BINAURAL BEAMFORMING BASED ON AUTOMATIC INTERFERER SELECTION

Costas A. Kokke, Richard C. Hendriks and Andreas I. Koutrouvelis

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

ABSTRACT

Binaural cues are important for sound localization. In addition, spatially separated sound sources are more intelligible than when they are co-located. Binaural cue preservation in multi-microphone hearing assistive devices is therefore important for the user's listening experience and safety. A number of linearly-constrained-minimum-variance (LCMV) based methods exist for this purpose. These are all limited in the number of sources for which they can preserve the binaural cues. We propose a method of automatically selecting the most important interfering sources using convex optimization. The proposed method is compared, using simulation experiments, to existing methods in terms of noise suppression and localization errors. It improves the performance of the joint binaural LCMV beamformer, by giving it more degrees of freedom for noise reduction and allows a larger number of (virtual) sources present in the scene.

Index Terms— Noise reduction, spatial cue preservation, binaural beamforming, interferer selection, convex optimization.

1. INTRODUCTION

The use of hearing assistive devices (HADs) has significantly increased in our aging society. The need for them is expected to further increase, partly due to increased recreational exposure to loud sound [2]. Typically, HADs come in pairs, called bilateral or binaural HADs, each equipped with multiple microphones. These microphones can be used for beamforming. An important aspect of binaural HADs is their ability to preserve the binaural cues of the sound field, resulting in a more natural user experience. More importantly, binaural cue preservation helps the HAD user to localize sound in day-to-day situations, for example in traffic. Another important reason to preserve binaural cues is the ability of the human auditory system to distinguish spatially separated sources better than spatially co-located sources. This helps to focus on sources by their location. The cocktail party effect is an example of this [3].

Beamforming can be performed in the short-time discrete Fourier transform (STFT) domain by changing the magnitude and phase of the STFT coefficients of the different microphones prior to combining them. However, this could distort the spatial cues of the output signals. Such binaural cue distortions may have considerable consequences and should be prevented.

To this end, binaural beamformers are developed for simultaneous noise reduction and binaural cue preservation [4, 5, 6, 7]. In this paper, we consider binaural beamformers where the target is undistorted, as in [4, 5, 7]. This is beneficial for the speech intelligibility. Most binaural beamformers are limited in the amount of binaural cues they can retain, as the degrees of freedom for noise reduction and binaural cue preservation are limited and depend on the number of microphones. In general, noise reduction and binaural cue preservation comes with a trade off of the available degrees of freedom. Spending all degrees of freedom on interferer binaural cue preservation implies no controlled noise reduction and vice versa. An important aspect of binaural noise reduction is therefore to efficiently use the constraints for binaural cue preservation.

Existing binaural beamformers spend these degrees of freedom in different manners. The binaural minimum variance distortionless response (BMVDR) beamformer [7], for example, spends all degrees of freedom on noise reduction, and none on binaural cue preservation. This causes all sound to appear to come from the target direction.The BMVDR can be generalized to the general binaural linearly constrained minimum variance (GBLCMV) beamformer, which can have at most $2M-1$ constraints and still perform controlled noise reduction, where M is the amount of microphones on both HADs. Typically, two constraints are used to binaurally constrain the target source, leaving $2M-3$ constraints to be used for other purposes. The joint binaural linearly constrained minimum variance (JBLCMV) beamformer [4, 7] is a distortionless binaural beamformer that fits in the GBLCMV framework. It uses one constraint to binaurally constrain an interfering point source, meaning it can at most binaurally constrain $2M-3$ interfering point sources, while still being able to do noise suppression.

Typically, these methods use one constraint per interferer in all time-frequency tiles. Even in the tiles where an interferer might be (almost) inaudible, due to suppression by the beamformer, or due to masking by other sources. Since the amount of microphones on current HADs is relatively low, typically $M = 4$, degrees of freedom are scarce. This not only means that degrees of freedom are quickly exhausted, but it also means that degrees of freedom are spent on inperceptual, and thus unnecessary, binaural cue constraints in some time-frequency tiles. To be able to take into account even more interfering sources, while still being able to do controlled noise reduction, we present in this paper a method to automatically select the most important (e.g., audible or perceptually important) binaural constraints for the final beamformer output. By doing so, interferers inaudible at the output are not binaurally preserved, leaving degrees of freedom to preserve other sources, or, apply more noise reduction.

