Rate-constrained noise reduction in wireless acoustic sensor networks

Amini, Jamal; Hendriks, Richard; Heusdens, Richard; Guo, Meng; Jensen, Jesper

DOI
10.1109/TASLP.2019.2947777

Publication date
2020

Document Version
Accepted author manuscript

Published in
IEEE/ACM Transactions on Audio Speech and Language Processing

Rate-Constrained Noise Reduction in Wireless Acoustic Sensor Networks

Jamal Amini, Richard C. Hendriks, Richard Heusdens, Meng Guo and Jesper Jensen

***Abstract*—Wireless acoustic sensor networks (WASNs) can be used for centralized multi-microphone noise reduction, where the processing is done in a fusion center (FC). To perform the noise reduction, the data needs to be transmitted to the FC. Considering the limited battery life of the devices in a WASN, the total data rate at which the FC can communicate with the different network devices should be constrained. In this paper, we propose a rate-constrained multi-microphone noise reduction algorithm, which jointly finds the best rate allocation and estimation weights for the microphones across all frequencies. The optimal linear estimators are found to be the quantized Wiener filters, and the rates are the solutions to a filter-dependent reverse water-filling problem. The performance of the proposed framework is evaluated using simulations in terms of mean square error and predicted speech intelligibility. The results show that the proposed method is very close in performance to that of the existing optimal method based on discrete optimization. However, the proposed approach can do this at a much lower complexity, while the existing optimal reference method needs a non-tractable exhaustive search to find the best rate allocation across microphones.**

***Index Terms*—Wireless acoustic sensor networks, multi-microphone noise reduction, rate-distortion trade-off.**

I. Introduction

WIRELESS acoustic sensor networks (WASNs) can provide increased spatial diversity [1], [2], leading to better noise reduction performance compared to single-microphone noise reduction systems. As a realistic example, consider binaural hearing aids (HAs), potentially extended with additional assistive devices, collaborating with each other through a wireless link [3]. Thanks to the increased number of microphones as well as the increased spatial diversity, they can enhance the speech intelligibility and quality for hearing-impaired listeners [4], [5]. This can be achieved by performing the noise reduction (estimation) process in a distributed way, e.g., [6]–[8] or by aggregating the microphone observations of the network nodes at a fusion center (FC) followed by estimation of the source of interest and suppression of the environmental noise. In the case of an FC, in practice, one of the nodes in the network (e.g., one of the HAs) could be selected as the FC.

J. Amini, R. C. Hendriks and R. Heusdens are with the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, the Netherlands e-mails: {j.amini, r.c.hendriks, r.heusdens}@tudelft.nl

J. Jensen and M. Guo are with Oticon A/S, Kongebakken 9, 2765 Smørum, Denmark, e-mails:{megu, jesj}@oticon.com

J. Jensen is also with Electronic Systems Department, Aalborg University, 9100 Aalborg, Denmark

This work was supported by the Oticon Foundation and NWO, the Dutch Organisation for Scientific Research.

One common approach for noise reduction is the multi-channel Wiener filter (MWF) [9], which is the linear minimum mean square error (MMSE) estimator [10], [11]. Although the original typical MWF considers situations where all microphones are integrated into the same device, many examples exist, where the microphones are distributed over multiple wirelessly connected devices. A well-known example is the binaural MWF [11]–[14], where the microphone recordings of both HAs are combined to calculate two target signal estimates, one for each ear of the user. Another more general example can be found in [15] where an MWF-based filter is proposed for spatially distributed microphones. Note that in all these methods, the microphone signals are assumed to be available error free at the fusion center.

To limit the scope of this work, we consider the situation where the processing of the microphone signals in the WASN is performed in an FC. To combine the observations at the FC, the actual (realization of the) microphone signals must be transmitted to the FC. As the transmission powers of the devices may be limited due to limited battery life-time, the data needs to be compressed/quantized at a certain data rate. The process of quantization, however, introduces errors in the representation of the microphone signals, and therefore errors in the final target signal estimation. This introduces a trade-off between the data rate and the estimation accuracy (or error) [16], which links the noise reduction problem to the data compression problem.

Several rate-constrained beamforming (noise reduction) algorithms have been introduced in the literature to consider the rate of transmission as a resource constraint in the beamforming process, e.g., [16]–[19]. Assuming all sources to be jointly Gaussian random processes and using Wyner-Ziv coding [20], [21], a binaural rate-constrained beamformer has been proposed in [17, Sec. III-A]. This beamformer is limited to two devices (i.e., two HAs), which efficiently trades off the data rate against the beamforming performance. The method inevitably assumes that the joint statistics (for example cross-correlations) between the two HAs are known in both devices, which is limiting in practice. Moreover, an infinitely long sequence with a sophisticated decoder is needed to implement the proposed framework, which essentially provides a bound on the possible performance. Finally, this method is limited to the case of only two processing nodes (potentially with multiple microphones per node). The more generalized setup, which may include assistive devices is not considered in this method. Unlike [17, Sec. III-A], sub-optimal rate-constrained beamformers have been proposed in [17, Sec. III-B], [16], [18], [19], which do not suffer from the requirement that the

joint statistics should be known. Typically, these approaches also only consider two collaborating devices. Although these methods are simpler and computationally less expensive than [17, Sec. III-A], they combine all the observations from one device (HA), say, device A, into a single-channel observation, without considering the correlation of the HA observations with the observations from the other HA, say device B, and transmit it to the other device (which serves as an FC). With such a sub-optimal combination, important information may get lost and the performance does not approach the optimal performance, not even asymptotically, at infinitely high data rates [16]. In fact, due to the local combination of the multiple realizations into a single realization, the acoustic scene dependency is not taken into account in the existing sub-optimal approaches.

Assuming the WASN consists of more than two devices (e.g., two hearing aids and multiple additional assistive devices), in this paper, we obtain a generalized rate-constrained noise reduction formulation, which can be interpreted as a chief executive officer (CEO) problem (as in information theory), first introduced in [22]. The FC can be thought of as a CEO and the microphones as agents. Each agent records a version of the signal of interest to be transmitted to the FC. As the devices in the WASN have limited battery lifetime, and that the power usage is proportional to the data rate (measured in bits) [23], there will be a limited bit rate available for transmitting/receiving the information to/from the agents. Agents should be prioritized (for the estimation task) based on the importance of the information they may have about the target signal. In addition, in our setup, as microphone signals may have generally non-flat power spectral densities, the rate-constrained estimation problem should be frequency dependent. Therefore, depending on the acoustic scene, it is reasonable to share the total data rate across different agents and different frequency components. In [24] a similar problem is studied for rate allocation and strategy selection in an operational rate-constrained beamforming task, given discrete sets of strategy candidates and operating rates. The method uses a discrete optimization algorithm, based on the Lagrange multiplier technique [25], to select the best candidates and operating rates in different frequencies. However, because of the discrete nature of the optimization problem, an exhaustive search is necessary for the rate allocation across agents, which is practically affordable only for a small-size microphone array.

In this paper, we propose a joint quantization-estimation algorithm for the rate-constrained noise reduction task. We consider a linear estimation task at the FC and propose an optimization problem to both, allocate the total bit rate budget to different microphones in different frequencies (i.e., the quantization part), as well as to find the best filter weights (i.e., the estimation part), minimizing a rate-constrained estimation error. Unlike [24] which treated the problem sequentially with separate quantization and estimation tasks, in this work we consider the joint quantization-estimation problem. Moreover, unlike the exhaustive search for rate allocation across microphones proposed in [24], which is only good for small microphone arrays, we propose to optimize the rate allocations across frequency and space (i.e., devices). The proposed solution is scalable to arbitrarily big microphone arrays. For an MSE criterion, under certain assumptions, the optimal weights are found to be rate-constrained Wiener filter coefficients and the optimal rate allocation is the solution to a reverse "water-filling" problem. An MSE-based performance measure and an instrumental speech intelligibility measure are used to evaluate the proposed framework and the proposed method outperforms equal/random rate allocation strategies. Moreover, the proposed method performs almost as good as the optimal non-polynomial discrete optimization that involves the infeasible exhaustive search [24], in most practical scenarios.

The paper is organized as follows. In Sec. II-A the acoustical signal model is stated and the linear estimation task is introduced in Sec. II-B. The quantization aware beamforming problem is introduced in Sec. II-C. In Sec. II-D the proposed rate-constrained noise reduction problem formulation is presented in a unified framework and the proposed solution is described in Sec. III. The performance analysis of the proposed and existing methods is carried out in Sec. IV. Finally, Sec. V concludes the paper.

