DISTRIBUTED GSC BEAMFORMING USING THE RELATIVE TRANSFER FUNCTION

Shmulik Markovich-Golan[1], Sharon Gannot[1] and Israel Cohen[2]

[1] Faculty of Engineering
Bar-Ilan University
Ramat-Gan, 52900, Israel
shmulik.markovich@gmail.com; sharon.gannot@biu.ac.il

[2] Department of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel
icohen@ee.technion.ac.il

ABSTRACT

A speech enhancement algorithm in a noisy and reverberant enclosure for a wireless acoustic sensor network (WASN) is derived. The proposed algorithm is structured as a two stage beamformers (BFs) scheme, where the outputs of the first stage are transmitted in the network. Designing the second stage BF requires estimating the desired signal components at the transmitted signals. The contribution here is twofold. First, in spatially static scenarios, the first stage BFs are designed to maintain a fixed response towards the desired signal. As opposed to competing algorithms, where the response changes and repeated estimation thereof is required. Second, the proposed algorithm is implemented in a generalized sidelobe canceler (GSC) form, separating the treatment of the desired speech and the interferences and enabling a simple time-recursive implementation of the algorithm. A comprehensive experimental study demonstrates the equivalent performance of the centralized GSC and of the proposed algorithm for both narrowband and speech signals.

1. INTRODUCTION

Distributed signal processing algorithms for WASNs have recently gained increased attention, raising research questions regarding power constraints, computational burden in the nodes, and maintenance of wireless links. A survey on the subject matter by Bertand is given in [1]. In this work, we consider the problem of speech enhancement in reverberant enclosures in WASNs.

Two criteria are commonly used for designing BFs [2] in this context: the minimum mean square error (MMSE), which minimizes the variance of the error at the output, and the minimum variance distortionless response (MVDR) which minimizes the noise power at the output while maintaining undistorted desired signal at the output. Doclo et al. [3] proposed the speech distortion weighted (SDW)-multi-channel Wiener filter (MWF) BF which generalizes both BFs. Doclo et al. [4] proposed a distributed algorithm for binaural hearing aid comprising multiple microphones where the left and right apparatuses are connected with a bidirectional wireless audio link. Bertrand and Moonen [5] generalized the problem and proposed the distributed adaptive node-specific signal estimation (DANSE)-k, which considers N nodes and k desired speakers. Their algorithm requires $N \times k$ audio channels.

In the current contribution, the MVDR criterion is considered. The MVDR is a special case of the linearly constrained minimum variance (LCMV), which is capable of constraining the response of multiple speakers. Gannot et al. [6] proposed a frequency domain MVDR criterion implemented in a GSC structure (A seminal work by Griffiths and Jim, 1982), namely a transfer function generalized sidelobe canceler (TF-GSC). Markovich-Golan et al. [7] proposed a distributed MVDR algorithm for the binaural case. Bertrand et al. [8] addressed the more general problem and introduced the linearly constrained (LC)-DANSE which minimizes the noise power at the output while maintaining P linear constraints. Their algorithm requires transmission of $N \times P$ audio channels.

Here, only the case of a single desired speaker is addressed. The efficient GSC-form implementation of the MVDR, rather than its closed-form, is considered. The GSC-form relaxes the requirement of the LC-DANSE algorithm to re-estimate the speech and noise spectra at each iteration. Both iterative and time recursive procedures are derived in this paper. The proposed algorithm, denoted distributed single-constraint generalized sidelobe canceler (DS-GSC), is based on a two-stage GSC. In the first stage, N local GSC-BFs are applied only to the local microphones at each node, yielding N signals which are broadcasted in the WASN. The second stage, comprises a global GSC BF which processes the N output signals of the first stage. A replica of the global filtering stage is applied simultaneously and independently in all the nodes of the WASN. The main advantages of the proposed scheme are its ability to adapt in speech-absent segments, and that it relaxes the requirement of closed-form MVDR algorithms to estimate the speech spectrum repeatedly in static environments. It is experimentally shown that the proposed algorithm is equivalent to the centralized TF-GSC. A comprehensive theoretical convergence proof and a robustness analysis is beyond the scope of this paper.

