An Integrated Active Noise Control and Crosstalk Cancellation System Designed Under a Generalized Model-Matching Framework

Michael Anthony[1*], Chih Yen Wang[1], Ching En Huang[1], You Siang Chen[2], and Mingsian R. Bai[1,2]
[1]Department of Electrical Engineering, [2]Department of Power Mechanical Engineering
National Tsing Hua University, Taiwan
[*]E-mail: michaelzhang220@gmail.com Tel: +886-966638837

***Abstract*— In this study, active noise control (ANC) and crosstalk cancellation (XTC) are formulated in a generalized model-matching framework. A unified system that jointly cancels undesired noise and synthesizes a binaural audio effect using a loudspeaker array is proposed. To design the ANC controller, we explore two approaches: a model-based approach based on the underdetermined multichannel inverse filtering and a learning-based approach utilizing the fully-convolutional time-domain audio separation network (Conv-TasNet). For the XTC, the frequency-domain underdetermined multichannel inverse filtering (FUMIF) approach is employed for the design of the inverse filters. The outputs from both subsystems are then merged to cancel the primary noise while preserving a binaural effect for the listener. The results of the study confirm the efficacy of the proposed system, demonstrating satisfactory noise reduction and binaural audio effects in several objective metrics.**

I. INTRODUCTION

Active noise control (ANC) generates anti-noise signals through secondary loudspeakers to cancel undesired noise [1–4]. Adaptive filtering techniques, such as the filtered-x least mean squares (FxLMS) algorithm [5–7], effectively address the time-varying noise issues. However, these methods are usually constrained by the overdetermined inverse filtering problem, limiting the achievable noise reduction performance. Although time-domain underdetermined multichannel inverse filtering (TUMIF) [8] mitigates this limitation, it relies on the assumption that the noise input signal is accessible. This is an impractical condition in real-world scenarios where the noise input is typically inaccessible. Inspired by the global audio telepresence (GOAT) system [9], we reformulate ANC as a generalized model matching (GMM) problem, enabling the noise source to be picked up by microphone array and the anti-noise to be generated by loudspeaker array. Additionally, the learning-based approach, employing the fully-convolutional time-domain audio separation network (Conv-TasNet) [10] is explored. This method replaces traditional DSP-based filters with a trainable neural network, enabling direct prediction of the anti-noise signal from reference microphone inputs.

The rendering of binaural audio was designed primarily for headphone playback. When loudspeakers are used instead, crosstalk cancellation (XTC) becomes essential to mitigate the undesired crosstalk that occurs along the contralateral paths between the loudspeakers and the listener's ears. Previous studies have explored various XTC approaches [11–16], including the use of TUMIF [17] which effectively broadens the sweet spot. In this paper, we employ frequency-domain underdetermined multichannel inverse filtering (FUMIF) method inspired by [18] to design the XTC controllers.

Both ANC and XTC can be viewed as a model matching problem for loudspeaker array. With ANC reformulated as GMM problem, the combined system can now be viewed entirely within a GMM framework for an integrated microphone-loudspeaker array system, cancelling undesired primary noise while synthesizing a binaural audio effect.

II. GENERALIZED MODEL MATCHING FRAMEWORK

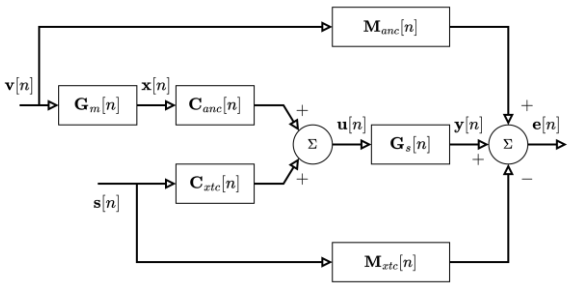


Fig. 1 The block diagram of the GMM framework.

