A Unified Approach to Generating Sound Zones Using Variable Span Linear Filters

Lee, Taewoong; Nielsen, Jesper Kjær; Jensen, Jesper Rindom; Christensen, Mads Græsbøll

Published in:
2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

DOI (link to publication from Publisher):
10.1109/ICASSP.2018.8462477

Publication date:
2018

Document Version
Accepted author manuscript, peer reviewed version

A UNIFIED APPROACH TO GENERATING SOUND ZONES USING VARIABLE SPAN LINEAR FILTERS

Taewoong Lee, Jesper Kjær Nielsen, Jesper Rindom Jensen, and Mads Græsbøll Christensen

Audio Analysis Lab, CREATE, Aalborg University, Aalborg, Denmark

{tlee,jkn,jrj,mgc}@create.aau.dk

ABSTRACT

Sound zones are typically created using Acoustic Contrast Control (ACC), Pressure Matching (PM), or variations of the two. ACC maximizes the acoustic potential energy contrast between a listening zone and a quiet zone. Although the contrast is maximized, the phase is not controlled. To control both the amplitude and the phase, PM instead minimizes the difference between the reproduced sound field and the desired sound field in all zones. On the surface, ACC and PM seem to control sound fields differently, but we here demonstrate they are actually extreme special cases of a much more general framework. The framework is inspired by the variable span linear filtering framework for speech enhancement. Using this framework, we demonstrate that 1) ACC gives the best contrast, but the highest signal distortion in the bright zone, and 2) PM gives the smallest signal distortion in the bright zone, but the worst contrast. Aside from showing this mathematically, we also demonstrate this via a small toy example.

***Index Terms*—** Sound zones, joint diagonalization, variable span linear filter, speech enhancement, personal sound.

1. INTRODUCTION

Sound zones enable multiple people to enjoy different audio content in the same acoustic environment without disturbing each other. This effect can be obtained using headphones, but headphones hinder social interaction between people. Therefore, a large loudspeaker array is instead employed to create the sound zones, and the control strategy of this loudspeaker array has been an active research field since the first attempt was made by Druyvesteyn and Garas in [1] two decades ago. They proposed that the control strategy should be frequency dependent. Specifically, they suggested to use active control for low frequencies, beamforming for mid frequencies, and directional loudspeakers for high frequencies.

Acoustic contrast control (ACC) and the concept of a bright zone and a dark zone were proposed by Choi and Kim [2] a few years later. In the bright zone, the desired sound should be present whereas the dark zone should be silent. By superposition, the concept of the bright and dark zones can be used to create two zones with different desired sound fields so most of the subsequent research has adopted this concept. In ACC, the acoustic contrast, which is defined as the acoustic potential energy ratio between the bright and dark zones, is the criterion that is maximized. Consequently, the ACC method does not try to match the reproduced sound field in the bright zone to the desired sound field, and this often results in audible distortions [3]. To minimize the distortion, the pressure matching (PM) method proposed by Poletti [4] minimizes the squared difference between the desired sound field and the reproduced sound field simultaneously in both zones. Since their introduction, ACC and PM have been used in many applications [5–9]. Although a number of different methods exist [10–17], including mode matching, most methods are based on ACC or PM.

ACC maximizes the contrast, but the difference between the desired sound field and the reproduced sound field is not taken into account. On the other hand, PM minimizes this differences, but at the cost of a reduced contrast. Although it has not been proved so far, there seem to be an empirical evidence for that ACC gives the best contrast, but the highest distortion, and that PM gives the smallest distortion, but the worst contrast. Therefore, many methods have tried to modify either ACC or PM to get an intermediate solution which allows us to trade-off distortion for contrast. Examples of such methods can be found in [18–23].

Although ACC and PM seem to be two fundamentally different approaches [9, 16, 18, 24], they are actually two special cases of a much more general framework. In this paper, we present this framework which is adopted from speech enhancement where it has recently been introduced as variable span linear filtering (VSLF) [25, 26]. The framework has several well-known speech enhancement filters, i.e., minimum distortion, Wiener, and maximum SNR as the special cases. When used in the context of sound zone design, we also obtain a general framework where ACC, PM, and some of their variations are the special cases. The framework can also be used to prove that ACC indeed gives the best contrast, but the highest distortion, and that PM gives the smallest distortion, but the worst contrast. Throughout this paper, we consider the sound zone design problem in the time domain since the sound zone design problem in the frequency domain is a special case of the time domain formulation. We also give the conditions under which the time domain formulation reduces to the frequency domain formulation.

