RATE-DISTRIBUTED BINAURAL LCMV BEAMFORMING FOR ASSISTIVE HEARING IN WIRELESS ACOUSTIC SENSOR NETWORKS

Jie Zhang, Richard Heusdens, and Richard C. Hendriks

Circuits and Systems (CAS) Group, Delft University of Technology, the Netherlands

ABSTRACT

In this paper, we propose a rate-distributed linearly constrained minimum variance (LCMV) beamformer for joint noise reduction and spatial cue preservation for assistive hearing in wireless acoustic sensor networks (WASNs). The WASN can consist of wireless communicating hearing aids, extended with additional wireless microphones. Due to the fact that each sensor node has a limited power budget, it is essential to consider the energy usage when designing algorithms for such WASNs. As the energy usage in terms of data transmission is directly affected by the communication rate, the proposed method optimally distributes the bit rate for each microphone node. The rate distribution is obtained by minimizing the total transmission costs under constraints on the noise reduction performance and spatial cue preservation of interfering sources. In contrast to *sensor selection*, i.e., binary decisions on the usefulness of a node, rate distribution allows for soft decisions, and, will lead to more degrees of freedom for joint noise reduction and spatial cue preservation than sensor selection. Numerical results show that given a certain noise reduction requirement, the proposed method displays improved energy efficiency and can preserve the spatial cues of more interferers compared to sensor selection approaches.

Index Terms— Rate allocation, sensor selection, LCMV, noise reduction, energy usage, binaural cue preservation, hearing aids.

1. INTRODUCTION

With the introduction of wireless communication, binaural processing for hearing assistive devices has attracted an increasing interest, e.g., [1–3]. The traditional hearing-aid (HA) configuration consists of two HAs that are mounted on the two ears, but operate independently. Although this setup can very well suppress noise, it does not take interaural information between the two HAs into account. That is, traditional HAs cannot preserve the spatial cues in the sound field. However, in many scenarios the user needs to be able to identify the direction of the audible sound sources, which can be obtained from the spatial cues (e.g., interaural level/phase difference).

In order to jointly suppress noise and preserve spatial cues, binaural HA algorithms have been proposed assuming the availability of wireless communication channels, e.g., [4–7]. In this work, we consider a general framework where the HAs are part of a bigger wireless acoustic sensor network (WASN) with additional assistive wireless microphones, see Fig. 1. The microphones can thus be part of the HA itself, or positioned somewhere in the vicinity. The microphone recordings are transmitted via wireless links to a fusion center (FC), which we consider in this work to be one of the HAs, see Fig. 1. Subsequently, the FC computes the binaural outputs for both HAs and transmits the output to the contralateral HA. As such, the FC can preserve the interaural information in the binaural outputs. The larger number of microphones in such a setup can potentially lead to both better noise reduction and spatial cue preservation. These advantages of binaural HAs in a WASN setup come with higher battery costs for transmission of data, and, introduction of quantization noise. These facts are typically neglected in most contributions on binaural speech enhancement, with the exception of e.g., [8–12].

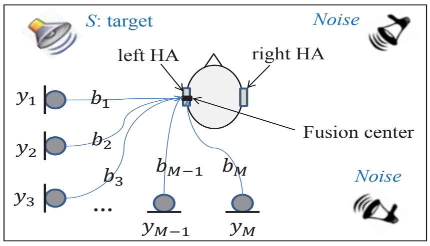

Fig. 1. A general binaural HA configuration in WASN.

In practice, HAs and assistive microphones in a WASN are battery driven, so that the trade off between the increased performance and energy usage for communication over such WASNs should be taken into account. Typically, the network lifetime needs to be maximized. In order to reduce the energy usage, generally there are two techniques that can be employed: *sensor selection* [13–15] and *rate allocation* [8, 9, 16, 17]. Sensor selection approaches lead to sparse networks, as only the most informative sensors are involved such that the energy usage in terms of data processing is saved effectively. Compared to sensor selection, rate allocation approaches can be used to distribute communication rates optimally to save the energy usage in terms of data transmission, since the transmission power between nodes and the FC is directly affected by the rate. The relationship between sensor selection and rate allocation was investigated in [17].

In this work, we apply the rate allocation approach in [17] to a binaural HA setting in a WASN. The problem is formulated by minimizing the total transmission power and constraining the noise reduction performance. The spatial cues are preserved using linear constrains within a binaural linearly constrained minimum variance (BLCMV) beamformer framework. Simulations show that although both the sensor selection and rate allocation approaches satisfy the performance requirement, the proposed rate allocation method is more efficient in energy usage and can preserve more interferers' spatial cues by including more sensors, each at a relatively low rate.

