KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Jesper Brunnström, Toon van Waterschoot, Marc Moonen Title: Signal-to-interference-plus-noise ratio based optimization for sound zone control Journal/Conference/Book: IEEE Open Journal of Signal Processing Vol/Book Chapter: Pages: ISBN:
Archived version	☐ Submitted version ☐ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ■ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ■ EU H2020 ☐ EU Horizon Europe ☐ FWO ■ KULeuven Internal Funds ☐ VLAIO ☐ AI ☐ Other
Published version	https://ieeexplore.ieee.org/abstract/document/10048524
Copyright statement	
Author contact	jesper.brunnstroem@kuleuven.be



IEEE Open Journal of Signal Processing

Received 27 October 2022; revised 8 February 2023; accepted 9 February 2023. Date of publication 17 February 2023; date of current version 26 May 2023. The review of this article was arranged by Associate Editor Michael Muma.

Digital Object Identifier 10.1109/OJSP.2023.3246398

Signal-to-Interference-Plus-Noise Ratio Based Optimization for Sound Zone Control

JESPER BRUNNSTRÖM (Graduate Student Member, IEEE), TOON VAN WATERSCHOOT (Member, IEEE), AND MARC MOONEN (Fellow, IEEE)

STADIUS Center for Dynamical Systems, Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium

CORRESPONDING AUTHOR: JESPER BRUNNSTRÖM (e-mail: jesper.brunnstroem@kuleuven.be).

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Council KU Leuven C14-21-0075. "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices". This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369: 'Service-Oriented Ubiquitous Network-Driven Sound—SOUNDS'. The scientific responsibility is assumed by the authors. The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

ABSTRACT Sound zone control is posed as an optimization problem where finite impulse response control filters are jointly optimized for low transmit power, while maintaining a sufficiently high signal-to-interference-plus-noise ratio (SINR) in all zones. This problem statement in particular allows the consideration of sound zone control under the additional influence of external noise, which is rarely taken into account and indeed necessitates an alternative problem statement. In addition, the spectral characteristics of the audio signals are taken into account, both for the transmit power minimization and computing the SINR. The optimization problem is shown to be solved optimally by semidefinite relaxation, but the computational cost is high. An alternate method is proposed, inspired by a duality between transmit and receive beamforming commonly exploited in digital communications. A virtual receive optimization problem for sound zone control is derived, and it is shown to have the same optimal solution as the original transmit optimization problem. The receive optimization problem is solved efficiently through a fixed point iteration algorithm. The two proposed methods are compared against acoustic contrast control in a simulated reverberant environment, where it is shown that the proposed methods require less transmit power to produce similar SINRs.

INDEX TERMS Array signal processing, convex optimization, duality, semi-definite relaxation, signal-to-interference-plus-noise ratio, sound zone control.

I. INTRODUCTION

In sound zone control different audio content is reproduced in different zones of the same space using a loudspeaker array [1]. Popular methods include acoustic contrast control (ACC) [2], [3], [4], [5], [6], [7] and pressure matching [8], [9], [10], both of which can be described as special cases of the variable span trade-off filter [11], [12]. Recent methods considering more general problem formulations include [13], [14], [15]. The task is to find a control filter for each loudspeaker and audio signal, such that when the filtered signals are transmitted, only the desired signal is heard in each zone.

Sound zone control is generally solved using superposition, a simplified problem formulation where only one zone has a desired signal that should be reproduced, whereas silence should be maintained in the other zones. Solving such a problem for each zone individually and superimposing the obtained solutions, the joint sound zone control problem is solved. In practice however, there is likely a non-negligible level of noise from external sources, which can be beneficial to take into account. In that case, the superposition strategy is no longer well suited, necessitating an alternative problem statement where all zones are considered jointly.

The sound zone control problem without superposition has a close resemblance to the transmit beamforming problem in digital communications [16], [17]. In contrast to transmit beamforming, in sound zone control the signal of interest is

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

the sound field rather than the microphone signals, meaning an operation analogous to receive beamforming to extract the transmitted signal from the noise and interference is not of interest. In addition, the transmitted signals are broadband, and the transmission channels are generally stationary on a longer time scale.

There is a duality between transmit and receive beamforming commonly exploited in digital communications, also referred to as uplink-downlink duality [18], [19], [20], [21]. It is particularly useful for solving transmit beamforming problems, which are in general more difficult than the corresponding receive beamforming problems [22]. An analogous duality exists for sound field reproduction, but is seldom used. It is used in [23] for sound zone control, however the method places significant restrictions on the loudspeaker and microphone arrays.

A common simplifying assumption in sound zone control is that the audio signals are spectrally white, which avoids the control filter being dependent on the likely non-stationary audio signals [5], [6], [7]. This is suboptimal, and sound zone control performance can be improved by taking the spectral characteristics of the audio signals into account.

Because the signals are broadband, the problem can be solved in the time domain or for a number of frequencies separately in the discrete Fourier transform domain. The latter generally has lower computational cost, but can give rise to degraded performance between frequency bins, especially when the control filter is short [5], [12], [24]. A time domain formulation avoids this problem, and allows for a natural way of taking the audio signal characteristics into account. However, it can introduce a new problem if the spectral distribution of the resulting audio is not controlled. The result can be a narrow bandpass effect at a frequency where the optimization criteria can be minimized. This problem is addressed in [5], [6], [7] for ACC, but the same ideas can be applied to the proposed methods.

The first contribution of this paper is the formulation of the sound zone control problem without superposition, taking both external noise and the spectral characteristics of the audio signals into account. Two methods are proposed to solve the problem, the first based on semi-definite relaxation (SDR), which is optimal but computationally costly. For the second method a virtual receive optimization problem for sound zone control is derived, and it is shown to have the same optimal solution as the original transmit optimization problem. It is solved efficiently through a modified version of the fixed point iteration algorithm from [20].

A. NOTATION

Vectors are denoted by lowercase bold letters (such as $\boldsymbol{a}$), matrices by uppercase bold letters (such as $\boldsymbol{A}$ or $\mathbb{A}$), and scalar variables by non-bold letters (such as a or A). Time-dependent quantities are functions of the discrete time index n. The inequality $\geq$ of a vector should be interpreted element-wise, and $\succcurlyeq$ denotes a generalized inequality over the cone of positive semi-definite matrices. The expectation operator is denoted by $\mathbb{E}$, and the trace operator by tr. The col operator creates a column vector from its arguments, with the ordering $\text{col}\{a(i)\}_{i=0}^{I} = [a(0) \quad a(1) \quad \dots \quad a(I)]^\top$. The diag and blkdiag operators create a diagonal and block diagonal matrix respectively with the analogous ordering. The identity matrix is denoted by $\boldsymbol{I}$, and a column vector with all ones is denoted by $\mathbf{1}$, the sizes of which should be clear from context.

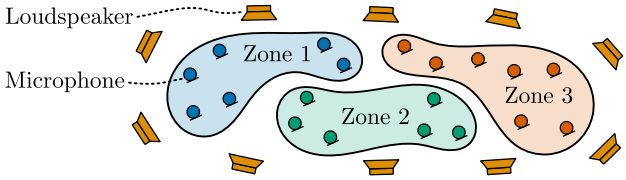

FIGURE 1. Depiction of the constituent parts of a sound zone control system with 3 zones.