2. GENERATION OF SOUND ZONES

The problem of generating sound zones is often formulated as that of generating a bright zone and a dark zone. A listening zone with high acoustic potential energy is referred to as a bright zone and the listener located in this zone should hear the desired audio content. On the other hand, a quiet zone has the acoustic potential energy as low as possible and is referred to as a dark zone. As alluded to in the introduction, a solution for the bright and dark zones design problem can straightforwardly be used to create two different bright zones through the superposition principle.

We start by setting up the mathematical model corresponding to Fig. 1. A zone can be explained as a region whose sound field is controlled by a loudspeaker array. Every zone is sampled at a number of microphone positions, which are not necessarily the same in each zone. Throughout this paper, the subscript $_B$ and $_D$ represent the bright zone and the dark zone, respectively, and $_C$ represents the

union of the bright and dark zones. In the derivation below, we focus on the bright zone, but the same derivation can be made for the dark zone. Assume that we have M_B microphones in the bright zone. In the absence of noise, the mth microphone measures the signals emitted by the L loudspeakers, convolved with the room impulse response (RIR) $h_{ml}[n]$ with $n \in \{0, \cdots, K-1\}$ from loudspeaker l to microphone m. The input of the lth loudspeaker is the signal $x[n]$ convolved with a finite impulse response (FIR) function $q_l[n]$ with $n \in \{0, \cdots, J-1\}$. We know $x[n]$ and $h_{ml}[n]$ (the latter is typically measured), but the control filters $\{q_l[n]\}_{l=1}^{L}$ are unknown. Thus, the objective is to design $q_l[n]$ to generate the sound zones. The reproduced sound pressure on the mth microphone position in the bright zone at the nth time sample can be written as

$$\begin{aligned} p_{ml}[n] &= \sum_{k=0}^{K-1} h_{ml}[k] \sum_{j=0}^{J-1} q_l[j] x[n-k-j] \\ &= \boldsymbol{h}_{ml}^T \boldsymbol{X}[n] \boldsymbol{q}_l, \end{aligned} \tag{1}$$

where $(\cdot)^T$ denotes the transpose of a vector or a matrix,

$$\boldsymbol{h}_{ml} = \begin{bmatrix} h_{ml}[0] & \cdots & h_{ml}[K-1] \end{bmatrix}^T \in \mathbb{R}^{K \times 1} \tag{2a}$$

$$\boldsymbol{q}_l = \begin{bmatrix} q_l[0] & \cdots & q_l[J-1] \end{bmatrix}^T \in \mathbb{R}^{J \times 1}, \tag{2b}$$

and $\boldsymbol{X}[n] = \big\{x[n-k-j+2]\big\}_{k=1,\cdots,K, j=1,\cdots,J}$. By summing the contribution from all L loudspeakers, we obtain

$$p_m[n] = \sum_{l=1}^{L} p_{ml}[n] = \sum_{l=1}^{L} \boldsymbol{h}_{ml}^T \boldsymbol{X}[n] \boldsymbol{q}_l = \boldsymbol{h}_m^T \mathbb{X}[n] \boldsymbol{q}, \tag{3}$$

where $\boldsymbol{h}_m = \begin{bmatrix} \boldsymbol{h}_{m1}^T & \cdots & \boldsymbol{h}_{mL}^T \end{bmatrix}^T$, $\boldsymbol{q} = \begin{bmatrix} \boldsymbol{q}_1^T & \cdots & \boldsymbol{q}_L^T \end{bmatrix}^T$, and $\mathbb{X}[n] = \boldsymbol{I}_L \otimes \boldsymbol{X}[n]$ are of size $LK \times 1$, $LJ \times 1$, and $LK \times LJ$, respectively. $\otimes$ is the Kronecker product, and $\boldsymbol{I}_L$ is the identity matrix of size L. Thus, the reproduced sound field in the bright zone $\boldsymbol{p}_B[n] = \begin{bmatrix} p_1[n] & \cdots & p_{M_B}[n] \end{bmatrix}^T$ can be written as

$$\boldsymbol{p}_B[n] = \boldsymbol{H}_B^T[n] \boldsymbol{q} \in \mathbb{R}^{M_B \times 1}, \tag{4}$$

where $\boldsymbol{H}_B[n] = \mathbb{X}^T[n] \begin{bmatrix} \boldsymbol{h}_1 & \cdots & \boldsymbol{h}_{M_B} \end{bmatrix} \in \mathbb{R}^{LJ \times M_B}$ is referred to as a spatial information matrix of the bright zone and is known. The reproduced sound field in the dark zone $\boldsymbol{p}_D[n]$ is defined in the same manner. Thus, the reproduced sound field and the spatial information matrix for the total zone are given by

$$\boldsymbol{p}_C[n] = \begin{bmatrix} \boldsymbol{p}_B^T[n] & \boldsymbol{p}_D^T[n] \end{bmatrix}^T \in \mathbb{R}^{(M_B+M_D) \times 1} \tag{5a}$$

$$\boldsymbol{H}_C[n] = \begin{bmatrix} \boldsymbol{H}_B[n] & \boldsymbol{H}_D[n] \end{bmatrix} \in \mathbb{R}^{LJ \times (M_B+M_D)}. \tag{5b}$$

