GENERATION OF PERSONAL SOUND FIELDS IN REVERBERANT ENVIRONMENTS USING INTERFRAME CORRELATION

Liming Shi,[1] *Guoli Ping,*[2] *Xiaoxiang Shen,*[2] *Mads Græsbøll Christensen*[1]

[1]Audio Analysis Lab, CREATE, Aalborg University, {ls, mgc}@create.aau.dk
[2]Acoustic Engineering Lab, Huawei Technologies Co., Ltd, {pingguoli, shenxiaoxiang3}@huawei.com

ABSTRACT

Personal sound field control techniques aim to produce sound fields for different sound contents in different places of an acoustic space without interference. The limitations of the state-of-the-art methods for sound field control include high latency and computational complexity, especially in the cases when the reverberation time is long and number of loudspeakers is large. In this paper, we propose a personal sound field control approach that exploits interframe correlation. Considering the past frames, the proposed method can accommodate long reverberation time with a low latency. To find the optimal parameters for the physical meaningful constraints, the subspace decomposition and Newton's method are applied. Furthermore, a sound field distortion oriented subspace construction method is proposed to reduce the subspace dimension. Compared with traditional methods, simulation results show that the proposed algorithm is able to obtain a good trade-off between acoustic contrast and reproduction error with a low latency for measured room impulse responses.

Index Terms— Sound field control, sound field reproduction, reverberant environments, interframe correlation, loudspeaker arrays

1. INTRODUCTION

Personal sound fields [1] has many and diverse applications in mobile devices, car cabins, and home entertainment [2–5]. The objective of sound field control techniques is to create multiple sound zones for different sound contents in an acoustic environment using a loudspeaker array. The spatial areas for listeners to enjoy one specific sound content is referred to as bright zones, and the rest of the control areas are referred to as dark zones for this content. Therefore, the definitions of "bright" and "dark" are relative and are interchangeable for different sound contents.

Sound field control techniques can be broadly classified into a number of categories, as described next. Acoustic contrast control (ACC) [2,6–9] aims to maximize the energy ratio between the bright and dark zones, neglecting the sound quality in the bright zones [6]. To improve the robustness of ACC, a regularized method is proposed in [7]. To deal with the causality problem of the frequency-domain ACC, the time-domain ACC method has been proposed and studied [8, 9]. Pressure matching (PM) [10–16] works by defining a desired sound field in the zones for which the optimal filter is obtained by minimizing the mean squared error, leading to a low reproduction error (RE) but also a low acoustic contrast (AC). The robustness of the PM, which can be improved using the regularization method, is studied in [11]. An AC constrained PM method is proposed in [12]. Meanwhile, the modal-domain and time-domain based PM algorithms have also been proposed [15–17], and methods combining the advantages of the ACC and PM are proposed in [18, 19]. Recently, motivated by the variable span trade-off (VAST) filter for speech enhancement [20, 21], the frequency-domain and time-domain subspace-based sound field control methods using generalized eigenvalue decomposition (GEVD) have been proposed [22, 23]. The subspace-based methods have the advantage of enabling an explicit trade-off between the AC and RE by adjusting the subspace dimension. The main limitations of the state-of-the-art sound field control methods are the high latency and computational complexity, especially in cases when the reverberation time of the room is long and the number of loudspeakers is large, since the standard way of dealing with reverberation is to use a long window length.

In this paper, we propose a sound field control method in the short time Fourier transform (STFT)-domain that exploits interframe correlation. By considering the interframe correlation, the proposed sound field control method can accommodate to reverberant acoustic environments without introducing additional latency. The optimal control filter is obtained by solving inequality constrained convex optimization problems with physically meaningful parameter settings. The subspace decomposition and Newton's method are applied to obtain the optimal Lagrange multipliers, and a sound field distortion oriented subspace construction method is proposed to reduce the subspace dimension, and thus the computational complexity.

2. FUNDAMENTALS

In this section, we formulate the sound field generation problem in the STFT domain and review the variable span trade-off filter that form the foundation of the proposed method.

