A NOVEL BINAURAL BEAMFORMING SCHEME WITH LOW COMPLEXITY MINIMIZING BINAURAL-CUE DISTORTIONS

Andreas I. Koutrouvelis[†], *Richard C. Hendriks*[†], *Richard Heusdens*[†], *Jesper Jensen*[‡,⋆], *and Meng Guo*[⋆]

[†]Circuits and Systems (CAS) Group, Delft University of Technology, the Netherlands
[‡]Electronic Systems Department, Aalborg University, Denmark
[⋆]Oticon A/S, Denmark
e-mails: {a.koutrouvelis, r.c.hendriks, r.heusdens}@tudelft.nl and {jesj, megu}@oticon.com

ABSTRACT

While the majority of binaural beamformers aim to minimize the output noise power while (approximately) preserving the binaural cues of the sources using constraints, we propose in this paper to minimize the binaural-cue distortions of the sources in the acoustic scene, such that the output noise power is below a predefined threshold. This new problem formulation is a convex QCQP problem, which leads to an efficient trade-off between noise reduction, binaural-cue preservation and complexity. In particular, the proposed beamformer provides a better trade-off between noise reduction and binaural-cue preservation (in terms of interaural level and phase differences) compared to the well-known binaural minimum variance distortionless response-η beamformer.

Index Terms— Binaural beamforming, binaural cues, convex optimization, noise reduction.

1. INTRODUCTION

Beamforming [1, 2] plays an important role in hearing assistive devices (HADs) [3, 4] such as hearing aids and cochlear implants. It aims at the reduction of background noise while preserving the signal coming from the target direction using multi-microphone recordings. The larger the number of microphones, the larger the degrees of freedom for noise reduction. Apart from noise reduction, some degrees of freedom may also be dedicated to preserve binaural cues of the acoustic sources after filtering [4]. Preservation of the directional binaural cues (interaural level and phase differences) is important for the HAD user to correctly localize the sources after filtering. Typically, this is achieved by using binaural beamformers which are implemented on binaural HAD systems [3, 4]. Usually, in binaural HAD systems, there are two collaborative HADs, with multiple microphones each, exchanging information via a wireless link [3].

Several binaural beamformers have been proposed in the literature, which can be classified into two main categories based on the information that is used. The first category consists of binaural beamformers (see e.g., [5–8]) that only need the noise cross power spectral density matrix (CPSDM) and the acoustic transfer functions (ATFs) of the target source with respect to the microphones. The binaural minimum variance distortionless response (BMVDR) beamformer has the best noise reduction performance among all distortionless linear spatial filters, since it spends all degrees of freedom on noise reduction [4]. As a result there may be severe binaural-cue distortions of the noise field after filtering [4]. Another example is the BMVDR-η beamformer [5,8], which adds a portion of the unprocessed acoustic scene to the BMVDR output. The larger the portion that is added, the better the binaural-cue preservation, but the worse the noise reduction performance.

The second category consists of binaural beamformers (see e.g., [9–12]) which also depend on the ATFs of the interfering sources. The joint binaural linearly constrained minimum variance (JBLCMV) beamformer [10, 11] uses one linear equality constraint per source/location to strictly preserve its binaural cues. Although this method has a closed-form solution, due to the limited number of available microphones, the equality constraints exhaust rapidly the degrees of freedom. This method is therefore inappropriate when a large number of sources is present. To tackle this problem, the relaxed JBLCMV (RJBLCMV) was proposed in [12], which uses inequality constraints instead of strict equality constraints to preserve the binaural cues. As a result, the optimization problem of the RJBLCMV has a larger feasibility set than the JBLCMV and, thus, the RJBLCMV allows more constraints to be used for more sources. Alternatively, for the same number of constraints, the RJBLCMV can lead to a better noise reduction performance by allowing some binaural-cue distortions. The optimization problem of the RJBLCMV method is not convex and does not have a closed-form solution. In [12], this problem was approximately solved with a series of convex optimization problems per time-frequency tile [12] leading to a large computational complexity.

In this paper, we propose an alternative problem formulation for binaural beamforming that belongs to the second category mentioned above. Instead of imposing constraints on binaural-cue preservation, we minimize the sum of the binaural-cue distortions under a constraint that the output noise power is below a user-defined threshold. The obtained formulation is a convex QCQP optimization problem [13], which leads to a beamformer that has a much lower computational complexity than the RJBLCMV optimization problem solver proposed in [12]. Moreover, the proposed method shows a better trade-off between directional binaural-cue preservation and noise reduction/predicted intelligibility compared to the BMVDR-η method and a similar trade-off as the RJBLCMV.

