Spatially Correct Rate-Constrained Noise Reduction For Binaural Hearing Aids in Wireless Acoustic Sensor Networks

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

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

DOI (link to publication from Publisher):
10.1109/TASLP.2020.3028264

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Amini, J., Hendriks, R. C., Heusdens, R., Guo, M., & Jensen, J. (2020). Spatially Correct Rate-Constrained Noise Reduction For Binaural Hearing Aids in Wireless Acoustic Sensor Networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *28*, 2731-2742. Article 9210875. https://doi.org/10.1109/TASLP.2020.3028264

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2020.3028264, IEEE/ACM Transactions on Audio, Speech, and Language Processing

Spatially Correct Rate-Constrained Noise Reduction For Binaural Hearing Aids in Wireless Acoustic Sensor Networks

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

***Abstract*—Compared to monaural hearing aids (HAs), binaural hearing aid systems, in which there is a communication link between the two devices, have improved noise reduction capabilities and the ability to preserve binaural spatial information. However, the limited HA battery lifetime puts constraints on the amount of information that can be shared between the two devices. In other words, the rate of transmission between the devices is an important constraint that needs to be considered, while preserving the spatial information. In this paper, a linearly constrained noise reduction problem is proposed, which jointly finds the optimal rate allocation and the optimal estimation (beamforming) weights across all sensors and frequencies, while preserving the binaural spatial cues of point sources. The proposed method considers a rate constraint together with linear constraints to preserve the binaural spatial cues of point sources. Minimizing the mean square error on the estimated target speech at the left and the right side beamformers, the optimal weights are found to be rate-constrained linearly constrained minimum variance (LCMV) filters, and the optimal rates are found to be the solutions to a set of reverse water filling problems. The performance of the proposed method is evaluated using the averaged binaural signal-to-noise ratio (SNR), the interaural level difference (ILD) error and the interaural time difference (ITD) error. The results show that the proposed method outperforms spatially correct noise reduction approaches that use naive/random rate allocation strategies.**

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

I. Introduction

Multi-microphone noise reduction techniques, e.g., [1], [2], can be used to increase the speech quality and intelligibility of hearing aids (HAs). One way to use multi-microphone noise reduction techniques in modern HAs is to enable the left-ear and right-ear mounted HAs to collaborate through a wireless link, leading to a binaural HA setup. The binaural HA system provides increased spatial diversity and may result in better noise suppression, compared to the case where the monaural HAs perform noise reduction independently [3], [4]. In addition to better noise suppression, multi-microphone processing in the binaural HA setup can preserve binaural spatial information if taken care of, see e.g., [5]–[7]. These spatial information preserving noise reduction algorithms typically aim to preserve the interaural level differences (ILDs) and the interaural time differences (ITDs) of the relevant signal components. ILDs and ITDs are known to help humans determine the perceived location of the sound sources [6].

A common approach to achieve multi-microphone noise reduction is to combine the spatial observations captured by the microphones at a fusion center (FC) to estimate the sources of interest, while reducing the amount of environmental noise [2]. In the binaural HA setup, it is often considered that there are two FCs, one at each HA. Over the last decade, several binaural multi-microphone noise reduction algorithms have been proposed (see e.g., [6], [8] for overview). Typically they differ in the objective function they optimize and whether they can preserve the spatial cues of the target source, interferers, and the diffuse noise component. They can also differ in the types of constraints used to preserve the spatial cues. Equality constraints (see e.g., [5], [9]–[11]) are used to preserve exactly the spatial cues of the sources, while inequality constraints (see e.g., [12], [13]) are used to approximately preserve the spatial cues of the sources. The latter category can typically achieve a larger amount of noise suppression. In this paper, we will focus on equality-constrained binaural multi-channel noise reduction filters.

A well known binaural minimum mean square error (MMSE)-based noise reduction algorithm is the binaural multi-channel Wiener filter (MWF) [14], which aims at minimizing the MSE of the target signal estimated at the reference microphones of the two FCs without imposing any source preserving constraints. This may result in significant noise reduction, but a distorted target signal. In contrast to the binaural MWF, the binaural minimum variance distortionless response (BMVDR) beamformer [8] minimizes the output noise power under two linear distortionless constraints that preserve the target signal at the two reference microphones leading to preservation of the binaural cues of the target source. These two constraints, however, reduce the noise reduction performance of the BMVDR, compared to the binaural MWF. Another example is the binaural linearly constrained minimum variance (BLCMV) beamformer [5], [15], which can preserve the ILDs and ITDs of the source of interest and multiple interferers. As another example, the optimal BLCMV (OBLCMV) [9] can achieve better noise reduction, compared to the BLCMV, however, can only preserve the ILD and ITD of one interferer. An LCMV-based approach

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.

is proposed in [10], [11] which tries to increase the degree of freedom of the optimization problem by introducing a set of linear equality constraints (firstly introduced in [16]) to enable preserving more interferers, for a given number of microphones, compared to the BLCMV and the optimal BLCMV. Most of the binaural LCMV-based methods differ in how the set of linear constraints is designed.

In all the above-mentioned methods, the two FCs of the binaural beamformers each estimate the target source with respect to their corresponding reference microphone. To calculate these estimates, both FCs are in need of the microphone recordings from all sensors. This means that observations from the contralateral devices, and potentially any other device included in the setup, should be transmitted to the FCs. As the devices have a limited amount of resources (here transmission bandwidth) due to the limited battery lifetime, the total bit-rate used for transmission should be constrained. Several methods have been proposed in the literature to cope with this problem [17]–[20]. In [19] a binaural rate-constrained noise reduction approach is proposed which finds the optimal trade-off between the rate of transmission and the amount of noise reduction. The method finds the bound on the performance in case there are only two processing nodes. In the present context, these two processing nodes are the HAs. Scenarios with more than two nodes are not considered in [19]. Besides this, the inevitable requirement of the knowledge of the, generally time varying, joint statistics of all microphone signals at both HAs and using impractical infinitely long vector quantization limit the application of the method in practice. As alternatives to the optimal solution, several sub-optimal methods have been presented [21]–[23]. In [24], such algorithms were described in a unified framework. These sub-optimal methods try to pre-filter the observation before quantization without knowing the joint statistics, which enables the process to be faster and simpler. For example, this pre-filtering could be done to obtain a local estimate of the target or the interferer by combining the local microphone signals at the corresponding device. However, the pre-filtering stage combines the multi-microphone observations into a single observation, which may lead to a loss of some important information that needs to be known to retrieve the signals at high rates. As a result, even at an infinitely high rate of transmission, some important information may be lost and the performance will not approach that of the optimal algorithm presented in [19], not even asymptotically.

To address the aforementioned limitations, an operational rate-constrained noise reduction framework was proposed in [25], which estimates the optimal rate allocation across different frequencies and sensors using an operational rate-distortion trade-off [26]. Unlike [19], it allows considering scenarios with some additional assistive devices along with the binaural HA setup , thereby forming a small-size wireless acoustic sensor network (WASN) with more than two nodes. Furthermore, for the two-node case, the performance of the algorithm in [25] approaches that of the optimal algorithm in [19] at high rates without any mismatch, as the observations are not pre-filtered before quantization and necessary information will not be removed. However, the exhaustive search, which is used in [25] to find the optimal allocation across sensors, becomes intractable when the size of the WASN grows. Therefore, this method is suitable for small-size networks only. To address this scalability issue, another approach based on non-convex optimization was proposed in [27]. This method jointly finds the best rate allocation and the best estimation (beamforming) weights across all frequencies and sensors for arbitrary sized WASNs. Based on the MSE criterion, the optimal estimation weights are found to be rate-dependent Wiener filters and the optimal rates are the solution to a filter-dependent "water filling" problem. An alternating optimization approach which is used in this method avoids an exhaustive search to find the best allocations and performs almost as good as the exhaustive search-based approach, in most practical scenarios, at the benefit of a much lower computational complexity [27].

The above-mentioned methods deal with the rate-distortion trade-off in the noise reduction problem based on the MSE criterion. However, these methods do not take into account the preservation of spatial information (cues) when dealing with rate-constrained noise reduction problems. The noise reduction performance is optimal when minimizing the MSE, but the spatial information may be destroyed and the estimated signals may sound unnatural and spatially incorrect. Therefore, this raises the question of how to incorporate spatial information preservation into the rate-constrained noise reduction problem proposed in [27].

In this paper, inspired by [27], we propose and solve a multi fusion-center spatially correct rate-constrained noise reduction problem, to find the best rate allocation and the best estimation (beamforming) weights across all sensors and frequencies such that the spatial information of the sources is preserved. The method links the LCMV-based beamformers to data compression by including a set of linear constraints to the original rate-distortion problem. Unlike [27], here, there are two FCs, therefore, the objective function is to minimize the sum of the distortions of the target estimation at both hearing aids, while considering the total rate budget and simultaneously preserving the spatial information of the sources. Using an alternating optimization approach, the optimal estimation weights are found to be the rate-dependent LCMV filters, and the rates for both fusion centers are the solutions to two water-filling problems. The performance of the proposed method is evaluated using output signal-to-noise ratio (SNR) gain measures, and ILD and ITD error measures. Simulation results show that the proposed method outperforms the methods with equal/random rate allocation strategies.

II. Problem Statement

A. Signal Model

In this paper, a generalized binaural hearing aid system is considered, which consists of two collaborating hearing aids along with a number of additional assistive devices. We assume that these assistive devices can only communicate with the two HAs and not with each other. In total $M = M^{\mathrm{L}} + M^{\mathrm{R}} + M^{\mathrm{A}}$ microphones are assumed to be embedded in the HAs and the assistive devices, including M^{L} microphones for the left HA, M^{R} microphones for the right HA, and M^{A}

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2020.3028264, IEEE/ACM Transactions on Audio, Speech, and Language Processing

microphones for additional assistive devices. It is assumed here that no pre-filtering is applied to the unprocessed microphone signals to be transmitted to the FC, i.e., the microphone signals per device are not combined (pre-filtered) to a single signal.

