IEEE Open Journal of Signal Processing

Received 13 March 2026; revised 1 July 2026; accepted 31 July 2026. Date of publication 6 August 2026;
date of current version 24 August 2026. The review of this article was arranged by Associate Editor Michele Geronazzo.

Digital Object Identifier 10.1109/OJSP.2026.3721064

Short Papers

Distributed Multichannel Wiener Filtering for Topology-Unconstrained Wireless Acoustic Sensor Networks

PAUL DIDIER [1], POURYA BEHMANDPOOR [2], HENRI GODE [3] (Graduate Student Member, IEEE), TOON VAN WATERSCHOOT [1], SIMON DOCLO [3,4] (Senior Member, IEEE), JÖRG BITZER [4], AND MARC MOONEN [1] (Fellow, IEEE)

[1]STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics, Electrical Engineering Department (ESAT), KU Leuven, 3001 Leuven, Belgium
[2]Department of Electronics and Informatics, Vrije Universiteit Brussel, B-1050 Brussels, Belgium
[3]Signal Processing Group, Department of Medical Physics, Acoustics and Cluster of Excellence Hearing4all Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany
[4]Fraunhofer IDMT, Project Group Hearing, Speech and Audio Technology, 26129 Oldenburg, Germany

CORRESPONDING AUTHOR: PAUL DIDIER (e-mail: phmdidier@proton.me).

This work was supported in part by A Holistic Approach to the Design of Integrated and Distributed Digital Signal Processing Algorithms for Audio and Speech Communication Devices through Research Council KU Leuven Project under Grant C14-21-0075 and in part by Service-Oriented Ubiquitous Network-Driven Sound—SOUNDS through European Union's Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie Grant 956369.

ABSTRACT This paper introduces the topology-independent distributed multichannel Wiener filter (TI-dMWF), a novel algorithm for distributed node-specific signal estimation in wireless acoustic sensor networks (WASNs) with unconstrained topologies. The TI-dMWF enables each node in the network to compute its centralized multichannel Wiener filter solution by exchanging only low-dimensional fused signals, without requiring iterative estimation, unlike state-of-the-art approaches such as the topology-independent distributed adaptive node-specific signal estimation (TI-DANSE) algorithm. The TI-dMWF is proven optimal when each source is observed by either all nodes or only one node. Theoretical analysis and numerical simulations confirm that it achieves centralized estimation performance in a single run. Its latency as a function of the pruned-tree depth and its computational complexity are also analyzed. Its robustness is assessed in reverberant-room simulations under estimated second-order statistics, various network topologies, and deviations from the assumed observability model.

INDEX TERMS Distributed signal processing, ad-hoc wireless acoustic sensor networks, multichannel Wiener filter, dimensionality reduction, distributed noise reduction.

I. INTRODUCTION

In recent years, devices capable of recording, processing, and exchanging audio data wirelessly have become ubiquitous, enabling the creation of wireless acoustic sensor networks (WASNs), which have gained attention for applications such as environmental monitoring and smart homes [1], [2]. Within a WASN, devices (nodes) equipped with one or more microphones (sensors), such as smartphones, laptops, or hearing aids, collaborate to perform audio signal processing tasks, e.g., noise reduction [3], [4], [5], dereverberation [6], [7], or sound source localization [8], [9]. Compared to traditional fixed microphone arrays, WASNs offer greater spatial coverage, deployment flexibility, and robustness to node failures [1], but operate under constraints such as communication cost and potentially dynamic topologies, making the design of distributed algorithms that respect these constraints a key challenge.

This paper addresses node-specific signal estimation in a WASN, where each node estimates a signal of interest (e.g., speech) from its own sensor signals and signals sent by other nodes. The optimal centralized linear minimum mean square error (LMMSE) solution to this problem is provided as the multichannel Wiener filter (MWF) [10, Ch.2], but computing the MWF directly requires access to all sensor signals in the network, which is often infeasible due to communication cost constraints and in WASNs that are not fully connected (FC). The distributed multichannel Wiener filter (dMWF) [11] lets each node of a FC WASN solve its node-specific LMMSE problem distributively by exchanging only dimensionally reduced (fused) signals, addressing the communication cost constraint. However, topologies may not be FC and may vary over time. We here propose the topology-independent distributed multichannel Wiener filter (TI-dMWF), which lets nodes

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

of a topology-unconstrained WASN compute the centralized MWF as if they had access to all network signals, while exchanging only fused signals. Here, *topology-unconstrained* is a property of the WASN and *topology-independent (TI)* is a property of an algorithm: the TI-dMWF is a topology-independent algorithm over topology-unconstrained WASNs. As the dMWF, it is iterationless, unlike the topology-independent distributed adaptive node-specific signal estimation (TI-DANSE) [12] and TI-DANSE$^+$ [13] algorithms.

The TI-DANSE algorithm belongs to the distributed adaptive signal fusion (DASF) framework [14], which the distributed adaptive node-specific signal fusion (DANSF) framework [15] extends to node-specific problems. Note that TI-DANSE$^+$ is closely related to DANSF [13]. Unlike all these iterative algorithms, the TI-dMWF reaches the centralized MWF in a single discovery-estimation pass in global-local subspaces (GLS) scenarios (defined hereafter).

Unlike the dMWF, which handles any partially overlapping desired subspaces (PODS) scenario [11], the TI-dMWF targets GLS scenarios, where each (speech or noise) source is observed by either all nodes (a *global* source) or only one node (a *local* source). Such scenarios arise, e.g., in airports or train stations: a global source is a signal propagating throughout the environment and picked up by all nodes (e.g., a public-address or alarm announcement), while a local source is observed by a single node (e.g., a nearby talker or localized noise). The more general case where a source is observed by an arbitrary subset of nodes is described by the PODS model; the GLS model considered here remains a relevant special case whenever network-wide and strictly local sources are clearly separated. The GLS assumption differs from the fully overlapping desired subspaces (FODS) assumption [11] (all speech sources observed by all nodes) needed for TI-DANSE and TI-DANSE$^+$ convergence and optimality [12], [13].

II. PROBLEM STATEMENT

We consider a WASN consisting of K nodes, where node k has M_k sensors ($k \in \mathcal{K} := \{1, \ldots, K\}$) such that the total number of sensors is $M = \sum_{k\in\mathcal{K}} M_k$. Signals are complex-valued, e.g., representing processing in one filter-bank bin. The acoustic scenario includes $Q^{(\mathrm{s})}$ desired sources whose latent signals at time t are grouped in $\mathbf{s}^{\mathrm{lat}}[t] \in \mathbb{C}^{Q^{(\mathrm{s})}\times 1}$, and $Q^{(\mathrm{n})}$ noise sources with latent signals $\mathbf{n}^{\mathrm{lat}}[t] \in \mathbb{C}^{Q^{(\mathrm{n})}\times 1}$, for a total of $Q = Q^{(\mathrm{s})} + Q^{(\mathrm{n})}$ sources. The sensor signals at all nodes are stacked in $\mathbf{y}[t] \in \mathbb{C}^{M\times 1}$. Omitting the time-index $[t]$ for conciseness, $\mathbf{y}$ can be written as $\mathbf{y} = \left[\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_K^{\mathrm{T}}\right]^{\mathrm{T}} = \mathbf{s} + \mathbf{n} + \mathbf{v}$, where $\mathbf{s} = \mathbf{A}\mathbf{s}^{\mathrm{lat}}$ and $\mathbf{n} = \mathbf{B}\mathbf{n}^{\mathrm{lat}}$ are the speech and noise components, with $\mathbf{A} \in \mathbb{C}^{M\times Q^{(\mathrm{s})}}$ and $\mathbf{B} \in \mathbb{C}^{M\times Q^{(\mathrm{n})}}$ the corresponding steering matrices, and $\mathbf{v} \in \mathbb{C}^{M\times 1}$ the self-noise. In GLS scenarios, this can be further expanded as:

$$\mathbf{y} = \overset{\circ}{\mathbf{s}} + \dot{\mathbf{s}} + \overset{\circ}{\mathbf{n}} + \dot{\mathbf{n}} + \mathbf{v}, \tag{1}$$

where $\mathbf{s} = \mathbf{s}^{\circ} + \dot{\mathbf{s}}$ and $\mathbf{n} = \mathbf{n}^{\circ} + \dot{\mathbf{n}}$ decompose into global-source components $\mathbf{s}^{\circ} = \mathbf{A}^{\circ}\mathbf{s}^{\mathrm{lat}}$, $\mathbf{n}^{\circ} = \mathbf{B}^{\circ}\mathbf{n}^{\mathrm{lat}}$ and local-source components $\dot{\mathbf{s}} = \dot{\mathbf{A}}\mathbf{s}^{\mathrm{lat}}$, $\dot{\mathbf{n}} = \dot{\mathbf{B}}\mathbf{n}^{\mathrm{lat}}$. The matrices $\mathbf{A}^{\circ}$, $\mathbf{B}^{\circ}$ (resp. $\dot{\mathbf{A}}$, $\dot{\mathbf{B}}$) equal $\mathbf{A}$, $\mathbf{B}$ with the local-source (resp. global-source) columns set to zero. As in the dMWF [11], the matrix obtained by removing the all-zeros columns from $[\mathbf{A}^{\circ} \mid \mathbf{B}^{\circ}]$ is assumed full column rank; its number of columns $\overset{\circ}{Q}$ is the *total number of global sources*, and $\dot{Q} = Q - \overset{\circ}{Q}$ the number of local sources. A root node k is assumed to satisfy $M_k \geq \overset{\circ}{Q}$, so that its upstream signal in Section III (cf. (7)) is well-defined; a node with $M_k < \overset{\circ}{Q}$ may still participate as a non-root node (cf. Section III-E). Relaxing this for root nodes is left for future work (cf. Section VII).