2.1. Background

We consider the problem of generating a bright zone B and a dark zone D in an enclosed space with known (i.e., measured) room impulse responses (RIRs) using L loudspeakers and L finite impulse response (FIR) filters (a.k.a., control filters). The proposed method operates in the frequency domain with $x_{t,k}$ denoting the input sound signal at the t^{th} frame and k^{th} frequency bin. The reproduced sound pressure at the $m^{\text{th}}, 1 \leq m \leq M^{\text{C}}$ control point in one of the sound zones can be written as

$$\begin{aligned} y_{t,k}^{m,\text{C}} &= x_{t,k} \sum_{l=1}^{L} q_k^l h_{m,k}^{l,\text{C}} \\ &= x_{t,k} (\mathbf{h}_{m,k}^{\text{C}})^T \mathbf{q}_k. \end{aligned} \tag{1}$$

where the superscript $(\cdot)^{\text{C}}, \text{C} \in \{\text{B}, \text{D}\}$ is the zone index, q_k^l denotes the control filter for the l^{th} loudspeaker, $h_{m,k}^{l,\text{C}}$ denotes the acoustic

transfer function (ATF) from the l^{th} loudspeaker to the m^{th} control point in zone C, $\mathbf{h}_{m,k} = [h_{m,k}^{1,\text{C}}, \cdots, h_{m,k}^{L,\text{C}}]^T$, $\mathbf{q}_k = [q_k^1, \cdots, q_k^L]^T$. For notational simplicity, we will leave out superscript C unless it is required. Using (1) and collecting the signals from all the M control points, we have

$$\mathbf{y}_{t,k} = x_{t,k}\mathbf{H}_k\mathbf{q}_k, \tag{2}$$

where $\mathbf{H}_k = [\mathbf{h}_{1,k}, \cdots, \mathbf{h}_{M,k}]^T$. Assuming the desired sound pressure at the m^{th} control point in the bright zone is generated by a virtual source and can be written as $d_{t,k}^m = x_{t,k}\ddot{h}_{m,k}$, where $\ddot{h}_{m,k}$ denotes the ATF from the virtual source to the m^{th} control point at the k^{th} frequency bin. The desired sound field from all the M control points in the bright zone can be expressed as

$$\mathbf{d}_{t,k} = x_{t,k}\ddot{\mathbf{h}}_k, \tag{3}$$

where $\ddot{\mathbf{h}}_k = [\ddot{h}_{1,k}, \cdots, \ddot{h}_{M,k}]^T$. For simplicity, we will leave out the subscript k when it is clear from the context.

The optimal control filter coefficients in the STFT-domain can be found by solving the following convex optimization problem, i.e.,

$$\begin{aligned} &\underset{\mathbf{q}}{\text{minimize}}\ \frac{1}{M^{\text{B}}T}\sum_{t=1}^{T}\|\mathbf{d}_t - \mathbf{y}_t^{\text{B}}\|^2 \\ &\text{s.t.}\quad \frac{1}{M^{\text{D}}T}\sum_{t=1}^{T}\|\mathbf{y}_t^{\text{D}}\|^2 \le \epsilon, \quad \|\mathbf{q}\|^2 \le \delta, \end{aligned} \tag{4}$$

where $\|\mathbf{q}\|^2$ is usually referred to as array effort [11], T denotes the total number of frames, ϵ and δ are user-defined parameters. Substituting (2) and (3) into (4), the constrained optimization problem can be further written as

$$\begin{aligned} &\underset{\mathbf{q}}{\text{minimize}}\ \mathbf{q}^H\mathbf{R}^{\text{B}}\mathbf{q} - \mathbf{q}^H\mathbf{r} - \mathbf{r}^H\mathbf{q} + \kappa \\ &\text{s.t.}\quad \mathbf{q}^H\mathbf{R}^{\text{D}}\mathbf{q} \le \epsilon / \left\langle |x|^2 \right\rangle, \quad \|\mathbf{q}\|^2 \le \delta, \end{aligned} \tag{5}$$