2. PRELIMINARIES

Let us assume that there are two HADs, each placed on one of the two ears, having M_L and M_R microphones on the left and right HAD, respectively. By allowing communication between the two devices, the two microphone arrays are merged into a single larger array of $M = M_L + M_R$ microphones. The multi-microphone noisy

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

signal model in the short-time Fourier transform domain is given by

$$\mathbf{y}(t,k)=\mathbf{a}(t,k)s(t,k)+\underbrace{\sum_{i=1}^{r}\mathbf{b}_i(t,k)u_i(t,k)+\mathbf{v}(t,k)}_{\mathbf{n}(t,k)}\in\mathbb{C}^{M\times 1}, \quad (1)$$

where t, k are the time-frequency indices, s and u_i are the target and i-th interfering signals at the original locations, and $\mathbf{a}$ and $\mathbf{b}_i$ their ATFs, respectively. The ambient noise vector $\mathbf{v}$ is the sum of microphone-self noise and diffuse noise. The total noise signal is denoted by $\mathbf{n}$. In binaural beamforming, two microphones are typically used as reference microphones, one at each HAD. The first M_{L} elements of all vectors in (1) are associated to the left HAD with the first element referring to its reference microphone, while the remaining M_{R} elements are associated to the right HAD, with the last element referring to its reference microphone. The vector elements corresponding to the left and right reference microphones have subscripts L and R, respectively. Since we perform the filtering operations for each time-frequency tile independently, we omit the time-frequency indices from now on for brevity.

3. EXISTING BINAURAL BEAMFORMERS

In binaural beamforming, there are two filters $\mathbf{w}_{\text{L}}, \mathbf{w}_{\text{R}} \in \mathbb{C}^{M\times 1}$ forming the binaural filter $\mathbf{w} = [\mathbf{w}_{\text{L}}^T\ \mathbf{w}_{\text{R}}^T]^T \in \mathbb{C}^{2M\times 1}$. Since there are two filters, there are two outputs, given by

$$x_{\text{L}} = \mathbf{w}_{\text{L}}^H\mathbf{y}, \quad x_{\text{R}} = \mathbf{w}_{\text{R}}^H\mathbf{y}, \quad (2)$$

where x_{L} and x_{R} are played back by the loudspeaker of the left and right HAD, respectively. All binaural filters examined in this paper belong to the family of distortionless filters (with respect to the target) satisfying the underdetermined system of equations given by

$$\mathbf{w}^H\mathbf{\Lambda}_{\text{A}} = \mathbf{f}_{\text{A}}^H, \quad (3)$$

where

$$\mathbf{\Lambda}_{\text{A}} = \begin{bmatrix}\mathbf{a} & \mathbf{0}\\ \mathbf{0} & \mathbf{a}\end{bmatrix} \in \mathbb{C}^{2M\times 2}, \qquad \mathbf{f}_{\text{A}} = \begin{bmatrix}a_{\text{L}}^*\\ a_{\text{R}}^*\end{bmatrix} \in \mathbb{C}^{2\times 1}. \quad (4)$$

Finally, a common goal of most binaural beamformers is to reduce the output noise power of the binaural filter given by

$$\mathbf{w}_{\text{L}}^H\mathbf{P}_{\mathbf{n}}\mathbf{w}_{\text{L}} + \mathbf{w}_{\text{R}}^H\mathbf{P}_{\mathbf{n}}\mathbf{w}_{\text{R}} = \mathbf{w}^H\mathbf{P}\mathbf{w}, \quad (5)$$

where $\mathbf{P}_{\mathbf{n}}$ is the CPSDM of the noise and $\mathbf{P}$ is a block-diagonal matrix with its two block matrices being each equal to $\mathbf{P}_{\mathbf{n}}$. In the sequel, we often use the binaural filter that performs no noise reduction, but simply presents the noisy signal of the reference microphones to the left and right HAD and is given by $\mathbf{w} = \mathbf{e}$, with

$$\mathbf{e} = [\mathbf{e}_{\text{L}}^T\ \ \mathbf{e}_{\text{R}}^T]^T,\ \mathbf{e}_{\text{L}} = [1\ \ 0\ \ \cdots\ \ 0]^T,\ \mathbf{e}_{\text{R}} = [0\ \ \cdots\ \ 0\ \ 1]^T. \quad (6)$$

