Binaural Beamforming Using Pre-Determined Relative Acoustic Transfer Functions

Andreas I. Koutrouvelis*, Richard C. Hendriks*, Richard Heusdens*, Jesper Jensen†‡ and Meng Guo‡
e-mails: {a.koutrouvelis, r.c.hendriks, r.heusdens}@tudelft.nl and {jesj, megu}@oticon.com
*Circuits and Systems (CAS) Group, Delft University of Technology, the Netherlands
†Dept. Electronic Systems, Aalborg University, Denmark, ‡Oticon A/S, Denmark

***Abstract*—Binaural beamformers (BFs) aim to reduce the output noise power while simultaneously preserving the binaural cues of all sources. Typically, the latter is accomplished via constraints relating the output and input interaural transfer functions (ITFs). The ITF is a function of the corresponding relative acoustic transfer function (RATF), which implies that RATF estimates of all sources in the acoustic scene are required. Here, we propose an alternative way to approximately preserve the binaural cues of the entire acoustic scene without estimating RATFs. We propose to preserve the binaural cues of all sources with a set of fixed pre-determined RATFs distributed around the head. Two recently proposed binaural BFs are evaluated in the context of using pre-determined RATFs and compared to the binaural minimum variance distrortionless response BF which can only preserve the binaural cues of the target.**

***Index Terms*—Binaural beamforming, interaural transfer function (ITF), relative acoustic transfer function (RATF).**

I. INTRODUCTION

Binaural hearing aid (HA) systems typically consist of two HAs, one at each ear, where each HA is typically equipped with multiple microphones. Allowing the HAs to collaborate, and combine their noisy microphone signals into a multi-microphone noise reduction algorithm, e.g., [1], [2], is an efficient way to achieve acoustic noise reduction. Unlike traditional monaural beamformers (BFs), e.g., [3], [4], which mainly focus on noise reduction, binaural BFs also aim to preserve the binaural cues of the sources in the acoustic scene [1]. This can be achieved through proper combination of the multi-microphone recordings of both HAs.

Many binaural BFs are based on the linearly constrained minimum variance (LCMV) framework [5]. This is due to the elegant and simple way in which constraints can be incorporated, as well as due to efficient adaptive implementations [4], [6]. The LCMV minimizes the output noise power under several linear equality constraints. In the case of binaural beamforming, these are often used to preserve the binaural cues of the present sources, while leaving the target signal undistorted at the two reference microphones. A different category of binaural noise reduction methods is based on the multi-channel Wiener filter (MWF) framework [7], [8]. The MWF-based methods [9]–[11] can achieve higher signal-to-noise-ratio (SNR) gains, but unlike the LCMV, they typically distort the target signal.

This work was supported by the Oticon Foundation and the Dutch Technology Foundation STW.

The binaural minimum variance distortionless response (BMVDR) BF [2] uses only two linear constraints to guarantee a distortionless response of the target at the two reference microphones. This results in binaural-cue preservation of the target source. Although this method achieves a relatively high binaural SNR gain, the price to pay is that the binaural cues of all interferers become identical to the binaural cues of the target after processing. The binaural LCMV (BLCMV) BF [12] preserves the binaural cues of the target, as well as of multiple interferers. This is achieved using two additional constraints per interferer. As a result, the degrees of freedom are exhausted fast for a small number of microphones. In contrast, the joint binaural LCMV (JBLCMV) BF [2], [13] achieves binaural-cue preservation using only one constraint per interferer. Thus, the JBLCMV can preserve the binaural cues of more interferers than the BLCMV [13].

Usually, the number of microphones per HA is relatively small, say, 2 or 3. As a result, the BLCMV and even the JBLCMV, suffer from the fact that the degrees of freedom are quickly exhausted with an increasing number of sources. This results in poor SNR gains and a small number of sources for which the binaural cues can be preserved. To overcome this problem, a relaxation of the JBLCMV method is proposed in [14]. In the current paper we refer to this method as relaxed JBLCMV (RJBLCMV). The equality constraints, used in the JBLCMV, which are meant to preserve the binaural cues of the interferers, are now replaced with inequality constraints. As a result, the binaural cues of the interferers are approximately preserved. The inequalities allow the RJBLCMV to use a larger number of constraints (and approximately preserve the binaural cues of more interferers) than other LCMV-based methods with equality constraints only, or, alternatively, to use the same number of constraints, but to trade-off binaural-cue accuracy against SNR gain.

