Acoust. Sci. & Tech. **46**, 1 (2025)

PAPER

Stepwise-based optimizing approaches for arrangements of loudspeaker in multi-zone sound field reproduction

Tong Zhou[1,*], Kazuya Yasueda[2], Ghada Bouattour[3],
Anthimos Georgiadis[3] and Akitoshi Kataoka[4]

[1]*Graduate School of Science and Technology, Ryukoku University,*
1–5 Yokotani, Seta Oe-cho, Otsu, 520–2194 Japan
[2]*Tokyo Healthcare University, 3–11–3 Setagaya, Setagaya-ku, Tokyo, 154–8568 Japan*
[3]*Leuphana Universität Lüneburg, Universitätsallee 1, Lüneburg 21335 Germany*
[4]*Ryukoku University, 1–5 Yokotani, Seta Oe-cho, Otsu, 520–2194 Japan*

(*Received 11 May 2024, Accepted for publication 10 September 2024,*
J-STAGE Advance published date: 27 September 2024)

Abstract: This study introduces bidirectional stepwise-based algorithms designed to optimize loudspeaker array configurations for Multizone Sound Field Reproduction systems. An initial arrangement selection method based on loudspeaker magnitude enhances the optimization process. These algorithms were validated using the Acoustic Contrast Control and Pressure Matching methods across free-field conditions and a comprehensive Room Impulse Response database including various room conditions. Comparative experiments against traditional unidirectional iterative strategies demonstrate that the proposed algorithms significantly outperform existing methods in terms of efficiency and effectiveness, especially in configurations with fewer loudspeakers. For example, in a small meeting room with 16 loudspeakers, the stepwise-based approaches achieved higher acoustic contrast and required substantially fewer iterations than conventional methods. Specifically, optimization efficiency improvements were about 55.2% and 77.8% in Acoustic Contrast Control and 36.7% and 68.6% in Pressure Matching, compared to conventional iteratively adding or removing approaches.

Keywords: Multi-zone sound field reproduction, Loudspeaker arrangement optimization, Stepwise, Pressure matching, Acoustic contrast control

1. INTRODUCTION

Multizone Sound Field Reproduction (MSFR), a sound field control technology, generates Personal Sound Zones (PSZs) in shared environments and has garnered significant interest over the past two decades [1–4]. This technology allows for delivering different audio contents to multiple listeners independently and simultaneously by creating multiple bright zones within the same acoustic environment, utilizing a set of loudspeakers to divide the sound field into acoustic bright and dark zones. MSFR has been applied in various settings, including car cabins [5,6], mobile devices [7], and outdoor concerts [8].

Research has explored generating an MSFR system using multiple sources (loudspeakers) through various sound field control methods, such as Acoustic Contrast Control (ACC) [9], Pressure Matching (PM) [10], Mode Matching (MM) [11,12], Amplitude Matching [13,14] and spatial Fourier transform-based approaches [15,16]. ACC and PM are the most widely implemented due to their simplicity and flexible array structure requirements, facilitating adaptation to practical applications. ACC focuses on maximizing the energy ratio between bright and dark zones, whilst PM aims to minimize the sum of reproduction error (RE) in a least-square sense across multiple control points (microphones) in the target zones. Studies have compared these methods, revealing that ACC achieves higher acoustic contrast (AC), whereas PM offers more accurate sound field reproduction in the bright zone [17]. To balance these methods, hybrid approaches like ACC-PM have been introduced, using a weighting factor to optimize evaluation metrics [18–20].

Most studies on ACC and PM utilize specific loudspeaker arrays: linear, circular, X-shaped, etc., based on empirical experience [9,17,21]. However, the arrangement

---

[*] e-mail: t22d501@mail.ryukoku.ac.jp

[doi:10.1250/ast.e24.56]

of loudspeakers critically influences the performance of ACC and PM, as the primary performance metrics (acoustic contrast for ACC and reproduction error for PM) are non-convex functions of loudspeaker positions, complicating the optimization process due to high computational complexity.

Various optimization algorithms have been applied to refine loudspeaker positions, seeking the optimal array structure [22–29]. Among these, iterative approaches are popular for their simplicity and low computational overhead compared to other strategies, such as genetic algorithms [22]. Optimization strategies requiring lower computational complexity have more potential for future applications, such as generating dynamically moving sound zones. Consequently, this study focuses on iterative approaches to optimize loudspeaker arrangements for MSFR.

The discussed iterative approaches adjust the number and configuration of loudspeakers by successively adding or removing them based on specific criteria. For example, Asano *et al.* and Enomoto *et al.* [23,24] determined the array structure by iteratively adding loudspeakers using the independence in transfer functions for active noise control and boundary surface control systems, utilizing Gram-Schmidt orthogonalization (GSO). Khalilian *et al.* [25] employed a Constrained Matching Pursuit (CMP) method to optimize the array by iteratively adding loudspeakers that reduce the RE. Optimizations have also focused on iteratively removing the least significant loudspeaker based on their impact on AC or RE in an Evolutionary Array Optimization (EAO) method [26,27].

However, a common limitation of those iterative approaches is their unidirectional nature, often leading to sub-optimal solutions as the optimization process progresses. The contribution of a loudspeaker can vary significantly if the array configuration changes; loudspeakers that were pivotal in earlier steps might become less impactful as new ones are added, yet the algorithm does not reconsider their removal. Furthermore, these approaches may continue adding/removing loudspeakers to reach a predetermined loudspeaker number requirement, regardless of whether these additions/eliminations enhance overall performance. Additionally, these unidirectional iterative algorithms potentially require extra process steps since they begin the optimization with either one or all candidates selected, without considering the practical requirement for a specific number of loudspeakers. Moreover, all previously mentioned iterative optimization strategies have been performed only in either ACC or PM, their effectiveness has not been universally validated across both methods under identical conditions.

