Binaural beam-forming with dominant spatial cue preservation for hearing aids

Vasudha Sathyapriyan
Circuits and Systems (CAS) Group
Delft University of Technology
Delft, The Netherlands
E-mail: vasudhasathya@gmail.com

Metin Çalış
Circuits and Systems (CAS) Group
Delft University of Technology
Delft, The Netherlands
E-mail: M.Calis@tudelft.nl

Richard C. Hendriks
Circuits and Systems (CAS) Group
Delft University of Technology
Delft, The Netherlands
E-mail: R.C.Hendriks@tudelft.nl

***Abstract*—Many beam-forming algorithms available for hearing aids, preserve both the interaural time differences (ITDs) and the interaural level differences (ILDs) of the interferers. Constraining both the spatial cues over all frequencies will exhaust the degrees of freedom (DoF) available for noise reduction in the filter design. The binaural cues, however, are frequency selective, i.e., the ITDs are dominant in the frequency range below 1.5 kHz and the ILDs are dominant at frequencies above 1.5 kHz. Hence in this paper, we propose two methods to preserve only the ILDs of the interferers, in the higher frequencies, while keeping the target undistorted. Since these formulations are non-convex, quadratically constrained quadratic programs (QCQPs), an approximate convex relaxation is proposed. The proposed methods preserve the ILDs in the higher frequencies while an available algorithm that preserves interaural transfer function (ITF) of the interferers is used for the lower frequencies. The performance of the methods proposed are evaluated through simulations and the localization performance is validated through informal listening tests.**

***Index Terms*—Binaural beam-forming, convex optimization, interaural level difference (ILD), noise reduction.**

I. Introduction

Multi-microphone speech processing has become an integral part of hearing aids. By using multiple microphones, noise reduction algorithms are able to perform spatial filtering in addition to spectro-temporal filtering [1]. For the hearing impaired users, it is important to have good speech intelligibility, while also being able to localize sound, for example, to avoid accidents in traffic. Moreover, the ability to localize sound offers additional improvement in speech intelligibility due to spatial release from masking (SRM) [2]. Thus, over the years, many beam-forming algorithms have been proposed, that try to minimize the noise, in combination with acoustic spatial scene preservation of the target and the interferers.

Binaural beam-formers jointly process the microphone measurements from both ears and they can be classified as spatial filtering and spatio-temporal filtering algorithms. The binaural multi-channel Wiener filter (MWF) is a spatio-temporal beam-former that generates the minimum mean square error (MMSE) estimate of the speech component in the reference microphone signal after processing [3]. With the binaural MWF, the interaural level difference (ILD) and the interaural time difference (ITD) cues of the target are preserved, however, the spatial cues of the noise components are distorted. To preserve the spatial cues of the noise components, additional ITF or ITD or/and ILD terms are added to the cost function (see e.g., [4], [5], [6]). Unlike spatio-temporal filters, spatial filtering algorithms keep the target signal undistorted after processing. The binaural minimum variance distortionless response (BMVDR) beam-former is a binaural extension of the minimum variance distortionless response (MVDR) beam-former, that achieves optimal noise reduction in the presence of background noise and interfering sources. While the spatial cues of the target are preserved, the interferers however, appear to be co-located with the target [7]. By introducing additional constraints on the interferers, the binaural linearly constrained minimum variance (BLCMV) beam-former can preserve the spatial cues of a limited number of interferers [8]. The BLCMV formulation was further simplified with the joint binaural linearly constrained minimum variance (JBLCMV) beam-former, which allows for the spatial cue preservation of more interferers [9]. In [10], by replacing the equality constraints of the JBLCMV formulation with inequality constraints, a relaxation to the ITF cue preservation has been introduced. This increases the degrees of freedom (DoF) of the filter, thereby, preserving the spatial cues of a greater number of interferers. Moreover, for a fixed number of interferers, the noise reduction performance also improves, due to the increased DoF. By choosing a suitable allowable error in the ITF cues, a good trade-off between noise reduction and spatial cue preservation of the interferers can be achieved.

