An Alternating Mode Strategy for Adaptive Sound Field Control and Acoustic Path Tracking

Junqing Zhang*, Jingli Xie*, Dongyuan Shi*, Wen Zhang*, Jingdong Chen*, and Jacob Benesty†

*CIAIC and Shaanxi Provincial Key Laboratory of Artificial Intelligence,
Northwestern Polytechnical University, Xi'an, Shaanxi, China
†INRS-EMT, University of Quebec, Montreal, Canada

***Abstract*—Sound field control (SFC) aims to accurately reproduce a desired sound field within a specified region, which requires both adaptation to input signal characteristics and precise estimation of acoustic paths between loudspeakers and microphones. To meet these demands, two adaptive algorithms are proposed. The first is a signal-adaptive multichannel filtered-x least-mean-square (MCFxLMS) filter designed to handle nonstationary input signals such as speech and music. The second is an acoustic path tracking algorithm that incorporates an input signal decorrelation strategy, enabling robust tracking of multichannel room impulse responses (RIRs) even under highly correlated excitation. Additionally, an alternating mode-switching mechanism is introduced to dynamically activate each algorithm based on predefined criteria. This approach can reduce computational complexity in large-scale multichannel systems while preserving sound field fidelity within the control region.**

I. Introduction

Sound field control (SFC) is the process of reproducing a desired sound field within a specified region using an array of loudspeakers [1]–[4]. Traditional methods generally assume linear, time-invariant acoustic systems, and spectrally white input signals. However, in practice, SFC must contend with nonstationary inputs and dynamically changing acoustic environments. This study tackles these real-world challenges and proposes strategies to improve control performance under such conditions.

To address the nonstationary nature of input signals, adaptive SFC (ASFC) techniques have been developed. For instance, a method was proposed in [5], which incorporates time-averaged input statistics and perceptual masking to improve SFC performance [5]. Following this principle, the work in [6] proposed to incorporate the instantaneous spectral characteristics of the input signal to further enhance acoustic contrast [6]. In automotive environments, an online ASFC approach was proposed using the filtered-x least-mean-square (FxLMS) framework [7]. In multiple-input multiple-output (MIMO) systems, where computational complexity becomes a major concern, block-based adaptive algorithms [8], [9] and a distributed pressure-matching approach using diffusion adaptation [10] have been developed to efficiently manage processing demands across networked nodes.

This work was supported in part by the National Key Research and Development Program of China under Grant No. 2024YFF0505502, in part by the NSFC Grants: 62171373, 61831019, 62192713, and 62501477, and in part by the China Postdoctoral Science Foundation (CPSF) under Grants: GZC20242259 and 2025M774361. E-mail: junqingzhang@nwpu.edu.cn.

However, most existing methods fail to account for the time-varying nature of room acoustics, which can result from factors such as positional mismatches, temperature changes, scattering effects, and fluctuations in sound speed [11]. One approach to mitigating this issue involves introducing robustness constraints based on prior knowledge of the acoustic transfer function [12]–[14]. While effective under stable conditions, these methods depend heavily on accurate priors and thus struggle in rapidly changing environments. To overcome this limitation, system identification techniques have been explored as a more adaptive and flexible solution [15]–[17]. Drawing inspiration from secondary path modeling in active noise control (ANC), a joint iterative strategy has been proposed to simultaneously perform SFC and track variations in the room impulse responses (RIRs) [18]. Expanding on this idea, an adaptive control framework was later introduced to maintain acoustic contrast while dynamically tracking RIR changes [19]. These approaches typically require concurrent updates to both the RIR modeling filters and the SFC filters.

Conventional RIR tracking algorithms commonly rely on white or colored noise as excitation signals due to their favorable uncorrelated properties. However, such signals are impractical in real-world applications, as they introduce perceptually irrelevant and potentially disruptive content. In practice, excitation signals are more likely to be speech or music, which exhibit strong self and inter-channel correlations. This high degree of correlation leads to non-uniqueness issues, severely hindering both the convergence speed and accuracy of adaptive tracking algorithms.

To address the aforementioned challenges, a range of preprocessing techniques have been developed, which can be broadly classified into nonlinear [20], time-variant [21]–[24], and resampling-based methods [25]–[28]. One of most well-known approaches, proposed by Benesty et al., introduced nonlinear distortions to decorrelate input signals in stereo acoustic echo cancellation (AEC) [29]. The half-wave rectifier (HWR) is a well-known example that balances decorrelation performance with perceptual fidelity [29]. More recent work has shifted toward psychoacoustically motivated techniques, aiming to preserve perceptual quality while enhancing signal decorrelation [30], [31]. In the context of MIMO system identification, the non-uniqueness issue has been tackled by decomposing excitation signals into correlated and uncorrelated components. Two extended filters are then used to model these components independently, demonstrating improved identification performance over traditional methods [32], [33].

979-8-3315-7206-8/25/$31.00

To advance the current state-of-the-art, this paper introduces an alternating mode-switching strategy that dynamically switches between adaptive SFC and MIMO acoustic path tracking under correlated excitation conditions. The main contributions are threefold. First, an online SFC algorithm is introduced that effectively handles input signals both with and without intra- and inter-channel correlations. Second, a decorrelation technique leveraging filtered loudspeaker outputs is developed to enable robust tracking of multichannel acoustic systems. Third, a state-driven coordination mechanism is designed to manage the two adaptive processes, facilitating more efficient resource utilization in practical deployments. Extensive simulations are presented to demonstrate the feasibility and effectiveness of the proposed approach.

