A Convex Approximation of the Relaxed Binaural Beamfomring Optimization Problem

Koutrouvelis, Andreas I.; Hendriks, Richard Christian; Heusdens, Richard; Jensen, Jesper

DOI
10.1109/TASLP.2018.2878618

Publication date
2019

Document Version
Accepted author manuscript

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

Citation (APA)
Koutrouvelis, A. I., Hendriks, R. C., Heusdens, R., & Jensen, J. (2019). A Convex Approximation of the Relaxed Binaural Beamfomring Optimization Problem. *IEEE/ACM Transactions on Audio Speech and Language Processing*, *27*(2), 321-331. [8514022]. https://doi.org/10.1109/TASLP.2018.2878618

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

A Convex Approximation of the Relaxed Binaural Beamforming Optimization Problem

Andreas I. Koutrouvelis, Richard C. Hendriks, Richard Heusdens and Jesper Jensen

***Abstract*—The recently proposed relaxed binaural beamforming (RBB) optimization problem provides a flexible trade-off between noise suppression and binaural-cue preservation of the sound sources in the acoustic scene. It minimizes the output noise power, under the constraints which guarantee that the target remains unchanged after processing and the binaural-cue distortions of the acoustic sources will be less than a user-defined threshold. However, the RBB problem is a computationally demanding non-convex optimization problem. The only existing suboptimal method which approximately solves the RBB is a successive convex optimization (SCO) method which, typically, requires to solve multiple convex optimization problems per frequency bin, in order to converge. Convergence is achieved when all constraints of the RBB optimization problem are satisfied. In this paper, we propose a semi-definite convex relaxation (SDCR) of the RBB optimization problem. The proposed suboptimal SDCR method solves a single convex optimization problem per frequency bin, resulting in a much lower computational complexity than the SCO method. Unlike the SCO method, the SDCR method does not guarantee user-controlled upper-bounded binaural-cue distortions. To tackle this problem we also propose a suboptimal hybrid method which combines the SDCR and SCO methods. Instrumental measures combined with a listening test show that the SDCR and hybrid methods achieve significantly lower computational complexity than the SCO method, and in most cases better trade-off between predicted intelligibility and binaural-cue preservation than the SCO method.**

***Index Terms*—Binaural beamforming, binaural cues, convex optimization, LCMV, noise reduction, semi-definite relaxation.**

I. Introduction

BINAURAL beamforming (see e.g., [1] for an overview), also known as binaural spatial filtering, plays an important role in binaural hearing-aid (HA) systems [2]. Binaural beamforming is typically described as an optimization problem, where the objective is to i) minimize the output noise power, ii) preserve the target sound source at the left and right HA reference microphone, and iii) preserve the binaural cues of all sound sources after processing. The microphone array, which is typically mounted on the HA devices, has only a few microphones and, thus, there is only limited freedom (i.e., a small feasibility set) to search for a good compromise between the three aforementioned goals. Besides the challenge in finding a good trade-off among all these goals, the complexity should remain as low as possible, due to the limited computational power of the HA devices.

The binaural minimum variance distortionless response (BMVDR) beamformer (BF) [1] provides the maximum possible noise suppression among all binaural target-distortionless BFs [3]. Unfortunately, the BMVDR severely distorts the binaural-cues of the residual noise at the output of the filter. Specifically, the residual noise inherits the interaural transfer function of the target and, hence, sounds as originating from the target's direction [1]. The lack of spatial separation between the target and the noise after processing, may not only provide an unnatural impression to the user, but may also negatively effect the intelligiblity [4]. In [5], [6], the BMVDR was compared with an oracle-based (i.e., non-practically implementable) method in several noise fields (diffuse [5] and diffuse plus directional [6]). The oracle-based method has the same noise suppression as the BMVDR, but does not cause any binaural-cue distortions of the acoustic scene. The spatially correct oracle-based method achieved an improvement of about 3 dB in the 50% speech reception threshold (SRT) over the BMVDR. Therefore, there are several reasons to seek for methods that simultaneously provide the maximum possible noise suppression and binaural-cue preservation of all sources in the acoustic scene.

Several modifications of the BMVDR BF have been proposed, which can be roughly categorized into two groups. The first group consists of BFs that add or maintain a portion of the unprocessed scene at the output of the filter (see e.g., [5], [7]–[10]). An interesting approach, which is referred to as BMVDR-η [10], adds a portion of the unprocessed scene to the output of the BMVDR BF such that the binaural cues of the noise will be preserved in a certain extent. The second group consists of BFs, whose optimization problems have the same objective function as the BMVDR, but introduce extra equality [3], [11], [12] or inequality [13] constraints in order to preserve the binaural cues of the interferers after processing. These constraints are functions of either i) the (relative) acoustic transfer functions (R)(ATFs) of the interferers which can be estimated (see e.g., [14] for an overview), or ii) pre-determined anechoic (R)ATFs forming a grid around the head of the user as proposed in [15]. Moreover, these additional constraints in the optimization problem results in less degrees of freedom for noise reduction. With equality constraints, closed-form solutions may be derived, but the degrees of freedom can be easily exhausted when multiple interferers exist in the acoustic scene, resulting in poor noise reduction. On the other hand, inequality constraints provide more flexibility and can approximately preserve the binaural cues of, typically, many more acoustic sources, or for the same number of acoustic sources provide a larger amount of noise reduction [13]. Unfortunately, closed-form solutions do not exist for the inequality-constrained binaural BFs and, thus, iterative methods with a larger complexity are used instead.

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

Recently, the relaxed binaural beamforming (RBB) optimization problem was proposed, which uses inequality constraints to preserve the binaural cues of the interfering sources [13]. The inequality constraints in the RBB are not convex, resulting in a non-convex optimization problem. In [13], a suboptimal successive convex optimization (SCO) method was proposed to approximately solve the RBB problem. In most cases, the SCO method needs to solve more than one convex optimization problem, per frequency bin, in order to converge. Convergence is achieved when all constraints of the RBB problem are satisfied. As a result, the SCO method guarantees an upper-bounded binaural-cue distortion of the interferers (as expressed by the interaural transfer function error), where the upper bound is controlled by the user.

Unfortunately, the SCO method is computationally very demanding due to its need to solve multiple convex optimization problems, per frequency bin, in order to converge. In this paper, we propose a semi-definite convex relaxation (SDCR) of the RBB optimization problem, which is significantly faster than the SCO method. This is because, the SDCR method requires to solve only one convex optimization problem per frequency bin. The main drawback of the SDCR method is that it does not guarantee user-controlled upper-bounded binaural-cue distortions as the SCO method. We solve this issue by combining the SDCR and SCO methods into a suboptimal hybrid method. The hybrid method guarantees user-controlled upper-bounded binaural-cue distortions, and still has a significantly lower computational complexity than the SCO method. Simulation experiments combined with listening tests show that both proposed methods, in most cases, provide a better trade-off between predicted intelligibility and binaural-cue preservation than the SCO method.

II. Signal Model and Notation

We assume that there is one target point-source signal, r point-source interferers, background noise, and two HAs with M microphones in total. The processing is accomplished per time-frequency bin independently. Neglecting time-frequency indices for brevity, the acquired M-element noisy vector in the DFT domain, for a single time-frequency bin, is given by

$$\mathbf{y} = \underbrace{s\mathbf{a}}_{\mathbf{x}} + \underbrace{\sum_{i=1}^{r} v_i \mathbf{b}_i + \mathbf{u}}_{\mathbf{n}} \in \mathbb{C}^{M\times 1}, \tag{1}$$

where s and v_i are the target and i-th interferer signals at the original locations; $\mathbf{a}$ and $\mathbf{b}_i$ the early acoustic transfer function (ATF) vectors of the target and i-th interferer, respectively; $\mathbf{u}$ the background noise, and $\mathbf{n}$ the total additive noise. The background noise is due to the diffuse late reverberation from all point sources and the microphone-self noise. Assuming statistical independence between all sources, the noisy cross-power spectral density matrix (CPSDM) is given by

$$\mathbf{P}_{\mathbf{y}} = \mathrm{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{P}_{\mathbf{x}} + \mathbf{P}_{\mathbf{n}} \in \mathbb{C}^{M\times M}, \tag{2}$$

with $\mathbf{P}_{\mathbf{x}} = \mathrm{E}[\mathbf{x}\mathbf{x}^H] = p_s\mathbf{a}\mathbf{a}^H$ and $\mathbf{P}_{\mathbf{n}} = \mathrm{E}[\mathbf{n}\mathbf{n}^H]$ the target and noise CPSDMs, respectively, and $p_s = \mathrm{E}[|s|^2]$ the power spectral density of the target signal.