The most important directional binaural cues are the interaural level differences (ILDs) and the interaural phase differences (IPDs) [14], which are the magnitude square and phase, respectively, of the interaural transfer function (ITF) [15]. The ITF of a point source, with ATF $\mathbf{b}_i$, before and after filtering is given by [15]

$$\text{ITF}_{\text{in}}^i = \frac{b_{i\text{L}}}{b_{i\text{R}}}, \quad \text{ITF}_{\text{out}}^i = \frac{\mathbf{w}_{\text{L}}^H\mathbf{b}_i}{\mathbf{w}_{\text{R}}^H\mathbf{b}_i}. \quad (7)$$

If the input and output ITFs are equal, there will be no directional binaural-cue distortions. Some of the binaural beamformers exploit this observation by setting $\text{ITF}_{\text{in}}^i = \text{ITF}_{\text{out}}^i$ as a constraint [10,11,16].

3.1. BMVDR Beamformer

The BMVDR [4] filter is the best performing linear binaural filter in terms of noise reduction among those filters satisfying the constraint in (3). It is the solution of the following optimizaiton problem:

$$\hat{\mathbf{w}}_{\text{BMVDR}} = \arg\min_{\mathbf{w}} \mathbf{w}^H\hat{\mathbf{P}}\mathbf{w} \text{ s.t. } \mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{A}} = \hat{\mathbf{f}}_{\text{A}}^H, \quad (8)$$

where the $\hat{}$ denotes quantities that need to be estimated in practice (i.e., the noise CPSDM and the target ATFs). For the BMVDR, the input and output ITFs are only equal for the target (if $\mathbf{a} = \hat{\mathbf{a}}$), but not for the interferers. More specifically, after filtering, the noise sounds as if it is coming from the target direction [4].

3.2. BMVDR-η Beamformer

The BMVDR-η [5, 8] adds a portion of the unprocessed acoustic scene to the BMVDR output in order to partially preserve the binaural cues of the noise field. The BMVDR-η filter is given by

$$\hat{\mathbf{w}}_{\text{BMVDR}-\eta} = \eta\hat{\mathbf{w}}_{\text{BMVDR}} + (1-\eta)\mathbf{e}, \quad (9)$$

where $\eta \in [0, 1]$ is the trade-off parameter that controls noise reduction and binaural-cue preservation. The $\hat{\mathbf{w}}_{\text{BMVDR}-\eta}$ also satisfies the constraint in (3). For $\eta = 1$, the BMVDR is obtained, while for $\eta = 0$, the unprocessed scene is obtained. If $\eta > 0$, the output ITF of the BMVDR-η is never equal to the ITF input of the interferers.

3.3. JBLCMV Beamformer

The joint binaural linearly constrained minimum variance (JBLCMV) beamformer [10, 11] adds constraints to the problem in (8) such that the directional binaural-cues of the interfering sources are exactly preserved by guaranteeing that the input and output ITFs are equal for all individual point source interferers. That is,

$$\hat{\mathbf{w}}_{\text{JBLCMV}} = \arg\min_{\mathbf{w}} \mathbf{w}^H\hat{\mathbf{P}}\mathbf{w} \text{ s.t. } \mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{A}} = \hat{\mathbf{f}}_{\text{A}}^H, \quad \mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{B}} = \mathbf{0}^T, \quad (10)$$

where

$$\hat{\mathbf{\Lambda}}_{\text{B}} = \begin{bmatrix}\hat{\mathbf{b}}_1\hat{b}_{1\text{R}} & \cdots & \hat{\mathbf{b}}_r\hat{b}_{r\text{R}}\\ -\hat{\mathbf{b}}_1\hat{b}_{1\text{L}} & \cdots & -\hat{\mathbf{b}}_r\hat{b}_{r\text{L}}\end{bmatrix} \in \mathbb{C}^{2M\times r}. \quad (11)$$

Let $\hat{\mathbf{\Lambda}} = [\hat{\mathbf{\Lambda}}_{\text{A}}\ \hat{\mathbf{\Lambda}}_{\text{B}}]$ and $\hat{\mathbf{f}} = [\hat{\mathbf{f}}_{\text{A}}^T\ \mathbf{0}^T]^T$. The JBLCMV problem in (10) has a closed-form solution (only for $r \le 2M-2$) given by

$$\hat{\mathbf{w}}_{\text{JBLCMV}} = \begin{cases}\hat{\mathbf{P}}^{-1}\hat{\mathbf{\Lambda}}\left(\hat{\mathbf{\Lambda}}^H\hat{\mathbf{P}}^{-1}\hat{\mathbf{\Lambda}}\right)^{-1}\hat{\mathbf{f}} & \text{if } r \le 2M-3\\ (\hat{\mathbf{\Lambda}}^H)^{-1}\hat{\mathbf{f}} & \text{if } r = 2M-2.\end{cases} \quad (12)$$

Unlike the BMVDR and BMVDR-η, the JBLCMV can simultaneously achieve perfect preservation of the directional binaural cues (if $\mathbf{b}_i = \hat{\mathbf{b}}_i\ \forall i$) and some noise reduction if $r \le 2M-3$.