II. Sound Field Control

A. Signal Model

The reproduced sound pressure at the mth microphone position, denoted by $p_m(n)$, is modeled as the linear convolution of the input signal $x(n)$ with the RIRs and the loudspeaker control filters $q_l(j)$ of length J, at time index n [6]; that is

$$p_m(n) = \sum_{l=1}^{L} \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} h_{ml}(k) x(n-k-j) q_l(j), \quad (1)$$

where $h_{ml}(k)$ represents the RIR of length K from the lth loudspeaker to the mth microphone, and L denotes the total number of loudspeakers. The corresponding matrix/vector form of (1) is

$$p_m(n) = \sum_{l=1}^{L} \mathbf{h}_{m,l} \mathbf{X}(n) \mathbf{q}_l = \mathbf{h}_m^T \mathbf{U}(n) \mathbf{q}, \quad (2)$$

where

$$\begin{aligned}
\mathbf{h}_{ml} &= [h_{ml}(0) \;\; h_{ml}(1) \;\; \cdots \;\; h_{ml}(K-1)]^T \in \mathbb{R}^{K\times 1}, \\
\mathbf{h}_m &= \left[\mathbf{h}_{m1}^T \;\; \mathbf{h}_{m2}^T \;\; \cdots \;\; \mathbf{h}_{mL}^T\right]^T \in \mathbb{R}^{LK\times 1}, \\
[\mathbf{X}(n)]_{k,j} &= x(n-K-J+2), \;\; \mathbf{X}(n) \in \mathbb{R}^{K\times J}, \\
\mathbf{U}(n) &= \mathbf{I}_L \otimes \mathbf{X}(n) \in \mathbb{R}^{LK\times LJ}, \\
\mathbf{q}_l &= [q_l(0) \;\; q_l(1) \;\; \cdots \;\; q_l(J-1)]^T \in \mathbb{R}^{J\times 1}, \\
\mathbf{q} &= \left[\mathbf{q}_1^T \;\; \mathbf{q}_2^T \;\; \cdots \;\; \mathbf{q}_L^T\right]^T \in \mathbb{R}^{LJ\times 1},
\end{aligned}$$

the superscript $\{\cdot\}^T$ and $\otimes$ denote, respectively, the transpose and Kronecker product operators, and $\mathbf{I}_L$ denotes the identity matrix of size $L \times L$.

By stacking the sound pressures from all M microphones into a vector, we obtain

$$\mathbf{p}(n) = \mathbf{H}\mathbf{U}(n)\mathbf{q}, \quad (3)$$

where $\mathbf{p}(n) = [p_1(n) \;\; p_2(n) \;\; \cdots \;\; p_M(n)]^T \in \mathbb{R}^{M\times 1}$ is the vector of sound pressure, and $\mathbf{H} = [\mathbf{h}_1 \;\; \mathbf{h}_2 \;\; \cdots \;\; \mathbf{h}_M]^T \in \mathbb{R}^{M\times LK}$ is the RIR matrix.

The desired response at the mth microphone within the control region is typically derived from a virtual source, such as the vth loudspeaker, with a modeling delay τ_{d} introduced to ensure causality [34]. That is

$$\begin{aligned}
p_m^{\mathrm{d}}(n) &= x(n) * \delta(n-\tau_{\mathrm{d}}) * h_{mv}(k) \\
&= \mathbf{x}^T(n) \boldsymbol{\Delta}_m \mathbf{h}_{mv},
\end{aligned} \quad (4)$$

where

$$\begin{aligned}
\mathbf{x}(n) &= [x(n) \;\cdots\; x(n-K+1)]^T \in \mathbb{R}^{K\times 1}, \\
[\boldsymbol{\Delta}_m]_{i,j} &= \begin{cases} 1 & \text{if } i = j+\tau_{\mathrm{d}} \\ 0 & \text{otherwise} \end{cases}, \quad \boldsymbol{\Delta}_m \in \mathbb{R}^{K\times K}, \\
\mathbf{h}_{mv} &= [h_{mv}(0) \;\; h_{mv}(1) \;\; \cdots \;\; h_{mv}(K-1)]^T \in \mathbb{R}^{K\times 1},
\end{aligned}$$

and $*$ denotes the linear convolution operator. By organizing $p_m^{\mathrm{d}}(n)$ from all the M microphones into a vector, we obtain

$$\mathbf{p}^{\mathrm{d}}(n) = \mathbf{X}^{\mathrm{d}}(n) \boldsymbol{\Delta} \mathbf{h}_v \in \mathbb{R}^{M\times 1}, \quad (5)$$

where

$$\begin{aligned}
\mathbf{p}^{\mathrm{d}}(n) &= \left[p_1^{\mathrm{d}}(n) \;\; p_2^{\mathrm{d}}(n) \;\; \cdots \;\; p_M^{\mathrm{d}}(n)\right]^T \in \mathbb{R}^{M\times 1}, \\
\mathbf{X}^{\mathrm{d}}(n) &= \mathrm{blkdiag}\left[\mathbf{x}^T(n) \;\; \cdots \;\; \mathbf{x}^T(n)\right]^T \in \mathbb{R}^{M\times MK}, \\
\boldsymbol{\Delta} &= \mathrm{blkdiag}\left[\boldsymbol{\Delta}_1 \;\cdots\; \boldsymbol{\Delta}_M\right] \in \mathbb{R}^{MK\times MK}, \\
\mathbf{h}_v &= \left[\mathbf{h}_{1v}^T \;\; \mathbf{h}_{2v}^T \;\; \cdots \;\; \mathbf{h}_{Mv}^T\right]^T \in \mathbb{R}^{MK\times 1},
\end{aligned}$$

and $\mathrm{blkdiag}[\cdot]$ denotes the block diagonal matrix operator.