III. Binaural Beamforming Preliminaries

Binaural BFs consist of two spatial filters, $\mathbf{w}_L$, $\mathbf{w}_R \in \mathbb{C}^{M\times 1}$, which are both applied to the noisy measurements producing two different outputs given by

$$\begin{bmatrix}\hat{x}_L\\ \hat{x}_R\end{bmatrix} = \begin{bmatrix}\mathbf{w}_L^H\mathbf{y}\\ \mathbf{w}_R^H\mathbf{y}\end{bmatrix}, \tag{3}$$

where $\hat{x}_L, \hat{x}_R$ are played back by the loudspeakers of the left and right HAs, respectively. Note that the subscripts L and R are also used to refer to the two elements of the vectors in (1) associated with the left and right reference microphones of the binaural BF. Here, we select the first and the M-th microphones as reference microphones and, thus, $y_L = y_1$ and $y_R = y_M$. The same applies to all vectors in (1).

All BFs considered in this paper are target-distortionless. Their goal is not only noise supression, but also preservation of the binaural cues of all sources in the acoustic scene. In this paper, we mainly focus on preserving, after processing, the perceived direction of all point sources. Therefore, in the following, we mean directional binaural cues when we use the term binaural cues. A simple way of measuring the binaural cues of a source is via the interaural transfer function (ITF), which is a function of the ATF vector of the source [16]. The ITF of the i-th interferer before and after applying the spatial filter is given by [16]

$$\mathrm{ITF}_i^{\mathrm{in}} = \frac{b_{iL}}{b_{iR}}, \quad \mathrm{ITF}_i^{\mathrm{out}} = \frac{\mathbf{w}_L^H\mathbf{b}_i}{\mathbf{w}_R^H\mathbf{b}_i}. \tag{4}$$

The input and output ITF of the target is expressed similarly. Ideally, to preserve the binaural cues of the point sources, a binaural BF will produce the same ITF output as the input for all point sources. In practice, this is very difficult to achieve, when the number of interferers, r, is large and the number of microphones, M, is small [13]. As a result, most BFs will introduce some distortion to the ITF output, resulting in a non-zero ITF error given by [13]

$$\mathrm{ITF}_i^{\mathrm{e}} = \left|\mathrm{ITF}_i^{\mathrm{out}} - \mathrm{ITF}_i^{\mathrm{in}}\right| = \left|\frac{\mathbf{w}_L^H\mathbf{b}_i}{\mathbf{w}_R^H\mathbf{b}_i} - \frac{b_{iL}}{b_{iR}}\right| \geq 0. \tag{5}$$

A. BMVDR Beamforming

The BMVDR BF [1] achieves the maximum possible noise suppression among all binaural BFs and is obtained from the following simple optimization problem [1], [3]:

$$\begin{aligned}\hat{\mathbf{w}}_L, \hat{\mathbf{w}}_R = \underset{\mathbf{w}_L,\mathbf{w}_R}{\arg\min} &\begin{bmatrix}\mathbf{w}_L^H & \mathbf{w}_R^H\end{bmatrix}\tilde{\mathbf{P}}\begin{bmatrix}\mathbf{w}_L\\ \mathbf{w}_R\end{bmatrix}\\ \text{s.t.} \quad &\mathbf{w}_L^H\mathbf{a} = a_L \quad \mathbf{w}_R^H\mathbf{a} = a_R,\end{aligned} \tag{6}$$

where

$$\tilde{\mathbf{P}} = \begin{bmatrix}\mathbf{P}_{\mathbf{n}} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{\mathbf{n}}\end{bmatrix}. \tag{7}$$

The optimization problem in (6) provides closed-form solutions to the left and right spatial filters given by [1], [3]

$$\hat{\mathbf{w}}_L = \frac{\mathbf{P}_{\mathbf{n}}^{-1}\mathbf{a}a_L^*}{\mathbf{a}^H\mathbf{P}_{\mathbf{n}}^{-1}\mathbf{a}}, \quad \hat{\mathbf{w}}_R = \frac{\mathbf{P}_{\mathbf{n}}^{-1}\mathbf{a}a_R^*}{\mathbf{a}^H\mathbf{P}_{\mathbf{n}}^{-1}\mathbf{a}}. \tag{8}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

It can easily be shown, that the output ITF of the i-th interferer of the BMVDR spatial filter is given by [3], [13]

$$\text{ITF}_i^{\text{out}} = \frac{a_L}{a_R}, \tag{9}$$

which is the ITF input of the target. Therefore, all interferers sound as coming from the target direction after applying the BMVDR spatial filter. The BMVDR ITF error of the i-th interferer is given by [13]

$$\text{ITF}_i^{\text{e,BMVDR}} = \left| \frac{a_L}{a_R} - \frac{b_{iL}}{b_{iR}} \right|. \tag{10}$$

B. Relaxed Binaural Beamforming

The relaxed binaural beamforming (RBB) optimization problem, introduced in [13], uses additional inequality constraints (compared to the BMVDR problem) to preserve the interferers' binaural cues. The RBB problem is given by [13]

$$\begin{aligned} \hat{\mathbf{w}}_L, \hat{\mathbf{w}}_R = \underset{\mathbf{w}_L, \mathbf{w}_R}{\arg\min} & \begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix} \tilde{\mathbf{P}} \begin{bmatrix} \mathbf{w}_L \\ \mathbf{w}_R \end{bmatrix} \\ \text{s.t.} \quad & \mathbf{w}_L^H \mathbf{a} = a_L \quad \mathbf{w}_R^H \mathbf{a} = a_R, \\ & \left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} - \frac{b_{iL}}{b_{iR}} \right| \leq \mathcal{E}_i, \ i = 1, \cdots, m \leq r, \end{aligned} \tag{11}$$

where

$$\mathcal{E}_i = c_i \text{ITF}_i^{\text{e,BMVDR}}, \quad 0 \leq c_i \ \leq 1.$$

Note that $\mathcal{E}_i$ is c_i times the ITF error of the i-th interferer of the BMVDR BF [13]. Recall that the BMVDR causes full collapse of the binuaral cues of the interferers towards the binaural cues of the target. Therefore, the inequality constraints in (11) control the percentage of collapse. A small c_i implies good preservation of binaural cues of the i-th interferer, but a smaller feasibility set and, thus, less noise reduction. On the other hand, a large c_i implies worse binaural-cue preservation, but more noise reduction.

It is clear from the above that the additional inequality constraints of the RBB problem require the knowledge of the (R)ATF vectors of the interferers. In practice, interferers' (R)ATF vectors are unknown and estimation is required. Several methods for estimating RATF vectors exist (see e.g., [14] for an overview). An alternative approach is to use pre-determined ancechoic (R)ATF vectors of fixed azimuths around the head of the user, as proposed in [15]. These pre-determined (R)ATF vectors are acoustic scene independent and need to be obtained once for each user. This is useful when the (R)ATF vectors of the interferers are difficult to estimate, because e.g., the locations of the interferers relative to the head of the user are non-static. It is worth noting that by using pre-determined (R)ATF vectors, a larger number of inequality constraints, $m > r$, is typically used in (11). This is because we do not know where the interferers are located and we would like to cover the entire space around the head of the user.

If $c_i > 0, i = 1, \cdots, m$, the inequality constraints of the optimization problem in (11) are non-convex. As a result, the optimization problem in (11) is non-convex. In [13], a sub-optimal successive convex optimization (SCO) method [13], described in Sec. III-C, was proposed to approximately solve the RBB problem.