2. SIGNAL MODEL AND NOTATION

We consider the binaural HAD setting, with two collaborating HADs that have a combined total of M microphones installed. The signals are assumed to be processed on a frame-by-frame basis in the frequency domain. Since processing takes place independently per frame, time-frame indices are omitted for convenience.

Assuming an additive distortion model, an STFT coefficient, $y_j[k]$, at the jth microphone is composed as follows,

$$y_j[k] = a_j[k]s[k] + \sum_{i=1}^{b} b_{ij}[k]u_i[k] + v_j[k]\,, \tag{1}$$

The first author performed the work as part of a thesis for the degree of Master of Science in Electrical Engineering at Delft University of Technology [1].

where $a_j[k]$ and $b_{ij}[k]$ are the acoustic transfer functions (ATFs) of the desired source and i-th interferer to the j-th microphone respectively, $s[k]$ and $u_i[k]$ are the desired source and i-th interferer respectively, b is the number of interferers, and $v_j[k]$ is additive uncorrelated noise. In the remainder of this work, the frequency variable k will be omitted for simplicity, since all processing will be done per frequency bin, assuming frequency bins are mutually independent. All microphone signals can be combined using vector notation. Let

$$\mathbf{y} = \mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{v} \in \mathbb{C}^{M\times 1}\,, \tag{2}$$

where $\mathbf{x} = \mathbf{a}s$ and $\mathbf{B} = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_b\end{bmatrix} \in \mathbb{C}^{M\times b}$.

Assuming all sources mutually uncorrelated, the cross power spectral density (CPSD) matrix $\mathbf{P}$ of all disturbances is defined as

$$\mathbf{P} = \sum_{i=1}^{b} \mathbf{P}_{\mathbf{n}_i} + \mathbf{P}_{\mathbf{v}} \in \mathbb{C}^{M\times M}\,, \tag{3}$$

with $\mathbf{P}_{\mathbf{n}_i}$ and $\mathbf{P}_{\mathbf{v}}$ the CPSD matrices of $\mathbf{n}_i = \mathbf{b}_i u_i$ and $\mathbf{v}$.

Finally, on each HAD we define a reference microphone. These microphones are used as a reference with respect to the preservation of the binaural cues of interfering point sources and the complete preservation of the target signal on both the left and right HAD.

3. PROBLEM STATEMENT

We aim to find a method that more efficiently constrains the binaural cues of interfering point sources. We propose to do this by automatic optimal interferer subset selection by choosing the optimal subset of known interferers for binaural cue preservation. As an example, we apply this to the JBLCMV beamformer [4, 7].

3.1. JBLCMV

The JBLCMV preserves the target point source and the binaural cues of up to $2M-3$ interfering point sources, while still doing controlled noise reduction [4, 7]. Let $\mathbf{w} = \begin{bmatrix}\mathbf{w}_L^H & \mathbf{w}_R^H\end{bmatrix}^H \in \mathbb{C}^{2M\times 1}$, where $\mathbf{w}_L$ and $\mathbf{w}_R$ are the filters that are both applied to all microphone signals to obtain the left and right HAD output signal, respectively. The optimization problem to obtain these filters is defined by

$$\mathbf{w} = \arg\min_{\mathbf{w}} \quad \mathbf{w}^H\tilde{\mathbf{P}}\mathbf{w} \tag{4a}$$

$$\text{s.t.} \quad \mathbf{w}^H\mathbf{C} = \mathbf{f}^H\,, \tag{4b}$$

where $\tilde{\mathbf{P}} = \begin{bmatrix}\mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}\end{bmatrix} \in \mathbb{C}^{2M\times 2M}$, $\mathbf{f} = \begin{bmatrix}a_L & a_R & \mathbf{0}_b^T\end{bmatrix}^H$,