The structure of the paper is as follows. The problem is formulated in Sec. 2. In Sec. 3, the TF-GSC BF is presented. In Sec. 4, we introduce the proposed GSC-based distributed algorithm, namely the DS-GSC. A comprehensive experimental study is given in Sec. 5. Some concluding remarks are given in Sec. 6.

2. PROBLEM FORMULATION

Consider an N-node WASN. Denote by M_n the number of microphones in the nth node, and by $M = \sum_{n=1}^{N} M_n$ the total number of microphones. The problem is formulated in the short time Fourier transform (STFT) domain, where k denotes the frequency index and ℓ denotes the segment index. Let $\mathbf{z}(\ell, k)$ be a vector constructed by all received microphone signals $\mathbf{z}(\ell, k) = \begin{bmatrix} \mathbf{z}_1^T(\ell, k) & \cdots & \mathbf{z}_N^T(\ell, k) \end{bmatrix}^T$ where $\mathbf{z}_n(\ell, k)$ is the local $M_n \times 1$ received signals vector of the nth node and $(\cdot)^T$ is the transpose operator. The term *local* (to the nth node) is associated with the signals and parameters calculated using only the microphones at the nth node. By the term *global* we designate signals and parameters which are calculated using data from other nodes shared via the wireless link. The global vector of received signals is formulated as:

$$\mathbf{z}(\ell, k) = \mathbf{h}(\ell, k)s(\ell, k) + \mathbf{v}(\ell, k) \quad (1)$$

where $s(\ell, k)$ is the desired speech source, and $\mathbf{h}(\ell, k)$ consists of a vector of $M \times 1$ acoustic transfer function (ATF) from the desired speaker to the microphones. The vector $\mathbf{v}(\ell, k)$ is a vector of interferences picked up by microphones. Assuming that the speaker and the noise are statistically independent, the covariance matrix of the received signals is:

$$\mathbf{\Phi}_{zz}(\ell, k) = \lambda_s(\ell, k)\mathbf{h}(\ell, k)\mathbf{h}^\dagger(\ell, k) + \mathbf{\Phi}_{vv}(\ell, k) \quad (2)$$

where $\lambda_s(\ell, k)$ is the variance of the desired speech signal, $\mathbf{\Phi}_{vv}(\ell, k)$ is the covariance matrix of the interferences and $(\cdot)^\dagger$ is the Hermitian operator. We assume that the network is fully connected, i.e., each transmitted signal is available to all nodes. In networks that are not fully connected a variation of the proposed algorithm can be derived, however, it is beyond the scope of this contribution. We assume that the location of the speaker is static and that the noise sources' statistics is slowly time-varying. Therefore, the speaker ATF and the noise covariance matrix are assumed time-invariant. Hence, in these quantities the frame index is omitted. For brevity, explicit frequency dependence is omitted hereafter. Finally, denote by $\mathbf{U}_n$ an $M \times M_n$ matrix that extracts the local microphones $\mathbf{z}_n(\ell) = \mathbf{U}_n^\dagger \mathbf{z}(\ell)$ from $\mathbf{z}(\ell)$:

$$\mathbf{U}_n = \left[\begin{array}{c|c|c} \mathbf{0}_{M_n \times (\sum_{n'=1}^{n-1} M_{n'})} & \mathbf{I}_{M_n} & \mathbf{0}_{M_n \times (\sum_{n'=n+1}^{N} M_{n'})} \end{array} \right]^\dagger \quad (3)$$

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

The problem at hand is to enhance the desired speech signal at each node, with access only to the local microphones and the transmitted signals.

