PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

Broadband loudspeaker placement optimization for personal sound zones systems

Min ZHU[1,2] and Sipei ZHAO[3]

[1] Tongda College of Nanjing University of Posts and Telecommunication, Nanjing, China

[2] Telecommunication and Information Engineering College, Nanjing University of Posts and Telecommunication, Nanjing, China

[3] Centre for Audio, Acoustics, and Vibration, Faculty of Engineering and IT, University of Technology Sydney, Sydney, Australia

ABSTRACT
Personal sound zones system has attracted considerable attention in the past decades due to its potential for private audio generation in public spaces. Various methods have been explored to optimize the driving signals of loudspeakers that are placed to form a regular array, such as circular, linear, and arc-shaped arrays. Recently, loudspeaker placement optimization has been investigated by researchers to reduce the number of loudspeakers without remarkable sacrifice in performance. Existing loudspeaker placement optimization algorithms have been designed in the frequency domain and the optimized loudspeaker arrangements depend on frequency, which is undesirable in practical applications. To overcome this problem, this paper explores broadband loudspeaker placement optimization for multizone sound field reproduction based on a time-domain evolutionary array optimization method. Simulations with measured room impulse responses are performed to select a smaller number of loudspeakers from 60 candidate loudspeakers that are uniformly placed along a circle. Simulation results demonstrate the optimized array achieves a higher acoustic contrast with a lower array effort than the empirical arc-shaped array, when the same number of loudspeakers are selected.

Keywords: Multizone sound, Personal audio, loudspeaker placement

1. INTRODUCTION
Personal Sound Zones (PSZ) systems, also known as personal audio systems, aim to generate independent listening zones for multiple users in a shared physical space, using an array of loudspeakers instead of headphones (1). Since its first proposal by Druyvesteyn and Garas (2) in 1997, numerous research have been carried out to optimize the driving signals of loudspeakers, and many methods have been developed, such as the acoustic contrast control (3), pressure matching (4), mode matching (5), variable span trade-off filtering (6), and tangent line (7) methods *etc*.

Loudspeaker placement optimization for PSZ systems have also been studied, and existing techniques can be categorized into sparse regularization and iterative methods (8). The sparsity regularization approaches apply a l_1-norm regularization term in the cost function to approximate the l_0-norm regularization motivated by the compressive sampling theory, including the Lasso (9) and the elastic net methods (10), which were originally proposed for single-zone sound field reproduction. The Lasso-based method was extended to multizone sound field reproduction in combination with pressure matching method to form a two-stage Lasso-LS scheme (11,12).

Different from the sparse regularization approaches, iterative methods select a loudspeaker from the candidate set in each iteration based on certain criteria. In the Gram-Schmidt Orthogonalization (GSO) method (13), the loudspeakers were added one by one based on the linear independence of the acoustic transfer functions. By contrast, the Singular Value Decomposition (SVD) method constructs

[1] zhumin@njupt.edu.cn
[3] Sipei.Zhao@uts.edu.au

a so-called ideal acoustic transfer function first, and then finds the candidate locations that best match the ideal acoustic transfer function one by one (14). The Constrained Matching Pursuit (CMP) method approximates the target sound field by iteratively adding the loudspeaker whose acoustic transfer function is the most correlated with the reproduction error (15).

Recently, Evolutionary Array Optimization (EAO) techniques have been developed to optimize loudspeaker placement in PSZ systems (16,17). In contrast to the iterative methods that select a loudspeaker from the candidate set at each time, the EAO method removes a loudspeaker from the candidate set in each iteration. Experimental results demonstrated the superiority of the EAO method over both the sparse regularization and the iterative methods (17). However, existing studies have been performed in the frequency domain and the optimized loudspeaker arrangements depend on frequency, which is undesirable in practical applications.