These methods preserve both the ITDs and the ILDs over the entire frequency spectrum. A human listener, however, does not rely at all frequencies on both the ITDs and the ILDs for the localization of sound. More specifically, the ITDs are the dominant cues at the frequencies below 1.5 kHz and the ILDs are the dominant cues at the higher frequencies [11]. Based on these facts, we propose two methods that try to preserve only the ILDs of the noise components, while keeping the target signal undistorted. These proposed methods are applied to preserve the ILDs of the noise components at higher frequencies and we use the JBLCMV method to preserve both the ITDs and ILDs of the noise components at lower frequencies. We investigate whether doing so saves DoF of the filter design, that can be used to improve the noise reduction performance, in contrast to preserving both the ITDs and the ILDs over the entire frequency spectrum as done with the JBLCMV method.

II. Signal Model

Consider a binaural hearing aid configuration, having one hearing aid (HA) on the left ear and one on the right ear, each with a microphone array containing $\frac{M}{2}$ microphones. The signals from the left HA are assumed to be transmitted wirelessly to the right HA, and vice versa, leading to a total of M signal measurements. In this work we neglect the fact that transmission comes with additional quantization noise. For optimal rate-constrained binaural beam-forming, we refer the reader to [12]. It is assumed that there is one target signal with 'r' additive, mutually uncorrelated interfering signals, and uncorrelated noise. All the measured noisy signals can be combined into a single vector $\mathbf{y} \in \mathbb{C}^{M\times 1}$ and can be represented in short-time Fourier transform (STFT) domain as

$$\mathbf{y}(l,k) = \mathbf{a}(l,k)s(l,k) + \sum_{i=1}^{r} \mathbf{b}_i(l,k)u_i(l,k) + \mathbf{v}(l,k), \quad (1)$$

where l, k are the time-frequency indices, $\mathbf{a} \in \mathbb{C}^{M\times 1}$ and $\mathbf{b}_i \in \mathbb{C}^{M\times 1}$ are the acoustic transfer function (ATF) vectors of the target and i^{th} interferer, s and u_i are the target and i^{th} interferer at its location, respectively. The vector $\mathbf{v} \in \mathbb{C}^{M\times 1}$ is the noise signal vector. The microphones $m = 1$ and $m = M$ are taken as the reference microphones for the left and the right HAs, respectively. Since the processing is done per time-frequency bin, the indices (l, k) are omitted in the rest of the work for convenience.

As the target, interferers and the noise are mutually uncorrelated, the cross power spectral density (CPSD) of the measured signal can be written as

$$\mathbf{P}_y = E\left[\mathbf{y}\mathbf{y}^H\right] = \mathbf{P}_x + \underbrace{\sum_{i=1}^{r} \mathbf{P}_{u_i} + \mathbf{P}_v}_{\mathbf{P}}, \quad (2)$$

where $\mathbf{P}_y$, $\mathbf{P}_x$, $\mathbf{P}_{u_i}$, $\mathbf{P}_v$, $\mathbf{P} \in \left(\mathbb{C}^{M\times M}\right)$ are the CPSD matrices of the noisy signal, the target signal, the i^{th} interferer, noise signal and the total noise signal, respectively.

The spatial filtering algorithms that will be discussed, estimate two complex spatial filters $\mathbf{w}_{\mathrm{L}}$ and $\mathbf{w}_{\mathrm{R}} \in \mathbb{C}^{M\times 1}$ of the combined binaural filter $\mathbf{w} = \left[\begin{array}{cc} \mathbf{w}_{\mathrm{L}}^H & \mathbf{w}_{\mathrm{R}}^H \end{array}\right]^H \in \mathbb{C}^{2M\times 1}$ such that, the output at the left and the right hearing aid can be given as

$$x_1 = \mathbf{w}_{\mathrm{L}}^H \mathbf{y}, \quad x_{\mathrm{M}} = \mathbf{w}_{\mathrm{R}}^H \mathbf{y}, \quad (3)$$

where x_1 and x_{M} are the target signals at the left and the right reference microphones, respectively.