$$\mathbf{C} = \Bigg[\underbrace{\begin{matrix}\mathbf{a} & \mathbf{0}\\ \mathbf{0} & \mathbf{a}\end{matrix}}_{\mathbf{\Lambda}_a} \quad \underbrace{\begin{matrix}\mathbf{b}_{1L} & \cdots & \mathbf{b}_{bL}\\ -\mathbf{b}_{1R} & \cdots & -\mathbf{b}_{bR}\end{matrix}}_{\mathbf{\Lambda}_b}\Bigg] \in \mathbb{C}^{2M\times(b+2)}\,, \tag{5}$$

with $\mathbf{b}_{iL} = \mathbf{b}_i b_{iL}^{-1}$ the ith left relative transfer function (RTF) and similarly for $\mathbf{b}_{iR}$. The first two columns of $\mathbf{C}$ in Eq. (5) provide the binaural distortionless constraints, while each other column provides (exact) binaural cue preservation for one interfering point source based on preservation of the interaural transfer function, i.e., $\frac{\mathbf{w}_L^H b_{iL}}{\mathbf{w}_L^H b_{iR}} = \frac{b_{iL}}{b_{iR}}$. We defined $\mathbf{\Lambda}_a$ and $\mathbf{\Lambda}_b$ to be the matrices containing all constraints on the target and interferers, respectively.

The solution to this problem formulation is given by

$$\mathbf{w} = \tilde{\mathbf{P}}^{-1}\mathbf{C}\Big(\mathbf{C}^H\tilde{\mathbf{P}}^{-1}\mathbf{C}\Big)^{-1}\mathbf{f}\,, \tag{6}$$

when $b \le 2M-3$. When $b > 2M-3$, there is no solution that provides controlled noise reduction and a subset should be chosen to satisfy the constraint $b \le 2M-3$. However, it is not clear which RTFs should be chosen for binaural cues preservation. A naive choice would be to select the $2M-3$ RTFs that have the highest associated powers. This is not optimal, as will be shown in Section 5, since it is the output power that determines the (perceptual) importance of the interferers. Moreover, this also changes with the filter. Furthermore, constraining the binaural cues of an interferer that is inaudible at the beamformer output needlessly lowers the noise reducing capabilities of the beamformer. In this work, we will therefore consider an optimal selection of the most important interferers. However, this is not obvious as this is a non-convex and integer problem.

3.2. Pre-determined RTFs

The problem formulation outlined above depends on the RTFs. Estimating RTFs of interferers in practical applications is challenging. Instead of actual RTFs, one could also use predetermined RTFs, where a number of virtual sources is placed in the far-field around the head of the user, each with a known predetermined relative acoustic transfer function (PRTF) [8].

To make sure that the mismatch between a PRTF and true RTF cannot become too large, we would like to have many PRTFs. This is problematic for the noise reduction versus binaural cue preservation trade-off of LCMV-based beamformers, because more constraints will lead to less noise reduction. Additionally, they can only handle a limited number of constraints, so a subset of the PRTFs has to be selected that is no larger than $2M-3$. This is, as mentioned in Section 3.1, a non-convex and integer problem.

3.3. Research Question

From the above, it is clear that the set of RTFs that are binaurally constrained, has a big impact on the final performance in terms of noise reduction and binaural cue preservation. This holds for both the framework using the true RTFs as well as when one uses the predetermined RTFs. In this paper we therefore investigate automatic selection of RTFs for binaurally constrained beamformers.

This method decides which RTFs are most important to constrain, based on the amount of noise reduction and the amount of binaurally unconstrained noise power after filtering. This will additionally allow for more potential noise reduction, since interfering sources that are deemed inaudible after filtering can be binaurally unconstrained, leaving more degrees of freedom for noise reduction.

4. PROPOSED METHOD

We propose a method that can select from all interferer constraints the $b \le 2M-3$ most important constraints. As such, a constraint matrix $\mathbf{C}$ as in Eq. (5) is constructed that always satisfies the limit on the number of constraints that the JBLCMV can handle.

