Generation of Personal Sound Zones With Physical Meaningful Constraints and Conjugate Gradient Method

Shi, Liming; Lee, Taewoong; Zhang, Lijun; Nielsen, Jesper Kjær; Christensen, Mads Græsbøll

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

DOI (link to publication from Publisher):
10.1109/TASLP.2021.3052564

Publication date:
2021

Document Version
Early version, also known as pre-print

Subspace-Based Methods for the Generation of Personal Sound Zones with Physically Meaningful Constraints

Liming Shi, *Student Member, IEEE,* Taewoong Lee, *Student Member, IEEE,* Lijun Zhang, Jesper Kjær Nielsen, *Member, IEEE,* and Mads Græsbøll Christensen, *Senior Member, IEEE*

***Abstract*—Personal sound zones provide users to experience independent listening and quiet areas in the same acoustic environment using multiple loudspeakers. Generally, this can be done either by maximizing an acoustic contrast (AC) that represents the acoustic potential energy ratio between the bright and dark zones or by minimizing a reproduction error or a signal distortion (SD) between the desired and reproduced sound fields. However, the former suffers from severe distortion in the reproduced sound field, whereas the latter suffers from poor acoustic contrast. Recently, a flexible and general framework for sound zone control referred to as the variable span trade-off (VAST) filter has been proposed. The VAST framework, which is a generalized eigenvalue decomposition (GEVD)-based method, allows the user to control the trade-off between AC and SD by adjusting two user parameters: the subspace rank and the Lagrange multiplier. Unfortunately, these parameters are not physically meaningful, and the user has to tune them for different source materials and acoustic environments. In this paper, we propose various strategies to control the reproduced sound field as precisely and accurately as possible by reformulating the problem using physically meaningful constraints, including SD and AC. Furthermore, a modified version of the conjugate gradient-based method is used to reduce computational complexity, to reduce SD when the subspace rank is small, and to persist the orthogonality of the basis functions as in VAST. The proposed method shows precise control over the reproduced sound field via extensive numerical validations in anechoic and reverberant environments for different physically meaningful constraints. For the reproducibility of the experimental results, a MATLAB implementation is available at https://tinyurl.com/rszern5.**

***Index Terms*—Personal sound zones, physically meaningful constraints, subspace-based approach, conjugate gradient method, variable span trade-off filters.**

I. Introduction

THE creation of personal sound zones became an active area of research ever since it was first introduced approximately two decades ago [1], [2]. Based on this concept, different applications such as cars and aircraft [3]–[9], outdoor concerts [10], mobile devices [11], etc. [12], [13] were studied. Typically, two different types of sound zones are considered: a bright zone (or a listening zone) and a dark zone (or a quiet or silent zone). The bright zone is where the acoustic potential energy is as high as possible, or the desired sound field is reproduced as well as possible. The dark zone is where the acoustic potential energy is as low as possible. Multiple loudspeakers are controlled by different control strategies to generate these zones for a single audio content. If the problem of generating sound zones is solved for each audio content individually, then multiple bright zones can be obtained by applying the superposition principle.

The control strategies for the generation of sound zones broadly fall into the following three categories: acoustic contrast control (ACC) [14], pressure matching (PM) [15], and mode matching [16]. The first two approaches can also be seen as optimization-based approaches that we focus on in this paper, and the mode matching can be considered to be an analytical approach. The mode matching (or the modal domain) approach describes the reproduced sound field via a spatial harmonic expansion to match such a field to the desired sound field.

ACC seeks the control filter that maximizes the acoustic contrast, which represents the acoustic potential energy ratio between the bright and dark zones [14]. This can be solved either by maximizing the energy in the bright zone with a constraint on the energy in the dark zone or by minimizing the energy in the dark zone with a constraint on the energy in the bright zone. Such approaches are termed as either direct or indirect acoustic contrast formulations, respectively [17]. Besides, a method that maximizes the energy difference between the bright and dark zones was also proposed [18], and it gives the same solution as the indirect method as described in [17]. Although this energy-based approach gives high acoustic contrast, the reproduced sound field often does not match the desired sound field since no constraint on the phase in the bright zone is considered [19]. This becomes more crucial, if the time domain ACC method, which is also referred to as the broadband ACC (BACC) [20], is considered since the method seeks for the solution which maximizes a few frequency components and filters out the rest. This results in a severe signal distortion but fulfills the maximization constraint. Several techniques were proposed, e.g., [21]–[23], to ease this distortion issue. Recently, to achieve different AC levels for different frequency bands, multiple constraints approaches also have been studied in [5], [9].

PM aims at reproducing desired sound fields for the bright and dark zones by minimizing the reproduction error, which is the difference between the reproduced and desired sound fields in both zones. This method gives a much more accurate

L. Shi, T. Lee, J. K. Nielsen and M. G. Christensen are with the Audio Analysis Lab, CREATE, Aalborg University, 9000 Aalborg, Denmark, e-mail: {ls, tlee, jkn, mgc}@create.aau.dk, L. Zhang is with the School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China, e-mail: zhanglj7385@nwpu.edu.cn

reproduction of the desired fields; however, it also gives a much worse contrast than that of ACC. A combination between ACC and PM was studied to obtain the best compromise of low distortion and high acoustic contrast such as [24] using [18] and [15] to provide a trade-off between acoustic contrast and signal distortion. Besides, in [25], [26], a modified version of PM was proposed to provide a trade-off between the signal distortion in the bright zone and the dark zone power. Often these methods are referred to as ACC-PM or PM-ACC; however, neither can reach a correct combination between ACC and PM.

Recently, inspired by subspace-based signal enhancement approaches [27]–[31], a general and flexible framework for the sound zone control referred to as the variable span trade-off (VAST) filter has been proposed [32]–[34]. This framework can be obtained by solving a convex optimization problem with a constraint on the dark zone power, and this allows the user to control the trade-off between the acoustic contrast and the signal distortion by adjusting two user parameters: the subspace rank and a regularization parameter (or the Lagrange multiplier). Moreover, BACC [20], the broadband PM, and BACC-PM [26] can be obtained from VAST by tuning these two user parameters.

The VAST approach, however, has multiple drawbacks. First, the influence of the regularization parameter on the performance of the VAST filter has not been well-studied. Moreover, the two user parameters are not physically meaningful, and the user has to tune them for different source materials and acoustic environments. Second, the desired sound field is not taken into account when constructing the subspace or computing the GEVD, which causes high signal distortion when the subspace dimension is small. Third, the VAST approach has high computational complexity due to the computation of a GEVD of a high dimensional matrix. Recently, a method that finds the optimal regularization parameter using Newton's method (similar to [35]) was proposed [36] with a performance degradation of $4-5$ dB in terms of acoustic contrast. This method exactly fulfills the constraint on the dark zone power and reduces the computational complexity of VAST using the conjugate gradient (CG) method. It should be noted that although such a method demands less computational complexity than that of VAST, this comes at the cost of the loss of controlling the orthogonality of the subspace when the subspace rank is large.

In this paper, we first investigate the influence of the regularization parameter on the performance of the VAST filter. Moreover, we present a variety of strategies to compute the optimal values of the regularization parameter, which fulfills different physically meaningful constraints, including the signal distortion and the acoustic contrast. These strategies will be more intuitive and direct than tuning the dark zone power, although such tuning is implicitly related to the acoustic contrast, and finally, the signal distortion. Furthermore, we propose a sound zone control method as an extended work of our previous work in [36] not only to take the desired sound field into account in the subspace but also to deal with the orthogonality issue.

The paper is organized as follows: in Section II, the problem of the generation of sound zones is formulated via VAST, which is a GEVD-based method that provides the mathematical foundation of this paper. Followed by this, various strategies to compute the optimal value of the regularization parameter based on physically meaningful constraints are presented in Section III. Furthermore, a sound zone control method based on the conjugate gradient method with the trade-off property as in VAST is proposed in Section IV. The proposed method is evaluated and compared to the VAST approach for different physically meaningful constraints in Section V. Finally, the paper is concluded in Section VI.

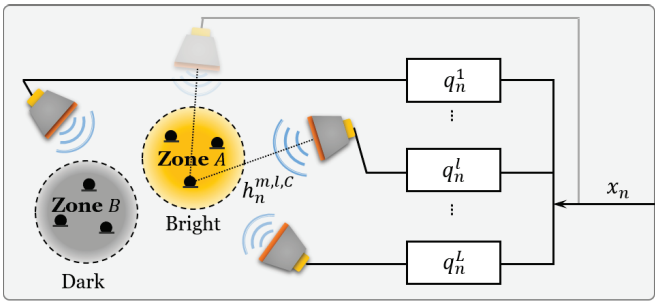

Fig. 1. An example of a system setup for sound zones. The input sound signal x_n is fed into L loudspeakers after filtering by the control filters $\{q_n^l\}_{l=1}^L$. The RIR from loudspeaker l to control point m in zone C is represented as $h_n^{m,l,C}$.

II. Generation of Sound Zones

In this section, we formulate the sound field generation problem in the time-domain and review the GEVD-based approaches for the generation of sound zones that form the fundamental structure of the proposed sound zone control algorithms.

A. Problem formulation

We consider the problem of generating a bright zone B and a dark zone D in an enclosed space with measured/known room impulse responses (RIRs) using L loudspeakers and L finite impulse response (FIR) filters (a.k.a., control filters) with the same length J, as shown in Fig. 1. The reproduced sound pressure at the $m^{\text{th}}, 1 \leq m \leq M^{\text{C}}$ control point in one of the sound zones can be written as

$$y_n^{m,\text{C}} = \sum_{l=1}^{L} x_n * q_n^l * h_n^{m,l,\text{C}} \ , \tag{1}$$

where the superscript $(\cdot)^{\text{C}}, \text{C} \in \{\text{B}, \text{D}\}$ is the zone index, x_n denotes the input sound signal, q_n^l denotes the control filter for the l^{th} loudspeaker, $h_n^{m,l,\text{C}}$ denotes the RIR of length K from the l^{th} loudspeaker to the m^{th} control point in zone C, $*$ denotes the linear convolution operator, and n denotes the sampling index. For simplicity, we have left out superscript C when not absolutely necessary. We assume that both the control filters and RIRs are linear and time-invariant and write (1) in matrix form, then we have

$$y_n^m = \sum_{l=1}^{L} \mathbf{q}_l^T \mathbf{H}^{m,l} \mathbf{x}_n = \mathbf{q}^T \mathbf{H}^m \mathbf{x}_n \ , \tag{2}$$

where

$$\mathbf{x}_n = [x_n, x_{n-1}, \cdots, x_{n-(K+J-2)}]^T \ ,$$
$$\mathbf{q}_l = [q_1^l, \cdots, q_J^l] \ ,$$

$\mathbf{H}^{m,l}$ is a $J \times (K+J-1)$ Toeplitz matrix with the first row vector and the first column vector are defined as $[h_1^{m,l}, \cdots, h_K^{m,l}, 0, \cdots, 0]$ and $[h_1^{m,l}, 0, \cdots, 0]^T$, respectively, and

$$\mathbf{H}^m = [(\mathbf{H}^{m,1})^T, \cdots, (\mathbf{H}^{m,L})^T]^T \ ,$$
$$\mathbf{q} = [\mathbf{q}_l^T, \cdots, \mathbf{q}_L^T]^T \ .$$