II. PROBLEM STATEMENT

A. SIGNAL MODEL

An array of L loudspeakers transmits audio signals to Z zones, each with M_z microphones. The problem setting is depicted in Fig. 1. The set of indices for the zones is $\mathcal{Z} = \{1, \dots, Z\}$ and for the loudspeakers $\mathcal{L} = \{1, \dots, L\}$. The monophonic audio signal $x_z(n)$ associated with zone z should be reproduced only in that zone by filtering it with the I-tap finite impulse response (FIR) filter $\boldsymbol{w}_{zl} \in \mathbb{R}^I$ before driving loudspeaker l. The audio propagates from loudspeaker l to microphone m through an acoustic path modelled by the time-invariant J-tap room impulse response (RIR) $\boldsymbol{h}_{lm} \in \mathbb{R}^J$. The signal received at microphone m is

$$y_m(n) = \eta_m(n) + \sum_{l\in\mathcal{L}} \sum_{z\in\mathcal{Z}} \boldsymbol{h}_{lm}^\top \boldsymbol{X}_z(n) \boldsymbol{w}_{zl}, \tag{1}$$

where $\eta_m(n)$ is additive external noise, and $\boldsymbol{X}_z(n) \in \mathbb{R}^{J\times I}$ is a Toeplitz matrix whose jth row and ith column is $x_z(n-i-j+2)$. The signal model can be expressed more compactly by introducing the RIR matrix

$$\boldsymbol{H}_z = \begin{bmatrix} \boldsymbol{h}_{11} & \dots & \boldsymbol{h}_{1M_z} \\ \vdots & \ddots & \vdots \\ \boldsymbol{h}_{L1} & \dots & \boldsymbol{h}_{LM_z} \end{bmatrix} \in \mathbb{R}^{JL\times M_z} \tag{2}$$

for all RIRs associated with zone z, the duplicated audio signal matrix $\mathbb{X}_z(n) = \text{blkdiag}\{\boldsymbol{X}_z(n)\}_{l=1}^{L}$, the control filter $\boldsymbol{w}_z = \text{col}\{\boldsymbol{w}_{zl}\}_{l\in\mathcal{L}}$, and the noise signal $\boldsymbol{\eta}_z(n) = \text{col}\{\eta_m(n)\}_{m=1}^{M_z}$. The desired signal $\boldsymbol{y}_z^D(n)$ and interference-plus-noise $\boldsymbol{y}_z^I(n)$ for zone z are then obtained as

$$\begin{aligned} \boldsymbol{y}_z^D(n) &= \boldsymbol{H}_z^\top \mathbb{X}_z(n) \boldsymbol{w}_z \\ \boldsymbol{y}_z^I(n) &= \boldsymbol{\eta}_z(n) + \sum_{\substack{i\in\mathcal{Z} \\ i\neq z}} \boldsymbol{H}_z^\top \mathbb{X}_i(n) \boldsymbol{w}_i. \end{aligned} \tag{3}$$

B. SIGNAL-TO-INTERFERENCE-PLUS-NOISE RATIO

With the signal model (3), a natural metric for the quality of the sound zone control is the ratio of desired to undesired sound in each zone, which is the signal-to-interference-plus-noise ratio (SINR). The SINR can be written as

$$\mathrm{SINR}_z = \frac{\mathbb{E}\left[\|\boldsymbol{y}_z^D(n)\|_2^2\right]}{\mathbb{E}\left[\|\boldsymbol{y}_z^I(n)\|_2^2\right]} = \frac{\boldsymbol{w}_z^\top \boldsymbol{R}_{zz}\boldsymbol{w}_z}{\sum_{\substack{i\in\mathcal{Z}\\ i\neq z}} \boldsymbol{w}_i^\top \boldsymbol{R}_{zi}\boldsymbol{w}_i + \sigma_z^2}, \tag{4}$$

where the noise power of zone z is $\sigma_z^2 = \mathbb{E}[\|\boldsymbol{\eta}_z(n)\|_2^2]$, and the spatial covariance matrices are defined as

$$\boldsymbol{R}_{zi} = \mathbb{E}\left[\mathbb{X}_i^\top(n)\boldsymbol{H}_z\boldsymbol{H}_z^\top\mathbb{X}_i(n)\right] \in \mathbb{R}^{LI\times LI}. \tag{5}$$

The definition of the SINR used here to characterize a zone is using the summed power of the signals at the microphones, which is likely satisfactory if the microphones are spaced densely and evenly. Otherwise, a method such as kernel interpolation [25] could be applied according to [26]. Defining the SINR for each microphone individually instead of the zone as a whole is also possible, but leads to a non-convex NP-hard problem [27].

C. OPTIMIZATION PROBLEM

The considered optimization problem is a SINR constrained transmit power minimization problem, which allows for the desired level of audio separation to be selected via the SINR constraints. Transmit power in this context refers to the expected sum power of the signals emitted by the loudspeakers. With the loudspeaker signals denoted by $\boldsymbol{s}(n) \in \mathbb{R}^L$, the optimization problem is

$$\begin{aligned} \underset{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_Z}{\text{minimize}} \quad & \mathcal{P} = \mathbb{E}[\|\boldsymbol{s}(n)\|_2^2] + \alpha \sum_{z\in\mathcal{Z}} \|\boldsymbol{w}_z\|_2^2 \\ \text{subject to} \quad & \mathrm{SINR}_z \geq \gamma_z, \end{aligned} \tag{6}$$

where the parameter $\gamma_z \in \mathbb{R}_{\geq 0}$ is the required SINR for the zone, and $\alpha \in \mathbb{R}_{\geq 0}$ is a parameter controlling the amount of regularization. The regularization can help mitigate problems in the case of imperfectly estimated covariance matrices. Introducing an audio signal vector $\boldsymbol{x}_z(n) = \mathrm{col}\{x_z(n-i)\}_{i=0}^{I-1}$, the loudspeaker signals are $\boldsymbol{s}(n) = \mathrm{col}\{\sum_{z\in\mathcal{Z}} \boldsymbol{w}_{zl}^\top \boldsymbol{x}_z(n)\}_{l\in\mathcal{L}}$, and the audio signal covariance for zone z is $\boldsymbol{A}_z = \mathrm{blkdiag}\{\mathbb{E}[\boldsymbol{x}_z(n)\boldsymbol{x}_z^\top(n)]\}_{l=1}^L$. Assuming the audio signals for each zone are mutually independent and zero mean, the objective function in (6) can be rewritten as

$$\mathcal{P} = \sum_{z\in\mathcal{Z}} \boldsymbol{w}_z^\top \tilde{\boldsymbol{A}}_z \boldsymbol{w}_z, \tag{7}$$

where $\tilde{\boldsymbol{A}}_z = \boldsymbol{A}_z + \alpha\boldsymbol{I}$. Any additional penalty terms, to promote for example spectral or spatial uniformity, can be added to the matrix $\tilde{\boldsymbol{A}}_z$ as well [5], [6], [7]. The weighting matrix $\tilde{\boldsymbol{A}}_z$ can be viewed as a description of the user's preference between all control filters satisfying the SINR constraints.

D. ESTIMATION OF EXPECTED VALUES

The optimization objective in (6) includes the expectation over the audio signals, the RIRs, and the external noise power. These values must in general be estimated from the available data. The audio signals are assumed to be locally stationary, such that the covariances are approximately constant over a fixed segment length. With a segment of N samples starting at time index n_0, the audio signal covariances can be estimated as

$$\boldsymbol{A}_z \approx \sum_{n=n_0}^{n_0+N-1} \mathrm{blkdiag}\{\boldsymbol{x}_z(n)\boldsymbol{x}_z^\top(n)\}_{l=1}^L. \tag{8}$$

The spatial covariance matrices in (5) are estimated in an analogous way, but filtering the audio signals through the estimated RIRs first, and then calculating the sample covariance matrix.