C. Successive Convex Optimization method

The successive convex optimization (SCO) method [13] approximately solves the RBB problem by solving multiple second-order cone program (SOCP) convex optimization problems per frequency bin. The SCO method converges, when all constraints of the RBB problem in (11) are satisfied. It has been shown that the SCO method always converges to a solution satisfying the constraints of the RBB problem if $m \leq 2M - 3$. This means that if the (R)ATF vectors of the interferers have been estimated accurately enough, the SCO method will guarantee user-controlled upper-bounded ITF error of the interferers [13]. For $m > 2M - 3$, no guarantees exist for convergence. In case the method does not converge, it stops after solving a pre-defined maximum number of convex optimization problems, $k_{\max}$. Nevertheless, for a reasonable number of inequality constraints, m, it has been experimentally shown that the SCO method always converges [13], [15]. It has been experimentally shown in [13], that for larger c_i values, the SCO method converges to solutions further away from the boundary of the inequality constraints of the RBB problem. This results in a better binaural-cue preservation and less noise reduction compared to the expected trade-off set by the user through the parameters $c_i, i = 1, \cdots, m$.

IV. Proposed Convex Approximation Method

The proposed method is a semi-definite convex relaxation (SDCR) of the optimization problem in (11). First, we review two important properties that will be useful for understanding the proposed optimization problem.

Property 1: Any quadratic expression can be expressed as [17]

$$\mathbf{q}^H \mathbf{Z} \mathbf{q} = \text{tr}\left(\mathbf{q}^H \mathbf{Z} \mathbf{q}\right) = \text{tr}\left(\mathbf{q}\mathbf{q}^H \mathbf{Z}\right). \tag{12}$$

Property 2: We have the following equivalence relation [18]

$$\begin{aligned} & \mathbf{Z} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{C} \end{bmatrix} \succeq 0 \Leftrightarrow \\ & \mathbf{A} \succeq 0, \quad \left(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger\right)\mathbf{B} = \mathbf{0}, \quad \mathbf{S}_1 \succeq 0, \qquad (13) \\ & \mathbf{C} \succeq 0, \quad \left(\mathbf{I} - \mathbf{C}\mathbf{C}^\dagger\right)\mathbf{B}^H = \mathbf{0}, \quad \mathbf{S}_2 \succeq 0, \qquad (14) \end{aligned}$$

with $\mathbf{S}_1 = \mathbf{C} - \mathbf{B}^H \mathbf{A}^\dagger \mathbf{B}$ the generalized Schur complement of $\mathbf{A}$ in $\mathbf{Z}$, $\mathbf{S}_2 = \mathbf{A} - \mathbf{B}\mathbf{C}^\dagger \mathbf{B}^H$ the generalized Schur complement of $\mathbf{C}$ in $\mathbf{Z}$, and $\mathbf{A}^\dagger$ is the pseudo-inverse of $\mathbf{A}$ [19].

Before, we present the proposed convex optimization problem, we first introduce an equivalent optimization problem to the problem in (11). That is,

$$\begin{aligned} \hat{\mathbf{w}}_L, \hat{\mathbf{w}}_R = \underset{\mathbf{w}_L, \mathbf{w}_R}{\arg\min} & \begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix} \tilde{\mathbf{P}} \begin{bmatrix} \mathbf{w}_L \\ \mathbf{w}_R \end{bmatrix} \\ \text{s.t.} \quad & \mathbf{w}_L^H \mathbf{a} = a_L \quad \mathbf{w}_R^H \mathbf{a} = a_R, \\ & \left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} - \frac{b_{iL}}{b_{iR}} \right|^2 \leq \mathcal{E}_i^2, \ i = 1, \cdots, m \leq r. \end{aligned} \tag{15}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

By reformulating the inequality in (15), we obtain an equivalent quadratic constraint given by

$$\left|\frac{\mathbf{w}_L^H\mathbf{b}_i}{\mathbf{w}_R^H\mathbf{b}_i}-\frac{b_{iL}}{b_{iR}}\right|^2\leq\mathcal{E}_i^2\Rightarrow \underbrace{\begin{bmatrix}\mathbf{w}_L^H & \mathbf{w}_R^H\end{bmatrix}}_{\mathbf{w}^H}\underbrace{\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}^H & \mathbf{C}\end{bmatrix}}_{\mathbf{M}_i}\underbrace{\begin{bmatrix}\mathbf{w}_L\\ \mathbf{w}_R\end{bmatrix}}_{\mathbf{w}}\leq 0, \tag{16}$$

where $\mathbf{A}=|b_{iR}|^2\mathbf{b}_i\mathbf{b}_i^H$, $\mathbf{B}=-b_{iL}^*b_{iR}\mathbf{b}_i\mathbf{b}_i^H$, $\mathbf{C}=\left(|b_{iL}|^2-|b_{iR}|^2\mathcal{E}_i^2\right)\mathbf{b}_i\mathbf{b}_i^H$. Therefore, the optimization problem in (15) can be re-written as

$$\begin{aligned}\hat{\mathbf{w}}=\arg\min_{\mathbf{w}}\ &\mathbf{w}^H\tilde{\mathbf{P}}\mathbf{w}\\ \text{s.t.}\quad &\mathbf{w}^H\begin{bmatrix}\mathbf{a} & \mathbf{0}\\ \mathbf{0} & \mathbf{a}\end{bmatrix}=\begin{bmatrix}a_L & a_R\end{bmatrix},\\ &\mathbf{w}^H\mathbf{M}_i\mathbf{w}\leq 0,\quad i=1,\cdots,m.\end{aligned} \tag{17}$$

The matrix $\mathbf{M}_i$ is not positive semi-definite and, therefore, the quadratic inequality constraint is not convex and, hence, the optimization problem in (17) is not convex. The proof of non positive semi-definiteness of $\mathbf{M}_i$ uses Property 2. Specifically, note that $\mathbf{A}\succeq 0$, but $\mathbf{S}_1=-|b_{iR}|^2\mathcal{E}_i^2\mathbf{b}_i\mathbf{b}_i^H\preceq 0$, because $\mathbf{b}_i\mathbf{b}_i^H\succeq 0$ and $-|b_{iR}|^2\mathcal{E}_i^2\leq 0$ and, therefore, $\mathbf{M}_i$ is not positive semi-definite.

The optimization problem in (17) is a non-convex quadratic-constrained quadratic program (QCQP) [18], [20]. Following the methodology described in [20], we use Property 1 to re-write the optimization problem in (17) into the following equivalent formulation:

$$\begin{aligned}\hat{\mathbf{w}},\hat{\mathbf{W}}=\arg\min_{\mathbf{w},\mathbf{W}}\ &\text{tr}\left(\mathbf{W}\tilde{\mathbf{P}}\right)\\ \text{s.t.}\quad &\mathbf{w}^H\begin{bmatrix}\mathbf{a} & \mathbf{0}\\ \mathbf{0} & \mathbf{a}\end{bmatrix}=\begin{bmatrix}a_L & a_R\end{bmatrix},\\ &\text{tr}\left(\mathbf{W}\mathbf{M}_i\right)\leq 0,\quad i=1,\cdots,m,\\ &\mathbf{W}=\mathbf{w}\mathbf{w}^H.\end{aligned} \tag{18}$$

The optimization problem in (18) is still not convex, but it has two differences with the problem in (17). The trace inequality is convex, but the new equality constraint, $\mathbf{W}=\mathbf{w}\mathbf{w}^H$ is not convex. Following [20], we apply the SDCR to the non-convex equality constraint of the problem in (18) and obtain the convex optimization problem given by

$$\begin{aligned}\hat{\mathbf{w}},\hat{\mathbf{W}}=\arg\min_{\mathbf{w},\mathbf{W}}\ &\text{tr}\left(\mathbf{W}\tilde{\mathbf{P}}\right)\\ \text{s.t.}\quad &\mathbf{w}^H\begin{bmatrix}\mathbf{a} & \mathbf{0}\\ \mathbf{0} & \mathbf{a}\end{bmatrix}=\begin{bmatrix}a_L & a_R\end{bmatrix},\\ &\text{tr}\left(\mathbf{W}\mathbf{M}_i\right)\leq 0,\quad i=1,\cdots,m.\\ &\mathbf{W}\succeq\mathbf{w}\mathbf{w}^H.\end{aligned} \tag{19}$$