This paper investigates the time-domain EAO method for loudspeaker placement optimization in PSZ systems, based on the Broadband Acoustic Contrast Control (BACC) method (18). Conventional BACC method maximizes the contrast of the average acoustic potential energy between the bright and dark zones; however, it suffers from uneven frequency responses (19) and nonuniform sound field distribution (20), leading to non-satisfactory sound quality in the bright zone. To flatten the frequency responses, various constraints have been proposed, including the response differential (21) and Response Trend Estimation (RTE) terms (22). To improve the spatial sound field distribution in the bright zone, a Spatial Uniformity Constraint (SUC) has been proposed recently (23). This paper optimizes loudspeaker placement in PSZ systems based on the BACC method with the RTE and SUC (BACC-RTE-SUC) terms. Simulations with measured Room Impulse Responses (RIRs) are performed to compare the performance of the optimized array with that of the empirical arc-shaped array, and the results demonstrate the advantages of the time-domain EAO method.

2. METHOD

2.1 BACC-RTE-SUC method

A PSZ system utilizes an array of loudspeakers to generate an acoustic bright zone and a dark zone, as illustrated in Figure 1. The input signal $x[n]$ (n denotes the time instant) is filtered with a Finite Impulse Response (FIR) filter $\mathbf{w}_l$ (l = 1, 2, …, L, L is the total number of loudspeakers) with a length of I before reproduced through each loudspeaker. The RIR from the l-th loudspeaker to the m-th microphone in the control zones is modelled as a FIR filter $\mathbf{h}_{ml}$ with a length of K.

To investigate the frequency response of the system, the input signal $x[n]$ is assumed to be a Dirac delta function, hence the sound pressure at the m-th microphone due to the l-th loudspeaker is written as (23)

$$p_{\mathrm{B},ml}[n]=\sum_{i=0}^{I-1}h_{\mathrm{B},ml}[n-i]\,w_l[i], \tag{1}$$

where the subscript $_\mathrm{B}$ denotes the bright zone. Arranging the sound pressure of all the time index into a vector, *i.e.*, $\mathbf{p}_{\mathrm{B},ml}=[p_{\mathrm{B},ml}[0], p_{\mathrm{B},ml}[1], \ldots, p_{\mathrm{B},ml}[I+K-2]]^{\mathrm{T}}$, Eq. (1) can be expressed concisely as

$$\mathbf{p}_{\mathrm{B},ml}=\mathbf{H}_{\mathrm{B},ml}\mathbf{w}_l, \tag{2}$$

where $\mathbf{w}_l=[w_l[0], w_l[1], \ldots, w_l[I-1]]^{\mathrm{T}}$ and

$$\mathbf{H}_{\mathrm{B},ml}=\begin{bmatrix} h_{\mathrm{B},ml}[0] & 0 & 0 \\ \vdots & \ddots & 0 \\ h_{\mathrm{B},ml}[K-1] & \ddots & h_{\mathrm{B},ml}[0] \\ 0 & \ddots & \vdots \\ 0 & 0 & h_{\mathrm{B},ml}[K-1] \end{bmatrix} \tag{3}$$

is a matrix of dimension $(I+K-1)\times I$. It should be noted that $P_{\mathrm{B},ml}$ in Eq. (1) is in fact the global impulse response with a length of $(I+K-1)$.

Summing up the contribution from all the L loudspeakers to the m-th microphone, one obtains (23)

$$\mathbf{p}_{\mathrm{B},m}=\sum_{l=1}^{L}\mathbf{p}_{\mathrm{B},ml}=\sum_{l=1}^{L}\mathbf{H}_{\mathrm{B},ml}\mathbf{w}_l=\mathbf{H}_{\mathrm{B},m}\mathbf{w}, \tag{4}$$