Each microphone records a version of the target speech signal filtered by the position dependent room impulse response. The recorded target signal is degraded by a number of interfering point sources present in the room, diffuse noise and/or microphone self noise. The target signal, in the short-time Fourier transform (STFT) domain, is denoted by $S_k \in \mathbb{C}$, where k denotes the discrete frequency index. The interfering point sources are indicated by $I_{ki} \in \mathbb{C}$, where i denotes the point noise source index. All other sources of noise captured at a particular microphone are indicated by $U_{km} \in \mathbb{C}$, with m the microphone index. All sources are assumed to be additive and mutually uncorrelated.

Let the subscript $(\cdot)_m$ denote the microphone index. The signal model can then be written as

$$Y_{km} = A_{km} S_k + \overbrace{\sum_{i=1}^{b} B_{kmi} I_{ki} + U_{km}}^{N_{km}}, \tag{1}$$

where $A_{km} \in \mathbb{C}$ is the acoustic transfer function (ATF) between the target signal and the mth microphone, and $B_{kmi} \in \mathbb{C}$ is the acoustic transfer function (ATF) between the ith point noise source and the mth microphone. The number of interferers is denoted by b.

Stacking all microphone signals in a vector, the signal model can be rewritten in vector notation as

$$\mathbf{y}_k = \overbrace{\mathbf{a}_k S_k}^{\mathbf{x}_k} + \overbrace{\sum_{i=1}^{b} \mathbf{b}_{ki} I_{ki} + \mathbf{u}_k}^{\mathbf{n}_k} = \mathbf{x}_k + \mathbf{n}_k, \tag{2}$$

where

$$\begin{aligned} \mathbf{y}_k =& [(\mathbf{y}_k^{\mathrm{L}})^{\mathrm{T}}, (\mathbf{y}_k^{\mathrm{A}})^{\mathrm{T}}, (\mathbf{y}_k^{\mathrm{R}})^{\mathrm{T}}]^{\mathrm{T}}, \\ \mathbf{y}_k^{\mathrm{L}} =& [Y_{k1}, \dots, Y_{kM^{\mathrm{L}}}]^{\mathrm{T}}, \\ \mathbf{y}_k^{\mathrm{A}} =& [Y_{k(M^{\mathrm{L}}+1)}, \dots, Y_{k(M^{\mathrm{L}}+M^{\mathrm{A}})}]^{\mathrm{T}}, \\ \mathbf{y}_k^{\mathrm{R}} =& [Y_{k(M^{\mathrm{L}}+M^{\mathrm{A}}+1)}, \dots, Y_{kM}]^{\mathrm{T}}, \end{aligned}$$

and similarly for $\mathbf{a}_k$, $\mathbf{b}_{ki}$ and $\mathbf{n}_k$. Let $\mathbf{y}_k^{\mathrm{L}}$, $\mathbf{y}_k^{\mathrm{A}}$, and $\mathbf{y}_k^{\mathrm{R}}$ denote the microphone signal vectors captured by the left side HA microphones, assistive microphones, and the right side microphones, respectively. The superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose operator on vectors/matrices, and the power spectral density (PSD) matrix $\mathbf{\Phi}_{\mathbf{y}_k} = \mathrm{E}[\mathbf{y}_k \mathbf{y}_k^{\mathrm{H}}]$ of vector $\mathbf{y}_k$ is given by

$$\mathbf{\Phi}_{\mathbf{y}_k} = \mathbf{\Phi}_{\mathbf{x}_k} + \mathbf{\Phi}_{\mathbf{n}_k}, \tag{3}$$

where

$$\begin{aligned} \mathbf{\Phi}_{\mathbf{x}_k} =& \mathrm{E}[\mathbf{x}_k \mathbf{x}_k^{\mathrm{H}}] = \Phi_{S_k} \mathbf{a}_k \mathbf{a}_k^{\mathrm{H}}, \\ \mathbf{\Phi}_{\mathbf{n}_k} =& \sum_{i=1}^{b} \Phi_{I_{ki}} \mathbf{b}_{ki} \mathbf{b}_{ki}^{\mathrm{H}} + \mathrm{E}[\mathbf{u}_k \mathbf{u}_k^{\mathrm{H}}], \end{aligned} \tag{4}$$

and where $\Phi_{I_{ki}} = \mathrm{E}[|I_{ki}|^2] \in \mathbb{R}$ is the PSD of the ith interferer, $\Phi_{S_k} = \mathrm{E}[|S_k|^2] \in \mathbb{R}$ is the PSD of the clean target speech, and $\mathrm{E}[\cdot]$ denotes the expectation operator. The conjugate transpose operator on complex vectors/matrices is denoted by the superscript $(\cdot)^{\mathrm{H}}$.

B. Linearly Constrained Estimation

A binaural beamformer estimates the signal of interest at both left side and right side reference positions by combining all the available noisy observations into a single estimate for each HA. Notice that in this paper we do not only consider the presence of the two HAs, but also the presence of additional assistive microphones. The two resulting beamformer outputs are constructed such that a fidelity criterion is satisfied and the binaural information is preserved. The target signals at the left and right HA, i.e., S_k^{L} and S_k^{R}, respectively, are estimated as

$$\hat{S}_k^{\mathrm{L}} = (\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}} \mathbf{y}_k, \quad , \hat{S}_k^{\mathrm{R}} = (\mathbf{w}_k^{\mathrm{R}})^{\mathrm{H}} \mathbf{y}_k, \tag{5}$$

where $\mathbf{w}_k^{\mathrm{L}} \in \mathbb{C}^M$ and $\mathbf{w}_k^{\mathrm{R}} \in \mathbb{C}^M$ are the filter coefficients of the left and right beamformers, respectively. Minimizing the sum of the output noise powers, for both beamformers, the binaural linearly constrained beamforming problem can be formulated as [5]

$$\begin{aligned} \min_{\mathbf{w}_i} \quad & \mathbf{w}_k^{\mathrm{H}} \mathbf{\Phi}_k \mathbf{w}_k \\ \text{subject to} \quad & \mathbf{\Lambda}_k^{\mathrm{H}} \mathbf{w}_k = \mathbf{f}_k, \end{aligned} \tag{6}$$

where

$$\mathbf{w}_k = [\mathbf{w}_k^{\mathrm{L}^{\mathrm{T}}} \mathbf{w}_k^{\mathrm{R}^{\mathrm{T}}}]^{\mathrm{T}} \in \mathbb{C}^{2M \times 1},$$

$$\mathbf{\Phi}_k = \begin{bmatrix} \mathbf{\Phi}_{\mathbf{n}_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_{\mathbf{n}_k} \end{bmatrix} \in \mathbb{C}^{2M \times 2M},$$

and $\mathbf{\Lambda}_k \in \mathbb{C}^{2M \times d}$ is the constraint matrix, with d the number of linear constraints. Different binaural LCMV-based beamformers can be constructed by changing the entries of $\mathbf{\Lambda}_k$. In this paper, we use the methodology from [10], [11], having an increased amount of degrees of freedom compared to [9]. These additional degrees of freedom can then be used to cancel more interferers, given a fixed number of microphones. Following [10], [11] matrix $\mathbf{\Lambda}_k$ and vector $\mathbf{f}_k$ are given by

$$\begin{aligned} \mathbf{\Lambda}_k =& \begin{bmatrix} \mathbf{a}_k & \mathbf{0} & \mathbf{b}_1 B_{k1}^R & \dots & \mathbf{b}_b B_{kb}^R \\ \mathbf{0} & \mathbf{a}_k & -\mathbf{b}_1 B_{k1}^L & \dots & -\mathbf{b}_b B_{kb}^L \end{bmatrix} \in \mathbb{C}^{2M \times (b+2)}, \\ \mathbf{f}_k^H =& [A_k^L \; A_k^R \; 0 \; \dots \; 0] \in \mathbb{C}^{1 \times (b+2)}, \end{aligned} \tag{7}$$

respectively. Solving the problem in (6), the optimal weights are computed as [10]

$$\mathbf{w}_k^{\star} = \mathbf{\Phi}_k^{-1} \mathbf{\Lambda}_k (\mathbf{\Lambda}_k^H \mathbf{\Phi}_k^{-1} \mathbf{\Lambda}_k)^{-1} \mathbf{f}_k, \tag{8}$$

and the optimal beamformer outputs are given by

$$\hat{S}_k^{\mathrm{L}\star} = (\mathbf{w}_k^{\mathrm{L}\star})^{\mathrm{H}} \mathbf{y}_k, \quad , \hat{S}_i^{\mathrm{R}\star} = (\mathbf{w}_k^{\mathrm{R}\star})^{\mathrm{H}} \mathbf{y}_k. \tag{9}$$

In order to compute the binaural outputs $\hat{S}_k^{\mathrm{L}\star}$ and $\hat{S}_k^{\mathrm{R}\star}$, the actual signal realizations $\mathbf{y}_k$ should be available error-free at both HAs. However, due to limited battery power, and therefore, limited transmission power, in practice, the bit-rate, denoted by r_{km} bits per sample (bps), which is used to represent the transmitted signals must be constrained. Using a fixed bit-rate over frequencies and microphones can be shown to be sub-optimal, see e.g., [27]. Instead, the bit-rate dependent quantization noise should be included in the signal model, and optimized for.

C. Quantization Aware Estimation

In this sub-section, we introduce bit-rate dependent quantization noise in the signal model in (1). In this paper, we assume that the microphone signals from all nodes in the WASN are being quantized using a uniform quantizer before transmission to the corresponding FC (HA). Note that for each FC, the local observations at the FC are assumed to be quantized at the highest possible resolution, such that additional quantization noise on microphone signals at the FC can be neglected. In other words, only quantization noise with respect to the observations from other nodes in the WASN will be considered.

Consider an arbitrary signal denoted by x and its quantized version denoted by $\tilde{x}$, with quantization noise $q = x - \tilde{x}$. If subtractive dithering is applied to the signal to be quantized at lower rates or under high bit rate assumptions [28], [29], the quantization error q will be uniformly distributed and uncorrelated to signal x. In this case, the variance of the quantization noise is given by [28] $\sigma_q^2 = \frac{\Delta^2}{12}$, where $\Delta = \frac{2x_{\max}}{2^r}$ is the quantization step size, which depends on the range of the signal (maximum absolute value $x_{\max}$) and the quantization rate r.