The methods discussed in the following section are spatial filtering algorithms, i.e., the target is maintained undistorted by using the linear constraint

$$\mathbf{w}^H \underbrace{\left[\begin{array}{cc} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{array}\right]}_{\mathbf{\Lambda}_A \in \mathbb{C}^{2M\times 2}} = \underbrace{\left[\begin{array}{cc} a_1 & a_M \end{array}\right]}_{\mathbf{f}_A^H \in \mathbb{C}^{1\times 2}}. \quad (4)$$

The objective of the beam-formers is to reduce the total output noise power, given by

$$\underset{\mathbf{w}}{\text{minimise}} \quad \mathbf{w}^H \underbrace{\left[\begin{array}{cc} \mathbf{P} & \mathbf{0}_{M\times M} \\ \mathbf{0}_{M\times M} & \mathbf{P} \end{array}\right]}_{\tilde{\mathbf{P}} \in \mathbb{C}^{2M\times 2M}} \mathbf{w}. \quad (5)$$

Additionally, the spatial cues of the interferers are maintained by preserving the input binaural cues of the interferers at the output. This can strictly be achieved by preserving both the ILDs and the interaural phase differences (IPDs) (which are equivalent to ITDs in the frequency domain), or the ITFs of the interferers. The ILD and the IPD can be defined as the magnitude square of the ITF and the phase of the ITF, respectively. The input and the output ITF of the i^{th} interferer can be given as

$$\mathrm{ITF}_{u_i}{}^{in} = \frac{b_{i,1}}{b_{i,M}}, \qquad \mathrm{ITF}_{u_i}{}^{out} = \frac{\mathbf{w}_{\mathrm{L}}^H \mathbf{b_i}}{\mathbf{w}_{\mathrm{R}}^H \mathbf{b_i}}. \quad (6)$$

The ILD and IPD is given as

$$\mathrm{ILD} = |\mathrm{ITF}|^2, \quad \mathrm{IPD} = \frac{\angle \mathrm{ITF}}{\pi}. \quad (7)$$

III. Previous Work

A. BMVDR beam-former

The BMVDR beam-former is the optimal beam-former in terms of its noise reduction performance. The problem formulation can be written as done in [7]

$$\min_{\mathbf{w}} \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \quad \text{s. t. } \mathbf{w}^H \mathbf{\Lambda}_A = \mathbf{f}_A^H, \quad (8)$$

and the closed form solution can be given as

$$\mathbf{w}_{\mathrm{BMVDR}} = \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda}_A \left(\mathbf{\Lambda}_A^H \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda}_A\right)^{-1} \mathbf{f_A}. \quad (9)$$

With this beam-former, the spatial cues of the target signal are preserved perfectly, however, the cues of the interferers are co-located with the target signal [7].

B. JBLCMV beam-former

To preserve the spatial cues of the interferers, additional constraints that correspond to $\mathrm{ITF}_{u_i}{}^{in} = \mathrm{ITF}_{u_i}{}^{out}$ are introduced to the BMVDR problem formulation. This is called the JBLCMV beam-former and the problem formulation can be given as [9]

$$\begin{aligned} &\min_{\mathbf{w}} \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \\ &\text{s. t. } \mathbf{w}^H \underbrace{\left[\begin{array}{cc} \mathbf{\Lambda}_A & \mathbf{\Lambda}_C \end{array}\right]}_{\mathbf{\Lambda}} = \underbrace{\left[\begin{array}{cc} \mathbf{f}_A^H & \mathbf{f}_C^H \end{array}\right]}_{\mathbf{f}^H}, \end{aligned} \quad (10)$$

where

$$\begin{aligned} \mathbf{\Lambda}_C &= \left[\begin{array}{ccc} \mathbf{b}_1 b_{1,M} & \dots & \mathbf{b}_r b_{r,M} \\ -\mathbf{b}_1 b_{1,1} & \dots & -\mathbf{b}_r b_{r,1} \end{array}\right] \in \mathbb{C}^{2M\times r}, \\ \mathbf{f}_C^H &= \left[\begin{array}{ccc} 0 & \dots & 0 \end{array}\right] \in \mathbb{C}^{1\times r}. \end{aligned} \quad (11)$$

The closed form solution is then

$$\mathbf{w}_{\mathrm{JBLCMV}} = \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \left(\mathbf{\Lambda}^H \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda}\right)^{-1} \mathbf{f}, \text{ for } \quad r \leq r_{max}. \quad (12)$$

The spatial cues of the target and up to $r_{max} = 2M - 3$ interferers are perfectly preserved here.