3. THE CENTRALIZED TF-GSC BF

Let $\mathbf{w}$ be the centralized TF-GSC BF. Recall that $\mathbf{w}$ is the output power $\mathbf{w}^\dagger \mathbf{\Phi}_{vv} \mathbf{w}$ minimizer, that satisfies the constraint $\mathbf{h}^\dagger \mathbf{w} = 1$. In many applications, when the goal is to reduce the noise level, while sacrificing dereverberation, it is sufficient to enhance the desired signal as received by a reference microphone (arbitrarily chosen here to be the first microphone). The relative transfer function (RTF) [6], $\tilde{\mathbf{h}}$, is defined as the vector of ATFs from the desired signal to the microphones normalized by the ATF to the reference microphone, $\tilde{\mathbf{h}} = \mathbf{h}/h_1$. The resulting modified constraint satisfies $\tilde{\mathbf{h}}^\dagger \mathbf{w} = 1$ (see [6]).

Now, we are ready to formulate the centralized GSC:

$$\tilde{\mathbf{w}} = \tilde{\mathbf{q}} - \tilde{\mathbf{B}}\tilde{\mathbf{f}} \quad (4)$$

where $\tilde{\mathbf{q}}$ denotes the fixed beamformer (FBF) and is parallel to the RTF, $\tilde{\mathbf{B}}$ denotes the blocking matrix (BM) and $\tilde{\mathbf{f}}$ denotes the noise canceller (NC). The FBF [2] is given by:

$$\tilde{\mathbf{q}} = \frac{\tilde{\mathbf{h}}}{\|\tilde{\mathbf{h}}\|^2}. \quad (5)$$

The matrix $\tilde{\mathbf{B}}$ is designed to block the RTF of the speaker, i.e., $\tilde{\mathbf{B}}^\dagger \tilde{\mathbf{h}} = \mathbf{0}_{(M-1)\times 1}$, but it is not unique. One way of constructing the BM is by calculating the singular value decomposition (SVD) of $\tilde{\mathbf{h}}$:

$$\tilde{\mathbf{h}} = \tilde{\mathbf{\Theta}}\tilde{\mathbf{\Gamma}}\tilde{\mathbf{\Psi}}^\dagger \quad (6)$$

and selecting the $M - 1$ left singular vectors (of $\tilde{\mathbf{\Theta}}$) which correspond to the zero singular values [2]. The NC is given in a closed-form by:

$$\tilde{\mathbf{f}} = \left(\tilde{\mathbf{B}}^\dagger \mathbf{\Phi}_{vv} \tilde{\mathbf{B}}\right)^{-1} \tilde{\mathbf{B}} \mathbf{\Phi}_{vv} \tilde{\mathbf{q}}. \quad (7)$$

An efficient time-recursive implementation for adapting the NC during speech-absent segments utilizes the normalized least mean squares (LMS) (NLMS) algorithm:

$$\tilde{\mathbf{f}}(\ell + 1) = \tilde{\mathbf{f}}(\ell) + \frac{\mu}{\tilde{\lambda}_u(\ell)} \tilde{\mathbf{u}}(\ell)\tilde{y}^*(\ell) \quad (8)$$

where the noise reference signals at the output of the BM are denoted $\tilde{\mathbf{u}}(\ell) = \tilde{\mathbf{B}}^\dagger \mathbf{z}(\ell)$, the output of the TF-GSC is denoted $\tilde{y}(\ell) = \tilde{\mathbf{w}}^\dagger \mathbf{z}(\ell)$, the adaptation step is $0 < \mu < 2$, $\tilde{\lambda}_u(\ell)$ is a power normalization factor:

$$\tilde{\lambda}_u(\ell + 1) = \rho \tilde{\lambda}_u(\ell) + (1 - \rho)\|\tilde{\mathbf{u}}(\ell)\|^2, \quad (9)$$

and ρ is a forgetting factor (typically $0.8 < \rho < 1$).

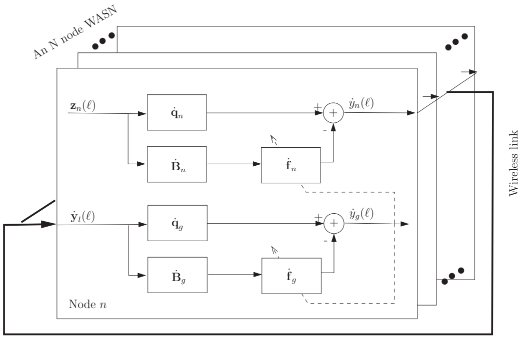

Fig. 1. A block-diagram of the DS-GSC.