Define the desired sound signal, generated from a virtual source, as $d_n^m = x_n * g_n^m = (\mathbf{g}^m)^T \mathbf{x}_n$, where $\mathbf{g}^m = [g_1^m, \cdots, g_K^m, 0, \cdots, 0]^T$ denotes the RIR from the virtual source to the m^{th} control point. Note that, in reverberated scenarios, the RIRs $\mathbf{g}^m$ can be truncated to contain only the direct path or early reflections. Then, all the algorithms for the generation of sound zones also perform dereverberation. The mean-squared-error (MSE) between the desired and reproduced signal over N time samples and M control points can be expressed as

$$\begin{aligned} J(\mathbf{q}) =& \frac{1}{NM} \sum_{m=1}^{M} \sum_{n=1}^{N} (d_n^m - y_n^m)^2 \\ =& \mathbf{q}^T \mathbf{R} \mathbf{q} - 2\mathbf{q}^T \mathbf{r} + \kappa, \end{aligned} \tag{3}$$

where $\mathbf{R} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{H}^m \mathbf{R}_x (\mathbf{H}^m)^T$ denotes the spatial autocorrelation matrix, $\mathbf{r} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{H}^m \mathbf{R}_x \mathbf{g}^m$ denotes the spatial cross-correlation vector, $\kappa = \frac{1}{M} \sum_{m=1}^{M} (\mathbf{g}^m)^T \mathbf{R}_x \mathbf{g}^m$ is a constant term, and $\mathbf{R}_x = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^T$ denotes the input signal autocorrelation matrix. Note that the desired sound signal for the dark zone is usually set to $d_n^{m,\text{D}} = 0$, leading to $\mathbf{r}^{\text{D}} = \mathbf{0}$ and $\kappa^{\text{D}} = 0$. We will use this setting throughout this paper. The remaining term $\mathbf{q}^T \mathbf{R}^{\text{D}} \mathbf{q}^T$ in (3) is commonly referred to as the acoustic power in the dark zone. The rank of $\mathbf{R}$ can be found as

$$\text{rank}(\mathbf{R}) \leq \min\{MN, M(K+J-1), LJ\} \ . \tag{4}$$

The objective of generating sound zones is to find $\mathbf{q}$ to minimize the MSEs (3) for both the bright and dark zones, simultaneously. It is also worth mentioning that the MSE cost function (3) is input data x_n-dependent. Most of the state-of-the-art approaches are data-independent. However, data-independent methods can also be obtained using (3) with a pre-defined $\mathbf{R}_x$. For example, by assuming the input signal is a white Gaussian noise with a unit variance, i.e., $\mathbf{R}_x = \mathbf{I}$, the MSE cost function is only dependent on the RIRs. Moreover, the perceptual weighting filters can also be integrated in $\mathbf{R}_x$ and $\mathbf{r}$ to take the characteristics of the human auditory system into account, as proposed in [34].

Only considering the bright zone, the MSE cost function (3) is minimized by setting its derivative w.r.t. $\mathbf{q}$ equal to zero, leading to the well-known Wiener-Hopf solution

$$\hat{\mathbf{q}}_{\text{opt}} = (\mathbf{R}^{\text{B}})^{-1} \mathbf{r}^{\text{B}} \ . \tag{5}$$

When the virtual source is placed at one of the loudspeakers (e.g., the l^{th} loudspeaker), setting the RIRs from the virtual source to all control points are identical to the RIRs from the l^{th} loudspeaker, the spatial cross-correlation vector becomes $\mathbf{r}^{\text{B}} = \mathbf{R}^{\text{B}} \mathbf{i}_{(J-1)l+1}$, where $\mathbf{i}_{(J-1)l+1}$ denotes an all zeros vector except that the $((J-1)l+1)^{\text{th}}$ element is one. In this case, the Wiener-Hopf solution is $\hat{\mathbf{q}}_{\text{opt}} = \mathbf{i}_{(J-1)l+1}$ and $J(\mathbf{q}) = 0$, that is playing the l^{th} loudspeaker and muting all the others. Only considering the dark zone, since $\mathbf{r}^{\text{D}} = \mathbf{0}$, the Wiener-Hopf solution is simply $\hat{\mathbf{q}}_{\text{opt}} = \mathbf{0}$, that is muting all the loudspeakers.

To minimize the MSEs in (3) for both the bright and dark zones, a combined cost function is typically used [37], i.e.,

$$\hat{\mathbf{q}} = \arg\min_{\mathbf{q}} J^{\text{B}}(\mathbf{q}) + \mu J^{\text{D}}(\mathbf{q}) \ , \tag{6}$$

where μ denotes the regularization/weighting parameter. For the generation of sound zones, subspace-based approaches have also been proposed. Next, we introduce these approaches from a subspace perspective and present their *pros* and *cons*.

B. GEVD-based approach

Let $\mathcal{K}_V$ be a V-dimensional subspace of $\mathcal{R}^{LJ}$ ($V \leq LJ$) with a basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_V\}$, and form a rank-$V$ matrix $\mathbf{U}_V = [\mathbf{u}_1, \cdots, \mathbf{u}_V]$. We assume that the control filter $\mathbf{q}$ is a vector in $\mathcal{K}_V$, i.e., $\mathbf{q} = \mathbf{U}_V \mathbf{z}$ and $\mathbf{z} \in \mathcal{R}^V$, and consider the optimization problem

$$\hat{\mathbf{z}} = \arg\min_{\mathbf{z}} J^{\text{B}}(\mathbf{U}_V \mathbf{z}) + \mu J^{\text{D}}(\mathbf{U}_{\mathbf{V}} \mathbf{z}) \ . \tag{7}$$

Plugging (3) into (7), the optimal control filter can be written as

$$\hat{\mathbf{q}} = \mathbf{U}_V \hat{\mathbf{z}} = \mathbf{U}_V (\mathbf{U}_V^T \mathbf{R}^{\text{B}} \mathbf{U}_V + \mu \mathbf{U}_V^T \mathbf{R}^{\text{D}} \mathbf{U}_V)^{-1} \mathbf{U}_V^T \mathbf{r}^{\text{B}} \ . \tag{8}$$

The subspace $\mathcal{K}_V$ can be chosen in multiple ways. One example is to set $V = LJ$, $\mu = 1$ and $\mathbf{U}_{LJ} = \mathbf{I}_{LJ}$, then (8) reduces to the time-domain PM method (originally proposed in the frequency-domain in [15]). Another example is to set $V = 1$ and the basis vector $\mathbf{u}_1$ to the eigenvector corresponding to the largest eigenvalue for the generalized eigenvalue problem $\mathbf{R}^{\text{B}} \mathbf{u}_i = \lambda_i \mathbf{R}^{\text{D}} \mathbf{u}_i, 1 \leq i \leq LJ$. In this case, the optimal filter (8) reduces to the time-domain ACC method [20] (originally proposed in the frequency-domain [14]). The motivation for setting the control filter $\mathbf{q}$ to the eigenvector corresponding to the largest eigenvalue is that the acoustic contrast ξ_{ac}, defined as the acoustic energy ratio between the bright zone and dark zone, i.e.,

$$\xi_{\text{ac}} = \frac{\mathbf{q}^T \mathbf{R}^{\text{B}} \mathbf{q}}{\mathbf{q}^T \mathbf{R}^{\text{D}} \mathbf{q}} \ , \tag{9}$$

is maximized for this choice of $\mathbf{q}$. Recently, motivated by a framework called the variable span linear filters which was proposed in speech enhancement [30], a variable span trade-off filter for the generation of sound zones has been proposed [32]. The essence of this approach is to form the V-dimensional subspace using V eigenvectors corresponding to the V largest eigenvalues after computing a GEVD. One of the important properties of the GEVD is that the rank-V matrix $\mathbf{U}_V$, formed by the eigenvectors, can jointly diagonalize both the matrices $\mathbf{R}^{\text{B}}$ and $\mathbf{R}^{\text{D}}$, i.e.,

$$\mathbf{U}_V^T \mathbf{R}^{\text{B}} \mathbf{U}_V = \mathbf{\Lambda}_V, \quad \mathbf{U}_V^T \mathbf{R}^{\text{D}} \mathbf{U}_V = \mathbf{I}_V, \tag{10}$$

where $\mathbf{\Lambda}_V = \mathrm{diag}(\lambda_1, \cdots, \lambda_V)$ is a diagonal matrix whose diagonal elements are eigenvalues sorted as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_V$. Substituting (10) into (8), the optimal control filter can be further written as

$$\begin{aligned}\hat{\mathbf{q}} =& \mathbf{U}_V(\mathbf{\Lambda}_V + \mu\mathbf{I}_V)^{-1}\mathbf{U}_V^T\mathbf{r}^{\mathrm{B}} \\ =& \sum_{i=1}^{V}\frac{\mathbf{u}_i^T\mathbf{r}^{\mathrm{B}}}{\lambda_i+\mu}\mathbf{u}_i,\end{aligned} \tag{11}$$

Obviously, when $V = 1$, (11) reduces to the time-domain ACC approach [20], and when $V = LJ$, (11) reduces to the time-domain ACC-PM approach [26]. Moreover, when $V = LJ$ and $\mu = 0$, (11) reduces to the Wiener-Hopf equation (5).

To explain the trade-off property of (11), we introduce another performance metric in addition to the acoustic contrast ξ_{ac} in (9), named the normalized signal distortion ξ_{sd}, defined as

$$\begin{aligned}\xi_{\mathrm{sd}} =& \frac{\frac{1}{NM^{\mathrm{B}}}\sum_{m=1}^{M^{\mathrm{B}}}\sum_{n=1}^{N}(d_n^{m,\mathrm{B}} - y_n^{m,\mathrm{B}})^2}{\frac{1}{NM^{\mathrm{B}}}\sum_{m=1}^{M^{\mathrm{B}}}\sum_{n=1}^{N}(d_n^{m,\mathrm{B}})^2} \\ =& \frac{\mathbf{q}^T\mathbf{R}^{\mathrm{B}}\mathbf{q} - 2\mathbf{q}^T\mathbf{r}^{\mathrm{B}} + \kappa^{\mathrm{B}}}{\kappa^{\mathrm{B}}}.\end{aligned} \tag{12}$$

Substituting the optimal control filter (11) into (12), we can obtain

$$\xi_{\mathrm{sd}}^{\mathrm{o}} = \frac{\kappa^{\mathrm{B}} - \sum_{i=1}^{V}\frac{\lambda_i+2\mu}{(\lambda_i+\mu)^2}c_i}{\kappa^{\mathrm{B}}}, \tag{13}$$

where $c_i = |\mathbf{u}_i^T\mathbf{r}^{\mathrm{B}}|^2$. Clearly, for the optimal control filter (11), the signal distortion metric ξ_{sd} is always within the range $[0, 1]$. A simpler form of the acoustic contrast metric can also be obtained by substituting (11) into (9), and we can obtain

$$\xi_{\mathrm{ac}}^{\mathrm{o}} = \frac{\sum_{i=1}^{V}\lambda_i c_i/(\lambda_i+\mu)^2}{\sum_{i=1}^{V}c_i/(\lambda_i+\mu)^2}. \tag{14}$$

As can be seen from (13) and (14), both the signal distortion and the acoustic contrast are functions of the regularization parameter μ and the subspace dimension V. For a fixed $\mu \geq 0$, with increasing V, the signal distortion $\xi_{\mathrm{sd}}^{\mathrm{o}}$ becomes smaller but the acoustic contrast $\xi_{\mathrm{ac}}^{\mathrm{o}}$ becomes lower, and vice versa. Therefore, the trade-off between $\xi_{\mathrm{sd}}^{\mathrm{o}}$ and $\xi_{\mathrm{ac}}^{\mathrm{o}}$ can be obtained by adjusting the subspace dimension V. The largest acoustic contrast λ_1 can be obtained by setting $V = 1$ (i.e., the time-domain ACC approach), but this also results in the highest signal distortion in the bright zone. Using all the eigenvectors by setting $V = LJ$, the smallest signal distortion is obtained, but the acoustic contrast is also the lowest. The advantage of the VAST approach is that, based on the users' preference, the subspace dimension V can be adjusted to strike a balance between high acoustic contrast and low signal distortion.

