A Time-domain Multi-channel Directional Active Noise Control System

Huawei Zhang[1], Jihui (Aimee) Zhang[2,1], Fei Ma[1], Prasanga N. Samarasinghe[1], Huiyuan Sun[3,1]
[1]Audio & Acoustic Signal Processing Group, The Australian National University, Canberra, Australia
[2]Institute of Sound and Vibration Research, University of Southampton, Southampton, U.K.
[3]School of Electrical and Information Engineering, The University of Sydney, Sydney, Australia
huawei.zhang@anu.edu.au, aimee_jihui.zhang@soton.ac.uk, fei.ma@anu.edu.au,
prasanga.samarasinghe@anu.edu.au, huiyuan.sun@sydney.edu.au

***Abstract*—This paper proposes a time-domain multi-channel directional active noise control (ANC) system to reduce noises selectively based on the direction of arrival (DOA). This system consists of three subsystems: the beamforming subsystem, the ANC subsystem, the sound reproduction subsystem. The beamforming subsystem is implemented to capture the desired sound for the sound reproduction subsystem. The ANC subsystem is developed to reduce noises from all directions and the sound reproduction subsystem is implemented to reproduce sound from the desired direction. As a preliminary work of directional spatial ANC, the proposed system enables the user's ears at error microphone locations hearing the desired sound with reduced noises without wearing headphones. The simulation results demonstrate that our system can reduce noises significantly and preserve the desired sound from desired direction, where the binaural localization cues of the desired sound is also preserved.**

***Index Terms*—Directional active noise control, multi-channel signal processing, time domain, beamforming, sound reproduction**

I. Introduction

Active noise control (ANC) systems reduce the primary noise by superposing it with an anti-noise generated by the secondary sources [1]–[4]. Different from conventional ANC systems [5]–[7], spatial ANC systems have been proposed to reduce noises over space in past years [8]–[11], which can improve the potential practical value of ANC systems. Multi-channel ANC systems [5] can be considered as a prior work of spatial ANC systems.

There is a limitation [12], [13] in the current spatial ANC systems [14]–[16]. They commonly aim to reduce noises from all directions [8], [17], but the user may want to hear sound coming from a certain desired direction [13]. Therefore, it is highly desirable for spatial ANC systems to reduce noises selectively.

Recently, directional ANC systems [13], [18]–[21] have been proposed to apply ANC selectively based on the direction of arrival (DOA) of the desired sound [22]. Patel *et al.* have designed an ANC headphone with directional hear-through capability, achieving directional ANC while preserving the desired sound from the look-up direction [13]. In [13], the system is implemented by reducing noises from all directions and playing the desired sound from the look-up direction (captured by the beamforming). Later, Xiao *et al.* proposed a joint function [19], which combines beamforming [22] and ANC [5] to reduce noises selectively and preserve the desired sound, without implementing sound reproduction, especially for smart glasses.

In our opinion, there is a potential of introducing directional ANC technology [13], [18]–[21] to relax the limitation in spatial ANC systems mentioned above. In this paper, as a preliminary work, we extend [13] to a multi-channel directional ANC system without wearing a headphone. Furthermore, we introduce a sound reproduction subsystem to release the limitation in [13] that the desired direction is set to be the look-up direction. Specifically, our system can reduce noises selectively and preserve the desired sound at multiple points based on the DOA of the desired sound. We assume that the DOA of the desired sound is known before our system processing. Firstly, our system can estimate the desired sound by implementing the time-domain beamforming algorithm (time delay beamforming) [22] to process reference signals received at a reference sensor array. Secondly, a time-domain multi-channel ANC subsystem with filtered-x least mean square algorithm (multi-channel FxLMS ANC subsystem) [5] is applied to reduce noises based on the received reference signals. Finally, amplitude panning [23] is introduced to the sound reproduction subsystem to reproduce the desired sound with binaural localization cues, based on the estimated desired sound and the DOA of the desired sound.

