \bar{\mathbf{z}}(\ell)$$

where

$$\mathbf{U}_n \triangleq \left[\begin{array}{cc} \mathbf{T}_n & \dot{\mathbf{T}}_n \end{array} \right] \tag{62a}$$

$$\dot{\mathbf{T}}_n \triangleq \left[\begin{array}{ccc} \mathbf{T}_1\mathbf{D}_1 & \cdots & \mathbf{T}_{n-1}\mathbf{D}_{n-1} \end{array} \right. \\ \left. \begin{array}{ccc} \mathbf{T}_{n+1}\mathbf{D}_{n+1} & \cdots & \mathbf{T}_N\mathbf{D}_N \end{array} \right]. \tag{62b}$$

From (62a), the number of transformed input signals at the nth node is given by:

$$M_n = \bar{M}_n + \dot{P}_n. \tag{63}$$

Note that $\mathbf{T}_n$ and $\mathbf{T}_{n'}$ $\forall n \neq n'$ are linearly independent, since they comprise different microphones. Now, since the rank of $\mathbf{D}_n$ in (57) is P_n, it follows that the rank of $\mathbf{T}_n\mathbf{D}_n$ is also P_n. Hence, we argue that the rank of $\dot{\mathbf{T}}_n$ is $\sum_{n' \neq n} P_{n'} = \dot{P}_n$. A similar argument can be applied to $\mathbf{U}_n$. Constructed as a concatenation of $\mathbf{T}_n$ and $\dot{\mathbf{T}}_n$, its rank equals M_n.

We designate the pth shared signal, $r^p(\ell)$, as the reference microphone for the pth source RTF (25). We identify the acoustic transfer function (TF) relating the pth source and the pth shared signal (54) as:

$$h_{\text{ref}}^p \triangleq (\mathbf{d}_n^p)^\dagger \, \mathbf{T}_n^\dagger \bar{\mathbf{h}}^p. \tag{64}$$

Now, the pth RTF (25) can be defined with respect to the pth shared signal. Considerations for constructing $\mathbf{d}_n^p;\ p = 1, \ldots, P$ will be discussed in Sec. V.

The proposed $\bar{M} \times M$ dimensional transformation matrix is finally given by:

$$\mathbf{U} \triangleq \left[\begin{array}{ccc} \mathbf{U}_1 & \cdots & \mathbf{U}_N \end{array} \right] \tag{65}$$

where we note that

$$M = \sum_{n=1}^{N} M_n = \sum_{n=1}^{N} \left(\bar{M}_n + \dot{P}_n \right) \\ = \bar{M} + (N-1)\,P. \tag{66}$$

It can be easily shown that the rank of the column-subspace of $\mathbf{U}$ is $\bar{M}$, since a permutation of $\begin{bmatrix} \mathbf{T}_1 & \cdots & \mathbf{T}_N \end{bmatrix} = \mathbf{I}_{\bar{M}}$ is its sub-matrix. Hence, $\mathbf{U}$ is a valid transformation matrix, rendering $\mathbf{w}$ and $\bar{\mathbf{w}}_{\text{LCMV}}$ equivalent (24).

According to (43) and (62a), the transformed input vector in the nth node is the M_n-dimensional vector:

$$\mathbf{z}_n(\ell) = \begin{bmatrix} \bar{\mathbf{z}}_n^T(\ell) & \dot{\mathbf{r}}_n^T(\ell) \end{bmatrix}^T \tag{67}$$

where the received shared signals at the nth node are given by the $\dot{P}_n \times 1$ vector $\dot{\mathbf{r}}_n(\ell) \triangleq \dot{\mathbf{T}}_n^\dagger \bar{\mathbf{z}}(\ell)$.

Examining (3) and (43), the transformed inputs vector of the nth node is given by:

$$\mathbf{z}_n(\ell) = \tilde{\mathbf{H}}_n \mathbf{s}(\ell) + \mathbf{v}_n(\ell) \tag{68}$$

where $\tilde{\mathbf{H}}_n = \mathbf{U}_n^\dagger \bar{\mathbf{H}}$ is an $M_n \times P$ matrix and $\mathbf{v}_n(\ell) = \mathbf{U}_n^\dagger \bar{\mathbf{v}}(\ell)$.

Define

$$\mathbf{H}_r \triangleq \mathbf{D}^\dagger \bar{\mathbf{H}} \tag{69}$$

where $\mathbf{D}$ is defined as

$$\mathbf{D} \triangleq \begin{bmatrix} \mathbf{T}_1 \mathbf{D}_1 & \cdots & \mathbf{T}_N \mathbf{D}_N \end{bmatrix} \tag{70}$$

and $\mathbf{D}_n$ is defined in (57). The elements of the $P \times P$ matrix $\mathbf{H}_r$ are the ATFs relating the speakers and the P shared signals. We assume that $\mathbf{H}_r$ is a full rank matrix. The condition for the invertibility of $\mathbf{H}_r$ is given is Sec. V.

We now show that the rank of $\tilde{\mathbf{H}}_n$ is P for $n = 1, \ldots, N$. Notice that the matrix $\mathbf{H}_r$ is a column permutation of the $P \times P$ matrix:

$$\left[\begin{array}{c|c} \mathbf{D}_n & \mathbf{0}_{\bar{M}_n \times \dot{P}_n} \\ \hline \mathbf{0}_{\dot{P}_n \times P_n} & \mathbf{I}_{\dot{P}_n \times \dot{P}_n} \end{array}\right]^\dagger \tilde{\mathbf{H}}_n.$$

Since $P = \text{rank}\left\{\mathbf{H}_r\right\} \le \text{rank}\left\{\tilde{\mathbf{H}}_n\right\} \le P$, we conclude that $\text{rank}\left\{\tilde{\mathbf{H}}_n\right\} = P$.

Determining $\mathbf{U}$ as above is instrumental for transforming, the centralized GSC-BF into a sum of N GSC-BFs in the transformed domain. The total output of the DGSC algorithm is available at each of the nodes in the WASN.

In the following sections we substitute $\tilde{\mathbf{H}}_n$ by the RTFs matrix

$$\mathbf{H}_n \triangleq \mathbf{U}_n^\dagger \mathbf{H} \tag{71}$$

for $n = 1, \ldots, N$. A block-diagram of the proposed algorithm is depicted in Fig. 1.