An important limitation of all the aforementioned binaural BFs is that they require estimates of the acoustic transfer functions (ATFs) or relative ATFs (RATFs) of the sources to form the constraints. This is rather impractical as estimation of these ATFs/RATFs is very challenging, in particular in dynamic scenarios. In this paper, we present a solution to this problem using fixed pre-determined RATFs, independent of the acoustical scenario. As a result, no tracking nor estimation of RATFs is needed. These pre-determined RATFs correspond to locations around the head. Each pre-determined RATF

ISBN 978-0-9928626-7-1 © EURASIP 2017

covers a small area in which some interferers might be present. As we use pre-determined RATFs instead of the true RATFs, steering vector mismatches (SVMs) are expected, potentially leading to a reduced preservation of the binaural cues. Increasing the number of pre-determined RATFs, however, leads to a lower expected SVM. We investigate both the JBLCMV and the RJBLCMV in the context of pre-determined RATFs, since these two methods can preserve the binaural cues of more locations than the BLCMV [13], [14]. It is to be expected that the RJBLCMV will be less sensitive to such SVMs as it, typically, allows to include much more constraints due to the introduced relaxation in binaural-cue preservation.

To guide the reader, in Section II the signal model and notation are presented. In Section III the idea of using pre-determined RATFs is introduced. In Section IV the JBLCMV method is reviewed in the context of the pre-determined RATFs. In Section V we provide a useful decomposition of the JBLCMV spatial filter that explains the SVM problem due to the usage of pre-determined RATFs. In the same section, we also propose how to mitigate the SVM problem. In Section VI the RJBLCMV method is reviewed in the context of the pre-determined RATFs. In Section VII, we evaluate the JBLCMV and RJBCMV using pre-determined RATFs. Finally, the conclusion is provided in Section VIII.

II. Signal Model & Notation

Without loss of generality, let us assume that each of the two HAs has $M/2$ microphones, i.e., a total of M microphones. The processing is done in the discrete Fourier transform domain on a frame-by-frame basis, independently for each frequency bin. The noisy vector acquired from the M-microphone array for a single frequency bin is given by

$$\mathbf{y} = s\mathbf{a} + \sum_{i=1}^{r} u_i \mathbf{b}_i + \mathbf{v} \in \mathbb{C}^{M\times 1}, \tag{1}$$

where r is the number of interferers, $\mathbf{a}$ and $\mathbf{b}_i$ are the ATFs of the target and the i-th interferer, s and u_i are the target and the i-th interferer at the original locations, respectively, and $\mathbf{v}$ represents the background noise vector. Note that the first $M/2$ elements and the last $M/2$ elements of all vectors in Eq. (1) correspond to the left and right HAs, respectively. The first and last microphone of the M-microphone array are considered as the left and right reference microphones for binaural beamforming. Thus, for convenience, the first and last element of all vectors of Eq. (1) are indexed with subscript L and R, respectively, i.e., $\mathbf{a} = [\alpha_L, \alpha_2 ..., \alpha_{M-1}, \alpha_R]^T$ and $\mathbf{b}_i = [b_{i,L}, b_{i,2}, ..., b_{i,M-1}, b_{i,R}]^T$, etc. Each ATF is typically associated with a couple of RATFs. The RATFs of the target with respect to the left and right reference microphones are given by $\bar{\mathbf{a}}_L = \mathbf{a}/a_L$ and $\bar{\mathbf{a}}_R = \mathbf{a}/a_R$, respectively, and for the i-th interferer $\bar{\mathbf{b}}_{iL} = \mathbf{b}_i/b_{i,L}$ and $\bar{\mathbf{b}}_{iR} = \mathbf{b}_i/b_{i,R}$.