Without loss of generality (w.l.o.g.), the desired signal at any node k contains the contribution to the first D sensors of node k of all speech sources observed by node k (global and local speech sources), i.e., $\mathbf{d}_k = \mathbf{E}_k^{\mathrm{T}}\mathbf{s} \in \mathbb{C}^{D\times 1}, \forall\, k \in \mathcal{K}$, where $\mathbf{E}_k \in \{0, 1\}^{M\times D}$ extracts the first D channels of $\mathbf{s}_k$ from $\mathbf{s}$. Also w.l.o.g., D is assumed to be the same for all nodes. To optimally estimate $\mathbf{d}_k$ based on $\mathbf{y}$ in the LMMSE sense in a centralized setting, the centralized MWF can be computed at the fusion center as:

$$\widehat{\mathbf{W}}_k = \underset{\mathbf{W}\in\mathbb{C}^{M\times D}}{\arg\min}\, \mathbb{E}\left\{\left\|\mathbf{d}_k - \mathbf{W}^{\mathrm{H}}\mathbf{y}\right\|_2^2\right\} = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{R}_{\mathbf{yd}_k}, \tag{2}$$

where $\mathbf{R}_{\mathbf{yy}} = \mathbb{E}\{\mathbf{y}\mathbf{y}^{\mathrm{H}}\}$ and $\mathbf{R}_{\mathbf{yd}_k} = \mathbb{E}\{\mathbf{y}\mathbf{d}_k^{\mathrm{H}}\}$ are spatial covariance matrices (SCMs). To estimate these SCMs in practical speech enhancement scenarios time-averaging strategies can be applied, assuming ergodic signals. The first SCM $\mathbf{R}_{\mathbf{yy}}$ can be estimated during speech-active periods using, e.g., exponential averaging over incoming time-frames of $\mathbf{y}$ [4]. The second SCM $\mathbf{R}_{\mathbf{yd}_k}$ cannot be estimated directly as it involves the unavailable desired signal $\mathbf{d}_k$. Instead, since $\mathbf{R}_{\mathbf{yd}_k} = \mathbf{R}_{\mathbf{ss}}\mathbf{E}_k$ with $\mathbf{R}_{\mathbf{ss}} = \mathbb{E}\{\mathbf{s}\mathbf{s}^{\mathrm{H}}\}$, and assuming wide-sense stationary noise, the on-off nature of speech is leveraged through a voice activity detector (VAD) or speech presence probability (SPP) [16], [17], [18], [19] to estimate the noise-only SCM $\mathbf{R}_{\mathbf{nn}}$ in speech-absent frames, giving $\mathbf{R}_{\mathbf{yd}_k} = (\mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{nn}})\mathbf{E}_k$. A generalized eigenvalue decomposition (GEVD)-based approach can alternatively be used [5]. In practice, we may use a forgetting factor $0 \ll \beta < 1$ to recursively estimate $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$ as:

$$\begin{aligned} &\text{if VAD=1: } \mathbf{R}_{\mathbf{yy}}[t] = \beta\mathbf{R}_{\mathbf{yy}}[t-1] + (1-\beta)\mathbf{y}_k[t]\mathbf{y}_k[t]^{\mathrm{H}}, \\ &\text{if VAD=0: } \mathbf{R}_{\mathbf{nn}}[t] = \beta\mathbf{R}_{\mathbf{nn}}[t-1] + (1-\beta)\mathbf{y}_k[t]\mathbf{y}_k[t]^{\mathrm{H}}, \end{aligned} \tag{3}$$

where $\mathbf{R}_{\mathbf{yy}}[t]$ and $\mathbf{R}_{\mathbf{nn}}[t]$ denote SCM at time-frame t.

III. TOPOLOGY-INDEPENDENT DISTRIBUTED MWF (TI-DMWF)

Without a fusion center, communication cost constraints and a possibly non-FC, time-varying topology inhibit the direct computation of the centralized MWF (2) at each node k. The TI-dMWF described in this section lets each node of any topology-unconstrained WASN compute the centralized MWF (2) distributively by exchanging only dimensionally reduced (fused) sensor signals. The TI-dMWF is defined with respect to one specific node k and only allows the estimation of the desired signal $\mathbf{d}_k$ at this node. The TI-dMWF steps must thus be repeated for each node in the WASN.

The desired signal estimate at a particular node k is obtained through cascaded estimation-fusion processes. To enable this cascade of operations, the WASN is first pruned to a tree. For instance, a minimum spanning tree (MST) can be computed from any WASN topology using Prim's algorithm [20], then oriented by choosing a root node. In the resulting rooted tree, a non-root node with a single neighbor is a leaf node and any other non-root node is an internal node; each non-root node has one downstream neighbor (towards the root) and each internal node at least one upstream neighbor (towards the leaves). We let $\mathcal{U}_q$ denote the upstream neighbors of node q and $\overline{\mathcal{U}}_q$ all nodes upstream of q, with $\mathcal{U}_q \subseteq \overline{\mathcal{U}}_q$.

Consider a node k and let it be the root of the tree after pruning. As in the dMWF [11], we seek a set of signals, stacked into an observation vector $\widetilde{\mathbf{y}}_k$ at node k, from which $\mathbf{d}_k$ can be estimated. To this end, a sum-and-send data flow from the leaf nodes to the

root node k is defined. To outline it, we focus on one non-root node $q \in \mathcal{K}_k = \mathcal{K}\backslash\{k\}$. In the considered GLS scenario, the TI-dMWF fused signal dimension is $\mathring{Q}$, the total number of *global* sources, so that the fused signal encompasses all accessible information about the global-sources subspace (recalling that all sources are mutually uncorrelated). If the total dimension $\widehat{M}_q$ of the signals available at node q (defined in (5)) does not exceed $\mathring{Q}$, i.e., $\widehat{M}_q \leq \mathring{Q}$, fusing them into a $\mathring{Q}$-dimensional signal would not reduce their dimensionality. Node q then directly forwards its available signals to its downstream neighbor, bypassing the fusion operation. This *mixed-network* case is discussed in Section III-E. For exposition, we assume $\widehat{M}_q > \mathring{Q}, \forall\, q \in \mathcal{K}_k$ in the remainder of this section.

Aiming at estimation of $\mathbf{d}_k$ at root node k, node q will send a fused signal $\widehat{\mathbf{z}}_q$ to its (only) downstream neighbor in the tree topology. This signal is recursively defined as a linear combination of the local sensor signals $\mathbf{y}_q$ with the signals received by node q from its own upstream neighbors (if q is a leaf node, then $\widehat{\mathbf{y}}_q$ below is simply $\mathbf{y}_q$), i.e.:

$$\forall\, q \in \mathcal{K}_k : \widehat{\mathbf{z}}_q = \widehat{\mathbf{P}}_q^{\mathrm{H}} \widehat{\mathbf{y}}_q \in \mathbb{C}^{\mathring{Q}\times 1}, \tag{4}$$

$$\text{with } \widehat{\mathbf{y}}_q = \left[\mathbf{y}_q^{\mathrm{T}}, \widehat{\mathbf{z}}_{u_1}^{\mathrm{T}}, \ldots, \widehat{\mathbf{z}}_{u_B}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{\widehat{M}_q \times 1}, \tag{5}$$

$$\text{where } \widehat{M}_q = M_q + |\mathcal{U}_q|\mathring{Q}, \tag{6}$$

where $\{u_1, \ldots, u_B\}$ are the indices of the upstream neighbors of node q (with $B = |\mathcal{U}_q|$), and $\widehat{\mathbf{P}}_q \in \mathbb{C}^{\widehat{M}_q \times \mathring{Q}}$ is the fusion matrix at node q directed towards root node k. For notation simplicity, we omit to refer to the root node k in $\widehat{\mathbf{z}}_q$, $\widehat{\mathbf{P}}_q$, and $\widehat{\mathbf{y}}_q$ but it is understood that these quantities are defined with respect to a particular root node k. When a tree with another root is pruned, $B = |\mathcal{U}_q|$ and $\widehat{\mathbf{y}}_q$ change (and so do $\widehat{\mathbf{P}}_q$ and $\widehat{\mathbf{z}}_q$) depending on the new set $\{\mathcal{U}_q\}_{q\in\mathcal{K}}$. It is noted that, for any leaf node l (i.e., any non-root node with a single neighbor), $\widehat{\mathbf{y}}_l = \mathbf{y}_l$ and thus $\widehat{M}_l = M_l$.