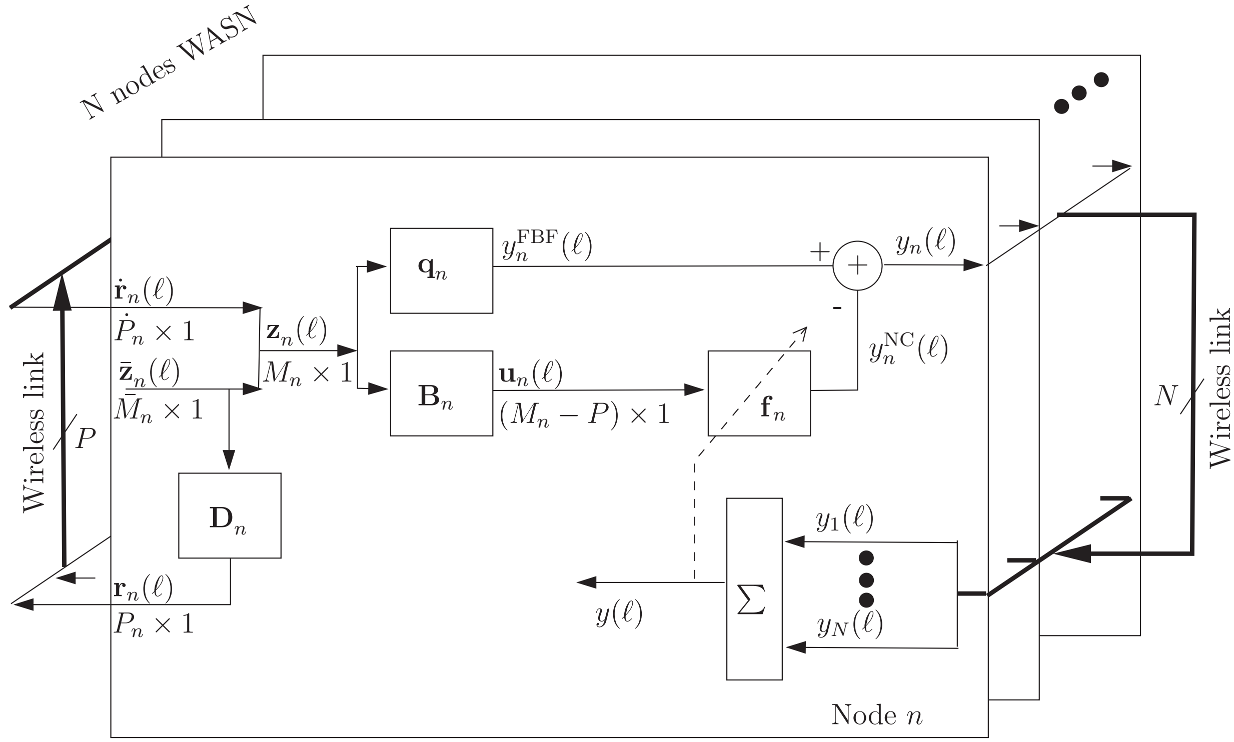


Fig. 1. The DGSC.

B. The distributed FBF

Had $\mathbf{H}$ been known to all nodes, it would have been possible to calculate the classic centralized FBF, $\mathbf{q}_{\mathrm{GSC}}$. In our case, we propose a distributed FBF consisting of a summation of local FBFs, which are calculated from the transformed RTFs at each node. Explicitly, the proposed distributed FBF at the nth node is defined as:

$$\mathbf{q}_n \triangleq \frac{1}{N}\mathbf{H}_n\left(\mathbf{H}_n^\dagger \mathbf{H}_n\right)^{-1}\mathbf{g}. \tag{72}$$

As $\mathbf{H}_n$ equals $\tilde{\mathbf{H}}_n$ up to a different column scaling, its rank equals P. Therefore, $\mathbf{H}_n^\dagger \mathbf{H}_n$ is an invertible matrix. As stated earlier, the FBF is not unique, and can have different forms with different selections of $\mathbf{a}, \mathbf{q}^{uc}, \mathbf{B}^{uc}$ in (35a),(35b). Various choices of the FBF will differ in their robustness to estimation errors.

It can be easily verified, by substituting (72) in (50a), that the global distributed FBF (35a) satisfies the global constraints set (27) since

$$\begin{aligned}\mathbf{H}^\dagger \mathbf{q} =& \left(\sum_{n=1}^{N} \mathbf{H}_n^\dagger \mathbf{q}_n\right)\\ =&\frac{1}{N}\sum_{n=1}^{N} \mathbf{g}\\ =&\mathbf{g}.\end{aligned} \tag{73}$$

This simple FBF design utilizes each of the WASN microphones and is not optimal in any sense. The robustness analysis of the proposed algorithm to estimation errors is out of the scope of the current contribution.

C. The distributed BM

As mentioned earlier, the BM is not unique, and several procedures for its construction are available. Recently, we have proposed an efficient implementation of a sparse BM [28]. Similarly to the construction of the BM in Sec. III, we propose that the nth node will construct a transformed BM by applying the SVD to $\mathbf{H}_n$, for $n = 1, \ldots, N$. The SVD of $\mathbf{H}_n$ is

$$\mathbf{H}_n = \begin{bmatrix} \mathbf{\Gamma}_n & \mathbf{B}_n \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_n \\ \mathbf{0}_{(M_n-P)\times P} \end{bmatrix} \mathbf{\Theta}_n^\dagger. \tag{74}$$

where the column-space of $\mathbf{H}_n$ is spanned by the column-space of $\mathbf{\Gamma}_n$. The null-subspace of $\mathbf{H}_n$ is spanned by the column-subspace of $\mathbf{B}_n$ and hence is an adequate BM at the nth node. Since the column rank of $\mathbf{H}_n$ is P, The dimensions of the BM at the nth node are $M_n \times (M_n - P)$, and its column rank is $M_n - P$.

Next, we prove that $\mathbf{B}$ is a valid BM. From its construction (51), it trivially blocks $\mathbf{H}$, hence, in order to complete the proof, we need to show that $\mathbf{B}^\dagger \mathbf{\Phi}_{vv} \mathbf{B}$ is of full-rank. From the definition of $\mathbf{\Phi}_{vv}$ in (14b), and since $\bar{\mathbf{\Phi}}_{vv}$ is full-rank (rank-$\bar{M}$), the latter condition is equivalent to showing that the column rank of $\mathbf{U}\mathbf{B} = \begin{bmatrix} \mathbf{U}_1\mathbf{B}_1 & \cdots & \mathbf{U}_N\mathbf{B}_N \end{bmatrix}$ is $\bar{M} - P$.

The rank of $\mathbf{U}_n$ is M_n. A one-to-one linear transformation from $\mathbf{U}_n$ to $\begin{bmatrix} \mathbf{Q}_n & \bar{\mathbf{H}} \end{bmatrix}$ exists for $n = 1, \ldots, N$ where $\mathbf{Q}_n$ is an $\bar{M} \times (M_n - P)$ matrix orthogonal to $\bar{\mathbf{H}}$. It follows that $\mathbf{U}_1, \ldots, \mathbf{U}_N$ share at least P degrees of freedom (the columns of $\bar{\mathbf{H}}$). Now, since the rank of $\mathbf{U}$ is $\bar{M}$, we conclude that $\mathbf{U}_1, \ldots, \mathbf{U}_N$ share exactly P degrees of freedom. Hence, the rank of $\begin{bmatrix} \mathbf{Q}_1 & \cdots & \mathbf{Q}_N \end{bmatrix}$ is $\bar{M} - P$. By construction, $\mathbf{Q}_n$ is an $\bar{M} \times (M_n - P)$ BM of $\bar{\mathbf{H}}$, and its $M_n - P$ outputs $\mathbf{Q}_n^\dagger \bar{\mathbf{z}}(\ell)$ are equivalent (represent the same noise signals) to the $M_n - P$ outputs $\mathbf{B}_n^\dagger \mathbf{z}_n(\ell) = \mathbf{B}_n^\dagger \mathbf{U}_n^\dagger \bar{\mathbf{z}}(\ell)$. Finally, $\mathbf{U}\mathbf{B}$ is a concatenation of the sub-matrices $\mathbf{U}_n\mathbf{B}_n$ for $n = 1, \ldots, N$. Hence, it has the same rank as the concatenation of $\mathbf{Q}_n$ for $n = 1, \ldots, N$. Based on the above discussion, it is guaranteed that $\mathbf{B}^\dagger\mathbf{\Phi}_{vv}\mathbf{B}$ is a full-rank $\left(\bar{M} - P\right) \times \left(\bar{M} - P\right)$ matrix.