3.4. RJBLCMV Beamformer

The relaxed JBLCMV (RJBLCMV) has a similar problem formulation as in (10), except for the fact that the equality constraints for the binaural-cue preservation of the interferers are replaced by non-convex inequality constraints. The inequality constraints upper bound the ITF error and allow a controlled trade-off between noise reduction and binaural-cue preservation. Due to the non-convex nature of the RJBLCMV optimization problem, a successive convex optimization method is used to approximately solve it in [12]. Solving multiple convex optimization problems per time-frequency tile leads to a large computational load for the application at hand.

4. PROPOSED METHOD

Unlike the methods reviewed in Sec. 3, which minimize the output noise power under controlled binaural-cue distortions, the proposed method does the opposite. It minimizes the distortions of the directional binaural cues of all interferers, given by $||\mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{B}}||_2^2$, while controlling the noise reduction performance using the constraint

$$\hat{p}_{\text{LL}}+\hat{p}_{\text{RR}}-\mathbf{w}^H\hat{\mathbf{P}}\mathbf{w}\geq\phi(\hat{p}_{\text{LL}}+\hat{p}_{\text{RR}}-\hat{\mathbf{w}}_{\text{BMVDR}}^H\hat{\mathbf{P}}\hat{\mathbf{w}}_{\text{BMVDR}}),\quad(13)$$

where $\hat{p}_{\text{LL}}+\hat{p}_{\text{RR}}=\mathbf{e}^T\hat{\mathbf{P}}\mathbf{e}$ is the sum of the estimated input noise power at the two reference microphones and $\phi\in[0,1]$ is the trade-off parameter that controls the output noise power. The constraint in (13) provides solutions that have a better noise reduction gain compared to a fraction ϕ of the BMVDR noise reduction gain. Altogether, the proposed convex QCQP problem is given by

$$\hat{\mathbf{w}}_{\text{prop}}=\arg\min_{\mathbf{w}}||\mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{B}}||_2^2 \text{ s.t. } \mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{A}}=\hat{\mathbf{f}}_{\text{A}}^H,$$
$$\mathbf{w}^H\hat{\mathbf{P}}\mathbf{w}\leq\epsilon(\phi),\qquad(14)$$

where $\epsilon(\phi)$ follows from (13) and is given by

$$\epsilon(\phi)=\phi\hat{\mathbf{w}}_{\text{BMVDR}}^H\hat{\mathbf{P}}\hat{\mathbf{w}}_{\text{BMVDR}}+(1-\phi)(\hat{p}_{\text{LL}}+\hat{\text{p}}_{\text{RR}}).\qquad(15)$$

If $\phi=0$, the optimization problem will provide a solution which is better or equal to the unprocessed scene in terms of noise reduction, while if $\phi=1$, the optimization problem will provide the BMVDR solution, which is the only feasible solution in this case. An interesting choice of ϕ is given by

$$\phi_{\text{JBLCMV}}=\frac{\hat{p}_{\text{LL}}+\hat{p}_{\text{RR}}-\hat{\mathbf{w}}_{\text{JBLCMV}}^H\hat{\mathbf{P}}\hat{\mathbf{w}}_{\text{JBLCMV}}}{\hat{p}_{\text{LL}}+\hat{p}_{\text{RR}}-\hat{\mathbf{w}}_{\text{BMVDR}}^H\hat{\mathbf{P}}\hat{\mathbf{w}}_{\text{BMVDR}}},\qquad(16)$$

which is only defined for $r\leq 2M-2$, since the JBLCMV filter is only defined in this case. If we select $\phi=\phi_{\text{JBLCMV}}$, we will constrain the problem in (14) to have at least the noise reduction performance of the JBLCMV. The feasibility set of the problem in (14), for a certain choice ϕ, is given by

$$\mathcal{S}(\phi)=\{\mathbf{w}:\ \ \mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{A}}=\hat{\mathbf{f}}_{\text{A}}^H\ \ \cap\ \ \mathbf{w}^H\hat{\mathbf{P}}\mathbf{w}\leq\epsilon(\phi)\}.\qquad(17)$$

Now we provide two properties that will be useful in the sequel to better understand the problem in (14).