The TI-dMWF is composed of two distinct steps. When aiming at a particular root node k, the first step is the *discovery step*, where the fusion matrix $\widehat{\mathbf{P}}_q$ and fused signal $\widehat{\mathbf{z}}_q$ are computed at each node $q \in \mathcal{K}_k$. This discovery step can itself be subdivided into two parts: (i) an *upstream data flow*, where fused signals are transmitted from the root node k to the leaf nodes, and (ii) a *downstream data flow*, where fused signals are transmitted from the leaf nodes to the root node k. The second step is an *estimation step* where the root node estimates $\mathbf{d}_k$ using the observation vector $\widehat{\mathbf{y}}_k$.

A. DISCOVERY STEP: UPSTREAM DATA FLOW

In the upstream data flow, the root node k floods a signal $\mathbf{y}_{q\uparrow}$ through the tree, such that every other node in $\mathcal{K}_k$ eventually has access to $\mathbf{y}_{q\uparrow}$. This signal is defined as a subset of $\mathbf{y}_k$:

$$\mathbf{y}_{q\uparrow} = \mathbf{E}_{\mathring{Q}}^{\mathrm{T}} \mathbf{y}_k \in \mathbb{C}^{\mathring{Q}\times 1}, \tag{7}$$

where $\mathbf{E}_{\mathring{Q}} = [\mathbf{I}_{\mathring{Q}} \mid \mathbf{0}]^{\mathrm{T}} \in \{0,1\}^{M_k \times \mathring{Q}}$ is a selection matrix. For simplicity, (7) assumes all $\mathring{Q}$ global sources are active and that $\mathring{Q}$ is known a priori or estimated by node k via source enumeration [21], [22], [23], [24].

B. DISCOVERY STEP: DOWNSTREAM DATA FLOW

The downstream data flow runs from the leaf nodes towards the root node k. Each node $q \in \mathcal{K}_k$ receives the fused signals $\{\widehat{\mathbf{z}}_u\}_{u\in\mathcal{U}_q}$ from its upstream neighbors, builds $\widehat{\mathbf{y}}_q$ via (5), and uses it with $\mathbf{y}_{q\uparrow}$ to compute its fusion matrix (towards root k) by solving the LMMSE problem:

$$\widehat{\mathbf{P}}_q = \underset{\mathbf{P}_q \in \mathbb{C}^{\widehat{M}_q \times \mathring{Q}}}{\arg\min}\ \mathbb{E}\left\{ \left\| \mathbf{y}_{q\uparrow} - \mathbf{P}_q^{\mathrm{H}} \widehat{\mathbf{y}}_q \right\|_2^2 \right\}, \tag{8}$$

$$\Leftrightarrow \widehat{\mathbf{P}}_q = \mathbf{R}_{\widehat{\mathbf{y}}_q\widehat{\mathbf{y}}_q}^{-1} \mathbf{R}_{\widehat{\mathbf{y}}_q\mathbf{y}_{q\uparrow}} \in \mathbb{C}^{\widehat{M}_q \times \mathring{Q}}, \forall\, q \in \mathcal{K}_k, \tag{9}$$

where $\mathbf{R}_{\widehat{\mathbf{y}}_q\widehat{\mathbf{y}}_q} = \mathbb{E}\{\widehat{\mathbf{y}}_q \widehat{\mathbf{y}}_q^{\mathrm{H}}\} \in \mathbb{C}^{\widehat{M}_q \times \widehat{M}_q}$ and $\mathbf{R}_{\widehat{\mathbf{y}}_q\mathbf{y}_{q\uparrow}} = \mathbb{E}\{\widehat{\mathbf{y}}_q \mathbf{y}_{q\uparrow}^{\mathrm{H}}\} \in \mathbb{C}^{\widehat{M}_q \times \mathring{Q}}$. Node q can then compute $\widehat{\mathbf{z}}_q$ via (4) and send it to its own downstream neighbor. For a leaf node l, (8) and (9) also hold, with $\widehat{\mathbf{y}}_l = \mathbf{y}_l$ and $\widehat{M}_l = M_l$.

When the tree-pruned WASN is a star (every node $q \in \mathcal{K}_k$ directly connected to the root k), the TI-dMWF reduces to the dMWF [11]: the upstream flow sends $\mathbf{y}_{q\uparrow}$ (equating $\mathbf{y}_{k\to q}$ in [11]) to every q, and the downstream flow returns $\widehat{\mathbf{z}}_q$ (equating $\mathbf{z}_q$) to k.

C. ESTIMATION STEP

At the end of the downstream data flow, root node k has access to $\{\widehat{\mathbf{z}}_q\}_{q\in\mathcal{U}_k}$ and can build its own observation vector $\widehat{\mathbf{y}}_k$ as in (5). Node k can then estimate its desired signal based on $\widehat{\mathbf{y}}_k$ via:

$$\widetilde{\mathbf{W}}_k = \underset{\mathbf{W}_k \in \mathbb{C}^{\widehat{M}_k \times D}}{\arg\min}\ \mathbb{E}\left\{ \left\| \mathbf{d}_k - \mathbf{W}_k^{\mathrm{H}} \widehat{\mathbf{y}}_k \right\|_2^2 \right\}, \tag{10}$$

$$\Leftrightarrow \widetilde{\mathbf{W}}_k = \mathbf{R}_{\widehat{\mathbf{y}}_k\widehat{\mathbf{y}}_k}^{-1} \mathbf{R}_{\widehat{\mathbf{y}}_k\mathbf{d}_k}, \tag{11}$$

where $\mathbf{R}_{\widehat{\mathbf{y}}_k\mathbf{d}_k} = \mathbb{E}\{\widehat{\mathbf{y}}_k \mathbf{d}_k^{\mathrm{H}}\}$. In a similar way as for the centralized MWF from (2), the two SCMs necessary to compute (11) may be estimated based on a VAD (or an SPP) and time-averaging as in (3). The TI-dMWF desired signal estimate is then obtained as $\widehat{\mathbf{d}}_{k,\text{TI-dMWF}} = (\widetilde{\mathbf{W}}_k)^{\mathrm{H}} \widehat{\mathbf{y}}_k$. All three steps are repeated for each node $k \in \mathcal{K}$ to estimate $\{\mathbf{d}_k\}_{k\in\mathcal{K}}$. When the root index k changes, so do $\widehat{\mathbf{y}}_q$ (and hence $\widehat{\mathbf{P}}_q$ and $\widehat{\mathbf{z}}_q$), through the new sets $\{\mathcal{U}_q\}_{q\in\mathcal{K}_k}$.

D. TI-DMWF OVERVIEW AND OPTIMALITY

Each node k performs the estimation step at every time-frame, which requires the fused signals $\{\widehat{\mathbf{z}}_q\}_{q\in\mathcal{U}_k}$ from the downstream data flow towards k. These steps are thus performed every frame. The fusion-matrix computation and the upstream data flow, however, may run less frequently (e.g., every $N_{\text{us}} > 1$ frames). The TI-dMWF is summarized in Algorithm 1.

Theorem 1. (TI-dMWF optimality): Under the signal model (1) (GLS scenario), the TI-dMWF desired signal estimate is equal to the centralized desired signal estimate at node $k, \forall\, k \in \mathcal{K}$, when $\widehat{\mathbf{P}}_q$ is defined as in (9), $\forall\, q \in \mathcal{K}_k$, i.e., $(\widetilde{\mathbf{W}}_k)^{\mathrm{H}} \widehat{\mathbf{y}}_k = \widehat{\mathbf{W}}_k^{\mathrm{H}} \mathbf{y}, \forall\, k \in \mathcal{K}$ where $\widehat{\mathbf{W}}_k$ is defined in (2) and $\widetilde{\mathbf{W}}_k$ is defined in (11). *Proof:* see Appendix A.

