Asymmetric Coding for Rate-Constrained Noise Reduction in Binaural Hearing Aids

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

***Abstract*—Binaural hearing aids (HAs) can potentially perform advanced noise reduction algorithms, leading to an improvement over monaural/bilateral HAs. Due to the limited transmission capacities between the HAs and given knowledge of the complete joint noisy signal statistics, the optimal rate-constrained beamforming strategy is known from the literature. However, as these joint statistics are unknown in practice, sub-optimal strategies have been presented. In this paper, we present a unified framework to study the performance of these existing optimal and sub-optimal rate-constrained beamforming methods for binaural HAs. Moreover, we propose to use an asymmetric sequential coding scheme to estimate the joint statistics between the microphones in the two HAs. We show that under certain assumptions, this leads to sub-optimal performance in one HA but allows to obtain the truly optimal performance in the second HA. Based on the mean square error distortion measure, we evaluate the performance improvement between monaural beamforming (no communication) and the proposed scheme, as well as the optimal and the existing sub-optimal strategies in terms of the information bit-rate. The results show that the proposed method outperforms existing practical approaches in most scenarios, especially at middle rates and high rates, without having the prior knowledge of the joint statistics.**

***Index Terms*—Binaural hearing aids (HAs), multi-microphone noise reduction, remote source coding.**

I. Introduction

NOWDAYS hearing aids (HAs) include digital signal processing algorithms to improve the intelligibility of the signal of interest. The HAs record the acoustic field with one or multiple microphones and try to improve the intelligibility of the desired speech signal while reducing environmental noise [1]. Using a wireless link, the two HAs can collaborate with each other and construct a binaural HA system to increase noise suppression and potentially preserve some important binaural (spatial) cues [2]. This leads to the notion of binaural multi-microphone noise reduction which may have a better speech intelligibility compared to single-microphone solutions [3], [4]. In addition, unlike bilateral systems or independent dual-channel systems, where a set of two monaural systems operate independently from each other (no collaboration between the devices), the binaural system exploits the spatial diversity, using more observations, and interaural information to increase the speech intelligibility [5]. A well known binaural multi-microphone noise reduction algorithm is the binaural multichannel Wiener filter (MWF) [6]–[9]. The binaural MWF includes two separate MWF beamformers for the two HAs. Each MWF beamformer combines its local observations with those of the contralateral HA to estimate its own version of the target signal such that the mean square error (MSE) between the target signal and its estimate is minimized. The binaural MWF allows for more degree of freedom for noise reduction [5], among all the MMSE-based single-channel or dual-channel speech enhancement methods [10], as it exploits the full potential of binaural processing by exchanging the signals between the HA devices (more microphone observations, the better the noise reduction performance) and as it outputs the best MSE estimate of the target signal. Binaural filters require that noisy observations from one HA are transmitted to the other one (e.g., through a wireless link) in order to be combined with local observations. Typically, transmission capacities are limited due to limited battery life-time [11], [12], which necessitates data compression. Ideally, the algorithm trades off the transmission bit-rate of contralateral HA observations against the estimation error of the target signal [12, Section I], which is remotely (i.e., indirectly after being filtered by the room channel) observable at the HAs .

From an information theoretic viewpoint, such an estimator can be seen as remote source coding [13]–[15]. Since the beamformer needs to decode the transmitted signals and combine these with its local observations (which are available error-free as "side information"), the more accurate problem formulation would be that of remote source coding with side information at the decoder. For directly observable sources this problem is referred to as Wyner-Ziv (WZ) coding [16] and for remote (i.e., indirectly observable) sources as remote WZ coding [17]. In fact, in remote source coding problems the (remote) target signals are of interest, and not necessarily the (noisy) direct observations. In [12, Section III-A] this problem is considered, assuming jointly Gaussian random sources, and an optimal tradeoff between the transmission rate and the MSE between the target signal and its estimate is derived. The method provides an upper bound

Manuscript received April 23, 2018; revised August 27, 2018; accepted October 1, 2018. Date of publication October 15, 2018; date of current version October 26, 2018. This work was supported in part by the Oticon Foundation and in part by NWO, the Dutch Organization for Scientific Research. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andy W. H. Khong. *(Corresponding author: Jamal Amini.)*

J. Amini, R. C. Hendriks, and R. Heusdens are with the Department of Electrical Engineering, Mathematics and Computer Engineering, Delft University of Technology, Delft 2628 CD, The Netherlands (e-mail: j.amini@tudelft.nl; r.c.hendriks@tudelft.nl; r.heusdens@tudelft.nl).

M. Guo is with Oticon A/S, Smørum 2765, Denmark (e-mail: megu@oticon.com).

J. Jensen is with Oticon A/S, Smørum 2765, Denmark, and also with the Electronic Systems Department, Aalborg University, Aalborg 9100, Denmark (e-mail: jesj@oticon.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2018.2876172

on the performance of the minimum MSE (MMSE)-based rate-constrained binaural beamforming algorithms. However, the requirement of knowing the (joint) statistics severely limits the application of the method in practice. In fact, the joint statistics between the two HA observations remain unknown in practice and can only be estimated if realizations are exchanged between the HAs.

Several sub-optimal approaches are proposed in [12, Section III-B], [18]–[20] in which local functions of the contralateral observations are transmitted, projecting the multiple signals on to a single signal. These methods provide practical alternatives to the optimal method in [12, Section III-A], as they do not need the knowledge of the (joint) statistics. However, the blind projection of the multi-microphone observations to one signal tends to a significant asymptotic mismatch in the performance even at sufficiently high rates [18]. An iterative reduced bandwidth MWF-based beamformer is proposed in [20], where local estimates of the target signal are assumed to be exchanged error-free between the HAs without any rate constraint. It is shown in [20] that for a single target signal, the iterative approach converges to the binaural MWF after sufficient transmissions between HAs. However, when analyzing the rate-constrained scenario [21], the total rate is distributed over transmissions (iterations) which results in a poor final performance (after convergence) in terms of bit-rate.

Typically, the aforementioned sub-optimal approaches do not make use of the joint statistics between the two HA observations. As a result, before exchanging the signals, some information will be removed which may be necessary for the other side to cancel out the noise sources, leading to an asymptotic sub-optimality. By asymptotic sub-optimality, we mean that the performance does not approach the optimal performance for increasing rate. Therefore, any knowledge (even incomplete) of the (joint) statistics between the nodes may be crucial to keep the necessary information when filtering the information, resolving the asymptotic sub-optimality, and to provide a good tradeoff between the rate and the distortion. This motivates trying to estimate the joint statistics.

In this paper, we study the performance of sub-optimal rate-constrained noise reduction techniques based on a unified encoding-decoding framework which can easily be translated to the existing sub-optimal schemes by changing certain parameters. Moreover, we propose an asymmetric sequential coding approach for the transmission of the information from one HA to the other HA (which we will refer to as *Link 1*) and vice versa (which we will refer to as *Link 2*). In addition, we propose an extension of the probability distribution preserving quantization method [22], to vector sources, to be used in *Link 1*. Using this distribution preserving quantization, the unquantized statistics can be retrieved and used to apply the optimal coding strategy [12, Section III-A] in *Link 2*. Based on the MSE criteria, the distortion gap between the monaural noise reduction approach, in which there is no communication between devices, and different sub-optimal/optimal noise reduction approaches are compared for both links. The results show that the proposed methods outperform the sub-optimal approaches in most practical scenarios and confirm the optimal asymptotic behavior of the proposed methods.

The paper is organized as follows. In Section II the binaural HA problem is stated and the well-known information theoretic rate-distortion tradeoff is introduced. In Section III we state the rate-constrained noise reduction problem in a unified framework and the optimal and some sub-optimal approaches are explained. The proposed asymmetric 2-way coding scheme is presented in Section IV. The performance analysis of the proposed and existing methods is carried out in Section V. Finally, Section VI concludes the paper.

II. Problem Statement

A. Signal Model

A typical binaural HA system consists of two wireless collaborating HAs. Assume that the left-side and right-side HAs include M_1, and M_2 microphones, respectively, with $M = M_1 + M_2$ microphones in total. All microphones record a filtered version of the target speech signal which is denoted in the frequency domain by $S[k]$, corrupted by and additive noise $N[k]$, with k the discrete frequency bin index. The frequency-domain description of the noisy observation captured by the ith microphone is given by

$$Y_i[k] = A_i[k]S[k] + N_i[k], \tag{1}$$

where $i = 1, \ldots, M$, A_i is the acoustic transfer function (ATF) between the target signal and the ith microphone. Stacking all noisy observations across the microphones in a vector, the signal model can be rewritten as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad \mathbf{y} \in \mathbb{C}^M. \tag{2}$$

where $\mathbf{y} = [Y_1[k], \ldots, Y_{M_1}[k], Y_{M_1+1}[k], \ldots, Y_M[k]]^{\mathrm{T}}$ denotes the total M noisy microphone signals, $\mathbf{x} = \mathbf{a}S$, and

$$\mathbf{a} = [A_1[k], \ldots, A_M[k]]^{\mathrm{T}}, \mathbf{n} = [N_1[k], \ldots, N_M[k]]^{\mathrm{T}}.$$

Note that the frequency index k is omitted, when defining the signal vectors, for ease of notation. To distinguish between the left-side and the right-side noisy microphone observations, vectors $\mathbf{y}_1 \in \mathbb{C}^{M_1}$ and $\mathbf{y}_2 \in \mathbb{C}^{M_2}$ are defined, respectively, as $\mathbf{y}_1 = [Y_1[k], \ldots, Y_{M_1}[k]]^{\mathrm{T}}$ and $\mathbf{y}_2 = [Y_{M_1+1}[k], \ldots, Y_M[k]]^{\mathrm{T}}$. The superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote transpose and conjugate transpose operators, respectively. All sources are assumed to be zero-mean and mutually uncorrelated. The cross-power spectral density (CPSD) matrix of the noisy signal vector $\mathbf{y}$, denoted by $\mathbf{\Phi}_{\mathbf{y}}$, can then be written as $\mathbf{\Phi}_{\mathbf{y}} = \mathbf{\Phi}_{\mathbf{x}} + \mathbf{\Phi}_{\mathbf{n}}$, $\mathbf{\Phi}_{(\cdot)} \in \mathbb{C}^{M \times M}$, with $\mathbf{\Phi}_{\mathbf{x}} = \Phi_S \mathbf{a}\mathbf{a}^{\mathrm{H}}$. Here, Φ_S is the power spectral density (PSD) of the clean speech signal S and $\mathbf{\Phi}_{\mathbf{n}} = E[\mathbf{n}\mathbf{n}^{\mathrm{H}}]$, where $E[\cdot]$ denotes the expectation.

The goal of the multi-microphone noise reduction algorithms is to estimate the clean speech signal while suppressing the environmental noise power. The binaural MWF [7], [8] consists of two filters (the left-side and the right-side filters). Let the left and right reference microphone indices be denoted by 1 and $M_1 + 1$, respectively. The filters estimate the target signal at the left-side and the right-side reference microphones, say $S_1 = A_1 S$ and $S_2 = A_{M_1+1} S$, respectively, by minimizing the MSE between the target signal and its estimates, say $\hat{S}_1$ and $\hat{S}_2$, respectively. Scalars A_1 and A_{M_1+1} denote the ATFs with

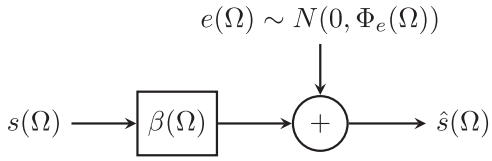


Fig. 1. Forward test channel representation of lossy source coding.

respect to the corresponding reference microphones. Finally, the estimates are given by $\hat{S}_1^\star = E[S_1|\mathbf{y}]$ and $\hat{S}_2^\star = E[S_2|\mathbf{y}]$.