As shown in Fig. 1, the generalized model matching framework is divided into two sections: ANC in the upper half and XTC in the lower half. Let $\mathbf{M}_{xtc}[n]$ denote the binaural target model, $\mathbf{M}_{anc}[n]$ the primary path between the noise source $\mathbf{v}[n]$ and the control points, $\mathbf{G}_m[n]$ the reference path between the noise source $\mathbf{v}[n]$ and the reference microphones, $\mathbf{G}_s[n]$ the secondary path between the secondary loudspeakers and the control points, $\mathbf{C}_{xtc}[n]$ the control filter for ANC, and

979-8-3315-7206-8/25/$31.00 ©2025 IEEE

$\mathbf{C}_{anc}[n]$ the control filter for XTC, then the goal of design in light of the GMM framework is to minimize

$$\mathfrak{I}_{anc} = \left\| \mathbf{G}_s \mathbf{C}_{anc} \mathbf{G}_m + \mathbf{M}_{anc} \right\|_F^2 \tag{1}$$

and

$$\mathfrak{I}_{xtc} = \left\| \mathbf{G}_s \mathbf{C}_{xtc} - \mathbf{M}_{xtc} \right\|_F^2 \tag{2}$$

such that the output signal $\mathbf{u}[n]$ comprises both the anti-noise and crosstalk-cancelling signals, and the error signal $\mathbf{e}[n]$ ideally contains no residual noise, preserving only the intended binaural audio effect, with n denoting the discrete time index and the notation $\|\cdot\|_F^2$ representing the Frobenius norm.

III. Proposed Method

The proposed method is divided into two stages: design stage and testing stage. During the design stage, the acoustic transfer functions (ATFs) of the reference path, primary path, and secondary path are measured. These measured ATFs are then utilized to compute the coefficients of the corresponding controllers, $\mathbf{C}_{anc}$ and $\mathbf{C}_{xtc}$. Next, these control filters are converted into causal finite impulse response (FIR) filters suitable for real-time implementation. During the testing stage, the sensors at control points are removed, and the proposed GMM framework operates to suppress unwanted noise while simultaneously synthesizing the desired binaural audio effect.

A. Causal GMM ANC

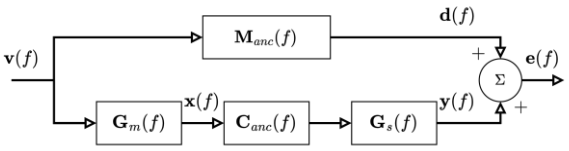


Fig. 2 The block diagram of the GMM problem for ANC.

Fig. 2 depicts the GMM architecture similar to the GOAT system [9], except without the modeling delay, which is impractical to introduce into the primary path $\mathbf{M}_{anc}(f)$ in the ANC scenario. The negative sign in (1) is due to the summing junction which represents the error microphone. Let J, N_m, N_s, and N_c represent the number of noise sources, reference microphones, secondary loudspeakers, and control points, respectively. The noise source $\mathbf{v}(f)$ propagates through the reference path $\mathbf{G}_m(f)$, which is an $N_m \times J$ matrix, and is captured by the reference microphone. The reference signal $\mathbf{x}(f)$ is multiplied by the ANC controller $\mathbf{C}_{anc}(f)$, an $N_s \times N_m$ matrix, producing the output signal, which drives the secondary loudspeaker. This signal subsequently propagates through the secondary path $\mathbf{G}_s(f)$, an $N_c \times N_s$ matrix. The anti-noise signal $\mathbf{y}(f)$ combines with the primary noise signal $\mathbf{d}(f)$ at the control points in a hope of minimizing the residual error signal $\mathbf{e}(f)$.

For simplicity, the frequency index f will be omitted hereafter. By applying a straightforward algebra to (1), the optimal solution can be calculated as

$$\hat{\mathbf{C}}_{anc} = -\mathbf{G}_s^{+} \mathbf{M}_{anc} \mathbf{G}_m^{+} \tag{3}$$

where the $\hat{\mathbf{C}}_{anc}$ is the optimal filter coefficients for ANC, and $\mathbf{M}_{anc}$, $\mathbf{G}_s$, and $\mathbf{G}_m$ correspond to the premeasured ATFs. Here, $\mathbf{G}_s$ and $\mathbf{G}_m$ are assumed to be a full row rank underdetermined system and full column rank overdetermined system, respectively, ensuring the invertibility of $\mathbf{G}_s \mathbf{G}_s^H$ and $\mathbf{G}_m^H \mathbf{G}_m$. Thus, (3) can be written as

$$\hat{\mathbf{C}}_{anc} = -\mathbf{G}_s^H \left(\mathbf{G}_s \mathbf{G}_s^H \right)^{-1} \mathbf{M}_{anc} \left(\mathbf{G}_m^H \mathbf{G}_m \right)^{-1} \mathbf{G}_m^H. \tag{4}$$

