Elsevier required licence: © <2025>. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
The definitive publisher version is available online at 10.1016/j.apacoust.2024.110341

Virtual quiet zone method for sound zone reproduction in coupled rooms⋆

Xiaoyi Gao[a,b], Qiaoxi Zhu[c], Xiangning Liao[d], Ming Wu[a,b,*], Lu Han[a,b] and Jun Yang[a,b,*]

[a]Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China
[b]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
[c]University of Technology Sydney, Sydney NSW 2007, Australia
[d]Chongqing Changan Automobile Company Limited, Chongqing 400023, China

ARTICLE INFO

Keywords:
Spatial audio
Sound zone reproduction
Array signal processing
Improved Grey Wolf Optimizer

ABSTRACT

Sound zone reproduction aims to deliver sound to target listeners in the listening zone while reducing unwanted sound for others in the quiet zone. Existing approaches primarily focus on reproducing sound zones within a single room. When employing existing methods in coupled spaces, the accurate measurement of cross-room acoustic transfer functions (ATFs) poses significant challenges. This paper proposes a virtual quiet zone method to improve sound zone performance in coupled rooms. The proposed method eliminates the need for complex cross-room ATF measurement. We use the Improved Grey Wolf Optimizer to establish an appropriate virtual quiet zone within the virtual space based on the measured sound pressures at the zones. The ATFs between the loudspeakers and the virtual quiet zone can be interpolated utilizing acoustic modeling and the measured ATFs between the loudspeakers and the listening zone. In the proposed algorithm, the optimization of sound zone reproduction in coupled rooms is transformed into the optimization of the position of the virtual quiet zone within the virtual space. Experimental results show that the proposed method outperforms the current acoustic contrast control method and brightness control method in real-world implementations in coupled rooms.

1. Introduction

Sound zone reproduction uses a loudspeaker array to deliver the audio content to the listeners in the specific zone while minimizing interference to others in a shared space [1, 2]. The zone with higher acoustic energy is called the listening zone, where the listeners can enjoy listening to the desired audio program. The zone with lower acoustic energy is called the quiet zone, where the disturbance from the reproduced audio program is controlled. Sound zone reproduction has various applications, including mobile devices [3], smart car cabins [4], and personal computers [5].

Many approaches have been developed for generating individual sound zones, such as acoustic contrast control (ACC) [6], brightness control (BC) [6], pressure matching [7] and hybrid approaches [8, 9]. All of these approaches require knowledge of the acoustic transfer function (ATF) between each loudspeaker and microphone to design the control filters, i.e., loudspeaker weights, that process the input audio signal.

Existing research typically focuses on single-room scenarios [Fig. 1(a)] [10], where there are no physical partitions between the listening and quiet zones. For example, assuming that the reverberation time is uniform throughout the room and that the sound energy decays at a consistent rate in all directions [11]. Therefore, the control filters for the sound zone reproduction system are generally designed by directly measuring the ATFs in single-room scenarios.

Apart from single rooms, coupled rooms are also common in real-world environments, such as open-plan offices, residential apartments, and public spaces. Coupled rooms can be two rooms of different volumes and reverberation characteristics connected by an aperture [12, 13, 14]. Sound energy in coupled rooms can decay at different rates. Moreover, due to the portal effects and boundary diffraction, which vary with inter-room position and coupling aperture size, the transition between coupled rooms is an intricate interaction [15]. Thus, accurately modeling the cross-room ATF is challenging, which makes the ATF measurement essential.

In this paper, we address sound zone reproduction in coupled rooms [Fig. 1(b)], using a loudspeaker array to create a listening zone in Room A and a quiet zone in the adjacent Room B. For example, creating a listening zone in a living room and a quiet zone in an adjoining study. The challenge arises when the loudspeakers and the quiet zone are located in different rooms with distinct acoustic conditions. This configuration presents difficulties for implementation, as it depends on measuring ATFs between the loudspeakers and the zones to design control filters. ATF measurement conducted in non-laboratory environments can suffer from varying background noise, such as environmental noise and human activity, which tend to intensify as the distance between the loudspeaker and microphone increases [11, 13, 14]. These factors make the accurate measurement of cross-room ATFs intricate.

⋆This work was supported by the Beijing Natural Science Foundation under Grant L223032.

*Corresponding author.
gaoxiaoyi@mail.ioa.ac.cn (X. Gao); qiaoxi.zhu@gmail.com (Q. Zhu); liaoxn@changan.com.cn (X. Liao); mingwu@mail.ioa.ac.cn (M. Wu); hanlu@mail.ioa.ac.cn (L. Han); jyang@mail.ioa.ac.cn (J. Yang)