B. Adaptive Sound Field Control Algorithm

To adapt to the statistical variations of the input signals, a time-varying control filter $\mathbf{q}(n)$ is employed to replace the fixed filter $\mathbf{q}$ in (3). The resulting error between the desired and the reproduced sound field is then given by

$$\begin{aligned}
\mathbf{e}_{\mathrm{q}}(n) &= \mathbf{p}^{\mathrm{d}}(n) - \mathbf{p}(n) \\
&= \mathbf{X}^{\mathrm{d}}(n)\boldsymbol{\Delta}\mathbf{h}_v - \mathbf{H}\mathbf{U}(n)\mathbf{q}(n).
\end{aligned} \quad (6)$$

The corresponding mean-squared error (MSE) is defined as

$$\mathcal{J}[\mathbf{q}(n)] \triangleq \mathbb{E}[\mathbf{e}_{\mathrm{q}}^T(n)\mathbf{e}_{\mathrm{q}}(n)], \quad (7)$$

where $\mathbb{E}[\cdot]$ denotes the mathematical expectation. Replacing the expectation in (7) with instantaneous values gives the following gradient approximation

$$\begin{aligned}
\frac{\partial \mathcal{J}[\mathbf{q}(n)]}{\partial \mathbf{q}(n)} &\approx \left\{\mathbf{U}^T(n)\mathbf{H}^T\left[\mathbf{H}\mathbf{U}(n)\mathbf{q}(n) - \mathbf{X}^{\mathrm{d}}(n)\boldsymbol{\Delta}\mathbf{h}_v\right]\right\} \\
&= -\mathbf{U}^T(n)\mathbf{H}^T\mathbf{e}_{\mathrm{q}}(n).
\end{aligned} \quad (8)$$

Based on (8), the multichannel filtered-x LMS (MCFxLMS) algorithm for SFC can be expressed as

$$\mathbf{q}(n+1) = \mathbf{q}(n) + \frac{\mu_{\mathrm{q}}}{\|\mathbf{H}\mathbf{U}(n)\|_F^2 + \delta_{\mathrm{q}}} \mathbf{U}^T(n)\mathbf{H}^T\mathbf{e}_{\mathrm{q}}(n), \quad (9)$$

where μ_{q} is the step size parameter, δ_{q} is a positive constant introduced for numerical stability, and $\|\cdot\|_F$ denotes the Frobenius norm.

III. RIRs tracking for MIMO SFC system

The SFC algorithm described in Sec. II-B assumes static RIRs between sources and sensors. In practice, however, these RIRs can vary over time due to changes in the environment. To preserve effective control performance, it is therefore necessary to identify and track such variations throughout the control process.

A. RIR Tracking

As illustrated in Fig. 1, the SFC system can be formulated as a MIMO system. Accordingly, the RIR tracking system with

