Blind Room Acoustic System Identification in Wireless Acoustic Sensor Networks

Matthias Blochberger, *Graduate Student Member, IEEE,* Filip Elvander, *Member, IEEE,* Randall Ali, *Member, IEEE,* Marc Moonen, *Fellow, IEEE,* Jan Østergaard, *Senior Member, IEEE,* Jesper Jensen, *Member, IEEE,* Toon van Waterschoot, *Member, IEEE*

***Abstract*—Blind Room Acoustic System Identification is an important task in applications involving Wireless Acoustic Sensor Networks (WASNs). In this work, we propose a distributed blind system identification (BSI) algorithm for room acoustic systems based on the cross-relation (CR) method and the Alternating Direction Method of Multipliers (ADMM). The algorithm operates over WASNs, enabling the identification of Room Impulse Responses (RIRs) without centralized data processing. We formulate the BSI problem as a modified general-form consensus problem, where sub-problems are solved by individual nodes in the network by a local quasi-Newton method, leading to a scalable solution for large sensor networks. The proposed method extends previous works by incorporating voice activity detection (VAD) for robust performance under non-stationary noisy signals. Additionally, we introduce regularization to improve convergence and estimation performance whenever identifiability conditions of the BSI problem are not met. Simulation results demonstrate that the algorithm achieves significant improvements in terms of estimation accuracy and adaptability in both simulated and measured acoustic environments.**

***Index Terms*—blind system identification, distributed signal processing, multichannel signal processing, alternating direction method of multipliers, wireless sensor networks, quasi newton methods**

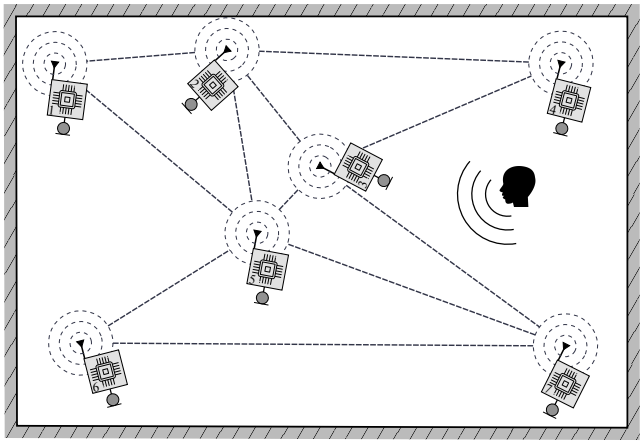


Fig. 1: A wireless acoustic sensor network within a room, where each sensor node comprises a microphone, a processing unit, and a transmission unit. Also shown is a speaker emitting sound into the room, i.e., the system input. The microphones receive the sound and each sensor signal represents a system output.

I. Introduction

WIRELESS sensor networks (WSN) are becoming more relevant in various technical fields, driven by the increase in network-connected devices and the growing Internet of Things (IoT). They find use in various applications such as environmental monitoring, industrial automation, and smart cities. They are also gaining traction in acoustic and audio signal processing applications. Some of the first proposals for networked audio sensors, such as [1], and [2], started the trend, and now, with the advent of smart homes and smart cities, the potential of using network-connected devices with built-in microphones and loudspeakers, such as phones, tablets and more is ever-increasing. Generally, devices described as "smart" in the current marketing language possess the functionality necessary to be used within a wireless acoustic sensor network (WASN); however, state-of-the-art devices do not use this full potential. Therefore, a valuable research direction is still to enable these devices' distributed signal processing capabilities fully. In the context of the latter, distributed signal processing tasks such as, e.g., distributed signal estimation [3], [4], noise control and echo cancellation [5], [6], and beamforming [6], [7] have been studied in recent years.

Blind system identification (BSI) refers to estimating the impulse responses of an unknown system without knowing the input signal, i.e., only the output signals. More specifically, we consider the single-input-multiple-output (SIMO) room acoustic system. In this case, the term system refers to the combination of the room, the microphones, and the source with their respective positions, which results in room impulse responses (RIR). They describe the acoustic path the sound waves take from the source to the microphone, which comprises the direct path - if existent - and reflections from the room boundaries and objects within the room and their effects. The task of BSI is to estimate these RIRs or RTFs, which can be used for signal enhancement tasks such as dereverberation, source separation, or acoustic echo cancellation or for room acoustics analysis, such as the computation of room acoustic parameters, such as the reverberation time RT60 or speech

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of the SOUNDS European Training Network. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369. This research received funding in part from the Research Foundation - Flanders (FWO) grant 12ZD622N as well as from the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (No. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. Source code available at https://github.com/SOUNDS-RESEARCH/XXXXXXXX

Manuscript received XXXX XX, 2023; revised XXXX XX, 2023.

intelligibility predictors C_{50} [8], room geometry estimation [9], [10], source localisation, or sound field analysis.

The BSI problem itself is a well-known problem and has been studied extensively in the literature. It was introduced in [11], and various algorithms have been proposed since. Early algorithms used higher-order statistics [12]–[14] for channel estimation; however, high computational complexity has motivated research into algorithms that only use second-order statistics. Such algorithms include the cross-relation (CR) algorithm [15], [16], subspace algorithms [17]–[20], and maximum-likelihood algorithms [21]. Out of the various multichannel BSI algorithms that have been proposed, adaptive cross-relation-based least-mean-squares (LMS) algorithms in the time and frequency domain are the ones most widely used. For instance, the normalized multichannel frequency-domain LMS (NMCFLMS) [22], [23] algorithm is an efficient algorithm utilising the fast Fourier transform (FFT) which has been extended to include constraints which improve robustness to noise and performance for acoustic impulse responses such as the noise-robust-NMCFLMS (RNMCFLMS) [24], l_p-RNMCFLMS [25] or phase-constrained-l_p-RNMCFLMS [26] algorithms. Further contributions are the multichannel quasi-Newton algorithm (MCQN) introduced in [27] and cross-correlation-based methods for early IR estimation [28]. Recently, [29] proposed a blind estimation algorithm for early-echo parameters in directional impulse responses. Significant contributions to BSI have been made by work such as [30], [31] analysing identifiability with near-common zeros and noise as well as state-space model-based approaches [32], [33] and Bayesian inference-based algorithms [34]–[40] partly applied in acoustic echo control, speech denoising, and dereverberation.

However, research into blind system identification in WASNs is limited, even more so in acoustics and audio signal processing applications. The work [41] proposes a distributed algorithm for estimating and equalising of room acoustics based on the common-acoustical pole (CAP) room model and the distributed averaging approach. Further, [42]–[44] introduce time-domain distributed gradient descent and recursive estimation algorithms for distributed BSI; however, they are not in an acoustic context and do not have an adaptive signal processing application in mind. Nonetheless, the distributed gradient descent-based (DGD) approach of [42] and the recursive approach in [44] based on the distributed stochastic approximation algorithm with expanding truncations (DSAAWET) [45] are taken into consideration in our study, implemented as adaptive algorithms. The algorithm proposed in [43] is intended for sparse channels and is not used as a direct comparison to our work. This paper extends our previous work [46]–[48], a distributed CR-based adaptive algorithm for BSI in WASNs.

In this paper, our contribution is to introduce a distributed algorithm for multichannel blind szstem identification in WASNs. We will refer to it as a ditributed adaptive alternating blind system identification (DABSI) algorithm. The BSI problem is formulated as a modified general-form consensus optimisation problem and solved in a distributed manner by the network nodes. For that, the alternating direction method of multipliers (ADMM) [49] is applied. This method is known to converge under mild conditions and is widely used in distributed optimisation problems. Our problem formulation separates the multichannel BSI problem into multiple lower-dimensional sub-problems with subsets of channels according to the sensor network's topology. The inter-channel cross-relations of the BSI problem are spread over nodes, such that each solves part of the original problem, i.e., each node solves a sub-problem using data from its network neighbours, and for each of the system's channels, multiple nodes form a consensus estimate. The ADMM update steps are applied in a block processing scheme forming an adaptive algorithm called Online-ADMM [50]–[52] with local quasi-Newton updates based on a Broyden class method [53], [54]. Compared to our previous work, in this paper, we extend the distributed algorithm by a few concepts to increase its robustness and performance, especially in speech-processing applications. The non-stationarity of a speech signal is challenging in adverse acoustic environments, with noise spectrally masking parts of the signal. We introduce a voice-activity detector-based (VAD) extension, which introduces separate estimates for "signal-plus-noise" and "noise-only" parts of the signal. These are then used to increase the estimation accuracy by solving a generalized eigenvalue problem (GEVD) to find the parameter subspace. This concept is also used for distributed signal estimation in existing literature, such as [55], [56]. In [44], the concept of known noise characteristics is also used, making the comparison to our work interesting. Furthermore, we introduce a regularisation term to the distributed BSI problem to improve convergence behaviour. This regularisation term is based on prior knowledge of the system's channel responses, i.e., a covariance matrix of channel responses for a specific set of channels, similar to [57]. In this case, the covariance matrix can be of a rather general set of channel responses, e.g., rooms within a specific size range. We conduct simulation studies using simulated rooms and measurement data to evaluate the performance of the proposed algorithm.

The paper is structured as follows: In Sec. II, we introduce the signal model and the well-known cross-relation method for BSI, which we then extend into its distributed formulation in Sec. III. In Sec. IV, we discuss extensions to the distributed BSI problem. In Sec. V, we provide theoretical comments on the convergence and computational complexity of the distributed BSI algorithm. In Sec. VI, we present simulation studies to evaluate the algorithm's performance. Finally, in Sec. VII, we give our concluding remarks.

TABLE I: Notation

object	related notation
boldface letter	lower case: vector; upper case: matrix
calligraphic letter	upper case: ordered set
matrix $\mathbf{A}$	$\mathbf{A}^\top$: transpose; $\mathbf{A}^*$: hermitian transpose; $\mathbf{A}^{-1}$: inverse; $\mathrm{tr}(\mathbf{A})$: trace
vector $\mathbf{a}$	column vector; $\|\mathbf{a}\|$ Euclidean norm
set $\mathcal{A}$	$\|\mathcal{A}\|$: cardinality
operators	$\mathrm{E}\{\cdot\}$: expectation; $\mathrm{Re}\{\cdot\}$: real part
constants	$\imath$: imaginary unit

II. Cross-Relation Blind System Identification

A. Signal Model

As our signal model for the system identification problem, we consider a single-input multiple-output (SIMO) system with M output signals and let $\mathcal{M} \triangleq (1, ..., M)$ be the ordered set of output channels. The observed discrete time signal $x_i[n]$ with discrete-time index n is modelled by the convolution of the system's unknown input signal $s[n]$ with the impulse response $h_i[n]$ defined by

$$x_i[n] = \sum_{m=0}^{L-1} h_i[m]s[n-m] + v_i[n], \quad i \in \mathcal{M}, \tag{1}$$

where L is the impulse response length, and $v_i[n]$ is additive noise with covariance matrix $\mathbf{\Sigma}_{\mathrm{v}}$. We will formulate the problem in a frame-based context; therefore, let τ be the frame index and define the discrete-time signal vector $\bar{\mathbf{x}}_i^{(\tau)} \in \mathbb{R}^{2L}$ as

$$\bar{\mathbf{x}}_i^{(\tau)} = \begin{bmatrix} x_i[\tau L] & x_i[\tau L-1] & \dots & x_i[\tau L-(2L-1)] \end{bmatrix}^\top \tag{2}$$

with a hop size of L samples. Let

$$\mathbf{x}_i^{(\tau)} = \mathbf{F}_{2L}\bar{\mathbf{x}}_i^{(\tau)} \tag{3}$$

be the DFT domain representation of the signal frame, where $\mathbf{F}_{2L} \in \mathbb{C}^{2L\times 2L}$ is the discrete Fourier transform (DFT) matrix for size $2L$. Under the assumption that the impulse responses are causal and finite (FIR) with a known length of L samples, we define the impulse response vector $\bar{\mathbf{h}}_i \in \mathbb{R}^L$ in the discrete-time domain as

$$\bar{\mathbf{h}}_i = \begin{bmatrix} h_i[0] & \dots & h_i[L-1] \end{bmatrix}^\top, \tag{4}$$

and analogously, we define its DFT as

$$\mathbf{h}_i = \mathbf{F}_L\bar{\mathbf{h}}_i, \tag{5}$$

where $\mathbf{F}_L \in \mathbb{C}^{L\times L}$ is the DFT matrix for size L. In practice, we replace the DFT with the fast Fourier transform (FFT).