where $\mathbf{w}=[\mathbf{w}_1^{\mathrm{T}}, \mathbf{w}_2^{\mathrm{T}}, \ldots, \mathbf{w}_L^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{H}_{\mathrm{B},m}=[\mathbf{H}_{\mathrm{B},m1}, \mathbf{H}_{\mathrm{B},m2}, \ldots, \mathbf{H}_{\mathrm{B},mL}]$. Denoting the sound pressure at all the M_B microphones in the bright zone in a vector form, *i.e.*, $\mathbf{p}_\mathrm{B}=[\mathbf{p}_{\mathrm{B},1}^{\mathrm{T}}, \mathbf{p}_{\mathrm{B},2}^{\mathrm{T}}, \ldots, \mathbf{p}_{\mathrm{B},M\mathrm{B}}^{\mathrm{T}}]^{\mathrm{T}}$, one obtains $\mathbf{p}_\mathrm{B}=\mathbf{H}_\mathrm{B}\mathbf{w}$, where $\mathbf{H}_\mathrm{B}=[\mathbf{H}_{\mathrm{B},1}^{\mathrm{T}}, \mathbf{H}_{\mathrm{B},2}^{\mathrm{T}}, \ldots, \mathbf{H}_{\mathrm{B},M\mathrm{B}}^{\mathrm{T}}]^{\mathrm{T}}$ is a $M_\mathrm{B}(I+K-1)\times IL$ matrix of the impulse responses from the loudspeaker array to the bright zone. Similarly, the sound pressure in the dark zone can be obtained as $\mathbf{p}_D = \mathbf{H}_D\mathbf{w}$ with $\mathbf{H}_D$ being the $M_D(I + K - 1) \times IL$ impulse response matrix from the loudspeaker array to the dark zone.

The cost function for the BACC-RTE-SUC is defined as (23)

$$\mathbf{w}_o = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{R}_B\mathbf{w}}{\mathbf{w}^T\left[\gamma\mathbf{R}_D + \alpha\mathbf{C}_{RTE} + \beta\mathbf{C}_{SUC}\right]\mathbf{w} + \lambda\mathbf{w}^T\mathbf{w}}, \tag{5}$$

where $\mathbf{R}_B = \mathbf{H}_B{}^T\mathbf{H}_B/M_B$, $\mathbf{R}_D = \mathbf{H}_D{}^T\mathbf{H}_D/M_D$, λ is the regularization parameter, α, β, γ are the weighting factors that add up to 1, and the two matrices $\mathbf{C}_{RTE}$ and $\mathbf{C}_{SUC}$ apply constraints to improve the flatness of frequency responses and increase the spatial sound pressure level uniformity in the bright zone, respectively. The detailed formulations of $\mathbf{C}_{RTE}$ and $\mathbf{C}_{SUC}$ can be found in (22) and (23), respectively, and are not shown here for brevity. The optimal solution to Eq. (5) is proportional to the eigenvector corresponding to the largest eigenvalue of the matrix $(\gamma\mathbf{R}_D + \alpha\mathbf{C}_{RTE} + \beta\mathbf{C}_{SUC} + \lambda\mathbf{I})^{-1}\mathbf{R}_B$ (23).

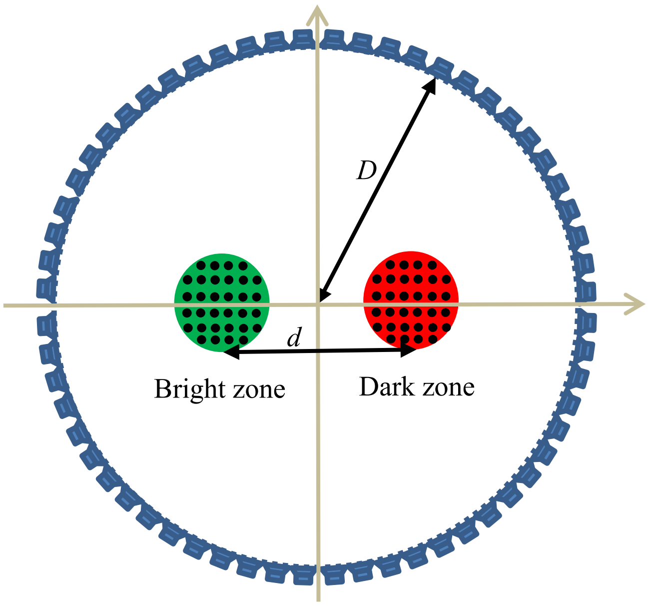

Figure 1 – Illustration of a PSZ system with a circular loudspeaker array.

2.2 Evolutionary Array Optimization

The idea of EAO is intuitive and straightforward. The performance of the system with the candidate set is evaluated in each step, and the loudspeaker that has the least contribution to the performance is removed; this process is repeated until the desired number of loudspeakers is retained. Various criteria in the frequency domain have been employed to evaluate the performance (16,17). In this paper, the Broadband Acoustic Contrast (BAC) is defined below to evaluate the system performance

$$BAC = 10\log_{10}\frac{\mathbf{w}_o{}^T\mathbf{R}_B\mathbf{w}_o}{\mathbf{w}_o{}^T\mathbf{R}_D\mathbf{w}_o}, \tag{6}$$

where $\mathbf{w}_o$ denotes the optimal solution to Eq. (5) in each iteration.