D. The distributed NC

The normalized LMS (NLMS) adaptation of the global NC in [21] is given by:

$$\mathbf{f}(\ell) = \mathbf{f}(\ell - 1) + \mu \frac{\mathbf{u}(\ell) y^*(\ell)}{P_u(\ell)} \tag{75}$$

where $P_u(\ell)$ is a recursive estimator of the power of the noise reference signals, i.e., $\mathrm{E}\left\{\|\mathbf{u}(\ell)\|^2\right\}$:

$$P_u(\ell) = \rho P_u(\ell - 1) + (1 - \rho)\|\mathbf{u}(\ell)\|^2 \tag{76}$$

where ρ is a forgetting factor (typically $0.8 < \rho < 1$). Due to inevitable estimation errors, some of the speech signals might leak to the noise reference signals. In order to prevent the self-cancelation phenomenon, which is manifested in a severe speech distortion, the NC is updated according to (75) only when the speakers are inactive. A perfect voice activity detector (VAD) is assumed for this purpose. The total output of the algorithm, $y(\ell)$, is available to all nodes as the summation in (49). As clearly seen in (75), the noise reference signals at the nth node, $\mathbf{u}_n(\ell)$, only affect $\mathbf{f}_n(\ell)$. Hence, updating the NC is equivalent to N simultaneous updates of the distributed NCs $\mathbf{f}_n(\ell);\ n = 1, \ldots, N$. Explicitly, the recursive update of the distributed NC is given by:

$$\mathbf{f}_n(\ell) = \mathbf{f}_n(\ell - 1) + \mu \frac{\mathbf{u}_n(\ell) y^*(\ell)}{P_{u,n}(\ell)} \tag{77}$$

where $P_{u,n}(\ell)$ is the estimated power of the global noise reference vector $\mathrm{E}\left\{\|\mathbf{u}(\ell)\|^2\right\}$ at the nth node. We assume that the power of the local noise reference signals at the various nodes are approximately the same, i.e $\mathrm{E}\left\{\|\mathbf{u}(\ell)\|^2\right\} = \frac{\bar{M}-P}{M_n-P_n}\mathrm{E}\left\{\|\mathbf{u}_n(\ell)\|^2\right\};\ n = 1, \ldots, N$. Hence the estimated power at the nth node is:

$$P_{u,n}(\ell) = \rho P_{u,n}(\ell - 1) + (1 - \rho)\frac{\bar{M} - P}{M_n - P_n}\|\mathbf{u}_n(\ell)\|^2. \tag{78}$$

The latter assumption can be circumvented by sharing estimates of the variance of the noise reference signals $\mathbf{u}_n(\ell);\ n = 1, \ldots, N$ in the WASN. Assuming that the noise statistics is slowly varying, the latter exchange of power estimates does not consume a large bandwidth.

V. Shared signals construction

Here, we propose a simple procedure for generating the shared signals, which is based on selecting the microphones with the highest SNR for each of the sources. Since the pth shared signal is used as the reference signal in the definition of the RTF (64), and since in practice the RTF is unknown and has to be estimated, it is desired that the SNR of the pth source will be maximal. The SNR of a microphone with respect to some source p is defined as the ratio between the source power and the power of the slowly time-varying noise.

As mentioned in Sec. IV, the shared signals should satisfy that the column rank of $\mathbf{H}_r$ is P. Therefore, a microphone that was selected as the shared signal of a certain source, cannot be chosen as a shared signal for another source, or else the rank of $\mathbf{H}_r$ will be lower than P.

During the initialization of the algorithm each node sets $\mathcal{J}_n \triangleq \left\{1, \ldots, \bar{M}_n\right\}$ the index set of candidate microphones for shared signals. For each source $p \in \{1, 2, \ldots, P\}$ the following procedure is applied. First, the nth node estimates $\gamma_n^p(j);\ j \in \mathcal{J}_n$, the pth source SNR at each of its available local microphones, $m_n(j);\ j \in \mathcal{J}_n$. Each node selects the microphone with the highest SNR. The SNR and the index of the candidate microphone of the nth node are:

$$\gamma_n^p \triangleq \max_{j \in \mathcal{J}_n} \gamma_n^p(j) \tag{79a}$$

$$j_n^p \triangleq \operatorname*{argmax}_{j \in \mathcal{J}_n} \gamma_n^p\left(j\right). \tag{79b}$$

Each node shares the maximal SNR γ_n^p with the rest of the nodes.

The node n' with the maximum SNR will be declared the owner of the source p, i.e., $\chi(p) = n'$:

$$\chi(p) \triangleq \operatorname*{argmax}_{n=1,\ldots,N} \gamma_n^p. \tag{80}$$

The n'th node constructs the BF that extracts the pth shared signal

$$\mathbf{d}_{n'}^p \triangleq \left[\begin{array}{ccc} \mathbf{0}_{1\times(j_{n'}^p-1)} & 1 & \mathbf{0}_{1\times(\bar{M}_{n'}-j_{n'}^p)} \end{array}\right]^T \tag{81}$$

and removes $j_{n'}^p$ from its set of candidate microphones to own a signal

$$\mathcal{J}_{n'} = \mathcal{J}_{n'} \backslash j_{n'}^p. \tag{82}$$

This way, it is guaranteed that a single microphone will not be chosen more than once. The procedure is repeated for all sources, resulting in the entire set of shared signals. Note that some nodes may be the owners of more than a single source, and some nodes may have no ownership on sources. The proposed method is very simple, and does not require any processing for constructing the shared signals.

In practice, $\mathbf{H}_r$ is usually full-rank, however, this is not guaranteed. In case, that $\mathbf{H}_r$ is rank-deficient, a simple procedure of replacing some of the shared signals until the rank is full can be applied.

VI. A COMPARISON BETWEEN THE DGSC AND THE LC-DANSE

We compare the proposed DGSC and the LC-DANSE [23]. Both algorithms converge to the centralized LCMV-BF. The LC-DANSE implements a distributed version of the closed-form LCMV, whereas the DGSC adopts the GSC implementation of the LCMV structure. In the DGSC a common objective to all nodes, i.e., the classification of desired and competing speakers, yields a single common constraints set. A more general approach is adopted by the LC-DANSE, which allows node-specific constraint sets. In practice, this enables each node to define its own objective, i.e., a set of desired and competing speakers. The LC-DANSE is an iterative algorithm (although, the iterations can be carried out recursively over time), while the DGSC is a time-recursive algorithm. The GSC structure conveniently decouples the task of noise reduction from the task of satisfying the constraints set. Hence, allowing the adaptive noise canceler (ANC) to adjust to variations in the noise statistics. The DGSC requires $N + P$ transmission channels, whereas the LC-DANSE requires $N \times P$ transmission channels. Both algorithms, require estimates of the sources RTFs. In static scenarios, the DGSC requires a single estimate thereof, whereas in the LC-DANSE, each iteration requires additional RTF estimates. In the following section, we experimentally compare the DGSC and the LC-DANSE.