IV. Methods Proposed

The methods discussed previously, preserve both ILDs and ITDs over the entire frequency spectrum. As mentioned before, human listeners are less sensitive to ITDs at the frequency range above 1.5 kHz [11]. Thus, we propose two methods that focus on preserving only ILDs of the interferers at frequencies above 1.5 kHz. The first method aims to perfectly preserve the ILDs, while the second method bounds the error in the ILDs of the interferers, as done with ITF in [10].

A. Perfect interaural level difference preservation (P-ILD)

With this method, the ILDs of the interferers are perfectly preserved by introducing the following additional constraints to the BMVDR optimization problem in (8)

$$\underbrace{\left|\frac{\mathbf{w}_{\mathrm{L}}^H\mathbf{b_i}}{\mathbf{w}_{\mathrm{R}}^H\mathbf{b_i}}\right|^2}_{\mathrm{ILD}_{u_i}{}^{out}}-\underbrace{\left|\frac{b_{i,1}}{b_{i,M}}\right|^2}_{\mathrm{ILD}_{u_i}{}^{in}}=0 \quad \text{for} \quad i=1,\ldots,r. \tag{13}$$

Expanding the ILD constraints in (13), the problem can be formulated as

$$\begin{aligned}
&(\mathbf{P}_1) \quad \min_{\mathbf{w}} \quad \mathbf{w}^H\tilde{\mathbf{P}}\mathbf{w}\\
&\text{s. t.} \quad \mathbf{w}^H\mathbf{\Lambda}_A=\mathbf{f}_A^H\\
&\mathbf{w}^H\underbrace{\begin{bmatrix}\mathbf{b_i}\mathbf{b_i}^H|b_{i,M}|^2 & \mathbf{0}_{M\times M}\\ \mathbf{0}_{M\times M} & -\mathbf{b_i}\mathbf{b_i}^H|b_{i,1}|^2\end{bmatrix}}_{\mathbf{M_i}\in\mathbb{C}^{2M\times 2M}}\mathbf{w}=0,\\
&\text{for} \quad i=1,\ldots,r.
\end{aligned} \tag{14}$$

Due to the quadratic ILD equality constraints on the interferers, (14) is a non-convex QCQP. Non-convex QCQP problems are NP-hard and are commonly overcome by implementing efficient approximation techniques using semi-definite relaxations [13]. The quadratic constraints can be linearized by using $\mathbf{W}=\mathbf{w}\mathbf{w}^H$, i.e., $\mathbf{W}\in\mathbb{C}^{2M\times 2M}$. Additionally, by relaxing it further as $\mathbf{W}\succeq\mathbf{w}\mathbf{w}^H$, the problem can be written as a semi-definite program (SDP).

$$\begin{aligned}
(\mathbf{SDR}_1) \quad \min_{\mathbf{W},\mathbf{w}} \quad & \mathrm{Tr}(\mathbf{W}\tilde{\mathbf{P}})\\
\text{s. t.} \quad & \mathbf{w}^H\mathbf{\Lambda}_A=\mathbf{f}_A^H\\
& \mathrm{Tr}\left(\mathbf{W}\mathbf{M}_i\right)=0 \quad \text{for } i=1,\ldots,r\\
& \begin{bmatrix}\mathbf{W} & \mathbf{w}\\ \mathbf{w}^H & 1\end{bmatrix}\succeq 0.
\end{aligned} \tag{15}$$

Since (15) is a relaxation, its optimal solution $\mathbf{p}^*_{\mathrm{SDR}_1}$ will be equivalent to the optimal solution $\mathbf{p}^*_1$ of the original problem in (14), only when $\mathbf{W}=\mathbf{w}\mathbf{w}^H$. Otherwise, $\mathbf{p}^*_{\mathrm{SDR}_1}$ provides a lower bound to the solution $\mathbf{p}^*_1$, i.e., $\mathbf{p}^*_{\mathrm{SDR}_1}\leq\mathbf{p}^*_1$ [14]. This bound can be tightened by introducing additional redundant constraints by the reformulation-linearization technique (RLT) proposed in [15]. These redundant constraints can be obtained by pre-multiplying the target distortionless constraints with $\mathbf{w}$ and squaring the distortionless constraint as