where $\mathbf{R}^{\text{B}}$ denotes the spatial autocorrelation matrix and is defined as $\mathbf{R}^{\text{B}} = \frac{1}{M^{\text{B}}}(\mathbf{H}^{\text{B}})^H\mathbf{H}^{\text{B}}$, $\left\langle |x|^2 \right\rangle = \sum_{t=1}^{T}\frac{1}{T}|x_t|^2$, $\mathbf{R}^{\text{D}} = \frac{1}{M^{\text{D}}}(\mathbf{H}^{\text{D}})^H\mathbf{H}^{\text{D}}$, $\mathbf{r} = \frac{1}{M^{\text{B}}}(\mathbf{H}^{\text{B}})^H\ddot{\mathbf{h}}$ denotes the spatial cross-correlation vector, and $\kappa = \frac{1}{M^{\text{B}}}\ddot{\mathbf{h}}^H\ddot{\mathbf{h}}$ is a constant. The Lagrangian function of (5) can be written as

$$\mathcal{L}(\mathbf{q}) = \mathbf{q}^H(\mathbf{R}^{\text{B}} + \mu\mathbf{R}^{\text{D}} + \theta\mathbf{I})\mathbf{q} - \mathbf{q}^H\mathbf{r} - \mathbf{r}^H\mathbf{q} + \text{C}, \tag{6}$$

where $\text{C} = \kappa - \mu\frac{\epsilon}{\left\langle |x|^2 \right\rangle} - \theta\delta$, μ and θ are the Lagrange multipliers. The minimizer of (6) w.r.t. $\mathbf{q}$ can be expressed as

$$\mathbf{q} = (\mathbf{R}^{\text{B}} + \mu\mathbf{R}^{\text{D}} + \theta\mathbf{I})^{-1}\mathbf{r}_k. \tag{7}$$

The Lagrange multipliers $\mu \ge 0$ and $\theta \ge 0$ can be found using the interior-point method [24].

2.2. Subspace-based sound field control

Recently, subspace-based methods using GEVD for generation of sound fields have been proposed [22, 23], a.k.a., VAST filter. Let $\mathcal{K}_V$ be a V-dimensional subspace of $\mathcal{C}^L$ ($V \le L$) with a basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_V\}$, and form a rank-$V$ basis matrix $\mathbf{U}_V = [\mathbf{u}_1, \cdots, \mathbf{u}_V]$. We assume that the control filter $\mathbf{q}$ is a vector in $\mathcal{K}_V$, i.e., $\mathbf{q} = \mathbf{U}_V\mathbf{z}$ and $\mathbf{z} \in \mathcal{C}^V$. The essence of the VAST [20] approach is to form the V-dimensional subspace using V eigenvectors corresponding to the V largest eigenvalues. The rank-V matrix $\mathbf{U}_V$ can jointly diagonalize both the matrices $\mathbf{R}^{\text{B}}$ and $\mathbf{R}^{\text{D}}$, i.e.,

$$\mathbf{U}_V^H\mathbf{R}^{\text{B}}\mathbf{U}_V = \mathbf{\Lambda}_V, \quad \mathbf{U}_V^H\mathbf{R}^{\text{D}}\mathbf{U}_V = \mathbf{I}_V, \tag{8}$$

where $\mathbf{\Lambda}_V = \text{diag}(\lambda_1, \cdots, \lambda_V)$ is a diagonal matrix whose diagonal elements are eigenvalues sorted as $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_V$. Substituting (8) and $\mathbf{q} = \mathbf{U}_V\mathbf{z}$ into (5), the optimization problem can be expressed as

$$\begin{aligned} &\underset{\mathbf{z}}{\text{minimize}}\ \mathbf{z}^H\mathbf{\Lambda}_V\mathbf{z} - \mathbf{z}^H\mathbf{c} - \mathbf{c}^H\mathbf{z} + \kappa \\ &\text{s.t.}\quad \mathbf{z}^H\mathbf{z} \le \epsilon / \left\langle |x|^2 \right\rangle, \quad \mathbf{z}^H(\mathbf{U}_V^H\mathbf{U}_V)\mathbf{z} \le \delta, \end{aligned} \tag{9}$$

where $\mathbf{c} = \mathbf{U}_V^H\mathbf{r}$. The minimizer of the corresponding Lagrangian function w.r.t. $\mathbf{z}$ can be expressed as

$$\mathbf{z} = (\mathbf{\Lambda}_V + \mu\mathbf{I}_V + \theta\mathbf{U}_V^H\mathbf{U}_V)^{-1}\mathbf{c}. \tag{10}$$