2. FUNDAMENTALS

2.1. Signal model

In this work, we assume that there are M microphones that are monitoring the sound field, see e.g. Fig. 1, where the FC allocates bit rates to each microphone node and computes the binaural output for each HA. In the short-term Fourier transform (STFT) domain, let l denote the frame index and ω the angular frequency bin. The noisy DFT coefficient of the quantized signal which is to be transmitted to the FC is given by

$$\hat{y}_k(\omega, l) = y_k(\omega, l) + q_k(\omega, l), \quad k = 1, 2, \cdots, M, \qquad (1)$$

where $q_k(\omega,l)$ denotes the quantization noise which is assumed to be uncorrelated with the microphone recording[1] $y_k(\omega,l)$ given by

$$y_k(\omega,l)=\sum_{i=1}^{\mathcal{I}}\underbrace{a_{ik}(\omega)s_i(\omega,l)}_{x_{ik}(\omega,l)}+\sum_{j=1}^{\mathcal{J}}\underbrace{h_{jk}(\omega)u_j(\omega,l)}_{n_{jk}(\omega,l)}+v_k(\omega,l),$$

where $a_{ik}(\omega)$ denotes the acoustic transfer function (ATF) of the ith target signal with respect to the kth microphone; $s_i(\omega,l)$ and $x_{ik}(\omega,l)$, the ith target source at the source location and at the kth microphone, respectively; $h_{jk}(\omega)$ the ATF from the jth interferer to the kth microphone; $u_j(\omega,l)$ and $n_{ik}(\omega,l)$, the jth interferer at the source location and at the kth microphone, respectively; $v_k(\omega,l)$ the kth microphone self noise. For notational brevity, we will omit the frequency variable ω and the frame index l now onwards. Using vector notation, the M channel signals are stacked in a vector $\hat{\mathbf{y}}=[\hat{y}_1,...,\hat{y}_M]^T$. Similarly, we define the vectors $\mathbf{y}$, $\mathbf{x}_i$, $\mathbf{n}_j$, $\mathbf{v}$, $\mathbf{q}$ for the microphone recordings, the ith target component, the jth interfering component, the additive noise and the quantization noise, respectively. Using this notation, (1) can be written compactly as

$$\hat{\mathbf{y}}=\sum_{i=1}^{\mathcal{I}}\mathbf{x}_i+\sum_{j=1}^{\mathcal{J}}\mathbf{n}_j+\mathbf{v}+\mathbf{q}=\mathbf{A}\mathbf{s}+\mathbf{H}\mathbf{u}+\mathbf{v}+\mathbf{q},\quad (2)$$

where $\mathbf{x}_i=\mathbf{a}_i s_i\in\mathbb{C}^M$ with $\mathbf{a}_i=[a_{i1},a_{i2},\cdots,a_{iM}]^T$ and $\mathbf{n}_j=\mathbf{h}_j u_j\in\mathbb{C}^M$ with $\mathbf{h}_j=[h_{j1},h_{j2},\cdots,h_{jM}]^T$. Further, $\mathbf{A}=[\mathbf{a}_1,\cdots,\mathbf{a}_{\mathcal{I}}]\in\mathbb{C}^{M\times\mathcal{I}},\mathbf{s}=[s_1,\cdots,s_{\mathcal{I}}]^T\in\mathbb{C}^{\mathcal{I}},\mathbf{H}=[\mathbf{h}_1,\cdots,\mathbf{h}_{\mathcal{J}}]\in\mathbb{C}^{M\times\mathcal{J}},\mathbf{u}=[u_1,\cdots,u_{\mathcal{J}}]^T\in\mathbb{C}^{\mathcal{J}}$. We assume in this work that the ATFs of the present sources (i.e., $\mathbf{A}$ and $\mathbf{H}$) are known. In practice, the target ATFs can be estimated using the generalized eigenvalue decomposition of the noise and noisy correlation matrices. The ATFs of the interferers can be replaced by pre-determined ATFs as in [19], at the cost of a small increase of the errors on the spatial cues. Assuming that all sources are mutually uncorrelated, the second-order statistics are then given by

$$\mathbf{R}_{\mathbf{yy}}=\mathbb{E}\{\mathbf{y}\mathbf{y}^H\}=\mathbf{R}_{\mathbf{xx}}+\underbrace{\mathbf{R}_{\mathbf{uu}}+\mathbf{R}_{\mathbf{vv}}}_{\mathbf{R}_{\mathbf{nn}}}\in\mathbb{C}^{M\times M},\quad (3)$$

where $\mathbf{R}_{\mathbf{xx}}=\sum_{i=1}^{\mathcal{I}}\mathbb{E}\{\mathbf{x}_i\mathbf{x}_i^H\}$ and $\mathbf{R}_{\mathbf{uu}}=\sum_{j=1}^{\mathcal{J}}\mathbb{E}\{\mathbf{n}_i\mathbf{n}_i^H\}$. In practice, $\mathbf{R}_{\mathbf{nn}}$ can be estimated using noise-only frames, and $\mathbf{R}_{\mathbf{yy}}$ during the speech+noise frames. The total noise second-order statistics in $\hat{\mathbf{y}}$ is given by $\mathbf{R}_{\mathbf{n+q}}=\mathbf{R}_{\mathbf{nn}}+\mathbf{R}_{\mathbf{qq}}$, under the assumption that the received noise sources and quantization noise are mutually uncorrelated. In case sensors utilize uniform quantizers to quantize their recordings, $\mathbf{R}_{\mathbf{qq}}$ then reads [16, 17, 20]