II. Problem Statement

A. Signal Model

We consider a microphone array consisting of M microphones, assumed to be embedded in different devices (i.e., HAs and/or assistive devices) placed at potentially different locations in space. Devices (agents) only communicate with an FC (and not with each other). Only the FC has access to the joint statistics. Each device can be equipped with more than one microphone. In this paper, it is assumed that for each device, the unprocessed microphone signals will be transmitted to the FC without pre-filtering stages, i.e., the microphone signals per device are not combined (pre-filtered) to a single signal. All microphones capture, in addition to the interferers, their version of the target speech signal, filtered by the acoustic channel, which is characterized by the room impulse response. In the short-time frequency transform (STFT) domain, we denote the target signal by $S_i \in \mathbb{C}$, with i the discrete frequency bin index. For notational convenience, the time-frame index is left out. The target speech is degraded by interfering noise, which might originate from, e.g., interfering point sources, diffuse noise, and/or microphone self-noise. The interfering noise observed at a particular microphone and at a particular frequency is indicated by $N_{ij} \in \mathbb{C}$, with $j = 1, \ldots, M$ being the microphone index. The signals S_i and N_{ij}, are assumed to be additive and mutually uncorrelated. Therefore, the microphone signal model can be written as

$$Y_{ij} = A_{ij} S_i + N_{ij} \in \mathbb{C}, \tag{1}$$

where $A_{ij} \in \mathbb{C}$ is the acoustic transfer function (ATF) between the target signal and the jth microphone. The signal model can be rewritten in vector notation by stacking all microphone signals in a vector, as

$$\mathbf{y}_i = \mathbf{a}_i S_i + \mathbf{n}_i = \mathbf{x}_i + \mathbf{n}_i \in \mathbb{C}^M, \tag{2}$$

where

$$\mathbf{y}_i = [Y_{i1}, ..., Y_{iM}]^{\mathrm{T}},$$

and similarly for $\mathbf{a}_i$ and $\mathbf{n}_i$, where the superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose operator on vectors/matrices. Since the signals S_i and N_{ij} are assumed to be uncorrelated, the power spectral density (PSD) matrix $\mathbf{\Phi}_{\mathbf{y}_i} = \mathrm{E}[\mathbf{y}_i\mathbf{y}_i^{\mathrm{H}}]$ of the vector $\mathbf{y}_i$ is given by

$$\mathbf{\Phi}_{\mathbf{y}_i} = \mathbf{\Phi}_{\mathbf{x}_i} + \mathbf{\Phi}_{\mathbf{n}_i} \in \mathbb{C}^{M\times M}, \tag{3}$$

where

$$\mathbf{\Phi}_{\mathbf{x}_i} = \mathrm{E}[\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}] = \Phi_{S_i}\mathbf{a}_i\mathbf{a}_i^{\mathrm{H}}, \quad \mathbf{\Phi}_{\mathbf{n}_i} = \mathrm{E}[\mathbf{n}_i\mathbf{n}_i^{\mathrm{H}}], \tag{4}$$

with $\Phi_{S_i} = \mathrm{E}[|S_i|^2] \in \mathbb{R}$ the PSD of the clean speech, and $\mathrm{E}[\cdot]$ the expectation operator. The conjugate transpose operator on complex vectors/matrices is indicated by the superscript $(\cdot)^{\mathrm{H}}$.

B. Linear Estimation Task

One way to increase speech intelligibility and quality of noisy signals is spatial filtering. The goal is to estimate the signal of interest at the FC by combining all the noisy observations into one single signal, such that a fidelity criterion is satisfied. In this paper, we consider linear estimation, i.e., S_i is estimated as $\hat{S}_i = \mathbf{w}_i^{\mathrm{H}}\mathbf{y}_i \in \mathbb{C}$, with $\mathbf{w}_i \in \mathbb{C}^M$ the weight vector. Minimizing the MSE, the best linear MSE estimator weights, say $\mathbf{w}_i^\star$, are given by the MWF [10]

$$\mathbf{w}_i^\star = \mathbf{\Phi}_{\mathbf{y}_i}^{-1}\mathbf{\Phi}_{\mathbf{y}_iS_i}, \quad i = 1,\dots,F, \tag{5}$$

where F is the number of frequency bins and $\mathbf{\Phi}_{\mathbf{y}_iS_i} \in \mathbb{C}^M$ is the CPSD vector between the observation vector $\mathbf{y}_i$ and the source S_i, which is given by $\mathrm{E}[\mathbf{y}_iS_i^*] = \mathbf{a}_i\mathrm{E}[|S_i|^2]$. The superscript $(\cdot)^*$ denotes the conjugate operator. Therefore, the optimal estimate, denoted by $\hat{S}_i^\star$, is given by $\hat{S}_i^\star = \mathbf{w}_i^{\star\mathrm{H}}\mathbf{y}_i$. Finally, the minimum MSE is computed as

$$D = \frac{1}{F}\sum_{i=1}^{F}\mathrm{E}[|S_i - \hat{S}_i^\star|^2] = \frac{1}{F}\sum_{i=1}^{F}\Phi_{d_i}, \tag{6}$$

with

$$\begin{aligned}\Phi_{d_i} =& \mathrm{E}[|S_i - \hat{S}_i^\star|^2]\\ =& \mathrm{E}[|S_i - \mathbf{w}_i^{\star\mathrm{H}}\mathbf{y}_i|^2]\\ =& \Phi_{S_i} - \mathbf{\Phi}_{\mathbf{y}_iS_i}^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{y}_i}^{-1}\mathbf{\Phi}_{\mathbf{y}_iS_i}, \quad i = 1,\dots,F.\end{aligned}$$

To compute the MWF output $\hat{S}_i^\star$, the noisy signal realizations should be available error-free at the FC. In practice, only a compressed/quantized version of the contralateral noisy signals are available. These signals are compressed at a certain rate, say r_{ij} bits per sample (bps). This leads to a modified signal model including quantization noise, as explained in the next subsection.

C. Quantization Aware Beamforming

As mentioned in the previous part of this section, the microphone signals are compressed prior to transmission to the FC. In this paper, we assume that the signals are being quantized using a uniform quantizer, which will be briefly explained in the following.

Let us consider an arbitrary signal x that is quantized, and the quantized version is denoted by $\tilde{x}$, with quantization noise $e = x - \tilde{x}$. Under high bit rate assumptions or by applying subtractive dithering to the signal to be quantized (at lower rates) [26], [27], the quantization error (noise) e will be uncorrelated to the signal x and will be uniformly distributed with variance $\sigma_e^2 = \frac{\Delta^2}{12}$. Here $\Delta = \frac{2x_{\max}}{2^r}$ is a step size, which depends on the range of the signal (maximum absolute value $x_{\max}$) and the quantization rate r. Applying this to the beamforming task, the quantization noise is taken into account and the signal model in (1) can be modified as

$$\tilde{Y}_{ij} = Y_{ij} + E_{ij} = A_{ij}S_i + N_{ij} + E_{ij} \in \mathbb{C}, \tag{7}$$

where $\tilde{Y}_{ij}$ is the quantized noisy signal and E_{ij} is the quantization noise. Similar to (2), using vector notation, we then have

$$\tilde{\mathbf{y}}_i = \mathbf{y}_i + \mathbf{e}_i = \mathbf{a}_iS_i + \mathbf{n}_i + \mathbf{e}_i \in \mathbb{C}^M, \tag{8}$$

where the quantization noise vector $\mathbf{e}_i = [E_{i1}, E_{i2}, \cdots, E_{iM}]^{\mathrm{T}}$ is assumed to be uncorrelated to the microphone signal vector $\mathbf{y}_i$, which is valid under the above-mentioned assumptions [26], [27]. Therefore, the CPSD matrix of the quantization noise vector $\mathbf{e}_i$ will be diagonal with elements

$$\Phi_{E_{ij}} = \frac{\Delta^2}{12} = \frac{(Y_{ij}^{\max})^2}{3\,2^{2\,r_{ij}}} = \frac{k_{ij}}{2^{2\,r_{ij}}}, \tag{9}$$

where $k_{ij} = \frac{(Y_{ij}^{\max})^2}{3}$. At the FC, the signal of interest S_i is estimated, given the compressed noisy microphone signals $\tilde{\mathbf{y}}_i$, as

$$\hat{S}_i = \mathbf{w}_i^{\mathrm{H}}\tilde{\mathbf{y}}_i. \tag{10}$$

The estimator $\hat{S}_i$ is a function of the estimation parameters $\mathbf{w}_i$ and the rates r_{ij}. In the next part of this section, we will propose a problem formulation to address the problem of finding the above-mentioned parameters, by minimizing the estimation error.

D. Rate-Distortion Trade-off in Noise Reduction Problems

As argued in the previous part of this section, at the FC, signals are available at a certain operating rate, say r_{ij} (bps). In fact, the receiver at the FC has a limited total capacity, say R_{tot}, due to limitations on transmission capabilities, to communicate with its agents [22] (here, microphones). Depending on this resource R_{tot} and the actual acoustic scene, different rate allocations across frequency and space are optimal [24]. In this paper, we address the problem of rate-constrained noise reduction in order to find the optimal rate allocation to each microphone signal at each specific frequency bin. We propose the following joint quantization-estimation problem.