This study introduces a bidirectional, stepwise (SW)-based iterative method that allows for the selection or deselection of loudspeakers as the array configuration evolves. In contrast, this work treats the unidirectional iterative optimization strategies as conventional approaches for comparative analysis. Stepwise-based optimizations are commonly utilized for selecting coefficients in statistical models [30]; however, to the best of the authors' knowledge, this study is the first to apply these methods to the loudspeaker arrangement for sound field reproduction systems. Additionally, an initialization arrangement selection is proposed based on the magnitude of loudspeakers within an array to prepare the optimization. Validation experiments are performed in both ACC and PM contexts in a free field and a published Room Impulse Responses (RIRs) database [31,32], aiming to:

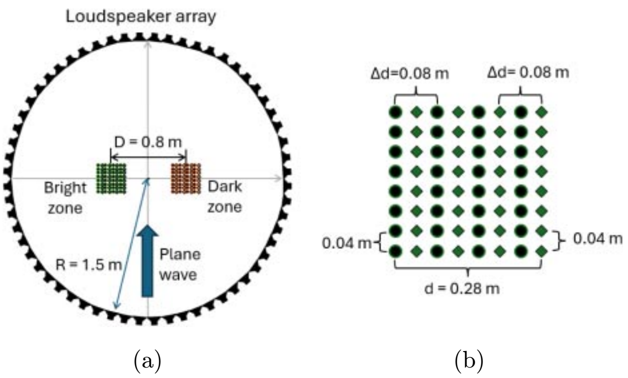
(1) Reproduce satisfactory performance.
(2) Optimize loudspeaker configurations efficiently.
(3) Maintain consistent performance in both ACC and PM.

We use a circular array of loudspeaker candidates as a case study (Fig. 1(a)). Still, the methodology is adaptable to any feasible loudspeaker arrangement due to the non-restrictive nature of ACC and PM concerning loudspeaker positioning. The loudspeakers can theoretically be positioned at any available location to suit practical applications.

The following structure of this paper is: Sect. 2 briefly introduces the theories of ACC and PM. Section 3 details the proposed SW-based algorithm procedure and the evaluation metrics for optimization. Section 4 presents the results of validation experiments. Finally, conclusions and future directions are discussed in Sect. 5.

2. MULTI-ZONE SOUND FIELD CONTROL METHODS

Figure 1 illustrates an MSFR system featuring a circular loudspeaker array and the arrangement of control points (circles) and evaluation points (diamonds) within the



(a) (b)

Fig. 1 Illustration of (a) example of an MSFR system featuring a circular loudspeaker array, (b) the arrangement of control points (circle) and evaluation points (diamond) within a single zone.

bright and dark zones. The synthesized sound pressures at N discrete points by M loudspeakers in the frequency domain can be formulated as:

$$\mathbf{p}(\omega) = \mathbf{G}(\omega)\mathbf{d}(\omega), \tag{1}$$

where $\mathbf{p}$ is an N-length column vector of synthesized sound pressures, $\mathbf{d}$ is an M-length column vector of the loudspeakers' driving signals, and $\mathbf{G}$ is an N × M matrix representing the transfer function from the loudspeakers to the control points. ω represents the angular frequency; this term will be omitted for simplicity in subsequent discussions. In a free-field environment, the transfer function from a loudspeaker to a control point can be obtained using Green's function

$$g_{nm} = \frac{e^{-jkr_{nm}}}{4\pi r_{nm}}, \tag{2}$$

where m and n are the indices of the loudspeakers and control points, k is the wavenumber, and r_{nm} denotes the distance between the corresponding loudspeaker and control point. j represents the imaginary unit. In real-world environments, the transfer function can also be measured as the impulse response [33].

2.1. Acoustic Contrast Control Method

The ACC method was first formulated by literature [9] to focus potential acoustic energy in a bright zone while minimizing it in a dark zone. The ACC method estimates the desired sound zones by maximizing the energy ratio between these zones. The corresponding cost function for ACC is expressed as:

$$J_{ACC}(\mathbf{d}) = \frac{\mathbf{p}_B^H\mathbf{p}_B}{\mathbf{p}_D^H\mathbf{p}_D} = \frac{\mathbf{d}^H\mathbf{G}_B^H\mathbf{G}_B\mathbf{d}}{\mathbf{d}^H\mathbf{G}_D^H\mathbf{G}_D\mathbf{d}}, \tag{3}$$

where $(*)^H$ denotes the Hermitian (complex conjugate) transpose, subscripts B and D indicate variables associated with the bright and dark zones, respectively. The driving signal for maximum acoustic potential energy in ACC is proportional to the maximum eigenvalue of $(\mathbf{G}_D^H\mathbf{G}_D)^{-1}(\mathbf{G}_B^H\mathbf{G}_B)$. Therefore, the optimal driving signal $\hat{\mathbf{d}}_{ACC}$ can be derived as:

$$\hat{\mathbf{d}}_{ACC} \propto \phi(\mathbf{d})[(\mathbf{G}_D^H\mathbf{G}_D + \lambda_1\mathbf{I})^{-1}(\mathbf{G}_B^H\mathbf{G}_B)], \tag{4}$$

where $\phi(*)$ denotes the function calculating the maximum eigenvalue, λ_1 is the regularization parameter.

2.2. Pressure Matching Method

The PM method is a sound field control method that aims to minimize the RE between the synthesized and desired sound pressure at multiple discrete control points within target areas. The RE is quantified as:

$$J_{PM}(\mathbf{d}) = \|\mathbf{G}\mathbf{d} - \mathbf{p}^{des}\|_2^2, \tag{5}$$

where $\| * \|_2$ denotes the l_2-norm operator, $\mathbf{p}^{des} = [\mathbf{p}_B^{des}, \mathbf{p}_D^{des}]^T$ represents the desired sound pressure in both bright and dark zones over control points. The desired plane wave sound field in the bright zone can be written as

$$p_{B_n}^{des} = Se^{jk\mathbf{r}_n\cdot\mathbf{u}_\varphi}, \tag{6}$$

n denotes the index of the control point. S is the amplitude of the incoming plane wave, $\mathbf{r}_n$ denotes the position of the n-th control point, $\mathbf{u}_\varphi$ is the unit vector in the direction of the incoming plane wave. For control points in the dark zone, the desired sound pressure $\mathbf{p}_D^{des} = \mathbf{0}$. Ideally, the minimum solution of Eq. (5) is zero, which leads to the driving signal obtainable if the inverse of the transfer function $\mathbf{G}^{-1}$ is known. Therefore, the inverse filter of the array can be represented as

$$\mathbf{d} = \mathbf{G}^{-1}\mathbf{p}^{des}. \tag{7}$$

However, in many scenarios $\mathbf{G}$ becomes ill-conditioned, and $\mathbf{G}^{-1}$ does not exist. To address this issue, a widely used regularized least squares method is employed to estimate the filter for PM. The equation can be reformulated as:

$$\hat{\mathbf{d}}_{PM} = (\mathbf{G}^H\mathbf{G} + \lambda_2\mathbf{I})^{-1}\mathbf{G}^H\mathbf{p}^{des}, \tag{8}$$

where λ_2 is the Tikhonov regularization parameter, $\mathbf{I}$ is an $M \times M$ identity matrix.