B. Cross-Relation Approach

The cross-relation (CR) approach for BSI aims to use only the set of output signals $x_i[n]$ of the system to identify the system, i.e., the impulse responses $h_i[n]$. They can be estimated by exploiting the relative channel information when more than one channel is available, and the identifiability conditions [16], [58] are satisfied. These conditions are:

(i) The channel transfer functions have no common zeros (i.e., the polynomials are not co-prime)
(ii) The covariance matrix of the input signal $s[n]$ is of full rank (i.e., the number of modes $\geq 2L+1$).

We want to comment on two types of ambiguity in the BSI problem: (i) the scaling ambiguity, where any arbitrary factor a can be multiplied with the impulse responses $h_i[n]$ and its reciprocal factor $1/a$ with the input signal while the output signals remain unchanged. Because the input signal is unknown, the impulse responses can only be estimated up to a scaling factor. Furthermore (ii), under the assumption that the impulse responses $h_i[n]$ are causal, the impulse responses can be shifted in time $h_i[n-n_0]$ simultaneously, i.e., preserving the relative time shifts between IRs, while the input signal is shifted by the same in the opposite direction $s[n+n_0]$. Without knowledge of the input signal, the system is only identifiable as a causal system, where the first non-zero time instant of all impulse responses is at $n=0$, which in the room-acoustic context corresponds to the direct path of the microphone that is closest to the source.

From the cross-convolved signal [23]

$$y_{ij}[n] \triangleq \sum_{m=0}^{L-1} w_j[n]x_i[n-m], \quad i,j \in \mathcal{M}, \tag{6}$$

where $w_j[n]$ is the estimate (model) of the impulse response $h_j[n]$ we can see that

$$y_{ij}[n] = y_{ji}[n], \quad i,j \in \mathcal{M}, i \neq j, \tag{7}$$

which is the fundamental equality of the cross-relation approach. To formulate this in the frame-based context, we denote (7) in vector form

$$\mathbf{y}_{ij}^{(\tau)} = \mathbf{y}_{ji}^{(\tau)}, \quad i,j \in \mathcal{M}, i \neq j \tag{8}$$

where the vector $\mathbf{y}_{ij}^{(\tau)} \in \mathbb{C}^L$ is the frequency-domain representation of the cross-convolved signal defined as the result of the overlap-save convolution

$$\mathbf{y}_{ij}^{(\tau)} = \mathbf{X}_i^{(\tau)}\mathbf{w}_j^{(\tau)} \tag{9}$$

with the data matrix $\mathbf{X}_i^{(\tau)} \in \mathbb{C}^{L\times L}$ given by

$$\mathbf{X}_i^{(\tau)} = \mathbf{W}_{01}\mathbf{D}_i^{(\tau)}\mathbf{W}_{10}. \tag{10}$$

The diagonal matrix

$$\mathbf{D}_i^{(\tau)} = \mathrm{diag}\left\{\mathbf{x}_i^{(\tau)}\right\} \tag{11}$$

contains the DFT transform of the sensor signal vector (3) on its diagonal and

$$\mathbf{W}_{01} = \mathbf{F}_L\begin{bmatrix}\mathbf{0}_L & \mathbf{I}_L\end{bmatrix}\mathbf{F}_{2L}^{-1} \in \mathbb{C}^{L\times 2L}, \tag{12a}$$

$$\mathbf{W}_{10} = \mathbf{F}_{2L}\begin{bmatrix}\mathbf{I}_L & \mathbf{0}_L\end{bmatrix}^\top\mathbf{F}_L^{-1} \in \mathbb{C}^{2L\times L} \tag{12b}$$

are the frequency-domain overlap-save matrices. We define the cross-relation error as

$$\mathbf{e}_{ij}^{(\tau)} = \mathbf{y}_{ij}^{(\tau)} - \mathbf{y}_{ji}^{(\tau)}, \quad i,j \in \mathcal{M}, i \neq j \tag{13}$$

which we then can put in matrix form

$$\mathbf{e}^{(\tau)} = \mathbf{X}^{(\tau)}\mathbf{w}^{(\tau)}, \tag{14}$$

where $\mathbf{X}^{(\tau)} \in \mathbb{C}^{ML\times ML}$ is defined by

$$\mathbf{X}^{(\tau)} \triangleq \begin{bmatrix} \sum_{i\neq 1}\mathbf{X}_i^{(\tau)} & -\mathbf{X}_2^{(\tau)} & \cdots & -\mathbf{X}_M^{(\tau)} \\ -\mathbf{X}_1^{(\tau)} & \sum_{i\neq 2}\mathbf{X}_i^{(\tau)} & \cdots & -\mathbf{X}_M^{(\tau)} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{X}_1^{(\tau)} & -\mathbf{X}_2^{(\tau)} & \cdots & \sum_{i\neq M}\mathbf{X}_i^{(\tau)} \end{bmatrix} \tag{15}$$

and $\mathbf{w} = \begin{bmatrix}\mathbf{w}_1^\top & \dots & \mathbf{w}_M^\top\end{bmatrix}^\top$ is a stacked vector of channel frequency responses. Note that in the noiseless case, $\mathbf{e}^{(\tau)} = \mathbf{0}$ and the equality (14) becomes a null space problem. The null space of the data matrix $\mathbf{X}^{(\tau)}$ is one-dimensional, and the

solution is unique up to a scalar factor. Due to the unavoidable presence of noise in acoustic environments, the data matrix $\mathbf{X}^{(\tau)}$ might be full rank, i.e., it has no null space, where the best estimate of the impulse responses is the eigenvector corresponding to the smallest eigenvalue. A straightforward approach to arrive at the estimate would be to compute an eigenvalue decomposition and select the corresponding eigenvector. However, the potentially high dimensionality of the problem and the goal of having an adaptive frame-based estimation algorithm make it a more appropriate approach to find an estimate by looking at it from an optimization perspective and solving a quadratic minimization problem with first-order methods. The problem is then formulated as

$$\begin{aligned} \min_{\mathbf{w}} \quad & f(\mathbf{w}) \\ \text{s.t.} \quad & \|\mathbf{w}\| = 1. \end{aligned} \tag{16}$$

with the objective function

$$f(\mathbf{w}) = \|\mathbf{X}^{(\tau)}\mathbf{w}\|^2 = \mathbf{w}^* \mathbf{C}^{(\tau)} \mathbf{w} \tag{17}$$

with

$$\mathbf{C}^{(\tau)} = \bar{\mathbf{X}}^{(\tau)*} \mathbf{X}^{(\tau)}, \tag{18}$$

which we will call the cross-relation (CR) matrix. The constraint $\|\mathbf{w}\| = 1$ is introduced to avoid the trivial solution $\mathbf{w} = 0$. It is to be noted that the eigenvectors of the CR matrix are the same as the eigenvectors of the data matrix $\bar{\mathbf{X}}^{(\tau)}$. We estimate a smoothed CR matrix by the recursive average

$$\mathbf{C}^{(\tau)} = \eta\, \mathbf{C}^{(\tau-1)} + (1-\eta)\, \mathbf{X}^{(\tau)*} \mathbf{X}^{(\tau)}, \tag{19}$$

where the real scalar $0 \ll \eta < 1$ is an exponential forgetting factor.

III. Distributed BSI using O-ADMM

A. Sensor Network

We consider a wireless acoustic sensor network (WASN) with M sensor nodes, as shown in Fig. 1, where we define a sensor node as a device comprising a single microphone, a processing unit, and a transmission unit. We will use the set $\mathcal{M}$ defined in Sec. II-B to refer to the set of nodes, each corresponding to one of the channels of the acoustic system. In section Sec. II-B, we introduced the cross-relation problem formulation. In a network context, this would involve a centralized algorithm where a central processing unit collects all signals $x_i[n]$ and solves the CR problem in the form of (16) with signals from M channels. This, on the one hand, necessitates the transmission of all signals to this central processing unit; on the other hand, it requires the central processing unit to solve a high-dimensional problem. In the distributed context, nodes solve lower-dimensional problems while sharing information with other nodes, i.e., a subset of all nodes, optimally arriving at the same global solution as the centralized algorithm. Let us consider two scenarios:

1) Neighbourhoods: each node transmits to a subset of nodes.
2) Broadcasting: each node transmits to all other nodes.

For the distributed cross-relation method, we will assume that in the broadcasting scenario, nodes only use the subset of signals relevant to their respective sub-problems. Let us define the graph of this network as $\mathcal{G} \triangleq \{\mathcal{M}, \mathcal{E}\}$, where $\mathcal{E}$ is the set of edges (i,j) connecting the channels/nodes i and j. We consider the edges as directed, i.e., the edge (i,j) represents the signal transmission from node i to node j, and define the sets of the transmit-neighbourhood $\mathcal{T}_i \subseteq \mathcal{M}$ and receive neighbourhood $\mathcal{R}_i \subseteq \mathcal{M}$ of node i as

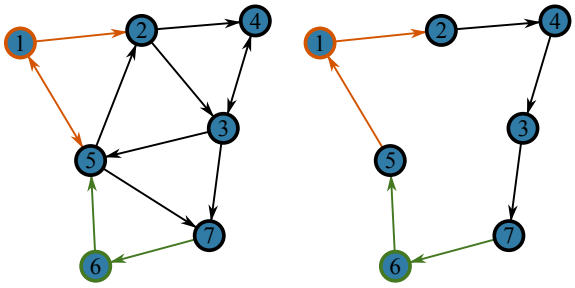


Fig. 2: Topology of channels and representation of sparse selection of cross relations. We assume that each node has an edge to itself, i.e., (i,i), and omit this in the graph representation.

$$\mathcal{T}_i = (o_1, ..., o_{T_i}) = (j \in \mathcal{M} \mid (i,j) \in \mathcal{E}), \tag{20a}$$

$$\mathcal{R}_i = (r_1, ..., r_{R_i}) = (j \in \mathcal{M} \mid (j,i) \in \mathcal{E}). \tag{20b}$$

ordered according to their positions in the ordered set $\mathcal{M}$. We denote the size of these sets as $T_i \triangleq |\mathcal{T}_i|$ and $R_i \triangleq |\mathcal{R}_i|$. Note that if $\mathcal{T}_i = \mathcal{R}_i$ for all $i \in \mathcal{M}$, the graph is equivalent to an undirected graph. Fig. 2 (left) and Fig. 2 (right) show examples of such network topologies, one densely, the other sparsely connected. For the remainder of this work, we will assume that the inter-node communication is error-free and instantaneous.

B. Online-ADMM

We formulate the distributed BSI problem as a modified general-form consensus problem [49], where the cost function is separable into local cost functions, and the variables are constrained to each other through consensus variables. Consequently, multiple nodes locally have estimates of the same variable, i.e., channel response estimates. The notation is best explained with an example: Let a variable v represent some quantity, e.g., an estimate. This variable is available at node i and node j, which has a copy of it. The variable is denoted $v_{i|i}$ at node i and $v_{j|i}$ at node j, respectively. In short, the subscript before the vertical line denotes the owner node (of the copy) and the subscript after the vertical line denotes the node it refers to.