However, μ and V are not physically meaningful and hard to tune. In the next section, we consider different performance metrics which can be set as constraints and present approaches to find an optimal value which fulfils the corresponding constraints.

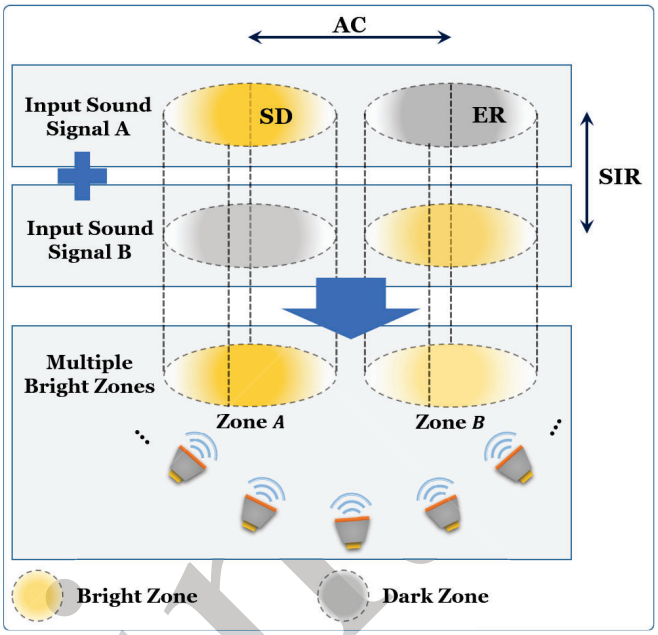

Fig. 2. An illustration for performance metrics in the scenario of multiple bright zones.

III. The Optimal Regularization Parameter for Generating Sound Zones

In this section, except for the acoustic contrast and signal distortion metrics respectively defined in (9) and (12), we introduce additional performance metrics, named dark zone energy reduction (ER) and signal to interference ratio (SIR)[1]. These metrics are depicted in Fig. 2. The influence of the regularization parameter μ on these performance metrics is analyzed and the solution space of μ for constraints on these performance metrics are derived. Finally, we show that the optimal regularization parameter can be obtained based on a variety of constrained optimization problems.

The dark zone energy reduction (ER) is defined as the ratio of energies between the unprocessed signal and the linearly filtered signal in the dark zone, i.e.,

$$\begin{aligned}\xi_{\mathrm{er}} =& \frac{\frac{1}{NM^{\mathrm{D}}}\sum_{m=1}^{M^{\mathrm{D}}}\sum_{n=1}^{N}|\mathbf{q}_{\mathrm{u}}^T\mathbf{H}^m\mathbf{x}_n|^2}{\frac{1}{NM^{\mathrm{D}}}\sum_{m=1}^{M^{\mathrm{D}}}\sum_{n=1}^{N}|y_n^{m,\mathrm{D}}|^2} \\ =& \frac{\mathbf{q}_{\mathrm{u}}^T\mathbf{R}^{\mathrm{D}}\mathbf{q}_{\mathrm{u}}}{\mathbf{q}^T\mathbf{R}^{\mathrm{D}}\mathbf{q}},\end{aligned} \tag{15}$$

where $\mathbf{q}_{\mathrm{u}} = \mathbf{1}_L \otimes \mathbf{i}_1$ and $\mathbf{1}_L$ denotes an all ones vector of length L. Clearly, for the generation of sound zones, it is desired to obtain a high ER. Note that the ER metric defined in (15) is similar to the noise reduction factor in signal enhancement defined in [39]. Substituting the optimal control filter (11) into

[1]In literature, for example, [38], the target-to-interferer ratio (TIR) is defined as the acoustic potential energy ratio or the loudness ratio between the reproduced and interfering sound fields in a given zone. Here, we specifically define SIR as the power ratio between the two sound fields in a given zone to avoid a confusion.

(15), we obtain

$$\xi_{\text{er}}^{\text{o}} = \frac{\mathbf{q}_{\text{u}}^T \mathbf{R}^{\text{D}} \mathbf{q}_{\text{u}}}{\sum_{i=1}^{V} c_i/(\lambda_i+\mu)^2} \ . \tag{16}$$

The signal to interference ratio (SIR) is defined as the ratio of energies between the linearly filtered target signal (e.g., source α) and the linearly filtered interference signal (e.g., source β) in the bright zone, i.e.,

$$\begin{aligned}\xi_{\text{sir}}(\mu_\alpha,\mu_\beta) =& \frac{\frac{1}{NM_\alpha^{\text{B}}}\sum_{m=1}^{M_\alpha^{\text{B}}}\sum_{n=1}^{N}|y_{n,\alpha}^{m,\text{B}}|^2/\sigma_\alpha^2}{\frac{1}{NM_\beta^{\text{D}}}\sum_{m=1}^{M_\beta^{\text{D}}}\sum_{n=1}^{N}|y_{n,\beta}^{m,\text{D}}|^2/\sigma_\beta^2} \\ =& \frac{\sigma_\beta^2 \mathbf{q}_\alpha^T \mathbf{R}_\alpha^{\text{B}} \mathbf{q}_\alpha}{\sigma_\alpha^2 \mathbf{q}_\beta^T \mathbf{R}_\beta^{\text{D}} \mathbf{q}_\beta} \ ,\end{aligned} \tag{17}$$

where the subscripts $(\cdot)_\alpha$ and $(\cdot)_\beta$ represent the target α and interference β, respectively, and $\sigma_\alpha^2 = \frac{1}{N}\sum_{n=1}^{N} x_{n,\alpha}^2$, $\sigma_\beta^2 = \frac{1}{N}\sum_{n=1}^{N} x_{n,\beta}^2$. Substituting the optimal control filter (11) for the target signal and interference signal into (17), we obtain

$$\xi_{\text{sir}}^{\text{o}}(\mu_\alpha,\mu_\beta) = \frac{\sigma_\beta^2 \sum_{i=1}^{V_\alpha} \lambda_{i,\alpha} c_{i,\alpha}/(\lambda_{i,\alpha}+\mu_\alpha)^2}{\sigma_\alpha^2 \sum_{i=1}^{V_\beta} c_{i,\beta}/(\lambda_{i,\beta}+\mu_\beta)^2} \ . \tag{18}$$

The SIR is closely related to the AC. In fact, it can be easily shown that

$$\xi_{\text{ac}}^{\text{o}}(\mu_\alpha)\xi_{\text{ac}}^{\text{o}}(\mu_\beta) = \xi_{\text{sir}}^{\text{o}}(\mu_\alpha,\mu_\beta)\xi_{\text{sir}}^{\text{o}}(\mu_\beta,\mu_\alpha) \ , \tag{19}$$

where $\xi_{\text{ac}}^{\text{o}}(\mu_\alpha)$ and $\xi_{\text{ac}}^{\text{o}}(\mu_\beta)$ denotes the AC for source α and β, respectively.

A. The regularization parameter for different constraints

In this subsection, we propose different methods to obtain the regularization parameter based on constraints on the physically meaningful metrics, i.e., AC, SD, ER, and SIR.

1) Constraint on the acoustic contrast (AC): A natural choice of the constraint on the acoustic contrast is

$$\xi_{\text{ac}}^{\text{o}} \geq g_{\text{ac}} \ , \tag{20}$$

where g_{ac} is a constant, and it represents the lowest allowed acoustic contrast. To solve (20), we first present the derivative of $\xi_{\text{ac}}^{\text{o}}$ w.r.t. μ ($\mu \geq 0$). When $V = 1$, $\frac{\partial \xi_{\text{ac}}^{\text{o}}}{\partial \mu} = 0$. When $V \geq 2$, we have

$$\begin{aligned}\frac{\partial \xi_{\text{ac}}^{\text{o}}}{\partial \mu} =& 2\frac{\sum_{i=1}^{V}\sum_{j=1}^{V}\frac{c_i c_j(\lambda_j-\lambda_i)}{(\lambda_i+\mu)^3(\lambda_j+\mu)^2}}{\left(\sum_{i=1}^{V}\frac{c_i}{(\lambda_i+\mu)^2}\right)^2} \\ =& 2\frac{\sum_{i=1}^{V-1}\sum_{j=i+1}^{V}\frac{c_i c_j(\lambda_i-\lambda_j)^2}{(\lambda_i+\mu)^3(\lambda_j+\mu)^3}}{\left(\sum_{i=1}^{V}\frac{c_i}{(\lambda_i+\mu)^2}\right)^2}, \\ \geq& 0 \ .\end{aligned} \tag{21}$$

Therefore, for a fixed V, $\xi_{\text{ac}}^{\text{o}}$ is a nondecreasing function for $\mu \geq 0$. The acoustic contrast $\xi_{\text{ac}}^{\text{o}}$ is in the range $[\lim_{\mu\to 0}\xi_{\text{ac}}^{\text{o}}, \lim_{\mu\to\infty}\xi_{\text{ac}}^{\text{o}})$. It can be readily shown that $\lim_{\mu\to 0}\xi_{\text{ac}}^{\text{o}} = (\sum_{i=1}^{V} c_i/\lambda_i)/(\sum_{i=1}^{V} c_i/\lambda_i^2)$. Based on the L'Hôpital's rule [40], for a fixed V, the upper bound for the acoustic contrast can be written as $\lim_{\mu\to\infty}\xi_{\text{ac}}^{\text{o}} = (\sum_{i=1}^{V} c_i\lambda_i)/(\sum_{i=1}^{V} c_i)$. In conclusion, for a fixed V, the solution for (20) exists when $g_{\text{ac}} \in [0, \lim_{\mu\to\infty}\xi_{\text{ac}}^{\text{o}})$. For simplicity, we assume that, for $V \geq 2$, not all the eigenvalues are equal and at least two elements in $\{c_i, 1 \leq i \leq V\}$ are not zeros. With this assumption, $\xi_{\text{ac}}^{\text{o}}$ is a strictly increasing function w.r.t. μ ($\mu \geq 0$). When $V \geq 2$, the solution space of μ for the acoustic contrast constraint (20) can be written as

$$\mu \in \begin{cases} [0,\infty) & g_{\text{ac}} < \lim_{\mu\to 0}\xi_{\text{ac}}^{\text{o}}, \\ [{\xi_{\text{ac}}^{\text{o}}}^{-1}(g_{\text{ac}}),\infty) & g_{\text{ac}} \in [\lim_{\mu\to 0}\xi_{\text{ac}}^{\text{o}}, \lim_{\mu\to\infty}\xi_{\text{ac}}^{\text{o}}), \end{cases} \tag{22}$$

where ${\xi_{\text{ac}}^{\text{o}}}^{-1}$ denote the inverse function of $\xi_{\text{ac}}^{\text{o}}$ defined in (14), and ${\xi_{\text{ac}}^{\text{o}}}^{-1}(g_{\text{ac}})$ can be obtained using Newton's method [41] for the problem $\xi_{\text{ac}}^{\text{o}} = g_{\text{ac}}$. When $V = 1$ and $g_{\text{ac}} \leq \lambda_1$, the solution space for μ is $[0,\infty)$.