$$\mathbf{w}\mathbf{w}^H\mathbf{\Lambda}_A=\mathbf{w}\mathbf{f}_A^H \tag{16}$$

and

$$\left(\mathbf{w}^H\mathbf{\Lambda}_A-\mathbf{f}_A^H\right)\left(\mathbf{w}^H\mathbf{\Lambda}_A-\mathbf{f}_A^H\right)^H=0, \tag{17}$$

respectively. On linearizing them both with $\mathbf{W}$, we get

$$\mathbf{W}\mathbf{\Lambda}_A-\mathbf{w}\mathbf{f}_A^H=0 \tag{18}$$

and

$$\begin{aligned}
&\mathrm{Tr}\left(\mathbf{W}\mathbf{\Lambda}_A\mathbf{\Lambda}_A^H\right)-\mathbf{w}^H\mathbf{\Lambda}_A\mathbf{f}_A\\
&-\left(\mathbf{\Lambda}_A\mathbf{f}_A\right)^H\mathbf{w}+\mathbf{f}_A^H\mathbf{f}_A=0,
\end{aligned} \tag{19}$$

respectively.

If the constraints in (18) and (19) are added to (15), the $\mathbf{SDR}_1$ problem can be re-written as

$$\begin{aligned}
(\textbf{SDR-RLT}_1) \quad \min_{\mathbf{W},\mathbf{w}} \quad & \mathrm{Tr}(\mathbf{W}\tilde{\mathbf{P}})\\
\text{s. t.} \quad & \mathbf{w}^H\mathbf{\Lambda}_A=\mathbf{f}_A^H\\
& \mathrm{Tr}\left(\mathbf{W}\mathbf{M}_i\right)=0 \quad \text{for} \quad i=1,\ldots,r\\
& \mathbf{W}\mathbf{\Lambda}_A-\mathbf{w}\mathbf{f}_A^H=0\\
& \mathrm{Tr}\left(\mathbf{W}\mathbf{\Lambda}_A\mathbf{\Lambda}_A^H\right)-\mathbf{w}^H\mathbf{\Lambda}_A\mathbf{f}_A\\
& -\left(\mathbf{\Lambda}_A\mathbf{f}_A\right)^H\mathbf{w}+\mathbf{f}_A^H\mathbf{f}_A=0\\
& \begin{bmatrix}\mathbf{W} & \mathbf{w}\\ \mathbf{w}^H & 1\end{bmatrix}\succeq 0.
\end{aligned} \tag{20}$$

The solution $\mathbf{p}^*_{\text{SDR-RLT}_1}$ of the $\textbf{SDR-RLT}_1$ problem, satisfies $\mathbf{p}^*_{\mathrm{SDR}_1}\leq\mathbf{p}^*_{\text{SDR-RLT}_1}\leq\mathbf{p}^*_1$. Thus, the above method provides a lower bound to the optimal solution of the problem $\mathbf{P}_1$, that achieves the same or better noise reduction performance than the problem $\mathbf{P}_1$. The maximum number of interferers r_{max}, whose cues can be preserved in problem $\mathbf{P}_1$ is $2M-3$. With the relaxation, however, the feasibility region widens and the problem $\textbf{SDR-RLT}_1$ can be solved for a larger r, which however, is difficult to determine.

B. Relaxed interaural level difference preservation (R-ILD)

With this method, the ILD cues of the interferers are bound within an upper limit $\mathcal{E}_i$, by introducing the following additional inequality constraints to the BMVDR optimization problem in (8)

$$\left|\left|\frac{\mathbf{w}_{\mathrm{L}}^H\mathbf{b_i}}{\mathbf{w}_{\mathrm{R}}^H\mathbf{b_i}}\right|^2-\left|\frac{b_{i,1}}{b_{i,M}}\right|^2\right|\leq\mathcal{E}_i, \quad \text{for} \quad i=1,\ldots,r, \tag{21}$$

$$\text{where} \quad \mathcal{E}_i=c_i\underbrace{\left|\left|\frac{a_1}{a_M}\right|^2-\left|\frac{b_{i,1}}{b_{i,M}}\right|^2\right|}_{\epsilon_{u_i}^{\mathrm{BMVDR}}}, \tag{22}$$