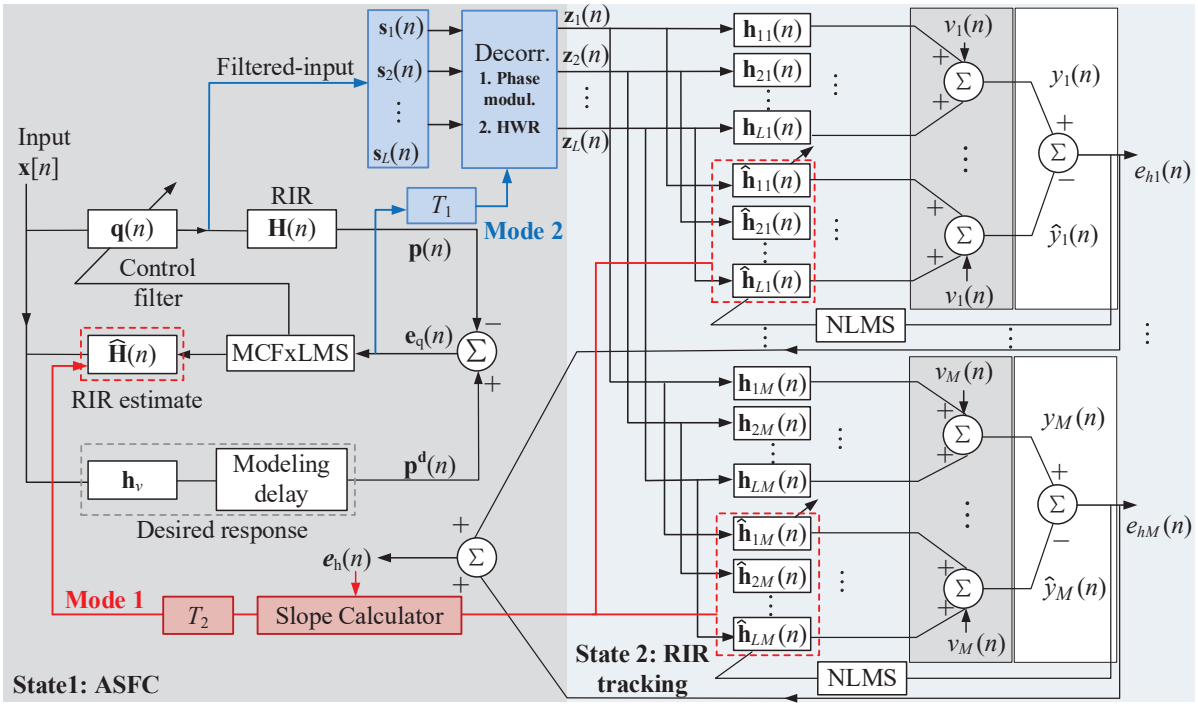

Fig. 1. Block diagram of the proposed alternating mode-switching method for adaptive sound field control and RIR tracking.

L inputs and M outputs is represented as an $L \times M$ system. The associated signal model follows the formulation in [35], [36], given by

$$\mathbf{y}(n) = \mathbf{H}(n)\mathbf{s}(n) + \mathbf{v}(n), \tag{10}$$

where

$$\mathbf{y}(n) = [y_1(n) \;\; y_2(n) \;\; \cdots \;\; y_M(n)]^T \in \mathbb{R}^{M\times 1},$$

$$\mathbf{H}(n) = \begin{bmatrix} \mathbf{h}_{11}^T(n) & \mathbf{h}_{12}^T(n) & \cdots & \mathbf{h}_{1L}^T(n) \\ \mathbf{h}_{21}^T(n) & \mathbf{h}_{22}^T(n) & \cdots & \mathbf{h}_{2L}^T(n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_{M1}^T(n) & \mathbf{h}_{M2}^T(n) & \cdots & \mathbf{h}_{ML}^T(n) \end{bmatrix} \in \mathbb{R}^{M\times LK},$$

$$\mathbf{s}(n) = \left[\mathbf{s}_1^T(n) \;\; \mathbf{s}_2^T(n) \;\; \cdots \;\; \mathbf{s}_L^T(n)\right]^T \in \mathbb{R}^{LK\times 1},$$

$$\mathbf{s}_l(n) = [s_l(n) \;\; s_l(n-1) \;\; \cdots \;\; s_l(n-K+1)]^T \in \mathbb{R}^{K\times 1},$$

$$\mathbf{v}(n) = [v_1(n) \;\; v_2(n) \;\; \cdots \;\; v_M(n)]^T \in \mathbb{R}^{M\times 1},$$

$s_l(n) = x(n) * q_l(n)$ is the control-filtered excitation signal[1], and $\mathbf{v}(n)$ is the observation noise vector.

At time n, the error signal vector encompassing all outputs is defined as

$$\begin{aligned} \mathbf{e}_\mathrm{h}(n) &= \mathbf{y}(n) - \hat{\mathbf{y}}(n) \\ &= \mathbf{y}(n) - \hat{\mathbf{H}}(n)\mathbf{s}(n), \end{aligned} \tag{11}$$

where $\mathbf{e}_\mathrm{h}(n) = [e_1(n) \;\; e_2(n) \;\; \cdots \;\; e_M(n)]^T \in \mathbb{R}^{M\times 1}$, and $\hat{\mathbf{H}}(n)$ is an estimate of $\mathbf{H}(n)$. The corresponding MSE is then written as

$$\mathcal{J}[\hat{\mathbf{H}}(n)] \triangleq \mathbb{E}\left[\mathbf{e}_\mathrm{h}^T(n)\mathbf{e}_\mathrm{h}(n)\right]. \tag{12}$$

It follows then that the derivative of (12) with respect to $\hat{\mathbf{H}}$ is given by

$$\frac{\partial \mathcal{J}[\hat{\mathbf{H}}(n)]}{\partial \hat{\mathbf{H}}(n)} = -2\mathbb{E}\left[\mathbf{e}_\mathrm{h}(n)\mathbf{s}^T(n)\right]. \tag{13}$$

Accordingly, the NLMS-type adaptive update equation for the proposed tracking system is expressed as

$$\hat{\mathbf{H}}(n) = \hat{\mathbf{H}}(n-1) + \frac{\mu_\mathrm{h}}{\|\mathbf{s}(n)\|^2 + \delta_\mathrm{h}}\mathbf{e}_\mathrm{h}(n)\mathbf{s}^T(n). \tag{14}$$

B. Decorrelation of the Excitation Signals

As discussed in Sec. I, when the excitation signals in the ASFC system consist of music or speech, the resulting loudspeaker signals $\mathbf{s}(n)$ remain highly self correlated and correlated across channels, even after being processed by the control filters $\mathbf{q}(n)$. To address this issue, signal decorrelation is necessary. Motivated by approaches employed in AEC, two decorrelation techniques are explored.