Let $\mathbf{\Lambda}_a = \begin{bmatrix}\mathbf{a} & \mathbf{0}\\ \mathbf{0} & \mathbf{a}\end{bmatrix} \in \mathbb{C}^{2M\times 2}$ and $\mathbf{\Lambda}_b \in \mathbb{C}^{2M\times l}$ be the constraint matrices for the target and all l (virtual) interferers, respectively. Let

$$\mathbf{\Lambda} = \begin{bmatrix}\mathbf{\Lambda}_a & \mathbf{\Lambda}_b\end{bmatrix} \in \mathbb{C}^{2M\times(2+l)}\,, \tag{7}$$

be the matrix containing all the constraints, defined similar as Eq. (5). We then define a selection vector and matrix $\mathbf{p} \in \{0,1\}^l$ and $\mathbf{\Phi}_{\mathbf{p}}$, respectively, that are used to select columns from Eq. (7). Matrix $\mathbf{\Phi}_{\mathbf{p}}$ is a submatrix of $\operatorname{diag}\big(\begin{bmatrix}1 & 1 & \mathbf{p}^T\end{bmatrix}\big)$ with its zero columns removed. The selected columns are put into the constraint matrix $\mathbf{C_p}$ that is presented to the JBLCMV. This allows us to define

$$\mathbf{C_p} = \mathbf{\Lambda \Phi_p} \in \mathbb{C}^{2M\times(|\mathbf{p}|+2)}. \tag{8}$$

We now define an optimization problem with $\mathbf{p}$ as the minimization variable, the total output noise power as cost function and as constraints the power of binaurally unconstrained sources and the maximum number of constraints selected by $\mathbf{p}$ [9]. Let

$$\begin{aligned} \mathbf{p} = \arg\min_{\mathbf{p}\in\{0,1\}^l} \quad & \mathbf{w_p}^H \tilde{\mathbf{P}} \mathbf{w_p} && \text{(9a)} \\ \text{s.t.} \quad & \left\| (\mathbf{I} - \operatorname{diag}(\mathbf{p}))\mathbf{U}\hat{\mathbf{B}}^H \mathbf{w_p} \right\|_2^2 \le \beta^{-1} P_x && \text{(9b)} \\ & \mathbf{1}^T \mathbf{p} \le 2M - 3\,, && \text{(9c)} \end{aligned}$$

where P_x is the estimated power of the target, β is the desired minimum signal-to-binaurally-unconstrained-interferer-power ratio, $\hat{\mathbf{B}} = \begin{bmatrix} \mathbf{B}^H & \mathbf{B}^H \end{bmatrix}^H \in \mathbb{C}^{2M\times l}$ contains the binaural (P)RTFs, $\mathbf{U} = \operatorname{diag}(\mathbf{u}) \in \mathbb{C}^{l\times l}$ with $\mathbf{u}$ a vector with the power of each interferer and we choose $\mathbf{w_p}$ as

$$\mathbf{w_p} = \tilde{\mathbf{P}}^{-1}\mathbf{C_p}\left(\mathbf{C_p}^H \tilde{\mathbf{P}}^{-1}\mathbf{C_p}\right)^{-1}\mathbf{f} \in \mathbb{C}^{2M}, \tag{10}$$

which is the JBLCMV filter for the interferer subset determined by $\mathbf{p}$. Eq. (9b) guarantees that the power of the unconstrained interferers is below a certain level. Although we do not explicitly incorporate formal models of audibility or perception in this paper in order to focus on solving the problem formulation itself, this could be taken into account in Eq. (9b) in future research.

The problem in Eq. (9) is non-convex and may have multiple local minima. To solve this optimization problem, we propose a number of relaxations to reformulate the problem into a convex form that always has a feasible solution.