Computing each of these MWF outputs requires the availability of the error-free contralateral noisy signal realizations. In practice, only a compressed/quantized version of the contralateral noisy signals are available. These signals are compressed at a certain rate. Therefore, the problem can be viewed as a rate-constrained estimation task which will be described in the next section.

B. Rate-Distortion Function (RDF) [13, Ch. 4]

Let $s^N = \{s[i]\}_{i=1}^N$ be a sequence of discrete stationary Gaussian source samples, where $s[i] \in \mathbb{C}$ and N is the number of samples. The sequence s^N can be thought of as a single microphone observation along the time-axis. Assume that the encoder maps the sequence with R bits per sample to a bit sequence. The decoder receives the bit sequence and produces the quantized sequence $\hat{s}^N = \{\hat{s}[i]\}_{i=1}^N$.

The direct rate-distortion problem is to find the minimum asymptotic achievable rate at which the sequence can be encoded such that the reconstruction error does not exceed a certain value D, as $N \to \infty$ [13]. The problem is solved in [13, Ch. 4] and a parametric rate-distortion tradeoff is found, analytically. To achieve the optimal tradeoff [13, Ch. 4], for a stationary Gaussian source, the optimal forward test channel interpretation is presented in [13, Ch. 4] and [18, Section 3], which is illustrated in Fig. 1, where the quantization procedure in the frequency domain can be thought of as a test channel with input $s(\Omega)$ and output $\hat{s}(\Omega)$. The channel noise $e(\Omega)$ is uncorrelated to the input source $s(\Omega)$. The quantization parameters are computed as [13]

$$\beta(\Omega) = \max\left(0, 1 - \frac{\theta}{\Phi_s(\Omega)}\right),$$
$$\Phi_e(\Omega) = \max\left(0, \theta\left(1 - \frac{\theta}{\Phi_s(\Omega)}\right)\right), \tag{3}$$

where Φ_s is the PSD of the sequence s^N, $N \to \infty$, and $\theta \in (0, \sup \Phi_s]$ denotes the "reverse water filling" threshold parameter [13], [23].

III. Rate-Constrained Noise Reduction

Binaural rate-constrained noise reduction (RCNR) aims at estimating the target signal at the reference microphones, given some local observations and the quantized contralateral observations, such that the communication rate between HAs is minimized, satisfying a certain constraint. For the right-side beamformer, the local observation vector $\mathbf{y}_2$ acts as the "side information" and $\mathbf{y}_1$ as the contralateral observations. A similar argument holds for the left-side beamformer.

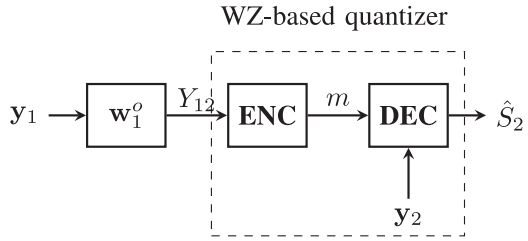


Fig. 2. Optimal rate-constrained beamforming.

Most approaches like [12], [18] use the following structure of three stages. First, the contralateral observations are filtered prior to being quantized, as we are interested in information about the target source (S_1 or S_2) and not necessarily in those of the contralateral noisy observations themselves (filtering stage). Second, the filtered signals are quantized and transmitted to the other side (quantization stage). Finally, the target signal is estimated given the side information and the filtered-quantized contralateral observations (estimation stage). Existing approaches differ in how different operators for these three processing stages of filtering, quantization, and final estimation are chosen, which we will explain using the above-mentioned unified description.

A. Optimal Rate-Constrained Noise Reduction

Encoding the sources for a decoder which has access to the side information is known as the Wyner-Ziv (WZ) problem [16]. Based on the WZ coding, In [12, Section III-A] the problem is optimally solved for the multiple microphones per HA setup (binaural setup) and the optimal rate-distortion tradeoff is found analytically. For stationary Gaussian sources, the interpretation of the optimal RCNR system is illustrated in Fig. 2 [12, Section III-A]. From the right-side beamformer's perspective, first, the left-side noisy signals ($\mathbf{y}_1$) are filtered as $Y_{12} = (\mathbf{w}_1^o)^{\mathrm{H}}\mathbf{y}_1$, using the joint statistics between the HA observations, with the optimal coefficients $\mathbf{w}_1^o$ computed as

$$\mathbf{w}_1^o = \mathbf{\Phi}_{\bar{\mathbf{y}}_1}^{-1}\mathbf{\Phi}_{\bar{\mathbf{y}}_1 S_2}, \quad \mathbf{w}_1^o \in \mathbb{C}^{M_1 \times 1} \tag{4}$$

where $\mathbf{\Phi}_{\bar{\mathbf{y}}_1}$ is the CPSD matrix of the direct innovation process $\bar{\mathbf{y}}_1$ defined as $\bar{\mathbf{y}}_1 = \mathbf{y}_1 - E[\mathbf{y}_1|\mathbf{y}_2]$ and $\mathbf{\Phi}_{\bar{\mathbf{y}}_1 S_2}$ is the CPSD vector between S_2 and $\bar{\mathbf{y}}_1$.

Second, using the WZ coding philosophy [16], the filtered signal Y_{12} will be optimally encoded, knowing the joint statistics and that the decoder has access to $\mathbf{y}_2$. The WZ-based decoder (Fig. 2) consists of an MMSE estimator, which estimates S_2, the target signal at the right-side reference microphone, given $\mathbf{y}_2$ and the quantized version of Y_{12}. See [12, Section III-A] for more details.

B. Sub-Optimal Rate-Constrained Noise Reduction

Achieving the optimal rate-distortion tradeoff as in [12, Section III-A] requires knowledge of the joint statistics between the noisy signals from both HAs, which are not available in practice. In [12, Section III-B] a sub-optimal method is presented in which a local estimate of the target signal, without using the correlation between the two HA observations, is transmitted to the contralateral device. However, in the presence of point

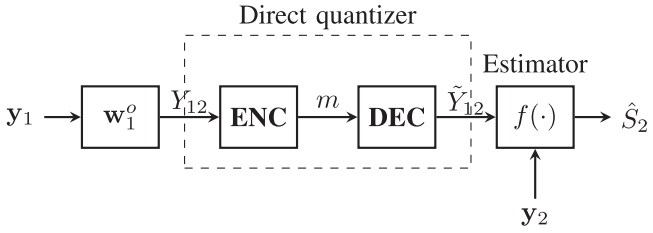


Fig. 3. Sub-optimal rate-constrained beamforming.

noise sources, the performance does not approach the ideal binaural performance, not even asymptotically (at infinite bit-rate) and a significant loss will occur at high rates, as confirmed by experiments in Section V. Two alternatives to the method presented in [12, Section III-B] were proposed in [18], [19]. We briefly explain these sub-optimal methods based on a unified communication scheme illustrated in Fig. 3.

Unlike [12, Section III-A], the sub-optimal filter $\mathbf{w}_1^s$, shown in Fig. 3, is only a function of the local observations. The above-mentioned sub-optimal methods differ from each other in how the filter $\mathbf{w}_1^s$ is chosen. For example, in [12, Section III-B], $\mathbf{w}_1^s$ denotes a filter which locally estimates the target signal, without any access to the side information.

The quantization stage in Fig. 3 can be represented by the forward test channel (Fig. 1) with input Y_{12} and output $\tilde{Y}_{12}$. The final MMSE estimate of the desired signal S_2 is given by $\hat{S}_2 = f(\mathbf{y}_2, \tilde{Y}_{12})$ and the corresponding MMSE by [18]

$$D_2(\theta) = \frac{1}{2\pi}\int_0^{2\pi} [\Phi_{S_2}(\Omega) - \mathbf{\Phi}_{S_2\tilde{\mathbf{y}}}(\Omega)\mathbf{\Phi}_{\tilde{\mathbf{y}}}^{-1}(\Omega)\mathbf{\Phi}_{\tilde{\mathbf{y}}S_2}(\Omega)]d\Omega, \tag{5}$$

where $\tilde{\mathbf{y}} = [\mathbf{y}_2^{\mathrm{T}}, \tilde{Y}_{12}]^{\mathrm{T}}$. Note that $\tilde{Y}_{12}$ depends on the rate of transmission.

IV. Asymmetric Coding for RCNR

As described in Section III-B, the main limitation of the sub-optimal methods is that they are not even asymptotically optimal, because of the blind filtering stage at the start of the communication chain. This results in a significant performance loss, as shown in Section V.

Our proposed idea is to leave all contralateral observations active, at least in one direction (for example, in the transmission from left-to-right) in order to exploit some statistics which will be helpful for informed coding in the other direction. This brings the notion of vector source quantization into the estimation process in one link. Assuming stationarity in the time domain and sufficiently large sample sequences, the quantization can be performed in the frequency domain assuming independent frequency bins. However, it is noteworthy that microphones at different spatial positions generally capture the sources with different powers, resulting in spatially non-stationary signals. Therefore the optimal quantization for microphone signals is done in the frequency domain along the time axis and in the eigenvalue domain of the CPSDs along the space axes. This will become more clear in Section IV-A1.

We propose a 2-way sequential coding scheme for communication between two HAs. This scheme is asymmetric in the sense that the quantization in one link is different from the other link. The proposed coding scheme is sequential meaning that some information is exploited in one link to be used in coding in another link. The scheme is illustrated in Fig. 4. The link where the communication starts is referred to as "*Link 1*" and vice versa as "*Link 2*". Let us assume the communication starts from the left HA to the right HA. In the following, we explain the proposed architecture.

A. Link 1: From Left-to-Right

Unlike the common RCNR techniques, in this method, the observation vector $\mathbf{y}_1$ is not projected onto a scalar signal (i.e., without filtering) for two reasons. First, we wish to resolve the asymptotic sub-optimality problem at high rates, and secondly, we wish to exploit the joint statistics at the right-side HA to reduce the redundancy in information transmission in *Link 2*. We introduce two methods based on the architecture in Fig. 4(a).

1) Method 1. RDF for Vector Sources With Memory: In Section II-B we explained the RDF for a time-stationary Gaussian source which accounts for one sensor observation in time. In order to quantize more than one observation, an extension to vector sources is required, which is presented in [24]. Recall that in the scalar case (Section II-B), the correlation matrix Φ_s is (for $N \to \infty$) diagonalizable by the Fourier transform and, hence, the RDF can be written in terms of the PSD of the stationary source, i.e., Φ_s [13]. Different from the scalar case in Section II-B, the correlation matrices involving multiple microphone observations are not diagonalizable by the (spatial) Fourier transform. Therefore, the resulting RDF for vector sources of such (spatially) non-stationary sources is different from that of the scalar case in Section II-B, and will be explained in the following.