Assuming that all sources in Eq. (1) are mutually uncorrelated, the cross power spectral density matrix (CPSDM) of the noisy measurements, $\mathbf{P_y} \in \mathbb{C}^{M\times M}$, is given by

$$\mathbf{P_y} = \mathrm{E}\left[\mathbf{y}\mathbf{y}^H\right] = p_s \mathbf{a}\mathbf{a}^H + \underbrace{\sum_{i=1}^{r} p_{u_i} \mathbf{b}_i \mathbf{b}_i^H + \mathbf{P_v}}_{\mathbf{P}}, \tag{2}$$

where $\mathrm{E}[\cdot]$ denotes statistical expectation, $\mathbf{P}$ is the CPSDM of the total noise, p_s and p_{u_i} are the power spectral densities of the target and the i-th interferer signals, respectively, and $\mathbf{P_v} = \mathrm{E}[\mathbf{v}\mathbf{v}^H]$ is the CPSDM of the background noise.

The binaural BFs consists of two spatial filters $\mathbf{w}_L, \mathbf{w}_R$ which are applied to $\mathbf{y}$, producing the outputs $x_L = \mathbf{w}_L^H\mathbf{y}$ and $x_R = \mathbf{w}_R^H\mathbf{y}$ at the left and right HAs, respectively.

III. Pre-Determined RATFs in Binaural Beamforming

In this section, we introduce the notion of using pre-determined RATFs in binaural beamforming. Specifically, we use m couples of pre-determined RATFs, i.e., $(\bar{\mathbf{q}}_{iL}, \bar{\mathbf{q}}_{iR})$, $i = 1, 2, \cdots, m$, where $\bar{\mathbf{q}}_{iL} = \mathbf{q}_i/q_{i,L}$ and $\bar{\mathbf{q}}_{iR} = \mathbf{q}_i/q_{i,R}$ are the pre-determined RATFs with respect to the left and right reference microphones, respectively, $\mathbf{q}_i$ is the corresponding pre-determined ATF, and $q_{i,L}$ and $q_{i,R}$ are the first and last elements of $\mathbf{q}_i$. Each pre-determined RATF couple, $(\bar{\mathbf{q}}_{iL}, \bar{\mathbf{q}}_{iR})$, corresponds to a pre-selected location in space with polar coordinates (θ_i, ϕ_i, h_i), where θ_i is the azimuth, ϕ_i the elevation, and h_i the distance from the center of the head. Note that the pre-determined RATFs are acoustic scene independent, but user dependent. Specifically, every user has its own set of anechoic head related transfer functions (HRTFs) which are used as pre-determined RATFs.

Without loss of generality, we examine the scenario where the m pre-selected locations are placed uniformly on the perimeter of a circle on the horizontal plane with radius h centered at the center of the head as shown in Fig. 1. As a result, we consider all azimuths equally important for binaural-cue preservation. The circle is selected to have a radius of $h > 2d^2/\lambda_{\text{min}}$ m, where d is the distance between the two HAs and $\lambda_{\text{min}} = 2c/F_s$, where c is the speed of sound and F_s is the sampling frequency. This is because, at this distance, the far field assumption is approximately met [15]. Consequently, the pre-determined RATFs are approximately distant invariant, i.e., there is no need to use more pre-determined RATFs for greater distances. Here, we assume that all sources are in the far-field, i.e., their distances are greater than h. A better approach, especially for nearby sources, is to have pre-determined RATFs for different elevations as well. For now we restrict ourselves to a single elevation.

If one of the m pre-determined RATF-couples a) matches with the actual RATF-couple of an interferer, i.e., $\exists j, i$: $(\bar{\mathbf{q}}_{jL}, \bar{\mathbf{q}}_{jR}) = (\bar{\mathbf{b}}_{iL}, \bar{\mathbf{b}}_{iR})$, and b) is included in the constraints of an LCMV-based BF, the binaural cues of the interferer will be preserved. However, more interestingly (and more likely) is the case where there are interferers in the acoustic scene whose RATF-couple does not match with one of the pre-determined RATF-couples. This results in SVMs. Obviously, the expected

ISBN 978-0-9928626-7-1 © EURASIP 2017

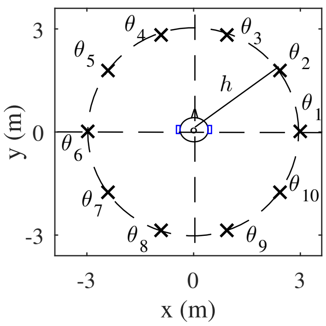

Fig. 1. Example: $m = 10$, $h = 3$ m, 'x' markers denote locations of pre-determined RATFs around the head which is centered at the origin $(0, 0)$.