4.1. Threshold Relaxation

When $l \le 2M - 3$, the norm in Eq. (9b) can always be made 0. As such, that constraint can never cause an infeasible problem as long as $l \le 2M - 3$. As discussed in Section 3.2, we typically want a larger amount of PRTFs, and as such, it might not always be possible to satisfy Eq. (9b), because of the cardinality constraint in Eq. (9c). Since an infeasible problem is generally useless, we implement a relaxation on the power constraint Eq. (9b), such that the feasible region is only expanded minimally to allow a solution to exist.

Let the updated problem be

$$\begin{aligned} \min_{\mathbf{p}\in\{0,1\}^l,\, t_1} \quad & \mathbf{w_p}^H \tilde{\mathbf{P}} \mathbf{w_p} + \alpha \max\left(0,\, t_1 - \beta^{-1} P_x\right) && \text{(11a)} \\ \text{s.t.} \quad & \left\| (\mathbf{I} - \operatorname{diag}(\mathbf{p}))\mathbf{U}\hat{\mathbf{B}}^H \mathbf{w_p} \right\|_2^2 \le t_1 && \text{(11b)} \\ & \mathbf{1}^T \mathbf{p} \le 2M - 3\,, && \text{(11c)} \end{aligned}$$

where t_1 is the new relaxation variable and α is the penalty weight for violation of the original threshold.

4.2. Cost Function Relaxation

To relax the cost function in Eq. (9a) to a convex form, we re-write it using Eqs. (8) and (10) and use the following properties of the selection matrix $\mathbf{\Phi_p}$: $\mathbf{\Phi_p}^T \mathbf{\Phi_p} = \mathbf{I}_{2+|\mathbf{p}|}$ and $\mathbf{\Phi_p}\mathbf{\Phi_p}^T = \operatorname{diag}\left(\begin{bmatrix} 1 & 1 & \mathbf{p}^T \end{bmatrix}\right)$. Consider the following decomposition [1]:

$$\mathbf{\Lambda}^H \tilde{\mathbf{P}} \mathbf{\Lambda} = \lambda \mathbf{I} + \mathbf{G}\,, \quad \text{using} \quad \lambda > \lambda_{\text{max}}\left(\mathbf{\Lambda}^H \tilde{\mathbf{P}} \mathbf{\Lambda}\right). \tag{12}$$

We use this decomposition such that we can formulate the relaxation of the cost function as a linear matrix inequality (LMI) using the Schur complement [10]. Using Eq. (12), the cost function is first re-written to

$$\begin{aligned} \mathbf{w_p}^H \tilde{\mathbf{P}} \mathbf{w_p} &= \mathbf{f}^H \mathbf{\Phi_p}\left(\mathbf{\Phi_p}^T (\lambda \mathbf{I} + \mathbf{G}) \mathbf{\Phi_p}\right)^{-1} \mathbf{\Phi_p}^T \mathbf{f}\,, && \text{(13)} \\ &= \mathbf{f}^H \mathbf{Q} \mathbf{f}\,, && \text{(14)} \end{aligned}$$

where

$$\mathbf{Q} = \mathbf{\Phi_p}\left(\lambda \mathbf{I} + \mathbf{\Phi_p}^T \mathbf{G} \mathbf{\Phi_p}\right)^{-1} \mathbf{\Phi_p}^T, \tag{15}$$

which is reformulated using the matrix inversion lemma [11] to

$$\mathbf{Q} = \mathbf{G}^{-1} - \mathbf{G}^{-1}\left(\mathbf{G}^{-1} + \lambda^{-1} \operatorname{diag}\left(\begin{bmatrix} \mathbf{1}_2^T & \mathbf{p}^T \end{bmatrix}\right)\right)^{-1} \mathbf{G}^{-1}. \tag{16}$$

The cost function in Eq. (14) is convex in $\mathbf{Q}$, but the definition of $\mathbf{Q}$ in Eq. (16) is not a convex constraint in $\mathbf{p}$. We therefore relax Eq. (16) by bounding the right-hand side from above and applying the Schur Complement [10] to obtain

$$\begin{bmatrix} -\mathbf{G}^{-1} - \lambda^{-1} \operatorname{diag}\left(\begin{bmatrix} \mathbf{1}_2^T & \mathbf{p}^T \end{bmatrix}\right) & \mathbf{G}^{-1} \\ \mathbf{G}^{-1} & \mathbf{Q} - \mathbf{G}^{-1} \end{bmatrix} \succeq 0\,, \tag{17}$$

which is convex in both $\mathbf{p}$ and $\mathbf{Q}$ [1]. The cost function in Eq. (14) and the above LMI are a convex relaxation of the original cost function in Eq. (9a), with the additional minimization variable $\mathbf{Q}$.