With the performance criterion in Eq. (6), the EAO method selects the desired N loudspeakers from L candidates iteratively. In each iteration, a loudspeaker is muted in sequence, and the optimal filter coefficients and the performance criteria are calculated; the loudspeaker that corresponds to the largest BAC is removed from the candidates. The complete description of the time-domain EAO method is summarized in Table 1.

Table 1 – Time-domain EAO based on the BACC-RTE-SUC method: select N out of L loudspeakers

Inputs	L candidate loudspeakers with transfer function matrices $\mathbf{H}_B$ and $\mathbf{H}_D$; the desired number of loudspeakers N; and $L' = L$.
Step 1	Mute the l-th loudspeaker from the candidates and calculate the optimal filter coefficients $\mathbf{w}_o(l)$ based on Eq. (5);
Step 2	Calculate the performance criterion $BAC(l)$ based on Eq. (6);
Step 3	Repeat Steps 1-2 for l = 1, 2, …, L' and find the l_0-th configuration with the maximum performance criterion, i.e., $l_0 = \arg\max_l \left[BAC(l)\right]$;
Step 4	Remove the l_0-th loudspeaker to select the ($L' - 1$) loudspeakers from the original L' loudspeakers and update $L' = L' - 1$;
Step 5	Repeat Steps 1-4 to remove more loudspeakers until the desired number of N loudspeakers are selected from the original L loudspeakers.

2.3 Evaluation Metrics

The performance of the array optimized with the EAO method is compared to that of the arc-shaped array in terms of Acoustic Contrast (AC) and Array Effort (AE), which are defined as

$$AC(f) = 10\log_{10}\frac{\hat{\mathbf{p}}_B(f)^H \hat{\mathbf{p}}_B(f)}{\hat{\mathbf{p}}_D(f)^H \hat{\mathbf{p}}_D(f)} \tag{7}$$

and

$$AE(f) = 10\log_{10}\frac{\hat{\mathbf{w}}_{opt}(f)^H \hat{\mathbf{w}}_{opt}(f)}{\left|\hat{w}_{ref}(f)\right|^2}, \tag{8}$$

respectively. $\hat{\mathbf{p}}_B(f)$ and $\hat{\mathbf{p}}_D(f)$ in Eq. (7) are the sound pressure in the bright and dark zones at frequency f, obtained by Fourier transforming the time series of $\mathbf{p}_B$ and $\mathbf{p}_D$, respectively. Similarly, $\hat{\mathbf{w}}_{opt}(f)$ in Eq. (8) denotes the optimized source strengths at frequency f, obtained by Fourier transforming the time series of the filter coefficients, and $\hat{w}_{ref}$ is the source strength that a single reference loudspeaker needs to generate the same acoustic potential energy in the bright zone. It is noted that the frequency-domain metrics in Eqs. (7) and (8) are only used to evaluate the performance of the optimized and arc-shaped arrays, and are not used in the EAO optimization process.

3. SIMULATIONS AND DISCUSSIONS

Simulations with measured RIRs are performed to evaluate the performance of the time-domain EAO method for broadband loudspeaker placement optimization.

3.1 RIRs measurements

RIRs were measured in the hemi-anechoic chamber at UTS Tech Lab, as shown in Figure 2, where L = 60 loudspeakers were placed along a circular truss with a radius of 1.5 m. The loudspeakers were 1.4 m above the ground. An 8×8 square microphone array with a dimension of 0.28 m (0.04 m interval between microphones) was used to capture the sound signals. The distance between the bright and dark zones were 0.8 m.