2) Constraint on the signal distortion (SD): A natural choice of the constraint on the signal distortion is

$$\xi_{\text{sd}}^{\text{o}} \leq g_{\text{sd}} \ , \tag{23}$$

where g_{sd} is a constant, and it denotes the maximum allowed signal distortion. The derivative of ξ_{sd} w.r.t. μ can be written as

$$\frac{\partial \xi_{\text{sd}}^{\text{o}}}{\partial \mu} = \frac{2\sum_{i=1}^{V}\frac{\mu}{(\lambda_i+\mu)^3}c_i}{\kappa^{\text{B}}} \ . \tag{24}$$

As can be seen from (24), for a fixed V, $\xi_{\text{sd}}^{\text{o}}$ is a strictly increasing function w.r.t. μ when $\mu > 0$, and it has a stationary point when $\mu = 0$. Therefore, the smallest and highest signal distortion can be obtained when $\mu = 0$ and $\mu \to \infty$, respectively. The signal distortion $\xi_{\text{sd}}^{\text{o}}$ is in the range $[\lim_{\mu\to 0}\xi_{\text{sd}}^{\text{o}}, 1)$, where $\lim_{\mu\to 0}\xi_{\text{sd}}^{\text{o}}$ can be computed by substituting $\mu = 0$ into (13). The solution space of μ for the signal distortion constraint (23) can be written as

$$\mu \in [0, {\xi_{\text{sd}}^{\text{o}}}^{-1}(g_{\text{sd}})] \quad \text{for} \quad g_{\text{sd}} \in [\lim_{\mu\to 0}\xi_{\text{sd}}^{\text{o}}, 1) \ , \tag{25}$$

where ${\xi_{\text{ac}}^{\text{o}}}^{-1}(g_{\text{sd}})$ can be obtained using Newton's method for the problem $\xi_{\text{sd}}^{\text{o}} = g_{\text{sd}}$.

3) Constraint on the dark zone energy reduction (ER): A natural choice of the constraint on the dark zone energy reduction is

$$\xi_{\text{er}}^{\text{o}} \geq g_{\text{er}} \ . \tag{26}$$

where g_{er} is a constant, and it denotes the minimum allowed dark zone energy reduction. For a fixed V, it can be easily verified that $\xi_{\text{er}}^{\text{o}}$ is a strictly increasing function of μ when $\mu \geq 0$. The dark zone energy reduction $\xi_{\text{er}}^{\text{o}}$ is in the range $[\lim_{\mu\to 0}\xi_{\text{er}}^{\text{o}}, \infty)$, where $\lim_{\mu\to 0}\xi_{\text{er}}^{\text{o}}$ can be computed by substituting $\mu = 0$ into (16). For a fixed V, the solution for (26) exists when $g_{\text{er}} \in [0,\infty)$. The solution space of μ can be written as

$$\mu \in \begin{cases} [0,\infty) & g_{\text{er}} < \lim_{\mu\to 0}\xi_{\text{er}}^{\text{o}}, \\ [{\xi_{\text{ac}}^{\text{o}}}^{-1}(g_{\text{er}}),\infty) & g_{\text{er}} \geq \lim_{\mu\to 0}\xi_{\text{er}}^{\text{o}}, \end{cases} \tag{27}$$

where $\xi_{\text{er}}^{\text{o}\,-1}(g_{\text{er}})$ can be obtained using Newton's method for the problem $\xi_{\text{er}}^{\text{o}} = g_{\text{er}}$.

4) Constraint on the signal to interference ratio (SIR): The constraints presented subsections in III-A1, III-A2 and III-A3 can be seen as source-independent approaches, since the optimal regularization parameters μ_α and μ_β for different sources (e.g., sources α and β) are obtained independently given the RIRs. Here, we present an approach to obtain μ_α and μ_β simultaneously by applying constraints on the SIRs for the considered two sources. The constraints can be formulated as follows:

$$\begin{aligned} \xi_{\text{sir}}^{\text{o}}(\mu_\alpha, \mu_\beta) &\geq g_{\alpha,\text{sir}}, \\ \xi_{\text{sir}}^{\text{o}}(\mu_\beta, \mu_\alpha) &\geq g_{\beta,\text{sir}}, \end{aligned} \tag{28}$$

where $g_{\alpha,\text{sir}}$ and $g_{\beta,\text{sir}}$ are constants and they denote the minimum allowed SIRs for source α and β, respectively. The solution space is shown in (52) (see details in Appendix A).

B. Optimization criterion and the optimal regularization parameter

In this subsection, we present some optimization criteria for the generation of sound zones and derive the optimal regularization parameter.

1) Minimizing SD with a constraint on ER (SD-ER): The optimization criterion can be written as

$$\arg\min_{\mu} \xi_{\text{sd}}^{\text{o}}, \quad \text{s.t.,} \quad \xi_{\text{er}}^{\text{o}} \geq g_{\text{er}}. \tag{29}$$

Since $\xi_{\text{sd}}^{\text{o}}$ is a nondecreasing function of μ, using (27), the solution for (29) can be written as

$$\hat{\mu} = \begin{cases} 0 & g_{\text{er}} < \lim_{\mu\to 0} \xi_{\text{er}}^{\text{o}}, \\ \xi_{\text{ac}}^{\text{o}\,-1}(g_{\text{er}}) & g_{\text{er}} \geq \lim_{\mu\to 0} \xi_{\text{er}}^{\text{o}}. \end{cases} \tag{30}$$

2) Maximizing ER with a constraint on SD (ER-SD): The optimization criterion can be written as

$$\arg\max_{\mu} \xi_{\text{er}}^{\text{o}}, \quad \text{s.t.,} \quad \xi_{\text{sd}}^{\text{o}} \leq g_{\text{sd}}. \tag{31}$$

Since $\xi_{\text{er}}^{\text{o}}$ is an increasing function of μ, using (25), the solution for (31) can be written as

$$\hat{\mu} = \xi_{\text{sd}}^{\text{o}\,-1}(g_{\text{sd}}) \quad \text{for} \quad g_{\text{sd}} \in [\lim_{\mu\to 0} \xi_{\text{sd}}^{\text{o}}, 1), \tag{32}$$

3) Minimizing SD with a constraint on AC (SD-AC): The optimization criterion can be written as

$$\arg\min_{\mu} \xi_{\text{sd}}^{\text{o}}, \quad \text{s.t.,} \quad \xi_{\text{ac}}^{\text{o}} \geq g_{\text{ac}}. \tag{33}$$

Since $\xi_{\text{sd}}^{\text{o}}$ is a nondecreasing function of μ, using (22), the solution for (33) can be written as

$$\hat{\mu} = \begin{cases} 0 & g_{\text{ac}} < \lim_{\mu\to 0} \xi_{\text{ac}}^{\text{o}}, \\ \xi_{\text{ac}}^{\text{o}\,-1}(g_{\text{ac}}) & g_{\text{ac}} \in [\lim_{\mu\to 0} \xi_{\text{ac}}^{\text{o}}, \lim_{\mu\to\infty} \xi_{\text{ac}}^{\text{o}}), \end{cases} \tag{34}$$

Algorithm 1 The GEVD algorithm for the generation of sound zones

1: Perform the GEVD for the generalized eigenvalue problem $\mathbf{R}^{\text{B}}\mathbf{u}_i = \lambda_i \mathbf{R}^{\text{D}}\mathbf{u}_i$.
2: Compute $c_i = |\mathbf{u}_i^T \mathbf{r}^{\text{B}}|^2, \quad 1 \leq i \leq LJ$
3: Select a subspace dimension V based on the performance bounds presented in Section III-A.
4: Obtain the optimal regularization parameter μ based on the optimization criteria presented in Section III-B.
5: Obtain the reduced-rank filter using (11).

4) Minimizing SD with constraints on SIR (SD-SIR): The optimization criterion can be written as

$$\begin{aligned} \arg\min_{\mu_\alpha,\mu_\beta} \quad & \xi_{\text{sd}}^{\text{o}}(\mu_\alpha) + \xi_{\text{sd}}^{\text{o}}(\mu_\beta), \\ \text{s.t.,} \quad & \xi_{\text{sir}}^{\text{o}}(\mu_\alpha, \mu_\beta) \geq g_{\alpha,\text{tir}}, \\ & \xi_{\text{sir}}^{\text{o}}(\mu_\beta, \mu_\alpha) \geq g_{\beta,\text{tir}}, \end{aligned} \tag{35}$$

Since $\xi_{\text{sd}}^{\text{o}}(\mu_\alpha)$ and $\xi_{\text{sd}}^{\text{o}}(\mu_\beta)$ are nondecreasing functions w.r.t. α and β, respectively, using (52), the solution for (33) can be obtained as follows:

1. When $k_2(\hat{\mu}_\alpha^{\text{start}}) \geq k(\hat{\mu}_\alpha^{\text{start}})$, $(\hat{\mu_\alpha}, \hat{\mu_\beta}) = (\hat{\mu}_\alpha^{\text{start}}, k(\hat{\mu}_\alpha^{\text{start}}))$ is the solution, where $k(\cdot)$, $k_2(\cdot)$, $\hat{\mu}_\alpha^{\text{start}}$ are defined in (47), (51) and (50), respectively.
2. When $k_2(\hat{\mu}_\alpha^{\text{start}}) < k(\hat{\mu}_\alpha^{\text{start}})$ and $k_2(\infty) \geq k(\infty)$, the solution of (33) can be obtained by solving $\xi_{\text{sir}}^{\text{o}}(\mu_\alpha, \mu_\beta) = g_{\alpha,\text{tir}}$ and $\xi_{\text{sir}}^{\text{o}}(\mu_\beta, \mu_\alpha) = g_{\beta,\text{tir}}$, which can be obtained using Newton's method.
3. When $k_2(\hat{\mu}_\alpha^{\text{start}}) < k(\hat{\mu}_\alpha^{\text{start}})$ and $k_2(\infty) < k(\infty)$, no solution exists for (33).

The LJ-dimensional GEVD algorithm for the generation of sound zones is summarized in Algorithm 1.

IV. The Conjugate Gradient Method for Generating Sound Zones

As noted before, the GEVD-based approach for the generation of sound zones suffers from a high signal distortion when the subspace dimension V is small. Moreover, when LJ is large, it is computationally complex to solve the LJ-dimensional generalized eigenvalue problem. To mitigate these problems, a conjugate gradient (CG) method for the generation of sound zones has recently been proposed in [36], which is referred to as RR-CG. In RR-CG, the subspace is constructed by using the search directions in the CG method for solving $\mathbf{R}^{\text{B}}\mathbf{q} = \mathbf{r}^{\text{B}}$. Clearly, compared to the GEVD-based method, $\mathbf{r}^{\text{B}}$ that contains the desired signal information is exploited when constructing the subspace. Unfortunately, a degradation of $4 - 5$ dB in performance in terms of AC or SD has been reported due to the loss of orthogonality of the search directions in the CG method [42] when V is large. Moreover, the RR-CG method does not have a trade-off property in terms of AC and SD by adjusting the subspace dimension, as in the GEVD-based method.