Additionally, two regularization parameters, α and β, are introduced to mitigate the ill-posedness:

$$\hat{\mathbf{C}}_{anc} = -\left(\mathbf{G}_s^H \mathbf{G}_s + \alpha \mathbf{I} \right)^{-1} \mathbf{G}_s^H \mathbf{M}_{anc} \mathbf{G}_m^H \left(\mathbf{G}_m \mathbf{G}_m^H + \beta \mathbf{I} \right)^{-1}. \tag{5}$$

Since no modeling delay is permitted here, directly converting (5) into FIR filters via inverse Fourier transform may result in non-causal filters. To convert a given frequency response function (FRF) into a causal FIR filter, the GMM procedure must satisfy a causality condition. Specifically, the effective delay of the primary path $\mathbf{M}_{anc}$ must be greater than the combined delay of the reference path $\mathbf{G}_m$ and the secondary path $\mathbf{G}_s$. That is,

$$\Delta = \text{delay}\left(\mathbf{M}_{anc} \right) - \left[\text{delay}\left(\mathbf{G}_m \right) + \text{delay}\left(\mathbf{G}_s \right) \right] > 0. \tag{6}$$

Hence, the FRF of a causal FIR filter can be written as

$$H\left(e^{j\omega} \right) = \sum_{n=0}^{N-1} c_n e^{-j\omega n} \tag{7}$$

where N is the FIR filter length estimated from Δ / T, with T being the sampling period. Sampling the FRF at $M \geq N$ frequencies leads to a linear system of equations

$$\begin{bmatrix} H\left(e^{j\omega_0} \right) \\ H\left(e^{j\omega_1} \right) \\ \vdots \\ H\left(e^{j\omega_{M-1}} \right) \end{bmatrix} = \begin{bmatrix} 1 & e^{-j\omega_0} & \cdots & e^{-j(N-1)\omega_0} \\ 1 & e^{-j\omega_1} & \cdots & e^{-j(N-1)\omega_0} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\omega_{M-1}} & \cdots & e^{-j(N-1)\omega_{M-1}} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{N-1} \end{bmatrix} \tag{8}$$

or in a more succinct matrix form

$$\mathbf{h} = \mathbf{E}\mathbf{c}. \tag{9}$$

In order to obtain real-valued filter coefficients, the real and imaginary parts of $\mathbf{h}$ and $\mathbf{E}$ are separated and stacked as follows

$$\mathbf{h}' = \begin{bmatrix} \mathbf{h}_R \\ \mathbf{h}_I \end{bmatrix}_{2M \times 1}, \quad \mathbf{E}' = \begin{bmatrix} \mathbf{E}_R \\ \mathbf{E}_I \end{bmatrix}_{2M \times N}, \tag{10}$$

where $\mathbf{h}_R$ and $\mathbf{h}_I$ denote the real and imaginary parts of vector $\mathbf{h}$, respectively, and $\mathbf{E}_R$ and $\mathbf{E}_I$ denote the real and imaginary parts of matrix $\mathbf{E}$, respectively. Thus, the optimal real-valued FIR filter coefficients $\hat{\mathbf{c}}$ can be obtained through a regularized least-squares solution:

$$\hat{\mathbf{c}} = \left(\mathbf{E}'^T \mathbf{E}' + \varepsilon \mathbf{I} \right)^{-1} \mathbf{E}'^T \mathbf{h}' \tag{11}$$

where ε is the regularization parameter.

B. Conv-TasNet ANC

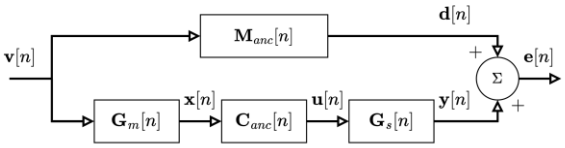


Fig. 3 The block diagram of the learning-based ANC system.