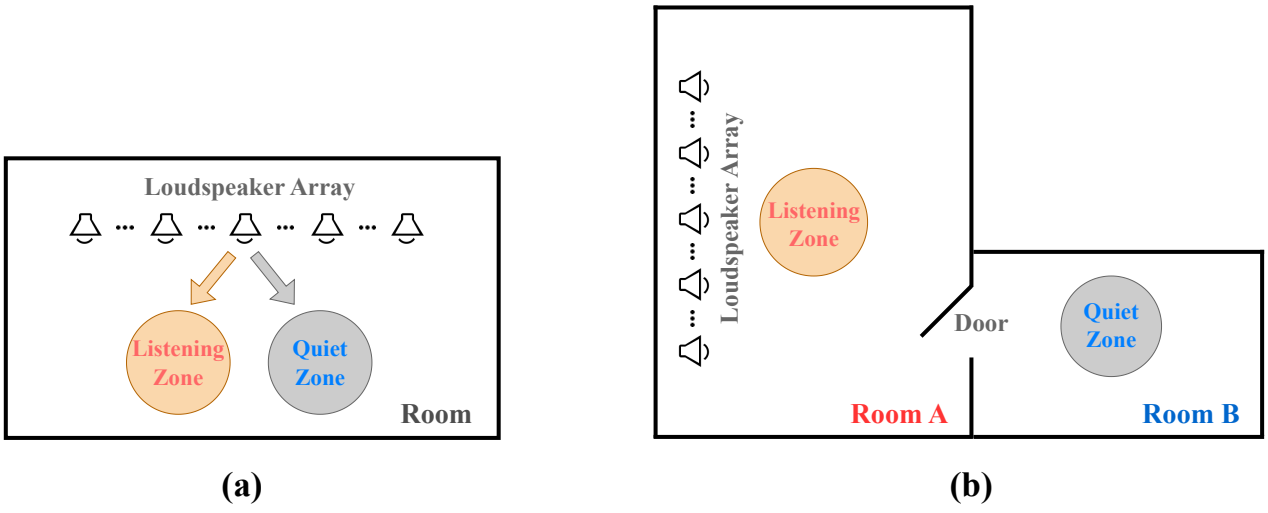


Figure 1: Sound zone reproduction systems in (a) a single room and (b) coupled rooms.

Existing sound zone reproduction techniques have been attempted for complex environments in single-room scenarios. Recent studies have explored robustness optimization [16, 17] and adaptive control [4, 18]. Robustness optimization designs control filters by measuring ATFs and estimating perturbations, while adaptive control continuously tracks ATFs and updates the filters in real time. However, these methods may be impractical for multi-room scenarios due to their complexity and the need for frequent ATF measurements to update the control filters.

This paper addresses a gap in the literature by developing simplified systems for personal sound zones in coupled rooms. We propose a virtual quiet zone method to enhance sound zone reproduction performance without requiring cross-room ATF measurement. Specifically, we introduce a virtual quiet zone, in which the ATFs are interpolated using acoustic modeling and the measured ATFs between the loudspeakers and the listening zone. Here, a point source model is utilized for acoustic modeling, considering the limited degrees of freedom that can be controlled with a finite number of loudspeakers in complex environments. The position of the virtual quiet zone is optimized using the Improved Grey Wolf Optimizer (IGWO) [19], with mean acoustic contrast (MeanAC) incorporating pre-emphasis [20, 21] as the fitness function. This method enables stable, real-time selection even during the reproduction of time-varying speech. By approximating the ATFs of the virtual quiet zone to those of the actual quiet zone, essential ATF information is preserved for calculating the loudspeaker weights. The properly regularized ACC is employed to design the control filter using the measured ATFs of the listening zone and the interpolated ATFs of the virtual quiet zone. This approach aims to reduce the impact of the ATF mismatches on performance.

The contributions of this paper are threefold. First, a virtual quiet zone is introduced to convert the task of sound zone reproduction in coupled spaces into the optimization of the position of the virtual quiet zone, which circumvents complex cross-room ATF measurement. Second, results from experiments show that our method outperforms ACC and BC in real-world implementations in coupled rooms. Third, building on the ACC method, our proposed approach can potentially extend other existing sound zone reproduction approaches from single rooms to coupled rooms with simplified implementation requirements.

2. Preliminary

In a typical sound zone reproduction system, a listening zone and a quiet zone are generated by a loudspeaker array composed of L loudspeakers. In the frequency domain, the sound pressure at N_L points in the listening zone and the sound pressure at N_Q points in the quiet zone are $\mathbf{p}_\mathrm{L} = \mathbf{G}_\mathrm{L}\mathbf{w} \cdot s$ and $\mathbf{p}_\mathrm{Q} = \mathbf{G}_\mathrm{Q}\mathbf{w} \cdot s$, respectively. Specifically, s is the input signal, $\mathbf{w} = [w_1, \ldots, w_l, \ldots, w_L] \in \mathbb{C}^{L\times 1}$ is the loudspeaker weight vector, $\mathbf{G}_\mathrm{L} \in \mathbb{C}^{N_\mathrm{L}\times L}$ and $\mathbf{G}_\mathrm{Q} \in \mathbb{C}^{N_\mathrm{Q}\times L}$ represent the ATF matrices corresponding to the listening zone and the quiet zone, respectively. As the theory in this paper is formulated in the frequency domain, the denotation of frequency is omitted in most equations for brevity.

2.1. Brightness control (BC)

BC optimizes the ratio of acoustic energy in the listening zone to the input power [6], that

$$\max_{\mathbf{w}} \frac{\|\mathbf{p}_\mathrm{L}\|_2^2}{\|\mathbf{w}\|_2^2} = \max_{\mathbf{w}} \frac{\mathbf{w}^\mathrm{H}\mathbf{G}_\mathrm{L}^\mathrm{H}\mathbf{G}_\mathrm{L}\mathbf{w}}{\mathbf{w}^\mathrm{H}\mathbf{w}}, \tag{1}$$

and the solution of Eq. (1) is

$$\mathbf{w} = \mathbf{\Phi}\left\{\mathbf{G}_\mathrm{L}^\mathrm{H}\mathbf{G}_\mathrm{L}\right\}, \tag{2}$$

where $\mathbf{\Phi}\{*\}$ denotes the eigenvector corresponding to the maximum eigenvalue of a matrix.

As discussed in Section 1, the accurate measurement of $\mathbf{G}_\mathrm{Q}$ poses challenges for sound zone reproduction in coupled rooms. BC does not require $\mathbf{G}_\mathrm{Q}$ and can be directly applied in coupled spaces. However, it performs less effectively than ACC in terms of acoustic contrast (AC) (see ELM in [16]).