The RIRs are slowly time-varying even if loudspeakers and zones are stationary, as temperature changes alone can cause a significant change [28]. The time-variations of the RIRs are likely slower than for the audio signals. The RIRs are therefore modelled as constant, and assumed to be measured to sufficient precision. If the variation of the RIRs is not excessive, they can be measured in an initial calibration stage and the microphones subsequently removed.

If the noise power is stationary over a long time, it could be directly estimated from the microphone signals before transmission of the audio signals is begun, for example when measuring the RIRs. In the general case, at least one microphone per zone is required to be present during operation. The transmitted audio signals at the microphones can be calculated according to (1) and subtracted from the recorded microphone signals, which will provide the true noise signal in the case of perfectly estimated RIRs.

E. ACOUSTIC CONTRAST CONTROL

For comparison, the sound zone control method ACC [5] can be obtained from the optimization problem

$$\begin{aligned} \underset{\boldsymbol{w}_z}{\text{maximize}} \quad & \boldsymbol{w}_z^\top \boldsymbol{R}_{zz}\boldsymbol{w}_z \\ \text{subject to} \quad & \boldsymbol{w}_z^\top \boldsymbol{R}_{-z}\boldsymbol{w}_z = \gamma_z, \end{aligned} \tag{9}$$

where $\boldsymbol{R}_{-z} = \alpha\boldsymbol{I} + \frac{1}{Z-1}\sum_{\substack{i\in\mathcal{Z}\\ i\neq z}} \boldsymbol{R}_{iz}$. The parameter $\alpha \in \mathbb{R}_{\geq 0}$ controls the regularization, applied according to [4]. The optimal control filter $\boldsymbol{w}_z$ is the principal generalized eigenvector of the pencil $(\boldsymbol{R}_{zz}, \boldsymbol{R}_{-z})$. The optimal control filter associated with each zone is computed independently, and then the obtained solutions are superimposed.

III. METHOD 1: SEMI-DEFINITE RELAXATION

The problem in (6) can be written in the form of a separable quadratically constrained quadratic program (QCQP). QCQPs are non-convex and NP-hard in general [29], however, some can be solved by SDR. Introducing the matrix $\boldsymbol{W}_z = \boldsymbol{w}_z\boldsymbol{w}_z^\top$ and using the fact that $\boldsymbol{w}_i^\top \boldsymbol{R}_{zi}\boldsymbol{w}_i = \mathrm{tr}(\boldsymbol{W}_i\boldsymbol{R}_{zi})$, the SDR of (6)

leads to the optimization problem

$$\begin{aligned} \underset{\boldsymbol{W}_1,\ldots,\boldsymbol{W}_Z}{\text{minimize}} \quad & \sum_{z\in\mathcal{Z}} \text{tr}(\boldsymbol{W}_z\tilde{\boldsymbol{A}}_z) \\ \text{subject to} \quad & \text{tr}(\boldsymbol{W}_z\boldsymbol{R}_{zz}) - \gamma_z \sum_{\substack{i\in\mathcal{Z}\\ i\neq z}} \text{tr}(\boldsymbol{W}_i\boldsymbol{R}_{zi}) \geq \sigma_z^2\gamma_z \\ & \boldsymbol{W}_z \succcurlyeq 0. \end{aligned} \tag{10}$$

The semi-definite program (10) is equivalent to (6) if the rank constraint $\text{rank}(\boldsymbol{W}_z) = 1$ is added for all z. However, the rank constraint is impractical as it is not convex. Without the rank constraint, the problem in (10) is a relaxation of the original problem (6), and therefore provides a lower bound for the optimal value. If a rank-1 solution is optimal for the relaxed problem, it is feasible for the original problem, and since the relaxed problem provides a lower bound, the solution is optimal for the original problem. The optimization problem (10) can also be identified as the Lagrangian bidual of (6).

It is shown in [29], [30], [31] that when SDR is used to solve a real-valued separable QCQP with K constraints and Z optimization variables, there is at least one optimal solution satisfying

$$\sum_{z=1}^{Z} \frac{\text{rank}(\boldsymbol{W}_z)(\text{rank}(\boldsymbol{W}_z)+1)}{2} \leq K. \tag{11}$$

As long as no optimal matrix is the zero matrix (a rank-0 matrix), a rank-1 solution is then guaranteed to exist if $K \leq Z+1$. The optimization problem (6) corresponds to a QCQP satisfying $K = Z$, and the zero matrix cannot be feasible for any $\gamma_z > 0$, so the problem (10) is equivalent to the original problem (6). When the optimal solution has been found, the control filter can be extracted as $\boldsymbol{w}_z = \sqrt{\lambda_z}\boldsymbol{u}_z$ where λ_z and $\boldsymbol{u}_z$ are the principal eigenvalue and eigenvector of the matrix $\boldsymbol{W}_z$ respectively.

IV. METHOD 2: VIRTUAL RECEIVE OPTIMIZATION PROBLEM

It can be shown that the optimal transmit control filters of (6) are the same up to a scaling as the optimal control filters for the receive optimization problem

$$\begin{aligned} \underset{\boldsymbol{q},\quad \bar{\boldsymbol{w}}_1,\ldots,\bar{\boldsymbol{w}}_Z}{\text{minimize}} \quad & \boldsymbol{\sigma}^\top\boldsymbol{q} \\ \text{subject to} \quad & \text{SINR}_z^{\text{rec}} \geq \gamma_z \\ & \boldsymbol{q} \geq 0 \\ & \|\bar{\boldsymbol{w}}_z\|_2 = 1, \end{aligned} \tag{12}$$

where the receive SINR is defined as

$$\text{SINR}_z^{\text{rec}} = \frac{q_z\bar{\boldsymbol{w}}_z^\top\boldsymbol{R}_{zz}\bar{\boldsymbol{w}}_z}{\bar{\boldsymbol{w}}_z^\top\left(\tilde{\boldsymbol{A}}_z + \sum_{\substack{i\in\mathcal{Z}\\ i\neq z}} q_i\boldsymbol{R}_{iz}\right)\bar{\boldsymbol{w}}_z}. \tag{13}$$

The power allocations associated with the different zones are collected in a vector $\boldsymbol{q} = \text{col}\{q_z\}_{z\in\mathcal{Z}}$, as are the noise powers $\boldsymbol{\sigma} = \text{col}\{\sigma_z^2\}_{z\in\mathcal{Z}}$. Two derivations of the receive optimization problem (12) is provided in Appendix A and B, together with proof that (12) and (6) are solved by the same control filters and with the same objective value.

A. FIXED POINT ITERATION

The problem in (12) can be solved using a fixed point iteration algorithm as in [20]. With an initial value of $\boldsymbol{q} = \boldsymbol{0}$, the following two steps are repeated until the control filters and power allocation have converged to a fixed value.

1) Calculate control filters $\bar{\boldsymbol{w}}_z$ according to Section IV-B.
2) Calculate power allocation $\boldsymbol{q}$ according to Section IV-C if (17) is satisfied, otherwise according to Section IV-D.

Once converged, the transmit power allocation for the original problem (6) can be calculated according to Section IV-E.

If the chosen SINR constraints can not be supported by any set of control filters, the problem will stay infeasible, meaning (17) will remain false. A maximum number of iterations can be set, after which the problem is considered infeasible. Because the algorithm generally only requires a few iteration to converge, the maximum number of iterations can be set to a moderately low number.