Property 1. *The matrix $\hat{\mathbf{\Lambda}}_B\hat{\mathbf{\Lambda}}_B^H\succeq 0$ is rank deficient, because the last row is minus the first one, so that $||\mathbf{w}^H\hat{\mathbf{\Lambda}}_B||_2^2$ has multiple minimizers. In this situation, the QCQP in (14) may have multiple solutions. If there are multiple solutions, they are all optimal.*

Property 2. $\forall\phi_1,\phi_2\in[0,1]$, $\phi_1\leq\phi_2\iff\epsilon(\phi_1)\geq\epsilon(\phi_2)\iff\mathcal{S}(\phi_2)\subseteq\mathcal{S}(\phi_1)$.

Corollary 1. *Let $r\leq 2M-2$. $\forall\phi\in[0,\phi_{JBLCMV}]$, $\hat{\mathbf{w}}_{JBLCMV}\in\mathcal{S}(\phi)$.*

Proof. Since $r\leq 2M-2$, $\hat{\mathbf{w}}_{\text{JBLCMV}}$ will be realizable (see (12)) and, thus, $\hat{\mathbf{w}}_{\text{JBLCMV}}\in\mathcal{S}(\phi_{\text{JBLCMV}})$. Hence, by Property 2, we have $\forall\phi\in[0,\phi_{\text{JBLCMV}}]$, $\mathcal{S}(\phi_{\text{JBLCMV}})\subseteq\mathcal{S}(\phi)$ and, thus, $\hat{\mathbf{w}}_{\text{JBLCMV}}\in\mathcal{S}(\phi)$. □

Proposition 1. *Let $r\leq 2M-2$. For all $\phi\in[0,\phi_{JBLCMV}]$, $\hat{\mathbf{w}}_{prop}=\hat{\mathbf{w}}_{JBLCMV}$ is a solution of the QCQP problem in (14).*

Proof. Since $r\leq 2M-2$, $\hat{\mathbf{w}}_{\text{JBLCMV}}$ will be realizable (see (12)). Since $||\hat{\mathbf{w}}_{\text{JBLCMV}}^H\hat{\mathbf{\Lambda}}_{\text{B}}||_2^2=0$ (see (10)) and $\forall\mathbf{w}$, $||\mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{B}}||_2^2\geq 0$, and since by Corollary 1, we have $\forall\phi\in[0,\phi_{\text{JBLCMV}}]$, $\hat{\mathbf{w}}_{\text{JBLCMV}}\in\mathcal{S}(\phi)$, a solution will be $\hat{\mathbf{w}}_{\text{prop}}=\hat{\mathbf{w}}_{\text{JBLCMV}},\forall\phi\in[0,\phi_{\text{JBLCMV}}]$. □

In words, Proposition 1 says that when there are enough degrees of freedom (i.e., for $r\leq 2M-2$), and the output noise is not constrained to be very small (i.e., $\phi\leq\phi_{\text{JBLCMV}}, e(\phi)\geq e(\phi_{\text{JBLCMV}})$), then the binaural-cue distortions will be zero and one of the solutions of the the problem in (14) will be the JBLCMV solution. This will not only happen for $\phi=\phi_{\text{JBLCMV}}$, but also for all smaller ϕ values since the binaural-cue distortions cannot be reduced any further. Recall that the problem formulation in (14) may have multiple optimal solutions (see Property 1). For instance, for $\phi=0$, both the $\hat{\mathbf{w}}_{\text{JBLCMV}}$ and $\mathbf{e}$ are optimal. In binaural beamforming, one would like the best trade-off between noise reduction and binaural-cue preservation and, therefore, among all possible solutions of the problem in (14) for $\phi\in[0,\phi_{\text{JBLCMV}}]$, we always select the $\hat{\mathbf{w}}_{\text{JBLCMV}}$ solution, which is better suited to our goal since it provides the minimum possible output noise.