We define the optimization problem as a modified general-form consensus problem [49]

$$\begin{aligned} \min_{\{\mathbf{w}_i, \mathbf{z}_{i|i}\}_{i\in\mathcal{M}}} \quad & \sum_{i\in\mathcal{M}} \left(f_i(\mathbf{w}_i) + g_i(\mathbf{w}_i)\right) \\ \text{s.t.} \quad & \mathbf{A}_{i|j}\mathbf{w}_i = \mathbf{B}_i \mathbf{z}_{j|j} \quad \forall\, i \in \mathcal{M}, j \in \mathcal{R}_i, \\ & \mathbf{z}_{i|i} \in \mathcal{C}_i \qquad\quad \forall\, i \in \mathcal{M}. \end{aligned} \tag{21}$$

The local primal variable $\mathbf{w}_i \in \mathbb{C}^{R_i L}$ is a stacked vector of channel response estimates

$$\mathbf{w}_i = \begin{bmatrix} \mathbf{w}_{i|r_1}^\top & \cdots & \mathbf{w}_{i|r_{R_i}}^\top \end{bmatrix}^\top \tag{22}$$

of all channels in the receive-neighbourhood $\mathcal{R}_i$ of node i, see (20b). Then, the variable $\mathbf{z}_{i|i} \in \mathbb{C}^L$ is the consensus estimate and $\mathbf{A}_{i|j}$ is the selection matrix for edge (i,j), mapping corresponding local estimates to the consensus estimate, with applied alignment matrix $\mathbf{B}_i$. The alignment matrix is necessary to compensate for time-shifts in the estimates due to time-shift ambiguity in the network. We discuss this in Sec. III-D. The objective function is the sum of local cost functions $f_i(\mathbf{w}_i)$, CR sub-problems using the channels with $j \in \mathcal{R}_i$ and regularisation terms $g_i(\mathbf{w}_i)$, which we describe in Sec. IV-B. The constraints on the consensus variables $\mathbf{z}_{i|i}$ are defined by the feasibility sets $\mathcal{C}_i$, a generalization of the non-triviality constraint in (16). This generalization is necessary in the distributed setting, and we discuss the reasoning and implications in Sec. III-C.

We formulate the real-valued augmented Lagrangian of (21) as

$$\mathcal{L}_\rho = \sum_{i \in \mathcal{M}} \left(f_i(\mathbf{w}_i) + g_i(\mathbf{w}_i) + r_i(\mathbf{w}_i) \right) \tag{23}$$

where the augmented term is

$$\begin{aligned} r_i(\mathbf{w}_i) = \sum_{j \in \mathcal{R}_i} & 2\,\mathrm{Re}\left\{ \mathbf{u}_{i|j}^* \left(\mathbf{A}_{i|j}\mathbf{w}_i - \mathbf{B}_i \mathbf{z}_{i|j} \right) \right\} \\ & + \frac{\rho}{2} \sum_{j \in \mathcal{R}_i} \left\| \mathbf{A}_{i|j}\mathbf{w}_i - \mathbf{B}_i \mathbf{z}_{i|j} \right\|^2 + \mathcal{I}_{\mathcal{C}_i}(\mathbf{z}_{i|i}). \end{aligned} \tag{24}$$

$\mathbf{u}_{i|j} \in \mathbb{C}^L$ is the dual variable associated with the consensus constraint for edge (i,j) and $\mathcal{I}_{\mathcal{C}_i}(\cdot)$ is the indicator function:

$$\mathcal{I}_{\mathcal{C}_i}(\mathbf{z}_{i|i}) = \begin{cases} 0 & \text{if } \mathbf{z}_{i|i} \in \mathcal{C}_i \\ \infty & \text{otherwise.} \end{cases} \tag{25}$$

Note that the non-triviality constraints as introduced here could also be on the local estimates $\mathbf{w}_{i|i}$. However, decoupling the constraints through the consensus estimates keeps the local cost functions smooth. The penalty parameter $\rho > 0$ is a scalar constant.

We denote the alternating minimization-maximization steps as [49], [59]

$$\mathbf{w}_i^{(\tau+1)} = \arg\min_{\mathbf{w}_i} \mathcal{L}_\rho \tag{26a}$$

$$\mathbf{z}_{i|i}^{(\tau+1)} = \Pi_{\mathcal{C}_i} \left(\arg\min_{\mathbf{z}_{i|i}} \mathcal{L}_\rho \right) \tag{26b}$$

$$\mathbf{u}_{i|j}^{(\tau+1)} = \arg\max_{\mathbf{u}_{i|j}} \mathcal{L}_\rho. \tag{26c}$$

We then define the update steps as

$$\mathbf{w}_i^{(\tau+1)} = \mathbf{w}_i^{(\tau)} - \mu \mathbf{V}^{(\tau)} \mathbf{g}_i^{(\tau)}(\mathbf{w}_i^{(\tau)}) \tag{27a}$$

$$\mathbf{y}_{i|j}^{(\tau+1)} = \mathbf{B}_i^* \left(\mathbf{A}_{i|j}\mathbf{w}_i^{(\tau+1)} + \rho^{-1}\mathbf{u}_{i|j}^{(\tau)} \right) \tag{27b}$$

$$\mathbf{y}_{j|i}^{(\tau+1)} = \mathbf{y}_{i|j}^{(\tau+1)} \tag{27c}$$

$$\mathbf{z}_{i|i}^{(\tau+1)} = \Pi_{\mathcal{C}_i} \left(\frac{1}{T_i} \sum_{j \in \mathcal{T}_i} \mathbf{y}_{i|j}^{(\tau+1)} \right) \tag{27d}$$

$$\mathbf{z}_{j|i}^{(\tau+1)} = \mathbf{z}_{i|i}^{(\tau+1)} \tag{27e}$$

$$\mathbf{u}_{i|j}^{(\tau+1)} = \mathbf{u}_{i|j}^{(\tau)} + \rho \left(\mathbf{A}_{i|j}\mathbf{w}_i^{(\tau+1)} - \mathbf{B}_i \mathbf{z}_{i|j}^{(\tau+1)} \right). \tag{27f}$$

The update steps are derived from the augmented Lagrangian, taking the gradients w.r.t. the primal variables and Lagrange multipliers. We use a quasi-Newton step based on the BFGS method [53], [54] with step size $\mu > 0$, for solving the local minimization problem (27a) with the gradient

$$\begin{aligned} \mathbf{g}_i^{(\tau)}(\mathbf{w}_i) = \nabla_{\mathbf{w}_i^*} \mathcal{L}_\rho^{(\tau)} = \nabla_{\mathbf{w}_i^*} f_i(\mathbf{w}_i) + \nabla_{\mathbf{w}_i^*} g_i(\mathbf{w}_i) \\ + \sum_{j \in \mathcal{R}_i} \left[\mathbf{A}_{i|j}^* \left(\mathbf{u}_{i|j} + \rho \left(\mathbf{A}_{i|j}\mathbf{w}_i - \mathbf{B}_i \mathbf{z}_{i|j} \right) \right) \right] \end{aligned}$$

and $\mathbf{V}^{(\tau)}$ is the estimated inverse Hessian according to the BFGS method for which we state the update equations (28a) - (28d), which can be implemented efficiently. Due to the potentially badly conditioned problem, this quasi-Newton method is more stable than a pure gradient descent while being computationally cheaper than a full Newton update step, which involves a matrix inversion. Further, we define the variable $\mathbf{y}_{i|j}^{(\tau)}$ (27b) as the local combination of the local estimate and Lagrange multiplier, which is then transmitted (copied) along the edge (i,j) (27c). The consensus update (27d) is an element-wise averaging of the aligned local variables within the transmit-neighbourhood, with subsequent transmission of consensus variables back along the edge. Ultimately, the dual update (27f) is a gradient ascent step.

The iteration index is denoted by (τ), which in this context also refers to the frame index of our adaptive system. This transforms this problem into an online optimization problem, where the estimates are adapted to changing data terms, i.e., similar to an iterative algorithm such as the stochastic gradient descent and the online quasi-Newton method in [27]. We refer the reader to literature such as [51], [52], [60] for details on distributed optimization in an online setting.

C. Distributed Non-Triviality Constraint

The optimization problem (21) includes a non-triviality constraint in a more general form than the one in (16). This section will discuss the implications of a network-wide non-triviality constraint, which must only be enforced with local-neighbourhood information. We introduced the concept in previous work [47]; here, we reiterate the main points and make minor modifications to the ordering of the update steps and the computation of the adaptive mixing factor.

$$\Delta \mathbf{w}_i^{(\tau+1)} = \mathbf{w}_i^{(\tau+1)} - \mathbf{w}_i^{(\tau)} \tag{28a}$$

$$\Delta \mathbf{g}^{(\tau+1)} = \mathbf{g}_i^{(\tau+1)}(\mathbf{w}_i^{(\tau+1)}) - \mathbf{g}_i^{(\tau+1)}(\mathbf{w}_i^{(\tau)}) \tag{28b}$$

$$\gamma^{(\tau+1)} = \Delta \mathbf{g}^{(\tau+1)*} \Delta \mathbf{w}_i^{(\tau+1)} \tag{28c}$$

$$\mathbf{V}^{(\tau+1)} = \left(\mathbf{I} - \frac{\Delta \mathbf{w}_i^{(\tau+1)} \Delta \mathbf{g}^{(\tau+1)*}}{\gamma^{(\tau+1)}} \right) \mathbf{V}^{(\tau)} \left(\mathbf{I} - \frac{\Delta \mathbf{g}^{(\tau+1)} \Delta \mathbf{w}_i^{(\tau+1)*}}{\gamma^{(\tau+1)}} \right) + \frac{\Delta \mathbf{w}_i^{(\tau+1)} \Delta \mathbf{w}_i^{(\tau+1)*}}{\gamma^{(\tau+1)}} \tag{28d}$$

We consider the orthogonal projection of the stacked vector $\mathbf{z} = \begin{bmatrix} \mathbf{z}_{1|1}^\top & \cdots & \mathbf{z}_{M|M}^\top \end{bmatrix}^\top$ onto the unit hypersphere in the parameter space $\mathbb{R}^{ML}$, i.e.,

$$\mathcal{C} = \left\{ \mathbf{z} \mid \|\mathbf{z}\|^2 = 1 \right\}. \tag{29}$$

The projection operator is

$$\Pi_{\mathcal{C}}(\mathbf{z}) = \frac{\mathbf{z}}{\|\mathbf{z}\|} = \frac{\mathbf{z}}{\sum_{i \in \mathcal{M}} \|\mathbf{z}_{i|i}\|^2}. \tag{30}$$

We separate the constraint into M individual constraints, as denoted in (21), where each feasible set

$$\mathcal{C}_i = \left\{ \mathbf{z}_{i|i} \mid \textstyle\sum_{j \in \mathcal{M}} \|\mathbf{z}_{j|j}\|^2 = 1 \right\}, \tag{31}$$

is the intersection of the hypersphere (29) and the parameter subspace $\mathbb{R}^L$ associated with $\mathbf{z}_{i|i}$. This intersection is a hypersphere within the subspace with a radius such that $\|\mathbf{z}_{i|i}\|^2 = 1 - \sum_{j \in \mathcal{M}, j \neq i} \|\mathbf{z}_{j|j}\|^2$. The projection operator (30) is not easily computable on each node independently, as the sum requires all $j \in \mathcal{M}$, which requires transmitting squared-norm values from each node to all other nodes, i.e., a fully connected network or broadcasting setup, which we generally do not assume to be the case. The chosen method to avoid this is to replace the denominator in (30) with an estimate such that

$$\tilde{\Pi}_{\mathcal{C}_i}^{(\tau)}(\mathbf{z}_{i|i}) = \frac{\mathbf{z}_{i|i}}{\sqrt{M \phi_i^{(\tau)}}}, \tag{32}$$

where

$$\phi_i^{(\tau)} \approx \frac{1}{M} \sum_{j \in \mathcal{M}} \|\mathbf{z}_{j|j}\|^2 \tag{33}$$

estimates the average of the squared norms of the channel response estimates. For this, we use an adaptive distributed averaging approach. Define

$$\boldsymbol{\eta}_i^{(\tau)} = \left[\{ \|\mathbf{z}_{j|j}^{(\tau)}\|^2 \}_{j \in \mathcal{T}_i} \right]^\top, \tag{34}$$