Taking into account the quantization noise, the signal model for each side can be modified as

$$\begin{aligned}\tilde{Y}^{\mathrm{L}}_{km} &= Y_{km} + Q^{\mathrm{L}}_{km} = A_{km}S_k + \overbrace{\sum_{i=1}^{b} B_{kmi}I_{ki} + U_{km}}^{N_{km}} + Q^{\mathrm{L}}_{km},\\ \tilde{Y}^{\mathrm{R}}_{km} &= Y_{km} + Q^{\mathrm{R}}_{km} = A_{km}S_k + \overbrace{\sum_{i=1}^{b} B_{kmi}I_{ki} + U_{km}}^{N_{km}} + Q^{\mathrm{R}}_{km},\end{aligned} \tag{10}$$

where Q^{L}_{km} and Q^{R}_{km} denote the quantization noise w.r.t. the left and right side FCs, with $\tilde{Y}^{\mathrm{L}}_{km}$ and $\tilde{Y}^{\mathrm{R}}_{km}$ being the quantized microphone signals for the left and right side FCs, respectively. Using vector notation, we have

$$\begin{aligned}\tilde{\mathbf{y}}^{\mathrm{L}}_k &= \mathbf{y}_k + \mathbf{q}^{\mathrm{L}}_k = \mathbf{x}_k + \mathbf{n}_k + \mathbf{q}^{\mathrm{L}}_k,\\ \tilde{\mathbf{y}}^{\mathrm{R}}_k &= \mathbf{y}_k + \mathbf{q}^{\mathrm{R}}_k = \mathbf{x}_k + \mathbf{n}_k + \mathbf{q}^{\mathrm{R}}_k,\end{aligned} \tag{11}$$

where the quantization noise vector $\mathbf{q}^{\mathrm{L}}_k = [Q^{\mathrm{L}}_{k1}, Q^{\mathrm{L}}_{k2}, \cdots, Q^{\mathrm{L}}_{kM}]^{\mathrm{T}}$ is uncorrelated to the microphone signal vector $\mathbf{y}_k$, under the above-mentioned assumptions [28], [29], and similarly for $\mathbf{q}^{\mathrm{R}}_k$. Note that the bit-rates at which the left side signals are quantized are not necessarily the same as those at which the right side signals are quantized and transmitted to the left side FC. Under the above assumptions, and using $\Delta = \frac{2Y^{\mathrm{L},\max}_{km}}{2^{r^{\mathrm{L}}_{km}}}$, the CPSD matrix of the quantization noise vector $\mathbf{q}^{\mathrm{L}}_k$ will be diagonal with elements

$$\Phi_{Q^{\mathrm{L}}_{km}} = \frac{\Delta^2}{12} = \frac{(Y^{\mathrm{L},\max}_{km})^2}{3\,2^{2\,r^{\mathrm{L}}_{km}}} = \frac{k^{\mathrm{L}}_{km}}{2^{2\,r^{\mathrm{L}}_{km}}}, \tag{12}$$

where $k_{km} = \frac{(Y^{\mathrm{L},\max}_{km})^2}{3}$. Similar expressions can be derived for the right side beamformer.

Applying the above mentioned quantization approach to the beamforming task, versions of the signal of interest S^{L}_k and S^{R}_k are estimated, given the quantized noisy microphone signals $\tilde{\mathbf{y}}^{\mathrm{L}}_k$ and $\tilde{\mathbf{y}}^{\mathrm{R}}_k$, as

$$\hat{S}^{\mathrm{L}}_k = (\mathbf{w}^{\mathrm{L}}_k)^{\mathrm{H}}\tilde{\mathbf{y}}^{\mathrm{L}}_k, \quad , \hat{S}^{\mathrm{R}}_k = (\mathbf{w}^{\mathrm{R}}_k)^{\mathrm{H}}\tilde{\mathbf{y}}^{\mathrm{R}}_k. \tag{13}$$

The beamformer outputs $\hat{S}^{\mathrm{L}}_k$ and $\hat{S}^{\mathrm{R}}_k$ depend on $\mathbf{w}^{\mathrm{L}}_k$, $\mathbf{w}^{\mathrm{R}}_k$, and on the rates r^{L}_{km} and r^{R}_{km}, respectively.

III. Proposed Spatially Correct Rate-Constrained Noise Reduction

In this sub-section, we propose and solve an optimization problem to jointly optimize the rates and the estimation weights across the sensors and frequencies. The FCs at the left and right HA have a limited total channel capacity of $R^{\mathrm{L}}_{\mathrm{tot}}$ and $R^{\mathrm{R}}_{\mathrm{tot}}$ bps, respectively, to receive information from the other nodes in the network, as argued in [30]. In addition to the transmission rate, in this paper, we also take into account the preservation of spatial information, beneficial for binaural hearing aids. Altogether, in this paper, we address the problem of joint rate-constrained noise reduction and spatial cue preservation to find the optimal filter coefficients and rate allocation for all sensors and frequencies.

A. Problem Formulation

Let K indicate the number of frequency bins. Let the rate matrix $\mathbf{R}^{\mathrm{L}}$ be defined as

$$\mathbf{R}^{\mathrm{L}} = \begin{bmatrix} {\mathbf{r}^{\mathrm{L}}_1}^{\mathrm{T}} \\ {\mathbf{r}^{\mathrm{L}}_2}^{\mathrm{T}} \\ \vdots \\ {\mathbf{r}^{\mathrm{L}}_K}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} r^{\mathrm{L}}_{11} & r^{\mathrm{L}}_{12} & \cdots & r^{\mathrm{L}}_{1M} \\ r^{\mathrm{L}}_{21} & r^{\mathrm{L}}_{22} & \cdots & r^{\mathrm{L}}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ r^{\mathrm{L}}_{K1} & r^{\mathrm{L}}_{K2} & \cdots & r^{\mathrm{L}}_{KM} \end{bmatrix},$$

which includes rates r^{L}_{km} to be allocated to frequency bin k and microphone signal m, for the left side FC. Please note that, here, the kth row of the matrix $\mathbf{R}^{\mathrm{L}}$ is defined as ${\mathbf{r}^{\mathrm{L}}_k}^{\mathrm{T}} = [(\mathbf{r}^{\mathrm{LL}}_k)^{\mathrm{T}}, (\mathbf{r}^{\mathrm{LA}}_k)^{\mathrm{T}}, (\mathbf{r}^{\mathrm{LR}}_k)^{\mathrm{T}}]^{\mathrm{T}}$, where $(\mathbf{r}^{\mathrm{LA}}_k)^{\mathrm{T}}$ includes the rates at which the assistive microphones must be quantized and transmitted to the left side FC, and $(\mathbf{r}^{\mathrm{LR}}_k)^{\mathrm{T}}$ includes the rates at which the right-side HA microphone signals must be quantized and transmitted to the left side FC, at kth frequency. A similar definition holds for the right side rate matrix $\mathbf{R}^{\mathrm{R}}$.

The weight matrix $\mathbf{W}^{\mathrm{L}}$ is similarly defined as

$$\mathbf{W}^{\mathrm{L}} = \begin{bmatrix} {\mathbf{w}^{\mathrm{L}}_1}^{\mathrm{T}} \\ {\mathbf{w}^{\mathrm{L}}_2}^{\mathrm{T}} \\ \vdots \\ {\mathbf{w}^{\mathrm{L}}_K}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} w^{\mathrm{L}}_{11} & w^{\mathrm{L}}_{12} & \cdots & w^{\mathrm{L}}_{1M} \\ w^{\mathrm{L}}_{21} & w^{\mathrm{L}}_{22} & \cdots & w^{\mathrm{L}}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ w^{\mathrm{L}}_{K1} & w^{\mathrm{L}}_{K2} & \cdots & w^{\mathrm{L}}_{KM} \end{bmatrix},$$

which includes the left side beamformer coefficients w^{L}_{km}. A similar definition holds for the the right side beamformer coefficient matrix $\mathbf{W}^{\mathrm{R}}$.

Inspired by [27], we propose to formulate a spatially correct noise reduction problem, which tries to minimize a sum-distortion function given by

$$D(\mathbf{R}^{\mathrm{L}}, \mathbf{R}^{\mathrm{R}}, \mathbf{W}^{\mathrm{L}}, \mathbf{W}^{\mathrm{R}}) = D(\mathbf{R}^{\mathrm{L}}, \mathbf{W}^{\mathrm{L}}) + D(\mathbf{R}^{\mathrm{R}}, \mathbf{W}^{\mathrm{R}}), \tag{14}$$

where

$$D(\mathbf{R}^{\mathrm{L}},\mathbf{W}^{\mathrm{L}})=\frac{1}{K}\sum_{k=1}^{K}d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})=\frac{1}{K}\sum_{k=1}^{K}\mathrm{E}[|S_k^{\mathrm{L}}-\hat{S}_k^{\mathrm{L}}|^2|\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}}],$$

$$D(\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{R}})=\frac{1}{K}\sum_{k=1}^{K}d(\mathbf{r}_k^{\mathrm{R}},\mathbf{w}_k^{\mathrm{R}})=\frac{1}{K}\sum_{k=1}^{K}\mathrm{E}[|S_k^{\mathrm{R}}-\hat{S}_k^{\mathrm{R}}|^2|\mathbf{r}_k^{\mathrm{R}},\mathbf{w}_k^{\mathrm{R}}].$$

Here, $d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})$ denotes the PSD of the estimation error at the kth discrete frequency bin for the left side fusion center, and similarly for $d(\mathbf{r}_k^{\mathrm{R}},\mathbf{w}_k^{\mathrm{R}})$.

To address the rate-constrained noise reduction problem, we need constraint functions over the rates. Let $R(\mathbf{R}^{\mathrm{L}})$ simply be defined as the sum-rate over all frequency bins and microphones with respect to the left HA, given by

$$R(\mathbf{R}^{\mathrm{L}})=\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M}r_{km}^{\mathrm{L}}. \tag{15}$$

and similarly for $R(\mathbf{R}^{\mathrm{R}})$.