4. THE DS-GSC

A two-stage distributed enhancement TF-GSC BF algorithm, denoted as DS-GSC, is now proposed. The first stage consists of N TF-GSC-BFs which process local signals at each node. The output signals of the first stage are broadcasted in the WASN. The second stage consists of a global TF-GSC BF which processes the N outputs of the first stage. A replica of the global BF is concurrently applied in all nodes.

Denote the first stage output at the nth node as $\dot{y}_n(\ell)$ for $n = 1, .., N$. Let the concatenated N outputs of the first stage be $\dot{\mathbf{y}}_l(\ell) = \dot{\mathbf{W}}_l^\dagger \mathbf{z}(\ell) = \left[\begin{array}{ccc} \dot{y}_1(\ell) \cdots & \dot{y}_N(\ell) \end{array} \right]^T$, where the subscript l denotes *local*, the N local BFs are given in a matrix notation by $\dot{\mathbf{W}}_l = \left[\begin{array}{ccc} \mathbf{U}_1 \dot{\mathbf{w}}_1 & \cdots & \mathbf{U}_N \dot{\mathbf{w}}_N \end{array} \right]$, and $\dot{\mathbf{w}}_n$ are filters of the first stage TF-GSC BF at the nth node:

$$\dot{\mathbf{w}}_n = \dot{\mathbf{q}}_n - \dot{\mathbf{B}}_n \dot{\mathbf{f}}_n. \tag{10}$$

Note that $\dot{\mathbf{W}}_l$ is an $M \times N$ matrix. The local FBF, BM and NC at the nth node are denoted $\dot{\mathbf{q}}_n$, $\dot{\mathbf{B}}_n$ and $\dot{\mathbf{f}}_n$, respectively. The total output of the DS-GSC is given by $\dot{y}_g(\ell) = \dot{\mathbf{w}}_g^\dagger \dot{\mathbf{y}}_l(\ell)$ where the subscript g denotes *global* and

$$\dot{\mathbf{w}}_g = \dot{\mathbf{q}}_g - \dot{\mathbf{B}}_g \dot{\mathbf{f}}_g \tag{11}$$

is the second stage global TF-GSC BF. The second stage FBF, BM and NC are denoted by $\dot{\mathbf{q}}_g$, $\dot{\mathbf{B}}_g$ and $\dot{\mathbf{f}}_g$, respectively. A block-diagram of the DS-GSC algorithm is depicted in Fig. 1.

The proposed algorithm consists of three phases: first, the local RTFs are estimated, and the local FBFs and BMs are constructed; second, the global RTFs are estimated, and the global FBF and BM are constructed; third, the local and global NCs are alternatingly updated until convergence. The first two phases are only applied once in static environments. We adopt in our implementation a standard subspace-based RTF estimation procedure [7], where a perfect voice activity detector (VAD) is assumed.

The local and global stage filtering are presented in Secs. 4.1 and 4.2, respectively. In Sec. 4.3 we present an iterative algorithm for updating the NCs, and prove its convergence. Later, in Sec. 4.4 we derive a time-recursive variant.

4.1. Local stage filtering

Denote by $\dot{\mathbf{h}}_n$ the local RTF, which equals the ATF relating the desired signal and the local microphone signals at the nth node normalized by its first component, $\dot{\mathbf{h}}_n = \mathbf{U}_n^\dagger \mathbf{h} / \left(\mathbf{U}_n^\dagger \mathbf{h} \right)_1$. The local TF-GSC BF at the nth node is designed to satisfy the constraint $\dot{\mathbf{h}}_n^\dagger \dot{\mathbf{w}}_n = 1$, therefore the desired signal component at its output is $h_{n,1} s(\ell)$. Similarly to (5) and (6) the local FBF is given by:

$$\dot{\mathbf{q}}_n = \frac{\dot{\mathbf{h}}_n}{\|\dot{\mathbf{h}}_n\|^2} \tag{12}$$

and the local BM, $\dot{\mathbf{B}}_n$, is constructed by the $M_n - 1$ left singular vectors, corresponding to the zero singular values in the SVD of $\dot{\mathbf{h}}_n$.