Remark 1 (Failure under PODS scenarios): The proof of Theorem 1 relies on the inter-node uncorrelatedness of the non-global components $\{\dot{\mathbf{y}}_q\}_{q\in\mathcal{K}}$, which holds under the GLS model (every source is global or observed by a single node). It is used twice: the block-diagonality of $\dot{\mathbf{\Gamma}}$ in (18) admits a factorization of the joint LMMSE filter $\widetilde{\mathbf{W}}_{l'l}$ into locally computable $\mathring{Q}$-dimensional fusion matrices, and it justifies replacing the unobservable $\mathbf{E}_{\mathring{Q}}^{\mathrm{T}} \mathring{\mathbf{y}}_k$ by the measurable $\mathbf{E}_{\mathring{Q}}^{\mathrm{T}} \mathbf{y}_k$ at every non-root node. Under a PODS model, a source observed by a strict subset of two or more nodes makes $\mathbb{E}\{\dot{\mathbf{y}}_q \dot{\mathbf{y}}_{q'}^{\mathrm{H}}\} \neq \mathbf{0}$ for some $q \neq q'$, so both steps fail: the fused signals are

Algorithm 1: The TI-dMWF.

for time-frame index $t = 0, 1, \ldots$ **do**
  for $k \in \mathcal{K}$ **do**
    The WASN is pruned to a tree with root k.
    if mod $(t, N_{\text{us}}) = 0$ **then**
      $\mathbf{y}_{k\uparrow}[t]$ is flooded from k through the tree.
    end if
    for q from leaves to root k (excluded) **do**
      Node q gathers $\{\widehat{\mathbf{z}}_{q'}[t]\}_{q'\in\mathcal{U}_q}$ and builds $\widehat{\mathbf{y}}_q[t]$ as in (5).
      Node q estimates $\mathbf{R}_{\widehat{\mathbf{y}}_q\widehat{\mathbf{y}}_q}$ using $\widehat{\mathbf{y}}_q[t]$ and $\mathbf{R}_{\widehat{\mathbf{y}}_q\mathbf{y}_{q\uparrow}}$ using $\mathbf{y}_{q\uparrow}[t]$ (or $\mathbf{y}_{q\uparrow}[t']$ where $t'<t$ if mod $(t, N_{\text{us}})\neq 0$).
      Node q computes $\widehat{\mathbf{P}}_q$ as in (9).
      Node q computes $\widehat{\mathbf{z}}_q[t]$ as in (4) and sends it downstream.
    end for
    Node k gathers $\{\widehat{\mathbf{z}}_q[t]\}_{q\in\mathcal{U}_k}$ and builds $\widehat{\mathbf{y}}_k[t]$ as in (5).
    Node k estimates $\mathbf{R}_{\widehat{\mathbf{y}}_k\widehat{\mathbf{y}}_k}$ and $\mathbf{R}_{\widehat{\mathbf{y}}_k\mathbf{d}_k}$.
    Node k computes $\widehat{\mathbf{W}}_k$ as in (11) then $\widehat{\mathbf{d}}_{k,\text{TI-dMWF}}[t]$.
  end for
end for

no longer a sufficient statistic and the algorithm does not reach the centralized MWF. This suboptimality is quantified in Section VI-B.

E. PRACTICAL CONSIDERATIONS

Since the pruned tree (and hence the fusion matrices) depends on the root, each non-root node q generally maintains a distinct fusion matrix $\widehat{\mathbf{P}}_q$ for each possible root $k \in \mathcal{K}_q$, i.e., $K-1$ in total, with the tree (re-)oriented per root.

When the topology changes (e.g., due to a link failure or a node joining/leaving), the pruned tree is recomputed for each affected root, and the fusion matrices $\widehat{\mathbf{P}}_q$ re-estimated only at nodes q whose upstream set $\mathcal{U}_q$ (hence $\widehat{\mathbf{y}}_q$) changes. The matrices of unaffected nodes remain valid. We advocate a shortest-path tree (SPT) [25] rooted at each k: since the downstream flow runs every frame while the upstream flow runs every N_{us} frames, the dominant per-frame latency grows with the tree *depth*, which an SPT minimizes.

The flooding of $\mathbf{y}_{q\uparrow}$ need not be synchronous: since the upstream flow is decoupled from the per-frame downstream flow, node q may compute $\widehat{\mathbf{P}}_q$ from a delayed $\mathbf{y}_{q\uparrow}[t']$ ($t' < t$), as the SCM estimation relies on time-averaged statistics (cf. (3)) that vary slowly relative to the frame rate.

The TI-dMWF also accommodates *mixed networks* where only some nodes apply a fusion operation. A node q for which fusion would not reduce the forwarded dimension ($\widehat{M}_q \leq \overset{\circ}{Q}$) instead forwards $\widehat{\mathbf{y}}_q$ unfused, which is then treated by its downstream neighbor as additional channels in (5). Theorem 1 optimality is preserved, as forwarding unfused signals retains the global-sources subspace information and the cascaded-fusion argument of Appendix A holds for any subset of fusing nodes.

The number of global sources $\overset{\circ}{Q}$, setting the dimension of $\mathbf{y}_{q\uparrow}$ in (7), must in practice be estimated by source enumeration (cf. Section III-A). For a mis-estimated $\overset{\circ}{Q}\prime \neq \overset{\circ}{Q}$, fixed network-wide for a given root k, every downstream node targets the same $\overset{\circ}{Q}\prime$-dimensional signal in (8). Underestimation ($\overset{\circ}{Q}\prime < \overset{\circ}{Q}$) makes $\mathbf{y}_{q\uparrow}$ span only a $\overset{\circ}{Q}\prime$-dimensional subset of the global-sources subspace, losing information and yielding a suboptimal estimate. Overestimation ($\overset{\circ}{Q}\prime > \overset{\circ}{Q}$) preserves optimality at the cost of a higher communication cost ($\overset{\circ}{Q}\prime$-dimensional exchanges), provided $\mathbf{R}_{\widehat{\mathbf{y}}_q\widehat{\mathbf{y}}_q}$ stays full-rank with the extra channels, which holds generically owing to per-sensor self-noise. This assumes $\overset{\circ}{Q}\prime \leq M_k$; the case $\overset{\circ}{Q}\prime > M_k$, for which (7) is ill-defined, is left for future work (cf. Section VII).

IV. COMMUNICATION COST ANALYSIS

We analyze and compare the communication cost of the TI-dMWF and TI-DANSE, assuming peer-to-peer signal exchanges and a TI-dMWF upstream data flow every $N_{\text{us}} \geq 1$ frames (cf. Section III). Note that TI-DANSE$^+$ has the same communication cost as TI-DANSE [13].

For one updating node k, the upstream data flow exchanges $N_{\text{up}} = (K-1)\overset{\circ}{Q}$ signal channels (propagating $\mathbf{y}_{q\uparrow}$ from root to leaves), and similarly for the downstream data flow, $N_{\text{down}} = (K-1)\overset{\circ}{Q}$. As the downstream flow runs continuously, the per-frame total is between N_{down} and $N_{\text{down}} + N_{\text{up}}$, depending on the upstream-flow frequency. Both flows are repeated for each node k. Some exchanges could be saved when nodes share upstream or downstream neighbors, but for simplicity we leave this optimization for future work.

For comparison, TI-DANSE exchanges $N_{\text{TI-D}} = 2(K-1)Q^{(\text{s})}$ signal channels per iteration [12]. While this is lower than the TI-dMWF, TI-DANSE requires *multiple iterations* to approach the centralized MWF. The single-pass nature of the TI-dMWF thus generally yields a lower *cumulative* communication cost than the iterative baselines whenever near-centralized accuracy is required. This has been verified in numerical experiments, omitted here for conciseness.

V. LATENCY AND COMPLEXITY ANALYSIS

Real-time feasibility depends on the per-frame latency L_k, dominated in the TI-dMWF by the downstream data flow at every frame. Let δ_{hop} be the average per-hop delay (transmission and per-node time to compute (4)). Since each node q must gather its upstream neighbors' fused signals before transmitting its own fused signals, the per-frame latency for root k is $L_k = \Delta_k \delta_{\text{hop}}$, with Δ_k the pruned-tree depth.

Topology-independence lets one choose a pruning strategy minimizing this depth: whereas an MST minimizes total communication power but may yield deeper trees, an SPT rooted at k minimizes Δ_k [25]. The real-time constraint $L_k < T_{\text{shift}}$ (with T_{shift} the STFT frame shift) requires $\delta_{\text{hop}} < T_{\text{shift}}/\Delta_k$; for $T_{\text{shift}} = 20$ ms and $\Delta_k = 4$, this allows δ_{hop} up to 5 ms, which is attainable by low-latency wireless protocols.

A separate *discovery latency* is associated with the upstream signal flooding. Since the discovery step runs only every N_{us} frames, independently of the per-frame estimation path, the upstream statistics may be delivered asynchronously or with delay without affecting optimality in slowly varying environments. Moreover, once the SCMs are well-estimated, the TI-dMWF reaches centralized performance in a single discovery cycle.