Equations (4) and (8) can only optimize the filters to generate the desired multizone if the array configurations are determined. However, it is known that the relative positions of the loudspeaker and control points can influence the synthesized sound field, as indicated in Eq. (2). Since target bright and dark zone locations are usually fixed, finding an optimal arrangement of the loudspeaker array is crucial for system performance. To enhance the efficacy of the desired multi-zone, this work proposes optimizing the array arrangement by a bidirectional stepwise approach.

3. OPTIMIZATION: STEPWISE-BASED APPROACHES

Stepwise (SW) algorithms represent a class of iterative optimization methods that develop a solution incrementally, adding or removing variables or elements step-by-step [34]. These approaches are generally categorized into forward selection, backward elimination, and stepwise selection (bidirectional elimination). Forward selection involves selecting the most contributive terms by attempting to add each potential element individually in each step. Conversely, backward elimination starts with all elements included and optimizes by continuously removing the least contributive term at each step.

Conceptually, the optimization approaches described in literature [23–25] can be classified as forward selection, whereas those in literature [26,27] align with backward elimination. These methods manipulate parameters by successively adding significant loudspeakers or eliminating less significant ones, each guided by differing cost functions. However, in the context of loudspeaker configuration optimization, a loudspeaker's contribution may vary if the overall configuration changes. This variability suggests that both forward and backward methods, owing to their unidirectional nature of either constantly adding or removing elements, may increasingly risk settling on suboptimal solutions as the optimization process progresses. In light of these considerations, this study proposes adopting a bidirectional stepwise-based approach for optimizing loudspeaker arrangements. This stepwise selection method integrates both forward and backward operations, allowing for a previously excluded term to be reconsidered and reselected in subsequent steps. Theoretically, this bidirectional approach offers a more extensive search space and a higher likelihood of identifying the optimal solution.

3.1. Optimization Procedure of SW-based Algorithms

One procedure of the proposed SW-based algorithm is detailed in Algorithm 1. The optimization process begins with establishing an initial loudspeaker arrangement. This arrangement is inspired by the findings of the importance of loudspeakers with significant strength magnitudes, as reported in literature [27]. Accordingly, this study adopts the strength magnitude of each loudspeaker within the array as the criterion for initial selection. For instance, in the PM method, the strength of each loudspeaker is calculated using Eq. (8) when all candidates are employed. Subsequently, a fixed number of loudspeakers with the greatest magnitudes are selected to form the initial array.

Once the initial array is established, the optimization progresses by continuously evaluating the contributions of loudspeakers both in the candidate pool and the selected set. This evaluation consists of a *forward* operation, where the most contributing loudspeaker from the candidate pool is added, and a *backward* operation, where the least contributing loudspeaker from the current set is removed. During a *forward* operation, if adding a loudspeaker from the candidate pool results in the most significant performance improvement, that loudspeaker is identified as the most contributing one. Conversely, during a *backward* operation, if removing a loudspeaker from the selected set causes the least reduction in performance, that loudspeaker is identified as the least contributing. The optimization process is iterative and concludes when the array structure stabilizes, specifically when a loudspeaker is repeatedly added and then removed in successive *forward* and *backward* operations. This study examines the impact of the sequence in which these operations are conducted by investigating two scenarios. In the first scenario, labeled SW-1 as detailed in Algorithm 1, the *forward* precedes the *backward* operation. Conversely, in the second scenario, labeled SW-2, the order of operations is reversed. For SW-2, the algorithm adheres to the same procedural steps as SW-1 but with the *forward* and *backward* operations (steps 4 and 5) interchanged.

Algorithm 1 SW-1

```
Input:
$K$: Desired loudspeaker number.
$\mathbf{C}$: Loudspeaker candidate pool.
Output: Optimized loudspeaker arrangement $\mathbf{L}$.
1: procedure STEPWISE($K, \mathbf{C}$)
2:   Initialization: Select $K$ loudspeakers from $\mathbf{C}$ to form the initial arrangement $\mathbf{L}$
3:   while $l_{m_1} \neq l_{m_2}$ do
4:     forward step: Add the most contributing loudspeaker $l_{m_1}$ from $\mathbf{C}$ to $\mathbf{L}$
5:     backward step: Remove the least contributing loudspeaker $l_{m_2}$ from $\mathbf{L}$ to $\mathbf{C}$
6:   end while
7:   return $\mathbf{L}$
8: end procedure
```

3.2. Evaluation Metrics

Two cost functions are used for the optimization process: AC for the ACC method and RE for the PM method, as defined in Eq. (3) and Eq. (5), respectively. For evaluating the results of optimized loudspeaker arrangements across different algorithms, the following metrics are employed:

1. Acoustic Contrast (AC) Between Zones: This metric is used to assess the performance of both ACC and PM. It focuses on the contrast in acoustic energy between the zones and is defined as

$$\mathbf{AC} = 10\log_{10}\frac{\|\mathbf{p}_B^{syn}\|_2^2}{\|\mathbf{p}_D^{syn}\|_2^2}, \tag{9}$$

where $\mathbf{p}_B^{syn}$ and $\mathbf{p}_D^{syn}$ denote the sum of synthesized sound pressures at evaluation points in the bright and dark zones, respectively, using the filters obtained by Eq. (4) for ACC and Eq. (8) for PM. Specifically for ACC, only the AC is used to evaluate its performance because this method primarily controls the sound field by manipulating acoustic energy alone, without explicitly addressing the fidelity of sound reproduction, as does PM.

2. Mean Square Error (MSE) in the Bright Zone: Specifically used for evaluating the PM performance, this metric assesses the fidelity of sound reproduction in the bright zones compared to the desired sound pressure. It is formulated as:

$$\mathbf{MSE} = 10\log_{10}\frac{\|\mathbf{p}_B^{syn} - \mathbf{p}_B^{des}\|_2^2}{\|\mathbf{p}_B^{des}\|_2^2}. \qquad (10)$$

3. Iterations: Additionally, the efficiency of different optimization approaches is compared using the number of iterations, which indicates how many array configurations are tested during the optimization process for a certain number of loudspeakers. For conventional Forward and Backward strategies, the number of iterations is predetermined based on the desired loudspeakers number K and the size of candidate pool M. Specifically:

- **Forward Approach:** $\sum_{sum=M-K+1}^{M} sum$
- **Backward Approach:** $\sum_{sum=K+1}^{M} sum$
- **Global Search:** $\frac{M!}{K!(M-K)!}$

For example, selecting $K = 8$ from $M = 60$ results in 452 iterations for the Forward strategy and 1,795 for the Backward strategy, compared to approximately 2.6×10^9 in a global search. With $K = 16$, the iterations are 840, 1,695, and about 1.5×10^{14}, respectively.