Using Property 2, the inequality constraint $\mathbf{W}\succeq\mathbf{w}\mathbf{w}^H$ can be re-written as a linear matrix inequality, and the optimization problem in (19) can be re-written into a standard-form semi-definite program (SDP) [20]. That is,

$$\begin{aligned}\hat{\mathbf{w}},\hat{\mathbf{W}}=\arg\min_{\mathbf{w},\mathbf{W}}\ &\text{tr}\left(\mathbf{W}\tilde{\mathbf{P}}\right)\\ \text{s.t.}\quad &\mathbf{w}^H\begin{bmatrix}\mathbf{a} & \mathbf{0}\\ \mathbf{0} & \mathbf{a}\end{bmatrix}=\begin{bmatrix}a_L & a_R\end{bmatrix},\\ &\text{tr}\left(\mathbf{W}\mathbf{M}_i\right)\leq 0,\quad i=1,\cdots,m.\\ &\begin{bmatrix}\mathbf{W} & \mathbf{w}\\ \mathbf{w}^H & 1\end{bmatrix}\succeq 0.\end{aligned} \tag{20}$$

This is a convex problem, which can be solved efficiently [20]. If the solutions are on the boundary, i.e., $\hat{\mathbf{W}}=\hat{\mathbf{w}}\hat{\mathbf{w}}^H$, the minimizer, $\hat{\mathbf{w}}$, of the problem in (20) is also the minimizer of the non-convex RBB problem. This means, that in the case of $\hat{\mathbf{W}}=\hat{\mathbf{w}}\hat{\mathbf{w}}^H$, the proposed problem in (20) is optimal and obtain solutions which satisfy the inequalities in Eqs. (17), (15) (11). Otherwise, if $\hat{\mathbf{W}}\succ\hat{\mathbf{w}}\hat{\mathbf{w}}^H$, the solution of the problem in (20) may or may not satisfy the inequalities of the RBB, which means that we lose the guarantee for user-controlled upper-bounded ITF error when the (R)ATF vectors of the interferers have been estimated accurately enough. In practice, $\hat{\mathbf{W}}=\hat{\mathbf{w}}\hat{\mathbf{w}}^H$ never occurred in our experiments and, thus, the two problems do not produce exactly the same solutions. However, we will experimentally show in Sec. V that the SDCR method always stays relatively close to the boundary of the inequality constraints of the RBB problem implying that it is a good approximation of the RBB problem.

Finally, the main advantage of the new proposed SDCR method is that it reduces significantly the computational complexity compared to the SCO method. Although SOCP problems (which are solved in the SCO method) are less computational complex than SDP [21], we will experimentally show in Sec. V that the proposed SDCR method is much less computational complex since a single convex problem is solved compared to the many more convex problems that must be solved in the SCO method per frequency bin.

A. Proposed Hybrid Method

In this section, we propose a hybrid method, which is a combination of the SDCR and the SCO methods. If the (R)ATF vectors of the interferers are estimated accurately enough, the hybrid method guarantees user-controlled upper-bounded binaural-cue distortions of the interferers as the SCO method. The proposed hybrid method is significantly faster than the SCO method and slightly slower than the SDCR method. We will experimentally show in Sec. V, that the hybrid proposed method achieves solutions closer to the boundary of the inequality constraints of the RBB problem compared to the SCO method.

For a particular frequency bin, the hybrid method first solves the SDCR problem and then checks if there is a feasible solution which satisfies the inequality constraints of (11). If all of them are satisfied, the SDCR method will be used to approximately solve the RBB problem. Otherwise the SCO method is used to approximately solve the RBB problem in this frequency bin. Note that the SCO method always obtains a feasible solution for $m\leq 2M-3$ (see Sec. III-C) and, thus, the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

Algorithm 1: Hybrid scheme

$\hat{\mathbf{w}}_1 \leftarrow$ SDCR Problem in (20)
if $\hat{\mathbf{w}}_1$ satisfies (21) **then**
  return $\hat{\mathbf{w}}_1$
else
  $\hat{\mathbf{w}}_2 \leftarrow$ SCO method [13]
  return $\hat{\mathbf{w}}_2$
end if

same holds for the hybrid method. In such a way, the hybrid method will always have a feasible solution (for $m \leq 2M-3$) which satisfies the constraints of the RBB problem, while at the same time reducing the overall computational complexity significantly. In order to avoid switching to the SCO method for just negligibly larger ITF errors than the user-controlled upper bounds $\mathcal{E}_i$, we use the following switching criterion:

$$\left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} - \frac{b_{iL}}{b_{iR}} \right| \leq \tilde{\mathcal{E}}_i, \ i = 1, \cdots, m, \tag{21}$$

where $\tilde{\mathcal{E}}_i$ is a slightly increased upper bound and is given by

$$\tilde{\mathcal{E}}_i = (c_i + \epsilon) \left| \frac{a_L}{a_R} - \frac{b_{iL}}{b_{iR}} \right|, \ i = 1, \cdots, m, \tag{22}$$

where ϵ is very small, e.g., $0 < \epsilon < 0.1$. This modification avoids possible switching to the SCO method for negligibly larger ITF errors than the $\mathcal{E}_i$. The hybrid method is summarized in Algorithm 1.

V. Experiments

We conducted three sets of experiments. The first set (referred to as *Experiment 1*) examines the theoretical performance differences between the SCO method [13] (with $k_{\max} = 50$), the proposed SDCR method, and the proposed hybrid method (with $\epsilon = 0.05$) when the true early RATF vectors of the target and interferers are used. The second more practical set of experiments (referred to as *Experiment 2*) examines the performance of the same methods, when estimated early RATFs are used. The third practical set of experiments (referred to as *Experiment 3*) examines the performance of the same methods, when the pre-determined anechoic RATFs are used for preserving the binaural cues of the interferers (as proposed in [15]) and an estimated early RATF vector is used for preserving the binaural cues of the target. We also included in all three sets of experiments the reference methods BMVDR [1] and the BMVDR-η [5], [10]. The BMVDR-η depends on the parameter η $(0 \leq \eta \leq 1)$ which controls the trade-off between noise reduction and binaural-cue preservation. Unlike the proposed methods in which a large c increases both the noise reduction performance and binaural-cue distortions, in the BMVDR-η, a large η decreases both the noise reduction performance and binaural-cue distortions.

A. Acoustic Scene Setup

The acoustic scene, considered in our experiments, was a reverberant office environment which consisted of one target female talker in the look direction (i.e., 0°), and 4 interferers, where each had the same average power at its original location, as the target signal at the original location. The first interferer was a male talker on the right-hand side of the HA user with azimuth of 80°; the second interferer was a music signal on the right-hand side of the HA user with azimuth of 50°; the third interferer was a vacuum cleaner on the left-hand side of the HA user with azimuth -35°; and the fourth interferer was a high-frequency ringing mobile phone on the left-hand side with azimuth -70°. The microphone self-noise was set to have a 40 dB SNR at the left reference microphone, and it had the same power in all microphones.

B. Hearing-Aid Setup and Processing

The total number of microphones was $M = 4$; two at each HA. The sampling frequency was 16 kHz. We used the overlap-and-add processing method [22] for analyzing and synthesizing our signals. The analysis and synthesis windows were square-root Hanning windows and the overlap was 50%. The frame-length was 10 ms, i.e., 160 samples, and the FFT size was 256. The microphone signals were created using the head impulse responses (with a length of 458 ms) from the reverberant office environment from the database in [23]. Note that the true early RATF vectors were based on the first 10 ms of the impulse responses. The late reverberation was generated from the convolution of the late (after 10 ms) part of the impulse responses and the corresponding source signals.