B. OPTIMAL CONTROL FILTER FOR GIVEN POWER ALLOCATION

For a given power allocation $\boldsymbol{q}$, the optimal control filter should maximize the SINR. Only the control filter $\bar{\boldsymbol{w}}_z$ affects $\text{SINR}_z^{\text{rec}}$ in (13), so the control filters can be found independently. Maximization of (13) is the maximization of a generalized Rayleigh quotient, which is solved by the principal generalized eigenvector of the pencil

$$\left(\boldsymbol{R}_{zz}, \tilde{\boldsymbol{A}}_z + \sum_{\substack{i\in\mathcal{Z}\\ i\neq z}} q_i\boldsymbol{R}_{iz}\right). \tag{14}$$

The principal generalized eigenvector is normalized to ensure $\|\bar{\boldsymbol{w}}_z\|_2 = 1$.

C. RECEIVE POWER ALLOCATION FOR FEASIBLE CONTROL FILTER

The optimal power allocation of (12) for a given feasible control filter will satisfy the SINR constraints with equality [20]. The optimal power allocation can therefore be directly solved for. Introducing the diagonal matrix $\boldsymbol{D} = \text{diag}\{\frac{\gamma_z}{\bar{\boldsymbol{w}}_z^\top\boldsymbol{R}_{zz}\bar{\boldsymbol{w}}_z}\}_{z\in\mathcal{Z}}$, the transmit power vector $\boldsymbol{s} = \text{col}\{\bar{\boldsymbol{w}}_z^\top\tilde{\boldsymbol{A}}_z\bar{\boldsymbol{w}}_z\}_{z\in\mathcal{Z}}$ and the interference matrix

$$\boldsymbol{\Psi}_{zi} = \begin{cases} \bar{\boldsymbol{w}}_i^\top\boldsymbol{R}_{zi}\bar{\boldsymbol{w}}_i & \text{if } z \neq i \\ 0 & \text{if } z = i \end{cases}, \tag{15}$$

the optimal power allocation for (12) is given by

$$\boldsymbol{q} = (\boldsymbol{I} - \boldsymbol{D}\boldsymbol{\Psi}^\top)^{-1}\boldsymbol{D}\boldsymbol{s}. \tag{16}$$

The power allocation problem is feasible if there is a power allocation $\boldsymbol{q} > 0$ that supports the requirement $\text{SINR}_z^{\text{rec}} \geq \gamma_z$ for all z. A sufficient and necessary criterion for feasibility is

$$\rho(\boldsymbol{D}\boldsymbol{\Psi}^\top) < 1, \tag{17}$$

where $\rho(\cdot)$ is the spectral radius [19]. The transmit power allocation is feasible if and only if the receive power allocation is feasible, because $\rho(\boldsymbol{D}\boldsymbol{\Psi}) = \rho(\boldsymbol{D}\boldsymbol{\Psi}^\top)$ [19].

D. RECEIVE POWER ALLOCATION FOR INFEASIBLE CONTROL FILTER

If the given control filter is not feasible according to (17), then the power allocation scheme in (16) cannot be used. The following power allocation scheme can then be used instead, which if iterated according to the algorithm in Section IV-A will find a feasible control filter if it exists.

While the control filter is infeasible, the power constrained SINR maximization problem will be solved instead, which can be stated as

$$\begin{aligned} C = \underset{\boldsymbol{q}}{\text{maximize}} \quad & \underset{z\in\mathcal{Z}}{\text{minimize}} \quad \frac{\text{SINR}_z^{\text{rec}}}{\gamma_z} \\ \text{subject to} \quad & \|\boldsymbol{q}\|_1 \leq P_{\max} \end{aligned} \tag{18}$$

for a given fixed control filter. The optimum of (18) will satisfy $\frac{\text{SINR}_z^{\text{rec}}}{\gamma_z} = \frac{\text{SINR}_i^{\text{rec}}}{\gamma_i}$ for all i, z [20]. The optimum objective can thereby be written in matrix form as

$$\boldsymbol{q}\frac{1}{C} = \boldsymbol{D}\boldsymbol{\Psi}^\top\boldsymbol{q} + \boldsymbol{D}\boldsymbol{s}. \tag{19}$$

The power constraint will also be satisfied exactly at the optimum, meaning that $\mathbf{1}^\top\boldsymbol{q} = P_{\max}$. Equation (19) can be multiplied by $\mathbf{1}^\top$ from the left, giving

$$\frac{1}{C} = \frac{1}{P_{\max}}\mathbf{1}^\top\boldsymbol{D}\boldsymbol{\Psi}^\top\boldsymbol{q} + \frac{1}{P_{\max}}\mathbf{1}^\top\boldsymbol{D}\boldsymbol{s}. \tag{20}$$

Together, (19) and (20) produces the eigenvalue problem

$$\boldsymbol{\Lambda}\begin{bmatrix}\boldsymbol{q}\\ 1\end{bmatrix} = \frac{1}{C}\begin{bmatrix}\boldsymbol{q}\\ 1\end{bmatrix} \tag{21}$$

in terms of the extended coupling matrix

$$\boldsymbol{\Lambda} = \begin{bmatrix}\boldsymbol{D}\boldsymbol{\Psi}^\top & \boldsymbol{D}\boldsymbol{s}\\ \frac{1}{P_{\max}}\mathbf{1}^\top\boldsymbol{D}\boldsymbol{\Psi}^\top & \frac{1}{P_{\max}}\mathbf{1}^\top\boldsymbol{D}\boldsymbol{s}\end{bmatrix}. \tag{22}$$

Only the maximum eigenvalue has a strictly positive eigenvector [20], so the power allocation is given by the first K components of the principal eigenvector of $\boldsymbol{\Lambda}$, scaled such that its last component is one.

If the ratio $\frac{\text{SINR}_z^{\text{rec}}}{\gamma_z} \geq 1$ is satisfied then so is the feasibility criterion (17). Because the ratio $\frac{\text{SINR}_z^{\text{rec}}}{\gamma_z}$ is equal for all z at the optimum, if the criterion can be satisfied for one zone, it can be satisfied for all zones. Therefore, if a feasible control filter exists, and $P_{\max}$ is set to a large enough value, the control filter will eventually be found. The dependence on $P_{\max}$ should not be restrictive in practice, as $P_{\max}$ can be set to a value much larger than a reasonable transmit power.

The global solution to the power constrained SINR maximization problem can be obtained instead of the SINR constrained power minimization problem by continuing to use this power allocation method until convergence.

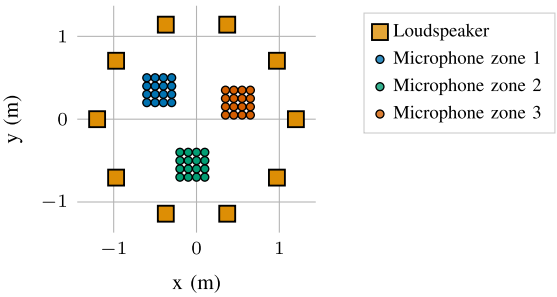

FIGURE 2. The positions of the microphones and loudspeakers.

E. TRANSMIT POWER ALLOCATION FOR OPTIMAL CONTROL FILTER

Once a set of normalized optimal control filters is obtained, the power allocation for (6) can be obtained directly as

$$\boldsymbol{p} = (\boldsymbol{I} - \boldsymbol{D}\boldsymbol{\Psi})^{-1}\boldsymbol{D}\boldsymbol{\sigma}. \tag{23}$$

This is a required last step, as the power allocation differs between the transmit and receive optimization problems. The final control filters are obtained as $\boldsymbol{w}_z = \sqrt{p_z}\bar{\boldsymbol{w}}_z$ for all z.