$$\boldsymbol{\phi}_i^{(\tau)} = \left[\{ \phi_j^{(\tau)} \}_{j \in \mathcal{T}_i} \right]^\top \tag{35}$$

which is the vector of squared norms of the current channel response estimates and the squared-norm estimates of the neighbours of node i, respectively. The adaptive distributed averaging update step is

$$\phi_i^{(\tau)} = \mathbf{d}^\top \left(\gamma_i^{(\tau)} \boldsymbol{\eta}_i^{(\tau)} + (1 - \gamma_i^{(\tau)}) \boldsymbol{\phi}_i^{(\tau-1)} \right) \tag{36}$$

where the vector $\mathbf{d}$ is a vector of weights computed by the fastest distributed linear averaging (FDLA) algorithm [61] and $\gamma_i^{(\tau)}$ is an adaptive forgetting factor between the instantaneous squared norm values and the previous distributed averaging estimate. We define the adaptive forgetting factor as

$$\gamma_i^{(\tau)} = \frac{\|\boldsymbol{\eta}_i^{(\tau)} - \boldsymbol{\eta}_i^{(\tau-1)}\|}{\|\boldsymbol{\eta}_i^{(\tau)}\| + \|\boldsymbol{\eta}_i^{(\tau-1)}\|} \tag{37}$$

where $0 \leq \gamma_i^{(\tau)} \leq 1$ is guaranteed by positive semi-definiteness of the norm and the triangle inequality on $\mathbb{C}^n$. Compared to [47], where the value had to be limited to $\gamma_i^{(\tau)} \leq 1$ explicitly, here it comes implicitly. It is easy to see that $\gamma_i^{(\tau)} \to 0$ if $\boldsymbol{\eta}_i^{(\tau)} - \boldsymbol{\eta}_i^{(\tau-1)} \to 0$, which is the case when the norms of the estimates are converging, which we assume to be the case when the estimates themselves are converging. Therefore, assuming convergence of response estimates, all squared-norm estimates at each node will converge to the same value, which is 1 due to the iterative projection

$$\lim_{(\tau) \to \infty} \|\mathbf{z}_{i|i}^{(\tau)}\|^2 = \lim_{(\tau) \to \infty} \phi_i^{(\tau)} = 1, \quad \forall i \in \mathcal{M}. \tag{38}$$

This approach allows us to enforce the network-wide non-triviality constraint adaptively in a distributed manner while only using information exchange within the node neighbourhoods.

D. Time-Shift Compensation

The time-shift ambiguity, as discussed in Sec. II-B, i.e., the inability to estimate leading zeros, gives rise to the misalignment of estimates of the same impulse response at different nodes. Fig. 3 shows a simple example of this in the time domain, from which it should be clear that finding a consensus between misaligned estimates is more than simple averaging. The estimation of delays of the sensor signals at each node is a problem that can be approached by various methods, such as cross-correlation, generalized cross-correlation, or phase difference-based methods. However, detailed discussion of these methods is beyond the scope of this paper. For this work, we assume that the delays are estimated or known at each node.

Under the assumption that the respective delay values $d_i^{(\tau)}$ are estimated or known at each node, we can compensate for the misalignment during the computation of the consensus estimate. In the DFT domain, the misalignment is equivalent to a phase shift, denoted for node i and frequency component k as

$$\theta_{i,k}^{(\tau)} = \frac{2\pi k}{L} d_i^{(\tau)}, \tag{39}$$

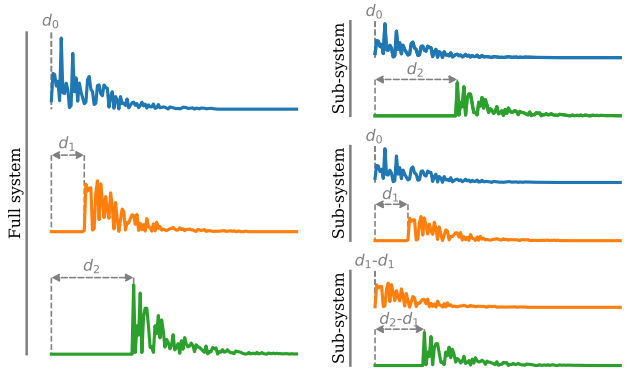


Fig. 3: The time-shift between sub-problem estimates. The time-shift ambiguity leads to misalignment of the impulse response estimates at nodes whose channel subsets have different initial arrival times.

The alignment matrix $\mathbf{B}_i$ applies the inverse phase shift to the estimates of the impulse responses at node i before the combination step and is defined as the diagonal matrix

$$\mathbf{B}_i^{(\tau)} = \operatorname{diag}\left\{\exp\left(-\imath\theta_{i,k}^{(\tau)}\right)\right\}_{k=0}^{L-1}, \tag{40}$$

where $\imath$ is the imaginary unit. Because it is diagonal, applying the alignment matrix is a simple element-wise multiplication in practice.

IV. Sub-problem extensions

A. Generalized Eigenvalue Problem

We assume that the source signal $s(t)$ has *"signal-plus-noise"* periods and *"noise-only"* periods, as would be the case for human speech. This approach requires a voice activity detection (VAD) algorithm to separate these periods. We assume that this algorithm is available and yields a binary VAD signal. Given this, it is a well-known approach to estimate covariance matrices for these periods separately and use this *"noise-only"* information to improve the algorithm's performance [55], [56], [62], [63].

Omitting the node index i, we introduce the *"signal-plus-noise"* CR matrix $\mathbf{C}_\mathrm{x}$ and the *"noise-only"* CR matrix $\mathbf{C}_\mathrm{v}$ as defined in (15). The matrices are estimated recursively, where each estimate is updated only in its respective VAD state:

$$\mathbf{C}_\mathrm{x}^{(\tau+1)} = \begin{cases} \eta\,\mathbf{C}_\mathrm{x}^{(\tau)} + (1-\eta)\mathbf{X}^{(\tau)*}\mathbf{X}^{(\tau)}, & \text{if VAD} = 1 \\ \mathbf{C}_\mathrm{x}^{(\tau)}, & \text{otherwise} \end{cases} \tag{41}$$

$$\mathbf{C}_\mathrm{v}^{(\tau+1)} = \begin{cases} \eta\,\mathbf{C}_\mathrm{v}^{(\tau)} + (1-\eta)\mathbf{X}^{(\tau)*}\mathbf{X}^{(\tau)}, & \text{if VAD} = 0 \\ \mathbf{C}_\mathrm{v}^{(\tau)}, & \text{otherwise.} \end{cases} \tag{42}$$

Let us consider the theoretical case where we have perfect knowledge of both matrices. For our signal model (1), under the assumption that noise and signal are independent, the CR matrix of the *clean* signal is

$$\mathbf{C}_\mathrm{s} = \mathbf{C}_\mathrm{x} - \mathbf{C}_\mathrm{v}. \tag{43}$$

The initial BSI problem in the noiseless case, restated here, gives us

$$\mathbf{C}_\mathrm{s}\mathbf{w} = (\mathbf{C}_\mathrm{x} - \mathbf{C}_\mathrm{v})\,\mathbf{w} = 0, \tag{44}$$

which is the generalized eigenvalue problem

$$\mathbf{C}_\mathrm{x}\mathbf{w} = \lambda\mathbf{C}_\mathrm{v}\mathbf{w} \tag{45}$$

with the smallest generalized eigenvalue $\lambda = 1$.

A full generalized eigenvalue decomposition (GEVD) is computationally expensive; however, since we are only interested in the smallest generalized eigenpair, we minimize the generalized Rayleigh quotient, which yields the minimal generalized eigenvalue as the minimum and eigenvector as the minimizer. We denote this minimization problem as

$$\begin{aligned} \min_{\mathbf{w}} \quad & f_\mathrm{GEV}(\mathbf{w}) \\ \text{s.t.} \quad & \|\mathbf{w}\| = 1 \end{aligned} \tag{46}$$

with the quotient

$$f_\mathrm{GEV}(\mathbf{w}) = \frac{\mathbf{w}^*\mathbf{C}_\mathrm{x}\mathbf{w}}{\mathbf{w}^*\mathbf{C}_\mathrm{v}\mathbf{w}}. \tag{47}$$

Considering (43) and that $\mathbf{C}_\mathrm{s}$ has a one-dimensional null space, i.e., its smallest eigenvalue is 0, this extension does not introduce bias. We assume the objective function $f_\mathrm{GEV}(\mathbf{w})$ to be analytic in both $\mathbf{w}$ and $\mathbf{w}^*$. The gradient wrt. $\mathbf{w}^*$ is then

$$\nabla_{\mathbf{w}^*} f_\mathrm{GEV}(\mathbf{w}) = c(\mathbf{w})\,(\mathbf{C}_\mathrm{x} - f_\mathrm{GEV}(\mathbf{w})\mathbf{C}_\mathrm{v})\,\mathbf{w} \tag{48}$$

where $c(\mathbf{w})$ is a real-valued scalar function. We disregard this scalar function, assuming it is sufficient to have a fixed gradient step size and look at the gradient direction

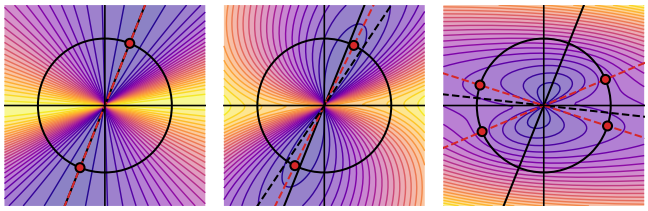
$$(\mathbf{C}_\mathrm{x} - f_\mathrm{GEV}(\mathbf{w})\mathbf{C}_\mathrm{v})\,\mathbf{w}. \tag{49}$$

The modification of the original update step (27a) is straightforward: We replace the gradient direction (28) with (49).

B. Regularization based on prior information

Here, we introduce a regularization term as foreshadowed in (21). The goal is to introduce prior information about the channel responses in order to mitigate the ill-posedness of the BSI problem when the identifiability conditions are not fully met. This appears when the input signal is not sufficiently exciting or when the impulse responses have common zeros [30], [31]. Due to the objective function not being a likelihood function, we cannot develop a proper Bayesian approach to include a prior distribution. However, we can construct a regularization term inspired by it: Consider a prior Normal distribution of channel responses in the DFT domain $\mathbf{h}_i \propto \mathcal{N}_{\mathbf{h}_i}(0, \mathbf{\Sigma}_{\mathbf{h}_i})$, where $\mathbf{\Sigma}_{\mathbf{h}_i} \in \mathbb{C}^{L\times L}$ is the covariance matrix - assumed known. This prior as a regularization term is consistent with the observations in [57] that the optimal regularizer comprises the inverse of the covariance matrix of the room impulse responses, here channel responses in the DFT domain. We define the regularization term as

$$g(\mathbf{w}) = \frac{\lambda}{2}\mathbf{w}^*\mathbf{\Sigma}_\mathbf{w}^{-1}\mathbf{w}, \tag{50}$$



(a) (b) (c)

Fig. 4: Illustration of the local objective function and the constraint. (a) shows the GEV case with $f_i(\mathbf{w}_i) = \frac{\mathbf{w}_i^* \mathbf{C}_\mathrm{x} \mathbf{w}_i}{\mathbf{w}_i^* \mathbf{C}_\mathrm{v} \mathbf{w}_i}$ and $g_i(\mathbf{w}_i) = 0$. (b) shows (a) with added non-zero regularizer $g_i(\mathbf{w}_i) > 0$. The black circle marks the hypershpere constraint, black line shows the signal-only null space, red dots are the local minima. (c) shows a bad regularizer that introduces additional local minima with large bias. Note: this plot shows a real-valued function for simplicity.

where $\lambda \in \mathbb{R}_+$ is the regularization parameter and $\mathbf{\Sigma}_{\mathrm{p},i} \in \mathbb{C}^{ML \times ML}$ is a block-diagonal matrix

$$\mathbf{\Sigma}_\mathbf{w} = \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{h}_i} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{\Sigma}_{\mathbf{h}_i} \end{bmatrix} \tag{51}$$

with as many blocks as channels in the (sub) problem, i.e., R_i. Given the non-stationary nature of the input signal and the resulting non-stationary power in $\mathbf{C}^{(\tau)}$, it becomes advantageous to employ an adaptive regularization parameter $\lambda^{(\tau)}$ to account for the varying scaling between the objective function and the regularization term. We found that defining the adaptive regularization parameter as

$$\lambda^{(\tau)} = \lambda_0 \frac{\mathrm{tr}(\mathbf{C}^{(\tau)})}{(M-1)\,\mathrm{tr}(\mathbf{\Sigma}_\mathbf{w}^{-1})} \tag{52}$$

is a sound choice, where $0 < \lambda_0 \ll 1$.