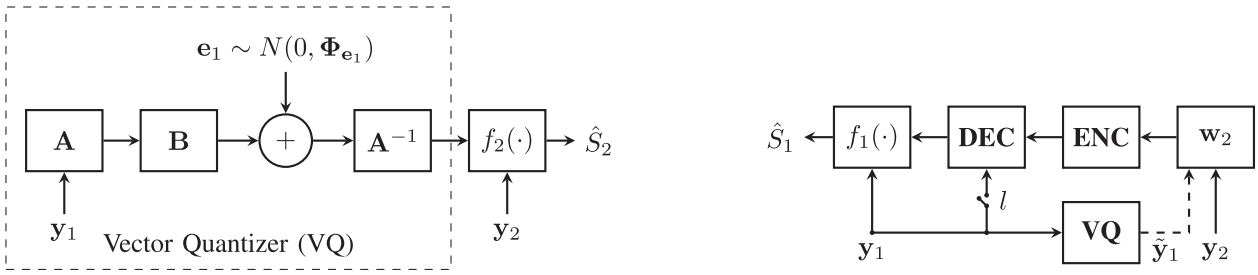
Given a discrete-time sequence of zero-mean time-stationary vector Gaussian sources, say $\{\mathbf{s}[n]\}_{n=0}^{N-1}$, where $\mathbf{s}[\cdot] \in \mathbb{R}^{M\times 1}$ can be any vector source (like the noisy observations), the cross correlation matrix is given by

$$\mathbf{\Sigma}_{\mathbf{s}} = \begin{bmatrix} \mathbf{\Sigma}_0 & \mathbf{\Sigma}_{-1} & \dots & \mathbf{\Sigma}_{-(N-1)} \\ \mathbf{\Sigma}_1 & \ddots & \ddots & \mathbf{\Sigma}_{-(N-2)} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{\Sigma}_{(N-1)} & \mathbf{\Sigma}_{(N-2)} & \dots & \mathbf{\Sigma}_0 \end{bmatrix} \tag{6}$$

where $\mathbf{\Sigma}_{\mathbf{s}} \in \mathbb{R}^{NM\times NM}$ is a block-Toeplitz matrix. Matrices $\mathbf{\Sigma}_i \in \mathbb{R}^{M\times M}$, $\quad i = -(N-1), \dots, (N-1)$ are of entries $[\mathbf{\Sigma}_i]_{uv} = E[s_u[n+i]s_v[n]]$, for all $0 \le n \le N-1$. The scalars $s_u[\cdot]$ is the uth entry in $\mathbf{s}[\cdot]$ and $[\mathbf{\Sigma}_i]_{uv}$ is the (statistical) cross correlation, $u, v = 1, \dots, M$. Stacking $\{\mathbf{s}[n]\}_{n=0}^{N-1}$ into a vector, say $\mathbf{s}_{\text{vec}} = [\mathbf{s}^T[0] ... \mathbf{s}^T[N-1]]^T$, the rate-distortion tradeoff for stationary vector $\mathbf{s}_{\text{vec}}$ is given by [13]

$$R_{MN}(\theta) = \sum_{i=1}^{MN} \max\left(0, \frac{1}{2}\log\frac{\lambda_i(\mathbf{\Sigma}_{\mathbf{s}})}{\theta}\right)$$

$$D_{MN}(\theta) = \sum_{i=1}^{MN} \min(\theta, \lambda_i(\mathbf{\Sigma}_{\mathbf{s}})), \tag{7}$$



(a) *Link 1*: from the left HA to the right HA.

(b) *Link 2*: from the right HA to the left HA.

Fig. 4. Proposed asymmetric 2-way coding scheme.

where $\lambda_i(\mathbf{\Sigma})$ is the ith eigenvalue of a matrix $\mathbf{\Sigma}$. From Szego's theorem [25], the asymptotic eigenvalue distribution of Toeplitz matrices corresponds to those of the PSD values in the frequency domain. The extensions to the Szego's theorem are proposed in [26] and [27] which state that for any Hermitian block-Toeplitz matrix (here $\mathbf{\Sigma}_{\mathbf{s}}$) the asymptotic behavior of an arbitrary function of eigenvalues follows that of corresponding CPSD matrices in the frequency domain, i.e.,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{MN}F(\lambda_i(\mathbf{\Sigma}_{\mathbf{s}})) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}F(\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega)))d\Omega, \tag{8}$$

where $\mathbf{\Phi}_{\mathbf{s}}$ is the CPSD matrix with respect to the vector sequence $\{\mathbf{s}[n]\}_{n=0}^{N-1}$ with elements $[\mathbf{\Phi}_{\mathbf{s}}]_{uv} = \sum_{k=-\infty}^{\infty}[\mathbf{\Sigma}_k]_{u,v}e^{-j\Omega k}$. $F(\cdot)$ is an arbitrary function applied on the eigenvalues. Based on (8), asymptotically as $N\to\infty$, the RDF in (7) can be rewritten as

$$\begin{aligned} R^*(\theta) &= \lim_{N\to\infty}\frac{1}{N}R_{MN} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}\max\left(0, \frac{1}{2}\log\frac{\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))}{\theta}\right)d\Omega \\ D^*(\theta) &= \lim_{N\to\infty}\frac{1}{N}D_{MN} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}\min(\theta, \lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega)))d\Omega, \end{aligned} \tag{9}$$

where the rate $R^*(\theta)$ is per source vector $\mathbf{s}[n]\in\mathbb{R}^{M\times 1}$. The entries of the vector $\mathbf{s}[n]$ can be thought of as spatial samples captured by M microphones. Microphone samples usually have different powers. In other words, matrices $\mathbf{\Sigma}_i$, which can be thought of as the cross correlation matrix across microphone samples, are not necessarily Toeplitz. Therefore $\mathbf{\Phi}_{\mathbf{s}}$ will not be spatially diagonalizable by the spatial Fourier transform, not even asymptotically, as the number of the microphones increases ($M\to\infty$) and the Karhunen-Love transform (KLT) matrix is in need for optimal coding in (9).

Note that the distortion $D^*(\theta)$ is the MSE between the vector source $\mathbf{s}[n]$ and its quantized version, say $\tilde{\mathbf{s}}[n]$. We use this RDF for vector source quantization (the dashed box in Fig. 4(a)). In *Link 1*, we are interested in estimating S_2. Therefore, with appropriate translation of the RDF in (9), the following procedure is described based on Fig. 4(a).

First, the observations $\mathbf{y}_1$ are spatially decorrelated as

$$\mathbf{z}_1 = \mathbf{A}\mathbf{y}_1,\ \mathbf{A}\in\mathbb{C}^{M_1\times M_1} \tag{10}$$

where $\mathbf{A}$ is a matrix whose rows are the eigenvectors of the CPSD matrix $\mathbf{\Phi}_{\mathbf{y}_1}$. The CPSD matrix $\mathbf{\Phi}_{\mathbf{z}_1}$ of the decorrelated vector $\mathbf{z}_1$ is diagonal with diagonal elements $[\mathbf{\Phi}_{\mathbf{z}_1}]_{uu} = \lambda_u(\mathbf{\Phi}_{\mathbf{y}_1})$, where $\lambda(\cdot)$ is the eigenvalue operator. Second, $\mathbf{z}_1$ is quantized, achieving the RDF presented in (9) by replacing $\mathbf{\Phi}_{\mathbf{s}}$ in (9) with $\mathbf{\Phi}_{\mathbf{y}_1}$. It can be shown that the following quantization model is obtained

$$\tilde{\mathbf{z}}_1 = \mathbf{B}\mathbf{z}_1 + \mathbf{e}_1, \tag{11}$$

where $\tilde{\mathbf{z}}_1$ is interpreted as a transformed-quantized left-side noisy observation. As the elements of $\mathbf{z}_1$ are uncorrelated, the vector quantizer in (9) can be interpreted as M_1 test channels corresponding to Fig. 1. The Vector $\mathbf{e}_1$ can be thought of as M_1 test channel noises. Therefore matrices $\mathbf{B}, \mathbf{\Phi}_{\mathbf{e}_1}\in\mathbb{R}^{M_1\times M_1}$ will be diagonal and the diagonal elements are computed based on (3), replacing $\Phi_s(\Omega)$ by $[\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)$, respectively. Applying the inverse-decorrelation matrix $\mathbf{A}^{-1}$ to reproduce quantized left-side observations, we obtain

$$\tilde{\mathbf{y}}_1 = \mathbf{A}^{-1}(\mathbf{B}\mathbf{z}_1 + \mathbf{e}_1) = \mathbf{C}\mathbf{y}_1 + \mathbf{A}^{-1}\mathbf{e}_1, \tag{12}$$

where $\mathbf{C} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}$. Finally, the estimator f_2 estimates the target signal S_2 as $\hat{S}_2 = f_2(\mathbf{y}_2, \tilde{\mathbf{y}}_1) = E[S_2|\mathbf{y}_2, \tilde{\mathbf{y}}_1]$. The direct distortion per frequency between $\mathbf{y}_1$ and its quantized version $\tilde{\mathbf{y}}_1$ is given by $\mathrm{tr}\{\mathbf{\Phi}_{\mathbf{d}_1}(\Omega)\}$, where $\mathbf{\Phi}_{\mathbf{d}_1}$ is the CPSD matrix of the direct error process $\mathbf{d}_1 = \mathbf{y}_1 - \tilde{\mathbf{y}}_1$, and $\mathrm{tr}\{\cdot\}$ denotes the trace operator on a matrix. We have

$$\mathrm{tr}\{\mathbf{\Phi}_{\mathbf{d}_1}(\Omega)\} = \sum_{u=1}^{M}\min(\theta, [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)). \tag{13}$$

As $[\mathbf{\Phi}_{\mathbf{z}_1}(\Omega)]_{uu} = \lambda_u(\mathbf{\Phi}_{\mathbf{y}_1}(\Omega))$, the overall MSE over all frequencies corresponds to the distortion in (9) with $\lambda_u(\mathbf{\Phi}_{\mathbf{s}})$ replaced by $\lambda_u(\mathbf{\Phi}_{\mathbf{y}_1}) = [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}$. The final (remote) distortion corresponds to (5) with $\tilde{\mathbf{y}} = [\mathbf{y}_2^T\tilde{\mathbf{y}}_1^T]^T$.

2) (Joint) Statistics Estimation: Based on the quantized left-side signal model in (12), the second order statistics in the frequency domain can be written as

$$\begin{aligned} \mathbf{\Phi}_{\tilde{\mathbf{y}}_1} &= \mathbf{C}\mathbf{\Phi}_{\mathbf{y}_1}\mathbf{C}^H + \mathbf{A}^{-1}\mathbf{\Phi}_{\mathbf{e}_1}\mathbf{A}^{-H} \\ \mathbf{\Phi}_{\mathbf{y}_1} &= \mathbf{C}^{-1}\left(\mathbf{\Phi}_{\tilde{\mathbf{y}}_1} - \mathbf{A}^{-1}\mathbf{\Phi}_{\mathbf{e}_1}\mathbf{A}^{-H}\right)\mathbf{C}^{-H}. \end{aligned} \tag{14}$$

To retrieve $\mathbf{\Phi}_{\mathbf{y}_1}$ we need to know $\mathbf{B}, \mathbf{A}$, and $\mathbf{\Phi}_{\mathbf{e}_1}$. The elements of the diagonal matrix $\mathbf{B}$ depend on the reverse water-filling parameter θ and the diagonal elements of $\mathbf{\Phi}_{\mathbf{z}_1}$. The scalar value θ is chosen by fixing the transmission bit-rate or the distortion. Moreover, using the backward test channel interpretation in [13], [23], for $\theta < [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)$ we have

$$[\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega) = [\mathbf{\Phi}_{\tilde{\mathbf{z}}_1}]_{uu}(\Omega) + \theta. \tag{15}$$