2.2. Acoustic contrast control (ACC)

ACC maximizes the AC between the listening zone and the quiet zone, where AC is defined as the ratio of the spatially averaged sound energy between the two zones. The objective of ACC is formulated as

$$\max_{\mathbf{w}} \mathrm{AC} = \max_{\mathbf{w}} \frac{N_{\mathrm{Q}}\|\mathbf{p}_{\mathrm{L}}\|_2^2}{N_{\mathrm{L}}\|\mathbf{p}_{\mathrm{Q}}\|_2^2} = \max_{\mathbf{w}} \frac{N_{\mathrm{Q}}\mathbf{w}^{\mathrm{H}}\mathbf{G}_{\mathrm{L}}^{\mathrm{H}}\mathbf{G}_{\mathrm{L}}\mathbf{w}}{N_{\mathrm{L}}\mathbf{w}^{\mathrm{H}}\mathbf{G}_{\mathrm{Q}}^{\mathrm{H}}\mathbf{G}_{\mathrm{Q}}\mathbf{w}}. \quad (3)$$

The ATF-based solution of Eq. (3) is

$$\mathbf{w} = \mathbf{\Phi}\left\{(\mathbf{G}_{\mathrm{Q}}^{\mathrm{H}}\mathbf{G}_{\mathrm{Q}} + \lambda\mathbf{I})^{-1}\mathbf{G}_{\mathrm{L}}^{\mathrm{H}}\mathbf{G}_{\mathrm{L}}\right\}, \quad (4)$$

where λ is the regularization parameter [22], which needs to be designed carefully for robustness improvement.

2.3. Robustness

The acoustic modeling-based regularized sound zone reproduction solutions are formulated in [16, 23]. Taking into account the additive errors present in the reproduction systems, the ATF between the lth loudspeaker and the nth control point is defined as:

$$\widehat{\mathbf{G}}_{n,l} = \mathbf{G}_{n,l} + ge^{j\phi}, \quad (5)$$

where $j = \sqrt{-1}$, g and ϕ are the amplitude and phase of the additive error, respectively.

Due to the challenge in estimating the error distribution, only the error bound estimate is employed to calculate the loudspeaker weights. Assuming that the amplitude g is uniformly distributed in the range $[0, g_{\max}]$ and phase ϕ is uniformly distributed in the range $[0, 2\pi]$. In this case, λ can be expressed by

$$\lambda = N_{\mathrm{Q}} \cdot g_{\max}^2/3. \quad (6)$$

2.4. Reduced in-situ measurement by acoustic modeling

The regularization approach incorporating acoustic modeling has been applied to reduce in-situ measurements of ATFs for sound zone reproduction in a single room [16]. In [16], the in-situ ATF measurements are reduced to a single measurement per loudspeaker, per zone. The ATFs of the other control points are obtained by interpolation, which are modeled as interpolated $\widetilde{\mathbf{G}}_{n,l}$ with mismatch error:

$$\widehat{\mathbf{G}}_{n,l} = \widetilde{\mathbf{G}}_{n,l} + ge^{j\phi}. \quad (7)$$

The corresponding regularization parameter λ in Eq. (6) can be determined with $g_{\max} = \max\{\mathbf{G}_{\mathrm{Q}}\} \times g_{\max,\mathrm{ME}}$, where $\max\{\mathbf{G}_{\mathrm{Q}}\}$ represents the maximum among all the ATFs in $\mathbf{G}_{\mathrm{Q}}$, and $g_{\max,\mathrm{ME}} = \sqrt{\gamma^2 - 2\cos(\phi_{\max})\gamma + 1}$ with $\gamma = 10^{\frac{3}{20}}$ and $\phi_{\max} = 10°$ [16].

In the proposed method, we extend this approach to the coupled rooms. Instead of measuring one ATF per loudspeaker per zone and then interpolating ATF towards the other control points at the same zone, this paper deals with measuring ATFs of the listening zone in Room A and then interpolating ATFs towards the virtual quiet zone in the virtual space, as detailed in the following section. This approach aims to preserve necessary ATF information to implement sound zone reproduction across rooms.

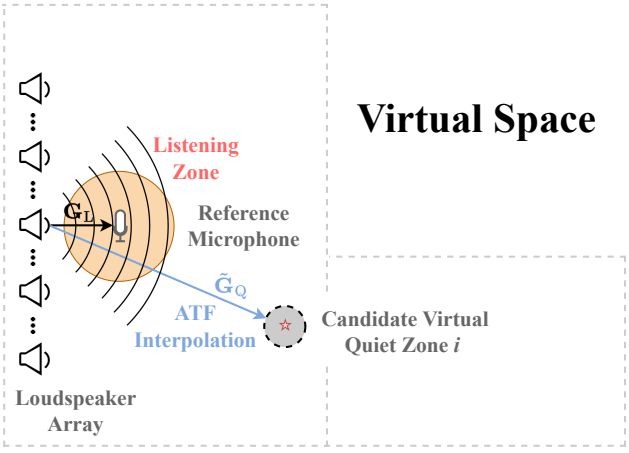


Figure 2: The proposed virtual quiet zone ($i = 1, 2, \ldots, I$) exists in the virtual space, and its acoustic transfer functions (ATFs) are estimated through interpolation using the point source model and measured ATFs between actual loudspeakers and the reference microphone in the actual listening zone.

3. Method

To circumvent complex cross-room ATF measurement, we present a feasible alternative solution. As shown in Fig. 2, we introduce a virtual space that combines the listening zone and the virtual quiet zone into a single connected space for simplified directional control. Introducing the virtual space allows us to map the beamforming from a complex environment onto the beamforming in a simplified environment.