Since the acoustic potential energy at the mth microphone position is calculated as $p_m^2[n]$, the average acoustic potential energy density in the bright zone is represented as the spatially-temporally averaged quantity

$$e_B = \frac{1}{M_B N} \sum_{n=0}^{N-1} \boldsymbol{p}_B^T[n] \boldsymbol{p}_B[n] = \frac{1}{M_B} \boldsymbol{q}^T \boldsymbol{R}_B \boldsymbol{q}, \tag{6}$$

where N is the number of time samples recorded by each microphone, $\boldsymbol{R}_B = N^{-1} \sum_{n=0}^{N-1} \boldsymbol{H}_B[n] \boldsymbol{H}_B^T[n]$ is a real-symmetric and positive (semi)definite, and is referred to as a spatial correlation matrix of the bright zone. If $M_B N \geq LJ$ is satisfied, $\boldsymbol{R}_B$ has full rank if $x[n]$ and $h_{ml}[n]$ are not trivial signals such as the zero vector. $\boldsymbol{R}_D$ is defined in the same manner for the dark zone. Finally, we also have that $\boldsymbol{R}_C = \boldsymbol{R}_B + \boldsymbol{R}_D$.

The acoustic contrast γ is an important quantity in order to describe the ratio of e_B to e_D. It is defined by [2]

$$\gamma = \frac{e_B}{e_D} = \frac{M_D \boldsymbol{q}^T \boldsymbol{R}_B \boldsymbol{q}}{M_B \boldsymbol{q}^T \boldsymbol{R}_D \boldsymbol{q}} = \kappa^2 \frac{\boldsymbol{q}^T \boldsymbol{R}_B \boldsymbol{q}}{\boldsymbol{q}^T \boldsymbol{R}_D \boldsymbol{q}}, \tag{7}$$

where $\kappa^2 = M_D / M_B$ and $e_D = M_D^{-1} \boldsymbol{q}^T \boldsymbol{R}_D \boldsymbol{q}$.

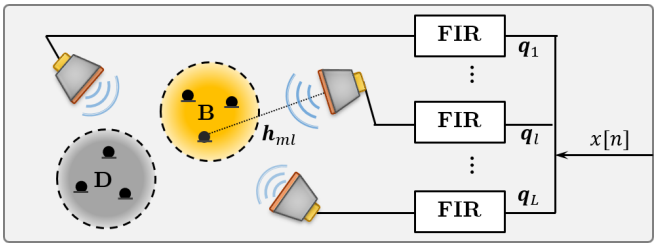

Fig. 1. System setup.

2.1. ACC

ACC, which was originally proposed in the frequency domain [2], maximizes e_B, while reducing e_D as much as possible. The frequency domain ACC is actually a special case of the time domain ACC, so we consider the latter case here. The ACC solution is obtained solving the constrained optimization problem

$$\text{maximize } e_B = M_B^{-1} \boldsymbol{q}^T \boldsymbol{R}_B \boldsymbol{q} \text{ s. t. } e_D = M_D^{-1} \boldsymbol{q}^T \boldsymbol{R}_D \boldsymbol{q}. \tag{8}$$

If we introduce a Lagrange multiplier γ, we can form the Lagrangian function [27] $\mathcal{L}_{\text{ACC}}(\boldsymbol{q}, \gamma) = M_B^{-1} \boldsymbol{q}^T \boldsymbol{R}_B \boldsymbol{q} + \gamma (M_D^{-1} \boldsymbol{q}^T \boldsymbol{R}_D \boldsymbol{q} - e_D)$. The solution $\boldsymbol{q}$ can be obtained by taking the derivative of $\mathcal{L}_{\text{ACC}}(\boldsymbol{q}, \gamma)$ with respect to $\boldsymbol{q}$ and setting the derivative equal to zero. Doing this, we obtain a generalized eigenvalue problem of the form

$$\kappa^2 \boldsymbol{R}_B \boldsymbol{q} = \gamma \boldsymbol{R}_D \boldsymbol{q}. \tag{9}$$

The solution $\boldsymbol{q}_{\text{ACC}}$ is given by the eigenvector corresponding to the largest eigenvalue $\kappa^{-2} \gamma_{\max}$ from the above equation. Note that $\boldsymbol{R}_D$ must be positive definite to calculate the inversion. To ensure this, regularization is sometimes performed by adding a scaled identity matrix to $\boldsymbol{R}_D$. Also note that when $x[n]$ is a J-periodic signal with $N = J \geq K$ and γ is optimized per frequency bin rather than globally as in (9), it decouples into J independent generalized eigenvalue problems, whose solutions are identical to the solution of ACC in the frequency domain of J frequency bins. For more on the ACC time domain approaches see [20, 21, 23, 28].