Equation (15) shows that at high rates (small θ) it is possible to retrieve the unquantized statistics of the transformed signal $\mathbf{z}_1$ from the quantized signal $\tilde{\mathbf{z}}_1$. Using the property in (15), $\mathbf{\Phi}_{\mathbf{z}_1}$ can be retrieved, estimating the quantized PSDs $[\mathbf{\Phi}_{\tilde{\mathbf{z}}_1}]_{uu}$ given the quantized realizations in the frequency domain, for a sufficiently small θ. Therefore, the matrix $\mathbf{B}$ can be computed at the decoder. Following a similar procedure, the diagonal matrix $\mathbf{\Phi}_{\mathbf{e}_1} = E[\mathbf{e}_1\mathbf{e}_1^H]$ can be computed. Computing $\mathbf{A}$ requires the data dependent KLT. As we often do not know this, we test in Section V the algorithm, next to the true $\mathbf{A}$ based on the KLT, also with an $\mathbf{A}$ based on the fixed discrete cosine transform (DCT). Finally, computing $\mathbf{B}$ and fixing $\mathbf{A}$, the matrix $\mathbf{C}$ is known and the local statistics $\mathbf{\Phi}_{\mathbf{y}_1} = E[\mathbf{y}_1\mathbf{y}_1^H]$ are retrieved at the decoder. Moreover, the joint statistics between the two side observations are retrieved as

$$E\left[\mathbf{y}_1\mathbf{y}_2^H\right] = \mathbf{C}^{-1}E\left[\tilde{\mathbf{y}}_1\mathbf{y}_2^H\right]. \tag{16}$$

Note that the statistics (joint or local) at a certain frequency Ω can be retrieved only if the matrix $\mathbf{C}(\Omega)$ is invertible and the PSD of the source at that frequency is positive. This implies that $\mathbf{B}(\Omega)$ should be invertible, as $\mathbf{A}(\Omega)$ is an orthogonal and invertible matrix. Since $\mathbf{B}(\Omega)$ is a diagonal matrix, all elements should be positive, implying that $\theta < \min_u [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)$, for a particular frequency. For $\mathbf{B}$ to be invertible in all frequencies, the condition is rewritten as $\theta < \min_\Omega \min_u [\mathbf{\Phi}_{\mathbf{z}_1}]_{uu}(\Omega)$. This condition is satisfied only at sufficiently high rates. In fact, at lower rates, the reverse-water filling algorithm tries to allocate more bit-rate to the frequency components with greater PSD values and zero bit rate to those with smaller (than the threshold θ) PSD values. In this case $\mathbf{B}$ becomes singular and, as a result, smaller PSDs cannot be retrieved at the decoder. In the next part of this section, another quantization method is proposed to address this limitation and guarantee the invertible $\mathbf{B}$ matrix for all frequencies.

3) Method 2. PDF Preserving Source Coding for Sources With Memory: Reverse-water filling for vector sources (*Method 1*) cannot guarantee positive bit-rates for strictly positive PSDs. To keep all frequency components of the signal active after the quantization, a constrained source coding approach was proposed in [28]. This method imposes an extra constraint to the original lossy source coding problem such that the probability distribution of the signal is preserved after the quantization process. The distribution preserving RDF (DP-RDF) is given in [22, Proposition 1] for a time-stationary Gaussian process, which can be thought of as a single microphone observation. As there are multiple microphones per HA, we extend this result to multiple observations and find the DP-RDF for vector sources with memory. Moreover, we propose a conceptual test channel interpretation to achieve such a rate-distortion tradeoff.

Proposition 1: The DP-RDF for a discrete-time sequence of zero-mean time-stationary vector Gaussian sources, say $\{\mathbf{s}[n]\}_{n=0}^{N-1}$, where $\mathbf{s}[\cdot] \in \mathbb{R}^{M\times 1}$ with the corresponding block-Toeplitz cross correlation matrix $\mathbf{\Sigma}_{\mathbf{s}}$, and the Hermitian CPSD matrix $\mathbf{\Phi}_{\mathbf{s}}$ is given by

$$R^{DP}(\mu) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}\log_2\frac{\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))}{\left(\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))D_u(\mu,\Omega) - \frac{D_u^2(\mu,\Omega)}{4}\right)^{\frac{1}{2}}}d\Omega$$

$$D^{DP}(\mu) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{u=1}^{M}D_u(\mu,\Omega)d\Omega, \tag{17}$$

where

$$D_u(\mu,\Omega) = 2\,\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega)) + \mu - (4\,\lambda_u^2(\mathbf{\Phi}_{\mathbf{s}}(\Omega)) + \mu^2)^{\frac{1}{2}}, \tag{18}$$

for $D_u(\mu,\Omega) < 2\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))$. The variable μ denotes a Lagrange parameter [22], which relates the rate to the distortion and satisfies the constraint on the distortion. Similar to "reverse water-filling" problems [13] with parameter θ, μ can be found by either fixing the total rate R^{DP} or the total distortion D^{DP}. The rate is per vector source $\mathbf{s}[\cdot] \in \mathbb{R}^{M\times 1}$. The distortion $D^{DP}(\mu)$ is the averaged MSE between the vector source $\mathbf{s}[\cdot]$ and the quantized source $\hat{\mathbf{s}}[\cdot]$. See Appendix A for derivations.

Equation (17) represents the proposed extension to the DP-RDF for vector sources. Note that (17) is only valid for strictly positive CPSD eigenvalues ($\lambda_u(\mathbf{\Phi}(\Omega)) > 0$). For $\lambda_u(\mathbf{\Phi}(\Omega)) = 0$ the rate allocated to such frequency component will be zero. Unlike *Method 1* (reverse water filling), here all frequency components with strictly positive CPSD eigenvalues are allocated with positive rates. Therefore, the PSDs can be retrieved from the quantized signal vector at the decoder. Comparing the direct distortions in (17) and (9), the gap in distortions is 3 dB at zero bit-rate and it vanishes asymptotically [28]. However, as we are interested in estimating the source S_2, it is not clear if *Method 1* is better than *Method 2* with respect to the final distortion, as the joint statistics are not available at the encoder in both methods.

Appendix B shows that the conceptual test channel interpretation, shown in the dashed box in Fig. 4(a), achieves the DP-RDF in (17), but with different quantization parameters from those in *Method 1*. First $\mathbf{y}_1$, is spatially decorrelated ($\mathbf{z}_1 = \mathbf{A}\mathbf{y}_1$). Then the decorrelated signals are quantized using the proposed distribution preserving quantization, $\tilde{\mathbf{z}}_1 = \mathbf{B}\mathbf{z}_1 + \mathbf{e}_1$ with quantization parameters given as (see Appendix B):

$$\mathbf{B}(\Omega) = \mathrm{diag}\{\beta_1(\Omega),\ldots,\beta_{M_1}(\Omega)\} = \mathrm{diag}\left\{1 - \frac{D_1(\mu,\Omega)}{2\,\lambda_1(\mathbf{\Phi}_{\mathbf{y}_1}(\Omega))},\ldots,1 - \frac{D_M(\mu,\Omega)}{2\,\lambda_M(\mathbf{\Phi}_{\mathbf{y}_{M_1}}(\Omega))}\right\}$$

$$\mathbf{\Phi}_{\mathbf{e}_1}(\Omega) = \mathrm{diag}\left\{\frac{\beta_1(\Omega)+1}{2}D_1(\mu,\Omega),\ldots,\frac{\beta_M(\Omega)+1}{2}D_M(\mu,\Omega)\right\}. \tag{19}$$

The quantized signal $\tilde{\mathbf{y}}_1$ is computed by applying the inverse transform matrix $\mathbf{A}^{-1}$ as $\tilde{\mathbf{y}}_1 = \mathbf{A}^{-1}\tilde{\mathbf{z}}_1$. Finally the target signal is estimated as $\hat{S}_2 = E[S_2|\tilde{\mathbf{y}}_1, \mathbf{y}_2]$.

The procedure to retrieve (joint) statistics is as follows. As the second order statistics of $\mathbf{y}_1$ are preserved, $\mathbf{\Phi}_{\mathbf{y}_1} = \mathbf{\Phi}_{\tilde{\mathbf{y}}_1}$ holds. $\mathbf{\Phi}_{\tilde{\mathbf{y}}_1}$ can be estimated using realizations in the frequency domain. By informing the decoder of the scalar parameter μ, the invertible matrices $\mathbf{B}$ and $\mathbf{\Phi}_{\mathbf{e}_1}$ in (19) are known at the decoder. Knowing $\mathbf{B}$ and fixing $\mathbf{A}$, $\mathbf{C}$ is known. Therefore, based on (16) the joint statistics are retrieved.

B. Link 2: From Right-to-Left

The goal in *Link 2* is to transmit the filtered-quantized right-side observations in order to estimate the target signal at the left-side reference microphone S_1. As the lossy (quantized) version of the left-side observation ($\tilde{\mathbf{y}}_1$) is available at the right-side, it acts as a (lossy) side information at the right-side encoder. We use this information to reduce the redundancy in the transmission of the information. This is done by the proposed coding architecture, illustrated in Fig. 4(b). As shown, different coding algorithms can be obtained by changing the switch l and by using different methods in *Link 1*. Note that if for example, the switch l is open, i.e., $l = 0$, the realization of $\mathbf{y}_1$ will not be used at the decoder. The side information at the left-side decoder is $\mathbf{y}_1$. We describe some possible scenarios.

1) Case a. Coding With Quantized Statistics and With $l = 0$: In this case we assume the *Method 1* is chosen in *Link 1*. The idea is to pre-filter the right-side observations $\mathbf{y}_2$ using quantized statistics retrieved from *Link 1* and directly quantize and transmit them to the other side. The sub-optimal filter coefficients $\mathbf{w}_2^a$ (compared to the optimal filter in (4)) are computed as $\mathbf{w}_2^a = \mathbf{\Phi}_{\bar{\mathbf{y}}_2^a}^{-1}\mathbf{\Phi}_{\bar{\mathbf{y}}_2^a S_1}$, where $\mathbf{\Phi}_{\bar{\mathbf{y}}_2^a}$ is the CPSD matrix of the innovation process $\bar{\mathbf{y}}_2^a = \mathbf{y}_2 - E[\mathbf{y}_2|\tilde{\mathbf{y}}_1]$. The filter coefficients are computed in a similar fashion to the one in optimal RCNR approach, described in Section III-A, except that here only the lossy side information $\tilde{\mathbf{y}}_1$ is available and not the lossless $\mathbf{y}_1$. In this way, we try to reduce some information redundancy in estimating S_1 of $\mathbf{y}_2$ given $\tilde{\mathbf{y}}_1$.

The filtered scalar signal $Y_{21}^a = (\mathbf{w}_2^a)^{\mathrm{H}}\mathbf{y}_2$ is encoded for a decoder that has no access to the side information $\mathbf{y}_1$ (the switch l is open). This means that Y_{21}^a is directly (blindly) quantized, i.e., $\tilde{Y}_{21}^a = \beta_2^a Y_{21}^a + E_2^a$. The quantization parameters correspond to (3) with replacing Φ_s with $\Phi_{Y_{21}^a} = (\mathbf{w}_2^a)^{\mathrm{H}}\mathbf{\Phi}_{\mathbf{y}_2}\mathbf{w}_2^a$. In fact, $\tilde{\mathbf{y}}_1$ is used for estimating the filter coefficients $\mathbf{w}_2^a$, but not used in the coding process. Finally the MWF filter is applied to the total observations $\tilde{\mathbf{y}} = [\mathbf{y}_1^T \tilde{Y}_{21}^a]^T$ and the target signal S_1 is estimated as $\hat{S}_1 = f_1(\tilde{\mathbf{y}}) = E[S_1|\tilde{\mathbf{y}}]$.