$$\mathbf{R}_{\mathbf{qq}}=\frac{1}{12}\mathrm{diag}\left(\left[\frac{\mathcal{A}_1^2}{4^{b_1}},\frac{\mathcal{A}_2^2}{4^{b_2}},...,\frac{\mathcal{A}_M^2}{4^{b_M}}\right]\right),\quad (4)$$

where $\mathcal{A}_k=\max\{|y_k|\}$ and $b_k,\forall k$ denotes the bit rate used by the kth microphone node. Note that the quantization in the sequel takes place in the STFT domain, e.g., the real and imaginary parts of the complex STFT coefficients are quantized separately.

2.2. BLCMV beamforming with binaural cue preservation

In [5], a general BLCMV beamforming framework was proposed for joint noise reduction and binaural cue preservation. Mathematically, this problem was formulated as

$$\hat{\mathbf{w}}_{\mathrm{BLCMV}}=\arg\min_{\mathbf{w}}\ \mathbf{w}^H\tilde{\mathbf{R}}_{\mathbf{n+q}}\mathbf{w},\quad \text{s.t.}\quad \mathbf{\Lambda}^H\mathbf{w}=\tilde{\mathbf{f}},\quad (5)$$

where

$$\tilde{\mathbf{R}}_{\mathbf{n+q}}=\begin{bmatrix}\mathbf{R}_{\mathbf{n+q}} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_{\mathbf{n+q}}\end{bmatrix}\in\mathbb{C}^{2M\times 2M},\quad (6)$$

$$\begin{aligned}\mathbf{\Lambda}&=\left[\begin{array}{c|c}\mathbf{\Lambda}_1 & \mathbf{\Lambda}_2\end{array}\right]\in\mathbb{C}^{2M\times(2\mathcal{I}+\mathcal{J})}\\ &=\left[\begin{array}{cc|ccc}\mathbf{A} & \mathbf{0} & \mathbf{h}_1h_{1R} & \cdots & \mathbf{h}_{\mathcal{J}}h_{\mathcal{J}R}\\ \mathbf{0} & \mathbf{A} & -\mathbf{h}_1h_{1L} & \cdots & -\mathbf{h}_{\mathcal{J}}h_{\mathcal{J}L}\end{array}\right],\end{aligned}\quad (7)$$

$$\begin{aligned}\tilde{\mathbf{f}}&=\left[\begin{array}{c|c}\mathbf{f}_1^H & \mathbf{f}_2^H\end{array}\right]^T\in\mathbb{C}^{2\mathcal{I}+\mathcal{J}}\\ &=\left[\begin{array}{cccccc|cccc}a_{1L}^* & \cdots & a_{\mathcal{I}L}^* & a_{1R}^* & \cdots & a_{\mathcal{I}R}^* & 0 & 0 & \cdots & 0\end{array}\right]^T,\end{aligned}$$

and the BLCMV beamformer is the concatenation of the LCMV beamformers at the two HAs, i.e., $\mathbf{w}_{\mathrm{BLCMV}}=[\mathbf{w}_L^T\ \mathbf{w}_R^T]^T$. In the BLCMV formulation, L and R are used to indicate the left and right beamformer or reference microphone for the two ears, respectively. Information on the spatial cues is contained in the interaural transfer function (ITF). The ITF of the ith target source with respect to the reference microphones can be defined as $\mathrm{ITF}_{\mathbf{x}_i}=\frac{a_{iL}}{a_{iR}},\forall i$, and the ITF of interferers can be defined similarly. Accordingly, we can see that the constraint $\mathbf{\Lambda}^H\mathbf{w}=\tilde{\mathbf{f}}$ in (5) consists of two components: 1) a constraint on the exact preservation of the $\mathcal{I}$ target sources, i.e., $\mathbf{\Lambda}_1^H\mathbf{w}=\mathbf{f}_1$, for which we know that full preservation requires

$$\mathrm{ITF}_{\mathbf{x}_i}^{\mathrm{in}}=\mathrm{ITF}_{\mathbf{x}_i}^{\mathrm{out}}=\frac{a_{iL}}{a_{iR}},i=1,\cdots,\mathcal{I};\quad (8)$$

2) A constraint on the preservation of the $\mathcal{J}$ interferers, i.e., $\mathbf{\Lambda}_2^H\mathbf{w}=\mathbf{f}_2$, for which we know that preserving the spatial cues requires

$$\mathrm{ITF}_{\mathbf{n}_j}^{\mathrm{in}}=\mathrm{ITF}_{\mathbf{n}_j}^{\mathrm{out}}=\frac{h_{jL}}{h_{jR}}=\frac{\mathbf{w}_L^H\mathbf{h}_j}{\mathbf{w}_R^H\mathbf{h}_j},j=1,\cdots,\mathcal{J}.\quad (9)$$