In this section, we propose an improved approach for the generation of sound zones based on the CG method. Unlike the GEVD-based method and RR-CG, the subspace in the proposed method is constructed with the search directions

in the CG method by utilizing all the available information contained in $\mathbf{R}^{\mathrm{B}}$, $\mathbf{R}^{\mathrm{D}}$, and $\mathbf{r}^{\mathrm{B}}$. Moreover, to deal with the loss of orthogonality problem in the CG method, a $V_1 < LJ$-dimensional GEVD (V_1 is the number of iterations of the CG method) is used to rotate the subspace and find the generalized eigenvalues. To have a trade-off property like the GEVD-based method, a V-dimensional subspace ($V \leq V_1$) is formed by using the V eigenvectors corresponding to the V largest eigenvalues. Furthermore, the approaches for computing the optimal regularization parameter, described in Section III, can be easily integrated into the proposed method.

A. Reduced-rank filtering and the CG method

Definition 1. For a square matrix $\mathbf{R}$ and a nonzero vector $\mathbf{r}$, the subspace defined by

$$\mathcal{K}_V(\mathbf{R},\mathbf{r}) = \mathrm{span}\{\mathbf{r}, \mathbf{R}\mathbf{r}, \mathbf{R}^2\mathbf{r}, \cdots, \mathbf{R}^{V-1}\mathbf{r}\} \ , \tag{36}$$

is referred to as a V-dimensional Krylov subspace associated with the pair $(\mathbf{R},\mathbf{r})$.

Consider solving the following MSE minimization problem:

$$\hat{\mathbf{q}} = \arg\min_{\mathbf{q}} \mathbf{q}^T\mathbf{R}\mathbf{q} - 2\mathbf{q}^T\mathbf{r} + \kappa \tag{37}$$

on a V-dimensional subspace $\mathcal{K}_V$ associated with a basis matrix $\mathbf{U}_V$. The V-dimensional reduced-rank Wiener filter with the basis matrix $\mathbf{U}_V$ for solving (37) can be written as

$$\hat{\mathbf{q}} = \mathbf{U}_V(\mathbf{U}_V^T\mathbf{R}\mathbf{U}_V)^{-1}\mathbf{U}_V^T\mathbf{r} \ . \tag{38}$$

The Krylov subspace $\mathcal{K}_V(\mathbf{R},\mathbf{r})$ is commonly used to build the basis matrix, i.e., $\mathbf{U}_V = [\mathbf{r}, \mathbf{R}\mathbf{r}, \mathbf{R}^2\mathbf{r}, \cdots, \mathbf{R}^{V-1}\mathbf{r}]$. The motivations for this particular choice of the subspace can be found in [43]. A variety of approaches was proposed to obtain (38) using the Krylov subspace (36), such as the powers of $\mathbf{R}$ receiver [44], the multi-stage Wiener filter [45], and the conjugate gradient approach. Mathematically, these approaches are equivalent and have the same results. The computational complexity and the robustness against round-off errors are major concerns among these algorithms. In this paper, we focus on the CG algorithm because the search directions form the Krylov subspace $\mathcal{K}_V(\mathbf{R},\mathbf{r})$, and they are mutually $\mathbf{R}$-orthogonal, which will be explained later.

The CG algorithm iteratively minimizes the quadratic cost function (37) along with a set of $\mathbf{R}$-orthogonal search directions. The CG algorithm and its computational complexity for each step is shown in Algorithm 2 where $\mathbf{d}_p$ and $\mathbf{r}_p$ denote the search direction and the residual vector at the p^{th} iteration, respectively. One important property of the CG algorithm is that the search directions $\mathbf{d}_n, n \geq 1$ are $\mathbf{R}$-orthogonal mathematically, i.e.,

$$\mathbf{D}_p^T\mathbf{R}\mathbf{D}_p = \mathbf{\Lambda}_p^{\mathrm{CG}}, \quad 1 \leq p \leq P, \tag{39}$$

where $\mathbf{\Lambda}_p^{\mathrm{CG}} = \mathrm{diag}\{\mathbf{d}_1^T\mathbf{R}\mathbf{d}_1, \cdots, \mathbf{d}_p^T\mathbf{R}\mathbf{d}_p\}$, P is the total number of iterations, and $\mathbf{D}_p = [\mathbf{d}_1, \cdots, \mathbf{d}_p]$.

Algorithm 2 The conjugate gradient (CG) algorithm

1:	Initiate $\mathbf{q}_1 = \mathbf{0}, \mathbf{d}_1 = \mathbf{r}_1 = \mathbf{r}$, $g_1 = \mathbf{r}_1^T\mathbf{r}_1$	
2:	**for** $p = 1, 2, \cdots, P$ **do**	
3:	$\mathbf{c}_p = \mathbf{R}\mathbf{d}_p$	$\mathcal{O}(P^2)$
4:	$\alpha_p = \frac{g_p}{\mathbf{d}_p^T\mathbf{c}_p}$	$\mathcal{O}(P)$
5:	$\mathbf{q}_{p+1} = \mathbf{q}_p + \alpha_p\mathbf{d}_p$	$\mathcal{O}(P)$
6:	$\mathbf{r}_{p+1} = \mathbf{r}_p - \alpha_p\mathbf{c}_p$	$\mathcal{O}(P)$
7:	$g_{p+1} = \mathbf{r}_{p+1}^T\mathbf{r}_{p+1}$	$\mathcal{O}(P)$
8:	$\beta_{p+1} = \frac{g_{p+1}}{g_p}$	$\mathcal{O}(1)$
9:	$\mathbf{d}_{p+1} = \mathbf{r}_{p+1} + \beta_{p+1}\mathbf{d}_p$	$\mathcal{O}(P)$
10:	**end for**	

B. Proposed CG algorithm for generating sound zones

We consider building the Krylov subspace $\mathcal{K}_{V_1}(\mathbf{R}^{\mathrm{B}} + \mu^{\mathrm{CG}}\mathbf{R}^{\mathrm{D}}, \mathbf{r}^{\mathrm{B}})$ using the search directions in the CG algorithm (with $V_1 - 1$ iterations) for solving the following optimization problem:

$$\arg\min_{\mathbf{q}} J^{\mathrm{B}}(\mathbf{q}) + \mu^{\mathrm{CG}}J^{\mathrm{D}}(\mathbf{q}) \tag{40}$$

where μ^{CG} is a pre-defined constant value. The basis matrix $\mathbf{U}_{V_1}$ can be obtained by collecting the search directions, i.e., $\mathbf{U}_{V_1} = \mathbf{D}_{V_1}$. Substituting $\mathbf{q} = \mathbf{U}_{V_1}\mathbf{z}$ into the acoustic contrast metric (9), we can obtain

$$\xi_{\mathrm{ac}} = \frac{\mathbf{z}^T\mathbf{U}_{V_1}^T\mathbf{R}^{\mathrm{B}}\mathbf{U}_{V_1}\mathbf{z}}{\mathbf{z}^T\mathbf{U}_{V_1}^T\mathbf{R}^{\mathrm{D}}\mathbf{U}_{V_1}\mathbf{z}} \ . \tag{41}$$

Similar to the VAST approach, we form a V_1-dimensional subspace $\mathbf{M}_{V_1}$ that can jointly diagonalize both of the matrices $\mathbf{U}_{V_1}^T\mathbf{R}^{\mathrm{B}}\mathbf{U}_{V_1}$ and $\mathbf{U}_{V_1}^T\mathbf{R}^{\mathrm{D}}\mathbf{U}_{V_1}$ using V_1 eigenvectors corresponding to the V_1 largest eigenvalues, i.e.,

$$\mathbf{M}_{V_1}^T\mathbf{U}_{V_1}^T\mathbf{R}^{\mathrm{B}}\mathbf{U}_{V_1}\mathbf{M}_{V_1} = \mathbf{\Lambda}_{V_1} \ , \tag{42a}$$

$$\mathbf{M}_{V_1}^T\mathbf{U}_{V_1}^T\mathbf{R}^{\mathrm{D}}\mathbf{U}_{V_1}\mathbf{M}_{V_1} = \mathbf{I}_{V_1} \ , \tag{42b}$$

where $\mathbf{\Lambda}_{V_1} = \mathrm{diag}(\lambda_1, \cdots, \lambda_{V_1})$ denotes a diagonal matrix whose diagonal elements are the eigenvalues sorted as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{V_1}$. Using only the first V columns of $\mathbf{M}_V$ and letting $\mathbf{z} = \mathbf{M}_V\mathbf{g}$, the optimization problem (7) can be written as

$$\hat{\mathbf{g}} = \arg\min_{\mathbf{g}} J^{\mathrm{B}}(\mathbf{U}_{V_1}\mathbf{M}_V\mathbf{g}) + \mu J^{\mathrm{D}}(\mathbf{U}_{V_1}\mathbf{M}_V\mathbf{g}) \ . \tag{43}$$

Using (42) and (8), the optimal control filter can be further written as

$$\hat{\mathbf{q}} = \mathbf{U}_{V_1}\sum_{i=1}^{V}\frac{\mathbf{m}_i^T\mathbf{U}_{V_1}\mathbf{r}^{\mathrm{B}}}{\lambda_i + \mu}\mathbf{m}_i, \tag{44}$$

where $\mathbf{m}_i$ denotes the i^{th} column of $\mathbf{M}_V$. Defining $c_i = |\mathbf{m}_i^T\mathbf{U}_{V_1}\mathbf{r}^{\mathrm{B}}|^2$ and substituting (44) into AC, SD, ER and SIR metrics defined in (9), (12), (15) and (17), respectively, it can be shown that these metrics can also be re-written in the same formulas as (14), (13), (16) and (18), respectively. Therefore, the analysis of influence of μ on these performance metrics presented in Section III can also be applied for the proposed approximated algorithm. Note that when $\mu^{\mathrm{CG}} = 0$, since the directions in the CG algorithm are $\mathbf{R}^{\mathrm{B}}$–orthogonal, one V dimensional EVD can be used to

Algorithm 3 The CG algorithm for the generation of sound zones

1: Initiate μ^{CG}, V_1.
2: Run V_1-1 iterations of the CG algorithm for solving (40).
3: Build the basis matrix $\mathbf{U}_{V_1}$ based on the search directions.
4: Perform the GEVD for the generalized eigenvalue problem $\mathbf{U}_{V_1}^T\mathbf{R}^{\text{B}}\mathbf{U}_{V_1}\mathbf{m}_i = \lambda_i(\mathbf{U}_{V_1}^T\mathbf{R}^{\text{D}}\mathbf{U}_{V_1})\mathbf{m}_i$.
5: Compute $c_i = |\mathbf{m}_i^T\mathbf{U}_{V_1}\mathbf{r}^{\text{B}}|^2, \quad 1 \le i \le V_1$
6: Select a subspace dimension V based on the performance bounds presented in Section III-A.
7: Obtain the optimal regularization parameter μ based on the optimization criterions presented in Section III-B.
8: Obtain the reduced-rank filter using (44).

obtain $\mathbf{M}_V$. In this case, the generalized eigenvalue problem $\mathbf{U}_V^T\mathbf{R}^{\text{B}}\mathbf{U}_V\mathbf{m}_i = \lambda_i(\mathbf{U}_V^T\mathbf{R}^{\text{D}}\mathbf{U}_V)\mathbf{m}_i$ reduces to $1/\lambda_i\mathbf{m}_i = (\mathbf{\Lambda}_V^{\text{CG}})^{-1}(\mathbf{U}_V^T\mathbf{R}^{\text{D}}\mathbf{U}_V)\mathbf{m}_i$. The proposed algorithm using CG for the generation of sound zones is summarized in Algorithm 3.