In the VAST approach for sound field control, a fixed μ is used and θ is set to zero. The VAST approach offers a trade-off between acoustic contrast and reproduction error by adjusting the subspace dimension V. It can be readily shown that when $V = 1$, the VAST approach reduces to the ACC method [6] leading to the largest acoustic contrast, defined as

$$\xi_{\text{ac}} = \frac{\mathbf{q}^H\mathbf{R}^{\text{B}}\mathbf{q}}{\mathbf{q}^H\mathbf{R}^{\text{D}}\mathbf{q}}. \tag{11}$$

When $V = L$, $\mu = 1$, $\theta = 0$, the VAST approach reduces to the PM method [10].

There are two main drawbacks of the VAST method for sound field control. First, when applied in a room with long reverberation time (e.g., RT60 $\ge$ 300ms), the VAST method has a long delay. Second, the dual variables μ and θ are not physically meaningful and difficult to tune in practice.

3. PROPOSED METHOD

In this section, we present the proposed sound field control method based on the subspace method and interframe correlation in the STFT-domain. The interframe correlation method is used to reduce the delay and accommodate long reverberation time. The subspace-based method and Newton's method are used for finding the optimal Lagrange multipliers efficiently.

3.1. The optimal control filter with interframe correlation

Instead of directly forming the reproduced sound in the STFT-domain like (1) [22], the input signal is first filtered with the RIRs in the time-domain, and the filtered signal can be written as $B_n^{l,m} = X_n * H_n^{l,m}$, where $*$ denotes the linear convolution operator. The filtered input signal is then transformed to the frequency domain using the STFT and let $b_{t,k}^{l,m}$ denote the corresponding frequency-domain signal at the k^{th} frequency bin. Then, using the past P frames, the signal at the mth control point can be expressed as

$$\begin{aligned} y_{t,k}^m &= \sum_{l=1}^{L}\sum_{p=1}^{P}(b_{t-p+1,k}^{l,m})^T q_{p,k}^l \\ &= (\mathbf{b}_{t,k}^m)^T\mathbf{q}_k, \end{aligned} \tag{12}$$

where $\mathbf{b}_{t,k}^m = [(\mathbf{b}_{t,k}^{1,m})^T, \cdots, (\mathbf{b}_{t,k}^{L,m})^T]^T$ and $\mathbf{b}_{t,k}^{l,m} = [b_{t,k}^{1,m}, \cdots, b_{t-P+1,k}^{1,m}]^T$. Collecting the signals from all the M control points, we obtain

$$\mathbf{y}_t = \mathbf{B}_{t,k}\mathbf{q}_k, \tag{13}$$

where $\mathbf{B}_{t,k} = [\mathbf{b}_{t,k}^1, \cdots, \mathbf{b}_{t,k}^M]^T$. Substituting (13) into (4) and following the derivations from the previous section, the optimal filter can be expressed as

$$\mathbf{q}_{\text{h}} = (\mathbf{R}_{\text{h}}^{\text{B}} + \mu\mathbf{R}_{\text{h}}^{\text{D}} + \theta\mathbf{I})^{-1}\mathbf{r}_{\text{h}}, \tag{14}$$

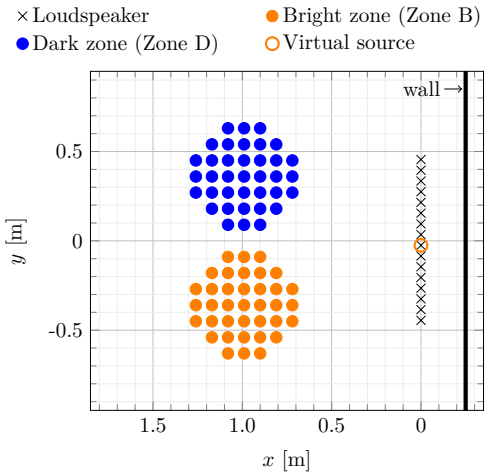

Fig. 1. Overview of the system setup for sound field control with 16 loudspeakers, 37 control points in each zone and the virtual source.