As shown in Fig. 3, the learning-based ANC architecture follows the structure of the model-based ANC. However, rather than analytically solving for the optimal control filter $\mathbf{C}_{anc}(f)$, a trainable neural network, Conv-TasNet, is employed to produce the anti-noise signal in the time-domain. During training, the reference signal $\mathbf{x}[n]$ is used as the network input. The output of the network is the estimated anti-noise signal, which subsequently propagates through the secondary path $\mathbf{G}_s[n]$ to cancel noise at the control points. The ground-truth target is defined as

$$\mathbf{d}[n] = \mathbf{M}_{anc}[n] * \mathbf{v}[n] \tag{12}$$

where the asterisk $*$ denotes the linear convolution. Thus, the error signal used during training is formulated as

$$\mathbf{e}[n] = \mathbf{d}[n] + \mathbf{G}_s[n] * \mathbf{C}_{anc}[n] * \mathbf{x}[n] \tag{13}$$

The network architecture closely follows that presented in [10] with slight modifications in the hyperparameters as detailed in Table I.

Table I. The hyperparameters of the Conv-TasNet architecture

Number of filters in autoencoder (N)	512
Length of the filters (L)	40
Number of channels in 1×1-conv block (B)	128
Number of channels in convolutional blocks (H)	512
Kernel size in convolutional blocks (P)	3
Number of convolutional blocks in each repeat (X)	3
Number of repeats (R)	3

C. FUMIF XTC

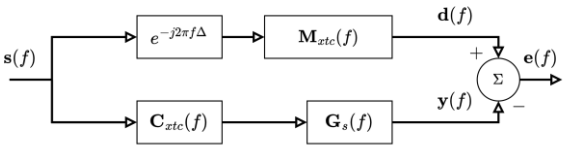


Fig. 4 The block diagram of the multichannel model matching problem for XTC.

The XTC aims to create a headphone-equivalent auditory experience by cancelling the crosstalk present in the contralateral paths while preserving the ipsilateral signals. Therefore, the target model is chosen as

$$\mathbf{M}_{xtc} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{14}$$

As illustrated in Fig. 4, the multichannel model matching architecture is formulated in the frequency domain, incorporating a modeling delay to ensure the causality of the XTC controllers. Let K denote the number of program signals. The program signal $\mathbf{s}(f)$ is multiplied by the XTC controller $\mathbf{C}_{xtc}(f)$, an $N_s \times K$ matrix, producing the output signal that drives the secondary loudspeaker. The objective is to minimize the mismatch between the generated crosstalk-cancelling signal $\mathbf{y}(f)$ and the desired signal $\mathbf{d}(f)$.

By neglecting the modeling delay for simplicity, the model matching cost function can be written as

$$\hat{\mathbf{C}}_{xtc} = \arg\min_{\mathbf{C}_{xtc}} \mathfrak{I} = \arg\min_{\mathbf{C}_{xtc}} \left\| \mathbf{M}_{xtc} - \mathbf{G}_s \mathbf{C}_{xtc} \right\|_F^2 \tag{15}$$

where the $\hat{\mathbf{C}}_{xtc}$ is the optimal filter coefficients for XTC. Taking the complex gradient, the optimal solution can be computed as

$$\hat{\mathbf{C}}_{xtc} = \left(\mathbf{G}_s^H \mathbf{G}_s \right)^{-1} \mathbf{G}_s^H \mathbf{M}_{xtc}. \tag{16}$$

As mentioned earlier, $\mathbf{G}_s$ is assumed to be full row rank ensuring the invertibility of $\mathbf{G}_s^H \mathbf{G}_s$. A regularization parameter γ is introduced to prevent control filter gain saturation due to ill-posedness:

$$\hat{\mathbf{C}}_{xtc} = \left(\mathbf{G}_s^H \mathbf{G}_s + \gamma \mathbf{I} \right)^{-1} \mathbf{G}_s^H \mathbf{M}_{xtc}. \tag{17}$$

To obtain causal FIR filters, an inverse Fourier transform along with a circular shift (hence the modeling delay) is required.

IV. Simulation

A. Simulation Setup

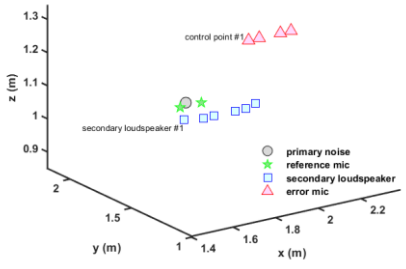


Fig. 5 Loudspeakers and microphones placement.