With the preservation of ITFs in (8-9), the binaural cues, e.g., interaural level difference (ILD) and interaural phase difference (IPD) are also preserved, because ILD and IPD are derived from ITF as

$$\mathrm{ILD}=|\mathrm{ITF}|^2,\quad \mathrm{IPD}=\angle\mathrm{ITF}.\quad (10)$$

Using the method of Lagrange multipliers, the closed-form solution of the above BLCMV problem is given by

$$\hat{\mathbf{w}}_{\mathrm{BLCMV}}=\tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1}\mathbf{\Lambda}(\mathbf{\Lambda}^H\tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1}\mathbf{\Lambda})^{-1}\tilde{\mathbf{f}}\in\mathbb{C}^{2M}.\quad (11)$$

For more details on BLCMV beamforming with binaural cue preservation, we refer to [5, 10, 11, 21, 22] and references therein.

3. RATE-DISTRIBUTED BLCMV BEAMFORMING

3.1. General problem formulation

Let V_k be the noise power spectral density (PSD) at the kth communication channel and d_k the distance over which transmission takes place. The transmission energy model is then given by [17]

$$g(\mathbf{b})=\sum_{k=1}^{M}d_k^2V_k(4^{b_k}-1),\quad (12)$$

where $\mathbf{b}=[b_1,\cdots,b_M]^T$. The above energy model holds under two conditions [23–25]: 1) in the context of band-limited applications (e.g., audio processing); 2) the microphone recordings are quantized at the channel capacity for reliable transmission. In this work, we intend to minimize $g(\mathbf{b})$ by allocating bit rates to microphone nodes, such that a prescribed noise reduction performance is obtained. With this, our initial problem can be formulated as:

$$\begin{aligned}\min_{\mathbf{w},\mathbf{b}}\ & g(\mathbf{b})=\sum_{k=1}^{M}d_k^2V_k(4^{b_k}-1)\\ \text{s.t.}\ & \mathbf{w}^H\tilde{\mathbf{R}}_{\mathbf{n+q}}\mathbf{w}\le\frac{\beta}{\alpha}\\ & \mathbf{\Lambda}^H\mathbf{w}=\tilde{\mathbf{f}},\quad b_k\in\mathbb{Z}_+,\quad b_k\le b_0,\forall k,\end{aligned}\quad \text{(P1)}$$

where β denotes the minimum output noise power that can be achieved, $\alpha\in(0,1]$ is to control the expected performance, $\mathbb{Z}_+$ denotes the set of non-negative integers, and b_0 the maximum number of bits per sample of each microphone signal. Note that β/α

[1] This assumption holds under high rate communication. At low rates, this can be achieved by subtractive dither [9, 18].

does not depend on the rate allocation strategy or statistics of the sensor network, because β/α is just a number that can be assigned by the users, e.g., 40 dB, to indicate a certain expected performance. By solving (P1), we can determine the optimal rate distribution that each microphone can utilize to quantize its recordings, such that the noise reduction system achieves a desired performance with minimum energy usage. One simple method to solve (P1) is exhaustive search, i.e., evaluating the performance for all $(b_0+1)^M$ choices for the rate distribution, but evidently this is intractable unless b_0 or M is very small. In the next section, we will propose an efficient solver for (P1) in the context of BLCMV beamforming.

3.2. Solver for rate-distributed BLCMV beamforming

Substituting the solution of the BLCMV beamformer from (11) to the general problem formulation in (P1), we can obtain a simplified optimization problem for rate-distributed BLCMV beamforming as

$$\begin{aligned} \min_{\mathbf{b}} \;& g(\mathbf{b}) = \sum_{k=1}^{M} d_k^2 V_k (4^{b_k} - 1) \\ \text{s.t.} \;& \tilde{\mathbf{f}}^H (\mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1} \mathbf{\Lambda})^{-1} \tilde{\mathbf{f}} \leq \frac{\beta}{\alpha} \\ & b_k \in \mathbb{Z}_+, \quad b_k \leq b_0, \forall k, \end{aligned} \tag{P2}$$

where $\mathbf{b}$ is implicit in the output noise power $\tilde{\mathbf{f}}^H (\mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1} \mathbf{\Lambda})^{-1} \tilde{\mathbf{f}}$, which is non-convex and non-linear in terms of $\mathbf{b}$. In what follows, we will explicitly express $\tilde{\mathbf{f}}^H (\mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1} \mathbf{\Lambda})^{-1} \tilde{\mathbf{f}}$ in terms of $\mathbf{b}$.