where $\mathbf{R}_{\text{h}} = \frac{1}{MT}\sum_{t=1}^{T}\mathbf{B}_t^H\mathbf{B}_t$ and $\mathbf{r}_{\text{h}} = \frac{1}{MT}\sum_{t=1}^{T}\mathbf{B}_t^H\mathbf{d}_t$. Assuming the frame shift is o seconds, to accommodate long reverberation time, $o \times P$ should be comparable to the RT60 of the acoustic environment, which will be verified by simulations. By using the information from the past frames, the time delay of the sound field control system is reduced by a factor of P. Thus, the proposed method has potentials in adaptive sound field control system. However, the dimension of $\mathbf{R}_{\text{h}}^{\text{B}}$ becomes $(LP) \times (LP)$, and thus the computational complexity for computing the inversion in (14) is $\mathcal{O}((LP)^3)$. When broadband processing is required, the high computational complexity issue becomes more severe, since we need to solve the optimal control filter for multiple frequency bins jointly [25]. The interior-point method [24] is often used for solving such inequality-constrained convex optimization problems. However, it can be computationally complex, especially for high-dimensional problems. In what follows, we present a low computational complexity method for solving the constrained optimization problem (4) with the proposed interframe correlation method based on the subspace method.

3.2. Finding the optimal Lagrange multipliers

Following the VAST approach, we first perform GEVD for each frequency bin index k in the range $[k_1, k_F]$ independently and build a V_k-dimensional ($V_k \leq LP$) basis matrix $\mathbf{U}_{V_k}$ using the eigenvectors, i.e.,

$$\mathbf{U}_{V_k}^H\mathbf{R}_h^{\text{B}}\mathbf{U}_{V_k} = \mathbf{\Lambda}_{V_k}, \mathbf{U}_{V_k}^H\mathbf{R}_h^{\text{D}}\mathbf{U}_{V_k} = \mathbf{I}_{V_k}. \tag{15}$$

Note that the basis matrix $\mathbf{U}_{V_k}$ can be built using any combination of the V_k eigenvectors. In the next subsection, we will discuss how to determine the number of eigenvectors and how to combine them for each frequency bin, which is referred to as subspace selection. Let $\mathbf{q}_{h,k} = \mathbf{U}_{V_k}\mathbf{z}_{h,k}$, follow the derivation to obtain (10), and the minimizer can be written as

$$\mathbf{z}_{h,k} = \mathbf{A}_k^{-1}\mathbf{c}_{h,k}, \tag{16}$$

where $\mathbf{A}_k = (\mathbf{\Lambda}_{V_k} + \mu\mathbf{I}_{V_k} + \theta\mathbf{U}_{V_k}^H\mathbf{U}_{V_k})$ and $\mathbf{c}_{h,k} = \mathbf{U}_{V_k}^H\mathbf{r}_{h,k}$. Distinct from the VAST filter, we intend to obtain the optimal Lagrange multipliers μ and θ. Using (15) and (16), the first and second constraints in (4) can be written as

$$\begin{aligned}\mathbf{c}_{h,k}^H\mathbf{A}_k^{-2}\mathbf{c}_{h,k} &\leq \epsilon \\ \mathbf{c}_{h,k}^H\mathbf{A}_k^{-1}(\mathbf{U}_{V_k}^H\mathbf{U})\mathbf{A}_k^{-1}\mathbf{c}_{h,k} &\leq \delta.\end{aligned} \tag{17}$$

The partial derivatives of (17) w.r.t. μ and θ can be easily obtained. Based on the complementary slackness condition [26], we can consider all four possible cases, use the Newton's method for each case to obtain μ and θ, and choose the one that leads to the lowest objective in (4). Due to the inversion operator in (17), the computational complexity of the Newton's method depends heavily on the subspace dimension V_k. In the VAST approach, the subspace is constructed by using the eigenvectors corresponding to the largest eigenvalues (i.e., acoustic contrast oriented). The subspace dimension is determined by considering both the acoustic contrast and reproduction error. In the next subsection, we propose a sound field distortion oriented subspace construction method.