However, the iteration number required for bidirectional SW-based algorithms cannot be expressed by a formula; rather, the iterations for individual *forward* and *backward* operations are fixed. For example, in SW-1, a single *forward* operation requires $M - K$ iterations, followed by a *backward* operation that requires $K + 1$ iterations. Similarly, SW-2 starts with a *backward* operation requiring K iterations, then a *forward* operation requiring $M - K + 1$. Therefore, each step toward becoming "wiser" involves $M + 1$ iterations, and the total number of iterations for SW-based approaches can be expressed as $1 + x(M + 1)$, where the first "1" accounts for the initial arrangement, x represents the number of cycles of *forward*/*backward* operations.

All the cost functions and evaluation metrics are applied at single frequencies, meaning the arrangements might differ when optimizing in different frequencies. Extending these metrics into a broadband context can be achieved by averaging each criterion over the frequency range, thereby accommodating variations in frequency response.

4. VALIDATION AND RESULTS

Validation experiments were conducted using ACC and PM methods to assess the efficacy of the proposed bidirectional SW-based optimization approach for an MSFR system. The experimental setup, depicted in Fig. 1, consists of a circular array of $M = 60$ loudspeakers with a radius $R = 1.5$ m; The target bright and dark zones are defined as $d = 0.28$ m squares, separated by a center-to-center distance of $D = 0.8$ m. The setup includes $N = 64$ microphones distributed at 0.04 m intervals within each zone, designated for control and evaluation purposes to prevent bias (as detailed in Fig. 1(b)). This microphone arrangement sets a spatial aliasing limit frequency of $f = \frac{c}{2\Delta d} = 2{,}125$ Hz, assuming a speed of sound $c = 340$ m/s. The input to the system comprises single-frequency plane waves with unit amplitude, originating from the direction illustrated in Fig. 1(a).

The validation of the proposed bidirectional stepwise-based (SW-based) optimization algorithms began in a free field setting, using the transfer function obtained by Eq. (2) to evaluate algorithm performance under ideal conditions. Subsequently, another validation was carried out using a multi-zone sound field reproduction Room Impulse Responses (RIRs) database published by UTS [31,32], which provided a real-environment test scenario. This database encompasses 260400 RIRs recorded across seven different room types. This paper presents validation results for three representative room types: an anechoic chamber, a small meeting room, and an open space to maintain conciseness. The sampling rate was set to 48 kHz, with the impulse response data fixed at 32,768 samples, resulting in a frequency resolution of approximately 1.465 Hz and integer frequency steps of 375 Hz. A frequency of 1,125 Hz was selected as the main evaluation frequency because it is an integer value within a practical range, neither too low nor too high, making it a suitable representative for our evaluations. The mean reverberation times (RTs) for these rooms are detailed in Table 1. Notably, the Open space is characterized by two different RTs due to its large, irregular polygonal shape, unlike the standard rectangular configuration typical of the other rooms [31].

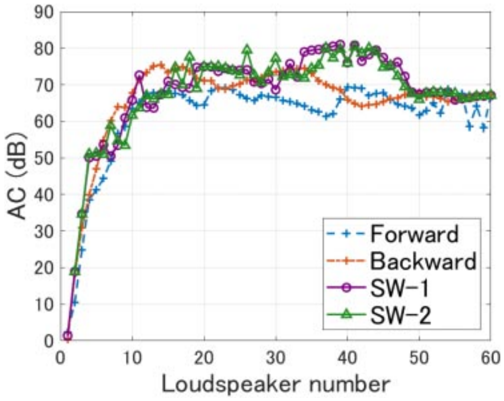
Additionally, as noted in prior studies [21], the regularization parameter significantly impacts the performance and robustness of an MSFR system. To establish a reference AC level as a baseline for comparison, the selections of λ_1, λ_2 for ACC and PM follow [31], set as $10^{-5} \times \alpha_{1,2}$. Here, α_1 and α_2 represent the maximum eigenvalues of the matrices $\mathbf{G}_D^H\mathbf{G}_D$ (Eq. (3)) and $\mathbf{G}^H\mathbf{G}$ (Eq. (8)), respectively.

4.1. Results in Free-field

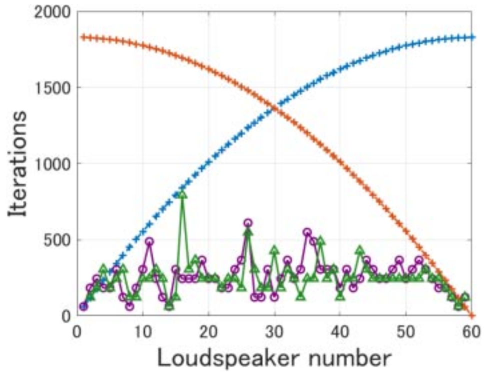
Figures 2 and 3 display the results of loudspeaker selections using different optimization algorithms in the ACC and PM contexts at 1,125 Hz in free-field, respectively. It is observed that the performance in terms of both AC and MSE tends to stabilize beyond a certain number of loudspeakers, 10 in ACC and 20 in PM, achieving

Table 1 Mean RT of each room.

Room type	Mean RT (s)
1. Free field	0.000
2. Anechoic chamber	0.045
3. Small meeting room	0.497
4. Open space	0.105, 0.998



(a) AC



(b) Iterations

Fig. 2 AC and iteration comparisons versus loudspeaker number in ACC in free-field at 1,125 Hz.

performance comparable to that when all 60 loudspeakers are deployed. This trend aligns with findings from other studies on loudspeaker arrangement optimization [26,27], underscoring the importance of identifying an optimal number and configuration of loudspeakers.