First of all, in order to reformulate (P2) as a convex optimization problem, we introduce a symmetric positive semi-definite matrix $\mathbf{Z} \in \mathbb{S}_+^{2\mathcal{I}+\mathcal{J}}$ with $\mathbb{S}_+$ denoting the set of symmetric positive semi-definite matrices, such that the first inequality constraint in (P2) can be recast to the following two new constraints equivalently, i.e.,

$$\mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1} \mathbf{\Lambda} = \mathbf{Z}, \tag{13}$$

$$\tilde{\mathbf{f}}^H \mathbf{Z}^{-1} \tilde{\mathbf{f}} \leq \frac{\beta}{\alpha}. \tag{14}$$

The inequality (14) can be rewritten as a linear matrix inequality (LMI) using the Schur complement [26, p.650], i.e.,

$$\begin{bmatrix} \mathbf{Z} & \mathbf{f} \\ \mathbf{f}^H & \frac{\beta}{\alpha} \end{bmatrix} \succeq \mathbf{O}_{2\mathcal{I}+\mathcal{J}+1}. \tag{15}$$

However, the equality constraint in (13) is both non-linear and non-convex in the unknown $\mathbf{b}$. The non-convexity can be tackled by relaxing it to

$$\mathbf{\Lambda}^H \mathbf{R}_{\mathbf{n+q}}^{-1} \mathbf{\Lambda} \succeq \mathbf{Z}, \tag{16}$$

since (14) and (16) are sufficient to obtain the original constraint in (P2). In order to linearize (16) in $\mathbf{b}$, we calculate $\tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1}$ as

$$\begin{aligned} \tilde{\mathbf{R}}_{\mathbf{n+q}}^{-1} &= (\tilde{\mathbf{R}}_{\mathbf{nn}} + \tilde{\mathbf{R}}_{\mathbf{qq}})^{-1} \\ &= \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} - \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} (\tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} + \tilde{\mathbf{R}}_{\mathbf{qq}}^{-1})^{-1} \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1}, \end{aligned} \tag{17}$$

where the second equality is derived from the matrix inversion lemma [27, p.18][2], and

$$\tilde{\mathbf{R}}_{\mathbf{nn}} = \begin{bmatrix} \mathbf{R}_{\mathbf{nn}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{nn}} \end{bmatrix}, \quad \tilde{\mathbf{R}}_{\mathbf{qq}} = \begin{bmatrix} \mathbf{R}_{\mathbf{qq}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{qq}} \end{bmatrix}. \tag{18}$$

Substituting $\mathbf{R}_{\mathbf{n+q}}^{-1}$ from (17) into (16), we obtain

$$\mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} \mathbf{\Lambda} - \mathbf{Z} \succeq \mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} (\tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} + \tilde{\mathbf{R}}_{\mathbf{qq}}^{-1})^{-1} \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} \mathbf{\Lambda}. \tag{19}$$

Using the Schur complement, we obtain the following LMI

$$\begin{bmatrix} \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} + \tilde{\mathbf{R}}_{\mathbf{qq}}^{-1} & \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} \mathbf{\Lambda} \\ \mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} & \mathbf{\Lambda}^H \tilde{\mathbf{R}}_{\mathbf{nn}}^{-1} \mathbf{\Lambda} - \mathbf{Z} \end{bmatrix} \succeq \mathbf{O}_{2M+2\mathcal{I}+\mathcal{J}}, \tag{20}$$

[2] $(\mathbf{A}+\mathbf{C}\mathbf{B}\mathbf{C}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{A}^{-1}$

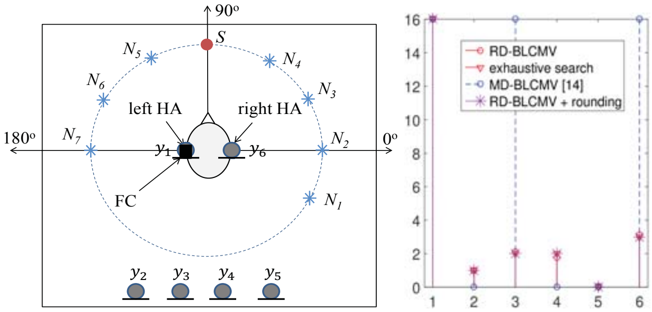

Fig. 2. Experimental setup and rate distribution obtained by RD-BLCMV, MD-BLCMV and exhaustive search, respectively.

where $\tilde{\mathbf{R}}_{\mathbf{qq}}^{-1} = \begin{bmatrix} \mathbf{R}_{\mathbf{qq}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\mathbf{qq}}^{-1} \end{bmatrix}$ and $\mathbf{R}_{\mathbf{qq}}^{-1}$ can be calculated from (4) directly. For notational convenience, we define a constant vector $\mathbf{e} = \left[\frac{12}{\mathcal{A}_1^2}, \cdots, \frac{12}{\mathcal{A}_M^2}\right]$. Further, we introduce a variable change $t_k = 4^{b_k} \in \mathbb{Z}_+, \forall k$, such that $\mathbf{R}_{\mathbf{qq}}^{-1} = \text{diag}(\mathbf{e} \odot \mathbf{t})$ and (20) are both linear in $\mathbf{t}$. In order to convexify the integer constraint $b_k \in \mathbb{Z}_+, \forall k$, we relax it to $b_k \in \mathbb{R}_+$, i.e., $t_k \in \mathbb{R}_+, \forall k$. Altogether, we arrive at

$$\begin{aligned} \min_{\mathbf{t}, \mathbf{Z}} \;& g(\mathbf{t}) = \sum_{k=1}^{M} d_k^2 V_k (t_k - 1) \\ \text{s.t.} \;& (15), (20), 1 \leq t_k \leq 4^{b_0}, \forall k, \end{aligned} \tag{21}$$

which is a standard semi-definite programming problem [26, p.128] and which can be solved efficiently in polynomial time using interior-point methods or solvers, e.g., CVX [28].