where, $\epsilon_{u_i}^{\mathrm{BMVDR}}$ is the ILD error of the i^{th} interferer, obtained with the BMVDR beam-former. By choosing a suitable $c_i\in[0,1]$, a good trade-off between the noise reduction and the ILD cue preservation can be achieved. As done with the P-ILD method, on expanding the ILD constraints and re-writing it as a joint optimization problem, the problem becomes a non-convex QCQP. Thus by following the approach used with $\mathbf{SDR}_1$, a semi-definite relaxation (SDR) that provides a lower bound to the problem with the constraints in (22) can be obtained as

$$
\begin{aligned}
(\textbf{SDR-RLT}_2) \quad & \min_{\mathbf{W},\mathbf{w}} \operatorname{Tr}(\mathbf{W}\tilde{\mathbf{P}}) \\
& \text{s. t. } \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\
& \operatorname{Tr}\left(\mathbf{W}\mathbf{M}_{A,i}\right) \leq 0 \quad \text{for} \quad i = 1, \ldots, r \\
& \operatorname{Tr}\left(\mathbf{W}\mathbf{M}_{B,i}\right) \leq 0 \quad \text{for} \quad i = 1, \ldots, r \\
& \mathbf{W}\boldsymbol{\Lambda}_A - \mathbf{w}\mathbf{f}_A^H = 0 \\
& \operatorname{Tr}\left(\mathbf{W}\boldsymbol{\Lambda}_A\boldsymbol{\Lambda}_A^H\right) - \mathbf{w}^H\boldsymbol{\Lambda}_A\mathbf{f}_A \\
& - \left(\boldsymbol{\Lambda}_A\mathbf{f}_A\right)^H \mathbf{w} + \mathbf{f}_A^H\mathbf{f}_A = 0 \\
& \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^H & 1 \end{bmatrix} \succeq 0.
\end{aligned} \tag{23}
$$

The optimization problems proposed are solved in polynomial time, using the CVX toolbox in Matlab [16].

V. Results

In this section, the simulations conducted with the proposed methods and the reference methods (BMVDR, JBLCMV) are compared with respect to the ILD cue preservation, noise reduction performance and predicted instrumental speech intelligibility. The ILD cue errors are measured and averaged over the frequencies $f \geq 1.5$ kHz, the range where the ITDs are less important, perceptually. The noise reduction performance is compared by evaluating the gain in the binaural global segmental signal-to-noise ratio (gsSNR). The ILD errors and gsSNR are evaluated as done in [17] and are omitted here due to space constraints. To measure the speech intelligibility performance, the speech intelligibility in bits (SIIB) metric is used, which was shown to be reliable when the speech is degraded by modulated point noise sources and reverberation [18], [19].

A. Performance

For the simulations, we use one target speech source $\mathbf{s}$, placed at $0°$ and 7 interfering speech sources $\mathbf{u}_1, \ldots, \mathbf{u}_7$ placed at $\{90°, -90°, 15°, -15°, 45°, -45°, 70°\}$, respectively. The interferers are considered incrementally, i.e., for 1 interfering source, $\mathbf{u}_1$ is considered, for 2 interfering sources $\mathbf{u}_1$ and $\mathbf{u}_2$ are considered and so on. The speech signals are taken from the TIMIT database [20]. For each HA, $\frac{M}{2} = 2$ microphones are considered and the point sources that reach the microphones of the HAs are spatialised using the head related impulse responses (HRIRs) from the multi-channel, behind-the-ear (BTE) database [21]. From this database, we consider the anechoic and the office environment HRIRs. The speech signals are 30 seconds long in duration, and each interferer is taken to be at a signal-to-noise ratio (SNR) = 0 dB with respect to the target source. Additionally, white gaussian noise (WGN) at SNR = 50 dB with respect to the target source is used to simulate the microphone self noise. To avoid additional sources of error due to ATF estimation, we use the true ATFs to evaluate the underlying statistics of the methods. The noise CPSD matrix $\mathbf{P}$ is computed using (2) in the noise only period, estimated by a voice activity detector (VAD). As the performance of the two proposed methods are similar in both the anechoic and the reverberant environment, only the results of the anechoic environment are presented here.