2.2. PM

As alluded to earlier, ACC does not try to control the phase in the two zones since it only optimizes the acoustic contrast [29, 30]. To control the phase and the amplitude, a desired (or target) sound field can be defined which should be reproduced. PM, proposed by Poletti [4], minimizes the difference between the desired and reproduced sound fields in a least squares sense. PM was originally formulated in the frequency domain, but we here describe the core concept of PM in the time domain. Let $\boldsymbol{d}_B[n]$ be the desired sound field of the bright zone. For instance, $\boldsymbol{d}_B[n]$ can be the sound field generated by a virtual source or a loudspeaker array. One specific example will be shown in Section 4. By virtue of being the dark zone, the desired signal in the dark zone is the zero vector $\boldsymbol{0}_{M_D}$. Thus, the desired sound field for the total zone is

$$\boldsymbol{d}_C[n] = \begin{bmatrix} \boldsymbol{d}_B^T[n] & \boldsymbol{0}_{M_D}^T \end{bmatrix}^T \in \mathbb{R}^{(M_B+M_D) \times 1}. \tag{10}$$

The reproduction error, i.e., the difference, between the desired and reproduced sound fields can then be defined as

$$\boldsymbol{\varepsilon}_C[n] = \boldsymbol{d}_C[n] - \boldsymbol{p}_C[n] = \boldsymbol{d}_C[n] - \boldsymbol{H}_C^T[n] \boldsymbol{q}, \tag{11}$$

and the average reproduction error energy for N samples recorded at each microphone is defined by

$$\begin{aligned} S_C &= \frac{1}{N} \sum_{n=0}^{N-1} \|\boldsymbol{\varepsilon}_C[n]\|^2 = \frac{1}{N} \sum_{n=0}^{N-1} \left(\|\boldsymbol{\varepsilon}_B[n]\|^2 + \|\boldsymbol{\varepsilon}_D[n]\|^2 \right) \\ &= S_B + S_D, \end{aligned} \tag{12}$$

where $\|\cdot\|$ is the ℓ_2 norm operator, S_B, S_D are the average distortion energy and the average residual energy, respectively. The least squares estimate of $\boldsymbol{q}_{\text{PM}}$ which minimizes S_C will then be

$$\begin{aligned}\boldsymbol{q}_{\text{PM}} &= \left(\sum_{n=0}^{N-1} \boldsymbol{H}_C[n]\boldsymbol{H}_C^T[n]\right)^{-1} \sum_{n=0}^{N-1} \boldsymbol{H}_C[n]\boldsymbol{d}_C[n] \\ &= \boldsymbol{R}_C^{-1}\boldsymbol{r}_B = \big(\boldsymbol{R}_B + \boldsymbol{R}_D\big)^{-1}\boldsymbol{r}_B, \end{aligned} \tag{13}$$

where $\boldsymbol{r}_B = N^{-1}\sum_{n=0}^{N-1} \boldsymbol{H}_B[n]\boldsymbol{d}_B[n]$. In order to avoid having to solve a large linear system, PM is typically considered in the frequency domain. Setting $N = J \geq K$ and assuming $x[n]$ to be J periodic make the time domain PM identical to the frequency domain PM, and the linear system in (13) decouples into J independent linear problems. Note that a regularization parameter is often introduced in (13) in order to ensure the inversion of $\boldsymbol{R}_B + \boldsymbol{R}_D$ [4]. Another variation consists in scaling the correlation matrices in order to trade-off the fitting to the desired signal in the bright zone to the average residual energy in the dark zone [18, 20]. Thus, the solution is $\boldsymbol{q}_{\text{HY}} \propto \big(\zeta\boldsymbol{R}_D + (1-\zeta)\boldsymbol{R}_B\big)^{-1}\boldsymbol{r}_B$ where ζ is a weighting factor. Although not being a combination of ACC and PM, such trade-off methods are often referred to as hybrid methods.