There are two main advantages of our system. First, the setup is highly flexible, where there are various choices of both the desired directions and system structure (each subsystem can be adjusted to satisfy different requirements). Second, the user can enjoy noise reduction with the desired sound hearable from the accurate direction.

II. Problem Formulation

The setup of the proposed system is shown in Fig. 1. It consists of M sensors as the linear reference sensor array with the inner distance L, K loudspeakers as the secondary sources, and J sensors as the error sensors with the inner distance d. We assume that there are multiple noise sources from different directions, and the direction of the desired sound source is

This research was funded by Australian Research Council (ARC) Discovery Project Grant No. DP180102375.

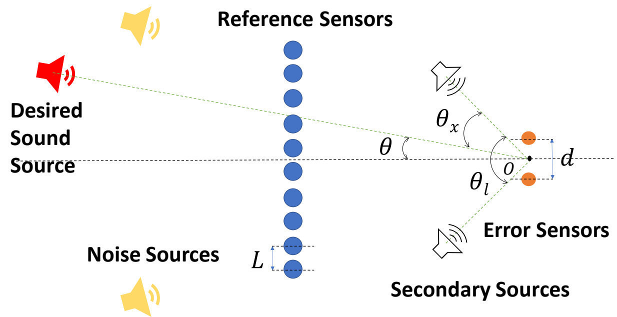

Fig. 1. The proposed system setup with a desired sound source, multiple noise sources, multiple reference sensors as the reference sensor array, multiple loudspeakers as secondary sources, and multiple error sensors as the error sensor array. The system is symmetric against the middle line (black dotted) across the origin (O).

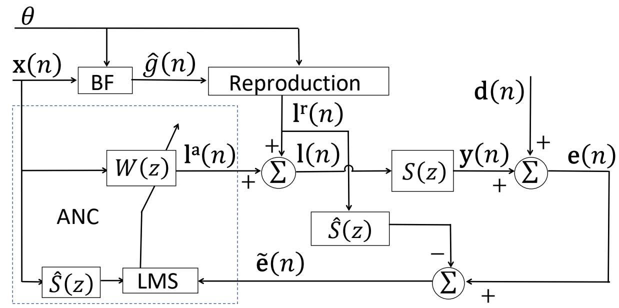

Fig. 2. The block diagram of our system, including beamforming (BF), active noise control (ANC) and reproduction in time domain.

different from any of noise sources. Here, θ is the DOA angle of the desired sound, θ_x is the angle between the DOA of the desired sound and the direction of the first secondary source, and θ_l is the aperture angle of the secondary sources. As shown in Fig. 1, the reference sensor array, the secondary sources and the error sensors are set on the middle line connected to the origin (O). The user's ears are assumed to be at the location of the error sensors, and the user is facing the sound sources.

In conventional ANC systems [5], the reference signals $\mathbf{x}(n) = [x_1(n) \quad x_2(n) \quad \cdots \quad x_M(n)]^T$ are received at the reference sensor array, and the error signals $\mathbf{e}(n) = [e_1(n) \quad e_2(n) \quad \cdots \quad e_J(n)]^T$ are received at the error sensor array. As there are both the desired sound and noises, the error signals $\mathbf{e}(n)$ consist of three components:

$$\begin{aligned}\mathbf{e}(n) &= \mathbf{y}(n) + \mathbf{d}(n)\\ &= \mathbf{y}(n) + \mathbf{g}(n) + \mathbf{v}(n),\end{aligned} \tag{1}$$

where $\mathbf{y}(n) = [y_1(n) \quad y_2(n) \quad \cdots \quad y_J(n)]^T$ is the secondary sound components and $\mathbf{d}(n) = \mathbf{g}(n) + \mathbf{v}(n)$ is the primary signal, which consists of two parts: the desired sound components $\mathbf{g}(n) = [g_1(n) \quad g_2(n) \quad \cdots \quad g_J(n)]^T$ and the noise components $\mathbf{v}(n) = [v_1(n) \quad v_2(n) \quad \cdots \quad v_J(n)]^T$. The secondary sound component $\mathbf{y}(n)$ is the convolution between the driving signals of the secondary sources, $\mathbf{l}(n) = [l_1(n) \quad l_2(n) \quad \cdots \quad l_K(n)]^T$, and the impulse responses of the secondary paths, $S(z)$. We denote the pre-estimated impulse responses of the secondary paths as $\hat{S}(z)$. (The secondary source feedback and the online secondary path estimation are neglected as they are not related to the central topic of this paper, but we will consider them in future works [5].)