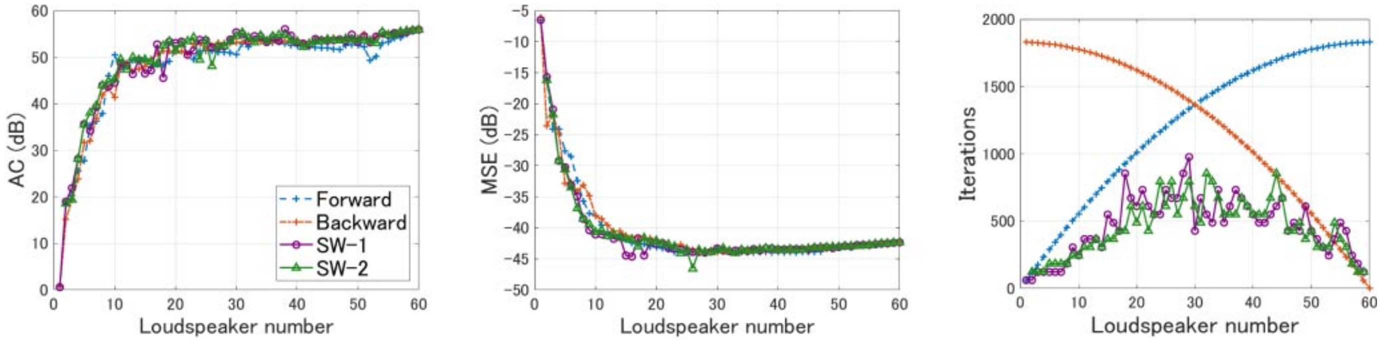
It can be observed from Fig. 2 that the proposed two SW-based algorithms outperformed AC and efficiency in overall loudspeaker numbers using ACC in a free field. Figure 3 indicates that no algorithm with the PM method demonstrates a distinct advantage regarding AC levels in the free field at 1,125 Hz. However, the proposed SW-based optimization achieved lower MSE than the traditional methods when 6–16 loudspeakers were utilized. Additionally, iterative efficiency is significantly improved.

Notably, a lower MSE in the bright zone did not always correlate with higher AC between zones, as exemplified by the results of SW-2 when $K = 26$. This discrepancy arises because the driving signal and evaluation results are derived from different microphone positions, as illustrated in Fig. 1(b). This phenomenon was not observed when the same microphones were used for both control and evaluation, and those results are omitted here for brevity.

4.2. Results on a RIRs Database

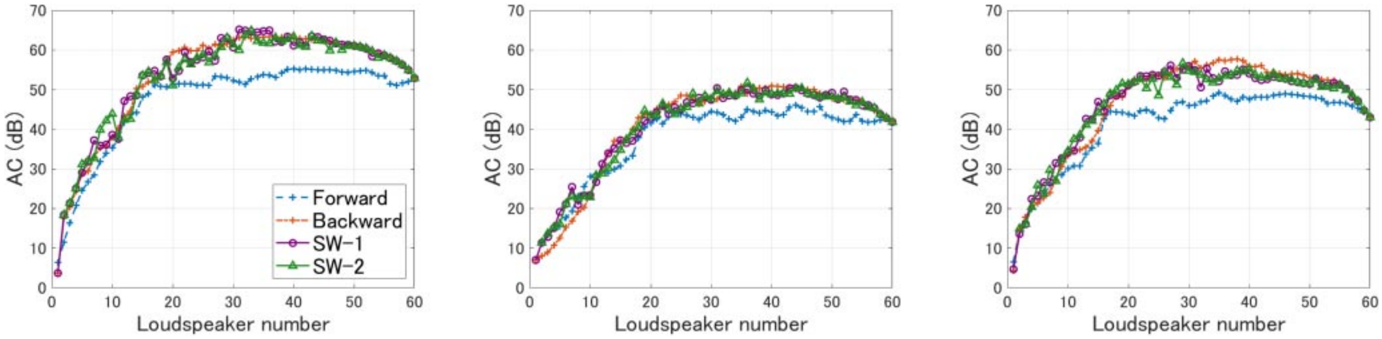
Figure 4 illustrates that the trend of AC performance at 1,125 Hz relative to the number of selected loudspeakers remains consistent across different room types within the ACC method. The AC levels showed little variation after selecting more than 20 loudspeakers in all optimization approaches. While the Forward approach generally found sub-optimal solutions across all loudspeaker numbers, SW-1 and SW-2 excelled, particularly when fewer than 20 were used. This is advantageous since fewer loudspeaker requirements are more favorable in practical applications. Conversely, Fig. 4(c) suggests the Backward approach yielded lower AC with fewer loudspeakers but achieved slightly higher AC when more than 30 were engaged in the open space.

The performance outcomes in PM are depicted in Fig. 5, demonstrating that the proposed SW-based algorithms consistently outperformed in terms of MSE performance overall loudspeaker numbers in all room types. In contrast, the Backward optimization approach yielded sub-optimal solutions in all three room types when less than 20 loudspeakers were utilized. This tendency is anticipated, as the Backward algorithm begins with all loudspeakers in place and may inadvertently eliminate



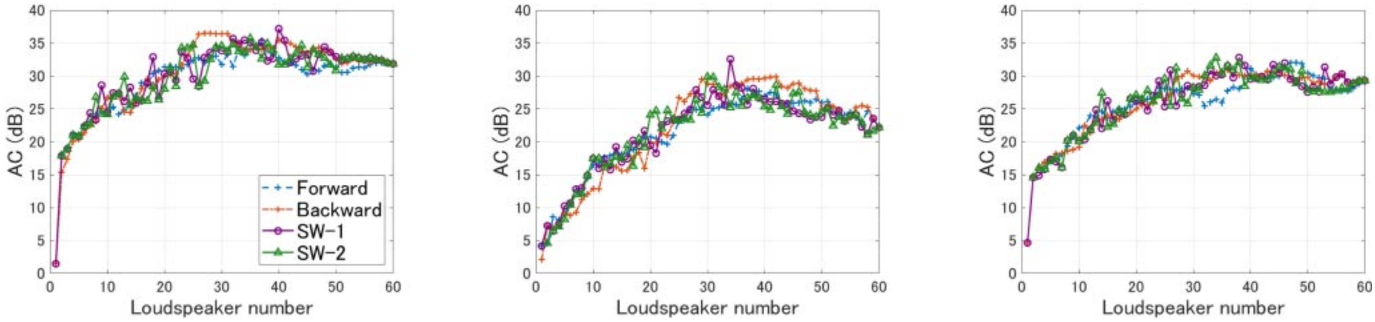
(a) AC (b) MSE (c) Iterations

Fig. 3 AC, MSE, and iteration comparisons versus loudspeaker number in PM in free-field at 1,125 Hz.