After (21) is solved, the allocated bit rates can be resolved by $b_k = \log_4 t_k, \forall k$ which are continuous values. In order to resolve the final integer rates, we apply the randomized rounding technique [13, 14, 17] to the solution of (21).

4. SIMULATION RESULTS

For the experiments, we place in total $M = 6$ microphones in a 2D room with dimensions (3×4) m, see Fig. 2 (left). From these M microphones, one is placed at each ear. These two microphones are taken as the reference microphone for the two HAs. The other four wireless microphones are placed as a (wireless) uniform linear array, whose x-coordinates are given by $\{0.9, 1.3, 1.7, 2.1\}$ m, and the y-coordinate is set to 1 m. The radius of the head of a user who wears the HAs is assumed to be 10 cm, and the FC is positioned at the left HA, i.e., the coordinate of the FC is $(1.4, 2.5)$ m. The ATFs are generated using [29] with reverberation time T_{60} = 200 ms. We consider one target source of interest which is located in front of the head. The target source signal is a 10 minute long concatenation of speech signals originating from the TIMIT database [30]. Seven speech shaped Gaussian interfering sources are present, and are placed at $-30°$, $0°$, $30°$, $60°$, $120°$, $150°$ and $180°$, respectively. All the sources are distributed on a circle (radius = 1 m) centered at the head and the elevation is set to be $0°$. All the signals are sampled at 16 kHz. We use a square-root Hann window of 20 ms for framing with 50% overlap. Microphone self noise is modeled at a signal-to-noise ratio (SNR) of 50 dB. The signal-to-(total)interference ratio (SIR) is set to be 0 dB. Furthermore, the PSD of the communication channel noise sources V_k in (21) are assumed to be the same for all microphones.

For comparison, we use as reference methods the model-driven BLCMV (MD-BLCMV) based microphone selection [14] and a solution to (P2) based on exhaustive searching. MD-BLCMV is an extension of [14] to our binaural setup. In order to validate the opti-

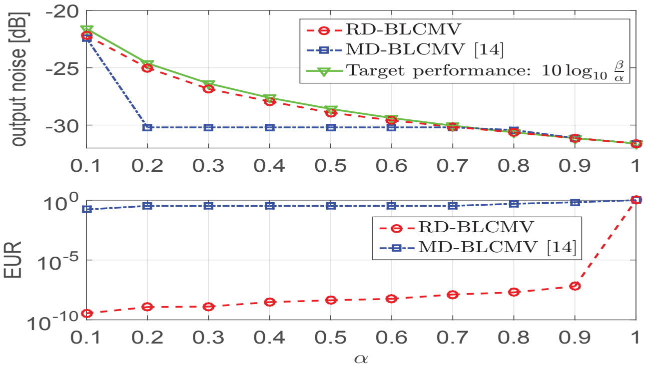

Fig. 3. Output noise power and energy efficiency in terms of α.

mality of the proposed method, the exhaustive search is employed to find out the optimal rate distribution. For the maximum rate $b_0 = 16$ bits and six microphones, there are 17^6 combinations for the exhaustive search. Fig. 2 (right) shows the rate distribution obtained by (21) (i.e., the proposed method referred as RD-BLCMV), MD-BLCMV and exhaustive search, respectively. The performance control parameter α for all methods is set to be 0.8. We observe that the proposed RD-BLCMV method is very close to the optimal solution that is obtained by the exhaustive search, and if we post-process the results from RD-BLCMV using randomized rounding, they are completely the same. For RD-BLCMV, five sensors are activated, and the first one is allocated with highest rate (i.e., 16 bits per sample), because it has no communication cost to the FC and has a higher SNR. The rates of the other activated sensors obtained by RD-BLCMV are much smaller than 16 bits, resulting in a saving of transmission costs. The MD-BLCMV method selects only three microphones, but each operates at the maximum rate of 16 bits per sample.