The difference between conventional ANC systems and the proposed system is that the proposed system aims to set the error signals $\mathbf{e}(n)$ close to the desired sound component $\mathbf{g}(n)$, by applying a set of control filters, $W(z)$, to reduce the driving signal of secondary sources $\mathbf{l}(n)$. In other words, our system aims to reduce all noises but preserve the desired sound at the error sensors. Ideally, the user at the location of the error sensors can hear the distortionless desired sound only.

III. System Structure

As shown in Fig. 2, our system can be divided into three subsystems, including the beamforming subsystem [22], the ANC [5] subsystem, the sound reproduction subsystem [23]. The ANC subsystem reduces both the desired sound components $\mathbf{g}(n)$ and the noise components $\mathbf{v}(n)$, and the sound reproduction subsystem aims to reproduce the desired sound component $\mathbf{g}(n)$, which is captured by the beamforming subsystem [13]. Therefore, the driving signals $\mathbf{l}(n)$ are divided into two parts:

$$\mathbf{l}(n) = \mathbf{l}^a(n) + \mathbf{l}^r(n), \tag{2}$$

where $\mathbf{l}^a(n) = [l_1^a(n) \quad l_2^a(n) \quad \cdots \quad l_K^a(n)]^T$ and $\mathbf{l}^r(n) = [l_1^r(n) \quad l_2^r(n) \quad \cdots \quad l_K^r(n)]^T$ are the control parts and the reproduction parts of the driving signals, respectively.

A. Beamforming

In this section, we introduce the beamforming subsystem to capture the desired sound, and the estimated desired sound signal is denoted as $\hat{g}(n)$.

In this subsystem, we apply time delay beamforming to capture the desired sound based on the received signals $\mathbf{x}(n)$ at the reference sensor array [22]. We set the inner distance of the reference sensor array to be much smaller compared with the distance between sound sources and the reference sensor array, so the sound sources can be considered as far-field sound sources. The received reference signal at the m-th sensor, $x_m(n)$, is related to the received reference signal at the first sensor, $x_1(n)$:

$$x_m(n) = x_1(n - \tau_m), \tag{3}$$

where $\tau_m = (m-1)L\sin(\theta)/c$ is time delay between $x_1(n)$ and $x_m(n)$, c is speed of sound propagation, θ is the DOA angle of the desired sound, and L is the inner distance of the reference sensor array. Therefore, the estimated desired sound signal $\hat{g}(n)$ can be obtained as:

$$\hat{g}(n) = \frac{1}{M}\sum_{m=1}^{M} x_m(n + \tau_m). \tag{4}$$

The estimated desired sound signal $\hat{g}(n)$ will be used for the sound reproduction subsystem in Section III-C. However, there is a mismatch issue: The desired sound signals $\mathbf{g}(n)$ are received at the error sensor array, but the estimated desired sound signal $\hat{g}(n)$ is obtained based on the reference signals $\mathbf{x}(n)$ received at the reference sensor array and then $\hat{g}(n)$ is used for sound reproduction at secondary sources. The mismatch issue is neglected as it is out of the scope of this paper, and we will introduce the characteristics of acoustic paths to deal with this issue in the future.

B. Active Noise Control

In this section, we introduce the ANC subsystem to reduce noises coming from all directions including the desired direction at the error sensor array, especially for noises whose frequency is lower than 1000 Hz.