3.1. Proposed method overview

We utilize the actual ATFs of the listening zone and employ point source-based interpolation to obtain the ATFs of the virtual quiet zone, as conceptually formulated in Eq. (8).

$$\widetilde{\mathbf{G}}_{\mathrm{Q}}^{i} = \mathrm{Interp}(\mathbf{G}_{\mathrm{L}}, \mathbf{X}_i), \quad (8)$$

where $\mathrm{Interp}(*)$ denotes the interpolation operation between the ATFs of the actual listening zone and the candidate virtual quiet zone. The ATF matrix of the ith virtual quiet zone $\widetilde{\mathbf{G}}_{\mathrm{Q}}^{i}$ is interpolated based on the measured $\mathbf{G}_{\mathrm{L}}$ and the position of the ith candidate virtual quiet zone $\mathbf{X}_i$, where $i = 1, 2, \ldots, I$ and I denotes the number of candidate positions of the virtual quiet zone. This approach contains nearly complete information about the primary energy radiation capability, i.e., directing a focused beam towards the listening zone while enabling directional control by creating a dark beam towards the virtual quiet zone.

The proposed virtual quiet zone method employs $\widetilde{\mathbf{G}}_{\mathrm{Q}}^{i}$ as a replacement for $\mathbf{G}_{\mathrm{Q}}$. Thus, the problem is then transformed

into finding the optimal position of the virtual quiet zone $\mathbf{X}_{\text{opt}}$, which corresponds to $\widetilde{\mathbf{G}}_{\text{Q}}^{\text{opt}}$, to preserve essential ATF information for loudspeaker weight calculation. As detailed in Section 3.4 and the Appendix, the IGWO algorithm is employed to select $\mathbf{X}_{\text{opt}}$ with the objective of maximizing the fitness function. The IGWO iteratively updates I candidate positions of the virtual quiet zone by monitoring sound pressures obtained at the actual listening and quiet zones, without requiring ATFs across multiple rooms. The candidate with the maximum $J(*)$ value is ultimately selected as $\mathbf{X}_{\text{opt}}$, that

$$\mathbf{X}_{\text{opt}} = \arg\max_{\mathbf{X}_i} J(\mathbf{X}_i). \tag{9}$$

Instead of capturing the complex sound propagation characteristics between sound zones across coupled rooms, the proposed method simplifies the process while remaining effective by considering the radiation capability of the loudspeaker array. The whole process is illustrated in Fig. 3, with details as follows.

3.2. ATF modeling of the virtual quiet zone

We use the point source model to interpolate the ATFs between the loudspeakers and the virtual quiet zone based on the measured listening zone ATFs. The interpolation process can be formulated as

$$\widetilde{G}_{i,l} = G_{\text{ref},l} \times e^{-jk(\widetilde{d}_{i,l} - d_{\text{ref},l})} \times d_{\text{ref},l} / \widetilde{d}_{i,l}, \tag{10}$$

where k represents the wave number of the sound wave. The wave number is defined as $k = 2\pi f/c$, where f is the frequency and c is the speed of sound. $\widetilde{G}_{i,l}$, $G_{\text{ref},l}$, $\widetilde{d}_{i,l}$ and $d_{\text{ref},l}$ denote the ATF and distance from the lth loudspeaker to the ith candidate virtual quiet zone and the reference microphone in the actual listening zone, respectively. The ATF matrix between the loudspeaker array and the ith candidate virtual quiet zone is then formulated as $\widetilde{\mathbf{G}}_{\text{Q}}^{i} = [\widetilde{G}_{i,1}, \widetilde{G}_{i,2}, \cdots, \widetilde{G}_{i,L}]$.

Point source interpolation offers advantages over pure point source model as it incorporates fundamental source equalization and partial loudspeaker directivity information. Depending on the condition of Room A, the measured room impulse responses from the loudspeaker array to the reference microphone may be truncated before deriving $G_{\text{ref},l}$ through FFT. This truncation retains only the direct sound and the main early reflection components, as described in [16].

3.3. Robust loudspeaker weight calculation

The proposed virtual quiet zone method introduces $\widetilde{\mathbf{G}}_{\text{Q}}^{i}$, which is the ATF matrix of the candidate virtual quiet zone, as a substitute for $\mathbf{G}_{\text{Q}}$. Thus, the loudspeaker weight solution of ACC is reformulated as

$$\widetilde{\mathbf{w}}^{i} = \mathbf{\Phi}\left\{[(\widetilde{\mathbf{G}}_{\text{Q}}^{i})^{\text{H}}\widetilde{\mathbf{G}}_{\text{Q}}^{i} + \lambda\mathbf{I}]^{-1}\mathbf{G}_{\text{L}}^{\text{H}}\mathbf{G}_{\text{L}}\right\}, \tag{11}$$

where λ is employed for regularization [16] against ATF mismatches.

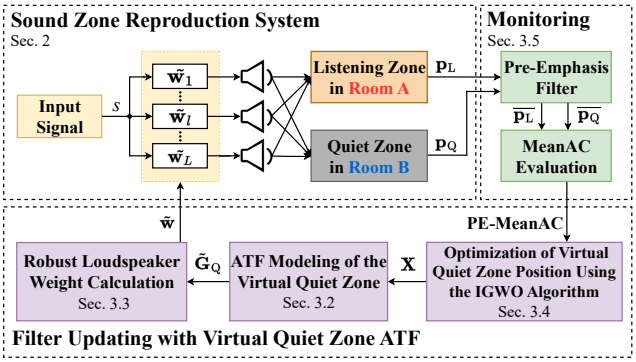


Figure 3: Proposed virtual quiet zone method for sound zone reproduction in coupled spaces. ATF refers to the acoustic transfer function. IGWO refers to the Improved Grey Wolf Optimizer.

Following the computation of loudspeaker weights for each frequency bin, the negative frequency bins are populated by complex conjugation, and an inverse FFT is performed to derive a time-domain control filter. Consequently, the signals inputted into the loudspeaker array can be obtained by filtering the input signal s using the designed control filter.