3.3. Sound field distortion oriented subspace construction

The normalized sound field distortion (NSFD) is defined as

$$\xi_{\text{nsfd}} = \frac{\frac{1}{M^{\text{B}}T}\sum_{t=1}^{T}\|\mathbf{d}_t - \mathbf{y}_t^{\text{B}}\|^2}{\frac{1}{M^{\text{B}}T}\sum_{t=1}^{T}\|\mathbf{d}_t\|^2}. \tag{18}$$

Substituting (16) and $\mathbf{q}_{h,k} = \mathbf{U}_{V_k}\mathbf{z}_{h,k}$ into the MSE-based objective function in (4), and assuming $\theta = 0$ (i.e., the array effort constraint is inactive), we can obtain $\xi_{\text{nsfd}}(V_k) = \frac{\kappa - \sum_{v=1}^{V_k}\frac{\lambda_v}{(\lambda_v+\mu)^2}c_{h,k}^v}{\kappa}$, where $c_{h,k}^v$ is the v^{th} element in $\mathbf{c}_{h,k}$. Therefore, in terms of minimizing the NSFD, with a fixed μ, the term $\frac{\lambda_v}{(\lambda_v+\mu)^2}c_{h,k}^v$ with the largest value leads to the fastest decrease of the NSFD. Motivated by

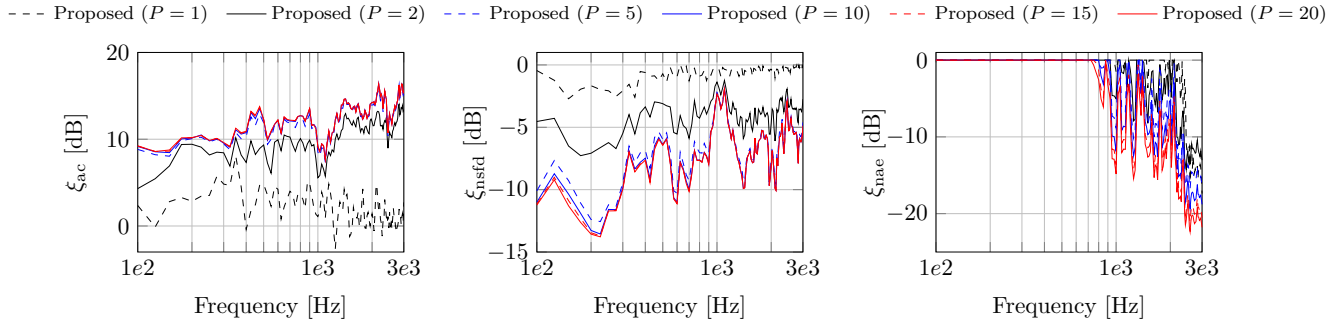

Fig. 2. The AC, NSFD and NAE performance of the proposed algorithm with different number of frames P.

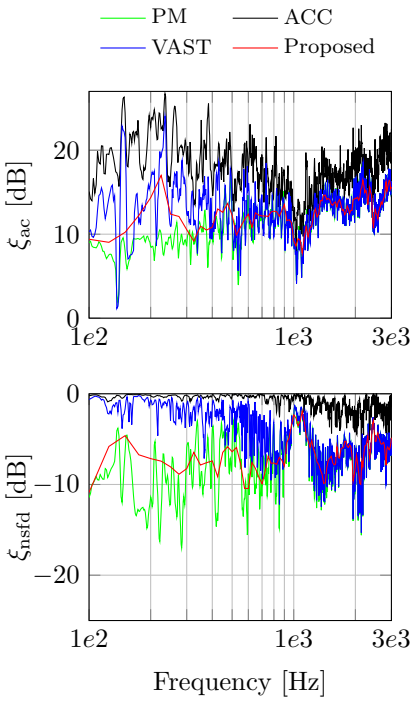

Fig. 3. The AC and NSFD performance of the PM, ACC, VAST and proposed method. The frame shift is set to 300 ms for PM, ACC and VAST. The frame shift for the proposed is set to 20 ms.