In Experiments 2 and 3, the early RATF vector of a point source was estimated using a time-segment of 5 s in which only this point source signal (including its late reverberation) and the microphone-self noise was active. Specifically, we estimated the CPSDM and its eigenvalue decomposition and then we assigned to the early RATF vector the most significant (corresponding to the largest eigenvalue) relative eigenvector of the estimated CPSDM. The noise CPSDM was estimated using 5 seconds of a noise-only segment, where all interferers were active, but the target source was inactive.

We used the CVX toolbox [24] to solve the convex optimization problems associated with the SCO, SDCR and hybrid methods. The CVX toolbox uses an interior point method to solve the convex optimization problems [18]. We also used a common c value for all interferers in the inequality constraints, i.e., $c_i = c, \forall i$. The spatial filters of all methods were estimated only once using the same estimated noise CPSDM and, thus, they were time invariant. In the Experiment 3, for the pre-determined RATF vectors, we used the RATF vectors corresponding to 24 pre-determined anechoic head impulse responses from the database in [23]. The pre-determined RATF vectors were associated with azimuths uniformly spaced around the head with a resolution of $360/24 = 15^\circ$, starting from -90°. The pre-determined RATF vector at 0° was omitted from the constraints, because it was in the same direction as the RATF vector of the target. Note that the true RATF vectors of all interferers had an azimuth mismatch with the pre-determined RATF vectors' azimuths.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

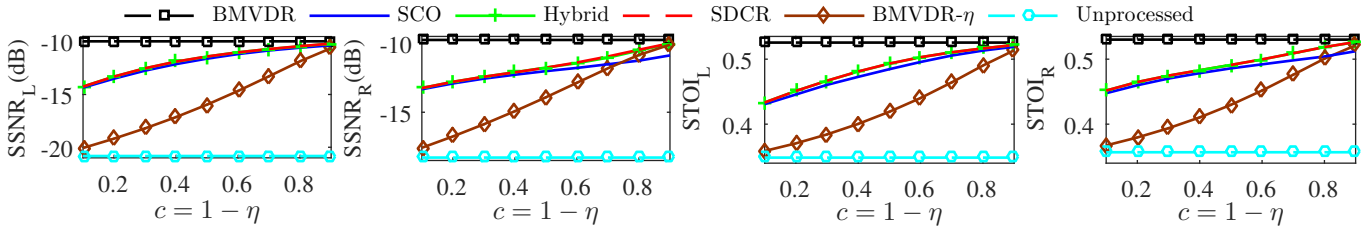

Fig. 1: Experiment 1: Noise reduction and intelligiblity prediction performances.

C. Evaluation Methodology

We measured the noise-reduction performance in terms of the segmental signal-to-noise-ratio (SSNR) only in target-presence time-regions. Let $\hat{\mathbf{X}}_L(t)$ and $\mathbf{Y}_L(t)$ denote the t-th time-frame of the estimated target and noisy signals, respectively, at the left reference microphone at the time domain, and $\mathcal{N}$ the set of the time-frames where the target is present. The SSNR at the left reference microphone is given by

$$\text{SSNR}_\text{L} = 10\log_{10}\frac{1}{|\mathcal{N}|}\sum_{t\in\mathcal{N}}\frac{||\hat{\mathbf{X}}_L(t)||_2^2}{||\hat{\mathbf{Y}}_L(t)-\hat{\mathbf{X}}_L(t)||_2^2}\ \text{dB}. \quad (23)$$

We also predicted intelligibility using the STOI measure [25].

We measured binaural-cue distortions with instrumental measures and a listening test. The instrumental measures were the average ITF error, interaural level difference (ILD) error and interaural phase difference (IPD) error per point source. These averages were calculated only over frequency (ommiting frequency bins with almost zero power), since we had fixed BFs over time. For the IPD error, we averaged only the frequency bins in the range of $0-1.5$ kHz, while for the ILD error, we averaged only the frequency bins in the range of $3-8$ kHz. This is because the ILDs are perceptually more important for localization above 3 kHz, while the IPDs are perceptually more important for localization below 1.5 kHz [26]. We used the expressions from [16] to compute the ILD and IPD errors for a single frequency bin.

The listening test was supplamentary to the Experiment 3 and is performed using the methodology described in [6]. Ten self-reported normal-hearing subjects participated (excluding the authors) and their age range was 26-37 years. They were asked to determine the azimuths of all point-sources in the acoustic scene when listening to signals processed by the compared methods as well as the unprocessed scene. The tested c values were 0.3 and 0.7 for the SCO, SDCR and hybrid methods. In addition to listening to the noisy and processed signals, the subjects also listened to the clean unprocessed point sources in isolation, in order to determine the reference azimuthms of the point sources. The localization errors were calculated with respect to the reference (and not the true) azimuths as in [6]. This is because we used only one set of head impulse responses from [23] to construct the binaural signals, which means that every subject will have a different reference azimuth. In this way, a significant estimation bias was removed. Two repetitions of the listening test were conducted. The reference azimuth of each source and every subject was computed as the average between the two repetitions, and the error was computed with respect to this averaged reference azimuth. The localization errors of the sources were averaged over subjects and repetitions. A two-way analysis of variance (ANOVA) test [27] was performed which involves the processing method and the point source as the two factors. The ANOVA test determines i) if there are at least two of the localization error mean values significantly different for the processing method factor, ii) if there are at least two of the localization error mean values significantly different for the point source factor, iii) if there is an interaction between the two factors. Finally, multiple pairwise comparisons were undertaken through the t-test with the Bonferroni correction [27] in order determine which specific methods resulted in significantly different localization error mean values. We also measured the complexity of the compared methods in terms of the average number of convex optimization problems and average execution time per-frequency bin. Note that the BFs are fixed over time and, therefore, we do not measure varying complexity over time.

D. Experiment 1: Results with True Early RATF Vectors

In this section, the compared methods use the true early RATF vectors of the sources in the constraints. Fig. 1 depicts the noise reduction performance and intelligibility prediction of the unprocessed scene, the SCO, SDCR, hybrid, BMVDR, and BMVDR-η methods at both reference microphones. The performance of SCO, SDCR and hybrid methods is measured for c values ranging from 0.1 to 0.9 with a step-size of 0.1. The performance of the BMVDR-η method is measured for η values ranging from 0.1 to 0.9 with a step-size of 0.1. In all figures, for illustration purposes, the η and c values are related as $c = 1-\eta$. As expected, as c increases (and η decreases), the noise reduction and predicted intelligibility increase. As expected the BMVDR has the best noise reduction performance and predicted intelligibility. All methods based on the RBB problem achieve similar performances for the left reference microphone, while for the right reference microphone the SCO method achieves the worst noise reduction performance among all, especially for $c \geq 0.5$. Note that the SDCR method has almost identical performance as the hybrid method. This is because, in this example the hybrid method switched to the SCO method only a few times. Finally, the BMVDR-η method has a comparable predicted intelligibility with the proposed methods only for small η values.