Next, we compare the output noise power and energy usage ratio (EUR) in terms of the performance control parameter α. The EUR is defined as the ratio between the energy used by RD-BLCMV or MD-BLCMV and the maximum transmission energy when all sensors are involved and each quantizes at b_0 bits. By inspection of Fig. 3, we see that both RD-BLCMV and MD-BLCMV [14] satisfy the desired amount of noise reduction, but RD-BLCMV is much closer to the target performance $10\log_{10}\frac{\beta}{\alpha}$, particularly when $0.2 \le \alpha \le 0.6$. Actually, for these α-values, the two microphones at the ears and the third microphone are chosen for MD-BLCMV, so that the output noise power and energy efficiency of MD-BLCMV remains the same over this α-range. More importantly, RD-BLCMV has much better energy efficiency compared to MD-BLCMV.

Fig. 4 shows the total preservation errors of the binaural cues (e.g., ILD and IPD) in terms of the number of activated interferers[3]. The errors ΔILD and ΔIPD are defined as

$$\Delta\text{ILD} = \sum_{j=1}^{\mathcal{J}}\sum_{\omega}\left(\text{ILD}_j(\omega) - \tilde{\text{ILD}}_j(\omega)\right)^2,$$

$$\Delta\text{IPD} = \sum_{j=1}^{\mathcal{J}}\sum_{\omega}\left(\text{IPD}_j(\omega) - \tilde{\text{IPD}}_j(\omega)\right)^2.$$

The RD-BLCMV method is compared to a BMVDR beamformer [11], a BLCMV framework [5] and the MD-BLCMV beamformer. The

[3] In [5], it was shown that the binaural cues of at most $2M - 2\mathcal{I} - 1$ interferers can be preserved with M microphones using the BLCMV beamformer in (5). In our case with $M = 6$ microphones, $\mathcal{I} = 1$ target source and $\mathcal{J} = 7$ interferers, the binaural cues of both the target source and all the interferers can be preserved by BLCMV or RD-BLCMV because $2M - 3 > \mathcal{J}$, and the degree of freedom devoted to noise reduction is $2M - \mathcal{J} - 2 = 3$.

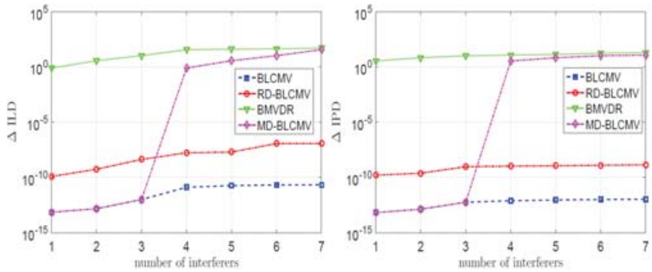

Fig. 4. Preservation errors versus the number of activated interferers.

BMVDR method is the worst preserving algorithm, as it does not consider binaural cue preservation constraints at all. More specifically, for the BMVDR method, the left and right MVDR beamformers can be formulated as

$$\mathbf{w}_L = \frac{\mathbf{R}_{\mathbf{n+q}}^{-1}\mathbf{a}a_L^*}{\mathbf{a}^H\mathbf{R}_{\mathbf{n+q}}^{-1}\mathbf{a}}, \quad \mathbf{w}_R = \frac{\mathbf{R}_{\mathbf{n+q}}^{-1}\mathbf{a}a_R^*}{\mathbf{a}^H\mathbf{R}_{\mathbf{n+q}}^{-1}\mathbf{a}} \tag{22}$$

for one target source that is identified by the ATF $\mathbf{a}$. Clearly, we have $\text{ITF}_{\mathbf{x}}^{\text{in}} = \text{ITF}_{\mathbf{x}}^{\text{out}} = \frac{a_L}{a_R}$ using the BMVDR beamformers. However, $\text{ITF}_{\mathbf{n}_j}^{\text{out}} = \frac{\mathbf{w}_L^H\mathbf{h}_j}{\mathbf{w}_R^H\mathbf{h}_j} = \frac{a_L}{a_R} = \text{ITF}_{\mathbf{x}}^{\text{in}}, \forall j$, which implies that the output binaural cues of the interfering sources collapse to the binaural cues of the target source. Hence, the BMVDR beamformer cannot preserve any binaural cues of interferers. The BLCMV method shows the best performance. However, it does not take the quantization into account and includes all microphones. This means it will be able to keep the spatial cues of all present sources, however, at the high battery cost of full-rate transmission. The MD-BLCMV method uses a hard selection, e.g., if it selects a subset of microphones that is too small, it will not be able to preserve the spatial cues of all sources. The RD-BLCMV approach applies the rate distribution and thus has a soft decision of microphones. In that sense, it usually activates more microphones (at the cost of more quantization noise), but this might lead to more degrees of freedom to preserve more spatial cues, while still satisfying the target noise reduction performance. In addition, all the methods can preserve the spatial cues of the target source because of the constraint $\mathbf{\Lambda}_1^H\mathbf{w} = \mathbf{f}_1$ being taken into account. From Fig. 4, we see that with an increasing number of interferers, the errors of RD-BLCMV or BLCMV only slightly increase, but the errors of MD-BLCMV suddenly increase when there are more than 3 interferers. This is because the BLCMV beamformers can preserve the binaural cues of up to $2M - 2\mathcal{I} - 1$ interferers using M microphones [5]. Using hard decisions on microphone selection, the degrees of freedom are much lower than when we use the rate allocation which is a soft decision. Therefore, the RD-BLCMV beamformer allows to use more constraints to preserve interferers than the MD-BLCMV beamformer: 7 versus 3 interferers in Fig. 2. Furthermore, similar to the BMVDR, the output binaural cues of the {4,5,6,7}th interferer based on MD-BLCMV will also collapse to those of the target source.