4.2. Global stage filtering

Denote by $\dot{\mathbf{h}}_g$ the global RTF, which equals the ATF relating the desired signal and the local output signals $\dot{\mathbf{y}}_l(\ell)$, normalized by its first component $\dot{\mathbf{h}}_g = \frac{\dot{\mathbf{W}}_l^\dagger \mathbf{h}}{h_1}$. Note, that the fixed response of the desired source at the local outputs, simplifies the global RTF estimation in static scenarios. Following similar arguments to (5) and (6) the global FBF is $\dot{\mathbf{q}}_g = \frac{\dot{\mathbf{h}}_g}{\|\dot{\mathbf{h}}_g\|^2}$ and the global BM, $\dot{\mathbf{B}}_g$, is constructed by the $N - 1$ left singular vectors, corresponding to the zero singular values in the SVD of $\dot{\mathbf{h}}_g$. Consequently, the desired signal component at the BF output is $h_1 s(\ell)$.

4.3. Iterative algorithm

Following the results of previous sections, the noise component at the output of the DS-GSC is:

$$\dot{v}(\ell) = \dot{\mathbf{w}}_g^\dagger \dot{\mathbf{W}}_l^\dagger \mathbf{v}(\ell) = \dot{\mathbf{w}}_l^\dagger \dot{\mathbf{W}}_g^\dagger \mathbf{v}(\ell). \tag{13}$$

where $\dot{\mathbf{W}}_g$ is an $M \times M$ diagonal matrix given by

$$\dot{\mathbf{W}}_g = \text{blkdiag} \left\{ \dot{w}_{g,1} \mathbf{I}_{M_1}, .., \dot{w}_{g,N} \mathbf{I}_{M_N} \right\} \tag{14}$$

and $\dot{\mathbf{w}}_l$ is a concatenation of the local GSC-BFs:

$$\dot{\mathbf{w}}_l = \dot{\mathbf{q}}_l - \dot{\mathbf{B}}_l \dot{\mathbf{f}}_l \tag{15}$$

and its components are given by:

$$\dot{\mathbf{q}}_l = \left[\begin{array}{ccc} \dot{\mathbf{q}}_1^\dagger & \cdots & \dot{\mathbf{q}}_N^\dagger \end{array} \right]^\dagger \tag{16a}$$

$$\dot{\mathbf{B}}_l = \text{blkdiag} \left\{ \dot{\mathbf{B}}_1, .., \dot{\mathbf{B}}_N \right\} \tag{16b}$$

$$\dot{\mathbf{f}}_l = \left[\begin{array}{ccc} \dot{\mathbf{f}}_1^\dagger & \cdots & \dot{\mathbf{f}}_N^\dagger \end{array} \right]^\dagger . \tag{16c}$$

The variance of the noise component (13) is:

$$\gamma = \dot{\mathbf{w}}_g^\dagger \dot{\mathbf{W}}_l^\dagger \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_l \dot{\mathbf{w}}_g \tag{17a}$$

$$= \dot{\mathbf{w}}_l^\dagger \dot{\mathbf{W}}_g^\dagger \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_g \dot{\mathbf{w}}_l. \tag{17b}$$

Since the FBFs and BMs have been already determined (recall that the acoustic scenario is static), only the NC filters $\dot{\mathbf{f}}_l$ and $\dot{\mathbf{f}}_g$ should be set for minimizing the residual noise power. The concatenated NCs vector is denoted by $\dot{\mathbf{f}} = \begin{bmatrix} \dot{\mathbf{f}}_l^\dagger & \dot{\mathbf{f}}_g^\dagger \end{bmatrix}^\dagger$.