3.4. Optimization of virtual quiet zone position using the IGWO algorithm

Utilizing the ATFs from various candidate virtual quiet zones, we can generate distinct control filters leading to diverse signals for reproduction. The value of the fitness function computed based on signals collected from the listening zone in Room A and the quiet zone in Room B reflects the performance of the designed control filters, thus providing insight into the suitability of the chosen virtual quiet zone. The IGWO algorithm is employed to iteratively update each candidate position of the virtual quiet zone, aiming to achieve a higher value of the fitness function. After several iterations, the optimal location for the virtual quiet zone is determined, as depicted in Eq. (9).

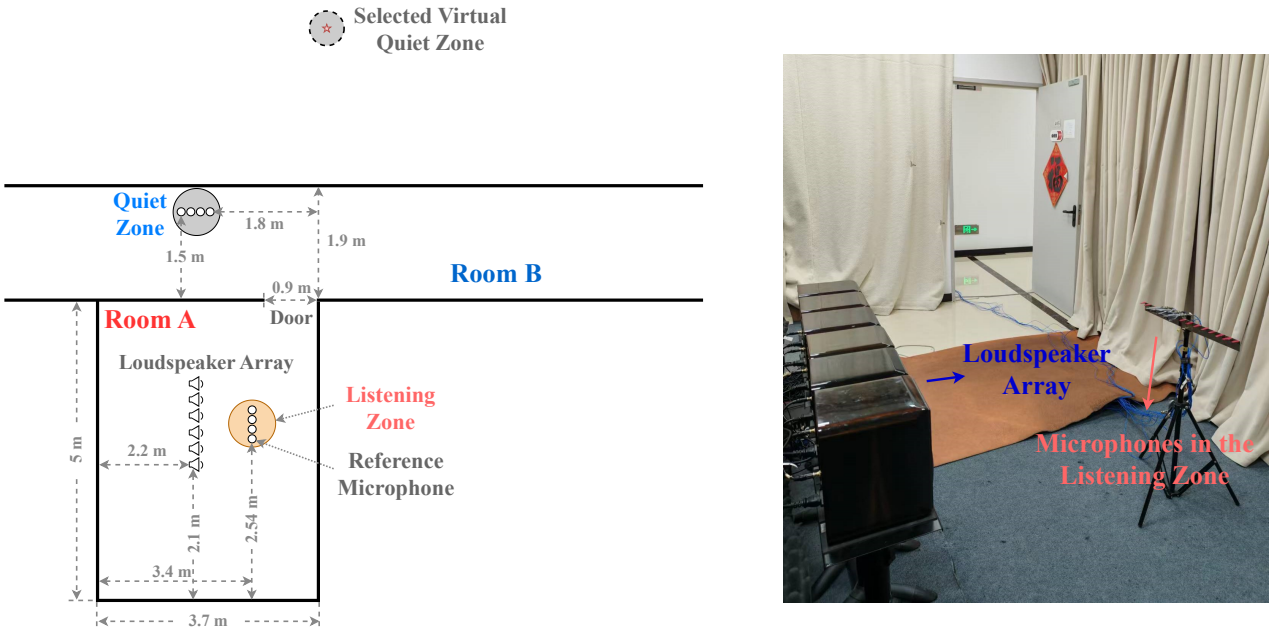
3.5. Fitness function for speech signal compatibility

In IGWO, robust convergence is crucial, especially when dealing with time-varying speech signals as the input signal s. OverallAC [24] and MeanAC [25] are common metrics in sound zone reproduction,

$$\text{OverallAC} = 10\log_{10}\left(\frac{N_{\text{Q}}}{N_{\text{L}}}\frac{\sum_{m=1}^{M}\|\mathbf{p}_{\text{L}}(f_m)\|^2}{\sum_{m=1}^{M}\|\mathbf{p}_{\text{Q}}(f_m)\|^2}\right), \tag{12}$$

$$\text{MeanAC} = 10\log_{10}\left(\frac{1}{M}\frac{N_{\text{Q}}}{N_{\text{L}}}\sum_{m=1}^{M}\frac{\|\mathbf{p}_{\text{L}}(f_m)\|^2}{\|\mathbf{p}_{\text{Q}}(f_m)\|^2}\right), \tag{13}$$

where f_m is the mth frequency, $m = 1, 2, \ldots, M$, and M is the number of frequency bins. Although both OverallAC and MeanAC show effectiveness with white noise as input,



(a) **(b)**

Figure 4: Experimental system. The listening zone is in a meeting room, while the quiet zone is in an adjacent corridor. (a) The schematic of the experimental system. The hollow circles denote microphones in the listening and quiet zones. The pink star denotes the optimal position of the virtual quiet zone, which is obtained by the Improved Grey Wolf Optimizer (IGWO). (b) Photo of the configuration within Room A.

MeanAC demonstrates greater consistency when handling varying speech signals.

Moreover, we propose an improved metric PE-MeanAC as the fitness function,

$$J(\mathbf{X}_i) = \text{PE-MeanAC} = 10\log_{10}\left(\frac{1}{M}\frac{N_\text{Q}}{N_\text{L}}\sum_{m=1}^{M}\frac{\|\overline{\mathbf{p}_\text{L}}(f_m,\mathbf{X}_i)\|^2}{\|\overline{\mathbf{p}_\text{Q}}(f_m,\mathbf{X}_i)\|^2}\right), \quad (14)$$

where $\overline{\mathbf{p}_\text{L}}$ and $\overline{\mathbf{p}_\text{Q}}$ denote the application of the pre-emphasis (PE) operation on signals received at the listening zone and the quiet zone, respectively.