SVM decreases when m is increased. An objective measure for the binaural-cue preservation of the i-th interferer after processing is the ITF error given by [14]

$$\mathcal{E}_i = \left|\text{ITF}_i^{\text{out}} - \text{ITF}_i^{\text{in}}\right| = \left|\frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} - \frac{b_{i,L}}{b_{i,R}}\right|, \; i = 1, \cdots, r. \quad (3)$$

If $\mathcal{E}_i = 0$, the binaural BF preserves exactly the binaural cues of the i-th interferer. Since we use pre-determined RATFs, we also define the pre-determined ITF error which is given as

$$\mathcal{E}_i^q = \left|\frac{\mathbf{w}_L^H \mathbf{q}_i}{\mathbf{w}_R^H \mathbf{q}_i} - \frac{q_{i,L}}{q_{i,R}}\right| = \left|\frac{\mathbf{w}_L^H \bar{\mathbf{q}}_{iR}}{\mathbf{w}_R^H \bar{\mathbf{q}}_{iR}} - \bar{q}_{iR,1}\right|, \; i = 1, \cdots, m. \quad (4)$$

The binaural BF methods, discussed in the sequel, constrain the error $\sum_{i=1}^r \mathcal{E}_i^q$. Ideally, $\sum_{i=1}^r \mathcal{E}_i$ should be constrained as well. Constraining $\sum_{i=1}^r \mathcal{E}_i$, by constraining $\sum_{i=1}^r \mathcal{E}_i^q$, depends on a) how close the sources are to the pre-determined RATFs or, equivalently, it depends on how many pre-determined RATFs are used, and b) the number of the available degrees of freedom for noise reduction (see Section V).

IV. JBLCMV

The joint binaural LCMV (JBLCMV) spatial filter [2], [13], [14] is obtained by the following LCMV problem

$$\hat{\mathbf{w}} = \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}}{\arg\min} \; \underbrace{\mathbf{w}_L^H \mathbf{P} \mathbf{w}_L + \mathbf{w}_R^H \mathbf{P} \mathbf{w}_R}_{\mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w}} \; \text{s.t. } \mathbf{w}^H \boldsymbol{\Lambda} = \mathbf{f}^H, \quad (5)$$

where $\mathbf{w} = [\mathbf{w}_L^T \;\; \mathbf{w}_R^T]^T \in \mathbb{C}^{2M \times 1}$, $\boldsymbol{\Lambda} \in \mathbb{C}^{2M \times (2+m)}$ is assumed full column rank, $\tilde{\mathbf{P}} = \text{diag}(\{\mathbf{P}, \mathbf{P}\}) \in \mathbb{C}^{2M \times 2M}$ is a block diagonal matrix, and $\mathbf{w}^H \boldsymbol{\Lambda} = \mathbf{f}^H$ is a set of $2 + m$ linear equality constraints. The constraints aim at a) the preservation of the target at the two reference microphones, which also implies that its binaural cues are preserved, and b) the preservation of the binaural cues of m pre-selected locations, as proposed in Section III. The first goal is accomplished via two linear constraints given by

$$\begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix} \begin{bmatrix} \bar{\mathbf{a}}_L & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{a}}_R \end{bmatrix} = [1 \quad 1]. \quad (6)$$

The second goal is accomplished by forcing the output ITF to be equal to the input ITF for each of the m pre-selected locations. This is accomplished using m linear constraints, i.e.,

$$\mathbf{w}_L^H \bar{\mathbf{q}}_{iL} - \mathbf{w}_R^H \bar{\mathbf{q}}_{iR} = 0, \; i = 1, \cdots, m. \quad (7)$$