We propose an iterative algorithm comprised of two alternating steps sequentially: first, updating the local NCs $\dot{\mathbf{f}}_l$; second, updating the global NC $\dot{\mathbf{f}}_g$. We denote values at the ith iteration with the superscript $(\cdot)^{(i)}$.

Consider the first update step, i.e., $\dot{\mathbf{f}}_g^{(i)}$ is updated to $\dot{\mathbf{f}}_g^{(i+1)}$ while $\dot{\mathbf{f}}_l^{(i+1)} = \dot{\mathbf{f}}_l^{(i)}$ remains fixed. An explicit expression of (17a) in terms of $\dot{\mathbf{f}}_g^{(i)}$ at the ith iteration is given by:

$$\gamma^{(i)} = \left(\dot{\mathbf{q}}_g - \dot{\mathbf{B}}_g \dot{\mathbf{f}}_g^{(i)}\right)^\dagger \left(\dot{\mathbf{W}}_l^{(i)}\right)^\dagger \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_l^{(i)} \left(\dot{\mathbf{q}}_g - \dot{\mathbf{B}}_g \dot{\mathbf{f}}_g^{(i)}\right). \tag{18}$$

Equation (18) is a quadratic function of $\dot{\mathbf{f}}_g^{(i)}$, allowing for a simple calculation of the gradient with respect to $\dot{\mathbf{f}}_g$:

$$\frac{\partial \gamma^{(i)}}{\partial \dot{\mathbf{f}}_g^\dagger} = -\dot{\mathbf{B}}_g^\dagger \left(\dot{\mathbf{W}}_l^{(i)}\right)^\dagger \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_l^{(i)} \left(\dot{\mathbf{q}}_g - \dot{\mathbf{B}}_g \dot{\mathbf{f}}_g\right). \tag{19}$$

The updated $\dot{\mathbf{f}}_g$ is obtained by equating the gradient to zero:

$$\dot{\mathbf{f}}_g^{(i+1)} = \left(\dot{\mathbf{B}}_g^\dagger \left(\dot{\mathbf{W}}_l^{(i)}\right)^\dagger \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_l^{(i)} \dot{\mathbf{B}}_g\right)^{-1} \cdot \dot{\mathbf{B}}_g^\dagger \left(\dot{\mathbf{W}}_l^{(i)}\right)^\dagger \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_l^{(i)} \dot{\mathbf{q}}_g. \tag{20}$$

Consider the second update step, i.e., $\dot{\mathbf{f}}_l^{(i)}$ is updated to $\dot{\mathbf{f}}_l^{(i+1)}$ while $\dot{\mathbf{f}}_g^{(i+1)} = \dot{\mathbf{f}}_g^{(i)}$ remains fixed. An explicit expression of (17b) at the ith iteration is given in terms of $\dot{\mathbf{f}}_l$ by:

$$\gamma^{(i)} = \left(\dot{\mathbf{q}}_l - \dot{\mathbf{B}}_l \dot{\mathbf{f}}_l^{(i)}\right)^\dagger \left(\dot{\mathbf{W}}_g^{(i)}\right)^\dagger \mathbf{\Phi}_{vv} \left(\dot{\mathbf{W}}_g^{(i)}\right)^\dagger \left(\dot{\mathbf{q}}_l - \dot{\mathbf{B}}_l \dot{\mathbf{f}}_l^{(i)}\right). \tag{21}$$