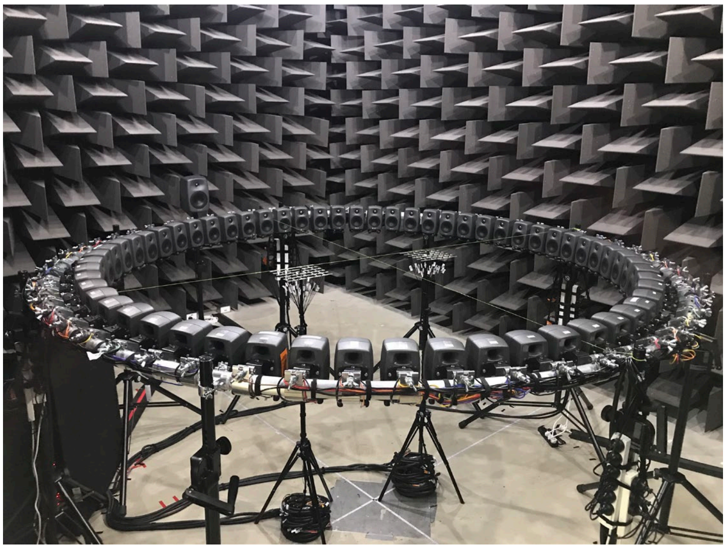

Figure 2 – Photo of the RIRs measurement system with 60 loudspeakers placed along a circular truss.

In the measurements, four Yamaha RIO1608-D2 and four Yamaha RIO8-D audio interfaces were daisy chained to form a 64-input-64-output audio system that communicates with a computer through a Dante virtual sound card. Both the loudspeaker and microphone arrays were connected to the audio interfaces for simultaneous sound generation and acquisition. During the measurements, a 3 s long logarithmic sine sweep signal from 20 Hz to 22 kHz was generated in MATLAB with a sampling rate of 48 kHz and reproduced through each loudspeaker, and the sound pressures in each zone were captured by the microphone array.

3.2 Simulation results

Simulations with the measured RIRs were performed to select different number of loudspeakers from the original L = 60 loudspeakers. To save computational burden, only the RIRs measured by 32 microphones were used in each zone, and the measured RIRs were down sampled to 2 kHz and modelled with 64-tap FIR filters. The length of the control filters was 32. The regularization parameter was set to 10^{-10} times the largest eigenvalue of the matrix $\mathbf{R}_{\mathrm{D}}$, and the weighting factors α, β and γ were 0.83, 0.083 and 0.087, respectively.

The simulated BAC for the EAO-optimized array and the arc-shaped array are compared in Figure 3. It is clear that, when the same number of loudspeakers are used, the performance of the EAO-optimized array is superior over the arc-shaped array in terms of BAC. For example, when 10 loudspeakers are used, the EAO-optimized array achieves a BAC of 42.7 dB and the arc-shaped array 32.3 dB. Similarly, when 30 loudspeakers are selected, the BAC for the EAO-optimized array and the arc-shaped array are 50.4 and 40.2 dB, respectively. With the number of loudspeakers increasing, the differences in the performance of the EAO-optimized and arc-shaped arrays decrease. For example, when 50 loudspeakers are chosen, the BAC for the EAO-optimized and arc-shaped arrays are 50.8 and 45.5 dB, respectively. This is expected because only 10 loudspeakers are removed from the original 60 loudspeakers.

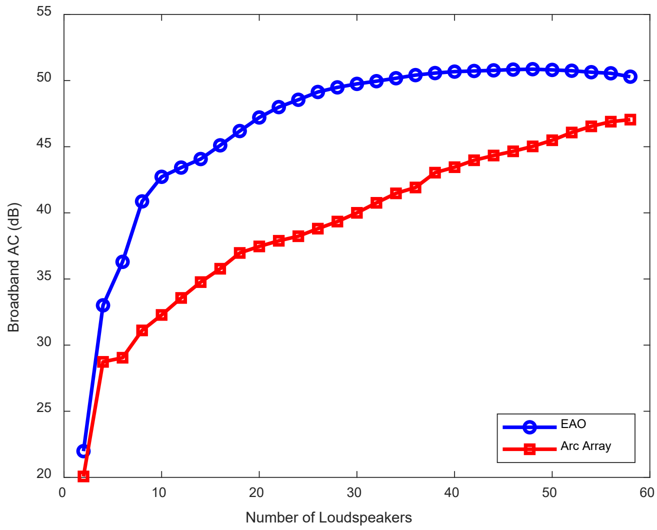

Figure 3 – Simulated BAC for the EAO optimized array and the arc-shaped array.