V. Theoretical Analysis

A. Convergence behaviour

In this section, we give some remarks on the convergence behaviour of the proposed algorithm. While we give no rigorous proof of convergence, we will give insight into the problem geometry and the properties of the objective function. Let us first consider the case of a static system, i.e., the channel responses do not change over time. Each local cost function $f_{i,\mathrm{GEV}}(\mathbf{w}_i) : \mathbb{C}^{R_i L} \to \mathbb{R}$ is a generalized Rayleigh quotient, and the local regularizer $g_i(\mathbf{w}_i) : \mathbb{C}^{R_i L} \to \mathbb{R}$ is a quadratic form. While the generalized Rayleigh quotient is a non-convex function, it exhibits favourable properties. Its minimization has been studied extensively [64]–[69], and it is known that the function constrained to the hypersphere $\|\mathbf{w}_i\| = c$ with $c > 0$ admits two local minima, which are both global [64], [65]. Fig. 4a shows this for a two-dimensional real-valued example; both minima $\mathbf{w}_i^{\mathrm{opt},1}, \mathbf{w}_i^{\mathrm{opt},2}$ are global minima, with $\mathbf{w}_i^{\mathrm{opt},1} = -\mathbf{w}_i^{\mathrm{opt},2}$ therefore, optimal. In this application's context, both solutions are equivalent estimates of the frequency responses up to a scalar factor a or $-a$, where a depends on the scale of signal and channel responses. This property is not necessarily preserved when a quadratic term, i.e., the regularizer $g_i(\mathbf{w}_i)$, is added to the generalized Rayleigh quotient, as shown in [70]. The sum of the two functions can admit local minima which are not global. However, if the choice of regularizer is appropriate, such that the minimal eigenvector of $\mathbf{\Sigma}^{-1}$ lines up with the null space of the CR matrix $\mathbf{C}_\mathrm{s}$, i.e., the prior information on channel responses is representative of the actual channel responses, then the local minima are still global. Fig. 4b and Fig. 4c show the effect of a "good" and a "bad" regularizer on the local minima. If the input signal does not lead to excitation in all frequency components or the frequency responses have common or near-common zeros (identifiability conditions), then the CR matrix has a null space or effective null space of dimension greater than one. In this case, the regularization is necessary. Further, the potential weak curvature of the generalized Rayleigh quotient leads to slow convergence with first order methods, which is why the DFP/BFGS method is used. While existing theoretical results on the convergence of the BFGS method are for convex static problems, practical results in [27] show that the method is also effective in an online setting.

The convergence of ADMM for convex functions is well-studied in the literature; however, in this case, the non-convexity of the objective functions and constraint sets does not allow us to apply those results directly. The analysis in [59, Corollary 2] shows that if the objective function is Lipschitz differentiable and the constraint set is compact, then for a sufficiently large ρ, the sequence $(\mathbf{w}_i^{(\tau)}, \mathbf{z}_{i|j}^{(\tau)}, \mathbf{u}_{i|j}^{(\tau)})_{\tau \in \mathbb{N}}$ has at least one limit point, of which each is a stationary point of the augmented Lagrangian. Assuming we have two global local minima, the stationary points are global minima. Therefore, even though ADMM has to be considered a local optimization method for non-convex problems and is usually not guaranteed to converge to a global minimum [49], the properties of the objective function and its local minima suggest that the algorithm converges to a global minimum. Further, considering the local cost functions could converge to either of the local minima, the consensus constraint on local estimates will force them to converge to equivalent local minima (the only scenario where this is not the case is when there is perfect symmetry in the local cost functions as well as the local estimate initialization, which is extremely unlikely).

In numerical simulations, with large enough ρ, the proposed algorithm shows convergence to minima that correspond to the best estimates considering the effect of noise. The optimal selection of the value of ρ depends on the acoustic system's noise level (SNR). Therefore, a conservatively large value is often advantageous in enabling convergence in general at the cost of convergence speed. We observed that $\rho = 1$ leads to stable behaviour in the numerical simulations. See also Sec. VI.

Here, I have removed the mention of dynamic systems for now, as the arguments were a little vague.

TABLE II: Table of operations and their respective computational complexity at a node i and frame (τ) as well as the number of bits transmitted. B denotes the number of bits per complex number. table is huge

	Eq.	**Add.**	**Mult.**	**Transm.**
$\mathbf{x}_i^{(\tau)}$	(3)	–	–	$(T_i-1)LB$
FFT	(3)	$L\log_2 L$	$\frac{L}{2}\log_2 L$	–
$\mathbf{X}_i^{(\tau)}$	(10)	–	$2R_iL^3$	–
$\mathbf{C}_i^{(\tau)}$	(19)	–	$R_i^3L^3$	–
$\mathbf{g}_i^{(\tau)}$	(28)	$4R_i^2L^2+3R_iL$	$3R_i^2L^2+2R_iL$	–
$\mathbf{w}_i^{(\tau+1)}$	(27a)	$R_i^2L^2+R_iL$	$R_i^2L^2$	–
$\mathbf{V}^{(\tau+1)}$	(28)	$5R_i^2L^2+2R_iL$	$6R_i^2L^2+2R_iL$	–
$\mathbf{y}_{i\|j}^{(\tau+1)}$	(27b)	R_iL	R_iL	–
$\mathbf{y}_{j\|i}^{(\tau+1)}$	(27c)	–	–	$(R_i-1)LB$
$\mathbf{z}_{i\|i}^{(\tau+1)}$	(27d)	T_iL	T_iL+2L	–
$\mathbf{z}_{j\|i}^{(\tau+1)}$	(27e)	–	–	$(T_i-1)LB$
$\mathbf{u}_{i\|j}^{(\tau+1)}$	(27f)	$2R_iL$	R_iL	–
$\boldsymbol{\eta}_i^{(\tau)}$	(34)	T_iL	–	T_i-1
$\gamma_i^{(\tau)}$	(37)	L	L	–
$\mathbf{B}_i$	(40)	–	L	–

B. Computational complexity

This section will comment on the proposed algorithm's computational complexity. Although our research has yet to focus on directly reducing computational complexity, the distributed problem formulation leads to some benefits. The sub-problems of the distributed algorithm are of lower dimension than the centralized counterpart, all of which can be solved in parallel by the processing units of the sensor nodes. Irrespective of the network's size, the reduction in complexity can be significant, thanks to the fixed size of receive and transmit neighbourhoods, $\mathcal{R}_i$ and $\mathcal{T}_i$, respectively, which determine the size of the sub-problems. This independence underscores the algorithm's scalability. Table II lists the arithmetic operations of the proposed algorithm's significant update steps. Fig. 5 (left) shows the relative reduction of arithmetic operations compared to a centralized M-channel gradient descent algorithm. The maximal reduction in complexity without removing nodes from the network is achieved for a ring topology as shown in Fig. 2 (right). There, the receive-neighbourhood of each node is of size $R_i = 2$, i.e., solving a 2-channel sub-problem.

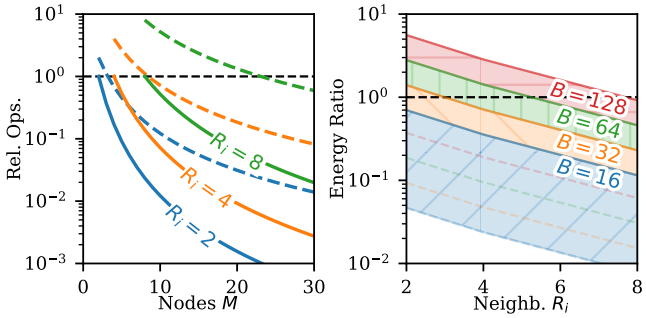


Fig. 5: Comparing the computational complexity at a node of the proposed distributed algorithm to a centralized algorithm with BFGS quasi-newton update. The plot shows the relative complexity of the distributed algorithm with respect to the centralized algorithm. (Left) The complexity is shown for different receive neighbourhood sizes R_i and a fixed length $L = 128$. Full line indicates complexity per node, dashed line of the same colour indicates the total complexity for the network. (Right) The approximate range of energy ratio between transmission and local computation for a selection of bits per complex number $B \in [16, 32, 64, 128]$.

Data transmission between sensor nodes is a significant operation in terms of energy consumption. The energy ratio between transmitting a single bit and computing a single instruction depends on the specific hardware implementation and can range from 200 to 3000 [71, Ch. 2.2.5]; we assume 2000 for the remainder of this section. In general, the argument is to reduce the number of bits that must be transmitted in favour of doing more computations locally, e.g., sophisticated compression and coding algorithms. While the task of bitrate reduction is out of this paper's scope, we would like to give a simple classification of this algorithm's energy ratio. In this application, several transmission steps contribute to the overall transmission cost. (1) The transmission of the microphone signals to the nodes in the transmit-neighbourhood, (2) the transmission of the local estimates to the nodes in the receive-neighbourhood, and (3) the transmission of the consensus estimates to the nodes in the transmit neighbourhood. Additionally, the transmission of the norm estimates for the non-triviality constraint and the delay estimates for time-shift

compensation must be transmitted. We define the number of bits per complex number as B and list the number of transmissions in Table II. Fig. 5 (right) shows the ratio between energy for transmission and local computation for a selection of $B \in [4, 16, 32, 128]$. There are several ways to further reduce bitrate. The transmission rate of the microphone signals can be reduced by employing a state-of-the-art audio codec, such as the Opus codec [72], [73] or recent neural network-based audio codecs, e.g., [74], [75]. The transmission rate of the local and consensus estimates can be approached using an application-specific coding scheme, such as the one proposed in [48] or neural network-based coding schemes.

VI. Numerical Simulations

Little note: The number of Monte Carlo runs is currently not at 50 yet, that is why the curves are still a bit "wiggly". This will improve later when I have more runs. It won't affect the overall effects and trends, however at this point, I don't want to make you wait any longer.

Let us define the error metrics, comparing the estimated channel responses $\mathbf{w}_{i|i}^{(\tau)}$ to the ground-truth channel responses $\mathbf{h}_i^{(\tau)}$, which we consider both time-varying. The normalized projection misalignment [76] for two vectors $\mathbf{a}, \mathbf{b}$ of equal length is defined as

$$\mathrm{npm}(\mathbf{a}, \mathbf{b}) = 20 \log_{10} \frac{\|\epsilon(\mathbf{a}, \mathbf{b})\|}{\|\mathbf{b}\|}, \tag{53}$$

with

$$\epsilon(\mathbf{a}, \mathbf{b}) = \mathbf{b} - \frac{\mathbf{b}^* \mathbf{a}}{\mathbf{a}^* \mathbf{a}} \mathbf{a}. \tag{54}$$

We define the global normalized projection misalignment (NPM) as

$$\mathrm{NPM}^{(\tau)} = \mathrm{npm}(\hat{\mathbf{h}}^{(\tau)}, \mathbf{h}^{(\tau)}), \tag{55}$$

where $\hat{\mathbf{h}}^{(\tau)}$ is the stacked vector of the estimated channel responses $\mathbf{w}_{i|i}^{(\tau)}$ for all $i \in \mathcal{M}$ and $\mathbf{h}^{(\tau)}$ is the stacked vector of the ground-truth channel responses $\mathbf{h}_i^{(\tau)}$. Additionally, we define the NPM for each channel i and frame (τ) as

$$\mathrm{NPM}_i^{(\tau)} = \mathrm{npm}(\mathbf{w}_{i|i}^{(\tau)}, \mathbf{h}_i^{(\tau)}). \tag{56}$$

Further, we also use microphone-averaged NPM, defined as

$$\mathrm{NPM}_{\mathrm{avg}}^{(\tau)} = \frac{1}{M} \sum_{i \in \mathcal{M}} \mathrm{NPM}_i^{(\tau)}, \tag{57}$$

which when compared with (55) is invariant to the relative scaling between the channels.