PE-MeanAC is calculated by passing speech signals received at the listening zone and the quiet zone through a pre-emphasis filter to enhance high-frequency components and improve spectral flatness [20, 21]. The transfer function of the first-order pre-emphasis filter is defined as $H(z) = 1 - \mu z^{-1}$, where μ is the pre-emphasis coefficient. In this paper, we set μ to 0.98, which is consistent with the common practices employed in [26, 27, 28]. It is noted that there may exist other filters with potentially superior effects besides the pre-emphasis filter we employed, which will be left for future exploration. The proposed PE-MeanAC facilitates a more accurate evaluation of the control filters during each iteration of the proposed method and enhances robustness to non-stationary speech inputs.

4. Experiment

Figure 4 depicts the experimental setup for the sound zone reproduction system in coupled rooms. Specifically, Room A was a meeting room, while the corridor adjacent to the meeting room served as Room B. The coupling between these two rooms was established through an open door.

The line array comprised six HiVi-Swans H5 loudspeakers with a spacing of 0.2 m. The listening zone was situated 1.2 m away from the loudspeaker array. Both the listening zone and the quiet zone were equipped with four Brüel & Kjær Type 4189 microphones that were evenly spaced at 0.04 m intervals. The impulse responses were measured using the swept sine method [29] with a sampling rate of 48 kHz, and the ATFs were computed using an 8192-point FFT. The measured impulse responses from the loudspeaker array to the reference microphone were cropped at 50 points after the impulse onset to avoid unnecessary room effects in the transfer functions for the proposed method [30].

Speech signals and white noise were tested separately as the input signals for the proposed method. The input speech signals for each iteration of the proposed system were randomly sourced from a Chinese podcast [31], while the white noise input remained constant. The duration of each input signal was 10 seconds. The IGWO in the proposed approach applied a population size of 5 and underwent 100 iterations.

The pink star in Fig. 4 denotes the optimal position of the virtual quiet zone determined by IGWO after 100 iterations. The optimal virtual quiet zone simulates the acoustic properties of a real quiet zone in complex environments. It is noted that the location of the optimal virtual quiet zone is related to the physical attributes of the actual room and environmental factors like reflection, scattering and diffractions. However, as it is not physically present, its position is not limited by

Table 1
Acoustic contrast (AC) performance averaged over 300-3400 Hz in the experiment.

Method	MeanAC
No Control	15.24 dB
BC	17.26 dB
ACC	18.51 dB
Proposed (white noise)	**21.17 dB**
Proposed (speech)	**21.10 dB**

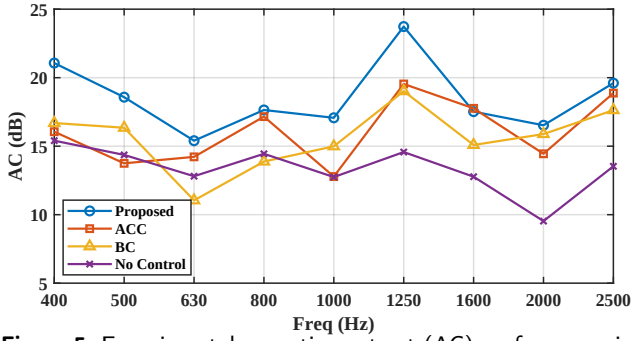


Figure 5: Experimental acoustic contrast (AC) performance in 1/3 octave bands centred 400-2500 Hz.

the boundaries of the actual rooms. The proposed method designed the control filter employing the measured ATFs of the actual listening zone and the interpolated ATFs of the aforementioned optimal virtual quiet zone.

We set up three baselines: No Control, BC and ACC. The No Control method reproduces the sound field without applying any weighting to the loudspeakers. BC does not require ATF measurement of the quiet zone, while ACC relies on ATF measurement of the quiet zone. After acquiring the control filter for each method, we evaluated the performance of each method by employing an identical segment of white noise signal as the system input. The measured acoustic contrast was subsequently derived utilizing the signals captured from the actual listening zone in Room A and the actual quiet zone in Room B, as depicted in Fig. 4.

Table 1 shows the MeanAC results over 300-3400 Hz for each method. Consistent MeanAC results for the proposed method using various input signals are presented in Table 1, indicating that the proposed PE-MeanAC greatly assists in accurately assessing the performance of the control filters. Additionally, Fig. 5 presents the detailed AC performance for each method across 1/3 octave bands considering human auditory perception. It is noted that Fig. 5 only displays the outcome of the proposed method using speech signals as input for brevity.

As shown in Table 1 and Fig. 5, the proposed method can improve AC performance, particularly in the lower frequency range below 1600 Hz. Specifically, in terms of MeanAC below 1600 Hz, the proposed method exhibits a superiority of 3.50 dB and 4.65 dB compared to the ACC and BC methods, respectively.

The ACC method relies on the ATFs between the loudspeakers in Room A and the actual quiet zone in Room B. However, the accurate measurement of cross-room ATFs is challenging due to its sensitivity to variations in the acoustic environment. BC method primarily concentrates the sound energy on the listening zone, and disregards controlling over the quiet zone. In contrast, the proposed virtual quiet zone method circumvents complex cross-room ATF measurement. This approach iteratively updates the control filter by optimizing the position of the virtual quiet zone using sound pressure measurement, which is more robust against perturbations in real-world applications. The experimental findings demonstrate the improved performance of the proposed method in sound zone reproduction within coupled spaces.

5. Conclusion

This paper extends personal audio system from single-space environments to coupled spaces by introducing the virtual quiet zone method. The proposed method converts the challenge of sound zone reproduction in coupled rooms into an optimization problem of the position of the virtual quiet zone, effectively addressing the difficulties related to cross-room ATF measurement. Experimental results confirmed the feasibility of the proposed approach in reducing sound interference between adjacent spaces. Notably, as a transfer function estimation strategy, the proposed method can be quickly adapted to other advanced personal sound zone reproduction approaches, which implies our approach has promising application prospects.