Putting together all constraints we have

$$\underbrace{\begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix}}_{\mathbf{w}^H} \underbrace{\begin{bmatrix} \bar{\mathbf{a}}_L & \mathbf{0} & \bar{\mathbf{q}}_{1L} & \cdots & \bar{\mathbf{q}}_{mL} \\ \mathbf{0} & \bar{\mathbf{a}}_R & -\bar{\mathbf{q}}_{1R} & \cdots & -\bar{\mathbf{q}}_{mR} \end{bmatrix}}_{\boldsymbol{\Lambda} \in \mathbb{C}^{2M \times 2+m}} = \underbrace{[1 \; 1 \; 0 \cdots 0]}_{\mathbf{f}^H}. \quad (8)$$

Note that the available degrees of freedom for noise reduction is $2M - m - 2$ and the maximum number of constraints that can be used for binaural-cue preservation of interferers/locations, while having at least one degree of freedom for noise reduction, is $2M - 3$. The problem in Eq. (5) has a closed-form solution given by [4]

$$\hat{\mathbf{w}} = \tilde{\mathbf{P}}^{-1} \boldsymbol{\Lambda} \left(\boldsymbol{\Lambda}^H \tilde{\mathbf{P}}^{-1} \boldsymbol{\Lambda}\right)^{-1} \mathbf{f}. \quad (9)$$

Note that the BMVDR BF [2] is also obtained from the optimization problem in Eq. (5), but with $m = 0$, i.e., the constraints in Eq. (7) are not used. It can be easily shown that the ITF error (see Eq. (3)) of the BMVDR is given by [14]

$$\mathcal{E}_i^{\text{BMVDR}} = \left|\frac{a_L}{a_R} - \frac{b_{iL}}{b_{iR}}\right| = \left|\bar{a}_{R,1} - \bar{b}_{iR,1}\right|, i = 1, \cdots, r, \quad (10)$$

while the pre-determined ITF error (see Eq. (4)) is given by

$$\mathcal{E}_i^{q,\text{BMVDR}} = \left|\frac{a_L}{a_R} - \frac{q_{iL}}{q_{iR}}\right| = \left|\bar{a}_{R,1} - \bar{q}_{iR,1}\right|, i = 1, \cdots, m. \quad (11)$$

V. SVM Problem

In this section, we provide a useful decomposition of the JBLCMV spatial filter that helps us to understand the SVM problem and how to handle it by using pre-determined RATFs. It is easy to show that if $\boldsymbol{\Lambda}$ and $\mathbf{f}$ in Eq. (8) are substituted into Eq. (9), the left and right spatial filters of the JBLCMV are given by

$$\hat{\mathbf{w}}_L = \rho_{L0} \mathbf{w}_{L0} + \rho_{L1} \mathbf{w}_{L1} + \cdots + \rho_{Lm} \mathbf{w}_{Lm}, \quad (12)$$

$$\hat{\mathbf{w}}_R = \rho_{R0} \mathbf{w}_{R0} + \rho_{R1} \mathbf{w}_{R1} + \cdots + \rho_{Rm} \mathbf{w}_{Rm}, \quad (13)$$

where

$$\mathbf{w}_{L0} = \frac{\mathbf{P}^{-1} \bar{\mathbf{a}}_L}{\bar{\mathbf{a}}_L^H \mathbf{P}^{-1} \bar{\mathbf{a}}_L}, \mathbf{w}_{R0} = \frac{\mathbf{P}^{-1} \bar{\mathbf{a}}_R}{\bar{\mathbf{a}}_R^H \mathbf{P}^{-1} \bar{\mathbf{a}}_R}, \quad (14)$$

$$\mathbf{w}_{Li} = \frac{\mathbf{P}^{-1} \bar{\mathbf{q}}_{iL}}{\bar{\mathbf{q}}_{iL}^H \mathbf{P}^{-1} \bar{\mathbf{q}}_{iL}}, \mathbf{w}_{Ri} = \frac{\mathbf{P}^{-1} \bar{\mathbf{q}}_{iR}}{\bar{\mathbf{q}}_{iR}^H \mathbf{P}^{-1} \bar{\mathbf{q}}_{iR}}, \quad (15)$$