Remark 1. *If $r>2M-2$ and $\phi=0$, the solution of the problem in (14) is selected as $\hat{\mathbf{w}}_{prop}=\mathbf{e}$.*

The problem in (14) may have multiple solutions (see Property 1). For $r>2M-2$ and $\phi=0$, we select the solution $\hat{\mathbf{w}}_{\text{prop}}=\mathbf{e}$, since it is the only known closed-form solution of the problem in (14) for $r>2M-2$ and $\phi=0$. Thus, by selecting $\hat{\mathbf{w}}_{\text{prop}}=\mathbf{e}$ the implementation is faster compared to an iteratively obtained solution from the proposed method.

4.1. Solver for the Proposed Problem

The Lagrangian of the problem in (14) is given by

$$L(\mathbf{w},\boldsymbol{\mu},\lambda)=\mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{B}}\hat{\mathbf{\Lambda}}_{\text{B}}^H\mathbf{w}+\lambda\left(\mathbf{w}^H\hat{\mathbf{P}}\mathbf{w}-\epsilon(\phi)\right)$$
$$+\Re\{\boldsymbol{\mu}^H(\hat{\mathbf{\Lambda}}_{\text{A}}^H\mathbf{w}-\hat{\mathbf{f}}_{\text{A}})\},\qquad(18)$$

where $\Re$ denotes the real part of a complex vector. Taking the gradient of (18) with respect to $\mathbf{w}$ and setting this equal to zero we obtain

$$\hat{\mathbf{w}}(\lambda,\boldsymbol{\mu})=-\frac{1}{2}\left(\hat{\mathbf{\Lambda}}_{\text{B}}\hat{\mathbf{\Lambda}}_{\text{B}}^H+\lambda\hat{\mathbf{P}}\right)^{-1}\hat{\mathbf{\Lambda}}_{\text{A}}\boldsymbol{\mu}.\qquad(19)$$

Combining $\mathbf{w}^H\hat{\mathbf{\Lambda}}_{\text{A}}=\hat{\mathbf{f}}_{\text{A}}^H$ and (19), we obtain

$$\boldsymbol{\mu}=-2\left(\hat{\mathbf{\Lambda}}_{\text{A}}^H(\hat{\mathbf{\Lambda}}_{\text{B}}\hat{\mathbf{\Lambda}}_{\text{B}}^H+\lambda\hat{\mathbf{P}})^{-1}\hat{\mathbf{\Lambda}}_{\text{A}}\right)^{-1}\hat{\mathbf{f}}_{\text{A}}.\qquad(20)$$

Combining (20) and (19) we obtain

$$\hat{\mathbf{w}}(\lambda)=\mathbf{T}(\lambda)\left(\hat{\mathbf{\Lambda}}_{\text{A}}^H\mathbf{T}(\lambda)\right)^{-1}\hat{\mathbf{f}}_{\text{A}},\qquad(21)$$

where

$$\mathbf{T}(\lambda)=\left(\hat{\mathbf{\Lambda}}_{\text{B}}\hat{\mathbf{\Lambda}}_{\text{B}}^H+\lambda\hat{\mathbf{P}}\right)^{-1}\hat{\mathbf{\Lambda}}_{\text{A}}.\qquad(22)$$

We can use bisection [13] to find the optimal $\hat{\lambda}\geq 0$, such that $\hat{\mathbf{w}}_{\text{prop}}=\hat{\mathbf{w}}(\hat{\lambda})$, where $\hat{\mathbf{w}}(\hat{\lambda})$ is given in (21). More specifically, the bisection method initially selects λ_{a}, λ_{b} such that $\hat{\mathbf{w}}(\lambda_{\text{a}})^H\hat{\mathbf{P}}\hat{\mathbf{w}}(\lambda_{\text{a}})\geq\epsilon(\phi)$ and $\hat{\mathbf{w}}(\lambda_{\text{b}})^H\hat{\mathbf{P}}\hat{\mathbf{w}}(\lambda_{\text{b}})\leq\epsilon(\phi)$, after which computes $\lambda_{\text{c}}=(\lambda_{\text{a}}+\lambda_{\text{b}})/2$. If $\hat{\mathbf{w}}(\lambda_{\text{c}})^H\hat{\mathbf{P}}\hat{\mathbf{w}}(\lambda_{\text{c}})\geq\epsilon(\phi)$, it assigns to λ_{c} the λ_{a} value, while if $\hat{\mathbf{w}}(\lambda_{\text{c}})^H\hat{\mathbf{P}}\hat{\mathbf{w}}(\lambda_{\text{c}})\leq\epsilon(\phi)$, then it assigns to λ_{c} the λ_{b} value. It keeps iterating until $|\lambda_{\text{a}}-\lambda_{\text{b}}|\leq t$, where t is a small threshold. A warm-start procedure is used for the initial choices of λ_{a} and λ_{b} for each frequency bin by using the converged λ_{a} and λ_{b} of the previous frequency bin. If these warm-started λ_{a} and λ_{b} do not satisfy the initial conditions mentioned above, they are increased and/or reduced gradually such that they do so. This warm-start approach typically reduces the average number of iterations that bisection takes to converge per frequency bin and combined with the fact that the proposed filter has a closed-form solution as a function of a single variable (the Lagrange multiplier λ) leads to a fast implementation.