1) Phase modulation: The first decorrelation technique considered in this work is subband phase modulation, selected for its proven effectiveness in enhancing convergence performance without significantly degrading learning quality [24]. In this method, an oversampled analysis filter bank with decimation factor D decomposes the excitation signal into U subbands. Each subband signal $s_l(\omega, u)$ from the lth channel is then modulated by a complex exponential function $e^{\jmath\phi_l(\omega,u)}$, resulting in the decorrelated subband-domain signal as

$$z_{l,\mathrm{PHI}}(w,u) = s_l(w,u)e^{\jmath\phi_l(w,u)}, \tag{15}$$

where w and u denote, respectively, the time and subband indices. A synthesis filter bank is then used to reconstruct the fullband signal $z_{l,\mathrm{PHI}}(n)$ from the modulated subbands.

The effectiveness of this decorrelation method largely relies on the phase function $\phi_l(\omega, u)$. The subband framework allows for the use of distinct phase functions across subbands, facilitating perceptually optimized designs that minimize signal

[1]Here, we set the length of RIR and control filter for each channel is equal, i.e., $K = J$.

degradation. To balance effective decorrelation with minimal audible distortion, a smooth phase function is typically adopted [21], [37], such that

$$\phi_l(w,u) = \alpha_u \sin(2\pi f_{l,u} w \frac{D}{f_s}), \tag{16}$$

where α_u represents the phase amplitude of the uth subband, $f_{l,u}$ is the corresponding modulation frequency for lth channel, and f_s is the fullband sampling rate.

2) Half-wave rectifier: The second decorrelation technique employed in this work is the widely used half-wave rectifier (HWR). This approach provides a favorable balance between effective decorrelation and low distortion, as its nonlinear transformation is inherently adapted to the excitation signal [20], [29]. The resulting decorrelated signal for the lth loudspeaker is given by

$$z_{l,\mathrm{HWR}}(n) = s_l(n) + \frac{\beta}{2}\left[s_l(n) + (-1)^{l+1}\,|s_l(n)|\right], \tag{17}$$

where the factor β controls the degree of nonlinearity applied to the excitation signal, and $|\cdot|$ is the absolute value operation.

IV. Proposed Alternating Mode-Switching Method

As discussed in Sec. I, conventional ASFC methods with online RIR tracking require the simultaneous adaptation of both the RIR modeling and SFC filters. This dual-update process incurs significant computational overhead, particularly in systems with many microphones and loudspeakers, thereby limiting practical deployment. To mitigate this issue, we propose a mode-switching strategy that activates only one update mode at a time based on predefined criteria, enabling more efficient use of system resources. This approach is motivated by similar techniques employed in ANC systems [38], [39].

Initially, the SFC update (Mode 1) operates using pre-calibrated RIRs, as illustrated in Fig. 2. When the system reaches a stable state and the SFC error $\mathbf{e}_\mathrm{q}$ falls below a predefined threshold ϵ_q, it indicates that the RIRs have remained relatively unchanged. In this scenario, the RIR tracking filters are left unaltered. Selecting an appropriate threshold (e.g., $\epsilon_\mathrm{q} = -20$ dB) ensures that the control performance within the target zone is preserved. Conversely, when the condition $T_1 : \mathbf{e}_\mathrm{q} \geq \epsilon_\mathrm{q}$ is met, it suggests that the RIRs may have changed over time. In response, the system switches to RIR tracking, while holding the SFC filters constant.

Upon activation by condition T_1, the system enters RIR tracking mode (Mode 2), during which the convergence behavior of the RIR modeling error must be monitored in real time. The control-filtered signals $\mathbf{s}(n)$ are processed through the 'Decorr.' block shown in Fig. 1, producing the loudspeaker playback signals, i.e., the decorrelated excitation signal $\mathbf{z}(n)$. When the error in the RIR modeling filters begins to stabilize, it indicates that the current RIR estimates are sufficiently accurate for effective SFC. To identify the point at which the RIR estimates become reliable, a second criterion T_2 is introduced. This criterion is based on the slope of the MSE function $\mathcal{J}[\hat{\mathbf{H}}(n)]$ defined in (12), and is given by

$$T_2 : \frac{\partial \mathcal{J}[\hat{\mathbf{H}}(n)]}{\partial n} = \mathbb{E}\left[2\mathbf{e}_\mathrm{h}(n)\frac{\partial \mathbf{e}_\mathrm{h}(n)}{\partial n}\right] \to 0. \tag{18}$$

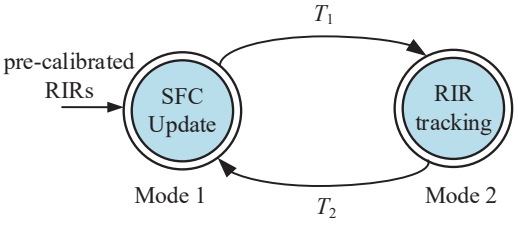

Fig. 2. Diagram illustrating mode-switching between SFC updates and RIR tracking.

In practical implementation, the above criterion can be approximated by substituting the expectation and derivative with a time average and a first-order difference, respectively [38], [39], i.e.,

$$T_2 : \frac{1}{N_\mathrm{a}} \sum_{n_\mathrm{a}=0}^{N_\mathrm{a}-1} \mathbf{e}_\mathrm{h}(n-n_\mathrm{a})\left[\mathbf{e}_\mathrm{h}(n-n_\mathrm{a}) - \mathbf{e}_\mathrm{h}(n-n_\mathrm{a}-1)\right] \to 0, \tag{19}$$

where N_a denotes the total number of samples used for averaging. Ideally, a near-zero slope signifies RIR convergence; however, this is seldom observed in practice. Therefore, a threshold ϵ_h is employed to trigger the switch back to Mode 1. The threshold parameters ϵ_q and ϵ_h were empirically tuned in our experiments, since their optimal setting depends on the specific system configuration and signal conditions.