V. Simulations

In this section, the performance of AC, SD, ER, and SIR accomplished by the proposed algorithms is evaluated under a specific sound zone setup with a circular array of loudspeakers. We first describe the experimental setup, data, different algorithms, and their acronyms in subsection V-A. The simulation results are shown in subsections V-B and V-C[2].

A. Experimental setup, data, algorithms and acronyms

Experimental setup: We consider a system that consists of a circular array with 16 evenly distributed loudspeakers with a radius of 2 m, two zones, and two virtual sources in a room of dimension 6.38 m × 5.4 m × 4.05 m, as shown in Fig. 3. It is assumed that zones A/B are the bright/dark zones for the source α and that zones B/A are the bright/dark zones for the source β, respectively. The number of control points (5 cm distance between the control points) used in each zone is set to $M = 25$. We assume that all the loudspeakers, control points, and virtual sources are located in the same plane at the height of 1.5 m. The sampling frequency is set to 16 kHz. The RIRs are generated using the RIR Generator toolbox [46], which is based on the image method [47]. The reverberation time is set to RT60 = 200 ms (i.e., $K = 3200$). The desired signal for the source α/β is generated by convolving the sound signal with the free-field RIRs, generated with the RIR Generator toolbox with RT60 = 0 s, from the virtual source α/β to the control points in zone A/B. The length of the control filters is set to $J = 240$ (15 ms). Therefore, the subspace dimension varies between $1 \le V \le LJ = 3840$.

Data: In the first and second experiments, two six-seconds long speech signals from the movie "Zootopia" in two different languages, i.e., English and Danish, are used as the sources α and β, respectively. In the third and fourth experiments, two ten-seconds long speech signal and music signal are used as the sources α and β, respectively.

[2]The MATLAB code and all the sound signals are available at https://tinyurl.com/rszern5

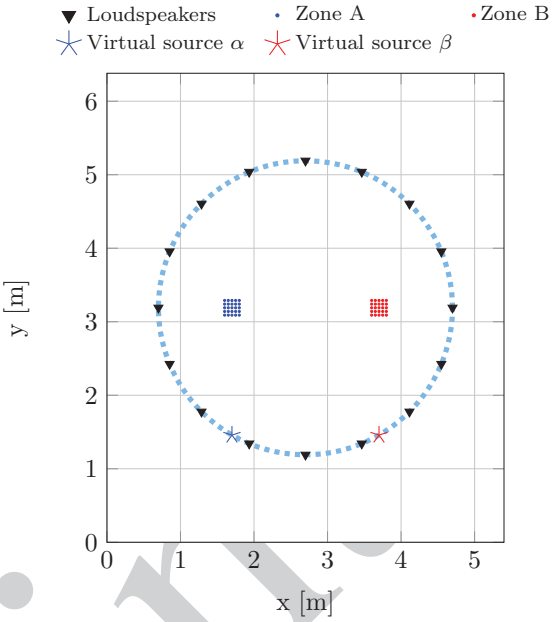

Fig. 3. An example of the system setup for sound zones with 16 loudspeakers, 25 control points in each zone and two virtual sources.

Algorithms and acronyms: In general, the performance of ER, SD, AC, and SIR performed by the proposed CG-based algorithms is compared to the traditional LJ-dimensional GEVD-based subspace construction method, i.e., VAST-GEVD [32] with a fixed μ but different choices of subspace dimension V. The proposed CG-based subspace construction algorithm with a fixed regularization parameter is referred to as VAST-CG. The SD-ER-GEVD and SD-ER-CG denote the GEVD-based and CG-based subspace construction algorithms using the criterion of minimizing the signal distortion with a constraint on energy reduction presented in subsection III-B1. Similarly, ER-SD-GEVD and ER-SD-CG denote the GEVD-based and CG-based subspace construction algorithms using the criterion of minimizing the energy reduction with a constraint on signal distortion presented in subsection III-B2. SD-AC-GEVD and SD-AC-CG denote the GEVD-based and CG-based subspace construction algorithms using the criterion of minimizing the signal distortion with a constraint on the acoustic contrast presented in subsection III-B3. SD-SIR-GEVD and SD-SIR-CG denote the GEVD-based and CG-based subspace construction algorithms using the criterion of minimizing the signal distortion with a constraints on the signal to interference ratio presented in subsection III-B4.

B. Results on the "speech and speech" example

In this subsection, we evaluate the performance of the GEVD and CG-based algorithms in terms of ER, SD, and AC with a fixed regularization parameter and optimal regu-

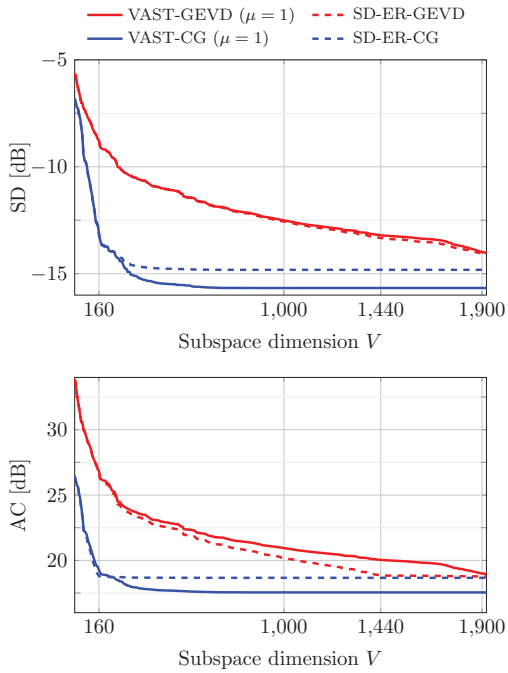

Fig. 4. The performance of SD and AC in different subspace dimensions using a fixed regularization parameter (solid) and using the SD-ER criterion (dashed) in Section III-B1. The GEVD-based and CG-based methods are in red and in blue, respectively.

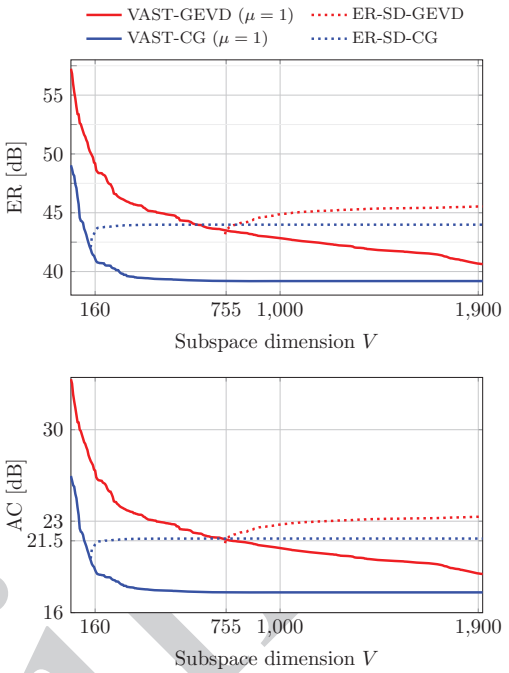

Fig. 5. The performance of ER and AC in different subspace dimensions using a fixed regularization parameter (solid) and using the ER-SD criterion (dashed) in Section III-B2. The GEVD-based and CG-based methods are in red and in blue, respectively.

larization parameters. Two speech signals are used, and the signal powers for them are set to be identical.

1) SD-ER criterion: In this section, we consider the optimization problem explained in Section III-B1; therefore, the performance of the VAST-GEVD, VAST-CG, SD-ER-GEVD, and SD-ER-CG in terms of SD followed by AC is tested for different choices of the subspace dimensions. The number of iterations for the CG-based methods is set to $V_1 = LJ/2$ (i.e., 1920), and so is the number of subspace dimension for the GEVD-based methods. The constraint g_{er} is set to 40 dB for SD-ER-CG and SD-ER-GEVD. A fixed regularization parameter is used for VAST-GEVD and VAST-CG, i.e., $\mu = 1$. The SD and AC of the GEVD-based and CG-based methods for source α with a fixed regularization parameter and an optimal regularization parameter are shown in Fig. 4. As can be seen, the GEVD-based methods, i.e., VAST-GEVD and SD-ER-GEVD, have a larger ER and AC but a lower SD than the CG-based methods, i.e., VAST-CG and SD-ER-CG, when $V \leq LJ/2$. Moreover, for SD-ER-GEVD, when $V \geq 1440$, the dark zone energy reduction ξ_{er} is about 40 dB. In contrast to this observation, for the CG-based method, a smaller V is required, i.e., $V \geq 160$, to reach $\xi_{\text{er}} \approx 40$ dB. Furthermore, the SD from both of the GEVD-based and CG-based methods is around -13 dB when $V = 1440$ and $V = 160$, respectively. By using the information of the spatial cross-correlation vector $\mathbf{r}$, the CG-based methods can obtain a low signal distortion with a smaller subspace dimension V compared to the GEVD-based methods.

2) ER-SD criterion: In the second experiment, we consider the optimization problem explained in Section III-B2; thus, the performance of the VAST-GEVD, VAST-CG, ER-SD-GEVD and ER-SD-CG in terms of ER and AC is tested for different choices of the subspace dimensions. The constraint g_{sd} is set to -12 dB for the SD-ER-CG and SD-ER-GEVD. All the other parameters are set to the same as the first experiment. ER and AC of the GEVD-based and CG-based methods with different regularization parameters for source α are shown in Fig. 5. As can be seen, for the SD-ER-GEVD, the signal distortion converges to the target value $\xi_{\text{sd}} \approx -12$ dB when $V \geq 755$; therefore, ER and AC increase. When $V \geq 1000$, ER and AC of the GEVD method with the optimal μ are around 2 dB higher than those of the CG based method with the optimal μ. Both ER and SD decrease with increasing V regardless of the methods if a fixed μ is used. Because the CG-based methods take $\mathbf{r}$ into account, they converge to the target optimization criterion faster than the GEVD-based methods. In this experiment, we can observe such a convergence from the CG-based and GEVD-based methods for $V \approx 160$ and $V \approx 755$, respectively. A similar observation can be drawn for the case of source β.

C. Results on the "speech and music" example

In this subsection, we evaluate the performance of the GEVD-based and CG-based algorithms with optimal regularization parameters using speech and music signals. The signal powers for the speech and music signals are set to be the same. Except for the audio examples, all the other parameters and the system setup are the same as the previous section.