VII. EXPERIMENTAL STUDY

In order to verify the equivalence between the centralized GSC and the proposed DGSC, a comprehensive experimental study is carried out. The validity of the proposed algorithm is tested for narrowband signals in Sec. VII-A and for speech signals in Sec. VII-B. We compare the following five algorithms, namely, the centralized closed-form LCMV, the centralized GSC, a single node local GSC (arbitrarily chosen as the first node), the LC-DANSE and the proposed DGSC algorithm. The comparison criteria are noise reduction and distortion of the constrained sources. Opposed to the global BFs and the DGSC algorithms where the number of constraints can be as large as the total number of microphones in the WASN ($P \leq \bar{M}$), the local GSC is constrained to handle only scenarios where $P \leq \bar{M}_1$. The performance is averaged over multiple Monte-Carlo experiments in various scenarios.

A. Narrowband signals

A WASN comprising $N = 4$ nodes, each consisting of $\bar{M}_n = 4$ microphones was simulated. We denote by *constrained* sources, sources for which desired responses exist and are maintained with a proper linear constraints set. Furthermore, we denote by *unconstrained* sources, all interfering sources that $\bar{\mathbf{v}}(\ell, k)$ comprises. We examine a total of 28 scenarios: all combinations of $P = 1, 2, \ldots, 7$ constrained sources and $P_i = 1, 3, 5, 7$ unconstrained sources. A spatially white Gaussian sensor noise is added to the microphone signals. In each scenario (a specific selection of P and P_i) ,10 sets of source ATFs and a vector of desired responses are randomized. For each set, 10 realizations of 10^5 samples of $P + P_i$ independent identically distributed (IID) Gaussian processes are randomized. These signals serve as the constrained and unconstrained sources. Note that in the narrowband case all sources are stationary. A total of 3200 Monte-Carlo experiments are used for the comparison of the various algorithms. The SNR, the ratio between the constrained signals power and the sensors spatially white noise, is set to 30dB, and the interference to noise ratio (INR), the ratio between the unconstrained sources power and the sensors noise, is set to 25dB. The step-size of the NLMS algorithms is set to $\mu = 0.25$. The results of the LC-DANSE algorithm are measured after 10 iterations. We assume that the RTFs are known without estimation errors, hence no distortion to the constrained signals is measured for the centralized LCMV, the centralized GSC, and the DGSC for all values of $P \leq \bar{M} = 16$. For the single node GSC, there is no distortion for $P = 1, 3 \leq \bar{M}_1 = 4$, but for $P = 5, 7$, due to lack of degrees of freedom (there are only $\bar{M}_1 = 4$ beams that can be steered), distortion is inevitable. The distortion measured in the LC-DANSE is also low (< -23dB) in all scenarios. The noise reduction (NR) of the various algorithms after convergence for $P_i = 3$ versus the number of constraints, P, is depicted in Fig. 2. The figure of merit is defined as the ratio between the slowly time varying noise power at the input and at the output. As expected, the NR of the centralized GSC is about 0.35dB lower than the centralized LCMV. This is a result of using the LMS algorithm, which suffers from excess MSE. It can be mitigated by reducing the step-size μ compromising convergence rate. The NR of the proposed DGSC is 0.52dB lower than the centralized GSC (probably since longer convergence time is required), whereas the of the NR single node GSC is much lower (from 7.7dB to 47.6dB, depending on the number of constraints). The NR performance of all BFs reduces as the number of constraints increases. The convergence of the NR versus the number of samples is depicted in Fig. 3 for a scenario with $P = 5$ and $P_i = 3$. Although the proposed DGSC and the centralized GSC converge to more or less the same NR as the centralized LCMV, the convergence time of the DGSC is higher. This may result from the higher condition number, defined as the ratio of the largest and smallest eigenvalues, of the noise references covariance matrix $\mathbf{B}^{\dagger}\mathbf{\Phi}_{vv}\mathbf{B}$. Higher condition number is known to increase the convergence time [29]. For example, in the depicted scenario, the average condition number of the noise references covariance matrix of the DGSC is 6.9dB higher than of the centralized GSC. The latter phenomenon may be attributed to the vector

$\mathbf{a} \neq \mathbf{0}$ in (35a), which increases the norm of the ANC in (35c), however, this subject requires further research.

The ratio of the noise level at the output of the DGSC and the noise level at the output of the centralized GSC is given in Table I for $P = 1, 2, \ldots, 8$ and $P_i = 1, 3, 5, 7$.

TABLE I

THE RATIO OF THE NOISE LEVEL AT THE OUTPUTS OF THE DGSC AND THE CENTRALIZED GSC [DB].

$P \backslash P_i$	1	3	5	7
1	0.03	0.20	0.18	0.26
2	0.08	0.22	0.37	0.44
3	0.32	0.22	0.39	0.62
4	0.25	0.56	0.48	0.75
5	0.51	0.50	0.89	0.66
6	0.38	0.90	0.45	1.01
7	0.15	1.29	0.91	0.95
8	0.31	0.75	0.90	0.84

B. Speech signals

The performance of the various BFs is tested in a simulated room scenario, by using a room impulse response (RIR) generator [30],[31]. The dimensions of the simulated room are $4\text{m} \times 3\text{m} \times 3\text{m}$, and its reverberation time is set to $T_{60} = 300\text{ms}$. An $N = 4$ nodes WASN where each node comprises $\bar{M}_n = 2$ microphones at a distance of 5cm is set. The nodes are located at the center of each of the four walls, 10cm from the walls surface and at a height of 1.5m. A desired female speaker and a competing male speaker, are located in the room as well as two white Gaussian stationary interferences. The figures of merit of the BFs are tested by 90 Monte Carlo experiments, where in each experiment the sources locations are randomly selected, and the microphone constellation remains fixed. The room setup of one of the Monte Carlo experiments is depicted in Fig. 4. The microphone signals are sampled at a sampling rate of 8kHz. The length of the STFT window is 4096 points with 75% overlap between frames. The estimated RTFs are double sided filters 3072 coefficient long. They are estimated using the subspace method as in [22]. In the DGSC algorithm, in order to save communication-bandwidth, the signals undergo inverse STFT prior to the broadcast in the network. We use the overlap and save scheme for applying the filters in the STFT domain [32],[21]. The SNR improvement, signal to interference ratio (SIR) improvement and distortion measures of the centralized GSC, the DGSC and the single node GSC for the various Monte