In the ANC subsystem, we revise the time-domain multichannel FxLMS-based adaptive algorithm [5]. Here, the control part of the i-th driving signal, $l_i^a(n)$, can be obtained as:

$$l_i^a(n) = \sum_{m=1}^{M} \mathbf{w}_{im}^T(n)\mathbf{x}_m(n), \tag{5}$$

where $\mathbf{w}_{im}(n)$ is the $\mathcal{L}$-order control filer for the i-th secondary source which filters the m-th reference microphone signal, and $\mathbf{x}_m(n) = [x_m(n) \quad x_m(n-1) \quad \cdots \quad x_m(n-\mathcal{L}+1)]^T$ is the reference signal vector. To avoid reducing the reproduced sound created by the reproduction subsystem (See in Section III-C), the pseudo error signals $\tilde{\mathbf{e}}(n) = [\tilde{e}_1(n) \quad \tilde{e}_2(n) \quad \cdots \quad \tilde{e}_J(n)]^T$ are introduced to directional ANC system [13]. The pseudo error signal $\tilde{e}_j(n)$ for the j-th error sensor can be obtained as the difference between the j-th received error signals $e_j(n)$ and the estimated reproduced sound at error microphones:

$$\tilde{e}_j(n) = e_j(n) - \sum_{i=1}^{K} \hat{s}_{ji}(n) * l_i^r(n), \tag{6}$$

where $l_i^r(n)$ is the reproduction components of the i-th driving signal, and $\hat{s}_{ji}(n)$ is the pre-estimated coefficients of corresponding secondary path between the j-th error sensor and the i-th secondary source. Therefore, we update the control filter $\mathbf{w}_{im}(n)$:

$$\mathbf{w}_{im}(n+1) = \mathbf{w}_{im}(n) - \mu \sum_{j=1}^{J} (\hat{s}_{ji}(n) * \mathbf{x}(n))\tilde{e}_j(n), \tag{7}$$

where μ is the step size.

C. Sound Reproduction

In this subsection, we introduce the sound reproduction subsystem to reproduce the desired sound with the binaural localization cues.

We want to reproduce sound at two ears of the user and avoid occupying too much space. Hence, we choose two secondary sources with the aperture angle $\theta_l = 90°$ and two error sensors for the sound reproduction subsystem. The angle between the DOA of the desired sound and the direction of the first secondary source, θ_x, can be obtained as:

$$\theta_x = \frac{1}{2}\theta_l - \theta, \tag{8}$$

where $0 \leq \theta_x \leq \theta_l$ and $-\frac{1}{2}\theta_l \leq \theta \leq \frac{1}{2}\theta_l$. As there are only two secondary sources, a time-domain amplitude panning algorithm [23] is implemented in this subsystem. The amplitude panning [23] can assign two different gains to corresponding secondary sources, based on both the prior knowledge of the DOA of the desired sound and the arrangement of hardware. Defining G_i as the gain for the i-th loudspeaker, the expressions of G_1 and G_2 can be given [23]:

$$\begin{cases} G_1 = \cos(\theta_x) \\ G_2 = \sin(\theta_x), \end{cases} \tag{9}$$

where θ_x is the angle between the DOA of the desired sound and the direction of the first loudspeaker, as shown in Fig. 1. Furthermore, the reproduction part of the driving signal for the i-th loudspeaker, $l_i^r(n)$, can be obtained [23]:

$$l_i^r(n) = G_i\hat{g}(n). \tag{10}$$

IV. Simulation and Result

In this section, we evaluate the noise reduction performance and the ability of preserving the desired sound of the proposed system by conducting simulation in the free space. A piece of recorded speech [24] is played as a desired sound source at $(-0.78, -2.90)$ m. Two pieces of recorded noises (a washer-dryer noise and an air-conditioner noise) [25] are played repeatedly as two noise sources at $(-1.00, -1.70)$ m and $(1.00, -1.70)$ m, respectively. We set 10 microphones as the linear reference sensor array between $(-0.45, -1.00)$ m and $(0.45, -1.00)$ m, with the inner distance $L = 0.1$ m and the center point at $(0, -1.00)$ m. We set two loudspeakers at $(-0.40, -0.40)$ m and $(0.40, -0.40)$ m as the secondary sources. We set two microphones at $(-0.09, 0)$ m and $(0.09, 0)$ m as the error sensors with the inner distance $d = 0.18$ m. The sampling frequency is 8000 Hz and the speed of sound is 343 m/s. The setup is shown in Fig. 1.