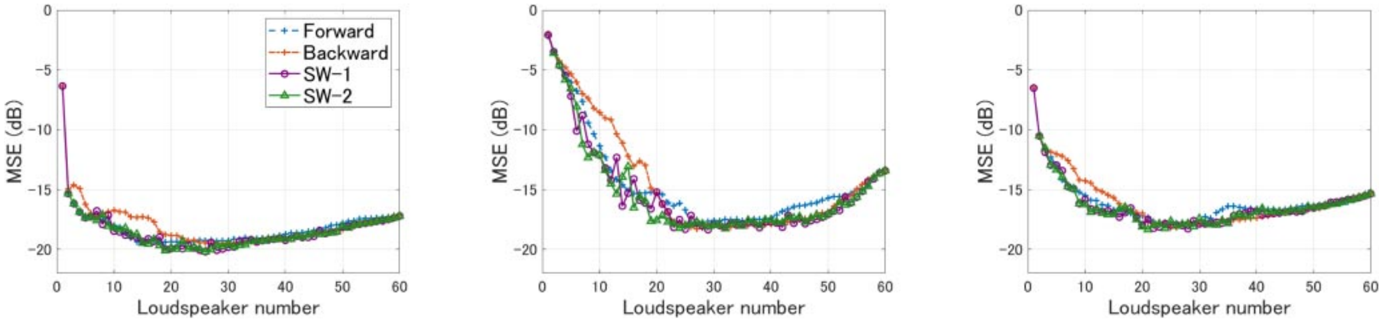


(a) Anechoic chamber (b) Small meeting room (c) Open space

Fig. 4 Comparison of AC level of different room types in ACC at 1,125 Hz.



(a) AC in Anechoic Room (b) AC in Small Meeting Room (c) AC in Open Space



(d) MSE in Anechoic chamber (e) MSE in Small meeting room (f) MSE in Open space

Fig. 5 Comparison of AC and MSE of different room types in PM at 1,125 Hz.

crucial loudspeakers early in the optimization process, thus reducing its effectiveness in subsequent steps. Regarding AC levels within PM, the proposed SW-based algorithms exhibited comparable performance with less than 20 loudspeakers selected across all room conditions. In a Small meeting room, the Backward algorithm had the lowest AC with fewer loudspeakers; it slightly improved and achieved higher AC when the selection was increased to more than 24 loudspeakers.

The iteration numbers used to quantify computational complexity during the optimization process in different room types are presented in Fig. 6 and Table 2. These results demonstrate that the proposed SW-based optimizations highly enhanced efficiency across most loudspeaker configurations, regardless of room type. Specifically, as shown in Table 2, with 16 loudspeakers at 1,125 Hz, the iteration numbers indicate that both SW-1 and SW-2 algorithms substantially improved optimization efficiency in the ACC; on average, efficiency improvements were approximately 60% and 80% over the Forward and Backward strategies, respectively. In PM, the improvements were similarly notable, with average enhancements of 47% and 73% compared to the Forward and Backward strategies. These findings underscore the effectiveness of the SW-based approaches in reducing computational demands while maintaining or enhancing performance.

The performance on AC and optimization efficiency of different algorithms in ACC and PM are given in Figs. 7 and 8, focusing on scenarios employing 16 loudspeakers in a small meeting room. Validation encompassed the

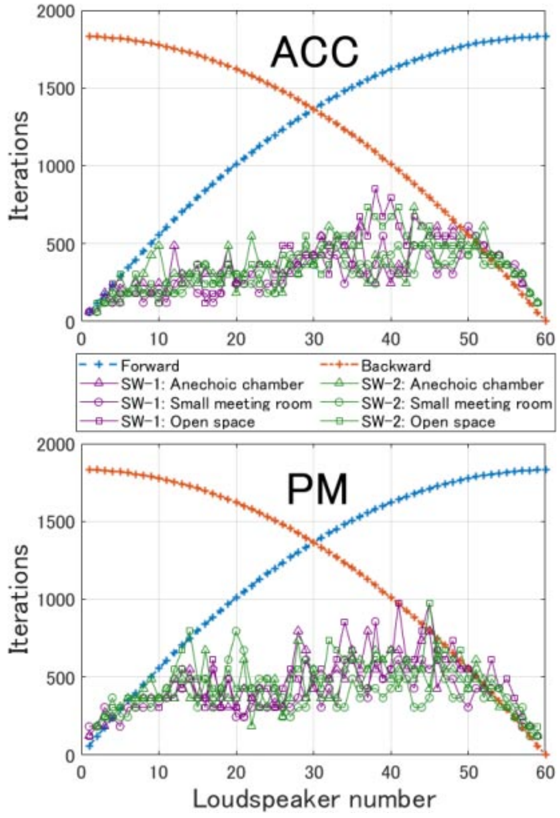


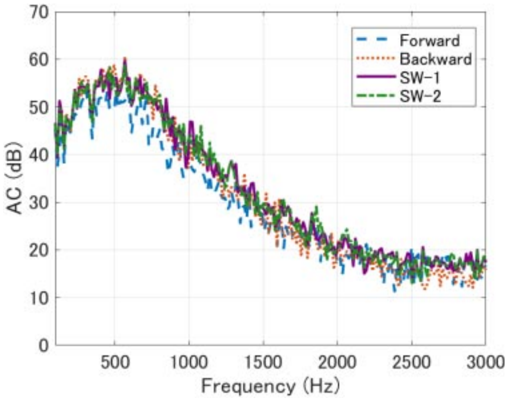
Fig. 6 Iteration comparison in ACC and PM of different room types at 1,125 Hz.

Table 2 Iterations with 16 loudspeakers were selected at 1,125 HZ in different room types. 1: free field, 2: anechoic chamber, 3: small meeting room, 4: open space.

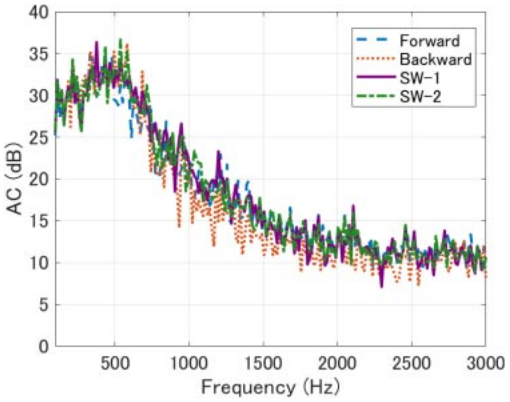
Method		ACC				PM			
Room type		1	2	3	4	1	2	3	4
Algo.	Forward	840							
	Backward	1,695							
	SW-1	245	306	184	123	489	367	367	306
	SW-2	794	367	367	245	367	672	489	489
Average of SW		328.875				443.250			

frequency range from 100 Hz to 3,000 Hz, covering the majority of the speech spectrum. This range is crucial for ensuring the speech messages are clearly understood. It is particularly relevant for applications where high speech quality is not critical, such as delivering zone-specific audio information in museums or employing sound zoning in open-plan offices to selectively mask speech frequencies and enhance privacy.