The configurations of the arc-shaped and the EAO-optimized arrays are compared in Figure 4 when 10, 30 and 50 loudspeakers are retained from the circular array. It can be observed that the loudspeakers in the EAO-optimized array spread over the circle when 10 and 30 loudspeakers are selected. This is different from the frequency dependent configurations obtained in (16,17), where the selected loudspeakers tend to gather around the bright zone. This demonstrates the need for broadband optimization instead of frequency dependent configuration for practical applications.

The AC and AE achieved by the EAO-optimized and arc-shaped arrays are compared in Figure 5 when 10, 30 and 50 loudspeakers are utilized. It can be seen that, when 10 loudspeakers are selected from the original 60 loudspeakers, the EAO-optimized array achieved a higher AC (by approximately 10 dB) and a lower AE (by over 10 dB) than the arc-shaped array, in the frequency range between 200 Hz and 800 Hz. With the increase in the number of the loudspeakers, the difference between the EAO-optimized and arc-shaped arrays becomes small. This is consistent with the broadband results in Figure 3.

The above results demonstrate that the EAO-optimized array is advantageous over the empirical arc-shaped array in terms of both AC and AE when the same number of loudspeakers are used. It is noted that the arc-shaped array is used as the comparison benchmark here because existing optimization methods, as summarized in (8), work in the frequency domain and cannot readily adapt to time domain.

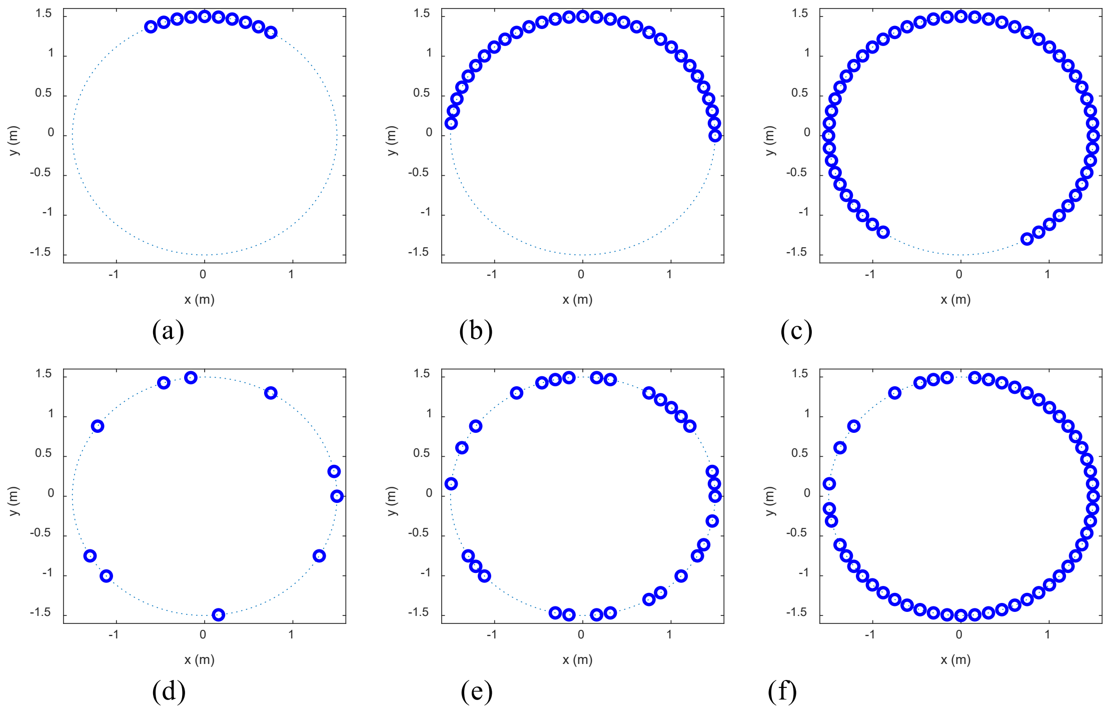

Figure 4 – Comparison of the arc-shaped array with (a) 10, (b) 30 and (c) 50 loudspeakers and the EAO-optimized array with (d) 10, (e) 30 and (f) 60 loudspeakers.