The spectrograms of both inputs and outputs of the proposed system after convergence are shown in Fig. 3. From the comparison of Fig. 3 (b), Fig. 3 (e) and Fig. 3 (f), we can find the proposed system (reproduction enabled) preserves the desired sound $g_1(n)$ compared to the conventional multi-channel ANC system (reproduction off). Here, the desired sound $g_1(n)$ is simulated by convolution between the clean speech signal and simulated acoustic path in the free field. Therefore, it shows that our system can reduce noises from undesired direction significantly and preserve the desired sound from desired direction at the error sensors. As shown in Fig. 3 (a), Fig. 3 (b) and Fig. 3 (e), the ANC subsystem [5] can reduce noises from all directions significantly (including parts of the desired sound) especially at low frequency below 1000 Hz, when reproduction is not applied (reproduction off). In addition, Fig. 3 (c) and Fig. 3 (d) shows that the beamforming [22] is able to capture the desired sound, but there are still some

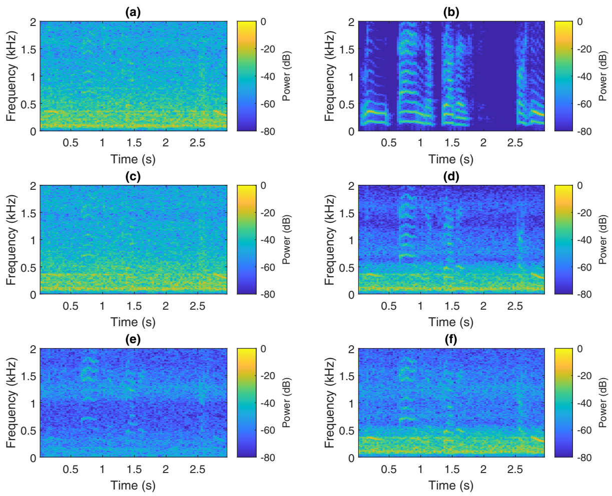

Fig. 3. The spectrograms of (a) the primary signal, $g_1(n)+v_1(n)$, at the first error microphone; (b) the desired sound $g_1(n)$ at the first error microphone; (c) the reference signal $x_3(n)$ received at the third reference microphone; (d) the beamforming output $\hat{g}(n)$; (e) the error signal $e_1(n)$ (reproduction off) at the first error microphone; (f) the error signal $e_1(n)$ (reproduction enabled) at the first error microphone.

TABLE I
THE INTERAURAL TIME DIFFERENCE (ITD) FOR THREE DIFFERENT DOA ANGLES FOR THE DESIRED SOUND θ: (A) $\theta = 0°$, (B) $\theta = 15°$ AND (C) $\theta = -30°$.

	Case (a)	Case (b)	Case (c)
DOA angle θ	$0°$	$15°$	$-30°$
Theoretical ITD (ms)	0	0.136	-0.262
Estimated ITD (ms)	0	0.125	-0.250

residual noises in the beamforming output (as the estimated desired sound $\hat{g}(n)$). Therefore, there are some residual noises, similar to a small part of the noise components $\mathbf{v}(n)$, at the the error sensors after the whole processing (reproduction enabled) [13], as shown in Fig. 3 (a) and Fig. 3 (f). There is a risk of enhancing the desired sound at the error sensors [13], but we consider this issue is acceptable as clearly hearing the desired sound is more desirable by users.

Figure 4 shows the comparison of the primary noise, the desired sound signal and the error signal (reproduction enabled) at the first error microphone after convergence. From Fig. 4 (a) and Fig. 4 (b), we can observe the noises are reduced after the ANC subsystem. Fig. 4 (a) and Fig. 4 (c) shows that compared with the primary noise, the error signal (reproduction enabled) is more similar to the desired sound signal. Therefore, we can observe the desired sound is preserved and noises are reduced when the reproduction subsystem is enabled.