To address the spatially correct noise reduction problem, we use the set of linear equality constraints defined in the previous section as

$$\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k=\mathbf{f}_k,\quad k=1,\cdots,K, \tag{16}$$

where,

$$\mathbf{w}_k=[(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{T}},(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{T}}]^{\mathrm{T}}.$$

Then, the proposed problem is defined as minimizing the estimation error, while satisfying the above-mentioned constraints. That is

$$\begin{aligned}\min_{\mathbf{R}^{\mathrm{L}},\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{L}},\mathbf{W}^{\mathrm{R}}}\quad & D(\mathbf{R}^{\mathrm{L}},\mathbf{W}^{\mathrm{L}})+D(\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{R}})\\ \text{subject to}\quad & R(\mathbf{R}^{\mathrm{L}})\le R_{\mathrm{tot}}^{\mathrm{L}},\\ & R(\mathbf{R}^{\mathrm{R}})\le R_{\mathrm{tot}}^{\mathrm{R}},\\ & \mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k=\mathbf{f}_k,\quad k=1,\cdots,K.\end{aligned} \tag{17}$$

The distortion function $D(\mathbf{R}^{\mathrm{L}},\mathbf{W}^{\mathrm{L}})=\frac{1}{K}\sum_{k=1}^{K}d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})$ is parameterized as a function of the estimator weights and allocated rates with $d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})$ defined as

$$\begin{aligned}d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})&=\mathrm{E}[|S_k^{\mathrm{L}}-\hat{S}_k^{\mathrm{L}}|^2|\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}}]\\ &=\mathrm{E}[|S_k^{\mathrm{L}}-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\tilde{\mathbf{y}}_k^{\mathrm{L}}|^2]\\ &=\mathrm{E}[|S_k^{\mathrm{L}}-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{a}_kS_k-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{n}_k-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{q}_k^{\mathrm{L}}|^2]\\ &=|A_k^{\mathrm{L}}-(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{a}_k|^2\Phi_{S_k}+(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\overbrace{[\mathbf{\Phi}_{\mathbf{n}_k}+\mathbf{\Phi}_{\mathbf{q}_k^{\mathrm{L}}}(\mathbf{r}_k^{\mathrm{L}})]}^{\mathbf{\Phi}_k^{\mathrm{L}}(\mathbf{r}_k^{\mathrm{L}})}\mathbf{w}_k^{\mathrm{L}},\end{aligned} \tag{18}$$

and similarly for the right side distortion function $D(\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{R}})$. Assuming a distortion-less response in the target signal direction, i.e., using the constraint $(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{a}_k=A_k^{\mathrm{L}}$, which is included in the linear equality constraints in (16), (17), and the fact that $\mathbf{\Phi}_{\mathbf{q}_k^{\mathrm{L}}}(\mathbf{r}_k^{\mathrm{L}})$ is diagonal (see (12)), the distortion function $d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})$ can be rewritten as

$$d(\mathbf{r}_k^{\mathrm{L}},\mathbf{w}_k^{\mathrm{L}})=(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_k}\mathbf{w}_k^{\mathrm{L}}+\sum_{m=M^{\mathrm{L}}+1}^{M}\frac{|w_{km}^{\mathrm{L}}|^2\,k_{km}^{\mathrm{L}}}{2^{2\,r_{km}^{\mathrm{L}}}}. \tag{19}$$

A similar expression can be written for the right side beamformer. Stacking both the variables for the left and the right FCs into matrices, we have

$$\mathbf{w}_k=[(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{T}},(\mathbf{w}_k^{\mathrm{R}})^{\mathrm{T}}]^{\mathrm{T}}\in\mathbb{C}^{2M\times1},$$
$$\mathbf{\Phi}_k=\begin{bmatrix}\mathbf{\Phi}_k^{\mathrm{L}} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Phi}_k^{\mathrm{R}}\end{bmatrix}\in\mathbb{C}^{2M\times2M}.$$

It is natural to assume positive rates, $r_{km}\ge0$ (e.g. $r_{\min}=0$ and $r_{\max}=\infty$). Therefore, the reformulated problem can further be written as

$$\begin{aligned}\min_{\mathbf{R}^{\mathrm{L}},\mathbf{R}^{\mathrm{R}},\mathbf{W}}\quad & \frac{1}{K}\sum_{k=1}^{K}[\mathbf{w}_k^{\mathrm{H}}\mathbf{\Phi}_k(\mathbf{r}_k^{\mathrm{L}},\mathbf{r}_k^{\mathrm{R}})\mathbf{w}_k]\\ \text{s.t.}\quad & \sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M}r_{km}^{\mathrm{L}}\le R_{\mathrm{tot}}^{\mathrm{L}},\\ & \sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}}r_{km}^{\mathrm{R}}\le R_{\mathrm{tot}}^{\mathrm{R}},\\ & r_{km}^{\mathrm{L}}\ge0,\quad r_{km}^{\mathrm{R}}\ge0,\\ & \mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k=\mathbf{f}_k,\end{aligned} \tag{20}$$

where the objective function includes the distortion function in (19), and also, includes a similar distortion function for the right-side FC. The function in (19) includes two terms: 1) the residual noise power $(\mathbf{w}_k^{\mathrm{L}})^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{n}_k}\mathbf{w}_k^{\mathrm{L}}$, which is a quadratic (convex) function of the weights and 2) the residual quantization noise $\sum_{m=M^{\mathrm{L}}+1}^{M}\frac{|w_{km}^{\mathrm{L}}|^2\,k_{km}^{\mathrm{L}}}{2^{2\,r_{km}^{\mathrm{L}}}}$, which is a summation of "quadratic-over-nonlinear" functions, which are non-convex. Therefore the problem in (20) is a non-convex optimization problem. However, fixing either $\mathbf{W}$ or $\mathbf{R}$, the problem will be convex in the remaining variable.

B. Proposed Solution

Although the problem formulated in (20) is non-convex, we can still find the necessary optimality conditions by writing the Karush-Kuhn-Tucker (KKT) conditions [31]. Considering the first and second inequality rate constraint functions in (20), it can be shown that the rate solutions actually lie on the boundary of the feasibility sets defined by the global rate budget constraints which are the first and the second constraints in (20) [27].

We solve the KKT conditions and the solution will be given in the following proposition.

Proposition. *The solution to the problem in (20) is given by*

$$\begin{cases}1)\mathbf{w}_k^{\star}(\mathbf{r}_k^{\mathrm{L}\star},\mathbf{r}_k^{\mathrm{R}\star})=\mathbf{\Phi}_k^{-1}\mathbf{\Lambda}_k(\mathbf{\Lambda}_k^{H}\mathbf{\Phi}_k^{-1}\mathbf{\Lambda}_k)^{-1}\mathbf{f}_k,\\ 2)\quad r_{km}^{\mathrm{L}\star}(\lambda_{\mathrm{L}}'^{\star},w_{km}^{\mathrm{L}\star})=[\frac{1}{2}\log_2(\frac{|w_{km}^{\mathrm{L}\star}|^2\;k_{km}^{\mathrm{L}}}{\lambda_{\mathrm{L}}'^{\star}})]^{+},\\ 3)\quad r_{km}^{\mathrm{R}\star}(\lambda_{\mathrm{R}}'^{\star},w_{km}^{\mathrm{R}\star})=[\frac{1}{2}\log_2(\frac{|w_{km}^{\mathrm{R}\star}|^2\;k_{km}^{\mathrm{R}}}{\lambda_{\mathrm{R}}'^{\star}})]^{+},\end{cases} \tag{21}$$

where $\lambda_L'^{\star}=\frac{K\lambda_L^{\star}}{2\;ln2}$ *and* $\lambda_R'^{\star}=\frac{K\lambda_R^{\star}}{2\;ln2}$ *are parameters, which satisfy the following equality constraints, respectively*

$$\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M}r_{km}^{\mathrm{L}}(\lambda_{\mathrm{L}}'^{\star})=R_{\mathrm{tot}}^{\mathrm{L}},$$

$$\sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}} r^{\mathrm{R}}_{km}(\lambda'^{\star}_{\mathrm{R}}) = R^{\mathrm{R}}_{\mathrm{tot}}.$$

Proof. See Appendix A. □

The rates are non-zero valued for $\lambda'^{\star}_{\mathrm{L}} \leq |w^{\mathrm{L}\star}_{km}|^2\ k^{\mathrm{L}}_{km}$ and $\lambda'^{\star}_{\mathrm{R}} \leq |w^{\mathrm{R}\star}_{km}|^2\ k^{\mathrm{R}}_{km}$ and are zero-valued otherwise. The non-linear operator $[\cdot]^{+}$ projects all negative valued rates to zero and the positive valued rates will remain unchanged, satisfying the set of inequality constraints in (20) ($r^{\mathrm{L}}_{km} \geq 0, \quad r^{\mathrm{R}}_{km} \geq 0,$).

As shown in the proposition, the optimal weights $\mathbf{w}^{\star}_{k}$ are the rate-constrained BLCMV coefficients, which, as a special case of the BLCMV coefficients, can be expressed as the BMVDR solutions. Note that, in general, $\mathbf{\Phi}_k^{-1}$ is a function of the bit-rates $\mathbf{r}^{\mathrm{L}\star}_{k}$ and $\mathbf{r}^{\mathrm{R}\star}_{k}$. The optimal rates r^{L}_{km} and r^{R}_{km} are the solution to the weighted reverse water filling problem. In other words, looking at the system of equations in (21), it turns out that to allocate the rates, we need to follow the reverse water filling approach while using the BLCMV filter coefficients. As explained, the BLCMV filters, when there is no quantization, can guarantee the preservation of the spatial cues of the target signal. Also here in (21), it is possible to preserve the spatial cues of the target signal, even when imperfect data, which is quantized at finite rate, is received by the corresponding beamformer and used to compute $\mathbf{\Phi}_k^{-1}$. Unlike the original water filling problem, where the rate allocation depends only on the microphone signal power, here, the rate allocation not only depends on the microphone signal power but also on the importance of the corresponding frequency component of the microphone signal to the estimation process. That is, the frequency bins which are more important in the target estimation stage, i.e., more informative, will be allocated more bits.