A cuboid room measuring $5 \times 6 \times 2.5$ m^3 is considered in this simulation. As shown in Fig. 5, the loudspeakers are positioned one meter above the floor, while the microphones are placed 30 cm higher. The ATFs are generated by using Image Source Method (ISM) [19], assuming a reverberation

time (T60) of 150 ms. Control points are numbered incrementally from left to right, starting with the first red triangle on the left. The regularization parameters α, β, γ, and ε are configured as 10^{-3}, 10^{-2}, 10^{-4}, and 10^{-4}, respectively.

To evaluate the ANC performance, a $2 \times 6 \times 4$ ANC system is implemented, comprising two reference microphones, six secondary loudspeakers, and four control points. The mean squared noise reduction (MSNR) [20] is used as the evaluation metric which is defined as

$$MSNR = 10 \log_{10} \frac{\sum_{n=0}^{N} |d_j[n]|^2}{\sum_{n=0}^{N} |e_j[n]|^2} \quad (18)$$

where $d_j[n]$ and $e_j[n]$ represent the primary noise signal and the residual error signal at the j-th control point, respectively. A higher MSNR value indicates better noise control performance.

The training configuration for the Conv-TasNet is summarized in Table II. As for the training data, two reference microphone signals are generated by convolving the noise dataset, which consists of fan noise from [21] and white noise, with the corresponding reference paths.

Table II. The setting of Conv-TasNet model training

Sampling rate	16 kHz
Loss function	Mean Square Error (MSE)
Activated function	tanh
Optimizer	Adam
Learning rate	0.0001
Batch size	8
Epoch	25

For evaluating the XTC performance, a 2×2 XTC system is employed, consisting of two secondary loudspeakers (loudspeaker #2 and #5) and two control points (control point #2 and #3). The configuration uses 1025 frequency bins and a modeling delay of 120 samples, without applying windowing or truncation in the inverse Fourier transform. The channel separation [22] is utilized as the evaluation metric which is given as

$$\text{channel separation} = 10 \log_{10} \frac{\sum_{f=0}^{fs/2} |P_B(f)|^2}{\sum_{f=0}^{fs/2} |P_D(f)|^2}, \quad (19)$$

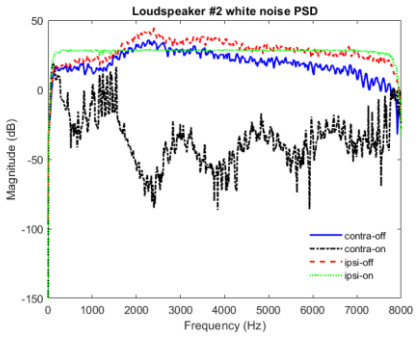
where $P_B(f)$ and $P_D(f)$ denote the signals measured in the bright zone and the dark zone. Higher channel separation value indicates better XTC performance.

B. Simulation Result

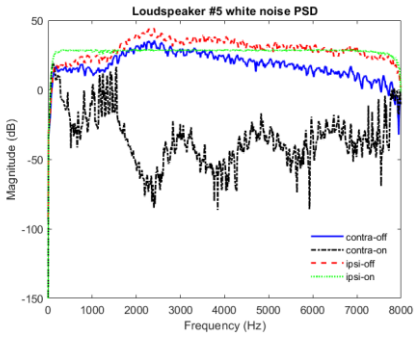
In order to evaluate the ANC performance, white noise is employed as the primary noise source. Table III summarizes the MSNR values obtained by the model-based and learning-based ANC methods across four control points. It can be observed that the learning-based approach consistently outperforms the model-based method, achieving significantly higher MSNR values at approximately 20 dB compared to 16 dB.

Table III. ANC performance evaluated in terms of MSNR (dB)

Control point	Model-based	Learning-based
Microphone #1	16.15	20.12
Microphone #2	16.3	20.02
Microphone #3	16.4	20.06
Microphone #4	16.32	20.04



(a)



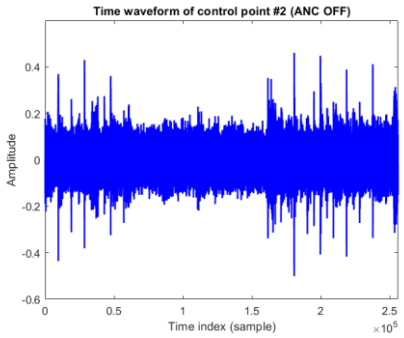
(b)

Fig. 6 Comparison of ipsilateral and contralateral paths (a) loudspeaker #2 to control points #2 and #3 (b) loudspeaker #5 to control points #2 and #3.