1) Proposed Problem Formulation: We are given a set of operating rates $\mathcal{Q} = \{\mathbf{R} \,|\, 0 \le r_{ij} \le \infty\}$, where the matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1M} \\ r_{21} & r_{22} & \dots & r_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{F1} & r_{F2} & \dots & r_{FM} \end{bmatrix} \in \mathbb{R}^{F\times M}.$$

includes rates r_{ij} to be allocated to each frequency bin i and microphone j. Let the distortion function $D(\mathbf{R})$ be defined as

the averaged (over frequency) power spectral density of the estimation error, given the rates, that is

$$D(\mathbf{R}) = \frac{1}{F}\sum_{i=1}^{F} d(\mathbf{r}_i), \tag{11}$$

where

$$d(\mathbf{r}_i) = \mathrm{E}[|S_i - \hat{S}_i|^2|\mathbf{r}_i], \quad \mathbf{r}_i \in \mathbb{R}^M,$$

denotes the PSD of the estimation error at the ith discrete frequency bin, given the rate vector $\mathbf{r}_i = [r_{i1}, \ldots, r_{iM}]^{\mathrm{T}}$, which is the ith row of the matrix $\mathbf{R}$ and includes the rates allocated to the different microphones for the specific frequency i. Furthermore, let $R(\mathbf{R})$ simply be defined as the sum-rate over all bins and microphones, given by

$$R(\mathbf{R}) = \sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij}. \tag{12}$$

Then, the problem is defined as minimizing the estimation error, while satisfying the total budget R_{tot} on the rates. That is

$$\begin{aligned} \min_{\mathbf{R}\in\mathcal{Q}} \quad & D(\mathbf{R}) \\ \text{subject to} \quad & R(\mathbf{R}) \leq R_{\text{tot}}. \end{aligned} \tag{13}$$

Assuming that the joint statistics are known only at the FC, and using (8) and (10), the distortion function $d(\mathbf{r}_i)$ can be further parameterized as a function of the estimator weights $\mathbf{w}_i$ as

$$\begin{aligned} d(\mathbf{r}_i, \mathbf{w}_i) &= \mathrm{E}[|S_i - \hat{S}_i|^2|\mathbf{r}_i] \\ &= \mathrm{E}[|S_i - \mathbf{w}_i^{\mathrm{H}}\tilde{\mathbf{y}}_i|^2|\mathbf{r}_i] \\ &= \mathrm{E}[|S_i - \mathbf{w}_i^{\mathrm{H}}\mathbf{a}_i S_i - \mathbf{w}_i^{\mathrm{H}}\mathbf{n}_i - \mathbf{w}_i^{\mathrm{H}}\mathbf{e}_i|^2|\mathbf{r}_i] \\ &= |1 - \mathbf{w}_i^{\mathrm{H}}\mathbf{a}_i|^2\Phi_{S_i} + \mathbf{w}_i^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i + \mathbf{w}_i^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{e}_i}(\mathbf{r}_i)\mathbf{w}_i. \end{aligned} \tag{14}$$

The diagonal matrix $\mathbf{\Phi}_{\mathbf{e}_i}(\mathbf{r}_i)$ is the CPSD matrix of the quantization noise with elements given by (9). Based on (9) and the fact that $\mathbf{\Phi}_{\mathbf{e}_i}(\mathbf{r}_i)$ is diagonal, the distortion function $d(\mathbf{r}_i, \mathbf{w}_i)$ can be rewritten as

$$d(\mathbf{r}_i, \mathbf{w}_i) = |1 - \mathbf{w}_i^{\mathrm{H}}\mathbf{a}_i|^2\Phi_{S_i} + \mathbf{w}_i^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i + \sum_{j=1}^{M}\frac{|w_{ij}|^2\, k_{ij}}{2^{2\, r_{ij}}}. \tag{15}$$

We define the weight matrix $\mathbf{W} \in \mathbb{C}^{F\times M}$ as

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1^{\mathrm{T}} \\ \mathbf{w}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{w}_F^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \ldots & w_{1M} \\ w_{21} & w_{22} & \ldots & w_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ w_{F1} & w_{F2} & \ldots & w_{FM} \end{bmatrix} \in \mathbb{C}^{F\times M},$$

i.e., the ith row of $\mathbf{W}$ contains the beamformer coefficients for frequency bin i. Substituting (15) into (11), and then into the original problem formulation (13), the reformulated problem can be rewritten as

$$\begin{aligned} \min_{\mathbf{R},\mathbf{W}} \quad & \frac{1}{F}\sum_{i=1}^{F}\left(|1 - \mathbf{w}_i^{\mathrm{H}}\mathbf{a}_i|^2\Phi_{S_i} + \mathbf{w}_i^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i + \sum_{j=1}^{M}\frac{|w_{ij}|^2\, k_{ij}}{2^{2\, r_{ij}}}\right) \\ \text{s.t.} \quad & \sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij} \leq R_{\text{tot}}, \\ & r_{ij} \geq 0. \end{aligned} \tag{16}$$

Note that the estimation error function in (15) includes three terms: 1) the target signal distortion, i.e., $|1 - \mathbf{w}_i^{\mathrm{H}}\mathbf{a}_i|^2\Phi_{S_i}$ 2) the residual noise power, i.e., $\mathbf{w}_i^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i$ and 3) the residual quantization noise, i.e., $\sum_{j=1}^{M}\frac{|w_{ij}|^2\, k_{ij}}{2^{2\, r_{ij}}}$. The first two terms are only functions of the weights and the last term is jointly a function of both the weights and the quantization rates. In fact, as the last term in (15) is a summation of "quadratic-over-nonlinear" functions, which are non-convex functions, the problem in (16) is a non-convex optimization problem. However, fixing $\mathbf{W}$ or $\mathbf{R}$, the problem will be convex in the remaining variable (component-wise convex).

III. Proposed Solution

In the following, we propose a solution to the non-convex problem in (16), presented in the previous section. The third term in (15), which is a summation of "quadratic-over-nonlinear" functions, causes the non-convexity in the objective function. Nevertheless, we can write the necessary Karush-Kuhn-Tucker (KKT) conditions [28] for the problem in (16) to find the necessary optimality conditions. It can be shown (see Appendix A) that the solution to (16) lies on the boundary of the feasibility set defined by the global budget constraint (first constraint in (16)). As a consequence, we can replace the inequality constraint on the total bit budget by an equality constraint. With this, the Lagrangian function is given by

$$\begin{aligned} L(\mathbf{R}, \mathbf{W}, \lambda, \mathbf{V}) = \frac{1}{F}\sum_{i=1}^{F}[\,&|1 - \mathbf{w}_i^{\mathrm{H}}\mathbf{a}_i|^2\Phi_{S_i} + \mathbf{w}_i^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i \\ &+ \sum_{j=1}^{M}\frac{|w_{ij}|^2\, k_{ij}}{2^{2\, r_{ij}}}] + \lambda\Big(\sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij} - R_{\text{tot}}\Big) - \sum_{i=1}^{F}\sum_{j=1}^{M} v_{ij} r_{ij}, \end{aligned} \tag{17}$$

where the matrix $\mathbf{V} \in \mathbb{R}^{F\times M}$ consists of non-negative entries v_{ij} which denote the Lagrangian multipliers, responsible for the element-wise non-negativity constraints, i.e., $r_{ij} \geq 0$. The Lagrangian multiplier λ is to assure the total rate constraint is met with equality.

In the following proposition, the solution to the KKT conditions w.r.t. the problem in (16) and the Lagrangian equation (17) is given as a system of equations.

Proposition. *Minimizing the constrained problem in (16) based on the Lagrangian function in (17), the parametric optimal weights and the optimal rates are given as*

$$\begin{cases} 1) & \mathbf{w}_i^{\star}(\mathbf{r}_i^{\star}) = \mathbf{\Phi}_{\tilde{\mathbf{y}}_i}^{-1}\mathbf{\Phi}_{\tilde{\mathbf{y}}_i S_i}(\mathbf{r}_i^{\star}), \\ 2) & r_{ij}^{\star}(\lambda'^{\star}, w_{ij}^{\star}) = \max(\frac{1}{2}\log_2(\frac{|w_{ij}^{\star}|^2\, k_{ij}}{\lambda'^{\star}}), 0), \end{cases} \tag{18}$$

where $i = 1, \ldots, F, \quad j = 1, \ldots, M$, *and* $\lambda'^{\star} = \frac{\lambda^{\star}}{2\ ln2}$ *is a parameter, which satisfies the equality constraint*

$$\sum_{i=1}^{F} \sum_{j=1}^{M} r_{ij}(\lambda'^{\star}) = R_{\text{tot}}.$$

Proof. See Appendix A. □

Note that the rates are zero-valued for $\lambda'^{\star} \geq |w_{ij}^{\star}|^2\ k_{ij}$. The operator $\max(\cdot, 0)$ assures that the rates are non-negative, satisfying the second set of inequality constraints in (16).

Looking at the system of equations in (18), the optimal weights $\mathbf{w}_i^{\star}$ are the rate-dependent multi-channel Wiener filter coefficients (first set of equations) and the optimal rates $r_{ij}^{\star}$ are the solution to the weighted reverse water-filling problem. In fact, the set of Wiener equations are responsible for the target estimation part and the rate equation for the quantization part (rate allocation). It is clear from (18) that the rate allocation is done across both frequencies and microphones, depending on both the microphone signal power (which is related to k_{ij}) and the contribution of components to the estimation process (which is related to $|w_{ij}^{\star}|^2$). The frequencies and devices that contribute most to the target estimation will be allocated more bits. Similar to the classical water-filling problems [23], [29], the components for which $|w_{ij}^{\star}|^2\ k_{ij} \leq \lambda'^{\star}$ will be allocated zero bits.