The per-frame computational complexity of the TI-dMWF stays low by decoupling the frequent downstream fusion from the infrequent upstream discovery. In the *estimation step*, each node performs a single matrix-vector product (4), of complexity $O(\overset{\circ}{Q}\widehat{M}_q)$, plus per-frame rank-1 SCM updates (3) costing $O(\widehat{M}_q^2 + \overset{\circ}{Q}\widehat{M}_q)$. The most

intensive task, solving (8) for $\widehat{\mathbf{P}}_q$ at $O(\widehat{M}_q^3)$, runs only in the *discovery step*, every N_{us} frames, giving a mean per-frame complexity $O(\overset{\circ}{Q}\widehat{M}_q + \widehat{M}_q^2 + \widehat{M}_q^3/N_{\text{us}})$. In contrast, TI-DANSE has a cheaper per-iteration update (an LMMSE problem of size $M_k + \overset{\circ}{Q}$ at the updating node) but needs many such updates across frames. Finally, a topology change costs only tree re-pruning ($O(K|E|\log K)$ per node via Dijkstra's algorithm [25] for K root-specific SPTs) plus fusion-matrix re-estimation at the affected nodes.

VI. NUMERICAL SIMULATIONS

This section reports four experiments. The first (Section VI-A) validates the framework using oracle SCMs in idealized scenarios. The second (Section VI-B) also uses oracle SCMs, isolating model mismatch from estimation errors to assess robustness to deviations from the GLS assumption. The third and fourth (Section VI-C) evaluate the TI-dMWF in simulated reverberant scenarios with SCMs estimated from the microphone signals, assessing respectively the convergence under estimated SCMs and the influence of the tree-pruning strategy.

A. VALIDATION UNDER ORACLE SCMS

As a first experiment, the framework is validated using oracle SCMs. The TI-dMWF is compared in an ideal setting to TI-DANSE [12] (with the filter-normalization of [26]) and TI-DANSE$^+$ [13], validating the TI-dMWF framework and assessing the convergence of the iterative baselines. Since TI-DANSE$^+$ is closely related to DANSF [13], [15], this also indirectly contrasts the TI-dMWF with the node-specific extension of the DASF framework. The SCMs at the updating node of TI-DANSE and TI-DANSE$^+$ may be singular in general GLS scenarios, so the comparison is conducted only in constrained global-local subspaces (cGLS) scenarios where all speech sources are observed by all nodes.

In the idealized setting, a topology-unconstrained WASN of $K = 6$ nodes with $M_k = 5, \forall\, k \in \mathcal{K}$ is deployed with $Q^{(\text{s})} = 2$ speech and $Q^{(\text{n})} = 2$ noise sources (so $M_k \geq Q, \forall\, k \in \mathcal{K}$). The steering matrix entries used in the signal model (1) are independently drawn from a standard normal distribution. Each node $k \in \mathcal{K}$ aims to estimate its own single-channel (D=1) desired signal d_k.

All SCMs needed for the MWFs in (2), (9), and (11) (and cf. [12], [13] for TI-DANSE and TI-DANSE$^+$) are computed from oracle knowledge of the steering matrices, source powers, and self-noise power. The centralized SCM is $\mathbf{R}_{\mathbf{yy}} = \mathbf{R}_{\mathbf{ss}} + \mathbf{R}_{\mathbf{nn}} + \mathbf{R}_{\mathbf{vv}} = \mathbf{A}\mathbf{R}_{\mathbf{ss}}^{\text{lat}}\mathbf{A}^{\text{H}} + \mathbf{B}\mathbf{R}_{\mathbf{nn}}^{\text{lat}}\mathbf{B}^{\text{H}} + \mathbf{R}_{\mathbf{vv}}$, with $\mathbf{R}_{\mathbf{ss}}^{\text{lat}}$, $\mathbf{R}_{\mathbf{nn}}^{\text{lat}}$ diagonal latent speech/noise power matrices and $\mathbf{R}_{\mathbf{vv}}$ the (diagonal) self-noise SCM. The latent power is set to 1 for every speech and noise source, and the self-noise power to 0.01 for every sensor. The SCMs required for the distributed algorithms are computed based on the centralized SCMs and the fusion matrices. For instance, in the TI-dMWF estimation step, the SCM $\mathbf{R}_{\widehat{\mathbf{y}}_k\widehat{\mathbf{y}}_k}$ in (11) is computed as $\mathbf{R}_{\widehat{\mathbf{y}}_k\widehat{\mathbf{y}}_k} = \mathbb{E}\{\widehat{\mathbf{y}}_k\widehat{\mathbf{y}}_k^{\text{H}}\} = \mathbf{D}_k^{\text{H}}\mathbf{R}_{\mathbf{yy}}\mathbf{D}_k$, where $\mathbf{D}_k$ arranges the fusion matrices $\{\mathbf{P}_q\}_{q\in\mathcal{K}_k}$ (such that $\mathbf{W}_k^{\text{TI}} = \mathbf{D}_k\widetilde{\mathbf{W}}_k$, with $\mathbf{W}_k^{\text{TI}}$ the network-wide TI-dMWF filter at node k; see Appendix A). The other SCMs (for the TI-dMWF and for each TI-DANSE/TI-DANSE$^+$ iteration) follow analogously, omitted for brevity.

In each simulation, the WASN is generated as a non-fully connected graph with connectivity $C = 0.5$ (as defined in [13]). This graph is pruned to a tree using the multiple max-$|\mathcal{U}_k^i|$ trees (MMUT) strategy [13], a constrained variant of Kruskal's MST algorithm [27] shown to speed up TI-DANSE$^+$ convergence relative to MST pruning.

Two evaluation metrics are used: the network-wide filter error $\text{MSE}_W = \frac{1}{K}\sum_{k\in\mathcal{K}} \|\mathbf{W}_k - \widehat{\mathbf{W}}_k\|_{\text{F}}^2$, measuring the difference between the distributed ($\mathbf{W}_k$) and centralized ($\widehat{\mathbf{W}}_k$) network-wide filter coefficients,[1] and the desired-signal error $\text{MSE}_d = \frac{1}{K}\sum_{k\in\mathcal{K}} \|\mathbf{d}_k - \widehat{\mathbf{d}}_k\|_2^2$, quantifying the error in the desired-signal estimates.

Fig. 1(a) compares the TI-dMWF to TI-DANSE and TI-DANSE$^+$ on both metrics. The MSE_d is also computed for $\widehat{\mathbf{d}}_k$ set to (i) the unprocessed first sensor signal $y_{k,1}$ (D=1 here) and (ii) the local MWF estimate from $\mathbf{y}_k$ alone. Results are averaged over 20 random cGLS scenarios, using a geometric mean for MSE_W (for a clear logarithmic-axis depiction [13]) and an arithmetic mean for MSE_d.

To simulate observability patterns in cGLS scenarios, $K \times Q$ *observability matrices* (with entries 0 for an unobserved and 1 for an observed source) are randomly generated such that all speech sources are observed by all nodes and each noise source is observed by a single node or by all nodes.

The results show that the TI-dMWF achieves the same performance as the centralized MWF and does so without relying on iterations, which is visible through the machine-precision MSE_W and the equal MSE_d between the centralized MWF and TI-dMWF. Conversely, TI-DANSE requires many iterations to converge towards the centralized MWF. While TI-DANSE$^+$ converges faster than TI-DANSE (we use MMUT pruning at connectivity $C = 0.5$), it still requires many iterations to reach the centralized MWF. After about 60 iterations, its MSE_d nearly matches the centralized MWF even though its MSE_W has not reached that of the TI-dMWF, suggesting that a moderately low MSE_W can suffice in practice. To outperform the local MWF, TI-DANSE$^+$ and TI-DANSE require around 30 and 82 iterations, respectively, in this 6-node setup.

B. ROBUSTNESS TO GLS DEVIATIONS

The second experiment assesses TI-dMWF's sensitivity to deviations from the GLS assumption via a leakage parameter $\alpha \in [0, 1]$. Sources unobserved under the assumed GLS pattern leak into the corresponding microphone signals with gain α, so $\alpha = 0$ is the strict GLS scenario and increasing α injects each source's energy into the nodes the pattern treats as non-observing. The algorithm keeps operating under the assumed pattern as the actual signals depart from it. To isolate model mismatch, it uses oracle SCMs and the idealized setup of Section VI-A with random steering matrices.

Fig. 1(b) shows the TI-dMWF MSE_W and MSE_d versus α, with the centralized, local, and unprocessed references. At $\alpha = 0$ (strict GLS), the TI-dMWF attains the centralized MWF exactly, its network-wide filter coinciding with the centralized MWF to machine precision ($\text{MSE}_W \approx 10^{-16}$) and its estimate reaching the centralized MSE_d, as predicted by Theorem 1. Any non-zero leakage immediately breaks this exact optimality, with no finite tolerance band: MSE_W leaves machine precision and rises continuously. This matches Remark 1 of Section III-D: once a source is observed by a strict subset of nodes, the non-global components $\{\dot{\mathbf{y}}_q\}_{q\in\mathcal{K}}$ become inter-node correlated, the block-diagonality of $\dot{\mathbf{\Gamma}}$ in (18) is lost, and the fused signals are no longer a sufficient statistic. Exact optimality is confined to the strict GLS scenario, but away from it the TI-dMWF degrades smoothly and monotonically with α. Nonetheless, across

[1] For full expressions of the network-wide filters please refer to (18) for the TI-dMWF, [12] for TI-DANSE, and [13] for TI-DANSE$^+$.