For large-scale MIMO SFC systems, simultaneously running both adaptive processes is computationally intensive. Additionally, their mutual interaction can impede the convergence of each algorithm. The proposed mode-switching strategy addresses these challenges by activating only one mode at a time, thereby reducing the overall computational burden and effectively minimizing interference between the two processes.

V. Evaluation

A. Performance Evaluation Metrics

Two metrics are used for evaluation: normalized signal distortion energy (NSDE) and normalized misalignment (NM).

- The NSDE at time n is defined as

$$\varepsilon_\mathrm{NSDE}(n) = 10\log_{10}\frac{\left\|\mathbf{p}(n) - \mathbf{p}^\mathrm{d}(n)\right\|^2}{\left\|\mathbf{p}^\mathrm{d}\right\|^2}. \tag{20}$$

- The NM quantifies the distance between the true and estimated impulse response matrices, normalized by the Frobenius norm of the true RIR matrix [40]. At time index n, it is defined as

$$\mathrm{NM}(n) = 10\log_{10}\frac{\left\|\mathbf{H}(n) - \hat{\mathbf{H}}(n)\right\|_F}{\left\|\mathbf{H}(n)\right\|_F}. \tag{21}$$

B. Simulation Setup

The well known image model is used to simulate the acoustic environments for evaluation [41]. The simulated room measures 5 m × 4 m × 2.5 m, with the speed of sound set to $c = 343$ m/s. RIRs are generated using the code provided in [42], featuring a reverberation time of $T_{60} \approx 500$ ms. The RIRs are truncated into a length of 128 samples. A 4-element uniform linear loudspeaker array with 0.15 m spacing is used, with the second speaker selected as the virtual source.

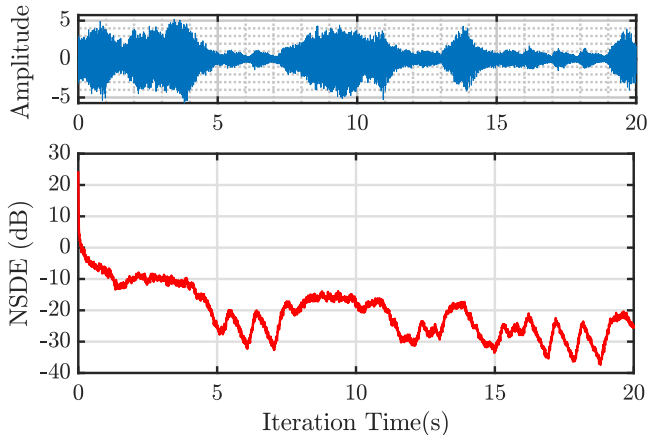

Fig. 3. Convergence behavior of the MCFxLMS algorithm applied to sound field control with a music signal as its input. The upper subplot presents the time-domain waveform of the input music signal at the 1st-channel loudspeaker.

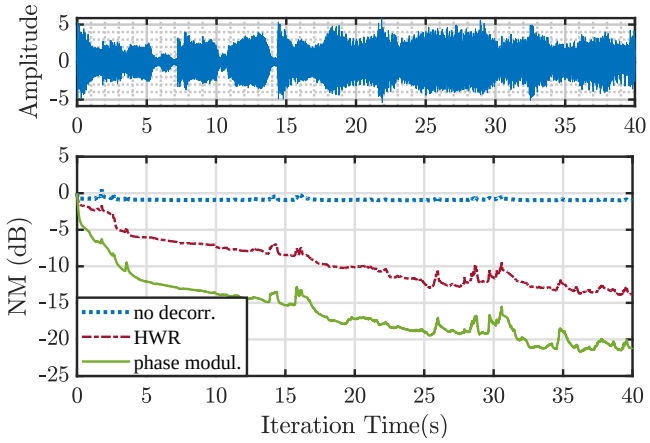

Fig. 4. Convergence performance of RIR tracking using various decorrelation methods. The upper subplot displays the time-domain waveform of the excitation signal applied to the 1st-channel loudspeaker.

The loudspeakers, microphones, and virtual source are all positioned in the same horizontal plane at a height of 0.5 m. A total of $M = 9$ microphones are uniformly distributed within a square control region, spaced at 0.1 m intervals. The loudspeaker and microphone arrays are aligned along the room's central axis and separated by 2 m. The loudspeaker array is located 0.5 m and 2.245 m from the nearest walls. The system operates at a sampling rate of $f_\mathrm{s} = 16$ kHz, and the loudspeaker control filters have a length of 128 taps. To simulate sensor noise, Gaussian white noise with an SNR of 30 dB is added at the microphones. Both the ASFC and RIR modeling filters are initialized to zero vectors and matrices, respectively. Step sizes are set to $\mu_\mathrm{q} = \mu_\mathrm{h} = 1$, with regularization parameters $\delta_\mathrm{q} = \delta_\mathrm{h} = 10^{-6}$. All results are averaged over 20 Monte Carlo runs.