V. EXPERIMENTS

A. EXPERIMENT DESIGN

To characterize the performance of the proposed methods, experiments in a simulated reverberant environment have been conducted. There were three zones with 16 microphones each, sharing 10 loudspeakers, all placed on the same plane in a room of size $5 \times 4.8 \times 2$ m. The geometry is shown in Fig. 2. The reverberation was simulated with the randomized image-source method [32], [33] with a maximum image position shift of 5 cm, using the implementation from [34].

The external noise was simulated as spatio-temporally white Gaussian noise. The same noise power was used for all microphones associated with the same zone. The noise of any two microphones was mutually independent. The noise powers for the three zones were $\boldsymbol{\sigma} = [3 \quad 1 \quad 5]^\top$, and were assumed to be known. The SINR constraints were set to $\boldsymbol{\gamma} = [10 \quad 20 \quad 30]^\top$.

The considered methods were the SDR of Section III, the fixed point iteration (FPI) of Section IV, and ACC of Section II-E. To illustrate the impact of regularization, both FPI and

TABLE 1. Control Filter Square Norm, Computed Transmit Power, and Cost Function Using Synthetic Stationary Audio Signals

	$\sum_{z\in\mathcal{Z}}\|\boldsymbol{w}_z\|_2^2$	$\sum_{z\in\mathcal{Z}}\boldsymbol{w}_z^\top\boldsymbol{A}_z\boldsymbol{w}_z$	$\mathcal{P}$
SDR $\alpha=10^{-2}$	27.5	38.9	**39.2**
FPI $\alpha=10^{-2}$	27.5	38.9	**39.2**
FPI $\alpha=0$	16 204.9	**36.5**	198.5
FPI $\tilde{\boldsymbol{A}}_z=\boldsymbol{I}$	**11.7**	41.6	41.7
FPI $x_z(n)=\delta(n)$	16.8	62.3	62.5
ACC $\alpha=10^{-2}$	19.5	52.0	52.2
ACC $\alpha=0$	158.4	53.0	54.6

TABLE 2. Mean Transmit Power and SINR Using Synthetic Stationary Audio Signals Over 1 min in a Simulated Environment

	$\mathbb{E}\|\boldsymbol{s}(n)\|_2^2$	SINR_1	SINR_2	SINR_3
SDR $\alpha=10^{-2}$	39.5	9.3	20.3	30.9
FPI $\alpha=10^{-2}$	39.4	9.4	20.3	30.8
FPI $\alpha=0$	**32.5**	11.8	0.1	34.0
FPI $\tilde{\boldsymbol{A}}_z=\boldsymbol{I}$	41.8	9.3	19.9	30.8
FPI $x_z(n)=\delta(n)$	62.5	9.3	38.5	36.2
ACC $\alpha=10^{-2}$	53.2	9.5	20.2	31.1
ACC $\alpha=0$	54.2	9.5	20.3	31.1

ACC were computed with regularization values $\alpha=10^{-2}$ and $\alpha=0$. In addition, two modifications of FPI were used for comparison. The first, referred to as FPI with $\tilde{\boldsymbol{A}}_z=\boldsymbol{I}$, has the optimization objective $\sum_{z\in\mathcal{Z}}\|\boldsymbol{w}_z\|_2^2$, and is designed to show it is suboptimal to only minimize the norm of the control filter in a broadband setting. The second, referred to as FPI with $x_z(n)=\delta(n)$, assumes the audio signals to be spectrally white when computing the control filters, and is designed to show that it is essential to take the characteristics of the audio signal into account.

To solve the optimization problem (10) for the SDR, the splitting conic solver [35] was used, available in the Python package CVXPY [36], [37]. The maximum number of iterations for the solver was set to 10000. The FPI method was considered converged when the entry-wise mean square difference between both subsequent control filters and subsequent power allocations was less than 10^{-12}.

The control filters obtained from all methods were normalized and then scaled according to the power allocation method (23). The SINR constraints were selected to be relaxed enough such that all methods generated feasible control filters, so that the power allocation method could be used. Because ACC solves an entirely different optimization problem, it is not guaranteed to find a feasible control filter even if it exists.

B. SYNTHETIC STATIONARY SIGNALS

The first experiment considered stationary audio signals. Each audio signal was generated as white Gaussian noise filtered through a sharp bandpass filter, with bandwidths 40 Hz to 450 Hz, 40 Hz to 250 Hz, and 80 Hz to 450 Hz, for zone 1 to 3 respectively. The signals were scaled after the bandpass filter to have mean power $\mathbb{E}[x_1(n)^2]=4$, $\mathbb{E}[x_2(n)^2]=3$, and $\mathbb{E}[x_3(n)^2]=2$. The spatial covariance matrices were calculated using 10 s of audio. The sample rate was 1000 Hz, the control filter length $I=32$, and the reverberation time $\text{RT}_{60}=0.18$ s.

The square norm of the control filters is shown in Table 1, as well as the computed transmit power, i.e. the transmit power that would be generated if the estimated covariance $\boldsymbol{A}_z$ was the true covariance. The cost function $\mathcal{P}$, which is a linear combination of the two using $\alpha=10^{-2}$ is shown as well. It can be seen that for both ACC and FPI, the control filter norm becomes significantly larger when the regularization is set to zero.

Using the calculated control filters, the sound in the room was then simulated. The resulting SINR for each zone and total transmit power is shown in Table 2. The values were obtained from the mean power estimated from 1 min of audio. The covariance estimation errors lead to SINR values slightly different from the specified level for all methods. Because FPI with $x_z(n)=\delta(n)$ assumes the audio is spectrally white, the resulting SINR is not close to the desired values. For FPI, the regularization was critical to obtain a reasonable control filter, as Table 2 shows that the SINR values are not close to the targets without regularization. Even in other cases, such as for ACC in this experiment, the regularization can also lower the transmit power in practice. ACC required more transmit power to achieve similar SINRs compared to all proposed methods except FPI with $x_z(n)=\delta(n)$.

1) EQUIVALENCE OF PROPOSED METHODS

In Table 1 it can be seen that SDR and FPI obtained the same objective value. Note that the optimal control filters are ambiguous with regards to sign. If $\boldsymbol{w}_z$ is optimal, so is $-\boldsymbol{w}_z$. Therefore, the solutions from SDR and FPI are identical up to a sign difference, which is independent for each zone z, i.e. the two solutions can be equal for some z while being negative inverses for others. This does not affect the norm or the computed transmit power, as Table 1 confirms. However, in using the filters in practice, there can be some differences in performance depending on the signs, which can be seen in Table 2 if the values for SDR and FPI are compared. This performance difference is entirely accounted for if the sign differences are corrected.