The interaural time difference (ITD) cue is one of the main localization cues, and the ITD can be used in evaluation of the performance of amplitude panning methods, especially at low frequency [23]. As the ANC subsystem reduces noises significantly at low frequency (less than 1000 Hz), the ITD

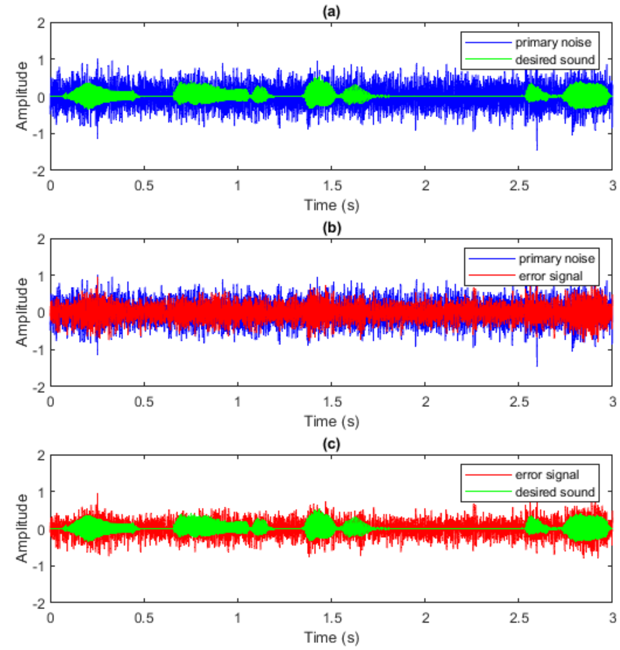

Fig. 4. The comparison between (a) the primary noise $g_1(n)+v_1(n)$ vs. the desired sound signal $g_1(n)$; (b) the primary noise $g_1(n)+v_1(n)$ vs. the error signal $e_1(n)$ (reproduction enabled); (c) the error signal $e_1(n)$ (reproduction enabled) vs. the desired sound signal $g_1(n)$ at the first error microphone.

is a suitable index to evaluate the performance of the sound reproduction subsystem. As shown in Table I, we evaluate the ITD of the final error signals (reproduction enabled) for three different cases, e.g., three different DOA angles of the desired sound θ: (a) $\theta = 0°$, (b) $\theta = 15°$ and (c) $\theta = -30°$. Three estimated ITD values are close to corresponding theoretical ITD values for the desired directions, which shows that our system can preserve the binaural localization cues of the desired sound.

V. CONCLUSION

In this paper, we proposed a time-domain multi-channel directional ANC system, which reduces noises and preserves the sound from a desired direction. Our system consists of three subsystems: the beamforming subsystem, the ANC subsystem, and the sound reproduction subsystem. The beamforming subsystem captures the desired sound for the sound reproduction subsystem. The ANC subsystem reduces noises from all directions, and the sound reproduction subsystem reproduces the captured desired sound. Simulation results showed that our system can reduce noises significantly and preserve sound from the desired direction with the binaural localization cues. In the future, we will design a near-field beamforming algorithm to develop a near-field directional ANC system, which is more suitable for our current sound reproduction subsystem.

REFERENCES

[1] Y. Kajikawa, W.-S. Gan, and S. M. Kuo, "Recent advances on active noise control: open issues and innovative applications," *APSIPA Trans. Signal Inf. Process.*, vol. 1, 2012.

[2] N. V. George and G. Panda, "Advances in active noise control: A survey, with emphasis on recent nonlinear techniques," *Signal Process.*, vol. 93, no. 2, pp. 363–377, 2013.

[3] L. Lu *et al.*, "A survey on active noise control in the past decade—part i: Linear systems," *Signal Process.*, vol. 183, p. 108039, 2021.