Recall that, if $r \leq 2M-2$ and $\phi \in [0, \phi_{\text{JBLCMV}}]$, the JBLCMV solution will be a solution to the proposed optimization problem. In order to speed up the computations even further, we use the closed-form solution in (12), instead of the iterative approach explained just above, to obtain the solution of the proposed problem for this range of ϕ and r values/settings. For $r > 2M-2$ we cannot do this, since the JBLCMV solution is unrealizable and, thus, the proposed filter will be computed iteratively as explained before. For $\phi = 0$ and $r > 2M-2$, we do not process the microphone signals (see Remark 1). Note that, unlike the JBLCMV filter which does not have a solution for $r > 2M-2$, the proposed framework guarantees a solution for any number r.

5. EXPERIMENTS

In this section, we compare the methods reviewed in Sec. 3 and the proposed method in Sec. 4. We compare the trade-offs between noise reduction/predicted intelligibility and directional binaural-cue distortions and the different computational times for each method. We compute the directional binaural-cue distortions as measured from the interaural level differences (ILDs) and interaural phase differences (IPDs) as in [12], averaged over frequency and sources. The ILD distortions were averaged for frequencies above 3 kHz and the IPD distortions were averaged below 1.5 kHz, since these ranges are perceptually relevant for localization [14]. The noise reduction was computed as the maximum of the segmental-signal-to-noise-ratio (SSNR) at the two ears, to take into account the better ear effect. Similarly, we determine the predicted intelligibility by calculating SIIB [17] at both ears and take the maximum to take the better ear effect into account. We use SIIB as our scene includes both point noise sources as well as reverberation and SIIB has shown to be a good intelligibility predictor in such cases [18].

The acoustic scene consists of five sources in total (1 m away from the head center), from which four sources are interferers (male talker “A” at 80^{o}, music signal at 50^{o}, vacuum cleaner at -35^{o}, and male talker “B” at -80^{o}) and a target (female talker at 0^{o}). The SNR of the target signal at its original location with respect to each interferer signal at its original location is 10 dB. The total number of microphones is $M = 4$ (two microphones per HAD, i.e., $M_{\text{L}} = M_{\text{R}} = 2$). The head impulse responses from the office environment from the database in [19] were used to construct the noisy signals. All methods used estimated ATFs using a simple estimation method. To do so, we gathered 5 s of each recorded point source in isolation (including its late reverberation and microphone-self noise) at the microphone array and we estimated its CPSDM. Then we assigned as its ATF, the eigenvector corresponding to the largest eigenvalue. The binaural filters from all methods were computed only once using a sample average noise CPSDM estimated using 5 s of noise-only microphone signals. We used the overlap-and-add method [20] with a square-root Hann window for the analysis and synthesis and a frame-length of 10 ms with an overlap 50%.

Fig. 1 shows the trade-off between noise reduction and directional binaural-cue distortions and the trade-off between predicted intelligibility and directional binaural-cue distortions as a function of η, c, ϕ, which are the trade-off parameters of the BMVDR-η, the RJBLCMV and the proposed method, respectively. It is clear from the results that for a certain amount of binaural-cue distortion, the proposed method and the RJBLCMV method provide a better noise reduction and predicted intelligibility compared to the BMVDR-η approach. In Fig. 2, the proposed method achieves a significantly lower computational load than the RJBLCMV. Remarkably, for $c, \phi \in (0, 1)$, the proposed approach is at least 4500 and at most 18700 times faster compare to the RJBLCMV method. Note however that the RJBLCMV filter is computed via CVX toolbox [21] and we expect that this difference will be smaller using a solver optimized for this problem. Moreover, for $\eta, \phi \in (0, 1)$, the proposed method is only 2 to 22 times slower compared to the BMVDR-η approach. Note that in Fig. 1, the JBLCMV did not preserve exactly the binaural-cues, due to estimation errors in the ATFs of the sources.