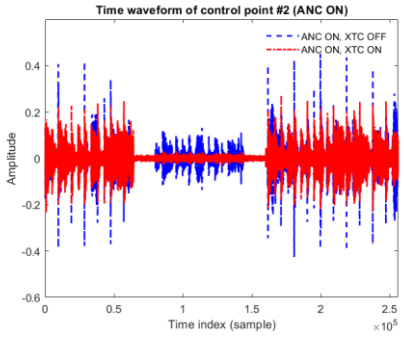
Fig. 6 presents the XTC performance, depicting the power spectral density (PSD) of loudspeakers under white noise excitation. Each plot compares ipsilateral and contralateral signals between loudspeakers and control points before ("ipsi-off" and "contra-off") and after ("ipsi-on" and "contra-on") the application of XTC processing. Significant attenuation in the contralateral paths across a broad frequency range is clearly observed, while ipsilateral signals remain nearly unaffected. Table IV quantitatively supports these observations, demonstrating a substantial improvement in channel separation of approximately 62 dB for both loudspeakers.

Table IV. The channel separation measured on the contralateral paths

Channel separation	Unprocessed (dB)	Processed (dB)
Loudspeaker #2	9.095	62.360
Loudspeaker #5	11.292	62.655

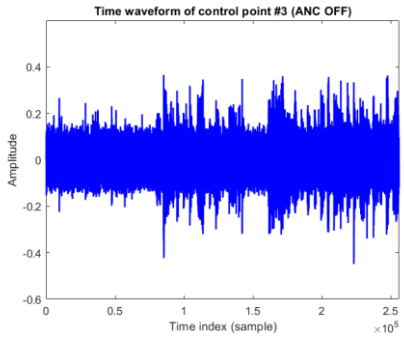


(a)

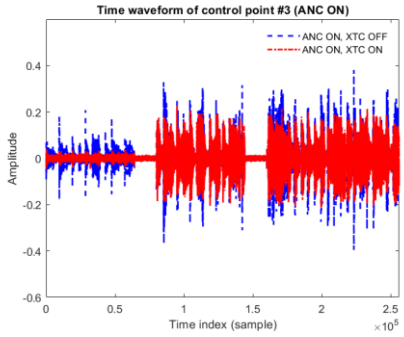


(b)

Fig. 7 Time waveform comparison at control point #2 (a) ANC off (b) ANC on.



(a)



(b)

Fig. 8 Time waveform comparison at control point #3 (a) ANC off (b) ANC on.

For verifying the combined ANC and XTC performance, learning-based ANC method is utilized due to its superior performance. Similar to previous tests, white noise serves as the primary noise source. The audio signal for the XTC evaluation is segmented into three distinct parts: male-only voice, female-only voice, and both voices combined, as illustrated in Figs. 7 and 8. After XTC processing the expected outcome at control point #2 is that only male voice is audible, whereas control point #3 exclusively presents the female voice.

As depicted in Figs. 7(a) and 8(a), the original audio signals are inaudible due to masking by the primary noise. In contrast, Figs. 7(b) and 8(b) clearly demonstrate the effectiveness of the noise cancellation. The blue waveform (dashed line) indicates the audio signal before the XTC processing takes place. After processed by the XTC, the result illustrated by the red waveform (dashed dotted line) shows that the male voice is predominantly audible at control point #2, while the female voice is clearly audible at control point #3. This confirms the effectiveness of the proposed unified system in achieving significant noise reduction while accurately rendering spatially separated binaural audio.

V. Conclusions

In this study, we presented a unified GMM framework to integrate ANC and XTC. We compared two ANC approaches: a model-based method inspired by the GOAT system and a learning-based method leveraging the Conv-TasNet. The findings of the present study demonstrated that the learning-based approach exhibited superior noise reduction performance in comparison with the model-based counterpart. For the XTC subsystem, employing the FUMIF technique enabled precise binaural audio rendering. The integrated system successfully generated an anti-noise signal capable of simultaneously suppressing undesired noise and accurately reproducing spatially separated binaural audio signals. Simulation results validated the effectiveness of our proposed method, highlighting its significant potential for applications requiring enhanced acoustic comfort and immersive spatial audio experiences. Future work will focus on real-time implementation and developing an end-to-end unified framework to further enhance practical applicability.