1) SD-AC criterion: In the third experiment, we consider the optimization problem described in Section III-B3; hence,

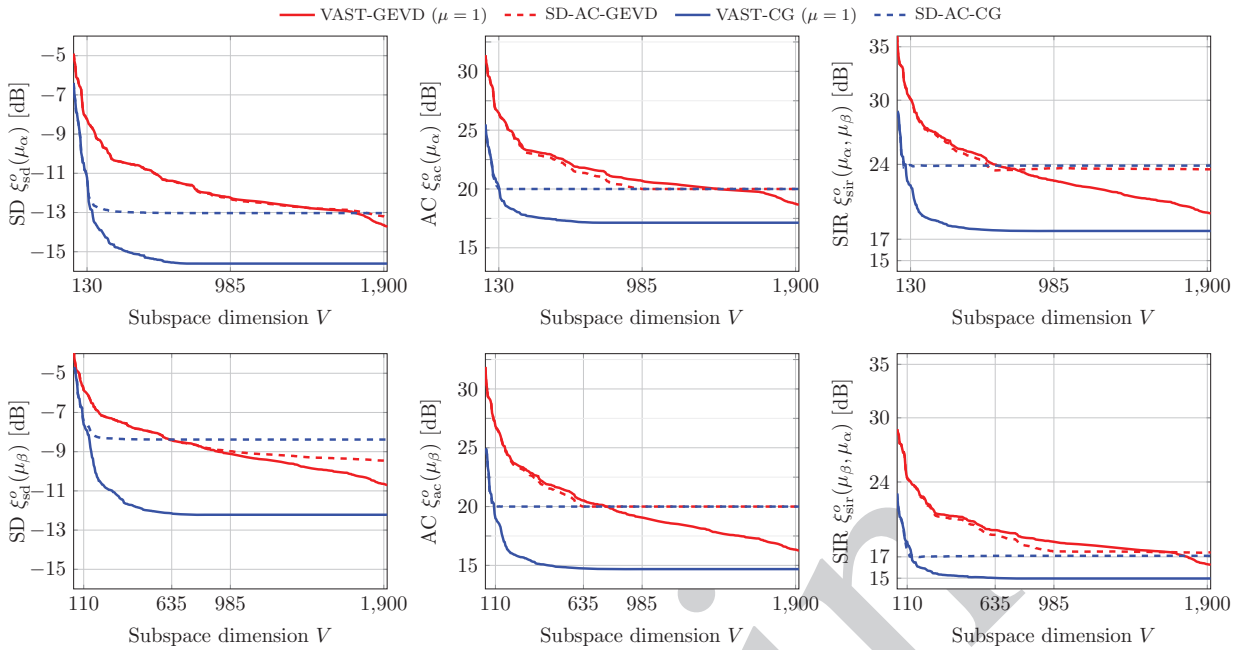

Fig. 6. SD, AC and SIR performance in different subspace dimensions using SD-AC criterion in Section III-B3. The upper and lower rows are the performance of each metric from zone A and zone B, respectively.

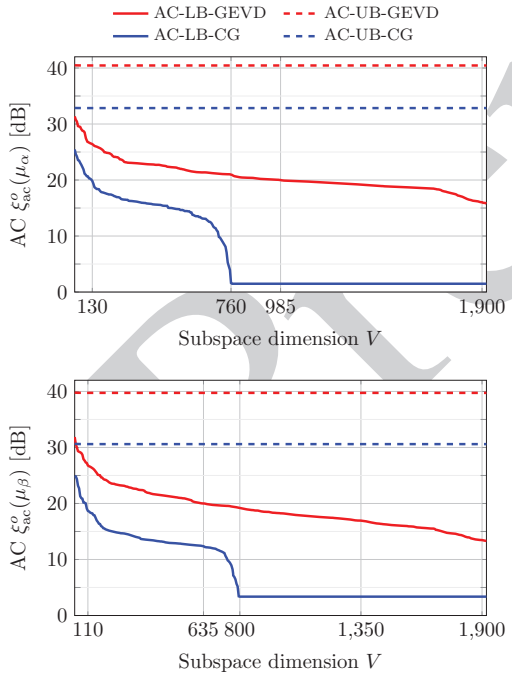

Fig. 7. The lower and upper bounds of AC in different subspace dimensions using SD-AC criterion in Section III-B3 for two different approaches.

the performance of the VAST-GEVD, VAST-CG, SD-AC-GEVD and SD-AC-CG in terms of SD, AC, and SIR is tested for different choices of the subspace dimensions. The upper and lower rows in Fig. 6 represent SD, AC, and SIR for source α and source β, respectively. The constraint on AC g_{ac} is set to 20 dB for the SD-AC-CG and SD-AC-GEVD. As can be seen, for the SD-AC-GEVD, when $V \geq 985$ for source α and $V \geq 635$ for source β, the acoustic contrast is $\xi_{\text{ac}} \approx 20$ dB. In contrast to the CG-based method, for the SD-AC-CG, when $V \geq 130$ for source α and $V \geq 110$ for source β, $\xi_{\text{ac}} \approx 20$ dB. For the speech source α, the SD, AC and SIR of the SD-AC-GEVD is close to the SD-AC-CG when $V = 1900$. For the music source β, the SD of the SD-AC-GEVD is around 1 dB lower than the SD-AC-CG when $V = 1900$. It is also shown that although the signal powers for the two sources are set to be the same and the AC for both sources are equal, there is a significant discrepancy between the SIRs of the two sources. The SIR is around 24 dB in zone A (the bright zone for source α), whereas the SIR in zone B (the bright zone for source β) is around 17 dB when $V = 1900$.

Besides, as shown in Fig. 7, we can observe the lower and upper bounds of AC for the GEVD-based and CG-based methods. AC-LB-GEVD, AC-UB-GEVD and AC-LB-CG, AC-UB-CG denote the lower and upper bounds of AC using the GEVD-based and CG-based methods, respectively. The upper bounds from both of the methods barely change with respect to the subspace dimension V. Also, the lower bound of AC for the CG-based method is constant when V is larger than a certain value (e.g., 760) in the upper panel. This is because the rank of $\mathbf{U}_V$ is 760. The similar trend can also be seen in the lower panel when V is approximately 800.

2) SD-SIR criterion: In the fourth experiment, we consider the optimization problem described in Section III-B4. It is worth noting that except for the optimization problem and the corresponding constraint, all the other parameters and the system setup are identical to the third experiment. The performance of the VAST-GEVD, VAST-CG, SD-SIR-GEVD and SD-SIR-CG in terms of SD, AC, and SIR is tested for

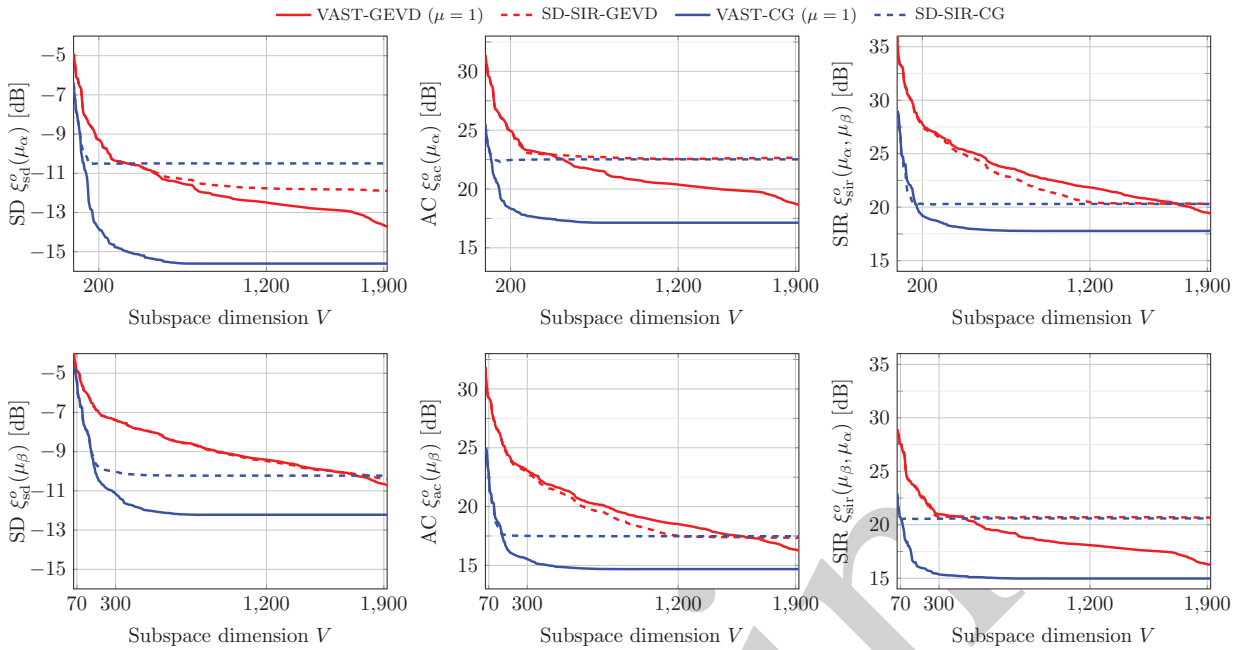

Fig. 8. SD, AC and SIR performance in different subspace dimensions using SD-SIR criterion in Section III-B4.

different choices of the subspace dimensions. The upper and lower rows in Fig. 8 represent SD, AC, and SIR for source α and source β, respectively. The constraint on SIR $g_{\rm sir}$ is set to 20 dB for the SD-SIR-CG and SD-SIR-GEVD. As can be seen, for the GEVD-based method, SIR $\xi_{\rm sir}$ converges to 20 dB when $V \geq 1200$ and $V \geq 300$ for sources α and β, respectively. In contrast, for the CG-based method, the SIR $\xi_{\rm sir}$ converges to 20 dB for sources α and β when $V \geq 200$ and $V \geq 70$, respectively. For source α, The SD performed by the GEVD-based method is less than 2 dB lower than the CG-based method when $V \geq 1200$.

VI. Conclusions

In this paper, subspace-based methods for sound zone control are proposed. Since it has proven in [36] that the CG-based sound zone control method is less computationally complex compared to the GEVD-based method in [32], but at the cost of performance degradation and loss of orthogonality of the basis functions, a modified version is proposed to persist the orthogonality of the basis functions as in the GEVD-based method and to introduce a fast convergence to the target response compared to the GEVD-based method.

The physically meaningful metrics, i.e., the acoustic contrast (AC), the signal distortion in the bright zone (SD), the energy reduction in the dark zone (ER), and the signal to interference ratio (SIR), are used as the constraints for reformulating the problem of generating sound zones. To this end, the control filters that accurately fulfill the constraints are successfully obtained. These diverse constraints are integrated into the two methods that based on GEVD and CG.

The performance of the proposed algorithms is evaluated on a circular array of loudspeakers for two bright zones scenario. Different constraints, which are summarized above, are considered for different audio data sets. It is shown that the CG-based method converges faster to the target constraints than the GEVD-based method because it takes the desired sound field into account when it seeks the basis functions.