2) Case b. Coding With Unquantized Statistics and With $l = 0$: In the previous case, we estimated the (joint) statistics based on the lossy side information $\tilde{\mathbf{y}}_1$ (*Method 1*). Therefore, the filtering coefficients are not estimated in an optimal manner as some frequency components are truncated, especially at lower rates. To estimate the optimal filter we do not need the actual realizations of the lossless side information $\mathbf{y}_1$. Instead, we only need to estimate the (joint) unquantized statistics from $\tilde{\mathbf{y}}_1$. To do so, *Method 2* is chosen in *Link 1*. The use of DP quantization enables retrieving the unquantized statistics $E[\mathbf{y}_1\mathbf{y}_2^H]$ and $E[\mathbf{y}_1\mathbf{y}_1^H]$ at all frequencies. The procedure to estimate the statistics is described in the Section IV-A2. The optimal right-side filter coefficients $\mathbf{w}_2^o$ are computed, similar to (4), as $\mathbf{w}_2^o = \mathbf{\Phi}_{\bar{\mathbf{y}}_2}^{-1}\mathbf{\Phi}_{\bar{\mathbf{y}}_2 S_1}$, where $\mathbf{\Phi}_{\bar{\mathbf{y}}_2} = \mathbf{\Phi}_{\mathbf{y}_2} - \mathbf{\Phi}_{\mathbf{y}_2\mathbf{y}_1}\mathbf{\Phi}_{\mathbf{y}_1}^{-1}\mathbf{\Phi}_{\mathbf{y}_1\mathbf{y}_2}$ is the CPSD matrix of the innovation process $\bar{\mathbf{y}}_2 = \mathbf{y}_2 - E[\mathbf{y}_2|\mathbf{y}_1]$. The direct quantization and final estimation stages resembles those of case a with a different filtered signal $Y_{21} = (\mathbf{w}_2^o)^{\mathrm{H}}\mathbf{y}_2$.

3) Case c. Optimal Coding with Unquantized Statistics and With $l = 1$: Like the Case b, in this case, again we use *Method 2* in *Link 1* since we want to preserve the statistics to compute optimal filter for *Link 2*. Following the optimal RCNR (when the switch l is closed) the right-side processor encodes the filtered signal $Y_{21} = (\mathbf{w}_2^o)^{\mathrm{H}}\mathbf{y}_2$ for a decoder which has access to the side information $\mathbf{y}_1$. Therefore unlike case a and case b, here the quantization stage is a side information informed process (remote Wyner-Ziv quantizer). The filtered-quantized signal is given by $\tilde{Y}_{21} = \beta_2^o Y_{21} + E_2^o$. The optimal (remote Wyner-Ziv) quantization parameters β_2^o and $\Phi_{E_2^o}$ correspond to (3) with Φ_s replaced by $(\mathbf{w}_2^o)^{\mathrm{H}}\mathbf{\Phi}_{\bar{\mathbf{y}}_2}\mathbf{w}_2^o$. It is important to note that here the quantization scaling factor β_2^o is now a function of the side information $\mathbf{y}_1$ (only a function of the statistics, not the realizations). This necessitates some extra information to be available at the decoder in order to decode indices which are computed knowing the fact that $\mathbf{y}_1$ would be available at the decoder. This extra information includes the joint entropy between $\mathbf{y}_1$ and $\tilde{Y}_{21}$. Knowing this information at the decoder, we can touch the performance bound of the optimal RCNR at least in one link (*Link 2*) in our proposed 2-way communication system. The final conditional mean estimator, which was included in the decoder box in the optimal RCNR architecture in Fig. 2, resembles that of the Case a, except with the different filtered-quantized signal $\tilde{Y}_{21}$.

V. Performance Evaluation

In this section, we compare the performance of the approaches, described in the previous sections, as a function of transmission bit-rate. We evaluate the methods based on two performance measures. The first performance measure presented in [12] and [18], is defined as the ratio of the MSE when there is no communication between the HAs to the one when the data is quantized before transmission. The output gains with respect to the two beamformers are given by

$$G_1(R_1) = \frac{D_1(0)}{D_1(R_1)}, \quad G_2(R_2) = \frac{D_2(0)}{D_2(R_2)}, \tag{20}$$

where $D_i(\cdot), i = 1, 2$ are defined in (5) but with different outputs $\hat{S}_i$ for different approaches. Note that the final outputs $\hat{S}_i$, $i = 1, 2$ are functions of the corresponding bit-rates as the data is quantized. For example, $D_1(R_1)$ denotes the MSE between the target source at the left-side reference microphone S_1 and its estimate $\hat{S}_1$ when the data is quantized and transmitted at R_1 bit-rate from the right-side HA to the left one (*Link 2*). $D_1(0)$ denotes the left-side MSE when there is no communication between HAs. i.e., $R_1 = 0$.

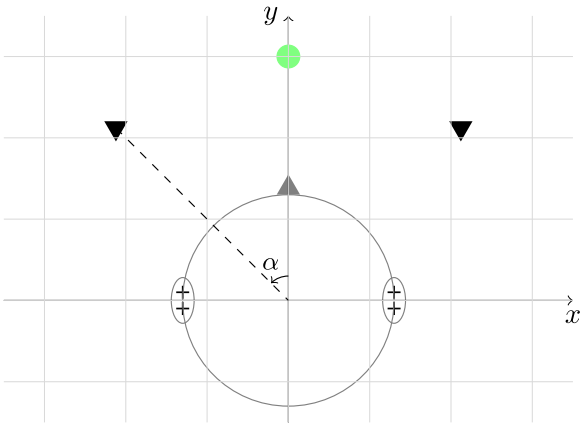


Fig. 5. Typical acoustic scene. The Target signal, interferers, and microphones are denoted by the Green circle, the black triangles, and the black "+" symbols, respectively.

Another performance measure which we refer to as "binaural gain" is proposed in [12] and is defined as the ratio of the sum of the MSEs with respect to the two HA reference microphones, when there is no communication between HAs to the one when the data is quantized and transmitted in both links at certain bit-rates, i.e.,

$$G_B(R_T) = \frac{D_1(0) + D_2(0)}{D_1(R_1) + D_2(R_2)}, \tag{21}$$

where $R_T = R_1 + R_2$ is the total rate budget for two links. The performance of the following approaches are compared throughout this section

- **B-MWF**: The full binaural MWF from [7], without quantization.
- **OPT**: Optimal approach from [12, Section III-A] (Section III-A)
- **SIG**: Sub-optimal approach from [12]: An estimate of the target signal is transmitted to the contralateral processor (Section III-B)
- **INT**: Sub-optimal approach from [18]: An estimate of the undesired (interfering) signal is transmitted to the contralateral processor (Section III-B)
- **RAW**: Sub-optimal approach from [18]: A raw reference microphone signal is transmitted to the contralateral processor, without any pre-filtering (Section III-B)
- **M1**: *Method 1* in *Link 1* (Section IV-A1)
- **M2**: *Method 2* in *Link 1* (Section IV-A3)
- **L2a**: Proposed sequential sub-optimal approach for *Link 2* using *Method 1* in *Link 1* (Case a, Section IV-B1)
- **L2b**: Proposed sequential sub-optimal approach for *Link 2* using *Method 2* in *Link 1* (Case b, Section IV-B2)
- **L2c**: Proposed sequential optimal approach for *Link 2* using *Method 2* in *Link 1* (Case c, Section IV-B3)

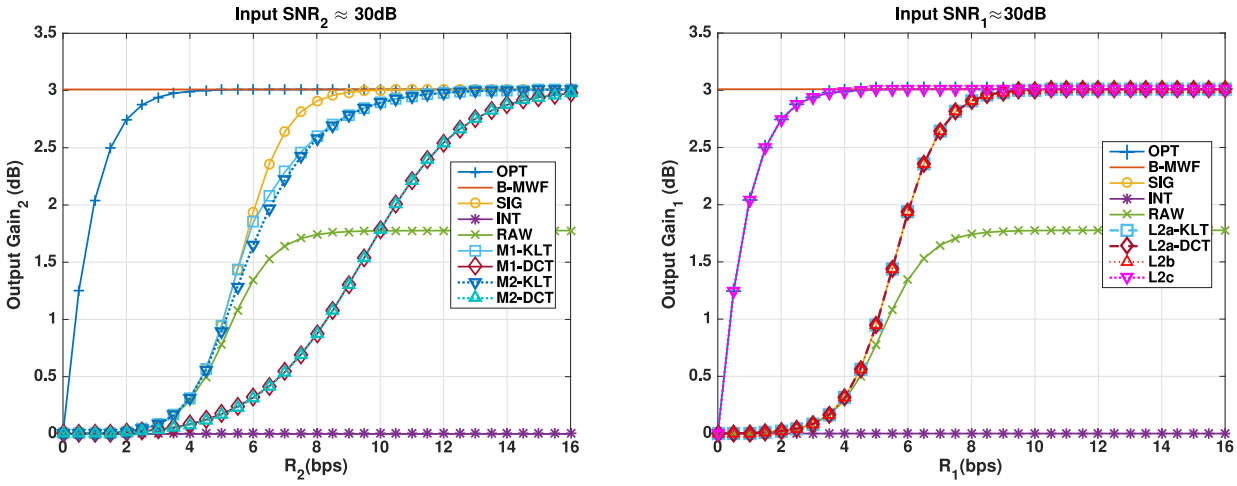
The acoustic scene used for the experiments is illustrated in Fig. 5. The four black "+" symbols indicate the microphones mounted on the virtual head. The planar distance between the two microphones per HA is 0.76 cm. The radius of the virtual head is set to 8.2 cm [29]. The Green circle denotes the desired (target) speech signal which is assumed to be fixed 0.8 m from the origin in front of the head for all experiments. The black triangles denote the interferers. The number and the position of the interferers vary in different experiments. Interferers are located randomly at different angles, say $\alpha = \tan^{-1}(\frac{y}{x}) - \frac{\pi}{2}$ and different distances from the origin ($(x, y) = (0, 0)$), say $r = \sqrt{x^2 + y^2}$. In this paper zero degrees corresponds to the direction straight ahead of the HA user and the angles are computed counterclockwise. All point noise sources have flat PSDs $\Phi_{(.)}(\Omega)$ over the interval $\Omega \in [-\pi F_s, \pi F_s]$ where $F_s = 16$ kHz. ATFs in (1) are found via head-related transfer functions (HRTF)s from the database in [29]. The PSD of the target speech signal is estimated based on the Welch's method using 12.5 seconds of the recorded speech at 16 kHz sampling frequency from the "CMU-ARCTIC" [30] database, without considering voice activity detection (VAD) errors. We used 512-point frames with 50% overlap and discrete Fourier transform (DFT) size of 1024 for the PSD estimation process.

A. Uncorrelated Noise

In this scenario, per microphone, the target signal is degraded by additive white Gaussian noise (AWGN), uncorrelated to the signal as well as across microphones, having the same variance among all microphones. Note that there is no point noise source (interferer) here. The uncorrelated noise power is set such that the input signal to noise ratio (SNR) at the corresponding left-side and right-side reference microphones, say SNR_1 and SNR_2, respectively, be approximately 30 dB. Based on the performance measures in (20), the output gains (in dB) in terms of bit per sample (bps) are shown in Fig. 6 for *Link 1* and *Link 2*, for the above-mentioned approaches. By M1-DCT we mean that the method M1 uses the fixed discrete cosine transform (DCT) matrix for the matrix $\mathbf{A}$ in (10), when spatially decorrelating the signals in the frequency domain. The similar explanation holds for M1-KLT, M2-KLT, and M2-DCT. Note that the method L2a-KLT in *Link 2* is sequentially related to the *Method 1* (M1-KLT) in *Link 1* as it uses the signal statistics quantized by *Method 1*. The similar relation holds between L2a-DCT and M1-DCT. L2b and L2c are related to the *Method 2* in *Link 1*. Their performances remain the same using KLT or DCT matrices in *Method 2*, as they use the retrieved (unquantized) statistics. These explanations hold also for other experiments in this section.