Fig. 2 shows the binaural-cue distortions of the compared methods per interfering source. The binaural-cue distortions

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

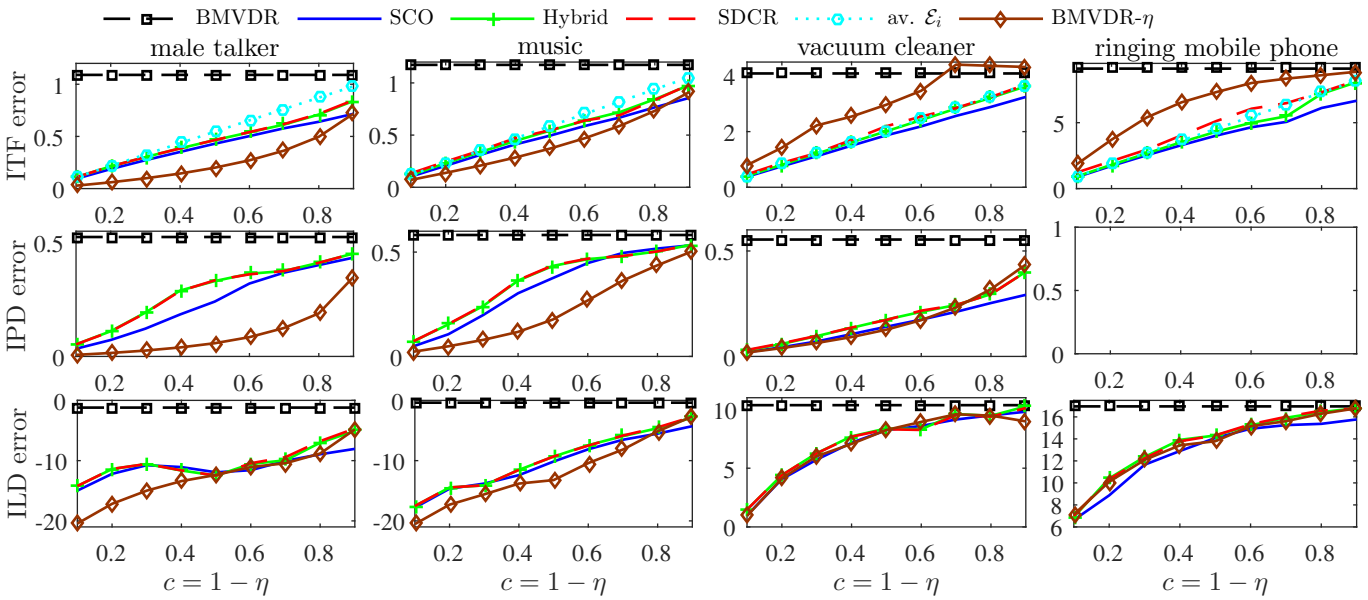

Fig. 2: Experiment 1: Binaural-cue distortions (averaged over frequency) of interferers.

of the target source are always zero in Experiment 1. As expected, as c increases (and η decreases), the binaural-cue distortions increase. For the ITF errors, we also display the c times the average ITF error of the BMVDR (which is labeled as av. $\mathcal{E}_i$) in order to visualize the closeness of the estimated spatial filters at the boundary of the inequality constraints of the RBB problem. It is clear that both SDCR and hybrid methods are closer to the boundary compared to the SCO method for the same c value. Moreover, the hybrid method is for all c values (on average) below the boundary, even if we used the extended switch criterion in (21). On the other hand, the ITF error of the SDCR method sometimes (see ringing mobile phone) is slightly above the boundary. As explained in Sec. IV, this is because the SDCR method does not guarantee a user-controlled upper-bounded ITF error as the SCO or the hybrid methods do. Notably, the SCO method for large c values (e.g., $c \geq 0.6$), is not close to the boundary, while the SDCR and hybrid methods are closer to the boundary. Thus, the SDCR and hybrid methods achieve more expected binaural-cue distortions according to the trade-off parameter set by the user compared to the SCO method. Note also that the IPD error for the ringing mobile phone was not computed because it has almost zero power below 1.5 kHz.

Fig. 3 shows the computational complexity of the compared methods in terms of average number of convex optimization problems required to solve for convergence and average cpu time in seconds per frequency bin. The SDCR method requires to solve much less convex problems than the SCO method (especially at larger c values) and slightly less compared to the hybrid method. The hybrid method requires to solve much less convex problems than the SCO method. The fastest method among all is obviously the BMVDR-η method because it has a closed-form solution while all the other methods are iterative.

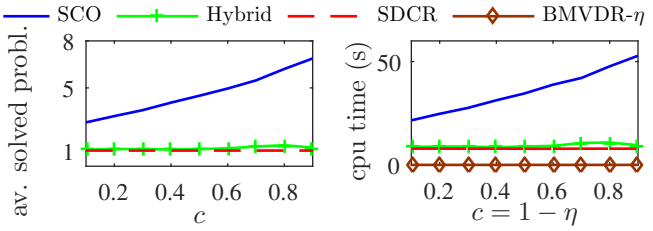

Fig. 3: Experiment 1: Computational complexity measured as the average number of solved convex optimization problems and average computation time (seconds) per frequency bin.

We can conclude from the above that, in most cases, both proposed methods are more optimal than the SCO method. Specifically, both proposed methods provide solutions that are closer to the expected solutions of the original RBB problem, since both proposed methods are closer to the boundary. This means that both methods provide a more user-controlled trade-off between noise reduction and binaural-cue preservation than the SCO method, especially in large c values. Finally both proposed methods are significantly less computationally demanding than the SCO method.

E. Experiment 2: Results with Estimated early RATF Vectors

In this section, the compared methods use estimated RATF vectors. Fig. 4 shows the noise reduction performance and intelligibility prediction of the compared methods which is very similar to the one in Fig. 1. Fig. 5 shows the binaural-cue distortions of the compared methods per point source (including the target source). As expected, here we have ITF errors which are sometimes above $\mathcal{E}_i$, because of the estimation errors in the RATF vectors. The computational complexity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

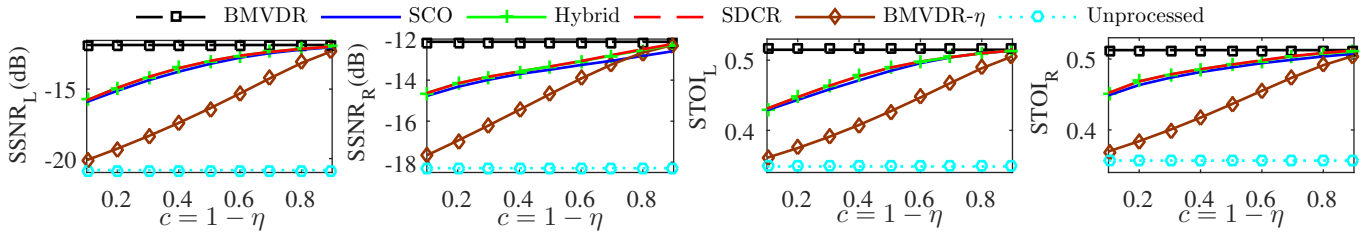

Fig. 4: Experiment 2: Noise reduction and intelligiblity prediction performances.

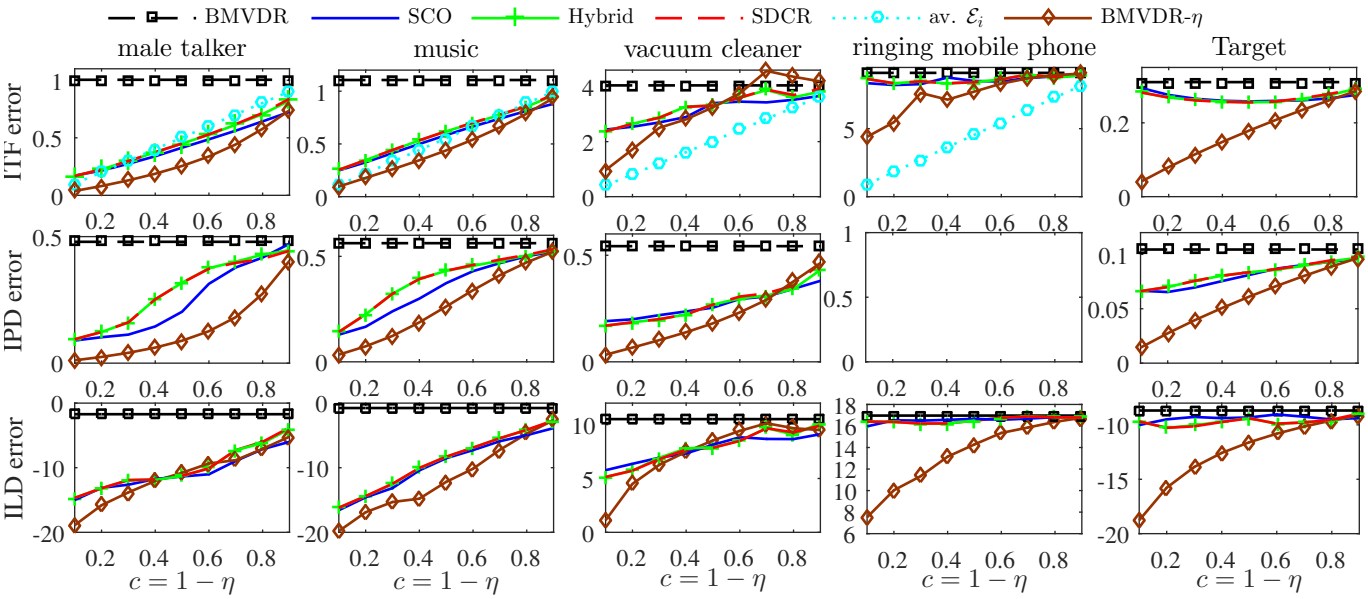

Fig. 5: Experiment 2: Binaural-cue distortions (averaged over frequency) of point sources.

performance is omitted because is very similar to Fig. 3. Finally, the BMVDR-η method has a similar performance as with Experiment 1, since the only thing that has changed is the estimation error in the target RATF vector.