this fact, we propose the following subspace construction method: 1) sort $\frac{\lambda_v}{(\lambda_v+\mu)^2}c_{h,k}^v$ for $1 \leq v \leq LP$, and form the basis matrix $\mathbf{U}_{V_k}$ using the eigenvectors corresponding to the largest $\frac{\lambda_v}{(\lambda_v+\mu)^2}c_{h,k}^v$; 2) determine the subspace dimension V_k by using the smallest V leading to $\xi_{\text{nsfd}}(V) \leq (1+\alpha)\xi_{\text{nsfd}}(LP)$, where α is a positive constant. When α is large, the subspace dimension V_k is smaller, and thus the computational complexity is lower. However, the achievable NSFD is getting higher than the lower bound $\xi_{\text{nsfd}}(LP)$.

4. SIMULATIONS

4.1. Simulation setup

We consider a system that consists of a linear array with 16 evenly distributed loudspeakers in a room of dimension 4.5 m × 4.5 m × 2 m as shown in Fig. 1 [22, 27]. The inter-element spacing of the loudspeaker array is 6 cm. The number of control points in the bright and dark zones is set to 37. All the loudspeakers, control points and virtual source are located at a height of 1.2 m with a reverberation time RT60 = 300 ms. The sampling frequency is set to 16 kHz. Measured RIRs are used for the performance evaluation. The eighth loudspeaker is marked with an orange circle in Fig. 1, and is used to generate the ATFs of the virtual source. More specifically, the ATFs $\mathbf{h}_k$ of the virtual source for the bright zone are obtained using the ATFs from the eighth loudspeaker (i.e. eighth column of $\mathbf{H}_k$) with 30 ms delay to assure causality [15]. Defining the normalized array effort (NAE) as $\xi_{\text{nae}} = \frac{\mathbf{q}^H\mathbf{q}}{\gamma_0 P}$, where γ_0 is set to the driving power of using only the eighth loudspeaker to produce averaged 70 dB sound pressure level (SPL) in the bright zone [28]. The upper bound δ is chosen to assure ξ_{nae} is smaller than 0 dB [29]. The frame shift is set to 20 ms, and the overlap is set to 50%. A 30-second news signal with one male and one female speakers is used for the performance evaluation, and the frequency range considered is set to $[100, 3000]$ Hz.

4.2. Simulation results

In the first simulation, we test the performance of the proposed algorithm for different number of frames P, and the results are shown in Fig. 2. The upper bound for the averaged dark zone energy ϵ is computed by making the averaged SPL in the dark zone be 40 dB lower than the case when all 16 loudspeakers are driving in-phase (i.e., 40 dB energy reduction (ER)) [30]. The subspace construction scheme is disabled, i.e., $\alpha = 0$. In this case, the proposed method reduces to the PM with interframe correlation. As can be seen, the NAE is below 0 dB as expected. Also, when the number of frames P becomes larger, the AC increases, NSFD decreases and the NAE tends to become smaller. However, when the frame shift (20ms) times P is close to the reverberation time (i.e. 300 ms), the performance of the proposed algorithm becomes stable. Therefore, $P = 15$ is used for the following simulations.

In the second simulation, we test the performance of the proposed algorithm with different α for subspace construction. The averaged subspace dimensions V for $\alpha = 0$, $\alpha = 0.2$ and $\alpha = 0.4$ are LP, $0.54 \times LP$ and $0.44 \times LP$, respectively. The relationship between subspace dimension and computational complexity is studied in [30]. Due to the page limit, the AC, NSFD and NAE performance is studied but not shown. From our experience, setting $\alpha = 0.2$ leads to a low subspace dimension with little performance degradation.

In the last simulation, we compare the AC and NSFD performance of the PM, ACC, VAST and the proposed. For the proposed algorithm, we set $\alpha = 0.2$, $P = 15$, and the frame shift is 20 ms. The PM is obtained using the proposed algorithm with $\alpha = 0$, $P = 1$ and the frame shift is 300 ms. The VAST is obtained using the AC-oriented subspace construction method with the same subspace dimension as the proposed algorithm, $P = 1$ and the frame shift is 300 ms. The ϵ for the PM, VAST and the proposed method is set to the same as the first simulation. The simulation results are shown in Fig. 3. As can be seen, the ACC method has the largest AC, but the NSFD is also the highest. Second, the PM has the lowest NSFD, but the AC is the lowest. Furthermore, the performance of the VAST and proposed is between the ACC and PM. Although the VAST method has 1.5 dB higher AC on average than the proposed from 100–700 Hz, the proposed method has an averagely -5 dB lower NSFD than the VAST. Therefore, the proposed method can deliver a good trade-off between AC, NSFD and computational complexity with a low latency.