Based on Fig. 6 we can make the following observations

- Method SIG is asymptotically optimal as argued in [18] in the presence of uncorrelated noise only. As the noise components at the two sides are independent, no necessary information will be removed by estimating the desired part of the signal and sending it to the other side.
- Method INT has no gain compared the monaural setup. With a similar argument, estimating the noise on one side has no added information for the other side, resulting in no increase in the performance.
- M1 and M2 outperform methods RAW and INT since they are not even asymptotically optimal. In the presence of only uncorrelated noise, any extra observation (extra microphone signal) can help to increase the performance.



(a) *Link 1*: from left to right. The gain is computed with respect to the right-side beamformer.

(b) *Link 2*: from right to left. The gain is computed with respect to the left-side beamformer.

Fig. 6. Output gains in the presence of uncorrelated noise only.

Method RAW in [18] chooses only one microphone signal (out of two), which degrades the performance.

- L2a and L2b have almost the same performance as the SIG method since almost no redundancy in information is remained after locally estimating the target signal itself.
- L2c is an optimal coding scheme in *Link 2*. It takes the correlation between the filtered observation and the side information into account and encodes the filtered signals knowing the fact that the decoder can revive the correlated information which is reduced during the encoding process. This approach assumes the joint entropy between the filtered-encoded signals and the side information is available at the decoder.

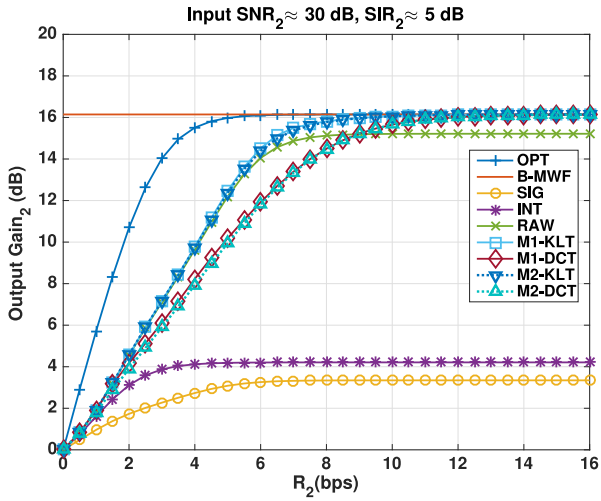
B. Correlated and Uncorrelated Noise

Two scenarios are considered in this section. First, one point noise source is added to the previous scenario at 30° and 0.8 m from the origin. The interfering signal power is set such that the input signal-to-interferer ratios (SIRs) with respect to the left and right reference microphones are approximately $\text{SIR}_1 \approx 0$ dB, and $\text{SIR}_2 \approx 5$ dB, respectively.

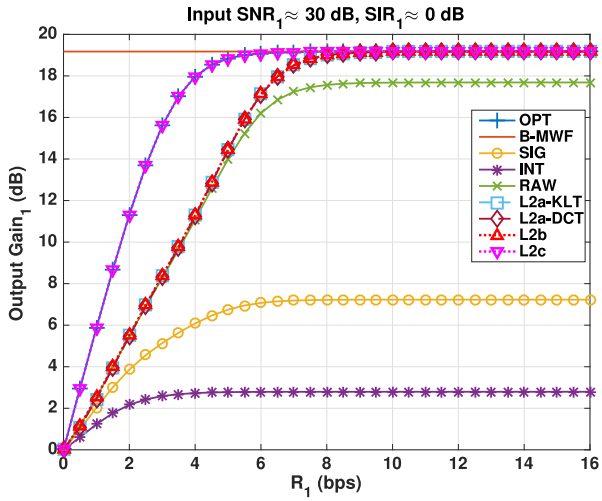
In the second scenario, four interferers are added at degrees $[-50°, -30°, 30°, 70°]$. The input SIRs at the corresponding reference microphones are $\text{SIR}_1 \approx 0$ dB and $\text{SIR}_2 \approx 0$ dB. The simulation results are shown in Fig. 7.

Based on Fig. 7 we can make the following observations

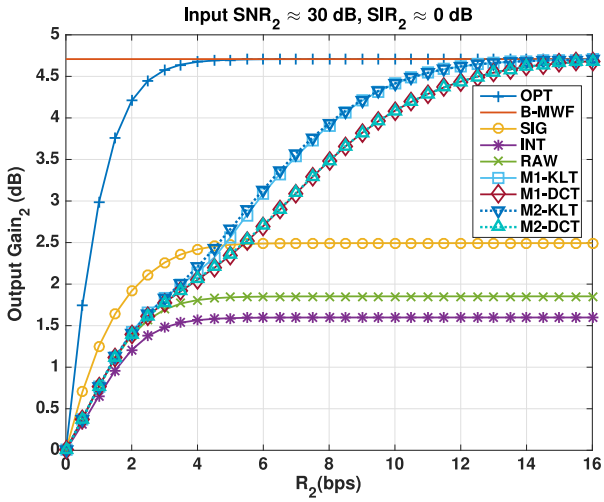
- As shown in Fig. 7(a) and 7(b), in the presence of one spatially correlated point noise source, the SIG method is not asymptotically optimal anymore. Some necessary (spatial) information about the interferer will be eliminated after the filtering stage before transmission. This information would be helpful for the left-side processor to cancel out the interferer [18]. In general, the loss in performance at high rates is significant.
- In *Link 1* for highly correlated signals (one interferer), the methods M1-KLT and M2-KLT outperform all other approaches (when using the optimal KLT matrix). Using the DCT matrix also results in a good performance, especially at high rates.
- L2a uses the quantized left-side signals from *Link 1* (*Method 1*) to reduce the redundancy in information transmission, and hence, outperforms almost all existing suboptimal approaches, especially at high rates. However, the use of quantized statistics results in a non-optimal filter in *Link 2* and degrades the performance, especially in the four interferers scenario, in comparison with the L2b and the L2c methods.
- L2b outperforms L2a, especially in the four interferers scenario. The use of *Method 2* in *Link 1* enables retrieval of the unquantized statistics (rather than quantized statistics), which helps to compute the optimal filter on the right side, and hence, results in a better estimation of the informative signal for the left-side beamformer. However, the filtered signal in both the L2a and the L2b methods is still correlated to the (left) side information $\mathbf{y}_1$ and the direct (blind) quantization of such a filtered signal does not take this correlation into account. As a result, the performance of both cases a and b are always worse than that of the case c. L2c touches the optimal performance as the side-information-aware quantization is used.
- The gap between M1 and the distribution preserving vector quantization method M2 is negligible in most scenarios and bit-rates. As the target estimation error is of interest (and not the direct estimation error between the noisy vector source $\mathbf{y}_i, i = 1, 2$ and its corresponding quantized version $\tilde{\mathbf{y}}_i$), it is not clear if the M2 performance should always be worse than that of M1. In fact, as the blind quantization process in *Link 1* does not use the joint information, both quantizers are not aware that which frequency components are more important and not predictable on the other side. However, M2 preserves the statistics of the contralateral observations which include spatial cues of sources and



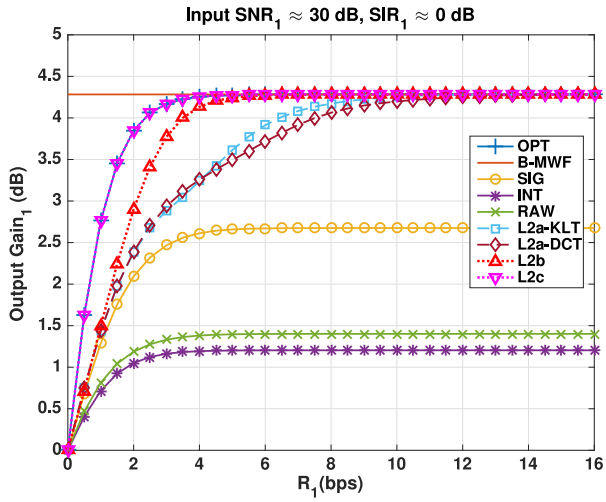
(a) One interferer: *Link 1*, from left to right.



(b) One interferer: *Link 2*, from right to left.



(c) Four interferers: *Link 1*, from left to right.



(d) Four interferers: *Link 2*, from right to left.

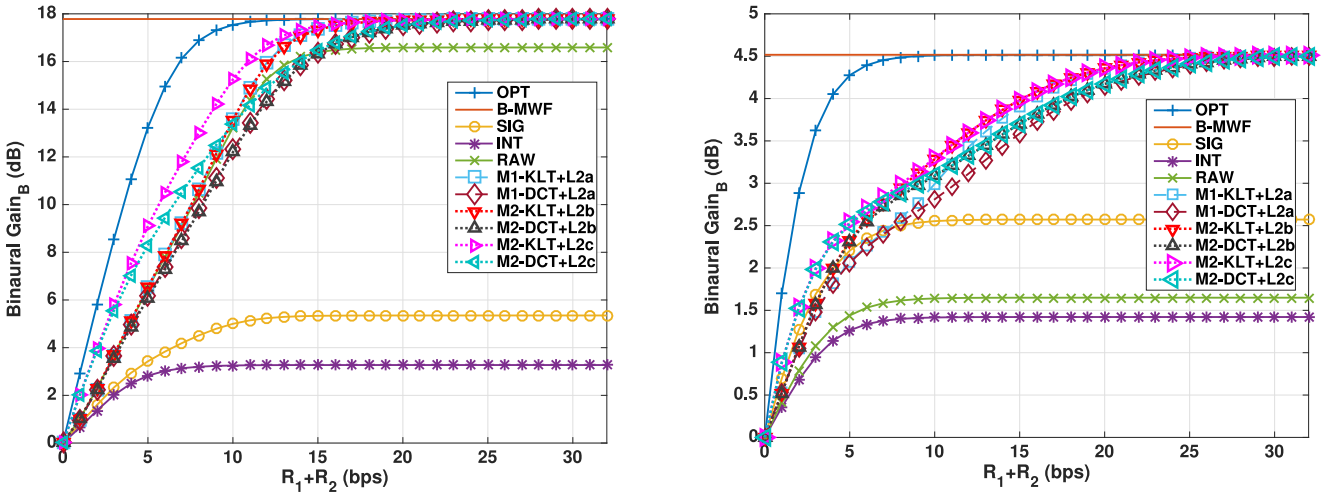
Fig. 7. Output gains for observations with correlated noise sources.

perceptually may help to get a more natural impression of the sound field.

- For all MWF-based methods, the performance is a function of the correlation between the observations. Therefore, it is clear that the performance will change by changing the source position, as the correlation between the observations will change. However, this does not affect the generality of the proposed method. as well as that of the optimal method. The proposed method tries to estimate the joint statistics, without any assumption on the source positions, and use it in another link to reduce the redundancy in information transmission.

C. Binaural Gain

In this section, we evaluate the methods based on the binaural gain measure in (21). We compute $G_B(R_T)$ for scenarios with correlated interferers, introduced in Section V-B. In fact, the same distortions as those computed for the results in Fig. 7 are used for computing $G_B(R_T)$. Fig. 8 shows the binaural gain in terms of the total bit-rate budget $R_T = R_1 + R_2$, for the scenario in which there is one interferer (left: Fig. 8(a)), and the scenario in which there are four interferers (right: Fig. 8(b)), along with the target signal and the uncorrelated noise. For example, when computing the binaural gain of "M2-KLT+L2c", the MSE of M2-KLT in *Link 1* is added to the MSE for the method L2c. The sequential-asymmetric method M2-KLT+L2c has the best performance among all other methods as it touches the optimal performance, at least in *Link 2*. However considering the loss in the performance of M2-KLT in *Link 1*, the binaural performance of M2-KLT+L2c is worse than that of the optimal performance [12, Section III-A] (which is optimal in both links). All proposed methods resolve the asymptotic sub-optimality issue of the existing sub-optimal methods and outperform them, especially at middle rates and high rates. The sequential nature of the proposed methods enables a smart use of the information at hand to reduce the bandwidth.