F. Experiment 3: Results with Pre-Determined RATF Vectors

In this section, the SCO, SDCR and hybrid methods use the pre-determined RATF vectors for the interferers' binaural-cue preservation and an estimated early RATF vector for the target. Fig. 6 shows the noise reduction performance and intelligibility prediction of the compared methods. Here the gap in performance (for the same c value) between the proposed methods and the SCO method is bigger compared to the case where the true RATF vectors were used. The proposed methods (especially the SDCR method) significantly improved both noise reduction and predicted intelligibility at both reference microphones for the same c value compared to the SCO method. The reason why the performance gap between the SDCR method and the hybrid method is increased compared to Experiment 1 is because the hybrid method switched many more times to the SCO method (see Algorithm 1) in Experiment 3. In conclusion, for the same c value, both proposed methods achieved in most cases a better noise reduction and predicted intelligibility than the SCO method, especially for larger c values. The BMVDR-η method has the same performance as with the Experiment 2 and now has a comparable intelligibility improvement for all η values compared to the proposed methods.

Fig. 7 shows the binaural-cue distortions of the compared methods per point source (including the target source). As expected, when pre-determined RATF vectors are used, all methods do not guarantee a user-controlled upper-bounded ITF error of the interferers which will be c times the BMVDR ITF error. Therefore, all methods, in many occasions (see vacuum cleaner and ringing mobile phone), result in a larger ITF error than the average $\mathcal{E}_i$. The SCO method has the lowest binaural-cue distortions compared to the proposed methods because it is further away from the boundary of the inequality constraints of the RBB problem. Nevertheless, we will see later on in the listening test that the compared methods do not have significantly different binaural-cue distortions for the same c value.

In Fig. 8, we show the computational complexities of the compared methods. The results are similar to the results in Fig. 3 with the only difference that now the hybrid method does not achieve significant computational savings over the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

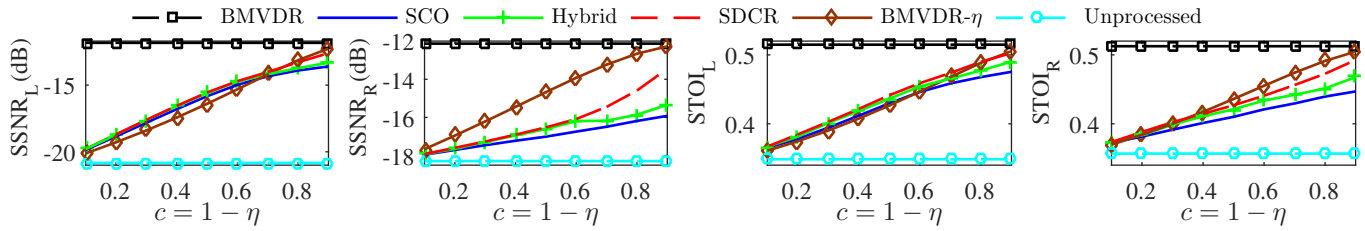

Fig. 6: Experiment 3: Noise reduction and intelligiblity prediction performances.

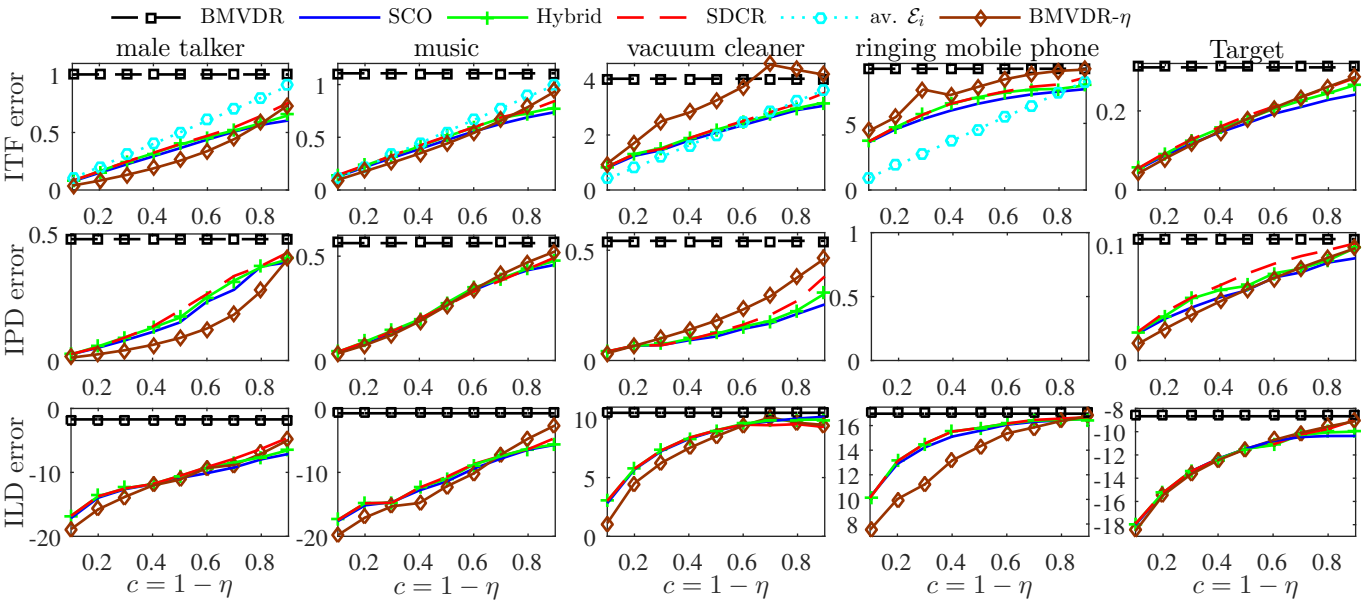

Fig. 7: Experiment 3: Binaural-cue distortions (averaged over frequency) of point sources.

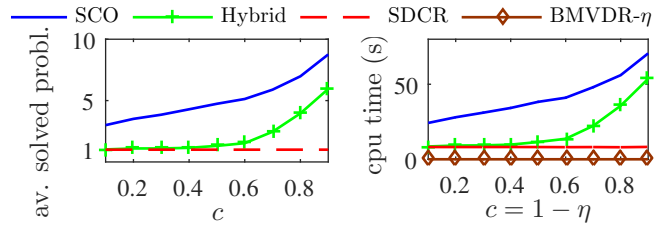

Fig. 8: Experiment 3: Computational complexity measured as the average number of solved convex optimization problems and average computation time (seconds) per frequency bin.

SCO method as with Experiment 1. However, the usage of the hybrid method using pre-determined RATF vectors is not critical, since no method can guarantee user-controlled upper-bounded ITF error of the interferers, unless the number of pre-determined RATF vectors is huge. However, this is not practical as it may result in non-feasible solutions or the noise reduction will be negligible.