One way to solve (18) is to apply alternating optimization [30]. First, the rates are initialized as $\mathbf{R}^0$, for example by an equal rate allocation where all components start to be allocated equal rates. Second, the optimal weight functions are computed, given $\mathbf{R}^0$, to find the updated weight matrix $\mathbf{W}^1$, where $\mathbf{W}^n$ denotes the updated matrix variable at nth iteration. Then the updated weights $\mathbf{W}^1$ are used to compute the updated rates $\mathbf{R}^1$. In this way, the equations are computed iteratively until a certain stopping criterion is met. As explained in Sec. II-D1, since the objective function in (16) is component-wise convex in the variables $\mathbf{W}$ and $\mathbf{R}$, as argued in [30], [31], any limit point (solution after sufficient iterations) is a critical point. Note that since the objective function is not jointly convex in $\mathbf{W}$ and $\mathbf{R}$, this critical point is not necessarily globally optimal. However, as confirmed by the simulation experiments in Sec. IV, the performance of the proposed method is almost as good as the (non-tractable) exhaustive search (for rate allocation across microphones) [24], for some representative example acoustic scenarios.

IV. Performance Evaluation

In this section, we perform simulations in several example acoustical scenarios to evaluate the performance of the proposed and existing approaches, as a function of the total communication rate R_{tot}.

In addition to predicted intelligibility by means of the short-time objective intelligibility (STOI) measure [32], we use the performance measure introduced in [17] and [16], which is defined as the ratio of the target signal estimation MSE, when there is no communication between the agents and the FC, say $D(0)$, to the MSE when the data is quantized before transmission, say $D(R)$. The output gain with respect to the beamformer (FC) is given by

$$G_{\text{FC}}(R) = \frac{D(0)}{D(R)}, \tag{19}$$

where $D(\cdot)$ is the MSE introduced in (11). $D(0)$ denotes the distortion when the devices do not communicate with the FC ($R_{\text{tot}} = 0$). In this case, the distortion is computed based on the local observations at the FC only.

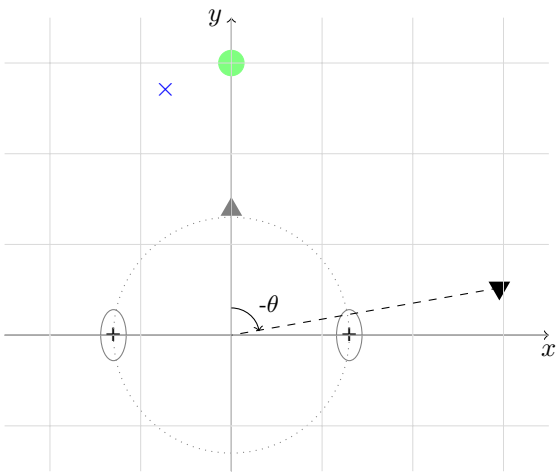


Fig. 1: Typical acoustic scene. The two HA microphones, the assistive microphone, the target signal, and the interferer are indicated by the black "+", the blue "×", the green circle, and the black triangle, respectively.

A. Example Generalized Binaural HA Setup

The first example acoustic scene is illustrated in Fig. 1. The binaural HA system includes two HA microphones (one per HA), denoted by the black "+" symbols, and are located with a distance of 10 cm w.r.t. the origin ($(x_{\text{o}}, y_{\text{o}}) = (0, 0)$), along the horizontal x-axis. The green circle indicates the target speech source, located in front of the HA system ($\theta = 0°$), at a distance of 3 m from the origin. In this paper, the location angles are computed counter-clockwise starting from the look direction. There is an assistive wireless microphone in this setup which is denoted by the blue "×" symbol, placed closer to the target speech at an angle $\theta = 15°$ and a distance of 2.8 m from the origin. The black triangle indicates the interfering signal, located at a distance of 3 m from the origin at an angle $\theta = -80°$, with a signal-to-interferer ratio (SIR) of 0 dB. In addition, simulated internal microphone noise is added to the microphone signals. The internal noise is assumed to be uncorrelated across microphones and is added with a signal-to-noise ratio (SNR) of 40 dB w.r.t. the target signal at the reference point.

In this experiment, without loss of generality, the FC is chosen to be the left side HA. Therefore, the left side microphone signal is considered as the reference local observation and the two other microphone signals as the agents' observations. The PSD of the target speech Φ_S is estimated based on Welch's method, using a 512-points discrete Fourier transform (DFT), computed frame-by-frame from 50% overlapping

speech frames, using around 10 s of the $F_s = 16$ kHz sampled speech signals taken from the "CMU-ARCTIC" database [33]. A flat PSD $\Phi_{n_1}(\omega)$ over the interval $\omega \in [-\pi, \pi]$ is assumed for the point noise source (interfering signal). Under the free-field assumption, the ATFs are generated using Habets' model [34], in a non-reverberant environment. The non-reverberant environment is chosen to get a more clear understanding of the effect of the number and location of the point noise sources on rate allocation behavior. Finally, the generated ATFs and the estimated PSDs are used to calculate the corresponding cross PSD matrices.

Based on the setup, the performance of the following approaches are compared throughout this section:

- **Equal Rate Allocation (2 Mics)**: Only the left-side and the right-side microphones (two microphones in total) are selected in this case (and thus not the assistive microphone). Therefore, there is only one microphone signal (from the right HA) which needs to be quantized before transmission. In this case, the rates are equally allocated over all frequencies.
- **Equal Rate Allocation (3 Mics)**: All three microphones are selected in this case. The rates are assumed to be equally allocated over all frequencies as well as across all microphones.
- **Discrete Optimization OPT [24]**: This method is based on discrete optimization, and optimally allocates the rates over all frequencies and across microphones. Note that, in this method, an exhaustive search is done to find the best allocations across the microphones, which is computationally very expensive and not tractable for big microphone arrays.
- **Proposed (2 Mics)**: The proposed method described in Sec. III. In this case, only the binaural setup (2-Microphone setup) is considered, meaning that the assistive microphone signal is not used. Therefore, the rate allocation is optimized only across frequency.
- **Proposed (3 Mics)**: The proposed method described in Sec. III. In this case, all microphones are used. Therefore, the rate allocation is optimized across both frequency and across microphones.
- **Remote Wyner Ziv (WZ) [17]**: The binaural rate-constrained beamforming presented in [17, Sec. III-A]. Note that only two processing nodes, i.e., in this setup two HAs, can be used in this method, joint statistics are needed at all processors (nodes) and impractical long-block vector quantizers are assumed.

1) Output Gains: In this part, we compare the above-mentioned approaches based on the performance measure in (19). Fig. 2 shows the output gain G_{FC} in dB as a function of the normalized (over frequency) total bit rate budget. The horizontal dash-dotted line denotes the performance of the 2-microphone MWF [11], [13], based on both the left and right microphone signals. It is assumed here that the right side observation is available (at an infinite rate) at the FC, i.e., without quantization noise. This method serves as a performance bound for the binaural setup. Similarly, the horizontal dashed line denotes the performance of the 3-microphone MWF [11], [13], where all microphone signals are used at an infinite rate. As shown, the performance of all methods approaches to the corresponding horizontal lines, at sufficiently high rates. The proposed method outperforms significantly the equal allocation strategies, as the rate allocation is optimized over frequency. The performance of the remote WZ method is computed based on the theoretical upper bound, described in [17]. As shown, the performance curve of the remote WZ method is upper-bounded by the 2-microphone MWF, as the assistive microphone is not considered in this method.