(a) One interferer + uncorrelated Noise.

(b) Four Interferers + Uncorrelated Noise.

Fig. 8. Binaural output gains for observations with correlated noise sources.

VI. Conclusion

In this paper, we studied the performance of the optimal/sub-optimal binaural rate-constrained noise reduction (beamforming) approaches based on the unified framework which can be interpreted as filtering, quantization, and final estimation stages. Moreover, we proposed a two-way asymmetric coding scheme which retrieves the statistics between two HA observations from quantized signals in one link (*Link 1*) to be used in another link (*Link 2*) and addresses two main limitations of existing methods. The first limitation is the strict requirement of the complete knowledge of the joint statistics in the optimal approach. The second limitation is the asymptotic sub-optimality of the existing sub-optimal approaches. Based on two performance measures, the proposed results outperform those of sub-optimal approaches. Moreover, the results confirm the asymptotic optimality of the proposed method.

Appendix A
DP-RDF for Vector Sources With Memory

We show the derivations that result in the proposed DP-RDF in (17) for vector sources with memory (multi-sensor observations) based on DP-RDF for a discrete-time independent scalar (one sensor) observation samples.

We are given a sequence of discrete-time zero-mean stationary vector Gaussian sources, denoted by $\{\mathbf{s}[n]\}_{n=0}^{N-1}$, where $\mathbf{s}[n] \in \mathbb{R}^{M\times 1}$, $n = 0, \ldots, N-1$, with the corresponding block-Toeplitz cross correlation matrix $\mathbf{\Sigma_s} \in \mathbb{R}^{MN\times MN}$, defined in (6), and the Hermitian CPSD matrix $\mathbf{\Phi_s} \in \mathbb{C}^{M\times M}$. The sequence $\{\mathbf{s}[n]\}_{n=0}^{N-1}$ is stacked into the vector $\mathbf{s}_{\text{vec}} = [\mathbf{s}^T[0] \ldots \mathbf{s}^T[N-1]]^T$. We define the following DP optimization problem based on the DP-RDF defined in [28]

$$\begin{aligned} \inf_{f(\tilde{\mathbf{s}}_{\text{vec}}|\mathbf{s}_{\text{vec}})} \quad & I(\mathbf{s}_{\text{vec}};\tilde{\mathbf{s}}_{\text{vec}}) \\ \text{subject to} \quad & E[\|\mathbf{s}_{\text{vec}}-\tilde{\mathbf{s}}_{\text{vec}}\|^2] \le D^{DP}, \\ & f(\mathbf{s}_{\text{vec}}) = f(\tilde{\mathbf{s}}_{\text{vec}}), \end{aligned} \tag{22}$$

where $I(\mathbf{x};\mathbf{y})$ is generally the mutual information between the random vector variables $\mathbf{x}$ and $\mathbf{y}$. The conditional distribution function of a random vector variable $\mathbf{x}$, given a random vector variable $\mathbf{y}$ is denoted by $f(\mathbf{x}|\mathbf{y})$. The problem in (22) tries to find the minimum rate R^{DP} at which the vector $\mathbf{s}_{\text{vec}}$ can be quantized such that the probabilty distribution of the source, say $f(\mathbf{s}_{\text{vec}})$, is preserved after the quantization, i.e., $f(\mathbf{s}_{\text{vec}}) = f(\tilde{\mathbf{s}}_{\text{vec}})$, and the MSE between $\mathbf{s}_{\text{vec}}$ and its quantized output $\tilde{\mathbf{s}}_{\text{vec}}$ does not exceed a certain value D^{DP}. Mutual information is invariant under unitary transformations [13] and the objective mutual information function $I(\mathbf{s}_{\text{vec}};\tilde{\mathbf{s}}_{\text{vec}})$ can be rewritten as a summation of separable functions [13], [23]

$$I(\mathbf{s}_{\text{vec}};\tilde{\mathbf{s}}_{\text{vec}}) = I(\mathbf{s}_{\text{dec}};\tilde{\mathbf{s}}_{\text{dec}}) = \sum_{i=1}^{MN} I(s_{\text{dec}}[i];\tilde{s}_{\text{dec}}[i]), \tag{23}$$

where $s_{\text{dec}}[i]$ is the ith element of the transformed vector $\mathbf{s}_{\text{dec}} = \mathbf{V}^{\text{H}}\mathbf{s}_{\text{vec}}$. Matrix $\mathbf{V}$ is derived by eigenvalue decomposition of the correlation matrix, i.e., $\mathbf{\Sigma_s} = \mathbf{V\Lambda V}^{\text{H}}$, where $\mathbf{\Lambda} = \text{diag}\{\lambda_1(\mathbf{\Sigma_s}), \ldots, \lambda_{MN}(\mathbf{\Sigma_s})\}$. The second equality in (23) holds as the elements of $\mathbf{s}_{\text{dec}}$ are statistically independent. Note that as $\mathbf{V^{-1}} = \mathbf{V^H}$, we have $\tilde{\mathbf{s}}_{\text{vec}} = \mathbf{V}\tilde{\mathbf{s}}_{\text{dec}}$, where $\tilde{\mathbf{s}}_{\text{dec}}$ denotes the transformed-quantized vector signal. Using the unitary transformation, the reformulated problem is given by

$$\begin{aligned} \inf_{f(\tilde{\mathbf{s}}_{\text{dec}}|\mathbf{s}_{\text{dec}})} \quad & \sum_{i=1}^{MN} I(s_{\text{dec}}[i];\tilde{s}_{\text{dec}[i]}) \\ \text{subject to} \quad & \sum_{i=1}^{MN} D_i \le D^{DP}, \\ & f(s_{\text{dec}}[i]) = f(\tilde{s}_{\text{dec}}[i]), \end{aligned} \tag{24}$$

where $D_i = E[|s_{\text{dec}}[i]-\tilde{s}_{\text{dec}}[i]|^2]$. Note that the unitary transformation preserves the MSE, i.e., $E[\|\mathbf{s}_{\text{vec}}-\tilde{\mathbf{s}}_{\text{vec}}\|^2] = E[\|\mathbf{s}_{\text{dec}}-\tilde{\mathbf{s}}_{\text{dec}}\|^2] = \sum_{i=1}^{MN} E[|s_{\text{dec}}[i]-\tilde{s}_{\text{dec}}[i]|^2]$.

As the elements of $\mathbf{s}_{\text{dec}}$ are statistically independent, based on Lemma 3 in [22], the problem in (24) for a decorrelated vector

$\mathbf{s}_{\text{dec}}$, can be solved as

$$R_{MN}^{DP}(\mu) = \begin{cases} \sum_{i=1}^{MN} \log_2 \frac{E[|s_{\text{dec}}[i]|^2]}{\left(E[|s_{\text{dec}}[i]|^2]D_i(\mu) - \frac{D_i^2(\mu)}{4}\right)^{\frac{1}{2}}} & D_{MN}^{DP}(\mu) < 2\,\sigma^2 \\ 0 & D_{MN}^{DP}(\mu) \geq 2\,\sigma^2 \end{cases}$$

$$D_{MN}^{DP}(\mu) = \sum_{i=1}^{MN} D_i(\mu), \tag{25}$$

where

$$D_i(\mu) = 2\,E[|s_{\text{dec}}[i]|^2] + \mu - \left(4E[|s_{\text{dec}}[i]|^2] + \mu^2\right)^{\frac{1}{2}}, \tag{26}$$

with $E[|s_{\text{dec}}[i]|^2] = \lambda_i(\mathbf{\Sigma}_{\mathbf{s}})$, $\sigma^2 = \sum_{i=1}^{MN} E[|s_{\text{dec}}[i]|^2]$ and μ a Lagrange variable relating the rate to the distortion [22]. The equation (26) is valid for $D_i(\mu) \leq 2\,E[|s_{\text{dec}}[i]|^2]$. In (25) $R_{MN}^{DP}(\mu)$ is the minimum achievable rate at which the source $\mathbf{s}_{\text{vec}}$ can be encoded and decoded with distortion not exceeding a certain value D_{MN}^{DP} such that its PDF is preserved after quantization. Note that the rate is per vector source $\mathbf{s}_{\text{vec}}$. $D_i(\cdot)$ is the corresponding MSE with respect to the ith element of the decorrelated vector source $\mathbf{s}_{\text{dec}}$.

For a given μ, R_{MN}^{DP} can be represented as a sum of non-linear functions of the eigenvalues of the block-Toeplitz matrix $\mathbf{\Sigma}_{\mathbf{s}}$ (not a Toeplitz matrix as in [22]). Let the non-linear function be

$$F_R(\lambda_i(\mathbf{\Sigma}_{\mathbf{s}}), \mu) = \log_2 \frac{\lambda_i(\mathbf{\Sigma}_{\mathbf{s}})}{\left(\lambda_i(\mathbf{\Sigma}_{\mathbf{s}})D_i(\mu) - \frac{D_i^2(\mu)}{4}\right)^{\frac{1}{2}}}, \tag{27}$$

where $D_i(\cdot)$ is also a non-linear function of $\lambda_i(\mathbf{\Sigma}_{\mathbf{s}})$, as shown in (26). We define $R^{DP}(\mu)$ as an asymptotic ($N \to \infty$) average of non-linear functions $F_R(\lambda_i(\mathbf{\Sigma}_{\mathbf{s}}), \mu)$, which is given by

$$R^{DP}(\mu) = \lim_{N\to\infty} \frac{1}{N} R_{MN}^{DP}(\mu) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{MN} F_R(\lambda_i(\mathbf{\Sigma}_{\mathbf{s}}), \mu) \tag{28}$$

With a similar argument, $D^{DP}(\mu)$ is defined as

$$D^{DP}(\mu) = \lim_{N\to\infty} \frac{1}{N} D_{MN}^{DP}(\mu) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{MN} D_i(\mu) \tag{29}$$

We use (8), the extension of the Szego's theorem, to find the corresponding rate-distortion tradeoff in the frequency domain. Substituting non-linear functions in (27) and (26) into (8), the corresponding equivalences of (28) and (29) are derived in the frequency domain, respectively, and consequently, the asymptotic DP-RDF for time-stationary Gaussian vector sources is given by

$$R^{DP}(\mu) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{u=1}^{M} \log_2 \frac{\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))}{\left(\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))D_u(\mu,\Omega) - \frac{D_u^2(\mu,\Omega)}{4}\right)^{\frac{1}{2}}} d\Omega$$

$$D^{DP}(\mu) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{u=1}^{M} D_u(\mu,\Omega) d\Omega, \tag{30}$$

where

$$D_u(\mu,\Omega) = 2\,\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega)) + \mu - \left(4\,\lambda_u^2(\mathbf{\Phi}_{\mathbf{s}}(\Omega)) + \mu^2\right)^{\frac{1}{2}}, \tag{31}$$

for $D_u(\mu,\Omega) < 2\lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega))$. The asymptotic rate $R^{DP}(\mu)$ in (30) is assumed to be set to zero for $D^{DP}(\mu) > \frac{1}{\pi}\int_{-\pi}^{\pi} \sum_{u=1}^{M} \lambda_u(\mathbf{\Phi}_{\mathbf{s}}(\Omega)) d\Omega$.

Appendix B
A Test Channel Achieving DP-RDF

We show that the conceptual test channel achieving (17) is based on the vector quantizer model shown in Fig. 4(a). We derive the respected quantization parameters $\mathbf{B}$ and $\mathbf{\Phi}_{\mathbf{e}_1}$ for *Method 2*. We introduce the following lemma.