Fig. 9 shows the results of the subjective localization test of Experiment 3. The examined values for the SCO, SDCR and hybrid methods are $c = 0.3, 0.7$, while for BMVDR-η we choose $\eta = 0.8$. A similar behavior as with the instrumental binaural-cue distortion measures is observed here. For a large c value we have in most cases a larger localization error. Moreover, as expected the BMVDR method has the largest localization error. Finally, the BMVDR-η method for $\eta = 0.8$ has a similar performance with the RBB-based methods for $c = 0.3$. Note that among all interferers the mobile ringing phone was the most difficult to localize for $c = 0.7$. Several users also reported difficulty in localizing the ringing phone after completing the test. We believe that this is because of the high frequency content of the ringing tone of the mobile phone and only the ILDs might have been used for localization.

Table I shows the results of the ANOVA test. We can conclude from the results that i) at least two of the mean values of the factor point source are significantly different, ii) at least two of the mean values of the factor processing method are significantly different and iii) there is a significant interaction between the two factors. Since there is a significant interaction between the two factors we have undertaken comparisons between pairs of methods for each interferer separately with several t-tests. The significance level was set to 1%. For the female talker all methods are not significantly different.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2018.2878618, IEEE/ACM Transactions on Audio, Speech, and Language Processing

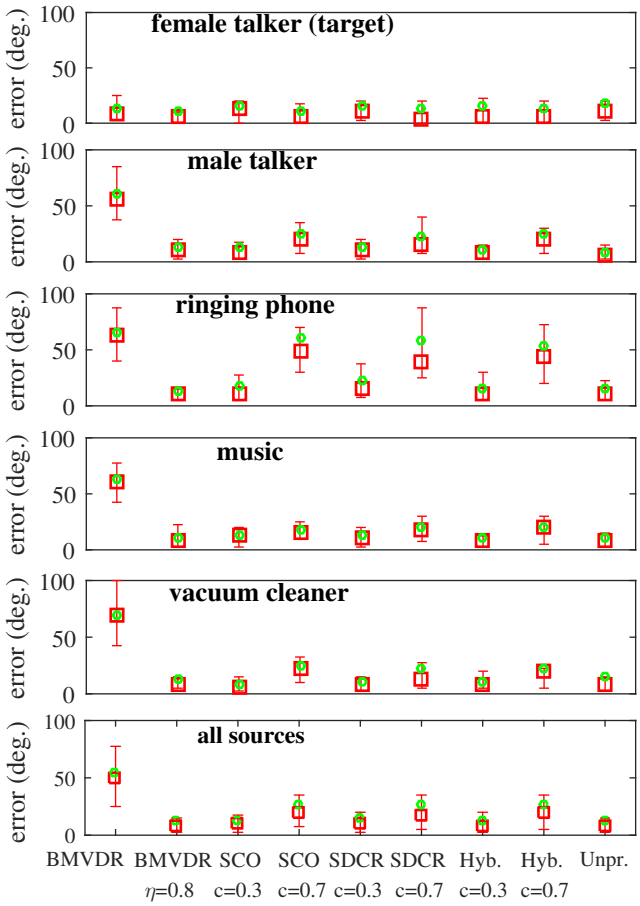

Fig. 9: Experiment 2: Localization test measuring the localization error in degrees for all compared methods and point sources. The bottom figure is the average localization error over all point sources.

For the male talker, music and vacuum cleaner, all methods are significantly different from the BMVDR method, but, surprisingly are not significantly different with each other. This means that even though in the instrumental measures we observed a not negligible difference in binaural-cue distortions between $c = 0.3$ and $c = 0.7$, the subjective evaluation contradicts that. For the mobile phone, the SDCR ($c = 0.3$), hybrid ($c = 0.3$), SCO ($c = 0.3$), BMVDR-η ($\eta = 0.8$) and unprocessed methods are all not significantly different, but are all significantly different with all the remaining methods. Furthermore, the SDCR ($c = 0.7$), hybrid ($c = 0.7$), SCO ($c = 0.7$) and BMVDR are not significantly different.

We can conclude from the above comparisons that the proposed methods do not cause significantly different binaural-cue distortions compared to the SCO method for the same c value and for all point sources in the acoustic scene. This means that even though we observed less binaural-cue distortions in the SCO method in Figs 2 and 7, compared to the proposed methods for the same c value, these differences are not perceptually important. However, recall that the proposed methods achieve a better noise reduction and predicted intelligibility compared to the SCO method. Thus, the proposed methods provide a better perceptual trade-off compared to the SCO method. Finally, note that the SCO, SDCR, hybrid for $c = 0.3$ and BMVDR-η for $\eta = 0.8$ methods are not statistically significantly different from the unprocessed scene for all point sources in the acoustic scene. This means that in all four methods the subjects managed (on average) to localize as good as in the unprocessed scene. However, unlike the unprocessed scene, all four methods improved noise reduction and predicted intelligibility.

TABLE I: Two-way ANOVA test with the point source and processing method as the two factors.

Source of variation	Sum Sq.	d.f.	Mean Sq.	F	p-value
Point-source(A)	165727.5	8	20715.9	39.77	8.1e-55
Proc.-method(B)	49935	4	12483.8	23.97	6.6e-19
AB	82551.8	32	2579.7	4.95	5.4e-17
Error	492196.6	945	520.8		
Total	790410.9	989			

VI. Conclusion

We proposed two new suboptimal methods for approximately solving the non-convex relaxed binaural beamforming (RBB) optimization problem. Both methods are significantly computationally less demanding compared to the existing successive convex optimization (SCO) method. For each frequency bin, the SCO method requires to solve many more convex optimization problems in order to converge compared to the proposed methods. Specifically, the first proposed method, which is a semi-definite convex relaxation (SDCR) of the RBB problem, solves only one convex optimization problem per frequency bin. Apart from the computational advantage, the SDCR method also achieves in most cases a better trade-off between intelligibility and binaural-cue preservation than the SCO method. However, the SDCR method does not guarantee user-controlled upper bounded ITF error when the RATF vectors of the interferers are estimated accurately enough. This problem is solved by the second proposed method, which is a hybrid combination of the SDCR and SCO methods. This method guarantees user-controlled upper-bounded ITF error, and at the same time is computationally much less demanding than the SCO method. Finally, listening tests showed that all three methods achieve not significantly different localization errors for the same amount of binaural-cue error relaxation.

References

[1] 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.
[2] J. M. Kates, *Digital hearing aids*. Plural publishing, 2008.
[3] 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, Dec. 2015.
[4] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acoustica*, vol. 86, no. 1, pp. 117–128, 2000.
[5] D. Marquardt, "Development and evaluation of psychoacoustically motivated binaural noise reduction and cue preservation techniques," Ph.D. dissertation, Carl von Ossietzky Universität Oldenburg, 2015.

[6] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, S. van de Par, J. Jensen, and M. Guo, "Evaluation of binaural noise reduction methods in terms of intelligibility and perceived localization," in *EURASIP Europ. Signal Process. Conf. (EUSIPCO)*, 2018.

[7] J. G. Desloge, W. M. Rabinowitz, and P. M. Zurek, "Microphone-array hearing aids with binaural output .I. Fixed-processing systems," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 529–542, Nov. 1997.

[8] D. P. Welker, J. E. Greenberg, J. G. Desloge, and P. M. Zurek, "Microphone-array hearing aids with binaural output .II. A two-microphone adaptive system," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 543–551, Nov. 1997.

[9] T. 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.

[10] 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.

[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] 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, Jan. 2016.

[13] 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, Jan. 2017.

[14] S. Gannot, E. Vincet, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multi-microphone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 4, pp. 692–730, April 2017.

[15] 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)*, Aug. 2017.

[16] 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.

[17] H. Anton, *Elementary linear algebra*. John Wiley & Sons, 2010.

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

[19] G. Golub and C. V. Loan, *Matrix Computations*, 3rd ed. Oxford: North Oxford Academic, 1983.

[20] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, Mar. 1996.

[21] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Mathematical programming*, vol. 95, no. 1, pp. 3–51, 2003.

[22] J. B. 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.

[23] H. Kayser, S. Ewert, J. Annemuller, 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.

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

[25] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.

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

[27] D. J. Sheskin, "Parametric and nonparametric statistical procedures," *Chapman & Hall/CRC: Boca Raton, FL*, 2000.