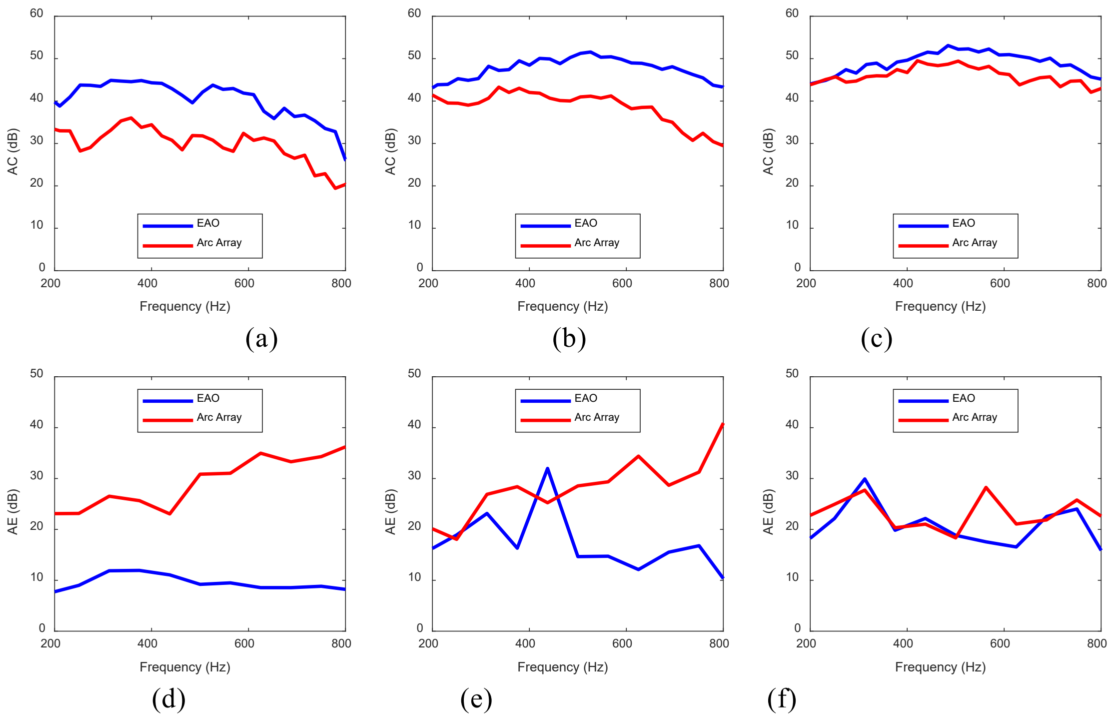

Figure 5 – Comparison of the Acoustic Contrast (AC) when (a) 10, (b) 30, and (c) 50 loudspeakers are used and the Array Effort (AE) with (d) 10, (e) 30, and (f) 60 loudspeakers are used.

4. CONCLUSIONS

This paper investigates loudspeaker placement optimization for personal sound zones systems based on the time-domain acoustic contrast control method, aiming to select a smaller number of loudspeakers from a large set of candidates. The recently proposed evolutionary array optimization technique is extended to the time domain for broadband loudspeaker placement optimization. Room impulse responses measured in a hemi-anechoic are used in simulations to evaluate the performance of the optimization method. Simulation results show that, when 10 loudspeakers are selected from the 60 candidates, the optimized array obtains a higher acoustic contrast by approximate 10 dB and a lower array effort by over 10 dB than the empirical arc-shaped array. Future work includes investigating the effect of room reverberation on the performance of the optimization algorithm.

ACKNOWLEDGEMENT

Sponsored by Tongda College of Nanjing University of Posts and Telecommunications (Grant No. XK201XZ22001).