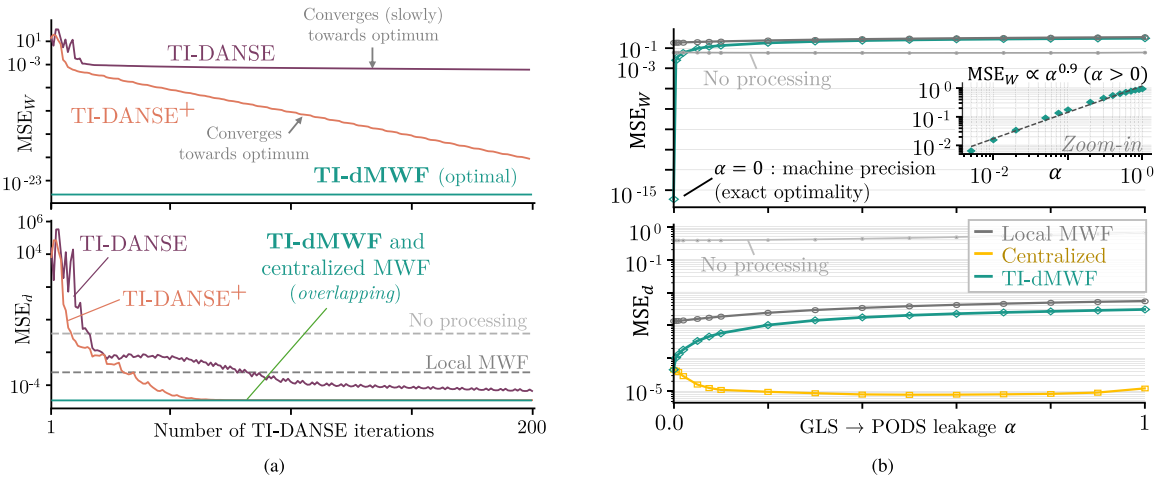


FIGURE 1. TI-dMWF performance, MSE_W (top) and MSE_d (bottom). (a) Versus iteration over 20 cGLS oracle scenarios ($C = 0.5$), comparing the TI-dMWF, TI-DANSE, and TI-DANSE$^+$. (b) Versus the observability-leakage α over 20 GLS oracle scenarios, with local, centralized, and unprocessed references (log-log inset).

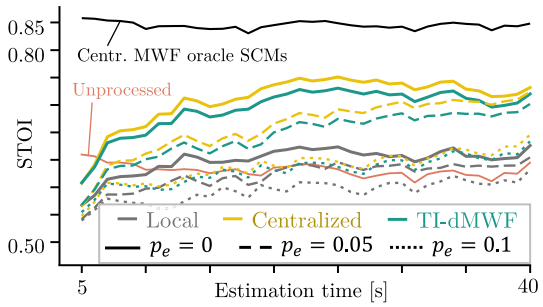


FIGURE 2. All-node mean STOI vs. time at $\beta = 0.99$, over 20 cGLS scenarios, for VAD error probabilities $p_e \in \{0, 0.05, 0.1\}$ (solid, dashed, dotted, respectively).

the whole α range the TI-dMWF network-wide filter stays closer to the centralized MWF than the local MWF, and its estimate stays more accurate than the local MWF and the unprocessed input.

C. PERFORMANCE UNDER ESTIMATED SCMS

Beyond the idealized setting of Section VI-A, the TI-dMWF is evaluated in reverberant scenarios with all SCMs estimated from the microphone signals. room impulse responses (RIRs) are generated with the image-source method (ISM) in Pyroomacoustics [28] for a $5 \times 5 \times 3$ m room with $T_{60} = 0.2$ s.

As in Section VI-A, the WASN has $K = 6$ nodes with $M_k = 5$ microphones (compact 3 cm-radius arrays), placed at least 0.25 m from the walls, with a $C = 0.5$ ad-hoc topology pruned via the SPT strategy unless stated otherwise. The scene contains $Q^{(\mathrm{s})} = 2$ speech sources (VCTK [29]) and $Q^{(\mathrm{n})} = 2$ babble noise sources, at least 1 m from any node, at 0 dB mean signal-to-noise ratio (SNR) (latent speech vs. noise), with sensor self-noise 20 dB below the source power. The cGLS observability follows Section VI-A: one noise source is observed by all nodes, one by a single node.

Microphone signals are obtained by time-domain convolution of the latent source signals with the RIRs and processed in a weighted overlap-add (WOLA) filter bank (1024-sample discrete Fourier transform (DFT), 50% overlap, Hann window). Speech sources follow a 3 s on-off cycle with randomized start times. The SCMs are estimated by VAD-based recursive averaging (3) with forgetting factor β, from a random initialization; an imperfect VAD flips each per-source frame decision with probability p_e. Each 40 s scenario is averaged over 20 runs.

Under estimated SCMs, a rank-$\mathring{Q}$ GEVD-based MWF [30] replaces the plain MWF of Section II (standard for robustness against estimation errors), applied equally to the centralized, local, and TI-dMWF estimators. Fusion matrices and filters are recomputed every $N_{\mathrm{us}} = 5$ frames (160 ms, cf. Section V) while the SCMs update every frame. Performance is assessed by means of the short-term objective intelligibility (STOI) [31] over sliding windows, averaged over all K nodes.

The third experiment tracks the TI-dMWF STOI over time, testing whether SCM estimation errors accumulate through its cascaded estimation-fusion structure (upstream fusion-matrix errors propagating towards the root). It is compared to the centralized GEVD-based multichannel Wiener filter (GEVD-MWF) (estimating a single $M \times M$ SCM), the local GEVD-MWF, and the unprocessed reference, at $\beta = 0.99$ and $p_e \in \{0, 0.05, 0.1\}$.

Fig. 2 shows that the online estimated-SCM filters reach a steady regime after an initial transient of about 15 s. At $\beta = 0.99$ and $p_e = 0$, the TI-dMWF closely tracks the centralized GEVD-MWF, with a converged all-node STOI of 0.76 versus 0.77, both clearly exceeding the local GEVD-MWF (0.66) and the unprocessed reference (0.63). The TI-dMWF thus matches the centralized estimated-SCM MWF to within a small margin, even though each of its nodes estimates only a small $\widehat{M}_q \times \widehat{M}_q$ SCM rather than the single $M \times M$ centralized SCM. SCM estimation errors thus do not accumulate through the cascaded estimation-fusion structure, despite upstream fusion-matrix errors propagating towards the root. Robustness to an imperfect VAD is shown by the $p_e = 0.05$ (dashed) and $p_e = 0.1$ (dotted) curves, for which the converged TI-dMWF STOI degrades monotonically to 0.74 and 0.66, respectively, while preserving the convergence behavior and the relative ordering of the estimators.

The fourth experiment assesses how the tree-pruning strategy affects SCM estimation-error propagation. The same scenarios are processed with five strategies: SPT, MMUT, and MST on the $C = 0.5$ ad-hoc topology, plus synthetic star and line topologies bounding the tree depth (1 for star, $K - 1$ for line). The per-frame downstream cost

is identical across strategies (any spanning tree of K nodes has $K-1$ edges, each carrying a $\overset{\circ}{Q}$-dimensional fused signal), so the choice is purely a latency-vs-performance trade-off.

The converged all-node TI-dMWF STOI and mean tree depth per strategy (20 cGLS scenarios, $\beta = 0.99$, $p_e = 0$) are summarized below; the tree-independent references are local 0.661, centralized 0.785, and unprocessed 0.626.

	star	SPT	MMUT	MST	line
Mean depth	1.0	2.6	3.1	3.6	4.0
STOI	0.792	0.767	0.767	0.755	0.757

The TI-dMWF performance varies only marginally, from 0.79 STOI for the shallow star topology (depth 1) to 0.76 for the deeper MST and line topologies, a spread within the inter-scenario variability, with a mild trend favoring shallower trees (fewer cascaded fusion stages). Since the per-frame communication cost is identical across strategies, the choice is governed by latency, for which the SPT is preferable as the minimum-depth tree for an arbitrary topology.

VII. CONCLUSION

We have introduced the TI-dMWF, an algorithm for distributed signal estimation in topology-unconstrained WASNs that lets each node compute its node-specific centralized MWF by exchanging only low-dimensional fused signals, without iterative filter estimation, unlike state-of-the-art algorithms. Simulations confirm that the TI-dMWF is optimal in GLS scenarios, where each source is observed by all nodes or only one. Latency and complexity analyses show that the TI-dMWF retains centralized optimality in the GLS setting with predictable latency, computational load, and communication cost, and degrades controllably when the GLS assumption only approximately holds. Future work may relax the GLS assumption and validate the algorithm in real-world scenarios beyond the present simulation study, and generalize the upstream signal construction (7) beyond the selection-matrix formulation to handle $\overset{\circ}{Q} > M_k$ at the root node.