We define the SNR in the acoustic scenarios as the mean SNR at the microphones, given by

$$\mathrm{SNR}_{\mathrm{dB}} = \frac{1}{M} \sum_{i=1}^{M} 10 \log_{10} \frac{\sigma_{y,i}^2 t_{\mathrm{y,\,i}}}{\sigma_{v,i}^2 t_{\mathrm{x}}}, \tag{58}$$

where $\sigma_{\mathrm{y,\,i}}^2$ is the convolved-source signal variance in signal-plus-noise periods and $\sigma_{\mathrm{v,i}}^2$ is the noise variance, at microphone i respectively. The scalars t_{x} and $t_{\mathrm{y,\,i}}$ are the total and cumulative signal length of signal-plus-noise periods, respectively.

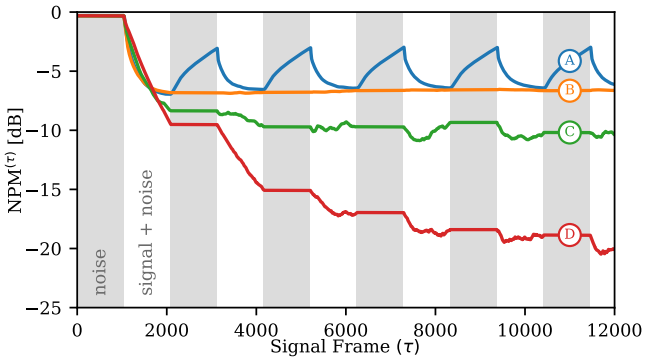


Fig. 6: Mean NPM for the distributed algorithm comparing the naive approach, i.e., no VAD and GEV update (A), using VAD to stop updating the estimates (B), a naive subtraction of the "noise-only" CR matrix from the "signal-plus-noise" CR matrix (C), and the proposed GEV problem update (D). Simulated here are WGN input and IRs and random covariance matrices. Shown is the median of 50 Monte-Carlo runs.

For the following simulation studies, we generate random positive definite noise covariance matrices by multiplying a random matrix where each entry is drawn from a Normal distribution with its transpose and scaling it to fulfil (58) for a given target SNR. To generate the noise signal, we generate random sequences of white Gaussian noise and filter it with the Cholesky decomposition of the noise covariance matrix.

We consider two types of input signals: (a) white Gaussian noise (WGN) signal with speech-like non-stationarity in amplitude and signal pauses (b) speech recordings from the LibriSpeech dataset [77]. Since the recordings in the dataset are of varying lengths, we choose a random set of recordings from the same speaker and concatenate them until reaching the desired length.

A. Evaluation of proposed extensions

The first set of results being presented here is an evaluation of the extensions proposed in this paper. We intend to show the effect each has on the estimation performance.

1) VAD & GEV: We compare the distributed algorithm with and without the GEV problem update as described in Sec. IV-A. We simulate 50 Monte-Carlo runs of random WGN IRs with $M = 3$ as described above, and the input signal is WGN with equally long signal-plus-noise and noise-only segments for visual explainability. The results in Fig. 6 show the clear advantage of including noise covariance information in the estimation process. We compare (A) the naive approach, i.e., no separate estimation of noise and signal-plus-noise CR matrices, (B) using VAD to stop updating the estimates, (C) a naive subtraction of the "noise-only" CR matrix from the "signal-plus-noise" CR matrix, and (D) the proposed GEV problem update. The results show that the proposed GEV problem update significantly improves the estimation performance in terms of NPM that is reached.

2) Regularization: We show the effect of the regularization term, which is described in Sec. IV-B. As a prior, we estimate

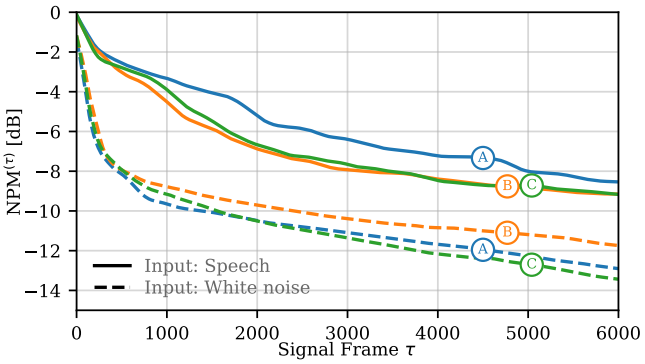


Fig. 7: NPM over times for the proposed algorithm with regularization parameter $\lambda_0 = 0$ (A), $\lambda_0 = 10^{-4}$ (B) and $\lambda_0 = 10^{-5}$ (C). Show is the median all runs for the SNR in the range [10, 40] dB. Regularization improves the estimation performance when the input signal is non-white.

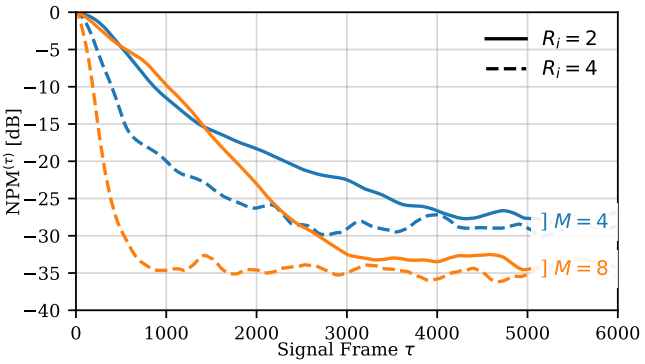


Fig. 8: NPM over time for the proposed algorithm with different WASN topology properties. Convergence speed increases and steady-state error decreases with a larger number of nodes M; convergence speed increases without affecting the steady-state error with a larger number of neighbours R_i. Shown is the median of 50 Monte-Carlo runs with randomly generated neighbourhoods. SNR is 20 dB.

the covariance matrix $\mathbf{\Sigma_h}$ (51) of the room channel responses in the DFT domain from 20000 randomly generated rooms. Then, 50 Monte-Carlo runs of random rooms with a randomly placed network with $M = 4$ are generated with speech-like WGN and speech input. We compare the proposed algorithm with regularization parameter $\lambda_0 = 0$, $\lambda_0 = 10^{-4}$, and $\lambda_0 = 10^{-5}$ and show the median of all runs for the SNR in the range [10, 40] dB. The simulation results plotted in Fig. 7 show that introducing prior information improves the estimation performance when the input signal is non-white by increasing convergence speed and reducing the steady-state error slightly. Unsurprisingly, the regularization term does not affect the estimation performance when the input signal is white.

3) Topology: We compare the estimation performance of the proposed algorithm for WASN topologies with different densities, i.e., more edges between nodes. The number of nodes is chosen as $M \in \{4, 8\}$, and we vary the number of neighbours such that $R_i \in \{2, 4\}$. We simulate 50 Monte-Carlo runs of random IRs with speech-like WGN input. Fig. 8 shows that the convergence speed increases significantly with a larger network size M and neighbourhood $\mathcal{R}_i$. However, the estimation performance in terms of steady-state NPM is only significantly affected by network size M, while only marginally by the neighbourhood size.

B. Evaluation in randomized simulated environment

As our first simulation study, we generate acoustic systems using the *pyroomacoustics* library [78]. For each simulation sun, we generate a random room with each of its dimensions sampled from a uniform distribution on the interval $[2, 6]$ meters. The absorption coefficients of the walls are sampled from a uniform distribution on the interval of $[0.05, 0.5]$. The single source and M sensor nodes are placed at random locations within the room with a minimum distance of 0.1 meters from the walls, again uniformly sampled. We construct the WASN topology as a directed graph by generating edges between each node and its nearest neighbours. We enforce a constraint that ensures the graph is connected, i.e., there is a path from each node i to each node j. In the case where $R_i = 2$, the smallest sub-problem size possible, the resulting topology is a ring topology; see Fig. 2 (right) for an example. We generate room impulse responses using the randomized image source method [79] contained in the "pyroomacoustics" library and shift the impulse responses such that the first non-zero impulse response sample with the earliest arrival time is at 0. As our ground truth and estimates, we use a 128 bin DFT of the impulse responses, appropriately downsampled and truncated. While this significantly simplifies the real-world scenario, it allows us to evaluate the algorithm's performance in a controlled environment while keeping simulations tractable.

We compare the proposed algorithm to various existing BSI algorithms:

- (MCQN) The centralized MCQN algorithm for the M channels problem as described in [27].
- (S-MCQN) The MCQN algorithm applied to the sub-problems independently.
- (DGD) An adaptive implementation of the distributed (stochastic) gradient descent algorithm (DGD) as described in [42].
- (DSAAWET) The distributed stochastic approximation algorithm with expanding truncations (DSAAWET) as described in [44], [45].
- (DABSI) Our proposed algorithm with local BFGS update step.

The MCQN algorithm is chosen as a representative of centralized algorithms, as it has been the best performing in the given scenario. We simulate 50 Monte-Carlo runs of random rooms as described above with both speech-like WGN input and speech input. The scenario is simulated at a set of SNR levels $\{10, 20, 30, 40\}$ dB. Fig. 9 and Fig. 10 show the error over time for the different algorithms and input types as well as the average NPM after convergence. It can be observed that the proposed algorithm outperforms the

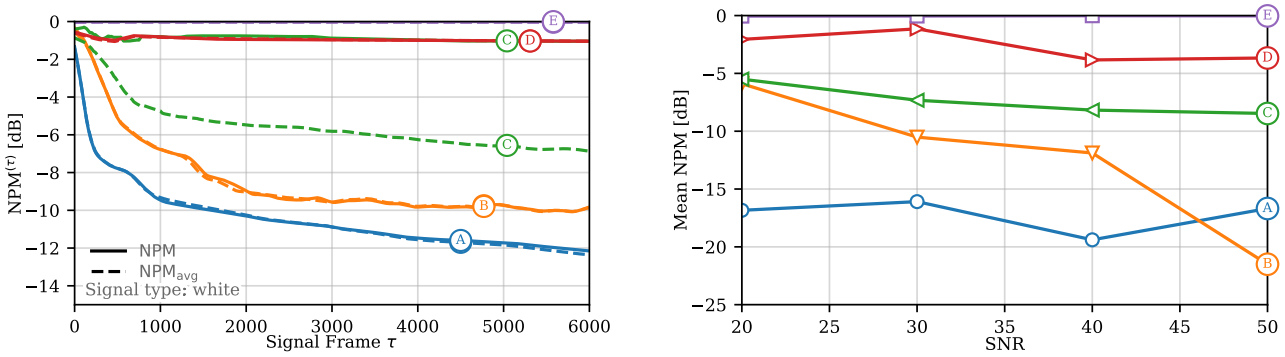


Fig. 9: Proposed DABSI algorithm (A) compared to MCQN (B), S-MCQN (C), DGD (D), and DSAAWET (E). Shown is the median of 50 Monte-Carlo runs with randomly generated rooms and WGN input. **(Left)** Convergence over time for SNR 20 dB. **(Right)** NPM after convergence for different SNR levels.