For the R-ILD method, the choice of the parameter c_i is taken to be $c_i = c = 0.2$, which is determined by comparing the localization performance against the noise reduction performance. Fig. 1 shows the results of the simulation in the anechoic environment. The $\text{gsSNR}^{\text{gain}}$ and $\text{SIIB}^{\text{gain}}$ are gain in the respective perfromance metrics, measured by the methods w.r.t. the unfiltered input signal. For interferers $r \leq 3$, the performance of the proposed methods emulate the performance of the JBLCMV method, in terms of ILD cue preservation, noise reduction and speech intelligibility. The P-ILD method preserves the ILDs perfectly, and the R-ILD method preserves the ILDs within the upper bound shown by Avg. $\mathcal{E}_i$, for $r \leq 3$. On increasing the number of interferers with $r > 4$, violations in the ILD cue constraints are observed with the proposed methods. Interestingly, with $r > 2M - 3 = 5$ ($r_{max} = 2M - 3$ for JBLCMV), it can be observed that the ILD errors with the proposed methods, are much lower than with the JBLCMV method. Moreover, with both the proposed methods, a $\text{gsSNR}^{\text{gain}}$ within 0.5 dB over the JBLCMV can be observed when a larger number of interferers is considered. Similarly, for a larger number of interferers present, the proposed methods show an improvement in speech intelligibility over the JBLCMV method.

B. Listening Test

To validate the localization performance of the two proposed methods, online informal listening tests were conducted with 18 subjects, with self reported normal hearing in the age group of 23-27 years. For these tests, one female target speaker and three interfering sources — male speaker, music signal and HF signal (2 kHz cut-off) were considered. The signals were 4 seconds long. The microphone self noise was simulated by additive WGN, such that the target is at a SNR of 50 dB with respect to the WGN. The overall SNR defined by the ratio of the target signal power to the total noise power is fixed to be 5 dB. The subjects were presented with signals filtered with the BMVDR, JBLCMV, P-ILD, R-ILD($c = 0.1$) and R-ILD($c = 0.3$) algorithms. Two values of c were chosen for the R-ILD method, to compare the localization performance between a smaller ($c = 0.1$) and a larger ($c = 0.3$) error in ILDs. The tests were conducted for both the anechoic and the office environment. The signals were presented in a random order to each subject, with two repetitions. Additionally, they were presented with the unprocessed signal, and the localization errors of the subjects are found relative to this reference position and are averaged over their repetitions. From Fig. 2, it can be observed that the localization performance of the proposed methods is similar to the JBLCMV method, where both ITDs and ILDs are preserved. As expected, female target signal is nearly preserved, due to the distortion-less target constraints in all the methods considered. The localization performance of HF signal with the proposed method, which

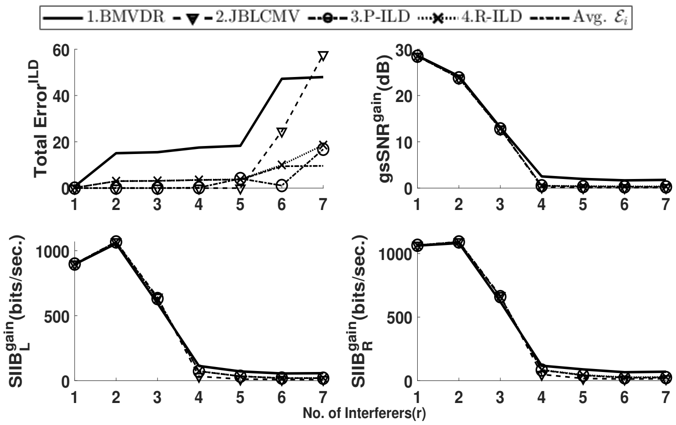

Fig. 1. Anechoic : ILD error, gsSNR$^{\text{gain}}$ and SIIB$^{\text{gain}}$ vs. No. of Interferers.