3. A UNIFIED APPROACH

In this section, we show how $\boldsymbol{q}$ can be obtained in a general framework which has ACC and PM as two extreme special cases. The framework is inspired by a recently introduced framework in speech enhancement which is referred to as VSLF [25, 26]. To present this framework in the context of sound zone design, we initially consider the generalized eigenvalue problem encountered in ACC. The solution to the generalized eigenvalue problem in (9) is a diagonal matrix $\boldsymbol{\Lambda}_{LJ}$ containing the LJ eigenvalues in descending order, $\lambda_1 \geq \cdots \geq \lambda_{LJ}$, and the square matrix $\boldsymbol{U} = \begin{bmatrix}\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_{LJ}\end{bmatrix}$ containing the LJ eigenvectors ordered according to the eigenvalues. These eigenvectors jointly diagonalize the spatial correlation matrices as [31, 32]

$$\boldsymbol{U}^T\boldsymbol{R}_B\boldsymbol{U} = \boldsymbol{\Lambda}_{LJ}, \quad \boldsymbol{U}^T\boldsymbol{R}_D\boldsymbol{U} = \boldsymbol{I}_{LJ}. \tag{14}$$

In ACC, the solution $\boldsymbol{q}_{\text{ACC}}$ is proportional to the first eigenvector $\boldsymbol{u}_1$. In speech processing, such a filter is referred to as a maximum SNR filter since it maximizes the noise suppression at the filter output. However, the maximum SNR filter is known to distort the speech signal significantly since it seeks to attenuate the noisy signal in all frequency bins, except for the frequency with the maximum SNR. This problem is also observed in sound zone designs where the filters seek to maximize the contrast at just one frequency [21–23].

In speech enhancement, there is generally a trade-off between signal distortion and noise suppression. Thus, if we want a small signal distortion, we get a low noise suppression and vice versa. In sound zone design, we have a similar trade-off between the signal distortion and the acoustic contrast. Despite its name, the hybrid method does not directly allow us to make a trade-off between the signal distortion and the acoustic contrast. Instead, it trades-off S_B for S_D through the scalar ζ. In other words, the extreme cases of the hybrid method are neither ACC nor PM. The hybrid method forces $\boldsymbol{q}$ towards the zero vector to minimize S_D when $\zeta \to 1$. On the other hand, the hybrid method only seeks to minimize the signal distortion in the bright zone for $\zeta \to 0$, but does not control the dark zone. Finally, PM is for $\zeta = 0.5$ [18].

The hybrid method is also a special case of the VSLF framework, but the framework also allows us to make a trade-off between

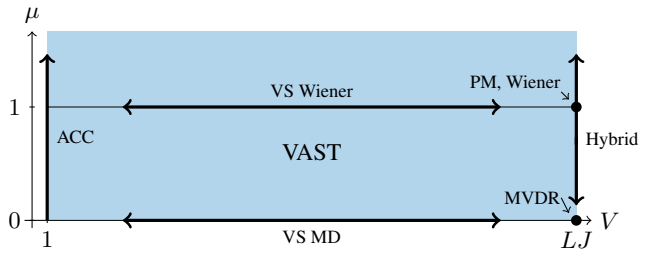

Fig. 2. An illustration of how the various special cases of the VAST solutions are related as a function of the user parameters V and μ.

S_B and γ. The main idea behind VSLF is to make a low rank approximation of the solution $\boldsymbol{q}$. Any vector can be written as a linear combination of basis functions as

$$\boldsymbol{q} = \boldsymbol{B}\boldsymbol{a}, \tag{15}$$

where $\boldsymbol{B} \in \mathbb{R}^{LJ\times LJ}$ and $\boldsymbol{a} \in \mathbb{R}^{LJ\times 1}$ contain the basis functions and the weights, respectively. To make an efficient V-rank approximation of $\boldsymbol{q}$, we use the first $1 \leq V \leq LJ$ eigenvectors in $\boldsymbol{U}$ as the basis functions. As in VSLF, the V-rank approximation $\boldsymbol{U}_V\boldsymbol{a}_V$ to $\boldsymbol{q}$ is then inserted in the models for the sound fields in the two zones, and we optimize over $\boldsymbol{a}_V$ instead of $\boldsymbol{q}$. Note that $\boldsymbol{U}_V$ and $\boldsymbol{a}_V$ are of size $LJ \times V$ and $V \times 1$, respectively. The rank V is a user-defined parameter, and, as we show later, it controls the trade-off between S_B and γ.