and where $\rho_{Li}, \rho_{Ri}, i = 0, \cdots, m$ are functions of several generalized inner products of the form $\mathbf{z}^H \mathbf{P}^{-1} \mathbf{g}$, where $\mathbf{z}$, $\mathbf{g}$ are RATFs of the target or of the i-th pre-selected location. Note that $\mathbf{w}_{L0}$ and $\mathbf{w}_{R0}$ are the left and right MVDR BFs, of the BMVDR BF, preserving the target at the two reference microphones, while suppressing the interferers. Moreover, $\mathbf{w}_{Li}$ and $\mathbf{w}_{Ri}$ are the left and right minimum power distortionless response (MPDR) BFs [5] preserving the possible interferers close to the i-th pre-selected location at the two reference microphones while suppressing the remaining interferers which are further away. Note that we used the term MPDR for $\mathbf{w}_{Li}$ and $\mathbf{w}_{Ri}$, because we can think of $(\bar{\mathbf{q}}_{iL}, \bar{\mathbf{q}}_{iR})$ as an estimate of one or more actual RATF-couples of interferers (which are possibly close to the i-th pre-selected location) which are also present in the CPSDM $\mathbf{P}$.

It is widely known that the MPDR BF is not robust to SVMs [5], [16], [17]. The two MVDR BFs $\mathbf{w}_{L0}$ and $\mathbf{w}_{R0}$ are robust to SVMs, because the target is not present in $\mathbf{P}$ [17]. However, $\mathbf{w}_{Li}$ and $\mathbf{w}_{Ri}$ are most likely not robust to SVMs, because the interferers are present in $\mathbf{P}$, but some of the interferers might be far away from the m pre-selected locations. Therefore, the probable SVMs will most likely result in an uncontrolled amount of suppression of the r interferers from the non-robust BFs $\mathbf{w}_{Li}$ and $\mathbf{w}_{Ri}$. This will probably result in binaural-cue distortions of the interferers. Obviously, the expected SVM increase when the number of pre-determined RATFs decreases. Moreover, in [17] it was shown that the sensitivity of the MPDR BF to SVM increases when the maximum possible SNR (i.e., the SNR that can be achieved with no SVM) of the MPDR BF increases. In other words, when the number of degrees of freedom for noise reduction increases (i.e., $2M - m - 2$ increases), the SVM sensitivity, typically, increases.

VI. RJBLCMV

The RJBLCMV [14] is a BF that relaxes some of the equality constraints and, thus, can, typically, use more constraints, for binaural-cue preservation, than the other LCMV-based methods. As a result, the RJBLCMV can preserve the binaural cues of more sources/locations than JBLCMV. The constraints in Eq. (8) can be partitioned as $\mathbf{w}^H \left[\mathbf{\Lambda}_1 \quad \mathbf{\Lambda}_2\right] = \left[\mathbf{f}_1^H \quad \mathbf{f}_2^H\right]$, where $\mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H$ contains the two constraints in Eq. (6) and $\mathbf{w}^H \mathbf{\Lambda}_2 = \mathbf{f}_2^H$ contains the pre-determined constraints in Eq. (7). The RJBLCMV makes use of this separation by having strict constraints with respect to the target, but inequality constraints on the m pre-selected locations, i.e.,

$$\hat{\mathbf{w}} = \underset{\mathbf{w}\in\mathbb{C}^{2M\times 1}}{\operatorname{argmin}} \, \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H,$$
$$\left| \frac{\mathbf{w}_L^H \mathbf{q}_i}{\mathbf{w}_R^H \mathbf{q}_i} - \frac{q_{i,L}}{q_{i,R}} \right| \leq c\mathcal{E}_i^{q,\text{BMVDR}}, \quad i = 1, \cdots, m, \quad (16)$$

where $\mathcal{E}_i^{q,\text{BMVDR}}$ is given in Eq. (11), and $0 \leq c \leq 1$ is a user-defined parameter that controls the trade-off between binaural-cue accuracy and SNR gain. The maximum and minimum allowable amount of relaxation are obtained for $c = 1$ and $c = 0$, respectively. Having $0 \leq c \leq 1$ allows to relax the amount of binaural-cue preservation and trade this off with SNR gain, while guaranteeing that the amount of ITF error for a certain pre-selected location is always a proportion c below the BMVDR pre-determined ITF error. This does not necessarily imply that the ITF errors of the interferers will be a proportion c below of the BMVDR ITF error (see Eq. (3)). However, in Section VII, we experimentally show that for a large enough number of pre-determined RATFs, m, the average ITF error of the interferers is approximately a proportion c below of the BMVDR average ITF error.