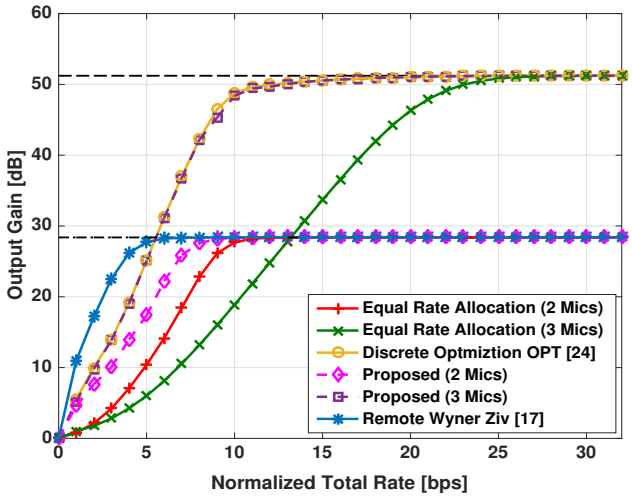
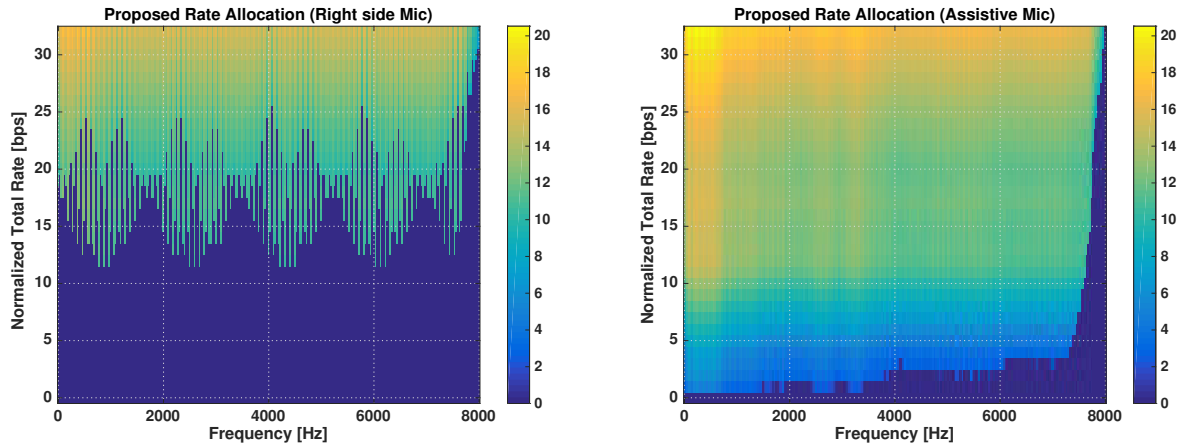


Fig. 2: Output Gain [dB] versus total rate [bit per sample] based on a generalized binaural setup in Fig. 1.

In this example setup, the proposed (3 Mics) method performs almost as good as the optimal discrete optimization method, which uses an exhaustive search to find the best allocations across microphones. Please note that, based on the complexity analysis which will be explained in Sec. IV-D, the computational complexity of the optimal discrete optimization method grows dramatically by increasing the number of the microphones. However, for the setup in Fig. 1 (with only three microphones) we could perform the exhaustive search for comparison. On average, the proposed alternating optimization approach needs less than 10 iterations to converge to a solution.

2) Rate Allocations Across Frequency: Based on the results, shown in Fig. 2, the rate distribution for each agent as a function of frequency and total bit rate is shown in Fig. 3. As shown in Fig. 3b, with a very small total rate, only lower frequency components are allocated non-zero rates. The effect of very high-frequency components on the final target estimation is negligible compared to the low-frequency components, as they have small PSD values, and therefore less rate is allocated. As the total rate increases, more high-frequency components can contribute to the estimation process.

Comparing Fig. 3a and Fig. 3b, for a small total rate, the right side microphone is barely used as the assistive microphone signal contains more information about the target signal (since it is located closer to the target source, based on Fig. 1). Therefore, more rate is allocated to the assistive microphone. As the total rate (total budget) increases, the right



(a) Right-side microphone contribution.

(b) Assistive microphone contribution.

Fig. 3: Rate distributions as a function of frequency and normalized total budget.

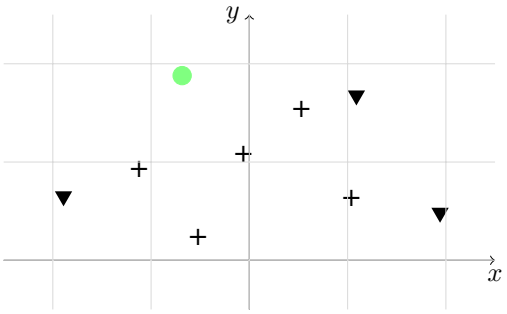


Fig. 4: An example acoustic scene: a general microphone array is shown by the black "+" symbols.

side microphone starts to contribute to the estimation process on its most important frequency components. The sinusoidal behavior of the rate distribution in Fig. 3a (at middle total rate values) is related to the shape of the squared value of the filtering weights ($|w_{ij}|^2$) over frequency.

B. Example General WASN Configuration

In this simulation experiment, we consider the second example acoustic scene, illustrated in Fig. 4. Five microphones are randomly located in space. The black triangles denote the interferers of which the number and location vary in different scenarios, which will be described later in this section. There is one target speech signal (Green circle) at $(2\,\mathrm{m}, 30^\circ)$. In this section, we consider the following three scenarios.

- **Scenario 1:** Only one interferer (point noise source).
- **Scenario 2:** Four interferers (point noise sources).
- **Scenario 3:** Four interferers along with diffuse noise.

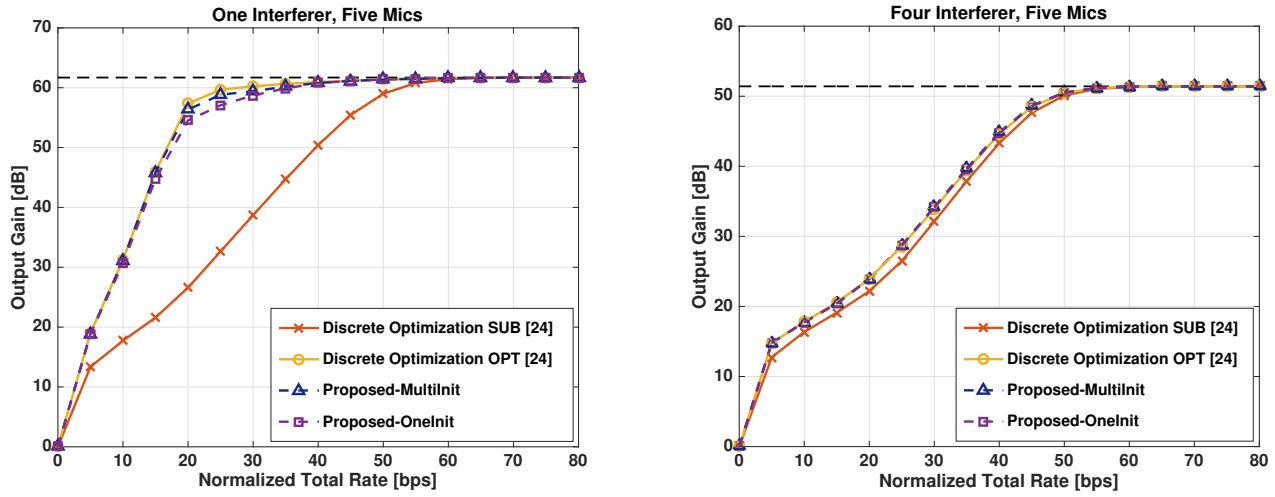
The FC is assumed to be located at the origin as a reference point (no local observations). For all scenarios, the interfering signals' power is chosen such that the SIR w.r.t. the target signal at the FC is 0 dB. In all experiments, uncorrelated internal noise is added to the microphone signals at 40 dB SNR w.r.t. the FC. For all sources, the ATFs and the power spectral densities are estimated/computed in a similar way as in the previous setup, in a non-reverberant environment.

Based on the setup, shown in Fig. 4, the following methods are compared:

- **Discrete Optimization SUB [24]**: This method is based on discrete optimization, and optimally allocates the rates over all frequencies. However, it assumes an equal rate allocation across microphones, as the optimal exhaustive search is very expensive and not tractable for big microphone arrays.
- **Discrete Optimization OPT [24]**: This method is based on discrete optimization, and optimally allocates the rates over all frequencies and across microphones. Based on our experiments and the complexity analysis, described in the Sec. IV-D, the exhaustive search used in this approach becomes intractable for more than five microphones.
- **Proposed**: The proposed method described in Sec. III.

1) Correlated Point Noise Sources: In this case, the scenarios 1 and 2 are considered. Scenario 1 contains only one interferer located at $(2\,\mathrm{m}, -60^\circ)$. Scenario 2 contains four interferers located at $(2\,\mathrm{m}, \{-80^\circ, -60^\circ, 40^\circ, 85^\circ\})$. Similar to Fig. 2, the output gains G_{FC} in dB as a function of total bit rate budget, are shown Fig. 5. Please note that at each normalized total bit budget, the budget will be distributed (maximally) across five microphones. For example, if the normalized total budget is 30 bps, it means that on average 30 bps may be allocated across five agents, and not necessarily six bps per agent. The dashed line denotes the performance of the 5-microphone MWF (which is an upper bound on the performance of the MSE-based methods), assuming all microphone signals are available at the FC, without quantization noise.

The proposed algorithm is based on alternating optimization which needs to be initialized. In the proposed-OneInit method, the algorithm is initialized based on reverse water filling on the power of the signals, assuming equal weights for all components. As we are not (theoretically) necessarily guaranteed to converge to the globally optimal solution, in the proposed-



(a) Scenario 1: One Interferer

(b) Scenario 2: Four Interferers

Fig. 5: Output Gain [dB] versus total rate [bit per sample] based on the second setup in Fig. 4.