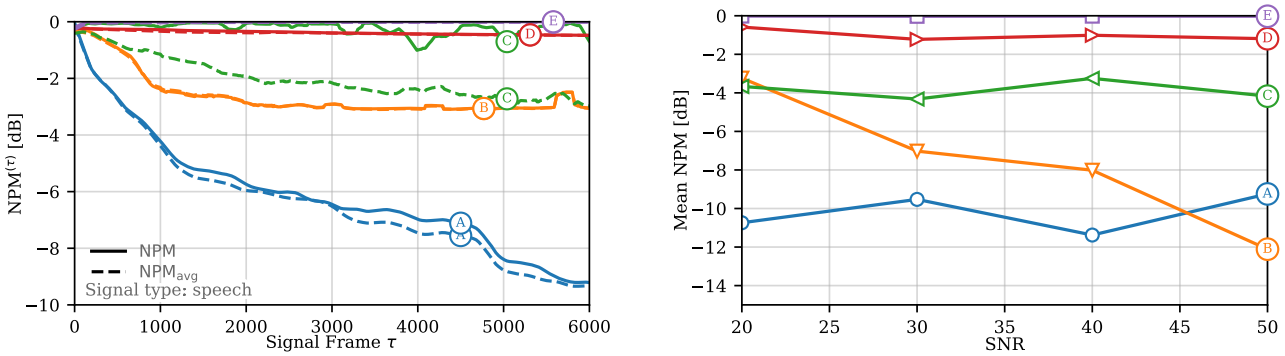


Fig. 10: Proposed DABSI algorithm (A) compared to MCQN (B), S-MCQN (C), DGD (D), and DSAAWET (E). Shown is the median of 50 Monte-Carlo runs with randomly generated rooms and speech input. **(Left)** Convergence over time for SNR 20 dB. **(Right)** NPM after convergence for different SNR levels.

compared algorithms in this scenario significantly, especially in the case of speech input. Improved convergence speed and estimation accuracy can be observed. To note is the difference in estimation performance of the S-MCQN algorithm in $\text{NPM}^{(\tau)}$ and $\text{NPM}^{(\tau)}_{\text{avg}}$ coming from the lost relative scaling between channels. The difference in performance between the MCQN and S-MCQN algorithms in both convergence speed and steady-state error is due to the lost cross-channel information by solving the sub-problems with fewer channels independently. This underlines the importance of the consensus step in the distributed algorithm: Even though the sub-problems are solved independently, it insures that cross-channel information that is lost by sub-problem independence is regained and improved upon. The convergence behaviour of the DGD algorithm is significantly slower than the other algorithms while the DSAAWET algorithm does not converge in the given scenario. Further, it can be observed that the proposed algorithm is less sensitive to the SNR level than the next best performing algorithm, the MCQN algorithm. In turn, however, for high SNR levels, the MCQN algorithm outperforms the proposed.

C. Measured environment

In addition to the simulated environments described in Sec. VI-B, we also consider a set of measured impulse responses from the MYRiAD database [80], [81]. We select random combinations of impulse responses from the dataset to construct the scenario. The source position is randomly selected from the given set of loudspeakers. The impulse responses are resampled to a sampling rate of 8000 kHz and truncated at 128 samples before convolving with the source signal, which is a speech-like WGN on the one hand and on the other, randomly selected speech signal from the librispeech dataset [77].

Again, we simulate 50 Monte Carlo runs of the constructed scenario using the measured impulse responses. The proposed DABSI algorithm converges faster and to a lower NPM than the other algorithms, as shown in Fig. 11 and Fig. 12. While overall the estimation accuracy is worse, the observed behaviour is consistent with the simulated environment. In the case of the speech input, the proposed DABSI algorithm and the centralized MCQN algorithm perform similarly, while the other algorithms perform worse.

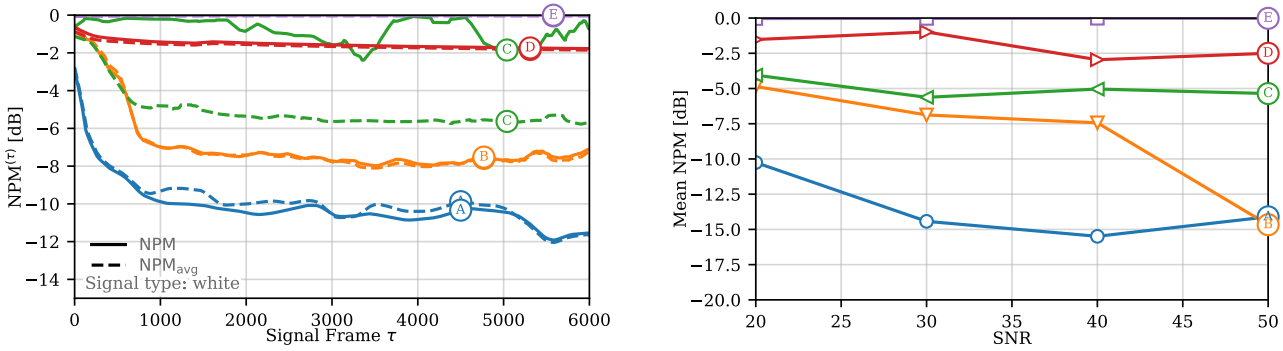


Fig. 11: Proposed DABSI algorithm (A) compared to MCQN (B), S-MCQN (C), DGD (D), and DSAAWET (E). Shown is the median of 50 Monte-Carlo runs with randomly selected impulse responses from the MYRiAD database and WGN input. **(Left)** Convergence over time for SNR 20 dB. **(Right)** NPM after convergence for different SNR levels.

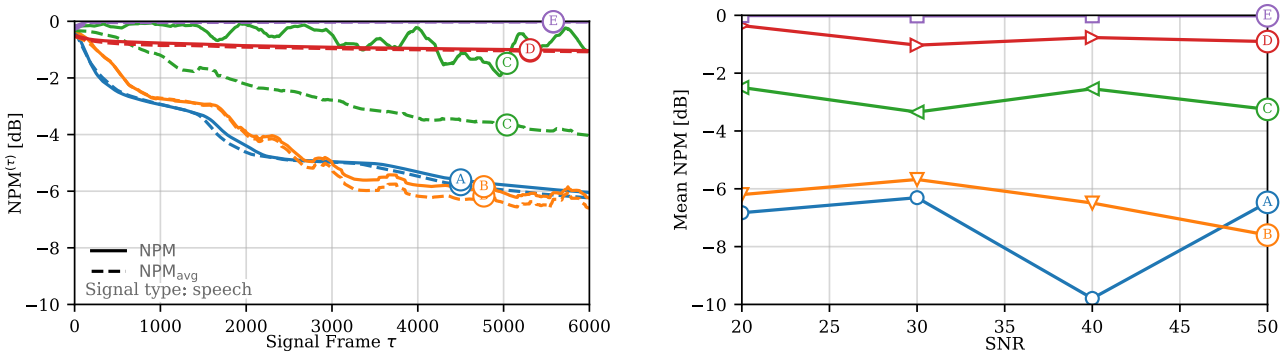


Fig. 12: Proposed DABSI algorithm (A) compared to MCQN (B), S-MCQN (C), DGD (D), and DSAAWET (E). Shown is the median of 50 Monte-Carlo runs with randomly selected impulse responses from the MYRiAD database and speech input. **(Left)** Convergence over time for SNR 20 dB. **(Right)** NPM after convergence for different SNR levels.

TABLE III: Simulation parameters

Parameter	**WGN IRs**	**Simulated Env.**		**Measured Env.**	
	WGN	WGN	Speech	WGN	Speech
Nr. Nodes M	$\{4, 8\}$	4	4	4	4
Step size μ	0.4				
Will fill this in...					

VII. Conclusions

The proposed distributed algorithm for blind acoustic system identification is a step towards a scalable solution for large sensor networks. The algorithm is based on the cross-relation method, which is well-established, and set into a distributed context using the alternating direction method of multipliers. For improved estimation performance, we extend the sub-problems with a VAD-based approach and a regularization term based on prior knowledge of general room channel responses. Moreover, we use a quasi-Newton method to solve the sub-problems which improves the convergence speed and estimation accuracy. The algorithm is evaluated on both synthetic and real-world data which shows good performance when compared to other BSI algorithms. Further, the effectiveness of the consensus approach is verified by comparing the results to a centralized algorithm applied to separate sub-problems. The algorithm is able to effectively estimate the channel responses in an adaptive way while avoiding the need for a centralized processing unit.

There are some limitations to the algorithm, such as the relatively high computational complexity and the assumption of a known length of the impulse responses. In simulated scenarios, the latter is dealt with by truncating impulse responses used to generate the data, however, in real-world applications, the length of the impulse responses is unknown. Moreover, the computational complexity of the algorithm is relatively high, due to matrix-matrix multiplications. For applicability in real-world scenarios, i.e., a real-time system, this need to be further investigated and optimized. Further, we made the assumption that there is a VAD at each node, yielding perfect data. In practice, this is not the case and the algorithm should be extended to be robust to inaccurate VAD decisions.

ACKNOWLEDGMENTS

This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. blabla...

REFERENCES

[1] M. Maroti, G. Simon, A. Ledeczi, and J. Sztipanovits, "Shooter localization in urban terrain," *Computer*, vol. 37, no. 8, pp. 60–61, 2004.

[2] W.-P. Chen, J. Hou, and L. Sha, "Dynamic clustering for acoustic target tracking in wireless sensor networks," *IEEE Trans. Mob. Comput.*, vol. 3, no. 3, pp. 258–271, 2004.

[3] A. Bertrand and M. Moonen, "Distributed Adaptive Node-Specific Signal Estimation in Fully Connected Sensor Networks—Part I: Sequential Node Updating," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5277–5291, Oct. 2010.

[4] ——, "Distributed Adaptive Node-Specific Signal Estimation in Fully Connected Sensor Networks—Part II: Simultaneous and Asynchronous Node Updating," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5292–5306, Oct. 2010.

[5] M. Ferrer, M. de Diego, G. Piñero, and A. Gonzalez, "Affine projection algorithm over acoustic sensor networks for active noise control," *IEEE Trans. Audio Speech Lang. Process.*, vol. 29, pp. 448–461, 2021.

[6] S. Ruiz, T. van Waterschoot, and M. Moonen, "Distributed combined acoustic echo cancellation and noise reduction using GEVD-based distributed adaptive node specific signal estimation with prior knowledge," in *Proc. 28th European Signal Process. Conf. (EUSIPCO '20)*. Amsterdam, Netherlands: IEEE, Jan. 2021, pp. 206–210.

[7] Y. Zeng and R. C. Hendriks, "Distributed Delay and Sum Beamformer for Speech Enhancement via Randomized Gossip," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 1, pp. 260–273, Jan. 2014.

[8] J. S. Bradley, "Predictors of speech intelligibility in rooms," *The Journal of the Acoustical Society of America*, vol. 80, no. 3, pp. 837–845, Sep. 1986.

[9] Y. E. Baba, A. Walther, and E. A. P. Habets, "3D Room Geometry Inference Based on Room Impulse Response Stacks," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 5, pp. 857–872, May 2018.

[10] K. MacWilliam, F. Elvander, and T. van Waterschoot, "Simultaneous Acoustic Echo Sorting and 3-D Room Geometry Inference," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23)*. Rhodes Island, Greece: IEEE, 2023, pp. 1–5.

[11] Y. Sato, "A Method of Self-Recovering Equalization for Multilevel Amplitude-Modulation Systems," *IEEE Trans. Commun.*, vol. 23, no. 6, pp. 679–682, Jun. 1975.

[12] D. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.

[13] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Conf. Rec. 25th Asilomar Conf. Signals, Syst. Computers*. Pacific Grove, CA, USA: IEEE Comput. Soc. Press, 1991, pp. 856–860.

[14] J. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.

[15] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

[16] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[17] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[18] S. Gannot and M. Moonen, "Subspace Methods for Multimicrophone Speech Dereverberation," *EURASIP J. Adv. Signal Process.*, vol. 2003, no. 11, p. 769285, Dec. 2003.

[19] K. I. Diamantaras and T. Papadimitriou, "An Efficient Subspace Method for the Blind Identification of Multichannel FIR Systems," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 5833–5839, Dec. 2008.

[20] Q. Mayyala, K. Abed-Meraim, and A. Zerguine, "Structure-Based Subspace Method for Multichannel Blind System Identification," *IEEE Signal Process. Lett.*, vol. 24, no. 8, pp. 1183–1187, Aug. 2017.