Appendix

A. The solution space for the SIR constraints

Using the SIR definition (18), it can be easily shown that, for any $\mu_\alpha \geq 0$, the SIR is a strictly increasing function of μ_β for $\mu_\beta \geq 0$ where the SIR is in the range of $[\xi^{\rm o}_{\rm sir}(\mu_\alpha, 0), \infty]$. For the convenience of derivation, it is able to write the nominator and denominator in (18) as

$$f(\mu_\alpha) = \sigma_\beta^2 \sum_{i=1}^{V_\alpha} \lambda_{i,\alpha} c_{i,\alpha} / (\lambda_{i,\alpha} + \mu_\alpha)^2, \tag{45}$$

$$t(\mu_\beta) = \sigma_\alpha^2 \sum_{i=1}^{V_\beta} c_{i,\beta} / (\lambda_{i,\beta} + \mu_\beta)^2.$$

The solution space for the first constraint in (26) can be expressed as

$$\{(\hat{\mu}_\alpha, \hat{\mu}_\beta) | \hat{\mu}_\alpha \geq 0, k(\hat{\mu}_\alpha) \leq \hat{\mu}_\beta \leq \infty\}, \tag{46}$$

where the boundary function is defined as

$$k(\hat{\mu}_\alpha) = \begin{cases} 0 & f(\hat{\mu}_\alpha) > g_{\alpha,\rm sir} t(0), \\ t^{-1}(f(\hat{\mu}_\alpha)/g_{\alpha,\rm sir}) & f(\hat{\mu}_\alpha) \leq g_{\alpha,\rm sir} t(0) \end{cases}. \tag{47}$$

By noting that $f(\mu_\alpha)$ is a strictly decreasing function w.r.t. μ_α and $t(\mu_\beta)$ is a strictly decreasing function w.r.t. μ_β, it can be easily shown that the boundary function $k(\hat{\mu}_\alpha)$ is a nondecreasing function w.r.t. $\hat{\mu}_\alpha$ for $\hat{\mu}_\alpha \geq 0$. Next, we present

the solution space for the second constraint in (26). Define the nominator and denominator of $\xi_{\text{sir}}^{\text{o}}(\mu_\beta, \mu_\alpha)$ as

$$\begin{aligned} f_2(\mu_\beta) &= \sigma_\alpha^2 \sum_{i=1}^{V_\beta} \lambda_{i,\beta} c_{i,\beta} / (\lambda_{i,\beta} + \mu_\beta)^2 \ , \\ t_2(\mu_\alpha) &= \sigma_\beta^2 \sum_{i=1}^{V_\alpha} c_{i,\alpha} / (\lambda_{i,\alpha} + \mu_\alpha)^2 \ . \end{aligned} \tag{48}$$

It can be easily shown that, for any $\mu_\alpha \geq 0$, $\xi_{\text{sir}}^{\text{o}}(\mu_\beta, \mu_\alpha)$ is a decreasing function w.r.t. μ_β, and its value is in the range of $(0, \xi_{\text{sir}}^{\text{o}}(0, \mu_\alpha)]$. Therefore, the solution space for the second constraint in (26) can be expressed as

$$\{(\hat{\mu}_\alpha, \hat{\mu}_\beta) | \hat{\mu}_\alpha \geq \hat{\mu}_\alpha^{\text{start}}, 0 \leq \hat{\mu}_\beta \leq k_2(\hat{\mu}_\alpha)\}, \tag{49}$$

where

$$\hat{\mu}_\alpha^{\text{start}} = \begin{cases} 0 & \xi_{\text{sir}}^{\text{o}}(0,0) > g_{\beta,\text{sir}}, \\ t_2^{-1}(f_2(0)/g_{\beta,\text{sir}}) & \xi_{\text{sir}}^{\text{o}}(0,0) \leq g_{\beta,\text{sir}}, \end{cases} \tag{50}$$

and the boundary function is defined as

$$k_2(\hat{\mu}_\alpha) = f_2^{-1}(t_2(\hat{\mu}_\alpha) g_{\beta,\text{sir}}). \tag{51}$$

It can be easily shown that the boundary function $k_2(\hat{\mu}_\alpha)$ is a strictly increasing function w.r.t. $\hat{\mu}_\alpha$. The solution space for (26) is the intersection area between the sets defined in (46) and (49), i.e.,

$$\{(\hat{\mu}_\alpha, \hat{\mu}_\beta) | \hat{\mu}_\alpha \geq \hat{\mu}_\alpha^{\text{start}}, k(\hat{\mu}_\alpha) \leq \hat{\mu}_\beta \leq k_2(\hat{\mu}_\alpha)\}. \tag{52}$$

REFERENCES

[1] W. F. Druyvesteyn, R. M. Aarts, A. Asbury, P. Gelat, and A. Ruxton, "Personal sound," in *Proc. Inst. Acoust.*, vol. 16, no. 2, 1994, pp. 571–585.

[2] W. F. Druyvesteyn and J. Garas, "Personal sound," *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 685–701, Sep. 1997.

[3] S. J. Elliott and M. Jones, "An active headrest for personal audio," *J. Acoust. Soc. Am.*, vol. 119, no. 5, pp. 2702–2709, May 2006.

[4] J. Cheer and S. J. Elliott, "Design and implementation of a personal audio system in a car cabin," in *21st Int. Congr. Acoust.*, Montreal, QC, Canada, Aug. 2013.

[5] J.-W. Choi, "Subband optimization for acoustic contrast control," in *Proc. 22nd Int. Congr. Sound Vib.*, Florence, Italy, 2015.

[6] X. Liao, J. Cheer, S. J. Elliott, and S. Zheng, "Design array of loudspeakers for personal audio system in a car cabin," in *Proc. 23rd Int. Congr. Sound Vib.*, Athens, Greece, Jul. 2016.

[7] H. So and J.-W. Choi, "Subband optimization and filtering technique for practical personal audio systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brighton, UK, May 2019, pp. 8494–8498.

[8] J.-H. Chang and W.-H. Cho, "Evaluation of independent sound zones in a car," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany, Sep. 2019.

[9] S. Widmark, "Causal MSE-optimal filters for personal audio subject to constrained contrast," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 5, pp. 972–987, May 2019.

[10] J. Brunskog, F. M. Heuchel, D. C. Nozal, M. Song, F. T. Agerkvist, E. F. Grande, and E. Gallo, "Full-scale outdoor concert adaptive sound field control," in *23rd Int. Congr. Acoust.* Aachen, Germany: German Acoustical Society (DEGA), Sep. 2019, pp. 1170–1177.

[11] J. Cheer, S. J. Elliott, Y. Kim, and J.-W. Choi, "Practical implementation of personal audio in a mobile device," *J. Audio Eng. Soc.*, vol. 61, no. 5, pp. 290–300, Jun. 2013.

[12] J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim, "A realization of sound focused personal audio system using acoustic contrast control," *J. Acoust. Soc. Am.*, vol. 125, no. 4, pp. 2091–2097, Apr. 2009.

[13] J.-M. Lee, T. Lee, J.-Y. Park, and Y.-H. Kim, "Generation of a private listening zone; acoustic parasol," in *20th Int. Congr. Acoust.*, Sydney, Australia, Aug. 2010.

[14] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700, Apr. 2002.

[15] M. Poletti, "An investigation of 2-D multizone surround sound systems," in *Proc. 125th Conv. Audio Eng. Soc.*, San Francisco, CA, USA, Oct. 2008.

[16] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 19, no. 6, pp. 1711–1720, Aug. 2011.

[17] S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 20, no. 7, pp. 2123–2133, Sep. 2012.

[18] M. Shin, S. Q. Lee, F. M. Fazi, P. A. Nelson, D. Kim, S. Wang, K. Park, and J. Seo, "Maximization of acoustic energy difference between two spaces," *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 121–131, Jul. 2010.

[19] W. Zhang, P. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," *Appl. Sci.*, vol. 7, no. 6, p. 532, May 2017.

[20] S. J. Elliott and J. Cheer, "Regularisation and robustness of personal audio systems," ISVR Technical Memorandum 995, Tech. Rep., 2011.

[21] Y. Cai, M. Wu, and J. Yang, "Design of a time-domain acoustic contrast control for broadband input signals in personal audio systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 341–345.

[22] Y. Cai, M. Wu, L. Liu, and J. Yang, "Time-domain acoustic contrast control design with response differential constraint in personal audio systems," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. EL252–EL257, Jun. 2014.

[23] D. H. M. Schellekens, M. B. Møller, and M. Olsen, "Time domain acoustic contrast control implementation of sound zones for low-frequency input signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, Mar. 2016, pp. 365–369.

[24] M. B. Møller, M. Olsen, and F. Jacobsen, "A hybrid method combining synthesis of a sound field and control of acoustic contrast," in *Proc. 132nd Conv. Audio. Eng. Soc.*, Budapest, Hungary, Apr. 2012, P. 8627.

[25] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4518–4525, Jun. 2012.

[26] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, pp. 1869–1878, Nov. 2015.

[27] S. H. Jensen, P. C. Hansen, S. D. Hansen, and J. A. Sørensen, "Reduction of broad-band noise in speech by truncated QSVD," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 6, pp. 439–448, Nov. 1995.

[28] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[29] 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, Sep. 2002.

[30] J. R. Jensen, J. Benesty, and M. G. Christensen, "Noise reduction with optimal variable span linear filters," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 24, no. 4, pp. 631–644, Apr. 2016.

[31] J. Benesty, M. G. Christensen, and J. R. Jensen, *Signal enhancement with variable span linear filters*. Springer, 2016, vol. 7.

[32] T. Lee, J. K. Nielsen, J. R. Jensen, and M. G. Christensen, "A unified approach to generating sound zones using variable span linear filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Calgary, AB, Canada, Apr. 2018, pp. 491–495.

[33] J. K. Nielsen, T. Lee, J. R. Jensen, and M. G. Christensen, "Sound zones as an optimal filtering problem," in *Proc. 52th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2018, pp. 1075–1079.

[34] T. Lee, J. K. Nielsen, and M. G. Christensen, "Towards perceptually optimized sound zones: A proof-of-concept study," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brighton, UK, May 2019, pp. 136–140.

[35] W. Zhang, T. D. Abhayapala, T. Betlehem, and F. M. Fazi, "Analysis and control of multi-zone sound field reproduction using modal-domain approach," *J. Acoust. Soc. Am.*, vol. 140, no. 3, pp. 2134–2144, Sep. 2016.

[36] L. Shi, T. Lee, L. Zhang, J. K. Nielsen, and M. G. Christensen, "A fast reduced-rank sound zone control algorithm using the conjugate gradient method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2020, pp. 436–440.

[37] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, Mar. 2015.

[38] J. Francombe, P. Coleman, M. Olik, K. R. Baykaner, P. J. B. Jackson, R. Mason, S. Bech, M. Dewhirst, J. A. Pedersen, and M. Dewhirst, "Perceptually optimized loudspeaker selection for the creation of personal sound zones," in *Proc. 52nd Int. Conf. Audio. Eng. Soc.*, Guildford, UK, Sep. 2013.

[39] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, 2006.

[40] A. E. Taylor, "L'Hospital's rule," *Amer. Math. Monthly*, vol. 59, no. 1, pp. 20–24, 1952.

[41] C. T. Kelley, *Solving nonlinear equations with Newton's method*. SIAM, 2003.

[42] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," Carnegie Mellon University, Pittsburgh, PA, USA, Tech. Rep., Aug. 1994.

[43] S. Burykh and K. Abed-Meraim, "Reduced-rank adaptive filtering using krylov subspace," *EURASIP J. on Advances in Signal Processing*, vol. 2002, no. 12, p. 535098, 2003.

[44] M. L. Honig and W. Xiao, "Adaptive reduced-rank interference suppression with adaptive rank selection," in *Proc. MILCOM*, vol. 2, Los Angeles, CA, USA, Oct. 2000, pp. 747–751.

[45] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the wiener filter based on orthogonal projections," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2943–2959, 1998.

[46] E. A. P. Habets, "Room impulse response generator," Technische Universiteit Eindhoven, Tech. Rep., 2006.

[47] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.