2) COMPUTATIONAL COST

The primary computational cost of FPI and ACC is the generalized eigenvalue decomposition (GEVD), which is performed once per iteration and once in total, respectively. Each GEVD has a complexity of $\mathcal{O}((LI)^3)$ [38]. The matrices are of the same size, so one iteration of the FPI algorithm has

TABLE 3. Control Filter Square Norm, Computed Transmit Power, and Cost Function Using Music Signals

	$\sum_{z\in\mathcal{Z}}\|\boldsymbol{w}_z\|_2^2$	$\sum_{z\in\mathcal{Z}}\boldsymbol{w}_z^\top\boldsymbol{A}_z\boldsymbol{w}_z$	$\mathcal{P}$
FPI $\alpha=10^{-2}$	13.1	**26.1**	**26.3**
FPI $\alpha=0$	301.0	**26.1**	29.1
FPI $\tilde{\boldsymbol{A}}_z=\boldsymbol{I}$	**5.8**	52.8	52.8
FPI $x_z(n)=\delta(n)$	47.7	163.8	164.3
ACC $\alpha=10^{-2}$	38.3	358.9	359.3
ACC $\alpha=0$	346.3	513.2	516.7

TABLE 4. Mean Transmit Power and SINR Using Music Signals Over 1 min in a Simulated Environment

	$\mathbb{E}\|\boldsymbol{s}(n)\|_2^2$	SINR_1	SINR_2	SINR_3
FPI $\alpha=10^{-2}$	**26.5**	10.1	20.0	30.8
FPI $\alpha=0$	**26.5**	10.1	20.0	30.8
FPI $\tilde{\boldsymbol{A}}_z=\boldsymbol{I}$	53.1	10.2	19.8	30.9
FPI $x_z(n)=\delta(n)$	165.3	20.2	3.0	88.6
ACC $\alpha=10^{-2}$	375.2	10.1	20.1	31.1
ACC $\alpha=0$	532.8	10.1	20.4	31.1

about the same computational cost as ACC. For the experiment presented here, it took 6 to 7 iterations for FPI to converge. The computational cost of SDR is dependent on the choice of solver, but has a considerably higher complexity in terms of *LI* [29]. For the experiment presented here, SDR took 1752 s, FPI took between 0.233 s to 0.303 s, and ACC took 0.046 s. These numbers only include the time required to calculate the control filter and power allocation, and not the covariance matrices which are similarly calculated for all methods.

C. MUSIC SIGNALS

More realistic signals were also considered, specifically music from the dataset [39]. The sample rate was increased to 8000 Hz, the control filter length to $I=256$, and the reverberation time to $\text{RT}_{60}=0.55$ s. Other parameters were identical to the previous experiment. SDR was not included because of its high computational cost and because it gives identical results as FPI. The covariance matrices were computed from 1 min of audio. The computed transmit power and the control filter square norm can be seen in Table 3. The same 1 min of audio used to estimate the covariances was transmitted, for which the resulting transmit power and SINR can be seen in Table 4.

Table 4 shows that for FPI with $x_z(n)=\delta(n)$ which does not take the audio signal characteristics into account at all, the resulting SINR is not close to the desired values. The music signals are further from spectrally white compared to the synthetic signals, so the method performs even worse here.

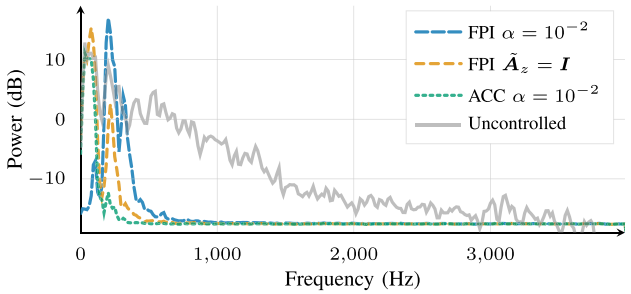

FIGURE 3. Mean spectrum of the microphone signals in zone 1 compared to the spectrum without sound zone control. The scaling of the uncontrolled audio signal is arbitrary, so is chosen to facilitate easy comparison.

Comparing FPI with $\tilde{\boldsymbol{A}}_z=\boldsymbol{I}$ to FPI with $\alpha=10^{-2}$, by including the audio signal characteristics in the power minimization, the resulting transmit power is approximately halved. Finally, ACC produces more than 10 times the transmit power compared to FPI at similar SINR.

The mean spectrum of the microphone signals in zone 1 is shown in Fig. 3 for a representative selection of the considered methods, compared against the resulting spectrum without sound zone control, which is defined as $x_1(n)$ transmitted unchanged from a single loudspeaker in the absence of interference and noise. Both FPI and ACC produces a narrowband response compared to the uncontrolled sound. To obtain a more spectrally smooth response, additional penalty terms can be added to the cost function (6).

VI. CONCLUSION

A formulation of sound zone control without superposition was proposed, where the control filters are jointly optimized for low transmit power, while maintaining a sufficiently high SINR. The optimization problem was shown to be solved optimally by semidefinite relaxation, but the computational cost was high. An alternate method was proposed, based on a virtual receive optimization problem derived from the original transmit optimization problem. The two problems were shown to have the same optimal control filters, while the former can be solved more efficiently. Through simulations it was shown that the proposed methods are effective, leading to lower transmit power at similar SINR compared to the popular ACC method. The importance of considering the spectral characteristics of the audio signals was demonstrated, as well as including appropriate regularization.

APPENDIX A
EQUIVALENCE OF TRANSMIT AND VIRTUAL RECEIVE OPTIMIZATION PROBLEMS

In this section the virtual receive optimization problem (12) is derived from the transmit optimization problem (6), demonstrating that the two problems have the same optimal control filters. The derivations follows the approach in [40].

The optimal solution of (6) will satisfy the SINR constraints with equality, meaning the problem can be equivalently expressed as

$$\begin{aligned} \underset{\boldsymbol{p},\bar{\boldsymbol{w}}_1,\ldots,\bar{\boldsymbol{w}}_Z}{\text{minimize}} \quad & \boldsymbol{p}^\top \boldsymbol{s} \\ \text{subject to} \quad & (\boldsymbol{I}-\boldsymbol{D\Psi})\boldsymbol{p} = \boldsymbol{D\sigma} \\ & \boldsymbol{p} > 0 \\ & \|\bar{\boldsymbol{w}}_z\| = 1. \end{aligned} \tag{24}$$

The positivity constraint for $\boldsymbol{p}$ can be replaced by a constraint on the spectral radius of the matrix $\boldsymbol{D\Psi}$. An irreducible non-negative square matrix $\boldsymbol{Q}$ has a strictly positive spectral radius $\rho(\boldsymbol{Q}) > 0$ which is an algebraically simple eigenvalue, and there is a unique vector that satisfies $\boldsymbol{x} > 0$, $\boldsymbol{x}^\top \boldsymbol{1} = 1$, and $\boldsymbol{Qx} = \rho(\boldsymbol{Q})\boldsymbol{x}$ [41, Th. 8.4.4]. If the constraints are fulfilled in (24), so that $(\boldsymbol{I}-\boldsymbol{D\Psi})\boldsymbol{p} = \boldsymbol{D\sigma}$, then looking at each row individually, all three of

$$p_z > 0, \qquad \frac{(\boldsymbol{D\Psi p})_z}{p_z} < 1, \qquad \frac{(\boldsymbol{D\sigma})_z}{p_z} > 0 \tag{25}$$

must be true. For a non-negative and irreducible matrix such as $\boldsymbol{D\Psi}$, the spectral radius is given by [42, Ch. 2, Eq. (2.11)]

$$\rho(\boldsymbol{D\Psi}) = \min_{x_z \geq 0} \max_{z \in \mathcal{Z}} \frac{(\boldsymbol{D\Psi x})_z}{x_z}. \tag{26}$$

By combining (25) and (26) the following can be asserted

$$\rho(\boldsymbol{D\Psi}) \leq \max_{z \in \mathcal{Z}} \frac{(\boldsymbol{D\Psi p})_z}{p_z} < 1. \tag{27}$$