A. APPENDIX A: TI-DMWF OPTIMALITY

We first show that a leaf node $l \in \mathcal{K}_k$ estimating $\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k$ from its local signals $\mathbf{y}_l$ obtains an LMMSE filter usable as a fusion matrix that compresses $\mathbf{y}_l$ into a lower-dimensional fused signal, which its downstream neighbor can then use to estimate $\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k$ without performance loss. Applying this recursively from the leaves to the root node k yields the centralized LMMSE estimator of $\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k$ at node k, achievable in practice by having each node transmit estimates of $\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\mathbf{y}_k$. The same process finally yields the centralized LMMSE estimator for $\mathbf{d}_k$ at node k.

For convenience, we rewrite the signal model as $\mathbf{y}_k = \overset{\circ}{\mathbf{y}}_k + \dot{\mathbf{y}}_k = \overset{\circ}{\mathbf{H}}_k\overset{\circ}{\mathbf{x}}^{\mathrm{lat}} + \dot{\mathbf{y}}_k$, where $\overset{\circ}{\mathbf{H}}_k \in \mathbb{C}^{M_k \times \overset{\circ}{Q}}$ is the steering matrix relative to all *global* sources in the environment and $\overset{\circ}{\mathbf{x}}^{\mathrm{lat}} \in \mathbb{C}^{\overset{\circ}{Q}\times 1}$ is the vector of their latent source signals. Consider the problem of estimating $\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k \in \mathbb{C}^{\overset{\circ}{Q}\times 1}$ at a node $l \in \mathcal{K}_k$ based solely on its local sensor signals $\mathbf{y}_l$. The corresponding LMMSE formulation is:

$$\widetilde{\mathbf{W}}_l = \underset{\mathbf{W}_l \in \mathbb{C}^{M_l \times \overset{\circ}{Q}}}{\arg\min}\ \mathbb{E}\left\{\left\|\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k - \mathbf{W}_l^{\mathrm{H}}\mathbf{y}_l\right\|_2^2\right\} \tag{12}$$

$$= \mathbf{R}_{\mathbf{y}_l\mathbf{y}_l}^{-1}\mathbf{R}_{\mathbf{y}_l\overset{\circ}{\mathbf{y}}_k}\mathbf{E}_{\overset{\circ}{Q}}, \tag{13}$$

where $\mathbf{R}_{\mathbf{y}_l\overset{\circ}{\mathbf{y}}_k} = \mathbb{E}\{\mathbf{y}_l\overset{\circ}{\mathbf{y}}_k^{\mathrm{H}}\} = \overset{\circ}{\mathbf{H}}_l\mathbf{R}_{\overset{\circ}{\mathbf{x}}\overset{\circ}{\mathbf{x}}}\overset{\circ}{\mathbf{H}}_k^{\mathrm{H}}$ with $\mathbf{R}_{\overset{\circ}{\mathbf{x}}\overset{\circ}{\mathbf{x}}} = \mathbb{E}\{\overset{\circ}{\mathbf{x}}^{\mathrm{lat}}(\overset{\circ}{\mathbf{x}}^{\mathrm{lat}})^{\mathrm{H}}\}$. The MWF (13) can be expanded using the matrix inversion lemma to obtain:

$$\widetilde{\mathbf{W}}_l = \left(\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l} + \overset{\circ}{\mathbf{H}}_l\mathbf{R}_{\overset{\circ}{\mathbf{x}}\overset{\circ}{\mathbf{x}}}\overset{\circ}{\mathbf{H}}_l^{\mathrm{H}}\right)^{-1}\overset{\circ}{\mathbf{H}}_l\mathbf{R}_{\overset{\circ}{\mathbf{x}}\overset{\circ}{\mathbf{x}}}\overset{\circ}{\mathbf{H}}_k^{\mathrm{H}}\mathbf{E}_{\overset{\circ}{Q}} \tag{14}$$

$$= \underbrace{\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\overset{\circ}{\mathbf{H}}_l}_{M_l\times\overset{\circ}{Q}}\underbrace{\left(\mathbf{I}_{\overset{\circ}{Q}} - \mathbf{X}_l^{-1}\overset{\circ}{\mathbf{H}}_l^{\mathrm{H}}\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\overset{\circ}{\mathbf{H}}_l\right)\mathbf{R}_{\overset{\circ}{\mathbf{x}}\overset{\circ}{\mathbf{x}}}\overset{\circ}{\mathbf{H}}_k^{\mathrm{H}}\mathbf{E}_{\overset{\circ}{Q}}}_{=\mathbf{M}_{lk}\in\mathbb{C}^{\overset{\circ}{Q}\times\overset{\circ}{Q}}}, \tag{15}$$

$$\text{with } \mathbf{X}_l = \mathbf{R}_{\overset{\circ}{\mathbf{x}}\overset{\circ}{\mathbf{x}}}^{-1} + \overset{\circ}{\mathbf{H}}_l^{\mathrm{H}}\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\overset{\circ}{\mathbf{H}}_l, \tag{16}$$

and where $\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l} = \mathbb{E}\{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l^{\mathrm{H}}\}$ is full-rank and square, thus invertible. The term $\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\overset{\circ}{\mathbf{H}}_l$ in (15) acts as a fusion matrix compressing $\mathbf{y}_l$ into a $\overset{\circ}{Q}$-dimensional signal, post-multiplied by $\mathbf{M}_{lk}$ to obtain the LMMSE filter $\widetilde{\mathbf{W}}_l$.

Now consider an internal node l' (direct downstream neighbor of l) estimating $\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k$ from $\mathbf{y}_{l'}$ and the $\mathbf{y}_l$ transmitted by node l. The corresponding LMMSE problem is:

$$\widetilde{\mathbf{W}}_{l'l} = \underset{\mathbf{W}_{l'l}\in\mathbb{C}^{(M_{l'}+M_l)\times\overset{\circ}{Q}}}{\arg\min}\ \mathbb{E}\left\{\left\|\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k - \mathbf{W}_{l'l}^{\mathrm{H}}\begin{bmatrix}\mathbf{y}_{l'}\\ \mathbf{y}_l\end{bmatrix}\right\|_2^2\right\}. \tag{17}$$

Following a similar reasoning as (12)–(16), an alternative expression for $\widetilde{\mathbf{W}}_{l'l}$ can be obtained:

$$\widetilde{\mathbf{W}}_{l'l} = \begin{bmatrix}\mathbf{R}_{\dot{\mathbf{y}}_{l'}\dot{\mathbf{y}}_{l'}}^{-1}\overset{\circ}{\mathbf{H}}_{l'}\\ \mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\overset{\circ}{\mathbf{H}}_l\end{bmatrix}\mathbf{M}_{l'lk}. \tag{18}$$

$$\mathbf{M}_{l'lk} = \left(\mathbf{I}_{\overset{\circ}{Q}} - \mathbf{X}_{l'l}^{-1}\overset{\circ}{\mathbf{H}}_{l'l}^{\mathrm{H}}\dot{\mathbf{\Gamma}}\overset{\circ}{\mathbf{H}}_{l'l}\right)\mathbf{R}_{\overset{\circ}{\mathbf{x}}\overset{\circ}{\mathbf{x}}}\overset{\circ}{\mathbf{H}}_k^{\mathrm{H}}\mathbf{E}_{\overset{\circ}{Q}} \tag{19}$$

$$\mathbf{X}_{l'l} = (\mathbf{R}_{\overset{\circ}{\mathbf{x}}\overset{\circ}{\mathbf{x}}})^{-1} + \overset{\circ}{\mathbf{H}}_{l'l}^{\mathrm{H}}\dot{\mathbf{\Gamma}}\overset{\circ}{\mathbf{H}}_{l'l}, \tag{20}$$

with $\mathbf{R}_{\dot{\mathbf{y}}_{l'}\dot{\mathbf{y}}_{l'}} = \mathbb{E}\{\dot{\mathbf{y}}_{l'}\dot{\mathbf{y}}_{l'}^{\mathrm{H}}\}$, $\overset{\circ}{\mathbf{H}}_{l'l} = [\overset{\circ}{\mathbf{H}}_{l'}^{\mathrm{T}} \mid \overset{\circ}{\mathbf{H}}_l^{\mathrm{T}}]^{\mathrm{T}}$, and $\dot{\mathbf{\Gamma}} = \mathrm{blkdiag}\{\mathbf{R}_{\dot{\mathbf{y}}_{l'}\dot{\mathbf{y}}_{l'}}^{-1}, \mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\}$. As in (15), the leftmost products $\mathbf{R}_{\dot{\mathbf{y}}_{l'}\dot{\mathbf{y}}_{l'}}^{-1}\overset{\circ}{\mathbf{H}}_{l'}$ and $\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\overset{\circ}{\mathbf{H}}_l$ in (18) act as fusion matrices compressing $\mathbf{y}_{l'}$ and $\mathbf{y}_l$ into $\overset{\circ}{Q}$-dimensional signals.