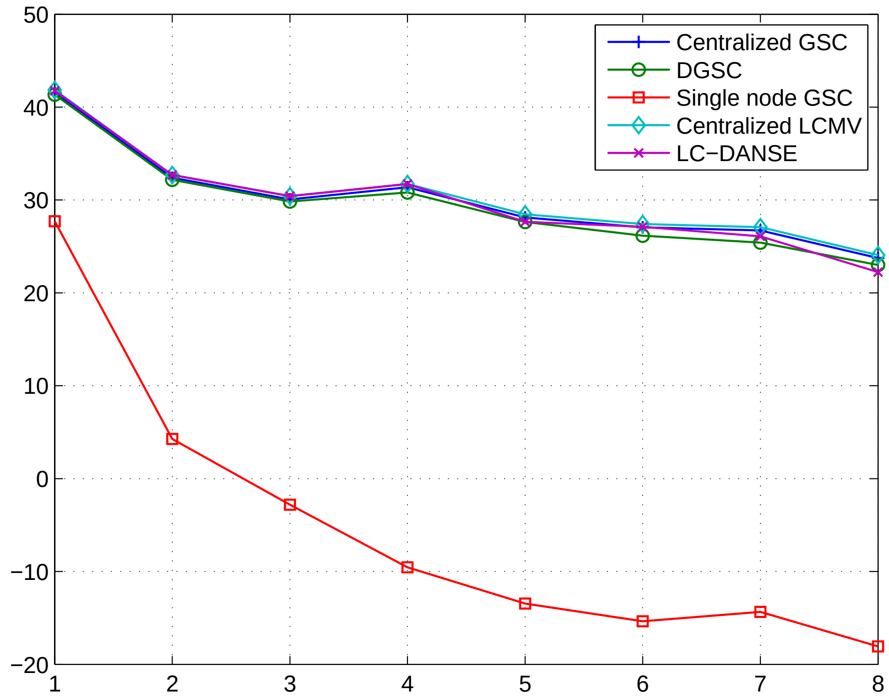


Fig. 2. The NR of the tested algorithms versus the number of constraints P, for $P_i = 3$.

Carlo experiments are depicted in Figs. 5,6,7, respectively. The SNR is the ratio between the powers of the desired speaker and the stationary noise, the SIR is the ratio between the powers of the desired speaker and the competing speaker, and the distortion is the ratio between the MSE of the desired speech at the output and the power of the desired speech signal. The SNR and the SIR at the input are set to 13dB and 0dB, respectively. It is clear from these figures that the NR values of the DGSC and the centralized GSC are equivalent, and that both outperform the single node GSC. The average figures of merit of the various algorithms is depicted in Table II. The SNR improvement of the DGSC and the centralized GSC are similar (20.1dB and 19.3dB, respectively. The slight differences may be explained as in the narrowband case), while the SNR improvement of the single node GSC is significantly lower (1.7dB). The SIR improvement and the distortion of the centralized GSC are 22.9dB and -23.0dB, respectively, whereas the corresponding measures of the DGSC are a bit worse 18.6dB and -20.3dB, respectively. This may be attributed to differences in the robustness of the BFs against RTF estimation errors (see discussion in the narrowband case). Due to the significantly lower number of microphones, the

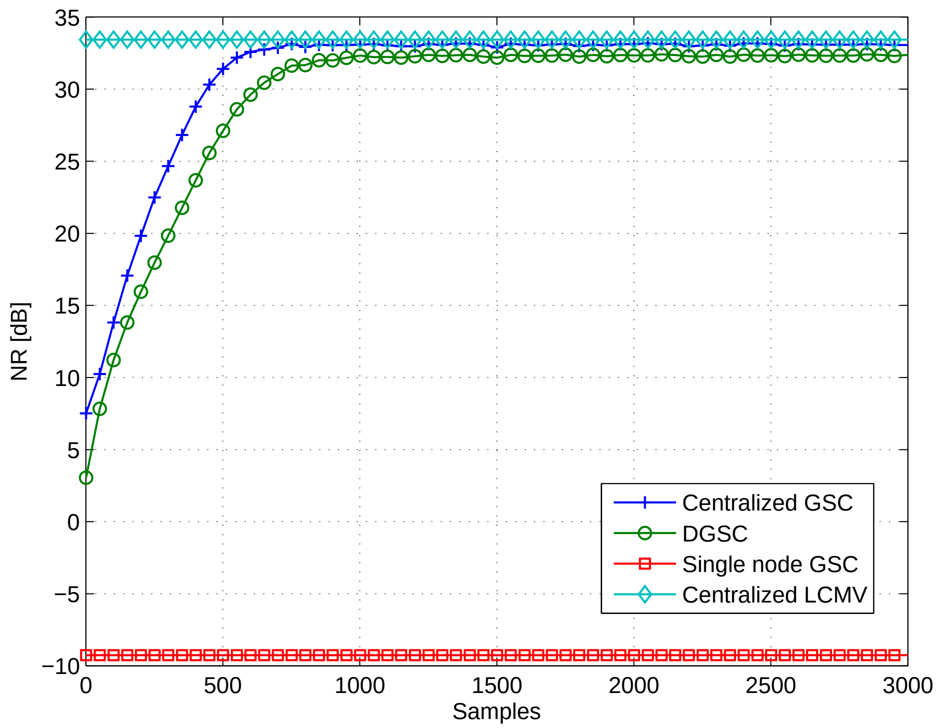


Fig. 3. The convergence of the tested algorithms versus the number of samples for $P = 5$ constraints and $P_i = 3$.

SIR improvement and distortion of the single node GSC (11.0dB and -14.1dB, respectively) are much worse than the centralized GSC. The centralized GSC and the DGSC exhibit comparable convergence behaviour as depicted in Fig. 8. Note, that the single node GSC converges much faster, but its overall performance is very poor.

Sonograms of the various components of the signal received in the first microphone, and the outputs of the centralized GSC, the DGSC and the single node GSC are depicted in Fig. 9. The equivalence of the DGSC and the centralized GSC and their superiority to the single node GSC can be deduced from the figures.

VIII. Conclusions

In this paper, we have introduced the DGSC, a novel distributed algorithm for speech enhancement in multiple speakers, noisy and reverberant environment. It is proven analytically that the proposed algorithm converges to the optimal centralized GSC-BF. The adaptive procedure of the DGSC is based on the low

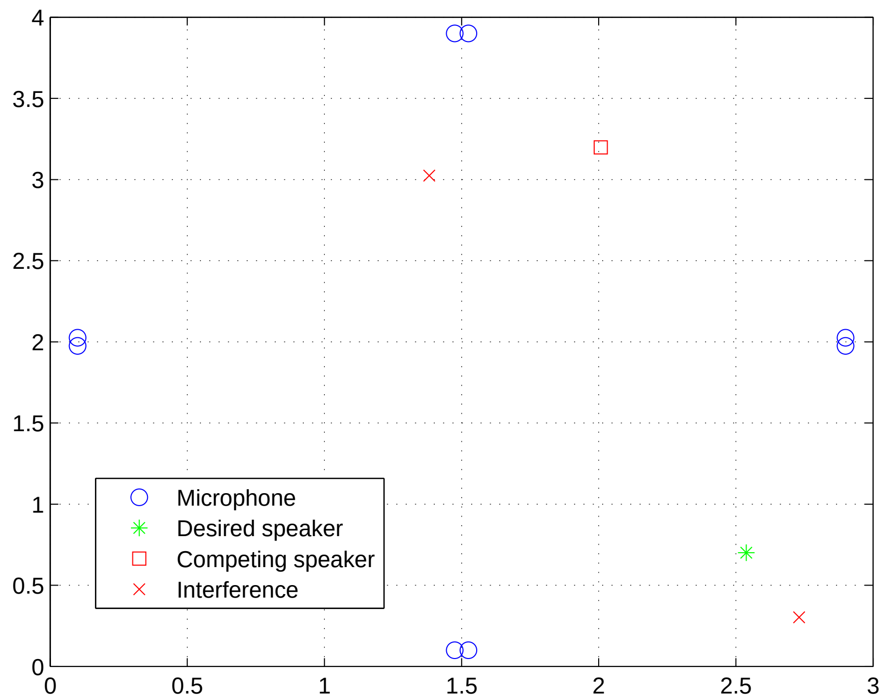


Fig. 4. The room setup of one of the Monte Carlo simulations.