Appendix

The Improved Grey Wolf Optimizer (IGWO) [19] is a population-based optimization algorithm. It iteratively searches for the optimal solution by mimicking the social leadership and hunting behavior of grey wolves. Compared to the original Grey Wolf Optimizer (GWO), the IGWO achieves an improved balance between local and global search while preserving diversity by incorporating a dimension learning-based hunting (DLH) search strategy. The IGWO includes three stages:

Initialization stage: I wolves (candidate positions of the virtual quiet zone) are randomly dispersed within the search space in the range $[lb_j, ub_j]$,

$$X_{ij} = lb_j + rand_j[0,1] \times (ub_j - lb_j), i \in [1, I], j \in [1, D]. \tag{15}$$

$\mathbf{X}_i(t) = \{X_{i1}, X_{i2}, \ldots, X_{iD}\}$ is the position of the ith wolf in the tth iteration, with the dimensionality $D = 2$ as we search for the optimal position of the virtual quiet zone over a plane.

Movement stage: IGWO utilizes both candidate wolves generated by the original GWO, $\mathbf{X}_{i-\mathrm{GWO}}(t+1)$, and the DLH search strategy, $\mathbf{X}_{i-\mathrm{DLH}}(t+1)$, to facilitate the movement of the wolf from its current position $\mathbf{X}_i(t)$ to an improved position $\mathbf{X}_i(t+1)$.

In GWO, the prey encircling behaviour is determined by $\mathbf{X}_\alpha, \mathbf{X}_\beta, \mathbf{X}_\delta$, which are the positions of the top three wolves α, β, δ from the population, that

$$\mathbf{X}_{i-\mathrm{GWO}}(t+1) = \frac{\mathbf{X}_{i1}(t) + \mathbf{X}_{i2}(t) + \mathbf{X}_{i3}(t)}{3}, \tag{16}$$

where

$$\begin{aligned} \mathbf{X}_{i1}(t) &= \mathbf{X}_\alpha(t) - \mathbf{A}_{i1} \times \mathbf{D}_\alpha, \\ \mathbf{X}_{i2}(t) &= \mathbf{X}_\beta(t) - \mathbf{A}_{i2} \times \mathbf{D}_\beta, \\ \mathbf{X}_{i3}(t) &= \mathbf{X}_\delta(t) - \mathbf{A}_{i3} \times \mathbf{D}_\delta. \end{aligned} \tag{17}$$

$$\begin{aligned} \mathbf{D}_\alpha &= |\mathbf{C}_{i1} \times \mathbf{X}_\alpha(t) - \mathbf{X}_i(t)|, \\ \mathbf{D}_\beta &= |\mathbf{C}_{i2} \times \mathbf{X}_\beta(t) - \mathbf{X}_i(t)|, \\ \mathbf{D}_\delta &= |\mathbf{C}_{i3} \times \mathbf{X}_\delta(t) - \mathbf{X}_i(t)|. \end{aligned} \tag{18}$$

The coefficients $\mathbf{A}_{i1}, \mathbf{A}_{i2}, \mathbf{A}_{i3}, \mathbf{C}_{i1}, \mathbf{C}_{i2}, \mathbf{C}_{i3}$ are computed by

$$\mathbf{A} = 2 \times a \times \mathbf{r}_1 - a, \tag{19}$$

$$\mathbf{C} = 2 \times \mathbf{r}_2, \tag{20}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are random vectors that fall within the range of $[0, 1]$, and $a = 2-(2\times t)/T$, where t is the current iteration, T is the maximum number of iterations.

The DLH search strategy assigns a neighbourhood to each wolf and generates $\mathbf{X}_{i-\mathrm{DLH}}(t+1)$. Specifically, the Euclidean distance between the current position $\mathbf{X}_i(t)$ and the candidate position $\mathbf{X}_{i-\mathrm{GWO}}(t+1)$ is

$$R_i(t) = ||\mathbf{X}_i(t) - \mathbf{X}_{i-\mathrm{GWO}}(t+1)||. \tag{21}$$

The neighborhood of $\mathbf{X}_i(t)$ with a radius of $R_i(t)$ is

$$\mathbf{U}_i(t) = \{\mathbf{X}_n(t) | d_E(\mathbf{X}_i(t), \mathbf{X}_n(t)) \leq R_i(t), \mathbf{X}_n(t) \in Pop\}, \tag{22}$$

where $d_E(\mathbf{X}_i(t), \mathbf{X}_n(t))$ represents the Euclidean distance between $\mathbf{X}_i(t)$ and $\mathbf{X}_n(t)$, Pop is an I-row by D-column matrix that stores the positions of the wolves. Subsequently, the candidate position $\mathbf{X}_{i-\mathrm{DLH}}(t+1)$ is

$$\mathbf{X}_{i-\mathrm{DLH}}(t+1) = \mathbf{X}_i(t) + rand \times (\mathbf{X}_n(t) - \mathbf{X}_r(t)), \tag{23}$$

where $\mathbf{X}_n(t)$ is randomly selected from $\mathbf{U}_i(t)$, and $\mathbf{X}_r(t)$ is a random wolf from Pop.

Selecting and updating stage: The IGWO selects and updates the winning candidate wolf in each iteration, based on comparing the fitness values of two candidates

$$\mathbf{X}_i(t+1) = \begin{cases} \mathbf{X}_{i-\mathrm{GWO}}(t+1), & if\, f(\mathbf{X}_{i-\mathrm{GWO}}) > f(\mathbf{X}_{i-\mathrm{DLH}}) \\ \mathbf{X}_{i-\mathrm{DLH}}(t+1), & otherwise \end{cases} \tag{24}$$

To update $\mathbf{X}_i(t+1)$, if the fitness value of $\mathbf{X}_i(t+1)$ is higher than that of $\mathbf{X}_i(t)$, $\mathbf{X}_i(t)$ is replaced; otherwise, it remains unchanged. The IGWO algorithm terminates when it reaches the predefined number of iterations T.