[21] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 661–672, Mar. 1996.

[22] Y. A. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Processing*, p. 12, 2002.

[23] ——, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[24] M. Hu, S. Doclo, D. Sharma, M. Brookes, and P. A. Naylor, "Noise robust blind system identification algorithms based on a Rayleigh quotient cost function," in *Proc. 23th European Signal Process. Conf. (EUSIPCO '15)*, Aug. 2015, pp. 2476–2480.

[25] H. He, J. Chen, J. Benesty, and T. Yang, "Noise robust frequency-domain adaptive blind multichannel identification with ℓ_p-norm constraint," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 9, pp. 1608–1619, Sep. 2018.

[26] B. Jo and P. Calamia, "Robust Blind Multichannel Identification based on a Phase Constraint and Different ℓp-norm Constraints," in *Proc. 28th European Signal Process. Conf. (EUSIPCO '20)*, Jan. 2021, pp. 1966–1970.

[27] E. A. Habets and P. A. Naylor, "An online quasi-Newton algorithm for blind SIMO identification," in *Proc. 2010 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '10)*. Dallas, TX: IEEE, Mar. 2010, pp. 2662–2665.

[28] M. Hu, "Cross-relation based blind identification of acoustic SIMO systems and applications," Ph.D. dissertation, Imperial College London, 2017.

[29] S. Kitić and J. Daniel, "Blind identification of ambisonic reduced room impulse response," *IEEE Trans. Audio Speech Lang. Process.*, vol. 32, pp. 443–458, 2024.

[30] D. Schmid and G. Enzner, "Cross-Relation-Based Blind SIMO Identifiability in the Presence of Near-Common Zeros and Noise," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 60–72, Jan. 2012.

[31] G. Enzner and P. Thüne, "On the statistics and the detection of multichannel common zeros," in *Proc. 2014 Int. Workshop Acoustic Signal Enhancement (IWAENC '14)*, Sep. 2014, pp. 21–25.

[32] G. Enzner and P. Vary, "Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones," *Signal Processing*, vol. 86, no. 6, pp. 1140–1156, Jun. 2006.

[33] S. Malik, D. Schmid, and G. Enzner, "A State-Space Cross-Relation Approach to Adaptive Blind SIMO System Identification," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 511–514, Aug. 2012.

[34] S. Malik and G. Enzner, "Recursive Bayesian Control of Multichannel Acoustic Echo Cancellation," *IEEE Signal Process. Lett.*, vol. 18, no. 11, pp. 619–622, Nov. 2011.

[35] D. Schmid, S. Malik, and G. Enzner, "A Maximum A Posteriori Approach to Multichannel Speech Dereverberation and Denoising," in *Proc. 2012 Int. Workshop Acoustic Signal Enhancement (IWAENC '12)*, Sep. 2012, pp. 1–4.

[36] S. Malik, "Bayesian learning of linear and nonlinear acoustic system models in hands-free communication," Ph.D. dissertation, Ruhr-Universität Bochum, Bochum, 2012.

[37] D. Schmid, S. Malik, and G. Enzner, "An expectation-maximization algorithm for multichannel adaptive speech dereverberation in the frequency-domain," in *Proc. 2012 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '12)*, Mar. 2012, pp. 17–20.

[38] S. Malik and G. Enzner, "A Variational Bayesian Learning Approach for Nonlinear Acoustic Echo Control," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5853–5867, Dec. 2013.

[39] D. Schmid, G. Enzner, S. Malik, D. Kolossa, and R. Martin, "Variational Bayesian Inference for Multichannel Dereverberation and Noise Reduction," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 8, pp. 1320–1335, Aug. 2014.

[40] P. Thuene and G. Enzner, "Maximum-Likelihood and Maximum-A-Posteriori Perspectives for Blind Channel Identification on Acoustic Sensor Network Data," in *Speech Communication; 13th ITG-Symposium*, Oct. 2018, pp. 1–5.

[41] T. van Waterschoot and M. Moonen, "Distributed estimation and equalization of room acoustics in a wireless acoustic sensor network," in *Proc. 20th European Signal Process. Conf. (EUSIPCO '12)*, Aug. 2012, pp. 2709–2713.

[42] C. Yu, L. Xie, and Y. C. Soh, "Distributed blind system identification in sensor networks," in *Proc. 2014 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '14)*, May 2014, pp. 5065–5069.

[43] Y. Liu, H. Liu, and C. Li, "Distributed blind identification of sparse channels in sensor networks," in *Proc. 35th Chinese Control Conf.(CCC '16)*, Jul. 2016, pp. 5122–5127.

[44] R. Liu and H.-F. Chen, "Distributed and recursive blind channel identification to sensor networks," *Control Theory Technol.*, vol. 15, no. 4, pp. 274–287, Nov. 2017.
[45] J. Lei and H.-F. Chen, "Distributed Stochastic Approximation Algorithm With Expanding Truncations," *IEEE Trans. Autom. Control*, vol. 65, no. 2, pp. 664–679, Feb. 2020.
[46] M. Blochberger, F. Elvander, R. Ali, M. Moonen, T. van Waterschoot, J. Østergaard, and J. Jensen, "Distributed Cross-Relation-Based Frequency-Domain Blind System Identification Using Online-Admm," in *Proc. 2022 Int. Workshop Acoustic Signal Enhancement (IWAENC '22)*. Bamberg, Germany: IEEE, Sep. 2022, pp. 1–5.
[47] M. Blochberger, F. Elvander, R. Ali, J. Østergaard, J. Jensen, M. Moonen, and T. van Waterschoot, "Distributed Adaptive Norm Estimation for Blind System Identification in Wireless Sensor Networks," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23)*. Rhodes Island, Greece: IEEE, Jun. 2023, pp. 1–5.
[48] M. Blochberger, J. Østergaard, R. Ali, M. Moonen, F. Elvander, J. Jensen, and T. van Waterschoot, "Adaptive coding in wireless acoustic sensor networks for distributed blind system identification," in *Conf. Rec. 55th Asilomar Conf. Signals, Syst. Computers*, 2023, pp. 1420–1424.
[49] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
[50] H. Wang and A. Banerjee, "Online alternating direction method," in *Proc. 29th Int. Conf. Machine Learning (ICML '12)*, 2012, pp. 1119–1126.
[51] S. Hosseini, A. Chapman, and M. Mesbahi, "Online distributed ADMM via dual averaging," in *Proc. 53rd IEEE Conf. Decision Control (CDC '14)*, Dec. 2014, pp. 904–909.
[52] ——, "Online Distributed Convex Optimization on Dynamic Networks," *IEEE Trans. Autom. Control*, vol. 61, no. 11, pp. 3545–3550, Nov. 2016.
[53] K. W. Brodlie, "An assessment of two approaches to variable metric methods," *Mathematical Programming*, vol. 12, no. 1, pp. 344–355, Dec. 1977.
[54] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research. New York: Springer, 2006.
[55] A. Hassani, A. Bertrand, and M. Moonen, "Distributed GEVD-based signal subspace estimation in a fully-connected wireless sensor network," in *Proc. 22nd European Signal Process. Conf. (EUSIPCO '14)*, Lisbon, Portugal, 2014, pp. 1292–1296.
[56] ——, "GEVD-Based Low-Rank Approximation for Distributed Adaptive Node-Specific Signal Estimation in Wireless Sensor Networks," *IEEE Transactions on Signal Processing*, vol. 64, no. 10, pp. 2557–2572, May 2016.
[57] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Processing*, vol. 88, no. 3, pp. 594–611, Mar. 2008.
[58] Y. Hua, "Blind methods of system identification," *Circuits Systems and Signal Process.*, vol. 21, pp. 91–108, 2002.
[59] Y. Wang, W. Yin, and J. Zeng, "Global Convergence of ADMM in Nonconvex Nonsmooth Optimization," *J Sci Comput*, vol. 78, no. 1, pp. 29–63, Jan. 2019.
[60] S. Hosseini, A. Chapman, and M. Mesbahi, "Online Distributed ADMM on Networks," Oct. 2015.
[61] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.
[62] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Low-rank Approximation Based Multichannel Wiener Filter Algorithms for Noise Reduction with Application in Cochlear Implants," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 4, pp. 785–799, Apr. 2014.
[63] R. Varzandeh, M. Taseska, and E. A. P. Habets, "An iterative multichannel subspace-based covariance subtraction method for relative transfer function estimation," in *2017 Hands-free Speech Communications and Microphone Arrays (HSCMA)*. San Francisco, CA, USA: IEEE, 2017, pp. 11–15.
[64] X.-B. Gao, G. H. Golub, and L.-Z. Liao, "Continuous methods for symmetric generalized eigenvalue problems," *Linear Algebra Its Appl.*, vol. 428, no. 2-3, pp. 676–696, Jan. 2008.
[65] G. H. Golub, "Eigenvalue computation in the 20th century," *J. Comput. Appl. Math.*, vol. 123, pp. 35–65, Mar. 2000.
[66] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *J. Res. Natl. Bur. Stand. (U. S.)*, vol. 49, pp. 409–435, 1952.
[67] DK. Faddeev and VN. Faddeeva, "Computational methods of linear algebra," *Journal of Soviet Mathematics*, vol. 15, pp. 531–650, 1981.
[68] Y. T. Feng and D. R. J. Owen, "Conjugate gradient methods for solving the smallest eigenpair of large symmetric eigenvalue problems," *Int. J. Numer. Meth. Engng.*, vol. 39, no. 13, pp. 2209–2229, Jul. 1996.
[69] P. Arbenz, U. L. Hetmaniuk, R. B. Lehoucq, and R. S. Tuminaro, "A comparison of eigensolvers for large-scale 3D modal analysis using AMG-preconditioned iterative methods," *Int. J. Numer. Meth. Engng*, vol. 64, no. 2, pp. 204–236, Sep. 2005.
[70] L.-H. Zhang, "On optimizing the sum of the Rayleigh quotient and the generalized Rayleigh quotient on the unit sphere," *Comput Optim Appl*, vol. 54, no. 1, pp. 111–139, Jan. 2013.
[71] H. Karl and A. Willig, *Protocols and Architectures for Wireless Sensor Networks*, 1st ed. Wiley, Apr. 2005.
[72] J.-M. Valin, K. Vos, and T. Terriberry, "Definition of the opus audio codec," Tech. Rep., 2012.
[73] J.-M. Valin, G. Maxwell, T. B. Terriberry, and K. Vos, "High-Quality, Low-Delay Music Coding in the Opus Codec," *J. Audio Eng. Soc.*, no. 8942, Oct. 2013.
[74] N. Zeghidour, A. Luebs, A. Omran, J. Skoglund, and M. Tagliasacchi, "SoundStream: An End-to-End Neural Audio Codec," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 30, pp. 495–507, 2022.
[75] R. Kumar, P. Seetharaman, A. Luebs, I. Kumar, and K. Kumar, "High-fidelity audio compression with improved RVQGAN," in *Proc. 37th Int. Conf. Neural Inf. Process. Syst.*, ser. Nips '23. Red Hook, NY, USA: Curran Associates Inc., 2024.
[76] D. Morgan, J. Benesty, and M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 174–176, Jul. 1998.
[77] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *Proc. 2015 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '15)*, South Brisbane, Australia, Apr. 2015, pp. 5206–5210.
[78] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *Proc. 2018 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '18)*, 2018, pp. 351–355.
[79] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE Trans. Audio Speech Lang. Process.*, vol. 23, no. 4, pp. 774–786, 2015.
[80] T. Dietzen, R. Ali, M. Taseska, and T. van Waterschoot, "MYRiAD: A multi-array room acoustic database," *EURASIP J. Audio, Speech, Music Process.*, vol. 2023, no. 1, p. 17, Apr. 2023.
[81] ——, "Data repository for MYRiAD: A multi-array room acoustic database," Dec. 2022.

Matthias Blochberger Use `\begin{IEEEbiographynophoto}` and the author name as the argument followed by the biography text.