TABLE II

PERFORMANCE COMPARISON OF THE CENTRALIZED GSC, THE DGSC AND THE SINGLE NODE GSC ALGORITHMS WITH SPEECH SIGNALS.

Algorithm	SNR imp. [dB]	SIR imp. [dB]	Dist. [dB]
Cent. GSC	20.1	22.9	-23.0
DGSC	19.3	18.6	-20.6
1 node GSC	1.7	11.0	-14.1

complexity, time recursive NLMS algorithm. A common P linear constraints set, comprising the speakers' ATF, is shared by all nodes in the network. The algorithm requires $N+P$ transmission channels. The GSC structure splits the BF into two components. The first component lies in the constraints (speakers) subspace and the second component lies in its corresponding null-space. The constraints subspace component of the DGSC is determined at the initialization phase of the algorithm where P shared signals are constructed

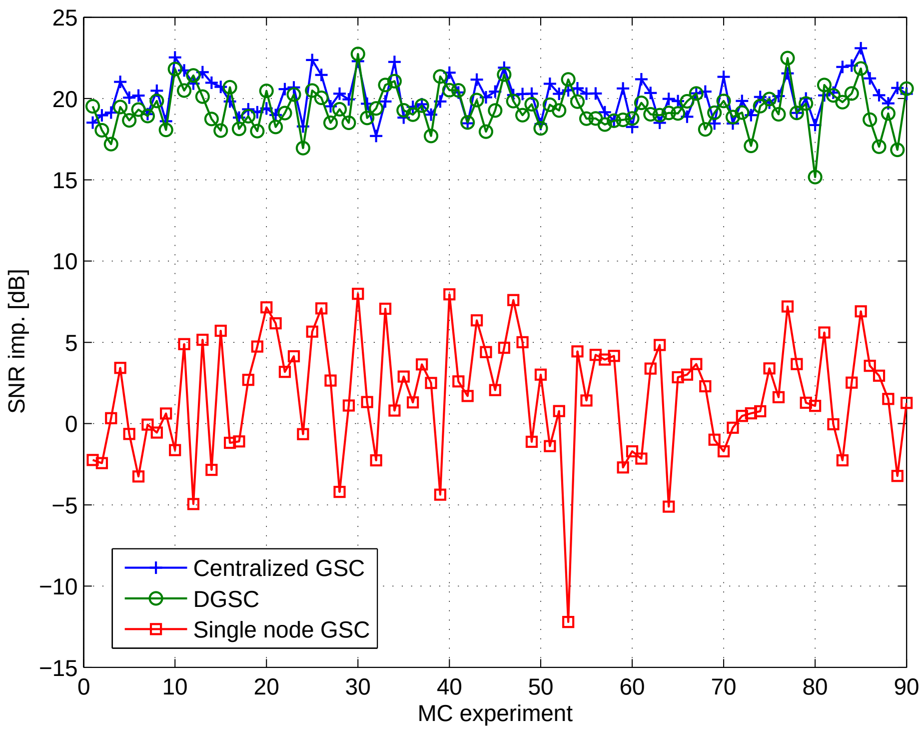


Fig. 5. The SNR improvement of the tested algorithms in various Monte Carlo experiments.

by a selection procedure in the WASN. In static environments this procedure should be applied only at the initialization stage. The second component is implemented as an adaptive algorithm which converges in speech-absent time segments.

A comprehensive experimental study validates the equivalence between the centralized GSC and the DGSC algorithms. The proposed algorithm was tested successfully for both narrowband and speech signals in multiple Monte Carlo experiments.

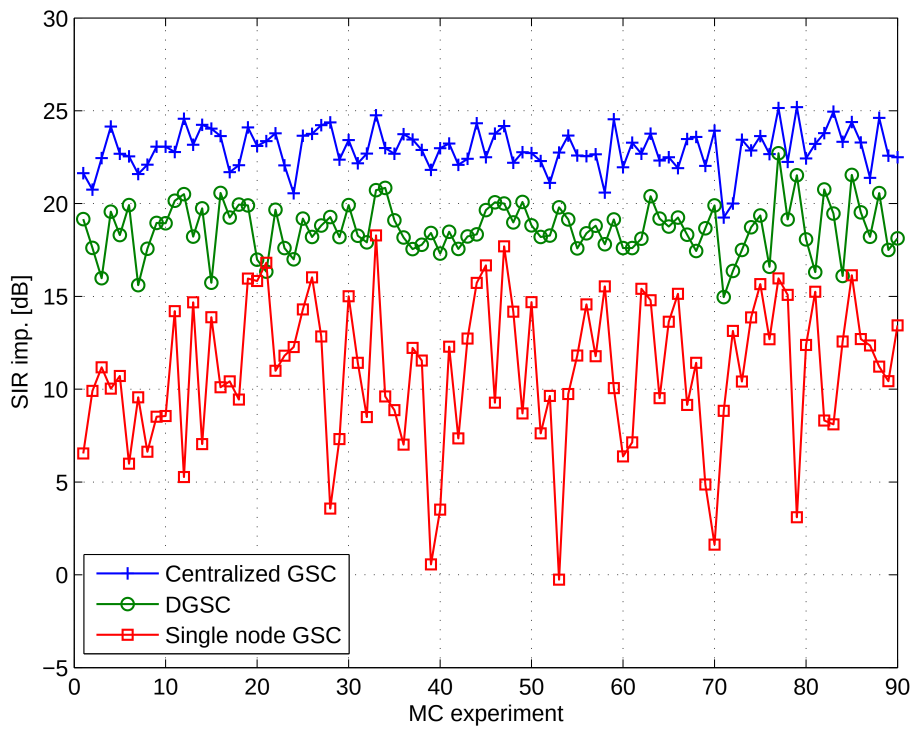


Fig. 6. The SIR improvement of the tested algorithms in various Monte Carlo experiments.