For the phase modulation-based decorrelation method, $U = 36$ filter banks are employed with a decimation factor of $D = 8$. The phase function is defined by parameters α_u and $f_{l,u}$. Phase amplitudes α_u are chosen according to Table I, based on the listening test results reported in [21], to introduce inter-channel phase offsets. The modulation frequency for the lth loudspeaker is set to $f_{l,u} = 10 + 2(l-1)$ Hz. In the case of the HWR method, the parameter β is fixed at 1.

C. ASFC Performance

Figure 3 presents the convergence behavior of the proposed MCFxLMS algorithm for SFC under time-invariant RIRs (i.e., operating in Mode 1). The error quickly drops below -20 dB, demonstrating effective adaptation to nonstationary inputs. As long as the RIR remains unchanged, the error continues to decrease gradually and stays at a consistently low level.

D. RIR Tracking Performance

Figure 4 shows the RIR tracking performance when the system switches to Mode 2. Without decorrelation, the RIR tracking performance is poor, which is likely due to the high temporal correlation and limited spectral diversity of the excitation signal, as music signals are typically harmonic and narrowband. To address this, phase modulation and HWR are employed for signal decorrelation. Among the methods, phase modulation delivers the best performance, achieving up to a 20-dB reduction in normalized misalignment by the end of the tracking period. This improvement is attributed to its strong decorrelation capability, as it alters the phase structure across the full frequency range while preserving the magnitude, enabling faster convergence and lower steady-state error. In comparison, HWR offers a favorable trade-off between performance and computational efficiency. However, as a simple, low-latency non-linear operation that maintains the temporal envelope of the signal, HWR is better suited for real-time or resource-constrained applications.

TABLE I
THE PHASE AMPLITUDE α_u FOR DIFFERENT SUBBANDS.

Subband u	$0-3$	4	5	6	≥ 7
α_u [Degrees]	20°	40°	70°	90°	180°

Due to space limitations, this work focuses on a proof-of-concept, while evaluations under dynamic RIR variations are left for future work.

VI. CONCLUSIONS

This paper presented a novel adaptive sound field control method within a MIMO framework, incorporating acoustic path tracking driven by realistically correlated excitation signals. The proposed approach supports online adaptation to input signals with or without self and inter-channel correlation and includes a decorrelation mechanism based on filtered loudspeaker outputs to enable accurate and robust tracking of room impulse responses. To further enhance practical applicability, a coordination strategy was developed to alternate between adaptive sound field control and RIR tracking, improving computational efficiency and adaptability.

REFERENCES

[1] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Speech, Audio Process.*, vol. 9, no. 6, pp. 697–707, 2001.

[2] S. Koyama, G. Chardon, and L. Daudet, "Optimizing source and sensor placement for sound field control: An overview," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 696–714, 2020.

[3] J. Yang, M. Wu, and L. Han, "A review of sound field control," *Appl. Sci.*, vol. 12, no. 14, p. 7319, 2022.

[4] Q. Xu, S. Wang, J. Tao, H. Zou, and X. Qiu, "Creating personal sound zones in car cabins with active noise control," *Appl. Acoust.*, vol. 235, p. 110670, 2025.

[5] 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.

[6] S. S. Bhattacharjee, A. J. Fuglsig, J. R. Jensen, L. Shi, G. Ping, H. Shen, and M. G. Christensen, "Low complexity signal adaptive sound zone control using subspace tracking," in *Proc. IEEE Int. Workshop Acoust. Signal Enhancement.*, 2024, pp. 309–313.

[7] 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. Amer.*, vol. 150, no. 3, pp. 1779–1793, 2021.

[8] M. B. Møller, J. Martinez, and J. Østergaard, "Reduced complexity for sound zones with subband block adaptive filters and a loudspeaker line array," *J. Acoust. Soc. Amer.*, vol. 155, no. 4, pp. 2314–2326, 2024.

[9] N. De Koeijer, M. B. Møller, J. Martinez, P. Martínez-Nuevo, and R. C. Hendriks, "Block-based perceptually adaptive sound zones with reproduction error constraints," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, 2024.

[10] M. Zhang, J. Zhang, J. Chen, and C. Richard, "Distributed pressure matching for personal sound zone control using diffusion adaptation," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, 2023, pp. 501–505.

[11] S. S. Bhattacharjee, J. R. Jensen, and M. G. Christensen, "Sound speed perturbation robust audio: Impulse response correction and sound zone control," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 33, pp. 2008–2019, 2025.

[12] 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.* IEEE, 2022, pp. 8712–8716.

[13] ——, "Robust acoustic contrast control with positive semidefinite constraint using iterative POTDC algorithm," in *Proc. IEEE Int. Workshop Acoust. Signal Enhancement.*, 2022, pp. 1–5.

[14] ——, "CGMM-based sound zone generation using robust pressure matching with ATF perturbation constraints," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 3331–3345, 2023.

[15] C. Paleologu, J. Benesty, and S. Ciochină, "Linear system identification based on a kronecker product decomposition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 10, pp. 1793–1808, 2018.

[16] G. Huang, J. Benesty, and J. Chen, "Dimensionality reduction of room acoustic impulse responses and applications to system identification," *IEEE Signal Process. Lett.*, vol. 30, pp. 1107–1111, 2023.