Figure 7 suggests that the conventional Forward approach resulted in lower AC at frequencies up to 1,500 Hz in ACC, while the Backward approach demonstrated lower AC in PM at frequencies above approximately 750 Hz. In contrast, the proposed SW-based approaches consistently achieved higher AC across the entire frequency range of interest, irrespective of the sound field control method. Additionally, Fig. 8 indicates that the SW-based approaches significantly reduced the number of iterations required during the optimization process, particularly in ACC. On average, the SW-based methods required approximately 376 iterations in ACC and 532 in PM across the entire frequency band of interest. Compared with the conventional Forward and Backward strategies, efficiency improvements were notable, about 55.2% and 77.8% in ACC, 36.7% and 68.6% in PM, respectively. These results underscore the superior efficiency and effectiveness of the SW-based approaches over traditional unidirectional iterative optimization strategies.



(a) ACC

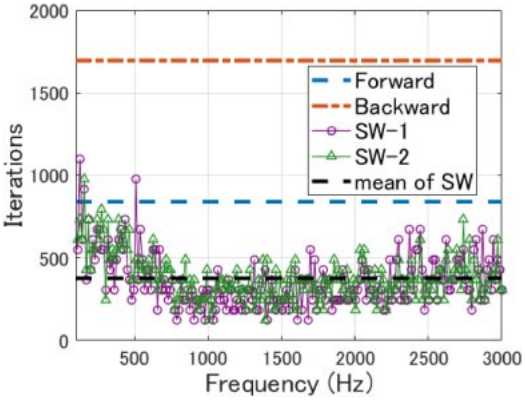


(b) PM

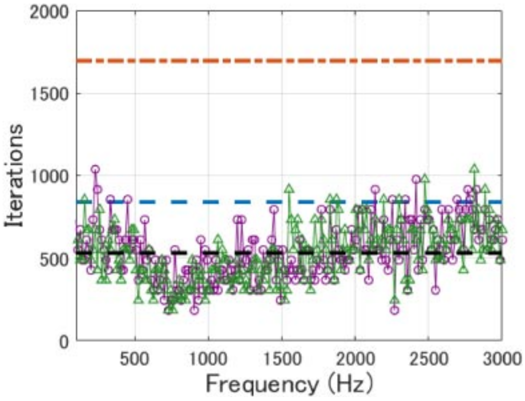
Fig. 7 AC level comparison versus frequency when 16 loudspeakers were selected in a small meeting room.

5. CONCLUSIONS

This study introduced bidirectional stepwise (SW)-based algorithms designed to optimize loudspeaker array configurations for Multizone Sound Field Reproduction (MSFR) systems. An initial arrangement based on the strength magnitude of loudspeakers, when all candidates are active, was also proposed to prepare for the optimiza-



(a) ACC



(b) PM

Fig. 8 Iteration comparison versus frequency when 16 loudspeakers are selected in a small meeting room.

tion process. To evaluate the effectiveness of these approaches, extensive experiments were conducted across different settings, including Acoustic Contrast Control (ACC) and Pressure Matching (PM), in free field conditions and a Room Impulse Responses (RIR) database, encompassing an anechoic chamber, a small meeting room, and an open space.

The validation experiments explored two operational orders of the proposed bidirectional SW-based algorithms. Results consistently showed that both configurations of the SW-based approach significantly outperformed traditional unidirectional iterative strategies. In contrast to the Forward approach, which often resulted in sub-optimal solutions in ACC, and the Backward approach, which led to higher Mean Square Error (MSE) in PM, the SW-based algorithms demonstrated superior performance, especially when fewer loudspeakers were utilized. Detailed performance comparisons in a small meeting room with 16 loudspeakers across various frequencies further substantiated that the SW-based strategies excelled in both acoustic contrast and optimization efficiency. These findings underscore the potential of bidirectional optimization strategies to significantly improve sound field control's practical applicability and effectiveness in MSFR systems. The SW-based approaches were particularly advantageous in scenarios where fewer loudspeakers are preferable, achieving an optimal balance between performance and computational efficiency.

Future research will expand these validations to broadband frequencies and incorporate additional metrics, such as array effort, to further enhance the robustness and versatility of these optimization techniques in diverse acoustic environments. Investigations will also be conducted to evaluate the impact of varying control positions on the performance of these techniques. Furthermore, the application of these methods in real-world scenarios will be systematically explored to validate their practical utility and effectiveness.

ACKNOWLEDGEMENT

The authors would like to acknowledge Kohei Asada for his invaluable advice on this study. This research was partly supported by the ICOM Foundation, Japan.

REFERENCES

[1] W. F. Druyvesteyn and J. Garas, "Personal sound," *J. Audio Eng. Soc.*, **45**, 685–701 (1997).

[2] M. Poletti, "An investigation of 2D multizone surround sound systems," *Proc. 125th Audio Eng. Soc. Conv.*, San Francisco, CA (Audio Engineering Society, New York), pp. 167–175 (2008).

[3] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Trans. Audio Speech Lang. Process.*, **19**, 1711–1720 (2011).

[4] T. Betlehem, W. Zhang, M. A. Poletti and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, **32**, 81–91 (2015).

[5] J. Cheer, S. J. Elliott and M. F. S. Galvez, "Design and implementation of a car cabin personal audio system," *J. Audio Eng. Soc.*, **61**, 412–424 (2013).

[6] X. Liao, J. Cheer, S. J. Elliott and S. Zheng, "Design of a loudspeaker array for personal audio in a car cabin," *J. Audio Eng. Soc.*, **65**, 226–238 (2017).

[7] J. Cheer, S. J. Elliott, Y. Kim and J. W. Choi, "Practical implementation of personal audio in a mobile device," *J. Audio Eng. Soc.*, **61**, 290–300 (2013).

[8] J. Brunskog, F. Heuchel, D. C. Nozal, M. Song, F. Agerkvist, E. Fernandez-Grande and E. Gallo, "Full-scale outdoor concert adaptive sound field control," *Proc. 23rd Int. Congr. Acoust.*, pp. 1170–1177 (2019).

[9] J. W. Choi and Y. H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.*, **111**, 1695–1700 (2002).

[10] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave soundfields," *J. Acoust. Soc. Am.*, **94**, 2992–3000 (1993).

[11] M. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc.*, **53**, 1004–1025 (2005).

[12] J. Zhang, W. Zhang, T. D. Abhayapala and L. Zhang, "2.5D multizone reproduction using weighted mode matching: Performance analysis and experimental validation," *J. Acoust. Soc. Am.*, **147**, 1404–1417 (2020).