MultiInit method, we also test the algorithm with multiple initializations. Initially, the total rate is randomly distributed to the components and the alternating optimization is carried out for each random initializations. The procedure is repeated and the allocation which results in a minimum distortion among all random initializations is selected. The proposed method with multiple initializations is very close, in performance, to the optimal discrete optimization approach. However, even with single initialization (proposed-OneInit) the performance of the proposed-OneInit method is not far from the optimal method. As shown in Fig. 5a, the proposed method performs significantly better than the sub-optimal discrete optimization method, as the optimal rates are also optimized across the agents. The remote Wyner Ziv approach is not included in the comparison, as it cannot consider more than two nodes, and therefore, it is not suitable for a general WASN setup.

In scenario 2, in Fig. 5b, instead of one point source, the scenario contains four interfering point sources. Increasing the number point sources has an interesting effect compared to the case of a single point source as in Fig. 5a. The performance gap between the sub-optimal approach, where the equal rate allocation is done across microphones, and the optimal methods is reduced. This can be explained as follows. Under mild differences in target signal powers captured by microphones, increasing the number of point sources, will reduce the spatial correlation (coherence) factor and makes the microphone signals more equally important in the target estimation process. Furthermore, in this case, all proposed and optimal curves are almost on top of each other, meaning that the proposed method managed to nearly achieve the optimal performance.

2) Diffuse Noise: In this scenario, there is a simulated diffuse noise along with four interferers. The diffuse noise is simulated as a cylindrical source array around the microphone array, for which the estimated spatial coherence function reasonably resembles the theoretical spatial coherence function between the microphone signals. Four interferers are located at $(2\,\mathrm{m}, \{-80°, -60°, 40°, 85°\})$. The powers of the sources are chosen such that the input signal to point noise and diffuse noise ratio (SIDR) is approximately 0 dB at the FC.

TABLE I: Computational complexity order

Method	Complexity
Discrete Optimization OPT [24]	$O(M^3\,F\,\lvert\mathcal{A}\rvert)$
Discrete Optimization SUB [24]	$O(M^3\,F\,q)$
Proposed	$O(M^3\,F\,K + MF\log(MF)\,K)$

$$\lvert\mathcal{A}\rvert = \tbinom{M-1}{M-1} + \tbinom{M}{M-1} + \ldots + \tbinom{q+M-2}{M-1}$$

Fig. 6 shows the output gains G_{FC} in dB as a function the total bit rate. The results show little difference between all competing methods, as almost the same (power-wise) impression of the environmental noise is received by each agent, and the observations become spatially less correlated. The sub-optimal discrete optimization, which is simple and fast, is therefore a suitable approach in this scenario. All proposed methods and the optimal method are almost on top of each other, and are asymptotically optimal meaning that the performance approaches that of the 5-microphone MWF method at a sufficiently high rate.

As mentioned in Sec. II-D1, the joint statistics need to be known only at the FC. Assuming that the statistics do not change rapidly over the number of consecutive frames, a piece of over-head information, which is needed to inform the agents about their allocated rates, can be averaged out over the frames, and hence, does not affect the proposed solution.

C. Computational Complexity

In this part, we compare the methods from a complexity point of view. The computational complexity of the competing methods in the previous part is listed in Table I, for a given total rate R_{tot}. Variable q denotes the number of all possible choices for the integer bit rate assigned to each frequency. Note

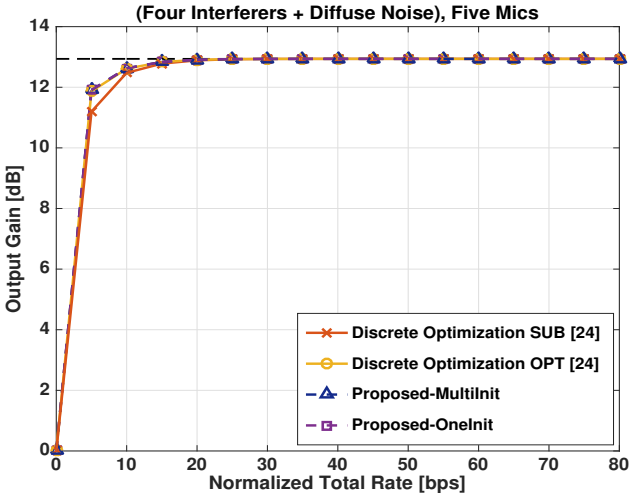


Fig. 6: Scenario 3: Diffuse noise + four interferers.

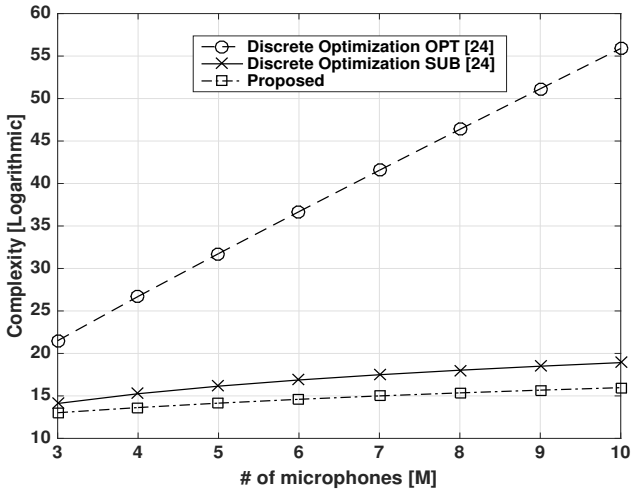


Fig. 7: Computational Complexity as a function of number of microphones [M].

that q generally may depend on the number of microphones M so that it may increase by increasing the number of microphones. The set $\mathcal{A}$ includes all possible allocations of the rate across microphones, for each frequency. When computing the cardinality $|\mathcal{A}|$, it is assumed that the rate (per frequency) can vary from zero bit to $(q-1)$ bits. In the optimal discrete optimization method (Discrete Optimization OPT), the exhaustive search is done over the set $\mathcal{A}$ to find the best bit allocation across microphones. In the sub-optimal discrete optimization method (Discrete Optimization SUB), the total bit rate (for each frequency) is distributed equally across microphones, therefore, the exhaustive search is not necessary. The computational complexity of the proposed method is based on (18) for K iterations. As shown, the proposed and sub-optimal methods have polynomial complexity order w.r.t. M and F. For the proposed method, for $\log(MF) \gg M^2$ the second term in the complexity order is dominant, therefore, the complexity will be of order $O(MF\log(MF))$ for one iteration ($K=1$). For a small M, the complexity is comparable to that of an FFT (complexity of order $F\log F$). In this case, the proposed method does not have a significant extra complexity, compared to FFT computations, which are unavoidable in frequency-domain noise reduction algorithms.

The complexity (in logarithmic scale) as a function of the number of microphones (M) is shown in Fig. 7, for $F = 512$, $q = 32M$, and $K = 15$ iterations over (18). As shown, the optimal method is computationally much more expensive than the other two methods. As shown in the simulations in the previous subsections, the proposed method is very close to the optimal method in terms of performance, although with much lower complexity.

In scenarios with highly correlated microphone signals (for example, scenario 1), there is a big performance gain in optimizing rate allocation across microphones (compared to the sub-optimal method). However, in scenarios with multiple sources and diffuse noise, the microphone signals become less correlated implying that the sub-optimal discrete optimization method becomes closer to the optimal discrete optimization method in terms of performance, with lower complexity.

D. Speech Intelligibility

In this section, we compare the competing methods in terms of speech intelligibility. Although all competing methods are based on optimizing the MSE criteria (and not based on speech intelligibility criteria) it is reasonable to see how they affect the speech intelligibility as a function of the bit rate.

In this paper, we choose the STOI measure [32] to evaluate the proposed method. Scenario 3 (as in the Sec. IV-B2) is chosen here based on the example acoustic scene shown in Fig. 4, which includes a simulated diffuse noise along with four interferers located at $(2\,\text{m}, \{-80°, -60°, 40°, 85°\})$. The SIDR w.r.t. the FC is set to 0 dB and the SNR is set to 40 dB. Uniformly distributed random realizations are added to the microphone signals as quantization noises. The variances of the quantization noises are computed using the corresponding optimized rate allocations for different methods.

The STOI measure as a function of the total rate is shown in Fig. 8. As shown, all curves approach (at high total rates) to the black dashed line which is the asymptotic STOI value when there is no quantization noise. Comparing Fig. 8 with Fig. 6, in this specific scenario, the STOI gaps between the sub-optimal discrete optimization method and the optimal methods are very low. In fact, under uniform quantization assumptions, small output gain differences between the competing methods at different total rates may not cause significant speech intelligibility gaps. As shown in Fig. 8, the proposed method performs as good as the optimal discrete optimization method in terms of the STOI objective measure, at much lower complexity.