[17] R. Longo, L. Simon, M. Melon *et al.*, "Online adaptive identification of multichannel systems for audio applications," *J. Acoust. Soc. Amer.*, vol. 155, no. 1, pp. 229–240, 2024.

[18] S. Zhao and I. Burnett, "Adaptive personal sound zones systems with online plant modelling," in *The 24th International Congress on Acoustics*, 2022.

[19] M. Hu, L. Shi, H. Zou, M. G. Christensen, and J. Lu, "Sound zone control with fixed acoustic contrast and simultaneous tracking of acoustic transfer functions," *J. Acoust. Soc. Amer.*, vol. 153, no. 5, pp. 2538–2538, 2023.

[20] S. Kühl, C. Antweiler, T. Hübschen, and P. Jax, "Kalman filter based stereo system identification with auto-and cross-decorrelation," in *Proc. IEEE Int. Conf. Hands-Free Speech Commun. Microphone Arrays*. IEEE, 2017, pp. 181–185.

[21] J. Herre, H. Buchner, and W. Kellermann, "Acoustic echo cancellation for surround sound using perceptually motivated convergence enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, 2007, pp. I–17.

[22] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza, "Low-complexity implementation of a real-time decorrelation algorithm for stereophonic acoustic echo cancellation," *Signal Process.*, vol. 92, no. 11, pp. 2668–2675, 2012.

[23] L. Romoli, S. Cecchi, and F. Piazza, "A combined approach for channel decorrelation in stereo acoustic echo cancellation exploiting time-varying frequency shifting," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 717–720, 2013.

[24] M. L. Valero and E. A. Habets, "Insight into a phase modulation technique for signal decorrelation in multi-channel acoustic echo cancellation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2016, pp. 519–523.

[25] T. S. Wada and B. Juang, "Multi-channel acoustic echo cancellation based on residual echo enhancement with effective channel decorrelation via resampling," in *Proc. IEEE Int. Workshop Acoust. Signal Enhancement.*, 2010.

[26] J. Wung, T. S. Wada, and B.-H. Juang, "Inter-channel decorrelation by sub-band resampling in frequency domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 29–32.

[27] ——, "On the performance of the robust acoustic echo cancellation system with decorrelation by sub-band resampling," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 635–638.

[28] J. Wung, T. S. Wada, M. Souden, and B.-H. Juang, "Inter-channel decorrelation by sub-band resampling for multi-channel acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 2127–2142, 2014.

[29] J. Benesty, D. R. Morgan, and M. M. Sondhi, "A better understanding and an improved solution to the specific problems of stereophonic acoustic echo cancellation,"*IEEE Trans. Speech, Audio, Process.*, vol. 6, no. 2, pp. 156–165, 2002.

[30] S. Cecchi, L. Romoli, P. Peretti, and F. Piazza, "A combined psychoacoustic approach for stereo acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1530–1539, 2010.

[31] L. Romoli, S. Cecchi, P. Peretti, and F. Piazza, "A mixed decorrelation approach for stereo acoustic echo cancellation based on the estimation of the fundamental frequency," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 2, pp. 690–698, 2011.

[32] P. Thüne and G. Enzner, "Trends in adaptive miso system identification for multichannel audio reproduction and speech communication," in *in Proc. IEEE Int. Symp. Image Signal Process. Anal.* 2013, pp. 767–772.

[33] ——, "Improved online identification of acoustic MISO systems based on separated input signal components," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.* 2013, pp. 413–417.

[34] S. Galvez, F. Marcos, S. J. Elliott, and C. Jordan, "Time domain optimisation of filters used in a loudspeaker array for personal audio," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 13, no. 11, pp. 1869–1878, 2015.

[35] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO signal processing*. Springer Science & Business Media, 2006.

[36] ——, "Identification of acoustic MIMO systems: Challenges and opportunities," *Signal Process.*, vol. 86, no. 6, pp. 1278–1295, 2006.

[37] M. Guo, S. H. Jensen, J. Jensen, and S. L. Grant, "On the use of a phase modulation method for decorrelation in acoustic feedback cancellation," in *Proc. IEEE Eur. Signal Process. Conf.*, 2012, pp. 2000–2004.

[38] J. Ji, D. Shi, W.-S. Gan, X. Shen, and Z. Luo, "A computation-efficient online secondary path modeling technique for modified fxlms algorithm," in *in *Proc. INTER-NOISE NOISE-CON Congr. Conf.*, vol. 268, no. 3, 2023, pp. 5216–5226.

[39] X. Shen, W.-S. Gan, and D. Shi, "Alternative switching hybrid ANC," *Appl. Acoust.*, vol. 173, p. 107712, 2021.

[40] J. Benesty, C. Paleologu, L.-M. Dogariu, and S. Ciochină, "Identification of linear and bilinear systems: A unified study," *Electronics*, vol. 10, no. 15, p. 1790, 2021.

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

[42] C. Pan, L. Zhang, Y. Lu, J. Jin, L. Qiu, J. Chen, and J. Benesty, "An anchor-point based image-model for room impulse response simulation with directional source radiation and sensor directivity patterns," *arXiv prepr. arXiv:2308.10543*, 2023.