[13] S. Koyama, T. Amakasu, N. Ueno and H. Saruwatari, "Amplitude matching: Majorization-minimization algorithm for sound field control only with amplitude constraint," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* (*ICASSP*) *'21*, pp. 411–415 (2021).

[14] T. Abe, S. Koyama, N. Ueno and H. Saruwatari, "Amplitude matching for multizone sound field control," *IEEE/ACM Trans. Audio Speech Lang. Process.*, **31**, 656–669 (2023).

[15] T. Okamoto, "Generation of multiple sound zones by spatial filtering in wavenumber domain using a linear array of loudspeakers," *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.* (*ICASSP*) *'14*, pp. 4733–4737 (2014).

[16] T. Okamoto, "Analytical methods of generating multiple sound zones for open and baffled circular loudspeaker arrays," *Proc. IEEE Workshop Applications of Signal Processing to Audio and Acoustics* (*WASPAA*), pp. 1–5 (2015).

[17] P. Coleman, P. J. B. 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.*, **135**, 1929–1940 (2014).

[18] 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.*, **135**, 734–741 (2014).

[19] J. H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *J. Acoust. Soc. Am.*, **131**, 4518–4525 (2012).

[20] J. H. Chang and F. Jacobsen, "Experimental validation of sound field control with a circular double-layer array of loudspeakers," *J. Acoust. Soc. Am.*, **133**, 2046–2054 (2013).

[21] P. Coleman, P. J. Jackson, M. Olik, M. Olsen, M. Møller and J. A. Pedersen, "The influence of regularization on anechoic performance and robustness of sound zone methods," *Proc. Meet. Acoust.*, Vol. 19, 055055 (2013).

[22] T. Zhou, K. Yasueda and A. Kataoka, "Optimization of parameters of control point and loudspeaker for quiet zone in genetic algorithm and multi-point control method," *Proc. 24th Int. Congr. Acoust.* (2022).

[23] F. Asano, Y. Suzuki and D. C. Swanson, "Optimization of control source configuration in active control systems using Gram-Schmidt orthogonalization," *IEEE Trans. Speech Audio Process.*, **7**, 213–220 (1999).

[24] S. Enomoto, Y. Ikeda, S. Ise and S. Nakamura, "Optimization of loudspeaker and microphone configurations for sound reproduction system based on boundary surface control principle," *Proc. 20th Int. Congr. Acoust.*, 7 pages (2010).

[25] H. Khalilian, I. V. Bajić and R. G. Vaughan, "Comparison of loudspeaker placement methods for sound field reproduction," *IEEE/ACM Trans. Audio Speech Lang. Process.*, **24**, 1364–1379 (2016).

[26] M. Zhu and S. Zhao, "An iterative approach to optimize loudspeaker placement for multi-zone sound field reproduction," *J. Acoust. Soc. Am.*, **149**, 3462–3468 (2021).

[27] S. Zhao and I. S. Burnett, "Evolutionary array optimization for multizone sound field reproduction," *J. Acoust. Soc. Am.*, **151**, 2791–2801 (2022).

[28] T. Zhou, K. Yasueda and A. Kataoka, "Enhancing acoustic contrast in multi-zone sound field reproduction through optimizing loudspeaker arrangements," *Proc. InterNoise '23*, pp. 673–682 (2023).

[29] 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.*, **28**, 696–714 (2020).

[30] M. J. Whittingham, P. A. Stephens, R. B. Bradbury and R. P. Freckleton, "Why do we still use stepwise modelling in ecology and behaviour?" *J. Anim. Ecol.*, **75**, 1182–1189 (2006).

[31] S. Zhao, Q. Zhu, E. Cheng and I. S. Burnett, "A room impulse response database for multizone sound field reproduction (L)," *J. Acoust. Soc. Am.*, **152**, 2505–2512 (2022).

[32] S. Zhao, Q. Zhu, E. Cheng and I. S. Burnett, "Erratum: A room impulse response database for multizone sound field reproduction (L) [J. Acoust. Soc. Am. 152(4), 2505–2512 (2022)]," *J. Acoust. Soc. Am.*, **155**, 2170 (2024).

[33] Y. Suzuki, F. Asano, H. Y. Kim and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, **97**, 1119–1123 (1995).

[34] M. A. Efroymson, "Multiple regression analysis," in A. Ralston and H. S. Wilf, Eds., *Mathematical Methods for Digital Computers* (John Wiley & Sons, New York, 1960).

Tong Zhou received the Bachelor's degree in Engineering from Anhui University Jianghuai College in 2014 and the Master's in Engineering from Ryukoku University in 2022. She is currently a doctoral student at the Graduate School of Science and Technology, Ryukoku University. Her research interests include sound field reproduction, audio signal processing, and optimization. She is a student member of the Acoustical Society of Japan (ASJ), the Acoustical Society of America (ASA), and the Audio Engineering Society (AES).

Kazuya Yasueda received the B.E., M.E., and Ph.D in engineering from Ryukoku University, Shiga, Japan, in 2013, 2015, and 2020, respectively. He is currently an Assistant Professor in the Division of Healthcare Informatics, Faculty of Healthcare, Tokyo Healthcare University. His research interests include sound field control, array signal processing, and speech privacy for public space. He is a member of ASJ, the Institute of Electronics, Information and Communication Engineers (IEICE), and the Japan Association for Medical Informatics (JAMI).

Ghada Bouattour has been a junior professor of Measurement and Sensor Technology in Production Engineering at Leuphana University Lüneburg since 2023. She holds a Ph.D. in Electrical Engineering from the Technical University of Chemnitz and a Master's in Embedded Systems from the National Engineering School of Sfax, Tunisia. Her research primarily explores smart sensor technology, including autonomous wireless sensors, inductive power transfer systems, and vibration energy harvesting. She has contributed to numerous publications and key projects in these areas.

Anthimos Georgiadis graduated in physics from RWTH Aachen University and the University of Cologne in 1980. In 1985, he was awarded a doctorate (Dr. rer. nat.). His research interests include laser, sensor, and measurement technologies. He is currently a Professor of metrology and intelligent systems at the Institute for Production Technology and Systems, Leuphana University Lüneburg.

Akitoshi Kataoka received the B.E., M.E., and Ph.D. in Electrical Engineering from Doshisha University in Kyoto in 1984, 1986, and 1999, respectively. Since joining NTT Laboratories in 1986, he has been engaged in research on noise and reverberation reduction, as well as medium bitrate speech coding algorithms. He contributed to the establishment of the ITU-T G.729 standards. In 2008, he joined the Faculty of Science and Technology at Ryukoku University as a Professor. His current interests include sound field reproduction and acoustic signal processing using machine learning.