4.3. Power Constraint Relaxation

To relax the power constraint in Eq. (9b) to a convex form, the norm is expanded, using the new minimization variable $\mathbf{Q}$, that is

$$\begin{aligned} &\left\| (\mathbf{I} - \operatorname{diag}(\mathbf{p}))\mathbf{U}\hat{\mathbf{B}}^H \mathbf{w_p} \right\|_2^2 = \\ &\mathbf{f}^H \mathbf{Q} \mathbf{\Lambda}^H \tilde{\mathbf{P}}^{-1} \tilde{\mathbf{B}} \mathbf{U} (\mathbf{I} - \operatorname{diag}(\mathbf{p})) \mathbf{U} \tilde{\mathbf{B}}^H \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \mathbf{Q} \mathbf{f}\,, \end{aligned} \tag{18}$$

which is a bilinear, and thus non-convex, expression in $\mathbf{p}$ and $\mathbf{Q}$. We relax it, by formulating Eq. (18) as an LMI after first introducing an approximation to $(\mathbf{I} - \operatorname{diag}(\mathbf{p}))$ to make it positive definite. After applying this approximation and the Schur Complement, we obtain

$$\begin{bmatrix} \mathbf{I} + (\frac{1}{\epsilon} - 1) \operatorname{diag}(\mathbf{p}) & \mathbf{U}\tilde{\mathbf{B}}^H \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \mathbf{Q} \mathbf{f} \\ \mathbf{f}^H \mathbf{Q} \mathbf{\Lambda}^H \tilde{\mathbf{P}}^{-1} \tilde{\mathbf{B}} \mathbf{U} & t_1 \end{bmatrix} \succeq 0\,, \tag{19}$$

where $\epsilon \ll 1$ is the discussed approximation.

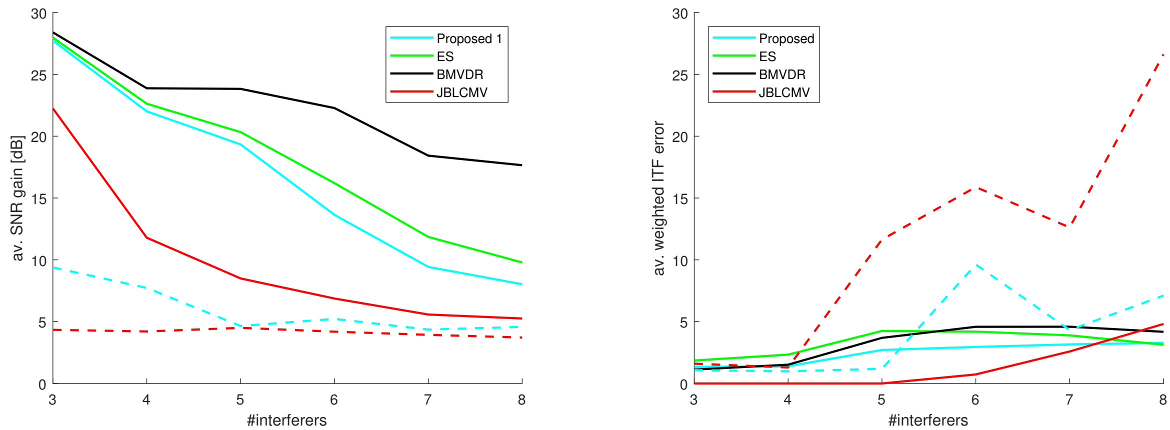
4.4. Binary Selection Relaxation

Since the optimization problem in Eq. (9) is an Integer Program (known to be non-convex and NP-hard to solve) we propose to relax the binary variable $\mathbf{p}$ to a continuous variable $\hat{\mathbf{p}} \in [0,1]^l$ [9, 12, 13].