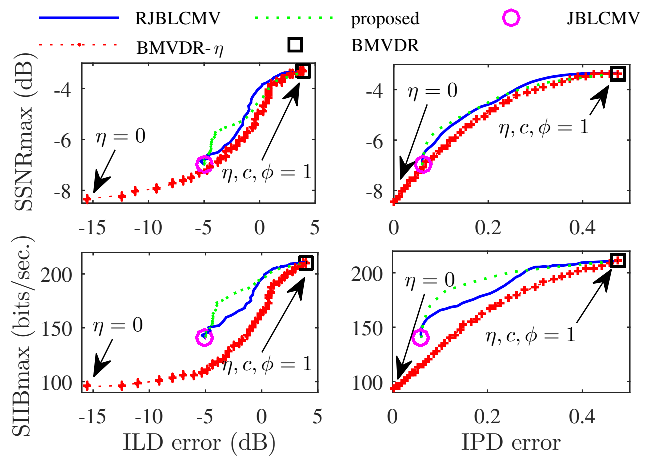

Fig. 1. SSNR vs ILD and IPD errors, and SIIB vs ILD and IPD errors.

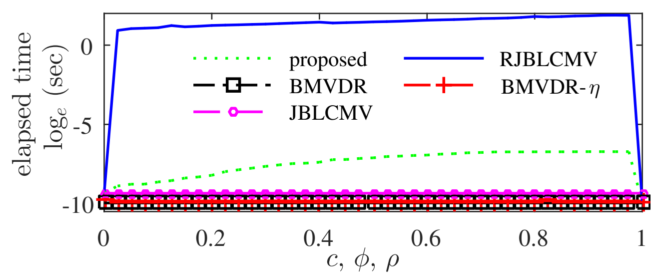

Fig. 2. Average (over frequency) computational time in log scale.

6. CONCLUSION

We proposed a binaural beamformer which provides an efficient trade-off between noise reduction, binaural-cue preservation and complexity. If there are enough degrees of freedom available, the proposed approach has as boundary solutions the BMVDR and JBLCMV, otherwise it has the BMVDR and unprocessed scene. In our simulation experiments, the proposed method shows a better noise reduction and predicted intelligibility performance given a certain amount of binaural-cue distortion compared to the BMVDR-η approach, and a significant computational complexity advantage over the RJBLCMV method.

7. REFERENCES

[1] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*, John Wiley & Sons, 2004.

[2] M. Brandstein and D. Ward (Eds.), *Microphone arrays: signal processing techniques and applications*, Springer, 2001.

[3] J. M. Kates, *Digital hearing aids*, Plural publishing, 2008.

[4] S. Doclo, W. Kellermann, S. Makino, and S. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[5] 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.

[6] J. Thiemann, M. Müller, D. Marquardt, S. Doclo, and S. van de Par, "Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene," *EURASIP J. Advances Signal Process.*, 2016.

[7] A. I. Koutrouvelis, J. Jensen, M. Guo, R. C. Hendriks, and R. Heusdens, "Binaural speech enhancement with spatial cue preservation utilising simultaneous masking," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, Aug. 2017.

[8] D. Marquardt and S. Doclo, "Interaural coherence preservation for binaural noise reduction using partial noise estimation and spectral postfiltering," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 7, pp. 1261–1274, 2018.

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

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

[11] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, "Improved multi-microphone noise reduction preserving binaural cues," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2016.

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

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[14] W. M. Hartmann, "How we localize sound," *Physics Today*, vol. 52, no. 11, pp. 24–29, Nov. 1999.

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

[16] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical analysis of linearly constrained multi-channel Wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids," *IEEE Trans. Audio, Speech, Language Process.*, vol. 23, no. 12, Sept. 2015.

[17] S. Van Kuyk, W. B. Kleijn, and R. C. Hendriks, "An instrumental intelligibility metric based on information theory," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 115–119, Jan. 2018.

[18] S. Van Kuyk, W. B. Kleijn, and R. C. Hendriks, "An evaluation of intrusive instrumental intelligibility metrics," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 11, pp. 2153–2166, 2018.

[19] H. Kayser, S. Ewert, J. Annemüller, 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. Advances Signal Process.*, vol. 2009, pp. 1–10, Dec. 2009.

[20] J. Allen, "Short-term spectral analysis, and modification by discrete Fourier transform," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 25, no. 3, pp. 235–238, June 1977.

[21] "Cvx: Matlab software for disciplined convex programming.," 2008.