To solve the system of equations in (21), a similar approach as in [27] is used. The approach is based on alternating optimization, where the system is initialized with, for example, equal rate allocation across all components for both the left and right FCs, say $\mathbf{R}^{\mathrm{L}}_{0}$ and $\mathbf{R}^{\mathrm{R}}_{0}$, respectively. Then the weight equation is computed based on the equal rates and the weight matrix $\mathbf{W}_1$ is updated. Then, the rates will be updated based on the computed weights to $\mathbf{R}^{\mathrm{L}}_{1}$ and $\mathbf{R}^{\mathrm{R}}_{1}$. This process will be repeated until a certain stopping criterion is met. As the problem in (20) is component-wise convex, it is shown in [32] that any limit point, which is the solution after sufficient iterations. is a critical point. This means that the obtained critical point is not necessarily globally optimal. However, as shown in [27], based on MSE and STOI measures, for certain types of noise reduction methods, the performance is almost as good as the method which uses an exhaustive search, but at the benefit of much lower computational complexity.

1) Special Cases of the Proposed Solution: In Table I, we highlight several special cases of the proposed solution in (21). As shown, (A) if the rate budgets go to infinity, then the solution will be equal to the joint BLCMV (JBLCM) filters [10], [11], using (7). (B) If the rate budgets go to infinity, and the matrix $\mathbf{\Lambda}_k$ is given by

$$\mathbf{\Lambda}_k = \begin{bmatrix} \mathbf{a}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{a}_k \end{bmatrix} \in \mathbb{C}^{2M\times 2}, \qquad (22)$$
$$\mathbf{f}^{H}_{k} = [A^{\mathrm{L}}_{k}\ A^{\mathrm{R}}_{k}] \in \mathbb{C}^{1\times 2}.$$

then the solution will become equal to the BMVDR filters [8]. (C) If the rate budgets are finite numbers, and the above-mentioned $\mathbf{\Lambda}_k$ in (22) is used, then the weight solution will be the rate-constrained BMVDR filters, which we refer to as "Proposed alternating optimization (AO)-BMVDR" in the next section. (D) Finally, when the rate budgets are finite, solving the equations in (21) and using (7) will lead to the proposed method, which we refer to as "ProposedAO-JBLCMV".

TABLE I: Special cases of the proposed solution in (21).

Method	Total Rate	Constraint Matrix $\mathbf{\Lambda}$
(A): JBLCMV [10], [11]	$\mathbf{R}^{\mathrm{L}}_{\mathrm{tot}} \to \infty$ $\mathbf{R}^{\mathrm{R}}_{\mathrm{tot}} \to \infty$	$\mathbf{\Lambda}_k$ as in (7)
(B): BMVDR [8]	$\mathbf{R}^{\mathrm{L}}_{\mathrm{tot}} \to \infty$ $\mathbf{R}^{\mathrm{R}}_{\mathrm{tot}} \to \infty$	$\mathbf{\Lambda}_k$ as in (22)
(C): ProposedAO-BMVDR	$\mathbf{R}^{\mathrm{L}}_{\mathrm{tot}}$ is finite $\mathbf{R}^{\mathrm{R}}_{\mathrm{tot}}$ is finite	$\mathbf{\Lambda}_k$ as in (22)
(D): ProposedAO-JBLCMV	$\mathbf{R}^{\mathrm{L}}_{\mathrm{tot}}$ is finite $\mathbf{R}^{\mathrm{R}}_{\mathrm{tot}}$ is finite	$\mathbf{\Lambda}_k$ as in (7)

IV. Performance Evaluation

In this section, we evaluate the proposed method as a function of the total bit rate budget by carrying out simulations in different acoustic scenarios. The proposed method will be compared to some existing methods using the binaural output SNR, and the ILD and ITD error measures, which will be defined in the next part of this section. In the evaluation, we will consider two different acoustic scenarios discussed in Sections IV-B and IV-C, respectively.

A. Performance Measures

We use the definitions presented in [6], [9], [10] for binaural input and output SNRs and ITD and ILD errors.

1) Binaural SNRs: The binaural input SNR and the binaural output SNR are defined as [9]

$$\begin{aligned}\mathrm{SNR}_{\mathrm{in}}(k) &= 10\log_{10}\Big(\frac{\mathbf{e}^{\mathrm{T}}_{\mathrm{L}}\mathbf{\Phi}_{\mathbf{x_k}}\mathbf{e}_{\mathrm{L}} + \mathbf{e}^{\mathrm{T}}_{\mathrm{R}}\mathbf{\Phi}_{\mathbf{x_k}}\mathbf{e}_{\mathrm{R}}}{\mathbf{e}^{\mathrm{T}}_{\mathrm{L}}\mathbf{\Phi}^{\mathrm{L}}_{k}\mathbf{e}_{\mathrm{L}} + \mathbf{e}^{\mathrm{T}}_{\mathrm{R}}\mathbf{\Phi}^{\mathrm{R}}_{k}\mathbf{e}_{\mathrm{R}}}\Big),\\ \mathrm{SNR}_{\mathrm{out}}(k) &= 10\log_{10}\Big(\frac{(\mathbf{w}^{\mathrm{L}}_{k})^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{x_k}}\mathbf{w}^{\mathrm{L}}_{k} + (\mathbf{w}^{\mathrm{R}}_{k})^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{x_k}}\mathbf{w}^{\mathrm{R}}_{k}}{(\mathbf{w}^{\mathrm{L}}_{k})^{\mathrm{H}}\mathbf{\Phi}^{\mathrm{L}}_{k}\mathbf{w}^{\mathrm{L}}_{k} + (\mathbf{w}^{\mathrm{R}}_{k})^{\mathrm{H}}\mathbf{\Phi}^{\mathrm{R}}_{k}\mathbf{w}^{\mathrm{R}}_{k}}\Big),\end{aligned} \qquad (23)$$

where k denotes the frequency index, and

$$\begin{aligned}\mathbf{e}^{\mathrm{T}}_{\mathrm{L}} &= [1, 0, \ldots, 0] \in \mathbb{R}^{M},\\ \mathbf{e}^{\mathrm{T}}_{\mathrm{R}} &= [\underbrace{0, \ldots, 0}_{M^{\mathrm{L}}+M^{\mathrm{A}}}, 1, 0 \ldots, 0] \in \mathbb{R}^{M}.\end{aligned}$$

The performance measure we use is defined as the binaural SNR gain, $\mathrm{SNR}_{\mathrm{gain}}(k)$, and is given by

$$\mathrm{SNR}_{\mathrm{gain}}(k) = \mathrm{SNR}_{\mathrm{out}}(k) - \mathrm{SNR}_{\mathrm{in}}(k). \qquad (24)$$

2) ILD and ITD Errors: To define the ILD and ITD errors, we first define the input and output interaural transfer functions (ITFs) w.r.t. the source of interest as [6], [10]

$$\begin{aligned}\mathrm{ITF}_X^{\mathrm{in}}(k) =& \frac{X_k^{\mathrm{L}}}{X_k^{\mathrm{R}}} = \frac{A_k^{\mathrm{L}}}{A_k^{\mathrm{R}}},\\ \mathrm{ITF}_X^{\mathrm{out}}(k) =& \frac{{\mathbf{w}_k^{\mathrm{L}}}^{\mathrm{H}}\mathbf{x}_k}{{\mathbf{w}_k^{\mathrm{R}}}^{\mathrm{H}}\mathbf{x}_k} = \frac{{\mathbf{w}_k^{\mathrm{L}}}^{\mathrm{H}}\mathbf{a}_k}{{\mathbf{w}_k^{\mathrm{R}}}^{\mathrm{H}}\mathbf{a}_k}.\end{aligned} \tag{25}$$

Note that to find the ITFs for the interferers, the signal X_k and the transfer function A_k should be replaced by I_{ki} and B_{ki}, respectively, in (25). With this, the input and output ILDs are defined as the squared magnitudes of the input and output ITFs. That is

$$\mathrm{ILD}_X^{\mathrm{in}}(k) = |\mathrm{ITF}_X^{\mathrm{in}}(k)|^2, \qquad \mathrm{ILD}_X^{\mathrm{out}}(k) = |\mathrm{ITF}_X^{\mathrm{out}}(k)|^2, \tag{26}$$

and the input and output ITDs defined as the phase of the input and output ITFs. That is

$$\mathrm{ITD}_X^{\mathrm{in}}(k) = \angle\mathrm{ITF}_X^{\mathrm{in}}(k), \qquad \mathrm{ITD}_X^{\mathrm{out}}(k) = \angle\mathrm{ITF}_X^{\mathrm{out}}(k). \tag{27}$$

The ILD and ITD errors are then defined as

$$\begin{aligned}\mathrm{ER}_{\mathrm{ILD}_X^{\mathrm{out}}}(k) &= |\mathrm{ILD}_X^{\mathrm{out}}(k) - \mathrm{ILD}_X^{\mathrm{in}}(k)|,\\ \mathrm{ER}_{\mathrm{ITD}_X^{\mathrm{out}}}(k) &= \frac{|\mathrm{ITD}_X^{\mathrm{out}}(k) - \mathrm{ITD}_X^{\mathrm{in}}(k)|}{\pi}.\end{aligned} \tag{28}$$

Note that $0 \leq \mathrm{ER}_{\mathrm{ITD}_X^{\mathrm{out}}}(k) \leq 1$. Please note that, in this paper, all defined measures will be rate-constrained, meaning that the measures are computed for a given total bit budgets $R_{\mathrm{tot}}^{\mathrm{L}}$ and $R_{\mathrm{tot}}^{\mathrm{R}}$, which will become more clear in the simulation results.

B. Example Binaural HA Setup using Head-Related Transfer Functions

1) Acoustic Scene 1: The first acoustic scene is based on the setup described in [33] and depicted in Fig. 1. The green circle in Fig. 1 denotes the target speech source, which is positioned at 3 m distance from the origin ((0,0)), in front of the binaural HA system. The binaural HA system consists of two HAs with two microphones per HA, with thus $M = 4$ microphones in total, mounted on a virtual head and denoted by the red "+" symbol. The zero degree corresponds to the looking direction of the virtual head and the angles are computed counterclockwise. The planar distance between the two microphones per HA is 0.76 cm and the radius of the typical head is 8.2 cm [33]. Interferers are indicated by the black triangles, assumed to be located at different positions in space, with a spatial resolution of 5°. The number and location of the interferers may vary in different experiments. Uncorrelated flat PSD noise is also added to the microphone signals at an SNR of 40 dB with respect to the corresponding reference microphones to simulate internal microphone noise.