The inequalities with the m pre-determined ATFs, $\mathbf{q}_i, i = 1, \cdots, m$, can be written in terms of pre-determined RATFs by multiplying both sides with $|q_{i,R}/q_{i,L}| = |\bar{q}_{iL,M}|$, where $\bar{q}_{iL,M}$ is the last element of $\bar{\mathbf{q}}_{iL}$. Therefore, the problem in Eq. (16) can be equivalently written as [14]

$$\hat{\mathbf{w}} = \underset{\mathbf{w}\in\mathbb{C}^{2M\times 1}}{\operatorname{argmin}} \, \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \text{ s.t. } \mathbf{w}^H \mathbf{\Lambda}_1 = \mathbf{f}_1^H,$$
$$|\mathbf{w}^H \mathbf{\Lambda}_{2,i}| \leq f_{2,i} \text{ for } i = 1, \cdots, m, \quad (17)$$

where $\mathbf{\Lambda}_{2,i}$ is the i-th column of $\mathbf{\Lambda}_2$, $f_{2,i} = |c\mathcal{E}_i^{q,\text{BMVDR}} \mathbf{w}_R^H \bar{\mathbf{q}}_{iR} \bar{q}_{iL,M}|$. The problem in Eq. (17) can be interpreted as a relaxed version of the JBLCMV. Note that if $c = 0$ in Eq. (17), the JBLCMV is obtained. The problem in Eq. (17) is non-convex and is approximately solved interatively as proposed in [14].

Unlike JBLCMV, the RJBLCMV is typically able to provide feasible solutions for $m > 2M - 2$ [14], which makes it applicable for the approximate preservation of binaural cues of more interferers/locations compared to the other strict equality constraint LCMV-based methods. Moreover, for the same number of constraints, the RJBLCMV can trade binaural-cue accuracy with improved SNR gain compared to JBLCMV [14]. As noted before, the very small number of available microphones in both HAs, limit the LCMV-based methods to preserve the binaural cues of only a very small number of pre-selected locations and, thus, the expected SVM is expected to be large. On the contrary, the RJBLCMV can typically approximately preserve much more locations and as we will see in Section VII, the average ITF error is smaller than with JBLCMV.

VII. Experiments

The JBLCMV and RJBLCMV methods, using pre-determined RATFs, are evaluated in terms of noise reduction and binaural-cue preservation and compared to the BMVDR. Noise reduction performance is measured with binaural SNR gain averaged overal all frequencies and frames (as in [14]) and the binaural cue preservation is measured with ITF error (see Eq. (3)) averaged over all frequencies and interferers as in [14]. In order to construct the microphone signals, and the pre-determined RATFs, we used the anechoic HRTFs from the database in [18]. The number of microphones that we used is $M = 6$, i.e., three microphones at each HA. The sampling frequency is 16 kHz, the frame length is 10 ms with an overlap of 50%, and the FFT length is 512.

The target is approximately in the look direction (i.e., $-5°$), with a distance of 0.8 m from the origin. We used its actual RATF (i.e., the ITF error of the target is zero for all methods) to form the distortionless constraints for all methods. The m pre-determined RATFs were selected as described in Section III with $h = 3$ m. We considered 8 simultaneously present speech shaped noise interferers with the same power as the target signal at the point that originates. Each one is randomly placed at one of the 72 possible angles of the HRTF database [18], with equal probability. The distance of the interferers from the origin is 3 m. The CPSDM $\mathbf{P}$ was computed using the true RATFs and estimated power spectral densities of the present interferers using all available data. Therefore, we examine the best possible performance of the