VI. Acknowledgements

This work was supported by the National Science and Technology Council (NSTC) in Taiwan, Republic of China, under the Project No. 113-2221-E-007-057-MY3.

REFERENCES

[1] S. M. Kuo and D. R. Morgan, "Review of DSP algorithms for active noise control," in *Proceedings of the 2000 IEEE International Conference of Control Applications*, 2000.

[2] S. M. Kuo and D. R. Morgan, Active noise control systems: algorithms and DSP implementations, New York: Wiley, 1996.

[3] P. Lueg, "Process of silencing sound oscillations". United States Patent 2,043,416, 9 June 1936.

[4] Y. Kajikawa, W. S. Gan and S. M. Kuo, "Recent applications and challenges on active noise control," in *8th International Symposium on Image and Signal Processing and Analysis*, Tieste, Italy, 2013.

[5] B. Widrow and S. D. Stearns, Adaptive signal processing, Englewood Cliffs, NJ: Prentice Hall, 1985.

[6] J. C. Burgess, "Active adaptive sound control in a duct: a computer simulation," *Journal of the Acoustical Society of America,* vol. 70, no. 3, pp. 715-726, 1981.

[7] E. Bjarnason, "Analysis of the filtered-x LMS algorithm," *IEEE Transaction on Speech and Audio Processing,* vol. 3, no. 6, pp. 501-514, 1995.

[8] M. R. Bai, H. Y. Chen, L. Yang and S. C. Huang, "Active control of noise in a duct using the sparsely coded time-domain underdetermined multichannel inverse filters," *Journal of the Acoustical Society of America,* vol. 146, no. 2, pp. 1371-1381, 2019.

[9] Y. S. Chen, S. Y. Chen and M. R. Bai, "Realization of global audio telepresence via a learning-based model-matching approach with an acoustic array system," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Nantes, France, 2024.

[10] Y. Luo and N. Mesgarani, "Conv-TasNet: surpassing ideal time-frequency magnitude masking for speech separation," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 27, no. 8, pp. 1256-1266, 2019.

[11] M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in rooms," in *Proceedings of the IEEE*, 1963.

[12] P. Damaske and V. Mellert, "A procedure for generating directionally accurate sound images in the upper half-space using two loudspeakers," in *Acustica*, 1969.

[13] D. H. Cooper, "Calculator program for head-related transfer functions," *Journal of the Audio Engineering Society,* vol. 30, pp. 34-38, 1982.

[14] W. G. Gardner, "Transaural 3D audio," MIT Media Laboratory Tech. Report, 1995.

[15] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *Journal of the Audio Engineering Society,* vol. 37, pp. 3-19, 1989.

[16] J. L. Bauck and D. H. Cooper, "Generalized transaural stereo and applications," *Journal of the Audio Engineering Society,* vol. 44, pp. 683-705, 1996.

[17] M. R. Bai, Y. W. Chen and Y. C. Hsu, "Robust binaural rendering with the time-domain underdetermined multichannel inverse prefilters," *Journal of the Acoustical Society of America,* vol. 146, no. 2, pp. 1302-1313, 2019.

[18] M. R. Bai and C. C. Lee, "Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction," *Journal of the Acoustical Society of America,* vol. 120, no. 4, pp. 1976-1989, 2006.

[19] J. B. Alley and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America,* vol. 65, no. 4, pp. 943-950, 1979.

[20] P. L. Zhong, Y. S. Chen and M. R. Bai, "Multichannel learning-based spatially extended active noise control via model matching and sensor transfer function interpolation," in *Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, Taipei, Taiwan, 2023.

[21] R. Tanabe, H. Purohit, K. Dohi, T. Endo, Y. Nikaido and T. Nakamura, "MIMII Due: sound dataset for malfunctioning industrial machine investigation and inspection with domain shifts due to changes in operational and environmental conditions," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New York, 2021.

[22] Y. L. Parodi and P. Rubak, "A subjective evaluation of the minimum channel separation for reproducing binaural signals over loudspeakers," *Journal of the Audio Engineering Society,* vol. 59, no. 7/8, pp. 487-497, 2011.