The left and right side HAs are considered as two FCs. For example, for the left side FC, the observations recorded at its microphones are thought as the local observations and the contralateral right side microphone signals are quantized and transmitted to the left side FC. A similar explanation holds for the right side FC. Welch's method is used to estimate the PSD of the target speech, using 512-discrete Fourier transform (DFT) points, which is computed frame-by-frame using 50% overlapping speech frames. Around 12s of recorded sampled speech (at $F_s = 16$ KHz) from the "CMU-ARCTIC" database [34] is used for the PSD estimation process. The head-related transfer functions (HRTFs) from the database in [33], with a spatial resolution of 5°, are used in this experiment. For the point noise sources, flat PSDs $\Phi_{I_k}(\omega)$ over the interval $\omega \in [-\pi, \pi]$ are considered. The cross-PSD matrices with respect to the target signal and the noises are computed using the estimated/computed PSDs and the HRTFs.

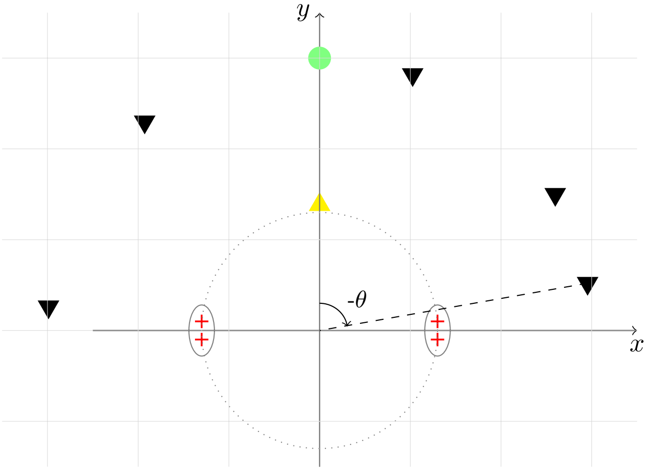


Fig. 1: Example acoustic scene. The target signal, the interferers, and the four HA microphones (two microphones per HA) are denoted by the green circle, the black triangles , and the red "+", respectively.

2) Competing Methods: The following methods are chosen as reference methods: a) **EQ-BMVDR:** the rate-constrained BMVDR. In this approach, we assume equal rate allocation across all sensors and frequencies, i.e., no optimization is done here. Note that when there is no quantization noise, this approach is equal to the BMVDR beamformer [8]. b) **EQ-JBLCMV:** The rate-constrained variation of the method proposed in [10], [11]. The equal rate allocation across all sensors and frequencies is considered in this approach. Note that when there is no quantization noise, which happens at infinitely high rates, this method will be the same as the one proposed in [10], [11]. c) **ProposedAO-BMVDR:** In this approach, the special case of the proposed alternating optimization (AO) method described in Sec. III-B will be used to allocate the rates in the BMVDR beamforming setup. The constraint matrix Λ will simply have two columns, taking into account the distortion-less response constraints with respect to the target signal. d) **ProposedAO-JBLCMV:** In this approach, the proposed method described in Sec. III-B will be used to allocate the rates with the constraint matrix Λ mentioned in (7). Please note that to run the proposed algorithm, as well as the competing methods, the ATFs and the joint statistic

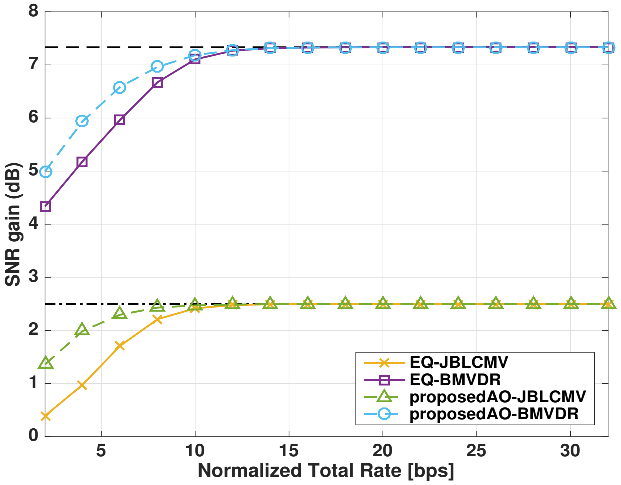


Fig. 2: SNR gain [dB] versus total rate [bit per sample] based on a binaural setup in Fig. 1 (Acoustic Scene 1).

are assumed to be known. Under stationary assumptions, and assuming that the spectral shape of the signal does not rapidly change over time, the over-head cost which is needed to inform the transmitters, on which bit-rate they should transmit the data, can be averaged out over consecutive frames.

3) Simulation Results: In this section, we will compare the methods described in the previous sub-section based on the measures introduced in Sec. IV-A. We consider the acoustical setup, shown in Fig. 1 with five interferers located at $(3\text{m}, \{-80°, -60°, -20°, 40°, 85°\})$. The signal to interferer ratio (SIR) with respect to both FCs are set to approximately 0 dB. Fig. 2 shows the SNR gains as a function of total bit budget for the above-mentioned scenario. Please note that in Fig. 2 and all the remaining results in this paper, the total bit-rate is normalized by the number of frequency samples, which is 512. The black horizontal dashed-line shows the upper bound on the performance of the BMVDR beamforming when there is no quantization noise, i.e., at infinitely high rates. Similarly, the black dashed-dotted horizontal line shows the upper bound on the performance of the JBLCMV beamforming at infinitely high rates. In fact, the BMVDR performs better than the JBLCMC in terms of SNR as it has more degrees of freedom for noise reduction, at the cost of losing some binaural information, which will be shown later in this section. The performance of the both the "EQ-BMVDR" and the "ProposedAO-BMVDR" approach that of the BMVDR at high rates without any mismatch. As shown, the proposed method significantly outperforms the methods with equal rate allocation as the alternating optimization approach is used to jointly optimize the rates and weights. A similar argument holds for the "ProposedAO-JBLCMV". The performance of the "ProposedAO-JBLCMV" is always worse than that of the "ProposedAO-BMVDR" as less degrees of freedom remain for the noise reduction, compared to BMVDR beamforming.

To see how the methods affect the preservation of the binaural spatial information, we compute the ILD and ITD errors, introduced in (28). The ILD and ITD errors are shown in Fig. 3. In this paper, the ILD and ITD errors are averaged among the target signal and the interferes.

The black dashed-line in both figures shows the asymptotic ILD and ITD errors for BMVDR beamforming, at infinitely high rates. Please note that the BMVDR method cannot preserve the spatial information with respect to the interferers, therefore there will be always ILD and ITD errors remaining in the processed signal. However, the JBLCMV beamformer can preserve the spatial information for up to $2M-3$ interferers, therefore, there is no ILD or ITD error with respect to the JBLCMV-based methods here. As shown in (21), in the proposedAO-JBLCMV method, as the weights are actually computed by the LCMV equations, it can also preserve the spatial information of $2M-3$ (which is five for $M=4$) interferers. As shown in Fig. 3a, in this specific scenario, the proposedAO-BMVDR method can perform better than the EQ-BMVDR method in terms of ILD errors at most total rates. However, as the problem proposed in (20) does not aim at optimizing the ILD or ITD errors, in general, it is not guaranteed to perform better than the equal rate allocation. The ILD and ITD errors w.r.t. both methods will approach that of the BMVDR beamforming at sufficiently high rates.

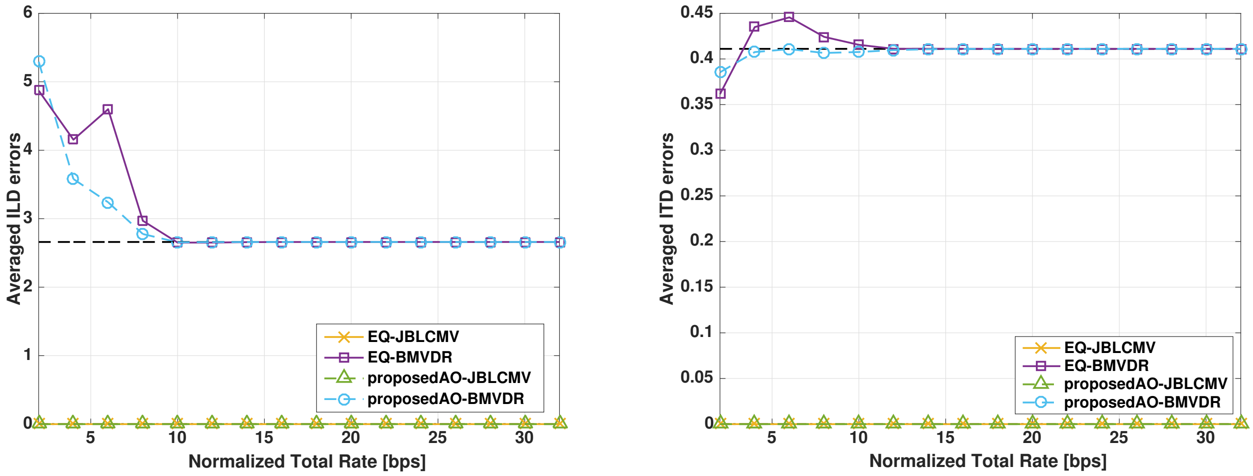
C. Example Generalized Binaural HA Setup Using Body-Related Transfer Functions

1) Acoustic Scene 2: In this section, we will compare the methods based on the generalized binaural HA setup from [35]. In addition to the binaural HA setup with four microphones as in Sec. IV-B, here, there is an assistive microphone, assumed to be mounted on the HA user's body (close to the left wrist). Therefore, this example includes five microphones. We use the body-related transfer functions (BRTFs) generated from the database presented in [35]. These impulse responses are measured with an adult human in an acoustically treated laboratory ($T_{60} \approx 200$ ms). All sources are assumed to be located at a planar distance of 2 m from the HA user. The target speech source is assumed to be located in front of the HA user and the six interferers are assumed to be located at $(2\text{m}, \{-15°, -30°, -60°, 30°, 60°, 90°\})$ with SIR set approximately to 0 dB w.r.t. both the left side and the right side reference microphones. Uncorrelated flat PSD noise is also added to the microphone signals with the SNR set to 40 dB to simulate internal microphone self noise. The PSD of the target speech and the other sources are estimated/assumed in the same fashion as described in the previous example setup in Sec. IV-B1.