Equation (21) is a quadratic function of $\dot{\mathbf{f}}_l$. The gradient of $\gamma^{(i)}$ with respect to $\dot{\mathbf{f}}_l$ is:

$$\frac{\partial \gamma^{(i)}}{\partial \dot{\mathbf{f}}_l^\dagger} = -\dot{\mathbf{B}}_l^\dagger \left(\dot{\mathbf{W}}_g^{(i)}\right)^\dagger \mathbf{\Phi}_{vv} \left(\dot{\mathbf{W}}_g^{(i)}\right)^\dagger \left(\dot{\mathbf{q}}_l - \dot{\mathbf{B}}_l \dot{\mathbf{f}}_l\right) \tag{22}$$

yielding

$$\dot{\mathbf{f}}_l^{(i+1)} = \left(\dot{\mathbf{B}}_l^\dagger \left(\dot{\mathbf{W}}_g^{(i)}\right)^\dagger \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_g^{(i)} \dot{\mathbf{B}}_l\right)^{-1} \cdot \dot{\mathbf{B}}_l^\dagger \left(\dot{\mathbf{W}}_g^{(i)}\right)^\dagger \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_g^{(i)} \dot{\mathbf{q}}_l. \tag{23}$$

It can be easily verified that the proposed algorithm converges. First, $\gamma^{(i+1)} \leq \gamma^{(i)}$ is monotonically non-increasing, since $\dot{\mathbf{f}}^{(i)}$ belongs to the minimization range of $\dot{\mathbf{f}}^{(i+1)}$. Second, $\gamma^{(i)}$ is trivially lower bounded by $0 \leq \gamma^{(i)}$. In practice the iterative algorithm cannot be implemented, since updating $\dot{\mathbf{f}}_l$ (23) involves non-local quantities unavailable at each node. However, a practical time-recursive algorithm can be derived, as presented in the sequel.

4.4. Time-recursive algorithm

A time-recursive procedure for updating the NCs is obtained by using two NLMS algorithms for updating $\dot{\mathbf{f}}_g(\ell)$ and $\dot{\mathbf{f}}_l(\ell)$ alternately, during speech-absent time segments. As in all stochastic approximation procedures, we propose to replace the iteration index i by the segment index ℓ. We further propose to perform L_u updates of each filter before switching to the other filter updates.

Consider an update step of $\dot{\mathbf{f}}_g(\ell)$ to $\dot{\mathbf{f}}_g(\ell+1)$ while $\dot{\mathbf{f}}_l(\ell+1) = \dot{\mathbf{f}}_l(\ell)$ is unaltered. An instantaneous estimate of the gradient (19) at the ℓth frame is $-\dot{\mathbf{u}}_g(\ell)\dot{v}^*(\ell)$ where $\dot{\mathbf{u}}_g(\ell) = \dot{\mathbf{B}}_g^\dagger \dot{\mathbf{v}}_l(\ell)$, and $\dot{\mathbf{v}}_l(\ell)$ is the noise component in $\dot{\mathbf{y}}_l(\ell)$. The updated $\dot{\mathbf{f}}_g(\ell+1)$ is therefore:

$$\dot{\mathbf{f}}_g(\ell+1) = \dot{\mathbf{f}}_g(\ell) + \frac{\mu}{\dot{\lambda}_{u,g}(\ell)} \dot{\mathbf{u}}_g(\ell)\dot{v}^*(\ell) \tag{24}$$

where, similarly to (9), $\dot{\lambda}_{u,g}(\ell)$ is a time-recursive estimate of the power normalization. Similarly, we can define an update step of $\dot{\mathbf{f}}_l(\ell)$ to $\dot{\mathbf{f}}_l(\ell+1)$ while $\dot{\mathbf{f}}_g(\ell+1) = \dot{\mathbf{f}}_g(\ell)$. An instantaneous estimate of the gradient (22) at the ℓth frame is $-\dot{\mathbf{W}}_g^\dagger(\ell)\dot{\mathbf{u}}_l(\ell)\dot{v}^*(\ell)$, where $\dot{\mathbf{u}}_l(\ell) = \dot{\mathbf{B}}_l^\dagger \mathbf{z}(\ell)$. Note that the calculation of the nth component of the estimated gradient can be done locally. Hence, $\dot{\mathbf{f}}_n(\ell+1);\ n = 1,..,N$ can be updated in parallel:

$$\dot{\mathbf{f}}_n(\ell+1) = \dot{\mathbf{f}}_n(\ell) + \frac{\mu}{\dot{\lambda}_{u,l}^n(\ell)} \dot{w}_{g,n}^*(\ell)\dot{\mathbf{u}}_n(\ell)\dot{v}^*(\ell) \tag{25}$$

where, as in (9), $\dot{\lambda}_{u,l}^n(\ell)$ is a power normalization factor.