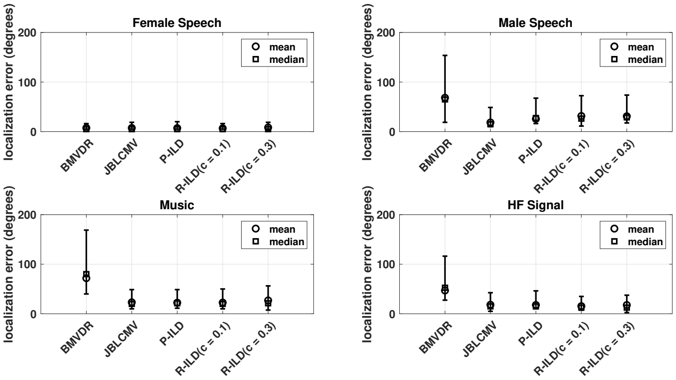

Fig. 2. Anechoic : localization error for each source across the proposed and reference methods.

only preserves the ILDs in the higher frequencies, is similar to the localization with JBLCMV. Furthermore, the localization performance of R-ILD (c = 0.1) was found to be better than R-ILD(c = 0.3). The results were also further verified with the help of T-test and analysis of variance (ANOVA) tests, the results of which can be found in [22].

VI. Conclusion

We proposed two binaural beam-formers, that preserve only the ILDs of the interferers while keeping the target undistorted. In comparison to preserving the ITF at all frequencies, the proposed methods showed a small improvement in noise reduction and speech intelligibility performance when a larger number of interferers is present. Moreover, the errors in ILDs are lower than with the JBLCMV for $r > 2M - 3$. Lastly, the informal listening tests helped to validate the ILD cue preservation performance of the methods proposed and also showed that the localization performance was good even with bounded errors in the ILD preservation as done with the R-ILD.

References

[1] T. Lotter, "Single and multimicrophone speech enhancement for hearing aids," Ph.D. dissertation, RWTH Aachen, Germany, Aug. 2004.

[2] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acust. United Acust.*, vol. 86, pp. 117–128, 2000.

[3] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. R. Liu, Eds. New York, NY, USA: Wiley, 2010, ch. 9, pp. 269–302.

[4] T. Van den Bogaert, J. Wouters, S. Doclo, and M. Moonen, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel wiener filter," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Honolulu, HI, USA, 2007, pp. 565–568.

[5] S. Doclo, Rong Dong, T. J. Klasen, J. Wouters, S. Haykin, and M. Moonen, "Extension of the multi-channel wiener filter with ITD cues for noise reduction in binaural hearing aids," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 2005, pp. 70–73.

[6] S. Doclo, R. Dong, T. J. Klasen, J. Wouters, S. Haykin, and M. Moonen, "Extension of the multi-channel wiener filter with localisation cues for noise reduction in hearing aids," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, vol. 4, Eindhoven, The Netherlands, 2005, pp. 221–224.

[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*, vol. 23, no. 12, pp. 2449–2464, 2015.

[8] E. Hadad, S. Gannot, and S. Doclo, "Binaural linearly constrained minimum variance beamformer for hearing aid applications," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, 2012, pp. 1–4.

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

[10] A. Koutrouvelis, "Multi-microphone noise reduction for hearing assistive devices," Ph.D. dissertation, Delft University of Technology, 2018.

[11] J. W. Strutt and L. Rayleigh, "On our perception of sound direction," *Philosophical Magazine*, vol. 13, pp. 214–232, 1907.

[12] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Spatially correct rate-constrained noise reduction for binaural hearing aids in wireless acoustic sensor networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2731–2742, 2020.

[13] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[14] J. Park and S. Boyd, "General Heuristics for Nonconvex Quadratically Constrained Quadratic Programming," Mar. 2017. [Online]. Available: arXiv:1703.07870

[15] H. D. Sherali and W. P. Adams, *A reformulation-linearization technique for solving discrete and continuous nonconvex problems*. Kluwer Academic Publishers, 1999.

[16] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[17] 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, 2016.

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

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

[20] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, "DARPA TIMIT acoustic-phonetic continuous speech corpus," *National Institute of Standards and Technology, Gaithersburg, MD*, 1993.

[21] 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, no. 1, p. 298605, 2009.

[22] V. Sathyapriyan, "Binaural beam-forming with dominant cue preservation for hearing aids," Master's thesis, Delft University of Technology, 2020. [Online]. Available: http://resolver.tudelft.nl/uuid:89fc770b-0ae9-425f-83ff-d0aaa4e0822d