2) Simulation Results: The SNR gain is shown in Fig. 4

Similar to Sec. IV-B3, The black horizontal dashed and the black dash-dotted lines denote the asymptotic BMVDR beamforming and JBLCMV beamforming SNR gains, respectively, at infinitely high rates. The performance of both "EQ-BMVDR" and "ProposedAO-BMVDR" follow a similar trend as in Fig. 2. Note that in this section, in addition to the generalized setup where there are five microphones (four microphones for the binaural HA setup and one additional assistive microphone), we also show the simulation results for the same acoustic scene, but with four microphones (without the assistive microphone), to show the benefit of having extra

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2020.3028264, IEEE/ACM Transactions on Audio, Speech, and Language Processing



(a) ILD errors (Acoustic Scene 1).

(b) ITD errors (Acoustic Scene 1).

Fig. 3: ILD and ITD errors versus total rate [bit per sample] based on the setup in Fig. 1 (Acoustic Scene 1).

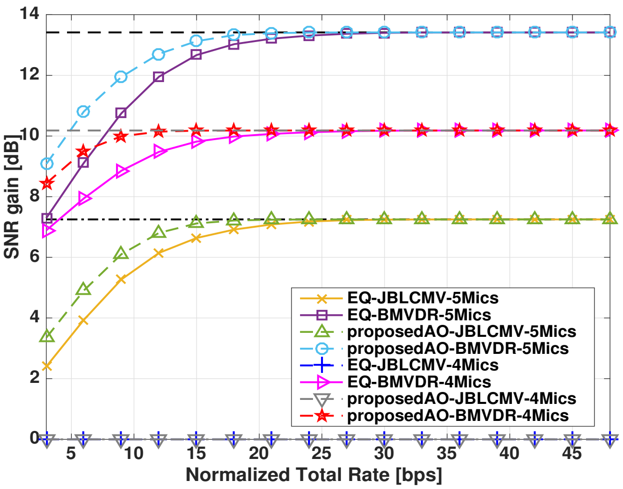


Fig. 4: SNR gain [dB] versus total rate [bit per sample] based on the generalized binaural setup using BRTFs (Acoustic Scene 2).

assistive microphone to increase the SNR gains. The methods which are based on the generalized setup are denoted by “x-5Mics”, and the methods that are based on the binaural setup are denoted by “x-4Mics”.

As shown in Fig. 4, with four microphones, the performance is always less than the case with five microphones. In fact, with six interferers, in this simulation with four microphones, all JBLCMV-based methods spend all their degrees of freedom to preserve the spatial cues of the sources and hence, there is no control over the noise reduction (i.e., no SNR gain in this case). However, the BMVDR-based methods with four microphones still have control over the amount of noise reduction. Using the proposed alternating optimization method allows for optimal rate allocation for generalized-extended binaural setups where the additional assistive microphone can help to increase the averaged SNR gain, compared to the binaural configuration with four microphones.

The ILD and ITD errors based on the generalized setup with five microphones, as well as for the binaural setup with four HA microphones, are shown in Fig. 5. As shown, All JBLCMV-based methods can guarantee the preservation of the spatial cues (the yellow, green, blue, and gray-colored curves lie on top of each other with zero ILD and ITD errors), where the BMVDR-based methods suffer from spatial cue errors. Especially, the BMVDR-based methods with five microphones, focus more on the noise reduction task, and therefore, they have slightly more ILD and ITD errors compared to the case with four microphones.

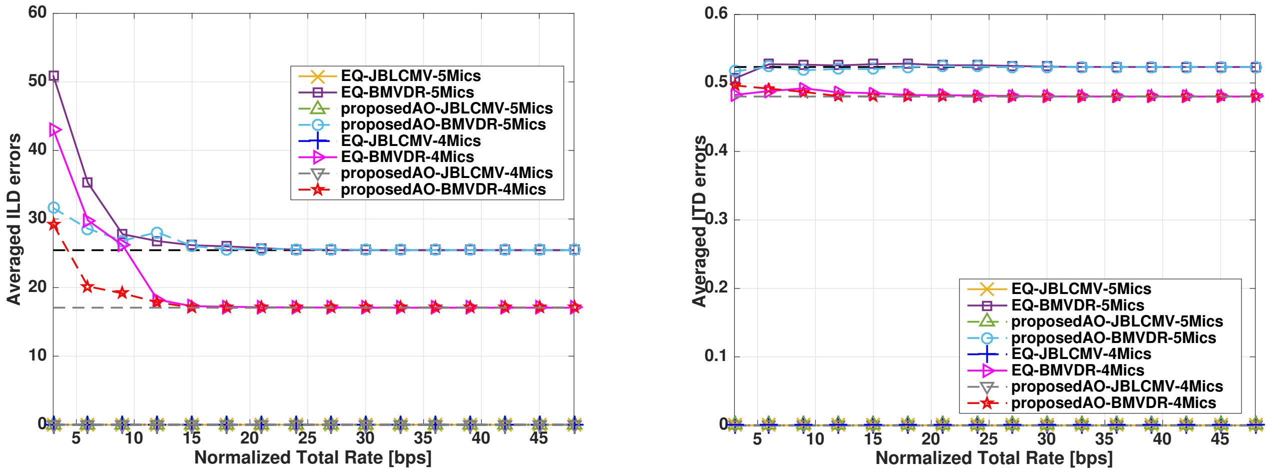
With a similar explanation as in Sec. IV-B3, the proposedAO-BMVDR, and the EQ-BMVDR methods are not able to preserve the spatial cues for all interferers as they do not impose any constraints to preserve the spatial cues of the interferers. As shown in Fig. 5b the proposedAO-BMVDR and the EQ-BMVDR methods have similar ITD errors at almost all rates, meaning that, if a certain amount of ITD error is of interest, then there is no need to send the high rate realizations to the FC, and hence, the observation can be quantized at lower rates and then transmitted. However, this argument is scenario-dependent.

Please note that similar to [27], here the proposed framework does not suffer from the scalability issue and can be applied to the more generalized scenarios including any number of microphones which can be located in random positions.

V. Conclusion

In this paper, we proposed a spatially correct rate-constrained noise reduction problem which jointly finds the best rate allocation and estimation weights across all frequencies and sensors. The problem is based on the modified rate-distortion trade-off where the optimization problem is modified to incorporate the preservation of binaural cues, which is an important factor for increasing the speech intelligibility for

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2020.3028264, IEEE/ACM Transactions on Audio, Speech, and Language Processing



(a) ILD errors (Acoustic Scene 2).

(b) ITD errors (Acoustic Scene 2).

Fig. 5: ILD and ITD errors versus total rate [bit per sample] based on the generalized binaural HA setup (Acoustic Scene 2).

$$
\begin{aligned}
L(\mathbf{R}^{\mathrm{L}},\mathbf{R}^{\mathrm{R}},\mathbf{W}^{\mathrm{L}},\mathbf{W}^{\mathrm{R}},\lambda_{\mathrm{L}},\lambda_{\mathrm{R}},\mathbf{V}^{\mathrm{L}},\mathbf{V}^{\mathrm{R}},\mathbf{M}) = {} & \frac{1}{K}\sum_{k=1}^{K}\mathbf{w}_k^{\mathrm{H}}\mathbf{\Phi}_k\mathbf{w} + \lambda_{\mathrm{L}}\left(\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M}[r_{km}^{\mathrm{L}}] - R_{\mathrm{tot}}^{\mathrm{L}}\right) + \lambda_{\mathrm{R}}\left(\sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}}[r_{km}^{\mathrm{R}}] - R_{\mathrm{tot}}^{\mathrm{R}}\right) \\
& - \sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M}[v_{km}^{\mathrm{L}}r_{km}^{\mathrm{L}}] - \sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}}[v_{km}^{\mathrm{R}}r_{km}^{\mathrm{R}}] + \sum_{k=1}^{K}\left(\mathrm{Re}\{\boldsymbol{\mu}_k\}^{\mathrm{T}}\mathrm{Re}\{\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k\} - \mathrm{Re}\{\boldsymbol{\mu}_k\}^{\mathrm{T}}\mathrm{Re}\{\mathbf{f}_k\}\right) \\
& + \sum_{k=1}^{K}\left(\mathrm{Im}\{\boldsymbol{\mu}_k\}^{\mathrm{T}}\mathrm{Im}\{\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k\} - \mathrm{Im}\{\boldsymbol{\mu}_k\}^{\mathrm{T}}\mathrm{Im}\{\mathbf{f}_k\}\right).
\end{aligned}
\tag{29}
$$

hearing aid users. Solving the proposed optimization problem, based on the set of linear cue preservation constraints, the estimation (beamformer) weights are found to be the rate-dependent LCMV filters, and the rates are the solutions to the set of water filling problems. We chose two different acoustic scenes to evaluate the performance of the proposed methods: 1) The binaural HA setup with four microphones using HRTFs. 2) The generalized binaural HA setup with five microphones using BRTFs, where an additional assistive microphone is collaborating with HAs. We compared the BMVDR-based methods with the JBLCMV-based methods. The performance of the proposed method is evaluated using SNR gains and ILD and ITD errors. The results showed that the proposed method outperforms the methods with naive/equal choices of rates. In addition, as shown in Fig. 2 and Fig. 4, the BMVDR-based methods perform better than JBLCMV-based methods in terms of SNR in both scenarios as there is more degree of freedom for noise reduction, at the cost of losing some spatial information of the sources. This behavior is consistent across different scenarios.

Appendix A
Derivations of the Proposed Solution in (21)

The solution to the optimization problem in (20) is given by (21). In this section, we show the derivations leading to (21). We solve the KKT conditions, derived based on the problem in (20).