The low rank approximation in VSLF can be employed in many different problem settings [25, 26]. As alluded to earlier, the hybrid method in [18] can be viewed as the solution to one such problem setting which is given by

$$\text{minimize } S_B \text{ subject to } S_D \leq \epsilon, \tag{16}$$

for $V = LJ$ where $\epsilon \geq 0$ controls how important it is for the sound zone system to suppress S_D. For $1 \leq V \leq LJ$, the solution to (16) is generally referred to as a variable span trade-off (VAST) filter and given by

$$\boldsymbol{q}_{\text{VAST}}(V,\mu) = \boldsymbol{U}_V\boldsymbol{a}_V(\mu) = \sum_{v=1}^{V} \frac{\boldsymbol{u}_v\boldsymbol{u}_v^T}{\mu+\lambda_v}\boldsymbol{r}_B, \tag{17}$$

where μ is a Lagrange multiplier. Typically, this Lagrange multiplier is controlled directly instead of indirectly through ϵ. The trade-off filter has many important special cases depending on how V and μ are selected. First, consider the case where $V = 1$. This gives

$$\boldsymbol{q}_{\text{VAST}}(1,\mu) = \frac{\boldsymbol{u}_1\boldsymbol{u}_1^T}{\mu+\lambda_1}\boldsymbol{r}_B \propto \boldsymbol{u}_1, \tag{18}$$

which is clearly the ACC solution. Another interesting special case is the case for $V = LJ$ where we obtain

$$\begin{aligned}\boldsymbol{q}_{\text{VAST}}(LJ,\mu) &= \boldsymbol{U}\big(\boldsymbol{\Lambda}_{LJ} + \mu\boldsymbol{I}_{LJ}\big)^{-1}\boldsymbol{U}^T\boldsymbol{r}_B \\ &= \big(\boldsymbol{R}_B + \mu\boldsymbol{R}_D\big)^{-1}\boldsymbol{r}_B, \end{aligned} \tag{19}$$

which is essentially equivalent to the hybrid method for a general μ and the PM method for $\mu = 1$. Other special cases can be found in Table 1 and their relationship is illustrated in Fig. 2. The listed names are inspired by the names from the speech enhancement community.

Earlier, we stated that the rank V controls the trade-off between S_B and γ. To show this, consider the acoustic contrast first. As detailed in Sec. 2.1, it is

$$\gamma_V(\mu) = \kappa^2\frac{\boldsymbol{a}_V^T(\mu)\boldsymbol{\Lambda}_V\boldsymbol{a}_V(\mu)}{\boldsymbol{a}_V^T(\mu)\boldsymbol{a}_V(\mu)}, \tag{20}$$

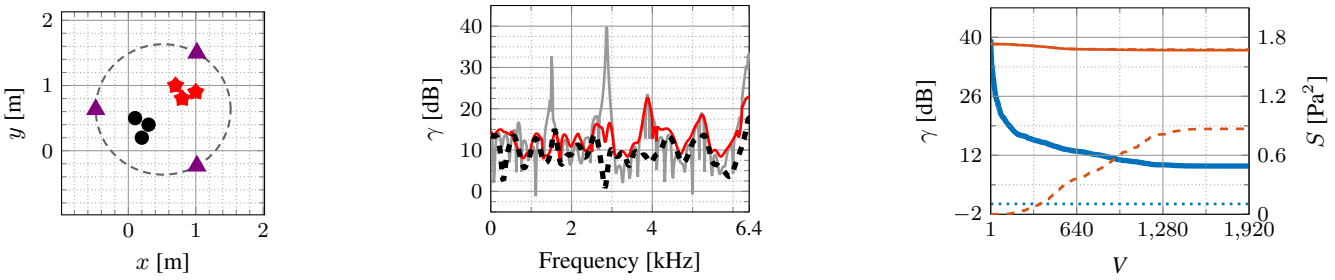

Fig. 3. (a) Geometry setup showing the loudspeaker positions (▲), the microphone positions in the bright (★) and dark (●) zones, (b) the acoustic contrast γ with respect to frequency when $V = 1$ which corresponds to ACC(—), $V = 800$ (—), and $V = 1920$ which corresponds to PM(▪▪), (c) the acoustic contrast before filtering(···), the acoustic contrast after filtering(—), S_B(—), S_D(- -), and S_C(-·-). Note that S_D is multiplied by 150 for better visualization.