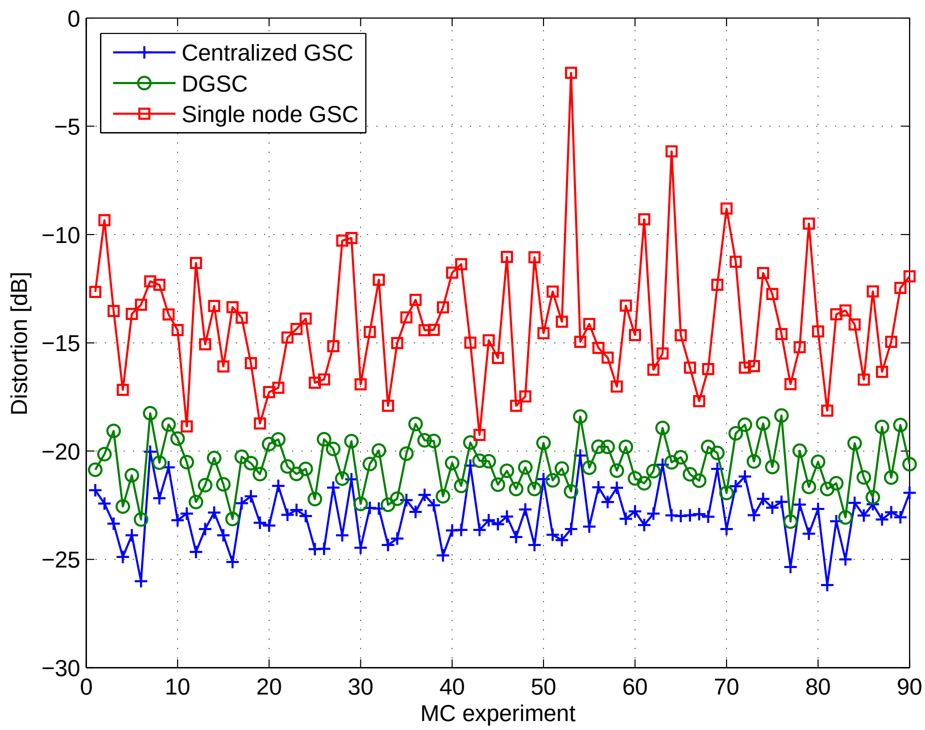


Fig. 7. The distortion of the tested algorithms in various Monte Carlo experiments.

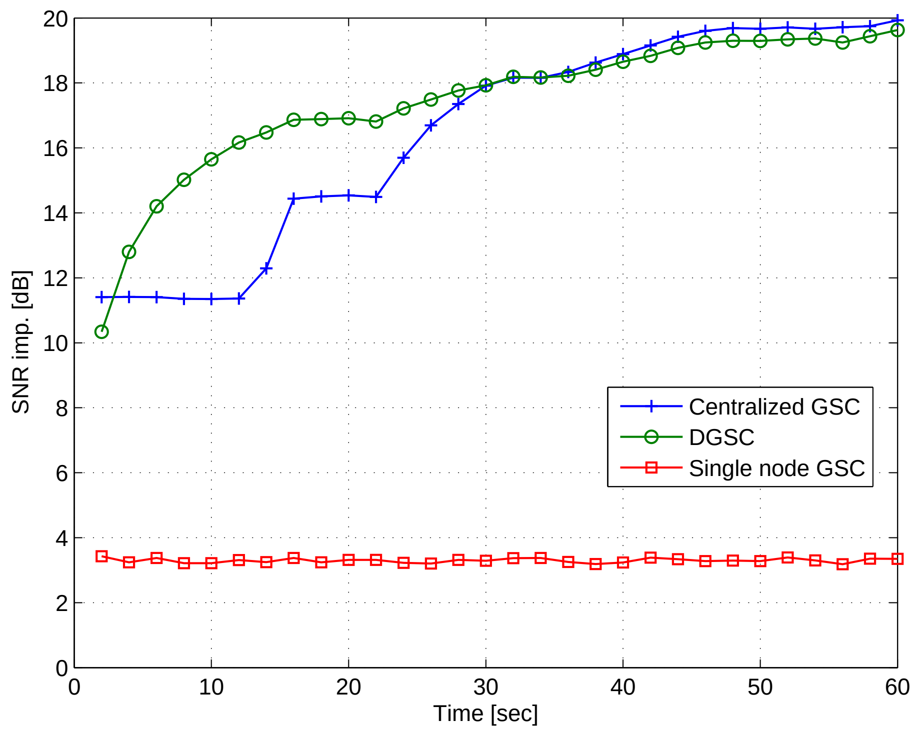
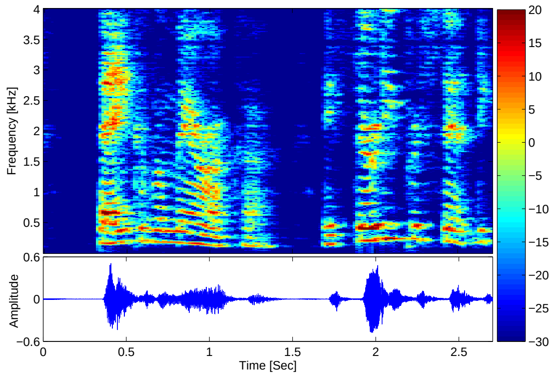
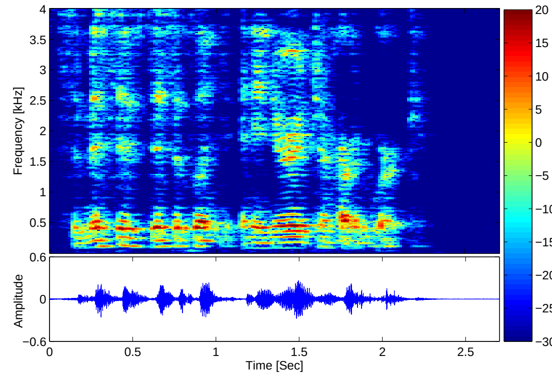


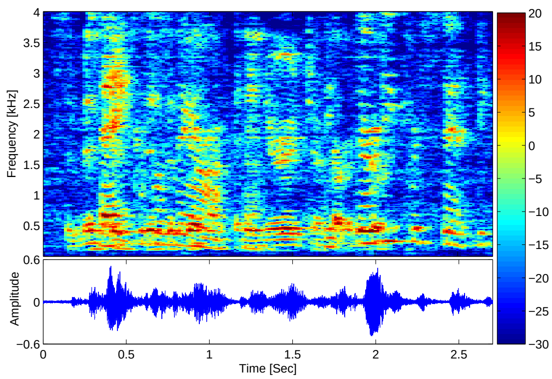
Fig. 8. The convergence of the tested algorithms versus time.



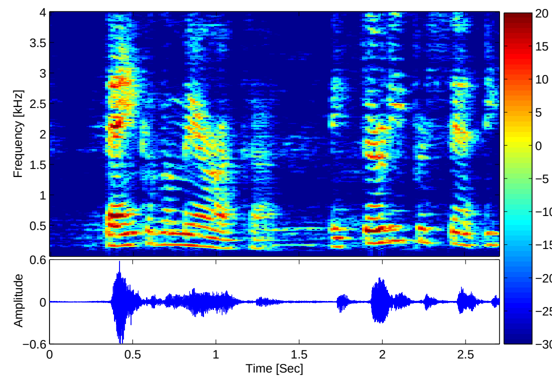
(a) Desired speaker



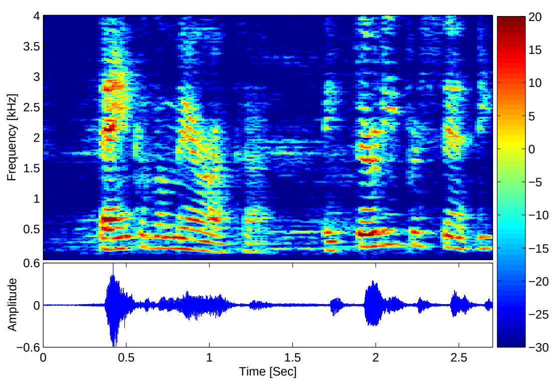
(b) Competing speaker



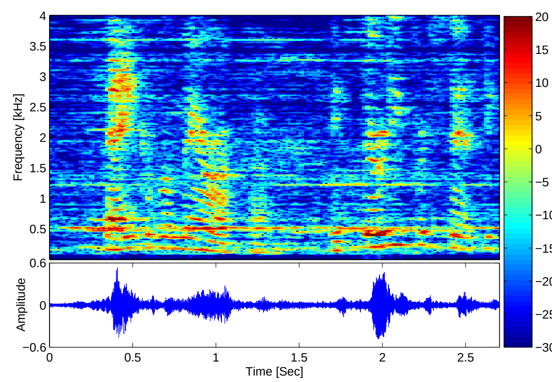
(c) First microphone



(d) Centralized. GSC



(e) DGSC



(f) Single node GSC

Fig. 9. Sonograms of the various components of the signal received in the first microphone, and the outputs of the centralized GSC, the DGSC and the single node GSC.