The Lagrangian function is given by (29). The matrix $\mathbf{M}$ includes the multipliers $\boldsymbol{\mu}_k$, i.e., $\mathbf{M} = [\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_K]$, and matrices $\mathbf{V}^{\mathrm{L}}$ and $\mathbf{V}^{\mathrm{R}}$ includes entries v_{km}^{L} and v_{km}^{R}, respectively. Given that

$$
\begin{aligned}
\mathrm{Re}\{\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k\} &= \frac{\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k + \mathbf{\Lambda}_k^{\mathrm{T}}\mathbf{w}_k^{*}}{2}, \\
\mathrm{Im}\{\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k\} &= \frac{\mathbf{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k - \mathbf{\Lambda}_k^{\mathrm{T}}\mathbf{w}_k^{*}}{2i},
\end{aligned}
\tag{30}
$$

the KKT condition w.r.t. the Lagrangian function in (29) is given by

$$
L_{\mathbf{w}_k^{*}} = \frac{1}{K}\mathbf{\Phi}_k\mathbf{w}_k + \frac{\mathbf{\Lambda}_k\mathrm{Re}\{\boldsymbol{\mu}_k\}}{2} - \frac{\mathbf{\Lambda}_k\mathrm{Im}\{\boldsymbol{\mu}_k\}}{2i} = 0, \tag{31a}
$$

$$
L_{r_{km}^{\mathrm{L}}} = \frac{-2\ln 2|w_{km}^{\mathrm{L}}|^2\, k_{km}^{\mathrm{L}}}{K2^{2r_{km}^{\mathrm{L}}}} + \lambda_{\mathrm{L}} - v_{km}^{\mathrm{L}} = 0, \tag{31b}
$$

$$
L_{r_{km}^{\mathrm{R}}} = \frac{-2\ln 2|w_{km}^{\mathrm{R}}|^2\, k_{km}^{\mathrm{R}}}{K2^{2r_{km}^{\mathrm{R}}}} + \lambda_{\mathrm{R}} - v_{km}^{\mathrm{R}} = 0, \tag{31c}
$$

$$
\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M} r_{km}^{\mathrm{L}} \le R_{\mathrm{tot}}^{\mathrm{L}}, \tag{31d}
$$

$$
\sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}} r_{km}^{\mathrm{R}} \le R_{\mathrm{tot}}^{\mathrm{R}}, \tag{31e}
$$

$$
\left(\sum_{k=1}^{K}\sum_{m=M^{\mathrm{L}}+1}^{M} r_{km}^{\mathrm{L}} - R_{\mathrm{tot}}^{\mathrm{L}}\right)\lambda_{\mathrm{L}} = 0, \tag{31f}
$$

$$\left(\sum_{k=1}^{K}\sum_{m=1}^{M^{\mathrm{L}}+M^{\mathrm{A}}} r_{km}^{\mathrm{R}} - R_{\mathrm{tot}}^{\mathrm{R}}\right)\lambda_{\mathrm{R}} = 0, \tag{31g}$$

$$\lambda_{\mathrm{L}} \geq 0, \quad \lambda_{\mathrm{R}} \geq 0, \tag{31h}$$

$$r_{km}^{\mathrm{L}} \geq 0, \quad r_{km}^{\mathrm{R}} \geq 0, \tag{31i}$$

$$r_{km}^{\mathrm{L}} v_{km}^{\mathrm{L}} = 0, \quad r_{km}^{\mathrm{R}} v_{km}^{\mathrm{R}} = 0, \tag{31j}$$

$$v_{km}^{\mathrm{L}} \geq 0, \quad v_{km}^{\mathrm{R}} \geq 0. \tag{31k}$$

$$\boldsymbol{\Lambda}_k^{\mathrm{H}}\mathbf{w}_k = \mathbf{f}_k. \tag{31l}$$

First, we solve the KKT conditions w.r.t. the estimation weights $\mathbf{w}_k$. Solving (31a) for $\mathbf{w}_k$, we have

$$\mathbf{w}_k^\star = K\boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k(\frac{\mathrm{Re}\{\boldsymbol{\mu}^\star\} + i\mathrm{Im}\{\boldsymbol{\mu}^\star\}}{2}) = \frac{K}{2}\boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k\boldsymbol{\mu}^\star. \tag{32}$$

Substituting (32) into the linear constraint (31l) and solving (31l), the optimal $\boldsymbol{\mu}^\star$ is given by

$$\boldsymbol{\mu}^\star = \frac{2}{K}(\boldsymbol{\Lambda}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k)^{-1}\mathbf{f}_k. \tag{33}$$

Finally, substituting (33) back into (32), the optimal weights are given by

$$\mathbf{w}_k^\star(\mathbf{r}_k^{\mathrm{L}\star}, \mathbf{r}_k^{\mathrm{R}\star}) = \boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k(\boldsymbol{\Lambda}_k^{\mathrm{H}}\boldsymbol{\Phi}_k^{-1}\boldsymbol{\Lambda}_k)^{-1}\mathbf{f}_k. \tag{34}$$

Note that, unlike the original BLCMV solution, here the optimal weights $\mathbf{w}_k^\star$, as well as the PSD matrix $\boldsymbol{\Phi}_k$ are functions of the optimal bit-rates, which will be derived in the following.

As the constraint functions for r_{km}^{L} and r_{km}^{R} are separable, we can independently solve the KKT equations w.r.t. the corresponding rates. We start with the solution for r_{km}^{L}. Solving (31b) for v_{km}^{L}, and substituting it into the complementary slackness condition in (31j), we have

$$(\frac{-2\ln2|w_{km}^{\mathrm{L}}|^2\, k_{km}^{\mathrm{L}}}{K2^{2r_{km}^{\mathrm{L}}}} + \lambda_{\mathrm{L}})r_{km}^{\mathrm{L}} = 0. \tag{35}$$

Looking at (35), there are two cases here: 1) the optimal rate r_{km}^{L} is set to zero, when based on (31j), the variable v_{km}^{L} has to be strictly greater than zero, which, by looking at (31b), implies $\frac{\lambda_{\mathrm{L}}K}{2\ln2} \geq |w_{km}^{\mathrm{L}}|^2\, k_{km}^{\mathrm{L}}$. 2) $v_{km}^{\mathrm{L}} = 0$, then solving (31b) for r_{km}^{L}, the optimal non-zero valued rates are given by

$$r_{km}^{\mathrm{L}\star} = \frac{1}{2}\log_2(\frac{|w_{km}^{\mathrm{L}\star}|^2\, k_{km}^{\mathrm{L}}}{\frac{K\lambda_{\mathrm{L}}^\star}{2\ln2}}), \tag{36}$$

which implies $\frac{\lambda_{\mathrm{L}}K}{2\ln2} < |w_{km}^{\mathrm{L}}|^2\, k_{km}^{\mathrm{L}}$. Combining cases 1 and 2, we have

$$r_{km}^{\mathrm{L}\star}(\lambda_{\mathrm{L}}'^\star, w_{km}^{\mathrm{L}\star}) = [\frac{1}{2}\log_2(\frac{|w_{km}^{\mathrm{L}\star}|^2\, k_{km}^{\mathrm{L}}}{\lambda_{\mathrm{L}}'^\star})]^+, \tag{37}$$

where $\lambda_{\mathrm{L}}'^\star = \frac{K\lambda_{\mathrm{L}}^\star}{2\ln2}$. The operator $[\cdot]^+$ assures positive rates and projects all negative values onto zero. The parameter $\lambda_{\mathrm{L}}'^\star$ must satisfy the KKT condition (31d) with equality, as argued in [27]. Note that the rates are functions of the weights $w_{km}^{\mathrm{L}\star}$ and the water-falling threshold parameter $\lambda_{\mathrm{L}}^\star$. Therefore, the alternating optimization is proposed to be used to solve theses equations in (37) and (34). A similar proof holds for $r_{km}^{\mathrm{R}\star}$.

Finally to find the optimal $\lambda_{\mathrm{L}}^\star$ and $\lambda_{\mathrm{R}}^\star$, a similar water-filling approach, as proposed in [27] (in the last part of the proof in the appendix), can be used.

References

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

[2] P. Vary and R. Martin, *Digital Speech Transmission: Enhancement, Coding And Error Concealment*, John Wiley & Sons, 2006.

[3] 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.

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

[5] E. Hadad, S. Gannot, and S. Doclo, "Binaural linearly constrained minimum variance beamformer for hearing aid applications," in *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, Sep. 2012, pp. 1–4.

[6] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 342–355, Feb 2010.

[7] 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.

[8] 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.

[9] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Optimal binaural lcmv beamformers for combined noise reduction and binaural cue preservation," in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2014, pp. 288–292.

[10] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, "Improved multi-microphone noise reduction preserving binaural cues," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 460–464.

[11] E. Hadad, D. Marquardt, D. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function mvdr beamformers with interference cue preservation constraints," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2449–2464, Dec 2015.

[12] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed binaural LCMV beamforming," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 1, pp. 137–152, 2017.

[13] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "A convex approximation of the relaxed binaural beamforming optimization problem," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 2, pp. 321–331, 2019.

[14] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel wiener filter for robust noise reduction," *Speech Communication*, vol. 49, no. 7-8, pp. 636–656, 2007.

[15] E. Hadad, S. Doclo, and S. Gannot, "The binaural lcmv beamformer and its performance analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 543–558, 2016.

[16] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical analysis of linearly constrained multi-channel wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2384–2397, Dec 2015.

[17] 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.

[18] S. Doclo, T. C. Lawin-Ore, and T. Rohdenburg, "Rate-constrained binaural MWF-based noise reduction algorithms," *in Proc. ITG Conference on Speech Communication, Bochum, Germany,*, Oct 2010.

[19] 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.

[20] 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.

[21] S. Srinivasan, "Low-bandwidth binaural beamforming," *Electronics Letters*, vol. 44, no. 22, pp. 1292–1293, Oct 2008.

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

[23] O. Roy and M. Vetterli, "Collaborating hearing aids," *in Proceedings of MSRI Workshop on Mathematics of Relaying and Cooperation in Communication Networks*, April 2006.

[24] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Asymmetric coding for rate-constrained noise reduction in binaural hearing aids," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 154–167, Jan 2019.

[25] 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.

[26] 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.

[27] J. Amini, R. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Rate-constrained noise reduction in wireless acoustic sensor networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1–12, 2020.

[28] 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.

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

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

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

[32] 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.

[33] 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.

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

[35] R. M. Corey, N. Tsuda, and A. C. Singer, "Acoustic impulse responses for wearable audio devices," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 216–220.