where the equality follows by applying the joint diagonalization of $\boldsymbol{R}_B$ and $\boldsymbol{R}_D$. From this expression we see that $\gamma_V(\mu) \geq \gamma_{V'}(\mu)$ if $V \leq V'$. Consequently, we achieve the best and worst contrasts for $V = 1$ and $V = LJ$, respectively, for a fixed μ. Second, we consider $S_B(V), S_D(V)$ as a function of V when we insert the VAST solution $\boldsymbol{q}_{\text{VAST}}(V, \mu)$ such that

$$S_B(V) = \sigma_d^2 - \sum_{v=1}^{V} \frac{\lambda_v + 2\mu}{(\lambda_v + \mu)^2} \|\boldsymbol{u}_v^T \boldsymbol{r}_B\|^2 \tag{21}$$

$$S_D(V) = \sum_{v=1}^{V} \frac{1}{(\lambda_v + \mu)^2} \|\boldsymbol{u}_v^T \boldsymbol{r}_B\|^2, \tag{22}$$

where $\sigma_d^2 = N^{-1} \sum_{n=0}^{N-1} \|\boldsymbol{d}_B[n]\|^2$. These expressions are very interesting since they show that $S_B(V)$ decreases and $S_D(V)$ increases with increasing V since the eigenvalues are nonnegative. Thus, we get the minimum S_B for $V = LJ$, but the minimum S_D for $V = 1$. Moreover, we also see that the ACC solution ($V = 1$) gives the highest signal distortion. The V that minimizes $S_C(V)$ can in some cases be between these endpoints. Specifically since $S_C(V)$ is the sum of $S_B(V)$ and $S_D(V)$, we see from

$$S_C(V) = \sigma_d^2 - \sum_{v=1}^{V} \frac{\lambda_v + 2\mu - 1}{(\lambda_v + \mu)^2} \|\boldsymbol{u}_v^T \boldsymbol{r}_B\|^2 \tag{23}$$

that $S_C(V)$ will start increasing from the smallest value V satisfying $\lambda_v < 1 - 2\mu$. Thus, $S_C(V)$ never increases with increasing V if we set $\mu \geq 1/2$ since all eigenvalues are nonnegative.

4. SIMULATION

The contribution made in this paper is theoretical, so we here only include a proof-of-concept by considering a toy example. As illustrated in Fig. 3 (a), a circular array with three loudspeakers is considered, and three microphones for each zone are used. The loudspeakers and the microphones are all assumed to be in the same plane. We also assume that the setup is situated in the free-field, that all loudspeakers behave as point sources, and that the microphones are ideal. The length of the control filters $\{\boldsymbol{q}_l\}_{l=1}^3$ is $J = 640$, and the sampling frequency is 12.8 kHz. The desired sound field of the bright zone is set as the sound field generated by the loudspeaker array denoted as $\boldsymbol{d}_B[n] = \boldsymbol{H}_B^T[n]\boldsymbol{i}_J^{(L)}$, where $\boldsymbol{i}_J^{(L)} = \boldsymbol{1}_L \otimes \boldsymbol{i}_J$, $\boldsymbol{1}_L = [1 \cdots 1]^T \in \mathbb{R}^{L \times 1}$ and $\boldsymbol{i}_J$ is the first column of $\boldsymbol{I}_J$. Finally, $\mu = 1$ to show the performance of PM when $V = LJ$, the input signal $x[n]$ is set as the Kronecker delta function.

As a performance measure, we use the acoustic contrast as a function of frequency. Specifically, we have used $\gamma[k] = \kappa^2 \sum_{m=1}^{M_B} \|\text{DFT}(p_m[n])\|^2 / \sum_{m=1}^{M_D} \|\text{DFT}(p_m[n])\|^2$ where $k \in \{0, \cdots, N-1\}$ represents the frequency index and DFT is the discrete Fourier transform [33]. Note that γ is equal to the largest eigenvalue since $\kappa^2 = 1$. As depicted in Fig. 3 (b), ACC, which corresponds to $V = 1$, has high contrasts at a few frequencies. As V increases toward the PM solution, however, the contrast becomes flatter across frequency, thus minimizing S_B.

Table 1. Various solutions for sound zone control

μ	V	Form
–	–	$\boldsymbol{q}_{\text{VAST}} = \sum_{v=1}^{V} \left[(\mu + \lambda_v)^{-1} \boldsymbol{u}_v \boldsymbol{u}_v^T \boldsymbol{r}_B\right]$
–	1	$\boldsymbol{q}_{\text{ACC}} = (\mu + \lambda_1)^{-1} \boldsymbol{u}_1 \boldsymbol{u}_1^T \boldsymbol{r}_B$
0	–	$\boldsymbol{q}_{\text{VS MD}} = \sum_{v=1}^{V} \left[\lambda_v^{-1} \boldsymbol{u}_v \boldsymbol{u}_v^T \boldsymbol{r}_B\right]$
0	LJ	$\boldsymbol{q}_{\text{MVDR}} = \boldsymbol{R}_B^{-1} \boldsymbol{r}_B$
1	–	$\boldsymbol{q}_{\text{VS W}} = \sum_{v=1}^{V} \left[(1 + \lambda_v)^{-1} \boldsymbol{u}_v \boldsymbol{u}_v^T \boldsymbol{r}_B\right]$
1	LJ	$\boldsymbol{q}_{\text{PM}} = \left(\boldsymbol{R}_B + \boldsymbol{R}_D\right)^{-1} \boldsymbol{r}_B$