Combining all relaxations from Sections 4.1 to 4.4, we obtain

$$\begin{aligned} \min_{\hat{\mathbf{p}}\in[0,1]^l,\, t_1,\, \mathbf{Q}\in\mathbf{S}_{++}^{r+2}} \quad & \mathbf{f}^H \mathbf{Q} \mathbf{f} + \alpha \max\left(0,\, t_1 - \beta^{-1} P_x\right) && \text{(20a)} \\ \text{s.t.} \quad & \begin{bmatrix} -\mathbf{G}^{-1} - \lambda^{-1} \operatorname{diag}\left(\begin{bmatrix} \mathbf{1}_2^T & \hat{\mathbf{p}}^T \end{bmatrix}\right) & \mathbf{G}^{-1} \\ \mathbf{G}^{-1} & \mathbf{Q} - \mathbf{G}^{-1} \end{bmatrix} \succeq 0 && \text{(20b)} \\ & \begin{bmatrix} \mathbf{I} + (\frac{1}{\epsilon} - 1) \operatorname{diag}(\hat{\mathbf{p}}) & \mathbf{U}\tilde{\mathbf{B}}^H \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \mathbf{Q} \mathbf{f} \\ \mathbf{f}^H \mathbf{Q} \mathbf{\Lambda}^H \tilde{\mathbf{P}}^{-1} \tilde{\mathbf{B}} \mathbf{U} & t_1 \end{bmatrix} \succeq 0 && \text{(20c)} \\ & \mathbf{1}^T \hat{\mathbf{p}} \le 2M - 3\,, && \text{(20d)} \end{aligned}$$

as the complete convex optimization problem. After solving Eq. (20) for $\hat{\mathbf{p}}$ we apply randomized rounding to obtain a binary $\mathbf{p}$. The resulting $\mathbf{p}$ we use to construct the constraint matrix $\mathbf{C_p}$ as described by Eq. (8) to use with the binaural LCMV solution in Eq. (6).



(a) Average SNR gains.

(b) Average weighted ITF errors of the interferers.

Fig. 1: Experimental results. Solid lines indicate the use of true RTFs in $\mathbf{\Lambda}_b$, while dashed lines indicate the use of PRTFs.

5. EXPERIMENTS

The proposed optimization problem in Eq. (20) is solved using CVX [14] and SeDuMi [15] in Matlab. The parameter ϵ is set to 100^{-1} and the penalty weight α is set to 10^6, such that violation of the original threshold is strongly discouraged. The desired binaurally unconstrained SIR is chosen to be $\beta = 10\,\mathrm{dB}$.

Experiments are done using either the true RTFs of the interferers or PRTFs. When using PRTFs, we use 24 PRTFs placed uniformly around the user's head in the horizontal plane at a distance of $3\,\mathrm{m}$. The RTFs are constructed using anechoic chamber head related impulse responses, using 2 microphones on each ear [16]. The intensity matrix $\mathbf{U}$ is determined using MVDR beamscanning [17, 1].

To evaluate, we use average SNR gain [4] and average weighted interaural transfer function errors [18, 8]. The weights are based on simultaneous masking [19], such that less audible interferers will be weighted less. Using true RTFs, we compare the proposed method with an exhaustive search (ES) implementation of the proposed framework, the BMVDR [7], and the JBLCMV [4, 7]. In addition, we evaluate the proposed method versus the JBLCMV using PRTFs.

5.1. Acoustic Scene

The simulated acoustic scene uses 3 to 8 interfering speech-shaped or white noise sources in arbitrary locations in the horizontal plane. They are all at $3\,\mathrm{m}$ distance from the user and their impulse responses are determined using head related impulse responses from [16]. The target is in front of the user at $80\,\mathrm{cm}$. Finally, additive white Gaussian noise, with an SNR of $50\,\mathrm{dB}$ compared to the target, is added to all microphone recordings independently, to simulate microphone self noise.