REFERENCES

[1] D. Estrin, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2033–2036, May 2001.

[2] D. Culler, D. Estrin, and M. Srivastava, "Overview of sensor networks," *Computer*, vol. 37, no. 8, pp. 41–49, Aug. 2004.

[3] H. Ochiai, P. Mitran, H. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4110 – 4124, nov. 2005.

[4] M. Ahmed and S. Vorobyov, "Collaborative beamforming for wireless sensor networks with Gaussian distributed sensor nodes," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 638 –643, feb. 2009.

[5] S. Wehr, I. Kozintsev, R. Lienhart, and W. Kellermann, "Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation," in *Proc. IEEE 6th Int. Symp. on Multimedia Software Eng.*, Dec. 2004, pp. 18 – 25.

[6] Y. Jia, Y. Luo, Y. Lin, and I. Kozintsev, "Distributed microphone arrays for digital home and office," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, May 2006.

[7] S. Doclo, M. Moonen, T. V. den Bogaert, and J. Wouters, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 1, pp. 38–51, Jan. 2009.

[8] 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, Oct. 2010.

[9] S. Markovich-Golan, S. Gannot, and I. Cohen, "A reduced bandwidth binaural MVDR beamformer," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.

[10] T. C. Lawin-Ore and S. Doclo, "Analysis of rate constraints for MWF-based noise reduction in acoustic sensor networks," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011.

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

[12] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[13] S. Doclo and M. Moonen, "GSVD-based optimal filtering for multi-microphone speech enhancement," in *Microphone arrays: Signal processing techniques and applications*. Springer, 2001, pp. 111–132.

[14] ——, "GSVD-Based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[15] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[16] M. Er and A. Cantoni, "Derivative constraints for broad-band element space antenna array processors," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, no. 6, pp. 1378–1393, Dec. 1983.

[17] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Speech distortion weighted multichannel Wiener filtering techniques for noise reduction," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. Springer, 2005, pp. 199–228.

[18] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[19] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, pp. 27–34, Jan. 1982.

[20] B. R. Breed and J. Strauss, "A short proof of the equivalence of LCMV and GSC beamforming," *IEEE Signal Processing Lett.*, vol. 9, no. 6, pp. 168–169, June 2002.

[21] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[22] S. Markovich-Golan, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[23] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 233 –246, jan. 2012.

[24] M. F. A. Ahmed and S. A. Vorobyov, "Sidelobe control in collaborative beamforming via node selection," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6168 –6180, Dec. 2010.

[25] 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, May 2011.

[26] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 451 – 459, Sept. 2004.

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

[28] ——, "A sparse blocking matrix for multiple constraints GSC beamformer," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Mar. 2012.

[29] B. Widrow and S. D. Stearns, "The LMS algorithm," in *Adaptive Signal Processing*, S. Haykin, Ed. Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985.

[30] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

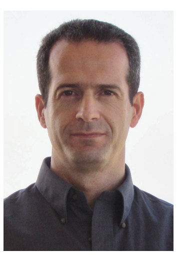
[31] E. Habets, "Room impulse response (RIR) generator," http://home.tiscali.nl/ehabets/rir_generator.html, July 2006.

[32] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

Shmulik Markovich-Golan Received the B.Sc. (Cum Laude) and M.Sc. degrees in electrical engineering from the Technion – Israel Institute of Technology, Haifa, Israel, in 2002 and 2008 respectively. He is currently pursuing the Ph.D. degree at the Engineering Faculty in Bar-Ilan University. His research interests include multi-channel signal processing, distributed sensor networks, speech enhancement using microphone arrays and distributed estimation.

Sharon Gannot (S'92-M'01-SM'06) received his B.Sc. degree (summa cum laude) from the Technion – Israel Institute of Technology, Haifa, Israel in 1986 and the M.Sc. (cum laude) and Ph.D. degrees from Tel-Aviv University, Israel in 1995 and 2000 respectively, all in electrical engineering. In the year 2001 he held a post-doctoral position at the department of Electrical Engineering (ESAT-SISTA) at K.U.Leuven, Belgium. From 2002 to 2003 he held a research and teaching position at the Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. Currently, he is an Associate Professor at the School of Engineering, Bar-Ilan University, Israel, where he is heading the Speech and Signal Processing laboratory. Prof. Gannot is the recipient of Bar-Ilan University outstanding lecturer award for 2010.

Prof. Gannot is currently an Associate Editor of IEEE Transactions on Speech, Audio and Language processing. He served as an Associate Editor of the EURASIP Journal of Advances in signal Processing between 2003-2001, and as an Editor of two special issues on Multi-microphone Speech Processing of the same journal. He also served as a guest editor of ELSEVIER Speech Communication journal and a reviewer of many IEEE journals and conferences. Prof. Gannot is a member of the Audio and Acoustic Signal Processing (AASP) technical committee of the IEEE since Jan., 2010. He is also a member of the Technical and Steering committee of the International Workshop on Acoustic Echo and Noise Control (IWAENC) since 2005 and the general co-chair of IWAENC held at Tel-Aviv, Israel in August 2010. Prof. Gannot will serve as the general co-chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2013. His research interests include parameter estimation, statistical signal processing, especially speech processing using either single- or multi-microphone arrays.



Israel Cohen (M'01-SM'03) is an Associate Professor of electrical engineering at the Technion – Israel Institute of Technology, Haifa, Israel. He received the B.Sc. (Summa Cum Laude), M.Sc. and Ph.D. degrees in electrical engineering from the Technion – Israel Institute of Technology, in 1990, 1993 and 1998, respectively.

From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT. In 2001 he joined the Electrical Engineering Department of the Technion. His research interests are statistical signal processing, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification and adaptive filtering. He is a coeditor of the Multichannel Speech Processing section of the *Springer Handbook of Speech Processing* (Springer, 2008), a coauthor of *Noise Reduction in Speech Processing* (Springer, 2009), a coeditor of *Speech Processing in Modern Communication: Challenges and Perspectives* (Springer, 20010), and a general co-chair of the 2010 International Workshop on Acoustic Echo and Noise Control (IWAENC).

Prof. Cohen is a recipient of the Alexander Goldberg Prize for Excellence in Research, and the Muriel and David Jacknow award for Excellence in Teaching. He served as Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS, and as Guest Editor of a special issue of the *EURASIP Journal on Advances in Signal Processing* on Advances in Multimicrophone Speech Processing and a special issue of the *Elsevier Speech Communication Journal* on Speech Enhancement.