Fig. 3 (c) shows the acoustic contrast with and without control filters on the left y-axis. We clearly see that the acoustic contrast decreases as V increases with ACC having the highest contrast and PM the smallest. The right y-axis, on the other hand, shows the average distortion energy S_B, the average residual energy S_D, and the average reproduction error energy S_C. Despite being a toy problem, it clearly illustrates how V can be chosen to trade-off S_B for γ. The average signal distortion energy S_B and the average reproduction error energy S_C decrease with V. Finally, the average residual energy S_D, which has been amplified by 150 in the figure, increases with V.

5. CONCLUSION

We have proposed a unified approach to generating sound zones in the time domain. This has been done by adopting a recent speech enhancement technique called variable span linear filtering. It is based on making a joint diagonalization of the spatial correlation matrices and has ACC and PM as extreme special cases corresponding to using the dominant or all eigenvectors. We also used the framework to show that ACC gives the best contrast, but the highest distortion, and that PM gives the smallest distortion, but the worst contrast. Finally, all solutions formed by using only a subset of eigenvectors will have an acoustic contrast and an average distortion energy which are upper and lower bounded by ACC and PM, respectively.

6. REFERENCES

[1] W. F. Druyvesteyn and J. Garas, "Personal sound," *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 685–701, 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. Am.*, vol. 111, no. 4, pp. 1695–1700, 2002.

[3] 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.

[4] M. A. Poletti, "An investigation of 2d multizone surround sound systems," in *Proc. 125th Conv. Audio. Eng. Soc.*, San Francisco, USA, 2008.

[5] 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.

[6] 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.*, 2010.

[7] J. Cheer, S. J. Elliott, and M. F. Simón-Gálvez, "Design and implementation of a car cabin personal audio system," *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 412–424, 2013.

[8] M. F. Simón-Gálvez, S. J. Elliott, and J. Cheer, "Personal audio loudspeaker array as a complementary tv sound system for the hard of hearing," *IEICE Trans.*, vol. E97-A, no. 9, pp. 1824–1831, 2014.

[9] 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, 2016.

[10] 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, 2011.

[11] W. Jin, W. B. Kleijn, and D. Virette, "Multizone soundfield reproduction using orthogonal basis expansion," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, may 2013, pp. 311–315.

[12] 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, 2016.

[13] 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.

[14] P. Coleman, P. J. B. Jackson, M. Olik, and J. A. Pedersen, "Personal audio with a planar bright zone," *J. Acoust. Soc. Am.*, vol. 136, no. 4, pp. 1725–1735, Oct. 2014.

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

[16] 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, 2015.

[17] T. Okamoto and A. Sakaguchi, "Experimental validation of spatial fourier transform-based multiple sound zone generation with a linear loudspeaker array," *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 1769–1780, Oct. 2017.

[18] 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.

[19] 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, 2012.

[20] 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.

[21] 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.

[22] M. B. Møller and M. Olsen, "Sound zones: On performance prediction of contrast control methods," in *AES Int. Conf. Sound Field Control*, 2016.

[23] D. H. 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. F. Simón-Gálvez, S. J. Elliott, and J. Cheer, "A superdirective array of phase shift sources," *J. Acoust. Soc. Am.*, vol. 132, no. 2, pp. 746–756, 2012.

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

[26] J. Benesty, M. G. Christensen, and J. R. Jensen, *Signal enhancement with Variable Span Linear Filters*. Springer, 2016, vol. 7.

[27] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

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

[29] F. Jacobsen, M. Olsen, M. B. Møller, and F. Agerkvist, "A comparison of two strategies for generating sound zones in a room," in *Proc. 18th Int. Congr. Sound Vib.*, Rio de Janeiro, Brazil, 2011.

[30] P. Coleman, P. Jackson, M. Olik, and J. A. Pedersen, "Optimizing the planarity of sound zones," in *Proc. 52nd Int. Conf. Audio. Eng. Soc.*, Guildford, UK, 2013.

[31] G. H. Golub and C. F. Van Loan, *Matrix computations*. The Johns Hopkins University Press, 1996.

[32] J. H. Wilkinson, *The algebraic eigenvalue problem*. Clarendon Press Oxford, 1965.

[33] A. H. Andersen, "On multiple sound zones for wideband signals," Master's Thesis, Aalborg University, 2014.