5. EXPERIMENTAL STUDY

An experimental study of three algorithms using multiple Monte-Carlo trials is presented. We compare the centralized TF-GSC, the time-recursive DS-GSC and the local TF-GSC (a TF-GSC BF which utilizes only the microphones of a the first node). A WASN comprised of $N = 4$ nodes, each consisting of $M_n = 2$ microphones was simulated. The performance of the BFs was evaluated by using the signal to total interference ratio (SIR) improvement.

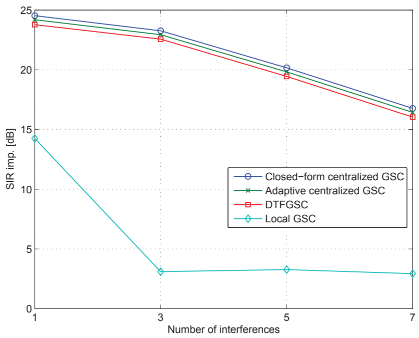

Fig. 2. SIR improvement versus the number of interferences with narrowband signals.

5.1. Narrowband signals

The performance of the algorithms was tested with various numbers of interfering sources $(1, 3, 5, 7)$. For each number of interfering sources, 20 different sets of desired source ATFs and noise covariance matrices were randomized at a particular frequency bin. For each set, 10 realizations of 10^5 samples of the desired and interfering narrowband sources were randomized (white Gaussian processes). Also, a spatially white noise was added to the received signals. The input SIR and signal to noise ratio (SNR) were set to 5dB and 30dB, respectively. In total, 800 Monte-Carlo experiments were tested. The algorithm parameters were set to: $\mu = 0.15$, $\rho = 0.95$ and $L_u = 12$. The SIR improvement of the proposed DS-GSC exhibits equivalent performance to the centralized TF-GSC in all Monte-Carlo experiments, and both significantly outperform the local TF-GSC. The average SIR improvement versus the number of interferences is depicted in Fig. 2.

5.2. Speech signals

A 4m $\times$ 3m $\times$ 3m room with a reverberation time of $T_{60} =$ 300ms was simulated. Four nodes were arbitrarily positioned at the center of the walls. The locations of the desired speaker and the two interfering sources were randomized with a minimum distance of 10cm from the walls in 100 Monte-Carlo experiments. An additive spatially white noise was added to the received signals. The input SIR and SNR were set to 6dB and 60dB, respectively. The sample rate was 8KHz and a 4096 point STFT with 75% overlap was used. The algorithm parameters were set to $\mu = 0.15$, $\rho = 0.95$ and $L_u = 12$. The average SIR improvement of the centralized TF-GSC, the DS-GSC and the local TF-GSC were 26.4dB, 28.6dB and 12.2dB, respectively. Theoretically, the DS-GSC cannot outperform the centralized TF-GSC, but in our experiments a 2.2dB improvement of the DS-GSC was recorded. This can be attributed to the finite segment length and the lengths of the different equivalent filters applied in each scheme.

6. CONCLUSIONS

A DS-GSC algorithm was proposed and both iterative and time-recursive adaptation procedures were derived. In static scenarios, only a single estimate of the speech RTF is required, as opposed to closed-form based distributed algorithms, which require repeated RTFs estimations. The convergence of the proposed algorithm was proven, however convergence to the optimal TF-GSC was beyond of the scope of the current contribution. A comprehensive experimental study demonstrates the equivalence between the centralized TF-GSC and the proposed DS-GSC.

7. REFERENCES

[1] 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.

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

[3] 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., pp. 199–228. Springer, 2005.

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

[5] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part I: Sequential node updating," *IEEE Trans. on Signal Process.*, vol. 58, no. 10, pp. 5277–5291, Oct. 2010.

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

[7] 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.

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