References

[1] Druyvesteyn WF, Garas J. Personal sound. J Audio Eng Soc 1997;45:685-701.

[2] Betlehem T, Zhang W, Poletti MA, Abhayapala TD. Personal sound zones: delivering interface-free audio to multiple listeners. IEEE Signal Process Mag 2015;32:81–91.

[3] Cheer J, Elliott SJ, Kim Y, Choi JW. Practical implementation of personal audio in a mobile device. J Audio Eng Soc 2013;61:290-300.

[4] Vindrola L, Melon M, Chamard JC, Gazengel B. Use of the filtered-x least-mean-squares algorithm to adapt personal sound zones in a car cabin. J Acoust Soc Am 2021;150:1779–93.

[5] Chang JH, Lee CH, Park JY, Kim YH. A realization of sound focused personal audio system using acoustic contrast control. J Acoust Soc Am 2009;125:2091-7.

[6] Choi JW, Kim YH. Generation of an acoustically bright zone with an illuminated region using multiple sources. J Acoust Soc Am 2002;111:1695–700.

[7] Kirkeby O, Nelson PA. Reproduction of plane wave sound fields. J Acoust Soc Am 1993;94:2992–3000.

[8] Cai Y, Wu M, Yang J. Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint. J Acoust Soc Am 2014;135:734–41.

[9] Lee T, Nielsen JK, Christensen MG. Signal-adaptive and perceptually optimized sound zones with variable span trade-off filters. IEEE/ACM Trans Audio Speech Lang Process 2020;28:2412–26.

[10] Wen X, Wu H, Fan Y, Jiang W. A multizone sound field reproduction method with constrained zone acoustic energy in the modal domain. Appl Acoust 2024;220:109959.

[11] Hold C, McKenzie T, Götz G, Schlecht SJ, Pulkki V. Resynthesis of spatial room impulse response tails with anisotropic multi-slope decays. J Audio Eng Soc 2022;70:526-38.

[12] Das O, Abel JS. Grouped feedback delay networks for modeling of coupled spaces. J Audio Eng Soc 2021;69:486-96.

[13] Nijs L, Jansens G, Vermeir G, Van der Voorden M. Absorbing surfaces in ray-tracing programs for coupled spaces. Appl Acoust 2002;63:611-26.

[14] Gül ZS, Eşmebaşı M, Özyurt ZB. The effects of stage house coupling on multipurpose auditorium acoustics. Appl Acoust 2022;198:108996.

[15] McKenzie T, Schlecht SJ, Pulkki V. Acoustic analysis and dataset of transitions between coupled rooms. ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)2021. p. 481-5.

[16] Zhu Q, Coleman P, Wu M, Yang J. Robust acoustic contrast control with reduced in-situ measurement by acoustic modeling. J Audio Eng Soc 2017;65:460–73.

[17] Zhang J, Shi L, Christensen MG, Zhang W, Zhang L, Chen J. CGMM-based sound zone generation using robust pressure matching with ATF perturbation constraints. IEEE/ACM Trans Audio Speech Lang Process 2023;31:3331-45.

[18] Hu M, Shi L, Zou H, Christensen MG, Lu J. Sound zone control with fixed acoustic contrast and simultaneous tracking of acoustic transfer functions. J Acoust Soc Am 2023;153:2538-44.

[19] Nadimi-Shahraki MH, Taghian S, Mirjalili S. An improved grey wolf optimizer for solving engineering problems. Expert Syst Appl 2021;166:113917.

[20] Vergin R, O'Shaughnessy D. Pre-emphasis and speech recognition. In: Proceedings 1995 Canadian Conference on Electrical and Computer Engineering; 1995. p. 1062-5.

[21] Huang B, Cui S, Huang J, Kang X. Discriminative frequency information learning for end-to-end speech anti-spoofing. IEEE Signal Process Lett 2023;30:185-9.

[22] Elliott SJ, Cheer J, Choi JW, Kim Y. Robustness and regularization of personal audio systems. IEEE Trans Audio Speech Lang Process 2012;20:2123-33.

[23] Cox H, Zeskind R, Owen M. Robust adaptive beamforming. IEEE Trans Acoust Speech Signal Process 1987;35:1365-76.

[24] Abe T, Koyama S, Ueno N, Saruwatari H. Amplitude matching for multizone sound field control. IEEE/ACM Trans Audio Speech Lang Process 2023;31:656–69.
[25] Coleman P, Jackson PJB, Olik M, Pedersen JA. Personal audio with a planar bright zone. J Acoust Soc Am 2014;136:1725–35.
[26] López-Espejo I, Edraki A, Chan WY, Tan ZH, Jensen J. On the deficiency of intelligibility metrics as proxies for subjective intelligibility. Speech Commun 2023;150:9-22.
[27] Reidy PF. A comparison of spectral estimation methods for the analysis of sibilant fricatives. J Acoust Soc Am 2015;137:EL248.
[28] Jiao Q, Wang X, Wang L, Bai H. Audio features based ADS-CNN method for flight attitude recognition of quadrotor UAV. Appl Acoust 2023;211:109540.
[29] Farina A. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In: Audio Engineering Society Convention 108; 2000. p. 1-15.
[30] Zhu Q, Qiu X, Coleman P, Burnett I. An experimental study on transfer function estimation using acoustic modelling and singular value decomposition. J Acoust Soc Am 2021;150:3557–68.
[31] Vistopia New Media Editorial Office. The editorial office without ideals. Available from: https://podtail.com/podcast/d6668b34, 2021.