V. Conclusion

In this paper, we proposed an MMSE-based rate-constrained noise reduction framework in wireless acoustic sensor networks (WASN) to jointly weight the contribution of the remote-microphone signals to the linear estimation task and allocate the bit rates across both frequency and spatial components (microphones). We introduced a joint estimation-compression optimization problem based on a rate-distortion

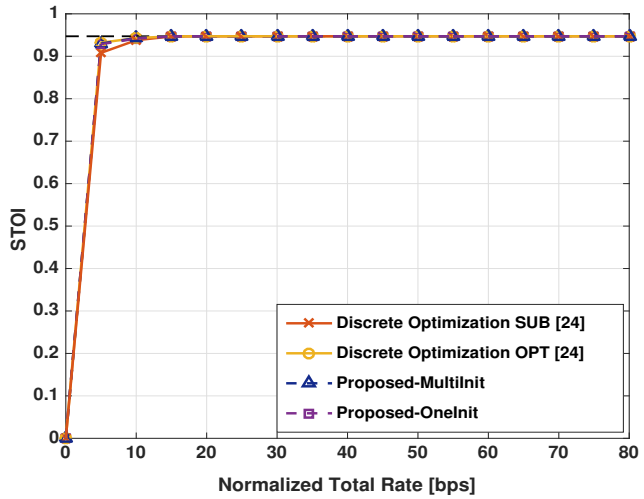


Fig. 8: STOI as a function of the total rate [bps] for Scenario 3: diffuse noise + four interferers.

trade-off to constrain the total rate at the fusion center. We proposed a solution to the component-wise convex estimation-compression problem based on alternating optimization. We found that the optimal estimation weights are actually the rate-constrained Wiener coefficients and the optimal rates are solutions to a filter-dependent reverse watering-filling problem. Based on the MSE criterion and the STOI intelligibility criterion, the performance of the proposed method is in most scenarios almost as good as the exhaustive search-based method, with lower complexity.

Appendix A
Derivations of the solution proposed in Sec. III (18)

In this section, we derive the necessary equations to solve the optimization problem, introduced in (16). Given the Lagrangian objective function in (17), the necessary KKT conditions for optimality are then given by

$$L_{\mathbf{w}_i^*} = \mathbf{\Phi}_{\mathbf{x}_i}\mathbf{w}_i - \Phi_{S_i}\mathbf{a}_i + \mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i + \mathbf{\Phi}_{\mathbf{e}_i}\mathbf{w}_i = 0, \tag{20a}$$

$$L_{p_{ij}} = \frac{-|w_{ij}|^2\, k_{ij}\, 2\ln 2}{2^{2r_{ij}}} + \lambda - v_{ij} = 0, \tag{20b}$$

$$\sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij} \le R_{\text{tot}}, \tag{20c}$$

$$\Big(\sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij} - R_{\text{tot}}\Big)\,\lambda = 0, \tag{20d}$$

$$\lambda \ge 0, \tag{20e}$$

$$r_{ij} \ge 0, \tag{20f}$$

$$r_{ij}v_{ij} = 0, \tag{20g}$$

$$v_{ij} \ge 0. \tag{20h}$$

We state that the optimal solution to this problem lies on the boundary of the budget constraint (20c). The proof of this statement is straightforward. Let us assume that an optimal solution, say $(\mathbf{W}^\star, \mathbf{R}^\star)$, is found such that $\mathbf{R}^\star$ lies strictly inside the feasibility set (and not on the boundary), with the corresponding objective distortion D^1. As the rates are constrained to be non-negative, one can increase the rates by a constant matrix, say $\mathbf{C}$, with non-negative entries to reach $\mathbf{R}^2 = \mathbf{R}^\star + \mathbf{C}$ such that the new solution, say $(\mathbf{W}^\star, \mathbf{R}^2)$ with a corresponding distortion D^2, still lies inside the set. As the distortion is a monotonically decreasing function over the rates, this implies $D^2 < D^1$. This shows that it is possible to increase rates until the full budget is used. Therefore, the third equation in the KKT conditions (20c) will be an equality constraint, and the fourth equation (complementary slackness over λ (20d)) and the fifth equation (20e) will be redundant.

We solve the KKT equations and find the optimal Lagrangian multiplier (λ) as a function of optimal weights. The first equation (20a) is actually the partial derivative with respect to the complex conjugate vector $\mathbf{w}_i^*$ [35], i.e.,

$$\begin{aligned} L_{\mathbf{w}_i^*} &= \mathbf{\Phi}_{\mathbf{x}_i}\mathbf{w}_i - \Phi_{S_i}\mathbf{a}_i + \mathbf{\Phi}_{\mathbf{n}_i}\mathbf{w}_i + \mathbf{\Phi}_{\mathbf{e}_i}\mathbf{w}_i \\ &= \left(\mathbf{\Phi}_{\mathbf{x}_i} + \mathbf{\Phi}_{\mathbf{n}_i} + \mathbf{\Phi}_{\mathbf{e}_i}\right)\mathbf{w}_i - \Phi_{S_i}\mathbf{a}_i \\ &= \mathbf{\Phi}_{\tilde{\mathbf{y}}_i}\,\mathbf{w}_i - \Phi_{S_i}\mathbf{a}_i \\ &= \mathbf{\Phi}_{\tilde{\mathbf{y}}_i}\,\mathbf{w}_i - \mathbf{\Phi}_{\tilde{\mathbf{y}}_i S_i} = 0, \end{aligned} \tag{21}$$

where the superscript $\{\cdot\}^*$ denotes the complex conjugate operator on matrices/vectors. The solution to (21) are, in fact, the multi-channel Wiener filter coefficients, given the optimal rate vector $\mathbf{r}_i^\star = [r_{i1}^\star, \ldots, r_{iM}^\star]^{\mathrm{T}}$, given by

$$\mathbf{w}_i^\star(\mathbf{r}_i^\star) = \mathbf{\Phi}_{\tilde{\mathbf{y}}_i}^{-1}\mathbf{\Phi}_{\tilde{\mathbf{y}}_i S_i}(\mathbf{r}_i^\star) \in \mathbb{C}^{M\times 1}, \qquad i = 1, \ldots, F. \tag{22}$$

To find the optimal rates, we solve (20b) for v_{ij} and substitute it into (20g) (complementary slackness), i.e.

$$r_{ij}\Big(\frac{-|w_{ij}|^2\, k_{ij}\, 2\ln 2}{2^{2r_{ij}}} + \lambda\Big) = 0, \qquad i = 1, \ldots, F. \tag{23}$$

Equality in (23) holds either by setting r_{ij} or $v_{ij} = \lambda - \frac{|w_{ij}|^2\, k_{ij}\, 2\ln 2}{2^{2r_{ij}}}$ to be zero. Considering the last three equations in (20) together with (23), the optimal rate value is zero, i.e., $r_{ij} = 0$ when $v_{ij} > 0$, which implies $\frac{\lambda}{2\ln 2} > |w_{ij}|^2\, k_{ij}$. Otherwise, the optimal r_{ij} will be strictly positive when $v_{ij} = 0$, which implies $\frac{\lambda}{2\ln 2} \le |w_{ij}|^2\, k_{ij}$, and we have

$$r_{ij}^\star(\lambda'^\star, w_{ij}^\star) = \begin{cases} \frac{1}{2}\log_2\Big(\frac{|w_{ij}^\star|^2\, k_{ij}}{\lambda'^\star}\Big) & \lambda'^\star \le |w_{ij}^\star|^2\, k_{ij}, \\ 0 & \lambda'^\star > |w_{ij}^\star|^2\, k_{ij}, \end{cases} \tag{24}$$

which simply can be rewritten as

$$r_{ij}^\star(\lambda'^\star, w_{ij}^\star) = \max\Big(\frac{1}{2}\log_2\Big(\frac{|w_{ij}^\star|^2\, k_{ij}}{\lambda'^\star}\Big), 0\Big), \tag{25}$$

where $i = 1, \ldots, F, \quad j = 1, \ldots, M$ with $\lambda'^\star = \frac{\lambda^\star}{2\ln 2}$ a rate reverse water filling parameter [23], [29]. In other words, the solution in (24) can be interpreted as if the equation (20b) is solved for r_{ij}, setting $v_{ij} = 0$, and the result is projected onto the non-negative orthant, i.e., $r_{ij} \ge 0$. Finally, to find an optimal $\lambda'^\star$ which satisfies the equality budget constraint (the equation (20c) with equality), i.e.,

$$\sum_{i=1}^{F}\sum_{j=1}^{M} r_{ij}^\star(\lambda'^\star, w_{ij}^\star) = R_{\text{tot}}, \qquad i = 1, \ldots, F. \tag{26}$$

we start by introducing a set $\mathcal{S}$ that contains the indices of components which are assumed to be allocated with positive rates

$$\mathcal{S} = \{(i,j)|\frac{|w_{ij}^{\star}|^2\ k_{ij}}{\lambda'^{\star}} > 0\}, \qquad i = 1,\ldots,F, \qquad (27)$$

where $i = 1,\ldots,F, j = 1,\ldots,M$. Given the set S, the budget constraint can be rewritten as

$$\sum_{(i,j)\in\mathcal{S}} (\frac{1}{2}\log_2(\frac{|w_{ij}^{\star}|^2\ k_{ij}}{\lambda'^{\star}})) = R_{\text{tot}}, \qquad i = 1,\ldots,F. \quad (28)$$

Taking the logarithm of both sides of (28) and solving for λ' we have

$$\lambda'^{\star} = \frac{(\prod_{(i,j)\in\mathcal{S}} |w_{ij}^{\star}|^2\ k_{ij})^{\frac{1}{|\mathcal{S}|}}}{2^{(\frac{2\ R_{\text{tot}}}{|\mathcal{S}|})}}, \qquad i = 1,\ldots,F. \qquad (29)$$

To find the set $\mathcal{S}$, we use the water-filling procedure [23] as follows.