5. CONCLUSION

In this paper, an STFT-domain based sound field control approach that uses interframe correlation has been proposed. In the proposed method, the window length of the STFT does not have to be chosen in accordance with reverberation time, unlike in state-of-the-art methods, and thus the proposed method reduces the latency of the algorithm for acoustic environments with long reverberation time. To obtain the optimal Lagrange multipliers, subspace-based and Newton's method are presented. Furthermore, a sound field distortion oriented subspace construction method is proposed to reduce the subspace dimension. The proposed method offers a good trade-off between acoustic contrast and reproduction error, and it features a low latency and computational complexity.

6. 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.-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.

[3] J.-H. Chang and W.-H. Cho, "Evaluation of independent sound zones in a car," in *Proc. 23rd Int. Congr. Acoust.*, Aachen, Germany, Sep. 2019.

[4] M. F. S. Gálvez, S. J. Elliott, and J. Cheer, "Personal audio loudspeaker array as a complementary tv sound system for the hard of hearing," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 97, no. 9, pp. 1824–1831, 2014.

[5] L. Vindrola, M. Melon, J.-C. Chamard, and B. Gazengel, "Use of the filtered-x least-mean-squares algorithm to adapt personal sound zones in a car cabin," *J. Acoust. Soc. Am.*, vol. 150, no. 3, pp. 1779–1793, 2021.

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

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

[8] 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. 252–257, 2014.

[9] M. Hu and J. Lu, "Theoretical explanation of uneven frequency response of time-domain acoustic contrast control method," *J. Acoust. Soc. Am.*, vol. 149, no. 6, pp. 4292–4297, 2021.

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

[11] P. Coleman, P. J. B. Jackson, M. Olik, M. B. Møller, M. Olsen, and J. A. Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 1929–1940, 2014.

[12] Y. Cai, M. Wu, and J. Yang, "Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint," *J. Acoust. Soc. Am.*, vol. 135, no. 2, pp. 734–741, 2014.

[13] N. Ueno, S. Koyama, and H. Saruwatari, "Three-dimensional sound field reproduction based on weighted mode-matching method," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 27, no. 12, pp. 1852–1867, 2019.

[14] Q. Zhu, X. Qiu, P. Coleman, and I. Burnett, "A comparison between two modal domain methods for personal audio reproduction," *J. Acoust. Soc. Am.*, vol. 147, no. 1, pp. 161–173, 2020.

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

[16] J. Zhang, W. Zhang, T. D. Abhayapala, and L. Zhang, "2.5 d multizone reproduction using weighted mode matching: Performance analysis and experimental validation," *J. Acoust. Soc. Am.*, vol. 147, no. 3, pp. 1404–1417, 2020.

[17] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 19, no. 6, pp. 1711–1720, 2010.

[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. 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.

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

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

[22] 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, and Lang. Process.*, vol. 29, pp. 363–378, 2020.

[23] 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, and Lang. Process.*, vol. 29, pp. 823–837, 2021.

[24] 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.

[25] G. N. Lilis, D. Angelosante, and G. B. Giannakis, "Sound field reproduction using the lasso," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 18, no. 8, pp. 1902–1912, 2010.

[26] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[27] M. Schneider and E. A. Habets, "Iterative dft-domain inverse filter optimization using a weighted least-squares criterion," *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 27, no. 12, pp. 1957–1969, 2019.

[28] P. Coleman, P. J. Jackson, M. Olik, M. Møller, M. Olsen, and J. A. Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 1929–1940, 2014.

[29] M. Zhu and S. Zhao, "An iterative approach to optimize loudspeaker placement for multi-zone sound field reproduction," *J. Acoust. Soc. Am.*, vol. 149, no. 5, pp. 3462–3468, 2021.

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