[4] ——, "A survey on active noise control in the past decade–part ii: Nonlinear systems," *Signal Process.*, vol. 181, p. 107929, 2021.

[5] S. M. Kuo and D. R. Morgan, "Active noise control: a tutorial review," *Proc. IEEE*, vol. 87, no. 6, pp. 943–973, 1999.

[6] S. J. Elliott and P. A. Nelson, "Active noise control," *IEEE Signal Process. Mag.*, vol. 10, no. 4, pp. 12–35, 1993.

[7] F. Yang, Y. Cao, M. Wu, F. Albu, and J. Yang, "Frequency-domain filtered-x lms algorithms for active noise control: A review and new insights," *Appl. Sci.*, vol. 8, no. 11, p. 2313, 2018.

[8] J. Zhang, T. D. Abhayapala, W. Zhang, P. N. Samarasinghe, and S. Jiang, "Active noise control over space: A wave domain approach," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 4, pp. 774–786, 2018.

[9] Y. Maeno, Y. Mitsufuji, and T. D. Abhayapala, "Mode domain spatial active noise control using sparse signal representation," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2018, pp. 211–215.

[10] H. Zhang and D. Wang, "Deep anc: A deep learning approach to active noise control," *Neural Networks*, vol. 141, pp. 1–10, 2021.

[11] S. Spors and H. Buchner, "Efficient massive multichannel active noise control using wave-domain adaptive filtering," in *Proc. IEEE 3rd Int. Symp. Communications, Control and Signal Processing (ISCCSP)*. IEEE, 2008, pp. 1480–1485.

[12] F. Ma, "Active control of outgoing noise fields in rooms," Ph.D. dissertation, The Australian National University, 2021.

[13] V. Patel, J. Cheer, and S. Fontana, "Design and implementation of an active noise control headphone with directional hear-through capability," *IEEE Trans. Consumer Electron.*, vol. 66, no. 1, pp. 32–40, 2020.

[14] K. Arikawa, S. Koyama, and H. Saruwatari, "Spatial active noise control based on individual kernel interpolation of primary and secondary sound fields," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2022, pp. 1056–1060.

[15] N. Murata, J. Zhang, Y. Maeno, and Y. Mitsufuji, "Global and local mode-domain adaptive algorithms for spatial active noise control using higher-order sources," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2019, pp. 526–530.

[16] H. Sun, T. D. Abhayapala, and P. N. Samarasinghe, "Time domain spherical harmonic analysis for adaptive noise cancellation over a spatial region," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. IEEE, 2019, pp. 516–520.

[17] F. Ma, W. Zhang, and T. D. Abhayapala, "Active control of outgoing broadband noise fields in rooms," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 529–539, 2020.

[18] K. Iwai, S. Kinoshita, and Y. Kajikawa, "Multichannel feedforward active noise control system combined with noise source separation by microphone arrays," *J. Sound Vib.*, vol. 453, pp. 151–173, 2019.

[19] T. Xiao, B. Xu, and C. Zhao, "Spatially selective active noise control systems," *J. Acoust. Soc. Am*, vol. 153, no. 5, pp. 2733–2733, 2023.

[20] T. Miyake, K. Iwai, and Y. Kajikawa, "Head-mounted multi-channel feedforward active noise control system for reducing noise arriving from various directions," *IEEE Access*, 2023.

[21] P. Zachos and J. Mourjopoulos, "Beamforming headphone anc for targeted noise attenuation," in *Audio Eng. Soc. Conv. 154*. Audio Engineering Society, 2023.

[22] H. L. V. Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2002.

[23] V. Pulkki, *Spatial sound generation and perception by amplitude panning techniques*. Helsinki University of Technology, 2001.

[24] Henry5.mp3. (2006). [Online]. Available: https://freesound.org/people/acclivity/sounds/24096/

[25] C. K. Reddy, E. Beyrami, J. Pool, R. Cutler, S. Srinivasan, and J. Gehrke, "A scalable noisy speech dataset and online subjective test framework," in *Proc. Interspeech*, 2019, pp. 1816–1820.