Algorithm 1: Linear Water-filling for optimal λ'

1 Sort the coefficients $|w_{ij}^{\star}|^2\ k_{ij}$ in descending order into set $\mathcal{P}$.
2 **Initialize** an empty set $\mathcal{S} = \emptyset$, $\lambda'_{\text{opt}} = -\infty$:
3 **Pick** the first element in $\mathcal{P}$.
4 **If** λ'_{opt} is less than the picked value
5 Add the corresponding index into $\mathcal{S}$;
6 Compute (29) and update λ'_{opt};
7 **Else**
8 Stop and return $\mathcal{S}$ and λ'_{opt} (Optimal value is found).
9 **Repeat** 3-8 until all members of $\mathcal{P}$ are picked.

REFERENCES

[1] H. L. Van Trees, *Optimum Array Processing. Part IV of Detection, Estimation and Modulation Theory*, New York, NY: Wiley, 2008.

[2] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Berlin, Germany: Springer Science & Business Media, 2001.

[3] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP J. Adv. Signal Process*, vol. 2009, pp. 6:1–6:10, Jan. 2009.

[4] R. Sockalingam, M. Holmberg, K. Eneroth, and M. Shulte, "Binaural hearing aid communication shown to improve sound quality and localization," *The Hearing Journal*, vol. 62, no. 10, pp. 46–47, 2009.

[5] D. Marquardt, *Development and Evaluation of Psychoacoustically Motivated Binaural Noise Reduction and Cue Preservation Techniques*, PhD Dessertation, Universiy of Oldenburg, 2015.

[6] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 343–356, Feb 2013.

[7] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networkspart i: Sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, Oct 2010.

[8] R. Heusdens, G. Zhang, R. C. Hendriks, Y. Zeng, and W. B. Kleijn, "Distributed MVDR beamforming for (wireless) microphone networks using message passing," in *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, Sept 2012, pp. 1–4.

[9] L. W. Brooks and I. S. Reed, "Equivalence of the likelihood ratio processor, the maximum signal-to-noise ratio filter, and the Wiener filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-8, no. 5, pp. 690–692, Sept 1972.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[11] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sep 2002.

[12] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, *Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction*, pp. 199–228, Berlin, Heidelberg: Springer, 2005.

[13] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1579–1585, April 2007.

[14] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18–30, March 2015.

[15] T. C. Lawin-Ore, S. Stenzel, J. Freudenberger, and S. Doclo, "Generalized multichannel Wiener filter for spatially distributed microphones," in *Speech Communication; 11. ITG Symposium*, Sept 2014, pp. 1–4.

[16] S. Srinivasan and A. den Brinker, "Rate-constrained beamforming in binaural hearing aids," *EURASIP Journal on Advances in Signal Processing*, pp. 1–9, 2009.

[17] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 645–657, Feb 2009.

[18] S. Srinivasan and A. C. den Brinker, "Analyzing rate-constrained beamforming schemes in wireless binaural hearing aids," in *2009 17th European Signal Processing Conference*, Aug 2009, pp. 1854–1858.

[19] 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 Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 38–51, Jan 2009.

[20] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, pp. 1–10, Jan 1976.

[21] H. Yamamoto and K. Itoh, "Source coding theory for communication systems with a remote source," *Trans. IECE Jpn*, vol. E63, no. 6, pp. 700–706, Oct 1980.

[22] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem [multiterminal source coding]," *IEEE Transactions on Information Theory*, vol. 42, pp. 887902, MAy 1996.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2006.

[24] J. Amini, R. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Operational rate-constrained beamforming in binaural hearing aids," in *26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018.

[25] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 9, pp. 1445–1453, Sep 1988.

[26] A. Sripad and D. Snyder, "A necessary and sufficient condition for quantization errors to be uniform and white," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 5, pp. 442–448, Oct 1977.

[27] S. P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, "Quantization and dither: A theoretical survey," *Audio Eng. Soc.*, vol. 40, pp. 355–375, 1992.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[29] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data Compression*, Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[30] A. Beck, "On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes," *SIAM Journal on Optimization*, vol. 25, no. 1, pp. 185–209, 2015.

[31] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear Gauss-Seidel method under convex constraints," *Operations Research Letters*, vol. 26, no. 3, pp. 127 – 136, 2000.

[32] C. H. Taal, R. C. Hendriks, Heusdens R., and J. Jensen, "An algorithm for intelligibility prediction of timefrequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.

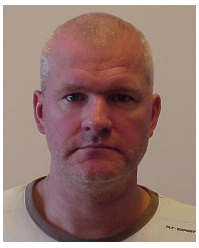
[33] J. Kominek, A. W. Black, and V. Ver, "CMU arctic databases for speech synthesis," Tech. Rep., 2003.

[34] E. A. P. Habets, "Room impulse response generator," https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator/, 2010.

[35] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proceedings F - Communications, Radar and Signal Processing*, vol. 130, no. 1, pp. 11–16, February 1983.

Jamal Amini received the B.Sc. degree in computer engineering from Shiraz University, Shiraz, Iran, in 2009, and the M.Sc. degree in electrical engineering from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2011. He is currently a Ph.D. student in the Circuits and Systems (CAS) Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology. His research interests are on speech enhancement, speech analysis and synthesis, source coding, and voice conversion.

Richard Christian Hendriks was born in Schiedam, The Netherlands. He received the B.Sc., M.Sc. (cum laude), and Ph.D. (cum laude) degrees in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 2001, 2003, and 2008, respectively. He is currently an Associate Professor in the Circuits and Systems (CAS) Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology. His main research interest is on biomedical signal processing, and, audio and speech processing, including speech enhancement, speech intelligibility improvement and intelligibility modelling. In March 2010, he received the prestigious VENI grant for his proposal Intelligibility Enhancement for Speech Communication Systems. He obtained several best paper awards, among which the IEEE Signal Processing Society best paper award in 2016. He is an Associate Editor for the IEEE/ACM Trans. on Audio, Speech, and Language Processing and the EURASIP Journal on Advances in Signal Processing, and serves as an elected member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing.



Richard Heusdens received the M.Sc. and Ph.D. degrees from Delft University of Technology, Delft, The Netherlands, in 1992 and 1997, respectively. Since 2002, he has been an Associate Professor in the Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology. In the spring of 1992, he joined the digital signal processing group at the Philips Research Laboratories, Eindhoven, The Netherlands. He has worked on various topics in the field of signal processing, such as image/video compression and VLSI architectures for image processing algorithms. In 1997, he joined the Circuits and Systems Group of Delft University of Technology, where he was a Postdoctoral Researcher. In 2000, he moved to the Information and Communication Theory (ICT) Group, where he became an Assistant Professor responsible for the audio/speech signal processing activities within the ICT group. He held visiting positions at KTH (Royal Institute of Technology, Sweden) in 2002 and 2008 and was a guest professor at Aalborg University from 2014-2016. He is involved in research projects that cover subjects such as audio and acoustic signal processing, speech enhancement, and distributed signal processing.

Meng Guo (S'10-M'13) received the M.Sc. degree in applied mathematics from Technical University of Denmark, Lyngby, Denmark, in 2006, and the Ph.D. degree in signal processing from Aalborg University, Aalborg, Denmark, in 2013.

From 2007 to 2010, he was with Oticon A/S, Smørum, Denmark, as a research and development engineer in the area of acoustic signal processing for hearing aids, especially in algorithm design of acoustic feedback cancellation. Currently, he is with Oticon A/S, Smrum, Denmark, as a research engineer. His main research interests are in the area of acoustic signal processing for hearing aid applications and wearable healthcare technology.

Jesper Jensen received the M.Sc. degree in electrical engineering and the Ph.D. degree in signal processing from Aalborg University, Aalborg, Denmark, in 1996 and 2000, respectively. From 1996 to 2000, he was with the Center for Person Kommunikation (CPK), Aalborg University, as a Ph.D. student and Assistant Research Professor. From 2000 to 2007, he was a Post-Doctoral Researcher and Assistant Professor with Delft University of Technology, Delft, The Netherlands, and an External Associate Professor with Aalborg University. Currently, he is a Senior Principal Scientist with Oticon A/S, Copenhagen, Denmark, where his main responsibility is scouting and development of new signal processing concepts for hearing aid applications. He is a Professor with the Section for Signal and Information Processing (SIP), Department of Electronic Systems, at Aalborg University. He is also a co-founder of the Centre for Acoustic Signal Processing Research (CASPR) at Aalborg University. His main interests are in the area of acoustic signal processing, including signal retrieval from noisy observations, coding, speech and audio modification and synthesis, intelligibility enhancement of speech signals, signal processing for hearing aid applications, and perceptual aspects of signal processing.