5. CONCLUSION

In this work, we studied rate-distributed BLCMV beamforming for wireless binaural hearing aids. The proposed method was formulated by minimizing the energy usage and constraining the noise reduction performance. Under the utilization of a BLCMV beamformer, the problem was solved by semi-definite programming with the capability of joint noise reduction and binaural cue preservation. The proposed method can achieve better energy efficiency by distributing bit rates, and preserve more interferers' spatial cues by activating more sensors as compared to sensor selection approaches.

6. REFERENCES

[1] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP J. Advances Signal Process.*, vol. 2006, pp. 175–175, 2006.

[2] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "Speech enhancement with multichannel wiener filter techniques in multimicrophone binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 125, no. 1, pp. 360–371, 2009.

[3] X. Zhang and D. Wang, "Deep learning based binaural speech separation in reverberant environments," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 5, pp. 1075–1084, 2017.

[4] T. J. Klasen, T. 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, 2007.

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

[6] A. Hassani, A. Bertrand, and M. Moonen, "Cooperative integrated noise reduction and node-specific direction-of-arrival estimation in a fully connected wireless acoustic sensor network," *ELSEVIER Signal Process.*, vol. 107, pp. 68–81, 2015.

[7] A. Hassani, J. Plata-Chaves, M. H. Bahari, M. Moonen, and A. Bertrand, "Multi-task wireless sensor network for joint distributed node-specific signal enhancement, LCMV beamforming and DOA estimation," *IEEE Journal of Selected Topics in Signal Process.*, vol. 11, no. 3, pp. 518–533, 2017.

[8] 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, 2009.

[9] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "On the impact of quantization on binaural MVDR beamforming," in *12th ITG Symposium of Speech Communication*. VDE, 2016, pp. 1–5.

[10] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 3, pp. 543–558, 2016.

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

[12] S. Srinivasan, A. Pandharipande, and K. Janse, "Beamforming under quantization errors in wireless binaural hearing aids," *EURASIP J. Audio, Speech, Music Process.*, vol. 2008, no. 1, pp. 824797, 2008.

[13] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, 2015.

[14] J. Zhang, S. P. Chepuri, R. C. Hendriks, and R. Heusdens, "Microphone subset selection for MVDR beamformer based noise reduction," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 8, pp. 550–563, 2018.

[15] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, 2009.

[16] F. de la Hucha Arce, F. Rosas, M. Moonen, M. Verhelst, and A. Bertrand, "Generalized signal utility for LMMSE signal estimation with application to greedy quantization in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1202–1206, 2016.

[17] J. Zhang, R. Heusdens, and R. C. Hendriks, "Rate-distributed spatial filtering based noise reduction in wireless acoustic sensor networks," *arXiv preprint arXiv:1712.07941*, 2017.

[18] R. M. Gray and T. G. Stockham, "Dithered quantizers," *IEEE Trans. Information Theory*, vol. 39, no. 3, pp. 805–812, 1993.

[19] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, J. Jensen, and M. Guo, "Binaural beamforming using pre-determined relative acoustic transfer functions," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*. IEEE, 2017, pp. 1–5.

[20] R. M. Gray, "Quantization noise spectra," *IEEE Trans. Information Theory*, vol. 36, no. 6, pp. 1220–1244, 1990.

[21] A. Bertrand and M. Moonen, "Distributed LCMV beamforming in a wireless sensor network with single-channel per-node signal transmission," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3447–3459, 2013.

[22] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 233–246, 2012.

[23] S. Shah and B. Beferull-Lozano, "Adaptive quantization for multihop progressive estimation in wireless sensor networks," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*. IEEE, 2013, pp. 1–5.

[24] Y. Huang and Y. Hua, "Multihop progressive decentralized estimation in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 1004–1007, 2007.

[25] Y. Huang and Y. Hua, "Energy planning for progressive estimation in multihop sensor networks," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4052–4065, 2009.

[26] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[27] K. B. Petersen, M. S. Pedersen, et al., "The matrix cookbook," *Technical University of Denmark*, vol. 7, pp. 15, 2008.

[28] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2008.

[29] E. A. P. Habets, "Room impulse response generator," Tech. Rep.

[30] J. S. Garofolo, "DARPA TIMIT acoustic-phonetic speech database," *National Institute of Standards and Technology (NIST)*, vol. 15, pp. 29–50, 1988.

978-1-5386-4752-3/18/$31.00 ©2018 IEEE