Any feasible point from (24) will therefore satisfy $\rho(\boldsymbol{D\Psi}) < 1$, and the reverse is also true, that any $\boldsymbol{D\Psi}$ satisfying $\rho(\boldsymbol{D\Psi}) < 1$ will give rise to a strictly positive $\boldsymbol{p}$. This means that (24) can be equivalently rewritten as

$$\begin{aligned} \underset{\boldsymbol{p},\bar{\boldsymbol{w}}_1,\ldots,\bar{\boldsymbol{w}}_Z}{\text{minimize}} \quad & \boldsymbol{p}^\top \boldsymbol{s} \\ \text{subject to} \quad & (\boldsymbol{I}-\boldsymbol{D\Psi})\boldsymbol{p} = \boldsymbol{D\sigma} \\ & \rho(\boldsymbol{D\Psi}) < 1 \\ & \|\bar{\boldsymbol{w}}_z\| = 1. \end{aligned} \tag{28}$$

By solving for $\boldsymbol{p}$, the resulting optimization problem is

$$\begin{aligned} \underset{\bar{\boldsymbol{w}}_1,\ldots,\bar{\boldsymbol{w}}_Z}{\text{minimize}} \quad & \boldsymbol{\sigma}^\top \boldsymbol{D}(\boldsymbol{I}-\boldsymbol{D\Psi})^{-\top}\boldsymbol{s} \\ \text{subject to} \quad & \rho(\boldsymbol{D\Psi}) < 1 \\ & \|\bar{\boldsymbol{w}}_z\| = 1. \end{aligned} \tag{29}$$

At this point, it is possible to introduce an alternative power allocation, $\boldsymbol{q} = \boldsymbol{D}(\boldsymbol{I}-\boldsymbol{D\Psi})^{-\top}\boldsymbol{s}$. It can be equivalently written as $\boldsymbol{q} = (\boldsymbol{I}-\boldsymbol{D\Psi}^\top)^{-1}\boldsymbol{Ds}$ [41, Section 0.7.4]. It can be seen that if $\boldsymbol{q} \geq 0$ and $(\boldsymbol{I}-\boldsymbol{D\Psi}^\top)\boldsymbol{q} = \boldsymbol{Ds}$, that implies the spectral radius constraint. The following optimization problem is therefore equivalent

$$\begin{aligned} \underset{\boldsymbol{q},\bar{\boldsymbol{w}}_1,\ldots,\bar{\boldsymbol{w}}_Z}{\text{minimize}} \quad & \boldsymbol{\sigma}^\top \boldsymbol{q} \\ \text{subject to} \quad & (\boldsymbol{I}-\boldsymbol{D\Psi}^\top)\boldsymbol{q} = \boldsymbol{Ds} \\ & \boldsymbol{q} \geq 0 \\ & \|\bar{\boldsymbol{w}}_z\| = 1. \end{aligned} \tag{30}$$

Expanding the first constraint, the same optimization problem becomes

$$\begin{aligned} \underset{\boldsymbol{q},\bar{\boldsymbol{w}}_1,\ldots,\bar{\boldsymbol{w}}_Z}{\text{minimize}} \quad & \boldsymbol{\sigma}^\top \boldsymbol{q} \\ \text{subject to} \quad & \frac{q_z \bar{\boldsymbol{w}}_z^\top \boldsymbol{R}_{zz} \bar{\boldsymbol{w}}_z}{\bar{\boldsymbol{w}}_z^\top \left(\tilde{\boldsymbol{A}}_z + \sum_{\substack{z \in \mathcal{Z} \\ i \neq z}} q_i \boldsymbol{R}_{iz} \right) \bar{\boldsymbol{w}}_z} \geq \gamma_z \\ & \|\bar{\boldsymbol{w}}_z\| = 1 \\ & \boldsymbol{q} \geq 0, \end{aligned} \tag{31}$$

where the equality constraint is relaxed into an inequality constraint, which does not change the optimum. This is the virtual receive optimization problem (12) in a familiar form, and it is thereby shown that the optimal control filters for (31) and (6) are the same.

APPENDIX B
DERIVATION OF VIRTUAL RECEIVE OPTIMIZATION PROBLEM FROM LAGRANGIAN DUAL

The virtual receive optimization problem (12) can also be derived from the Lagrangian dual, which will be done in this section, following the approach of [43]. Because the Lagrangian bidual (10) is shown to have the same optimal objective value as the primal (6) in Section III, it means that the Lagrangian dual, and by extension the virtual receive optimization problem, will have the same optimal objective value as well.

The Lagrangian of (6), where the non-negative $\boldsymbol{q} \geq 0$ is the dual variable, is

$$\begin{aligned} \mathcal{L}(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_Z, \boldsymbol{q}) &= \sum_{z \in \mathcal{Z}} \boldsymbol{w}_z^\top \tilde{\boldsymbol{A}}_z \boldsymbol{w}_z - \\ & \sum_{z \in \mathcal{Z}} q_z \left(\frac{\boldsymbol{w}_z^\top \boldsymbol{R}_{zz} \boldsymbol{w}_z}{\gamma_z} - \sum_{\substack{i \in \mathcal{Z} \\ i \neq k}} \boldsymbol{w}_i^\top \boldsymbol{R}_{zi} \boldsymbol{w}_i - \sigma_z^2 \right) \\ &= \boldsymbol{\sigma}^\top \boldsymbol{q} + \sum_{z \in \mathcal{Z}} \boldsymbol{w}_z^\top \boldsymbol{G}_z \boldsymbol{w}_z, \end{aligned} \tag{32}$$

using the following definition

$$\boldsymbol{G}_z = \tilde{\boldsymbol{A}}_z - \frac{q_z}{\gamma_z} \boldsymbol{R}_{zz} + \sum_{\substack{i \in \mathcal{Z} \\ i \neq z}} q_i \boldsymbol{R}_{iz}. \tag{33}$$

The Lagrangian dual function is obtained by minimizing (32) over $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_Z$ [44]. The value of the dual function is $\boldsymbol{\sigma}^\top \boldsymbol{q}$ if it is bounded below. The Lagrangian dual problem is the maximization of the Lagrangian dual function, which is

$$\begin{aligned} \underset{\boldsymbol{q}}{\text{maximize}} \quad & \boldsymbol{\sigma}^\top \boldsymbol{q} \\ \text{subject to} \quad & \boldsymbol{G}_z \succcurlyeq 0 \\ & \boldsymbol{q} \geq 0, \end{aligned} \tag{34}$$

where the constraint ensures that the dual problem is feasible, that $\min_{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_Z} \mathcal{L}(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_Z, \boldsymbol{q}) > -\infty$.