REFERENCES

1. Betlehem T, Zhang W, Poletti MA, Abhayapala TD. Personal sound zones: Delivering interface-free audio to multiple listeners. IEEE Signal Process Mag. 2015;32(2):81–91.
2. Druyvesteyn WF, Garas J. Personal sound. J Audio Eng Soc. 1997;45(9):685–701.
3. Choi J-W, Kim Y-H. Generation of an acoustically bright zone with an illuminated region using multiple sources. J Acoust Soc Am. 2002;111(4):1695–700.
4. Kirkeby O, Nelson PA. Reproduction of plane wave sound fields. J Acoust Soc Am. 1993;94(5):2992–3000.
5. Wu YJ, Abhayapala TD. Spatial multizone soundfield reproduction: Theory and design. IEEE Trans Audio, Speech Lang Process. 2011;19(6):1711–20.
6. Lee T, Nielsen JK, Jensen JR, Christensen MG. A unified approach to generating sound zones using variable span linear filters. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. 2018. p. 491–5.
7. Zhao S, Qiu X, Burnett IS. Acoustic contrast control in an arc-shaped area using a linear loudspeaker array. J Acoust Soc Am [Internet]. 2015;137(2):1036–9. Available from: http://scitation.aip.org/content/asa/journal/jasa/137/2/10.1121/1.4906184
8. Zhao S, Zhu Q. Comparative study of loudspeaker position optimization techniques for multizone sound field reproduction. In: Proceedings of Inter-Noise 2021. Washington, US; 2021. p. 1–10.
9. Lilis GN, Angelosante D, Giannakis GB. Sound field reproduction using the lasso. IEEE Trans Audio, Speech Lang Process. 2010;18(8):1902–12.
10. Gauthier P-A, Lecomte P, Berry A. Source sparsity control of sound field reproduction using the elastic-net and the lasso minimizers. J Acoust Soc Am. 2017;141(4):2315–26.
11. Radmanesh N, Burnett IS. Generation of isolated wideband sound fields using a combined two-stage lasso-LS algorithm. IEEE Trans Audio, Speech Lang Process. 2013;21(2):378–87.
12. Radmanesh N, Burnett IS, Rao BD. A lasso-ls optimization with a frequency variable dictionary in a multizone sound system. IEEE/ACM Trans Audio Speech Lang Process. 2016;24(3):583–93.
13. Asano F, Suzuki Y, Swanson DC. Optimization of control source configuration in active control systems using gram-schmidt orthogonalization. IEEE Trans Speech Audio Process. 1999;7(2):213–20.
14. Khalilian H, Bajic I V., Vaughan RG. Towards optimal loudspeaker placement for sound field reproduction. ICASSP, IEEE Int Conf Acoust Speech Signal Process - Proc. 2013;321–5.
15. Khalilian H, Bajic I V., Vaughan RG. Comparison of loudspeaker placement methods for sound field reproduction. IEEE/ACM Trans Audio Speech Lang Process. 2016;24(8):1364–79.
16. Zhu M, Zhao S. An iterative approach to optimize loudspeaker placement for multi-zone sound field reproduction. J Acoust Soc Am. 2021;149(5):3462–8.
17. Zhao S, Burnett IS. Evolutionary array optimization for multizone sound field reproduction. J Acoust Soc Am. 2022;151(4):2791–801.
18. Elliott SJ, Cheer J. Regularisation and Robustness of Personal Audio Systems. ISVR Technical Memorandum 995. 2011.
19. Hu M, Lu J. Theoretical explanation of uneven frequency response of time-domain acoustic contrast control method. J Acoust Soc Am. 2021;149(6):4292–7.
20. Coleman P, Jackson PJB, Olik M, Møller M, Olsen M, Abildgaard Pedersen J. Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array. J Acoust Soc Am [Internet]. 2014 [cited 2020 Aug 12];135(4):1929–40. Available from: http://dx.doi.org/10.1121/1.4866442]
21. Cai Y, Wu M, Liu L, Yang J. Time-domain acoustic contrast control design with response differential

constraint in personal audio systems. J Acoust Soc Am [Internet]. 2014 [cited 2020 Jul 27];135(6):EL252–7. Available from: https://doi.org/10.1121/1.4874236

22. Schellekens DHM, Moller MB, Olsen M. Time domain acoustic contrast control implementation of sound zones for low-frequency input signals. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. 2016. p. 365–9.
23. Zhao S, Burnett IS. Time-domain acoustic contrast control with spatial uniformity constraint for personal audio systems. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings. Singapore; 2022. p. 1–10.