ISBN 978-0-9928626-7-1 © EURASIP 2017

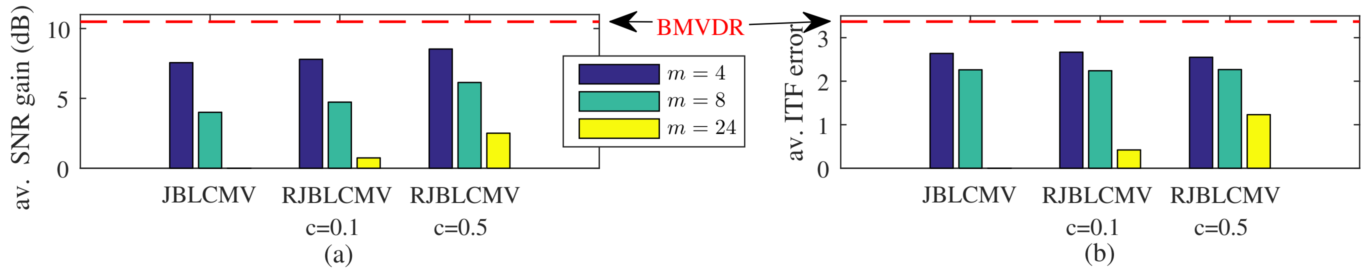

Fig. 2. Performance of RJBLCMV (using $c = 0.1, 0.5$, and $m = 4, 8, 24$), JBLCMV (using $m = 4, 8$) and BMVDR (denoted with red dashed line) with respect to (a) average binaural SNR gain, and (b) average ITF error.

competing methods, since no realistic estimation errors of $\mathbf{P}$ are considered. The background noise was simulated as white Gaussian noise with the same power at all microphones, with SNR $= 50$ dB with respect to the target signal at the left reference microphone. The maximum number of iterations for RJBLCMV were selected $k_{\text{max}} = 10$ as in [14].

Fig. 2 shows the average performance for 20 different random placements of all interferers, where each random placement is accomplished as explained before. The RJBLCMV is evaluated with $m = 4, 8, 24$ pre-determined RATFs, while the JBLCMV with only $m = 4, 8$. This is because the JBLCMV can use up to $2M - 3 = 9$ pre-determined RATFs. It is clear that the JBLCMV has poor performance, because it cannot use many constraints (i.e., many pre-determined RATFs) for $M = 6$, and the exptected SVM is large. On the other hand, RJBLCMV can achieve significantly better preservation of binaural cues even with just $M = 6$ microphones while still having quite a reasonable SNR gain. This is because the RJBLCMV can use much more constraints than the JBLCMV. As expected, the RJBLCMV with $c = 0.5$ has a larger SNR gain and ITF error, than with $c = 0.1$. Moreover, for $m = 24$, the RJBLCMV indeed achieves an average ITF error which is approximately c times below of the BMVDR average ITF error. Finally, both JBLCMV and RJBLCMV achieve a better binaural-cue preservation accuracy than the BMVDR.

VIII. Conclusion

A novel idea is presented for binaural-cue preservation without the need to estimate the relative acoustic transfer functions (RATFs). It is proposed to use pre-determined RATFs around the head of the hearing-aid user. The more pre-determined RATFs are used, the smaller the expected steering vector mismatch of the actual sources and the better the control of binaural-cue preservation. The pre-determined RATFs can be used in both the RJBLCMV and the LCMV-based methods. However, it is shown that only the RJBLCMV is promising in this context, because it can use much more constraints than the other LCMV-based BFs, for a small number of microphones.

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] 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.

[3] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[4] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. of the IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

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

[6] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas, Propag.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.

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

[8] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Process.*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.

[9] 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.

[10] 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 *Int. Workshop Acoustic Echo, Noise Control (IWAENC)*, Sep. 2006.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] R. Kennedy, T. Abhayapala, and D. Ward, "Broadband nearfield beamforming using a radial beampattern transformation," *IEEE Trans. Signal Process.*, vol. 46, no. 8, pp. 2147–2156, Aug. 1998.

[16] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *ELSEVIER Signal Process.*, vol. 93, no. 12, pp. 3264–3277, Dec. 2013.

[17] H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," *J. Acoust. Soc. Amer.*, vol. 54, no. 3, pp. 771–785, 1973.

[18] 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.

ISBN 978-0-9928626-7-1 © EURASIP 2017