The first constraint in (34) is equivalent to $\bar{\boldsymbol{w}}_z^\top \boldsymbol{G}_z \bar{\boldsymbol{w}}_z \geq 0$ for all $\bar{\boldsymbol{w}}_z$. The condition being true for all vectors is equivalent to $\max_{\bar{\boldsymbol{w}}_z} -\bar{\boldsymbol{w}}_z^\top \boldsymbol{G}_z \bar{\boldsymbol{w}}_z \leq 0$. Replacing the constraint in (34), the problem becomes

$$\begin{aligned} \underset{\boldsymbol{q}, \bar{\boldsymbol{w}}_1,\ldots,\bar{\boldsymbol{w}}_Z}{\text{maximize}} \quad & \boldsymbol{\sigma}^\top \boldsymbol{q} \\ \text{subject to} \quad & \frac{q_z \bar{\boldsymbol{w}}_z^\top \boldsymbol{R}_{zz} \bar{\boldsymbol{w}}_z}{\bar{\boldsymbol{w}}_z^\top \left(\tilde{\boldsymbol{A}}_z + \sum_{\substack{i \in \mathcal{Z} \\ i \neq z}} q_i \boldsymbol{R}_{iz} \right) \bar{\boldsymbol{w}}_z} \leq \gamma_z \\ & \boldsymbol{q} \geq 0. \end{aligned} \tag{35}$$

The constraints will be satisfied with equality at the optimum, so for a given control filter $\bar{\boldsymbol{w}}_z$ the power allocation $\boldsymbol{q}$ has a unique solution determined by the constraint. Therefore the solution of (35) is given by the fixed point of

$$\underset{\bar{\boldsymbol{w}}_1,\ldots,\bar{\boldsymbol{w}}_Z}{\text{maximize}} \frac{q_z \bar{\boldsymbol{w}}_z^\top \boldsymbol{R}_{zz} \bar{\boldsymbol{w}}_z}{\bar{\boldsymbol{w}}_z^\top \left(\tilde{\boldsymbol{A}}_z + \sum_{\substack{i \in \mathcal{Z} \\ i \neq z}} q_i \boldsymbol{R}_{iz} \right) \bar{\boldsymbol{w}}_z} = \gamma_z. \tag{36}$$

The same argument can be applied to the problem

$$\begin{aligned} \underset{\boldsymbol{q}, \bar{\boldsymbol{w}}_1,\ldots,\bar{\boldsymbol{w}}_Z}{\text{minimize}} \quad & \boldsymbol{\sigma}^\top \boldsymbol{q} \\ \text{subject to} \quad & \frac{q_z \bar{\boldsymbol{w}}_z^\top \boldsymbol{R}_{zz} \bar{\boldsymbol{w}}_z}{\bar{\boldsymbol{w}}_z^\top \left(\tilde{\boldsymbol{A}}_z + \sum_{\substack{i \in \mathcal{Z} \\ i \neq z}} q_i \boldsymbol{R}_{iz} \right) \bar{\boldsymbol{w}}_z} \geq \gamma_z \\ & \boldsymbol{q} \geq 0, \end{aligned} \tag{37}$$

so (35) and (37) have the same optimal solution. Scaling $\bar{\boldsymbol{w}}_z$ does not change the left-hand side of the constraint in (37). It is therefore possible to introduce the constraint $\|\bar{\boldsymbol{w}}_z\|_2 = 1$, meaning that the derived problem is exactly equal to (12).

REFERENCES

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

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

[3] 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. Acoustical Soc. America*, vol. 125, no. 4, pp. 2091–2097, Apr. 2009.

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

[5] 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.*, 2013, pp. 341–345.

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

[7] S. Zhao and I. S. Burnett, "Time-domain acoustic contrast control with a spatial uniformity constraint for personal audio systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 1061–1065.

[8] M. Poletti, "An investigation of 2-D multizone surround sound systems," in *Proc. Audio Eng. Soc. Conv.*, 2008, pp. 1–9.

[9] T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 437–440.

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

[11] 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.*, 2018, pp. 491–495.

[12] T. Lee, L. Shi, J. K. Nielsen, and M. G. Christensen, "Fast generation of sound zones using variable span trade-off filters in the DFT-domain," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 363–378, 2021.

[13] T. Lee, J. K. Nielsen, and M. G. Christensen, "Signal-adaptive and perceptually optimized sound zones with variable span trade-off filters," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 2412–2426, 2020.

[14] L. Shi, T. Lee, L. Zhang, J. K. Nielsen, and M. G. Christensen, "Generation of personal sound zones with physical meaningful constraints and conjugate gradient method," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 823–837, 2021.

[15] J. Zhang, L. Shi, M. G. Christensen, W. Zhang, L. Zhang, and J. Chen, "Robust pressure matching with ATF perturbation constraints for sound field control," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 8712–8716.

[16] F. Rashid-Farrokhi, K. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, Oct. 1998.

[17] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, 1999, pp. 987–996.

[18] E. Visotsky and U. Madhow, "Optimum beamforming using transmit antenna arrays," in *Proc. IEEE Veh. Technol. Conf.*, 1999, vol. 1, pp. 851–856.

[19] H. Boche and M. Schubert, "A general duality theory for uplink and downlink beamforming," in *Proc. IEEE Veh. Technol. Conf.*, 2002, vol. 1, pp. 87–91.

[20] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[21] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[22] E. Björnson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 142–148, Jul. 2014.

[23] G. Piñero, C. Botella, M. De Diego, M. Ferrer, and A. González, "On the feasibility of personal audio systems over a network of distributed loudspeakers," in *Proc. Eur. Signal Process. Conf.*, 2017, pp. 2729–2733.

[24] V. Molés-Cases, G. Piñero, M. de Diego, and A. Gonzalez, "Personal sound zones by subband filtering and time domain optimization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 2684–2696, 2020.

[25] N. Ueno, S. Koyama, and H. Saruwatari, "Directionally weighted wave field estimation exploiting prior information on source direction," *IEEE Trans. Signal Process.*, vol. 69, pp. 2383–2395, 2021.
[26] J. Brunnström, S. Koyama, and M. Moonen, "Variable span trade-off filter for sound zone control with kernel interpolation weighting," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2022, pp. 1071–1075.
[27] N. Sidiropoulos, T. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2239–2251, Jun. 2006.
[28] M. Olsen and M. B. Møller, "Sound zones: On the effect of ambient temperature variations in feed-forward systems," in *Proc. Audio Eng. Soc. Conv.*, 2017, pp. 1009–1018.
[29] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
[30] G. Pataki, "On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues," *Math. Oper. Res.*, vol. 23, no. 2, pp. 339–358, May 1998.
[31] A. I. Barvinok, "Problems of distance geometry and convex properties of quadratic maps," *Discrete Comput. Geometry*, vol. 13, no. 2, pp. 189–202, Mar. 1995.
[32] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Soc. America.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
[33] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 4, pp. 774–786, Apr. 2015.
[34] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulations and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 351–355.
[35] B. O'Donoghue, E. Chu, N. Parikh, and S. Boyd, "Conic optimization via operator splitting and homogeneous self-dual embedding," *J. Optim. Theory Appl.*, vol. 169, no. 3, pp. 1042–1068, Jun. 2016.
[36] S. Diamond and S. Boyd, "CVXPY: A python-embedded modeling language for convex optimization," *J. Mach. Learn. Res.*, vol. 17, pp. 1–5, 2016.
[37] A. Agrawal, R. Verschueren, S. Diamond, and S. Boyd, "A rewriting system for convex optimization problems," *J. Control Decis.*, vol. 5, no. 1, pp. 42–60, Jan. 2018.
[38] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 1996.
[39] Z. Rafii, A. Liutkus, S. Fabian-Robert, S. I. Mimilakis, and R. Bittner, "MUSDB18-HQ - an uncompressed version of MUSDB18," Dec. 2019, doi: 10.5281/zenodo.3338373.
[40] M. Bengtsson and B. Ottersten, "Optimum and suboptimum transmit beamforming," in *Handbook of Antennas in Wireless Communications*, L. C. Godara Ed., Boca Raton, FL, USA: CRC Press, 2001, ch. 18, pp. 18-1–18-33.
[41] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[42] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. Philadelphia, PA, USA: SIAM, 1994.
[43] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62–75, May 2010.
[44] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.