Lemma 1: Let $\tilde{z}(\Omega) \in \mathbb{C}$ and $z(\Omega) \in \mathbb{C}$ be zero-mean Gaussian random variables representing frequency domain signals. Then there exist a real-valued linear operator (scaling factor) β and a zero-mean Gaussian random variable $e(\Omega)$, uncorrelated to $z(\Omega)$, such that

$$\tilde{z} = \beta z + e, \tag{32}$$

and $E[|\tilde{z}|^2] = E[|z|^2]$, i.e., variables $\tilde{z}$ and $z(\Omega)$ have the same PSDs.

Proof: Denote the cross PSDs between $\tilde{z}$ and z by $\Phi_{\tilde{z}z} = E[\tilde{z}z^*]$ and $\Phi_{z\tilde{z}} = E[z\tilde{z}^*]$ and PSD of e by Φ_e, where $(\cdot)^*$ denotes the conjugate operator. Based on (32), the (cross) PSD relations are given by

$$\Phi_{\tilde{z}} = \beta^2 \Phi_z + \Phi_e, \tag{33a}$$

$$\Phi_{\tilde{z}z} = \Phi_{z\tilde{z}} = \beta \Phi_z, \tag{33b}$$

where we used the fact that z and e are uncorrelated, and that β is real. We define $D \triangleq E[|z-\tilde{z}|^2] \triangleq \Phi_{e'}$ where e' can be thought of as the error variable $e' = z - \tilde{z}$. As $\Phi_{\tilde{z}} = \Phi_z$, the distortion function D can be written as

$$D = \Phi_z + \Phi_z - 2\text{Re}\{\Phi_{z\tilde{z}}\} = 2\Phi_z - 2\Phi_{z\tilde{z}}. \tag{34}$$

Solving (34) and (33b) for β we have

$$\beta = 1 - \frac{D}{2\Phi_z}, \tag{35}$$

and substituting (35) into (33a) for Φ_e, we have

$$\Phi_e = \left(\frac{1+\beta}{2}\right) D. \tag{36}$$

The proof is complete.

Using Lemma 1, we derive the following distribution preserving quantization procedure for vector sources in the frequency domain, which achieves the DP-RDF in (30).

First, the left-side observations $\mathbf{y}_1$ are decorrelated as $\mathbf{z}_1 = \mathbf{A}\mathbf{y}_1$, using a unitary transformation matrix $\mathbf{A}$. Second, each element of the decorrelated vector $\mathbf{z}_1$, denoted by $Z_u(\Omega),;\ u = 1,..,M_1$, can be quantized in a probability distribution preserving manner based on the test channel model presented in Lemma 1 as $\tilde{Z}_u(\Omega) = \beta_u(\Omega)Z_u(\Omega) + E_u(\Omega)$. Let us denote the MSE $E[|Z_u(\Omega) - \tilde{Z}_u(\Omega)|^2]$ by $D_u(\Omega)$. Therefore, the distribution preserving quantization parameters $\beta_u(\Omega)$ and $\Phi_{E_u}(\Omega)$ correspond to (35) and (36), by replacing D and Φ_z with $D_u(\Omega)$ and $\Phi_{Z_u}(\Omega)$, respectively. Note that here the PSD of each element is preserved after the quantization, i.e., $E[|\tilde{Z}_u(\Omega)|^2] = E[|Z_u(\Omega)|^2] = [\Phi_{\mathbf{z}_1}]_{uu}$. We know from [22] and Appendix A that the optimal choices for the distortions $D_u(\Omega)$ are derived by minimizing the sum-rate with respect to the constraint on total distortion, i.e., $\sum_{u=1}^{M_1} D_u(\Omega) \leq D^{DP}(\Omega)$. Therefore the optimal values for $D_u(\Omega)$, which are derived based on (31) by replacing $\lambda_u(\mathbf{\Phi_s}(\Omega))$ with $[\Phi_{\mathbf{z}_1}]_{uu}$, i.e.,

$$D_u(\mu,\Omega) = 2\,[\Phi_{\mathbf{z}_1}]_{uu} + \mu - \left(4[\Phi_{\mathbf{z}_1}]_{uu} + \mu^2\right)^{\frac{1}{2}}, \qquad (37)$$

and hence, the optimal distribution preserving (DP) quantization of the decorrelated vector $\mathbf{z}_1$ is modeled by $\tilde{\mathbf{z}}_1 = \mathbf{B}\mathbf{z}_1 + \mathbf{e}_1$, where $\mathbf{B}$ is a diagonal matrix and the elements correspond to (35), replacing D and Φ_z with the optimal DP-MSE $D_u(\mu,\Omega)$ and $[\Phi_{\mathbf{z}_1}]_{uu}$, respectively. The vector $\mathbf{e}_1 = [E_1(\Omega),\ldots,E_{M_1}(\Omega)]^{\mathrm{T}}$ will have the diagonal PSD matrix $\mathbf{\Phi}_{\mathbf{e}_1}$ which correspond to (36) with similar substitutions to those for $\mathbf{B}$. Finally, the decorrelated-quantized vector $\tilde{\mathbf{z}}_1$ will be transformed back to the original quantized vector $\tilde{\mathbf{y}}_1$ applying inverse-transform matrix $\mathbf{A}^{-1}$ ($\tilde{\mathbf{y}}_1 = \mathbf{A}^{-1}\tilde{\mathbf{z}}_1$).

Following the above-mentioned procedure, we summarize the achievability proof of the distortion given in (17) by defining two error variables $\mathbf{d}_1 = \mathbf{y}_1 - \tilde{\mathbf{y}}_1$ and $\mathbf{e}'_1 = \mathbf{z}_1 - \tilde{\mathbf{z}}_1$. The direct distribution preserving MSE between $\mathbf{y}_1$ and $\tilde{\mathbf{y}}_1$ is denoted by $D^{DP}(\mu)$. We have

$$D^{DP}(\mu) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{\Phi}_{\mathbf{d}_1}\}d\Omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{\mathbf{\Phi}_{\mathbf{e}'_1}\}d\Omega \qquad (38a)$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{(\mathbf{I}-\mathbf{B})\mathbf{\Phi}_{\mathbf{z}_1}(\mathbf{I}-\mathbf{B})^{\mathrm{H}} + \mathbf{\Phi}_{\mathbf{e}_1}\}d\Omega \qquad (38b)$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{u=1}^{M_1}(1-\beta(\mu,\Omega))^2[\mathbf{\Phi}_{\mathbf{z}_1}]_{uu} + \frac{(1+\beta(\mu,\Omega))}{2}D_u(\mu,\Omega)d\Omega \qquad (38c)$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{u=1}^{M_1} D_u(\mu,\Omega)d\Omega, \qquad (38d)$$

which corresponds to the distortion function in (17), for the different vector source $\mathbf{y}_1$ (and not $\mathbf{s}$ in (17)). $D_u(\mu,\Omega)$ here corresponds to (37). Similar argument holds for the achievability proof of the parametric distribution preserving rate $R^{DP}(\mu)$.

REFERENCES

[1] V. Hamacher *et al.*, "Signal processing in high-end hearing aids: State of the art, challenges, and future trends," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 18, pp. 2915–2929, Nov. 2005.

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

[3] K. Eneman *et al.*, "Evaluation of signal enhancement algorithms for hearing instruments," in *Proc. 16th Eur. Signal Process. Conf.*, Aug. 2008, pp. 1–5.

[4] V. Hamacher, "Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2002, vol. 4, pp. 4008–4011.

[5] 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 Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

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

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

[8] 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 Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.

[9] D. Marquardt, "Development and Evaluation of Psychoacoustically Motivated Binaural Noise Reduction and Cue Preservation Techniques," Ph.D. dissertation, Signal Processing Group, Dept. Medical Phy. Acoust. Carl-von-Ossietzky, Univ. Oldenburg, Oldenburg, Germany, 2015.

[10] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP J. Adv. Signal Process.*, vol. 2006, no. 1, pp. 1–14, Mar. 2006.

[11] S. Srinivasan, "Low-bandwidth binaural beamforming," *Electron. Lett.*, vol. 44, no. 22, pp. 1292–1293, Oct. 2008.

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

[13] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.

[14] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 4, pp. 406–411, Jul. 1970.

[15] T. Flynn and R. Gray, "Encoding of correlated observations," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 773–787, Nov. 1987.

[16] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[17] 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.

[18] S. Srinivasan and A. den Brinker, "Rate-constrained beamforming in binaural hearing aids," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–9, 2009.

[19] S. Srinivasan and A. C. den Brinker, "Analyzing rate-constrained beamforming schemes in wireless binaural hearing aids," in *Proc. 17th Eur. Signal Process. Conf.*, Aug. 2009, pp. 1854–1858.

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

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

[22] M. Li, A. Ozerov, J. Klejsa, and W. B. Kleijn, "Asymptotically optimal distribution preserving quantization for stationary Gaussian processes," *IEEE Trans. Commun.*, 2011, arXiv: urn:nbn:se:kth:diva-38517.

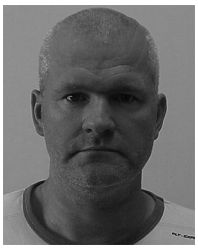
[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.

[24] V. Kafedziski, "Rate distortion of stationary and nonstationary vector Gaussian sources," in *Proc. IEEE/SP 13th Workshop Stat. Signal Process.*, Jul. 2005, pp. 1054–1059.

[25] U. Grenander and G. Szego, *Toeplitz Forms and Their Applications*. Berkeley, CA, USA: Univ. California Press, 1958.
[26] H. Gazzah, P. A. Regalia, and J. P. Delmas, "Asymptotic eigenvalue distribution of block Toeplitz matrices and application to blind SIMO channel identification," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1243–1251, Mar. 2001.
[27] P. Tilli, "Singular values and eigenvalues of non-Hermitian block Toeplitz matrices," *Linear Algebra Its Appl.*, vol. 272, no. 1, pp. 59–89, 1998.
[28] M. Li, J. Klejsa, and W. B. Kleijn, "On distribution preserving quantization," 2011, arXiv:1108.3728.
[29] 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.
[30] J. Kominek, A. W. Black, and V. Ver, "CMU arctic databases for speech synthesis," Language Technol. Inst., School Comput. Sci. Carnegie Mellon Univ., Tech. Rep. CMU-LTI-03-177, 2003.

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 the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2011. He is currently working toward the Ph.D. degree with the Circuits and Systems Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands. His research interests are on speech enhancement, speech analysis and synthesis, source coding, and voice conversion.

Richard C. 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 with the Circuits and Systems Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology. His main research interests are on biomedical signal processing, and audio and speech processing, including speech enhancement, speech intelligibility improvement, and intelligibility modeling. In March 2010, he received the prestigious VENI grant for his proposal Intelligibility Enhancement for Speech Communication Systems. He was the recipient of several best paper awards, among them is the IEEE Signal Processing Society Best Paper Award in 2016. He is an Associate Editor for the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and the *EURASIP Journal on Advances in Signal Processing*.



Richard Heusdens received the M.Sc. and Ph.D. degrees from the Delft University of Technology, Delft, The Netherlands, in 1992 and 1997, respectively. Since 2002, he has been an Associate Professor with 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, 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, 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 with Aalborg University from 2014 to 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 the 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. He is currently with Oticon A/S, Smørum, Denmark, as a Research Engineer. His main research interests are in the area of acoustic signal processing for hearing aid applications and healthcare wearable 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, Aalborg University, as an Assistant Research Professor. From 2000 to 2007, he was a Postdoctoral Researcher and an Assistant Professor with the Delft University of Technology, Delft, The Netherlands, and an External Associate Professor with Aalborg University. He is currently 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 also a Professor with the Section for Signal and Information Processing, Department of Electronic Systems, Aalborg University. He is also a co-founder of the Centre for Acoustic Signal Processing Research with 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.