A fusion matrix is *valid* (in the sense of this proof) if an LMMSE problem using the fused signals yields the same result as one using the original signals. The matrices $\mathbf{R}_{\dot{\mathbf{y}}_{l'}\dot{\mathbf{y}}_{l'}}^{-1}\overset{\circ}{\mathbf{H}}_{l'}$ and $\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\overset{\circ}{\mathbf{H}}_l$ are valid, being the LMMSE filters of nodes l' and l when estimating $\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k$ from $\mathbf{y}_{l'}$ and $\mathbf{y}_l$ alone, cf. (15) (proof omitted).

Hence node l''s estimate of $\mathbf{E}_{\overset{\circ}{Q}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k$ from $\mathbf{y}_{l'}$ and $\mathbf{y}_l$ is unchanged if node l transmits the fused $(\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\overset{\circ}{\mathbf{H}}_l)^{\mathrm{H}}\mathbf{y}_l \in \mathbb{C}^{\overset{\circ}{Q}\times 1}$ instead of $\mathbf{y}_l \in \mathbb{C}^{M_l\times 1}$. Moreover, post-multiplying $\mathbf{R}_{\dot{\mathbf{y}}_l\dot{\mathbf{y}}_l}^{-1}\overset{\circ}{\mathbf{H}}_l$ by any invertible $\overset{\circ}{Q}\times\overset{\circ}{Q}$ matrix (e.g., $\mathbf{M}_{lk}$ in (15) or $\mathbf{M}_{l'lk}$ in (18)) gives an equally valid fusion

matrix (proof omitted), so the estimate is likewise unchanged if l transmits $\widetilde{\mathbf{W}}_l^{\mathrm{H}}\mathbf{y}_l \in \mathbb{C}^{\overset{\circ}{Q}\times 1}$ instead of $\mathbf{y}_l$.

Formula (18) extends to any number of cooperating nodes. In a tree, each *leaf* node l first computes a valid fusion matrix $\widetilde{\mathbf{W}}_l$ from its local observation $\mathbf{y}_l$ as in (15). Each *internal* node p then combines the fused signals from its upstream neighbors with $\mathbf{y}_p$ to compute its own LMMSE filter for $\mathbf{E}_{\underset{Q}{\circ}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k$, of the form (18), which serves as its fusion matrix to compress all available signals into a single $\overset{\circ}{Q}$-dimensional signal. Repeating this up to the root node k yields the centralized LMMSE estimator for $\mathbf{E}_{\underset{Q}{\circ}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k$ there.

Since $\overset{\circ}{\mathbf{y}}_k$ is not directly available in practice, $\mathbf{E}_{\underset{Q}{\circ}}^{\mathrm{T}}\overset{\circ}{\mathbf{y}}_k$ can be replaced by $\mathbf{E}_{\underset{Q}{\circ}}^{\mathrm{T}}\mathbf{y}_k$ (cf. (7)) in the LMMSE estimation at all nodes but k without changing the obtained filters since $\dot{\mathbf{y}}_k$ is uncorrelated to $\dot{\mathbf{y}}_l$ by definition for any $l \in \mathcal{K}_k$. Once node k has received the same fused signals, the centralized LMMSE estimator for $\mathbf{d}_k = \mathbf{E}_k^{\mathrm{T}}\mathbf{s}$ is likewise computed at k, with a filter analogous to (18) (omitted for brevity). This concludes the proof of the optimality of the TI-dMWF.

REFERENCES

[1] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *Proc. IEEE Symp. Commun. Veh. Technol.*, 2011, pp. 1–6.

[2] S. Natarajan et al., "A comprehensive review of beamforming-based speech enhancement techniques, IoT, and smart city applications," in *Proc. Ind. Electron. Soc. Annu. On-Line Conf.*, 2023, pp. 1–6.

[3] 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/ACM Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 1, pp. 38–51, Jan. 2009.

[4] 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.

[5] A. Hassani, A. Bertrand, and M. Moonen, "GEVD-based low-rank approximation for distributed adaptive node-specific signal estimation in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2557–2572, May 2016.

[6] A. Lohmann, T. van Waterschoot, J. Bitzer, and S. Doclo, "Dereverberation in acoustic sensor networks using weighted prediction error with microphone-dependent prediction delays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2023, pp. 1–5.

[7] R. Chang, Z. Chen, and F. Yin, "Distributed Kalman filtering for speech dereverberation and noise reduction in acoustic sensor networks," *IEEE Sensors J.*, vol. 23, no. 24, pp. 31027–31037, Dec. 2023.

[8] A. Griffin, A. Alexandridis, D. Pavlidi, Y. Mastorakis, and A. Mouchtaris, "Localizing multiple audio sources in a wireless acoustic sensor network," *Signal Process.*, vol. 107, pp. 54–67, 2015.

[9] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Commun. Mobile Comput.*, vol. 20, no. 17, pp. 1–24, 2017.

[10] J. Benesty, S. Makino, and J. Chen, *Speech Enhancement*. Berlin, Germany: Springer, Mar. 2006.

[11] P. Didier et al., "Distributed multichannel wiener filtering for wireless acoustic sensor networks," 2025. Accessed: Aug. 17, 2026. [Online]. Available: https://arxiv.org/abs/2603.09735

[12] J. Szurley, A. Bertrand, and M. Moonen, "Topology-independent distributed adaptive node-specific signal estimation in wireless sensor networks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 1, pp. 130–144, Mar. 2017.

[13] P. Didier, T. van Waterschoot, S. Doclo, J. Bitzer, and M. Moonen, "Fast-converging distributed signal estimation in topology-unconstrained wireless acoustic sensor networks," *IEEE Trans. Signal Inf. Process. Netw*, vol. 12, pp. 568–584, 2026 .

[14] C. A. Musluoglu and A. Bertrand, "A unified algorithmic framework for distributed adaptive signal and feature fusion problems–Part I: Algorithm derivation," *IEEE Trans. Signal Process.*, vol. 71, pp. 1863–1878, 2023.

[15] C. A. Musluoglu and A. Bertrand, "A distributed adaptive algorithm for node-specific signal fusion problems in wireless sensor networks," in *Proc. Eur. Signal Process. Conf.*, 2023, pp. 1654–1658.

[16] M. H. Bahari et al., "Distributed multi-speaker voice activity detection for wireless acoustic sensor networks," 2017, *arXiv:1703.05782* .

[17] Y. Zhao, J. K. Nielsen, J. Chen, and M. G. Christensen, "Model-based distributed node clustering and multi-speaker speech presence probability estimation in wireless acoustic sensor networks," *J. Acoustical Soc. Amer.*, vol. 147, no. 6, pp. 4189–4201, 2020.

[18] M. Souden, J. Chen, J. Benesty, and S. Affes, "Gaussian model-based multichannel speech presence probability," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 5, pp. 1072–1077, Jul. 2010.

[19] M. Tammen, D. Fischer, B. T. Meyer, and S. Doclo, "DNN-based speech presence probability estimation for multi-frame single-microphone speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 191–195.

[20] R. C. Prim, "Shortest connection networks and some generalizations," *Bell Syst. Tech. J.*, vol. 36, no. 6, pp. 1389–1401, 1957.

[21] D. Williams and D. Johnson, "Using the sphericity test for source detection with narrow-band passive arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 38, no. 11, pp. 2008–2014, Nov. 1990.

[22] R. Brcich, A. Zoubir, and P. Pelin, "Detection of sources using bootstrap techniques," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 206–215, Feb. 2002.

[23] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[24] M. Wax and A. Adler, "Detection of the number of signals by signal subspace matching," *IEEE Trans. Signal Process.*, vol. 69, pp. 973–985, 2021.

[25] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[26] P. Didier, T. van Waterschoot, and M. Moonen, "Topology-independent GEVD-based distributed adaptive node-specific signal estimation in ad-hoc wireless acoustic sensor networks," in *Proc. Eur. Signal Process. Conf.*, 2024, pp. 1–5.

[27] J. B. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," *Proc. Amer. Math. Soc.*, vol. 7, no. 1, pp. 48–50, 1956.

[28] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 351–355.

[29] C. Veaux, J. Yamagishi, and K. MacDonald, "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit," The Centre for Speech Technology Research (CSTR), Univ. Edinburgh, Edinburgh, U.K., Tech. Rep., 2017. Accessed: Aug. 14, 2025.

[30] 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.

[31] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.