5.2. Results

The results are shown in Fig. 1. Using true RTFs (solid lines), the proposed method is close to the ES in terms of SNR gain and ITF error. This shows how close the convex relaxations are to the original problem formulation in Eq. (9). In addition, compared to the JBLCMV, the proposed method leads to significantly increased SNR gains, while still having lower spatial cue distortions compared to the BMVDR. Further increasing the number of interferers, the proposed method even obtains lower spatial cue distortions than the JBLCMV. Using PRTFs (dashed lines), the proposed method still has some improved SNR gain over the JBLCMV with PRTFs. More importantly, we see in Fig. 1b that the JBLCMV with naive constraint selection performs very poorly in terms of binaural cue preservation compared to our proposed selection method, which even leads to a lower ITF error than when using the true RTFs. This is because of the large number of PRTFs that it can select from. However, when the number of active interferers is larger than five, the proposed method starts to introduce spatial cue distortions as well. This is as expected, as the proposed method can only select $2M - 3 = 5$ interferers simultaneously. Additionally, there may not exist a PRTF that is colocated with an actual interferer. Using auditory and perceptual models within the proposed framework, this could be further improved. In [1], more detailed results are presented and discussed.

6. CONCLUSION

A new constraint selection method for binaural LCMV-based beamformers has been proposed. It has been shown to improve the performance in terms of noise reduction and binaural cue preservation compared to a naive selection method and improved binaural cue preservation compared to the BMVDR beamformer. This allows binaural beamforming methods like the JBLCMV to function with more than $2M - 3$ (virtual) interferers present. Furthermore, the proposed method improves its noise reducing capabilities overall, by unconstraining inaudible sources in relevant time-frequency bins.

Interesting areas of further research include combining interferer selection with other types of beamformers such as the RJBLCMV [5], which bounds binaural cue errors instead of exactly constraining them. Additionally, further research in binaural cue preserving methods would benefit greatly from binaural cue error measures more closely related the perception of such errors.

7. REFERENCES

[1] C. A. Kokke, "Interferer selection for binaural cue preservation in joint binaural linearly constrained minimum variance beam-

forming," M. Sc. thesis, TU Delft, Delft, the Netherlands, Aug. 2018.

[2] World Health Organization et al., "Hearing loss due to recreational exposure to loud sounds: a review," 2015.

[3] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acustica united with Acustica*, vol. 86, no. 1, pp. 117–128, 2000.

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

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

[6] S. Doclo, T. J. Klasen, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel wiener filtering and interaural transfer functions," in *Proc. Int. Workshop Acoust. Echo Noise Control (IWAENC)*, 2006.

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

[8] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, J. Jensen, and M. Guo, "Binaural beamforming using pre-determined relative acoustic transfer functions," in *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE, 2017, pp. 1–5.

[9] S. P. Chepuri and G. Leus, "Sparse sensing for distributed detection," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1446–1460, 2016.

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

[11] J. Franklin, *Matrix theory*, Dover Publications, Mineola, N.Y, 2000.

[12] J. Zhang, R. Heusdens, and R. C. Hendriks, "Rate-distributed spatial filtering based noise reduction in wireless acoustic sensor networks," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, no. 11, pp. 2015–2026, 2018.

[13] J. Zhang, S. P. Chepuri, R. C. Hendriks, and R. Heusdens, "Microphone subset selection for mvdr beamformer based noise reduction," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 550–563, 2018.

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

[15] J. F Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[16] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, pp. 6, 2009.

[17] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE signal processing magazine*, vol. 13, no. 4, pp. 67–94, 1996.

[18] A. I. Koutrouvelis, J. Jensen, M. Guo, R. C. Hendriks, and R. Heusdens, "Binaural speech enhancement with spatial cue preservation utilising simultaneous masking," in *Signal Processing Conference (EUSIPCO), 2017 25th European*. IEEE, 2017, pp. 598–602.

[19] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–515, 2000.