INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

A Deep Learning Method to Multi-Channel Active Noise Control

Hao Zhang[1], *DeLiang Wang*[1,2]

[1]Department of Computer Science and Engineering, The Ohio State University, USA
[2]Center for Cognitive and Brain Sciences, The Ohio State University, USA
{zhang.6720, wang.77}@osu.edu

Abstract

This paper addresses multi-channel active noise control (MCANC) on the basis of deep ANC, which performs active noise control by employing deep learning to encode the optimal control parameters corresponding to different noises and environments. The proposed method trains a convolutional recurrent network (CRN) to estimate the real and imaginary spectrograms of all the canceling signals simultaneously from the reference signals so that the corresponding anti-noises cancel or attenuate the primary noises in an MCANC system. We evaluate the proposed method under multiple MCANC setups and investigate the impact of the number of canceling loudspeakers and error microphones on the overall performance. Experimental results show that deep ANC is effective for MCANC in various scenarios. Moreover, the proposed method is robust against untrained noises and works well in the presence of loudspeaker nonlinearity.

Index Terms: multi-channel active noise control, deep learning, deep ANC, quiet zone, loudspeaker nonlinearity

1. Introduction

Active noise control (ANC) is a noise cancellation methodology based on the principle of acoustic superposition; specifically, noise is canceled by generating another noise with the same amplitude but opposite phase [1]. It has been used in applications such as headphones [2], automobiles [3], airplanes [4], and medical equipment [5]. A recent trend extends the control region of ANC to achieve noise cancellation at multiple spatial points or within a spatial zone [6, 5, 7]. However, the performance of single-channel ANC is degraded when it comes to noise control in a spatial noise environment [8]. MCANC that employs multiple controllers, loudspeakers and microphones has, therefore, been introduced to achieve ANC in such scenarios.

A general MCANC system with I reference microphones, J canceling loudspeakers, and K error microphones is shown in Figure 1. The active noise controller takes the reference signals and error signals, sensed by the reference microphones and error microphones respectively, as inputs to adapt weights so that the canceling signals generated can superpose with the primary noises at the error microphones. Conventionally, MCANC is accomplished by optimizing controller weights through adaptive algorithms so that the sum of the error signals is minimized. Adaptive filtering algorithms such as filtered-x least mean square (FxLMS) and its extensions are the most commonly used ANC algorithms and have been extended to MCANC modules [1, 9, 8]. Fast affine projection [10], mixed-error [7], and spline adaptive filter [11] algorithms have been proposed to address MCANC problems.

Standard adaptive filters based MCANC algorithms work by estimating $J \times K$ secondary paths (i.e. from loudspeakers to error microphones) during an initial stage and using the estimated secondary paths to update controllers. Each of the canceling loudspeakers is controlled by an adaptive controller, resulting in $I \times J$ controllers [8, 11]. To achieve noise control over multiple points or within a relatively large zone, more loudspeakers need to be utilized and the computational complexity of MCANC algorithms grows accordingly [11]. Numerous efforts have been made to alleviate the computational complexity of MCANC algorithms [10, 7, 12]. However, the reduction of complexity usually comes at the expense of noise attenuation performance [13].

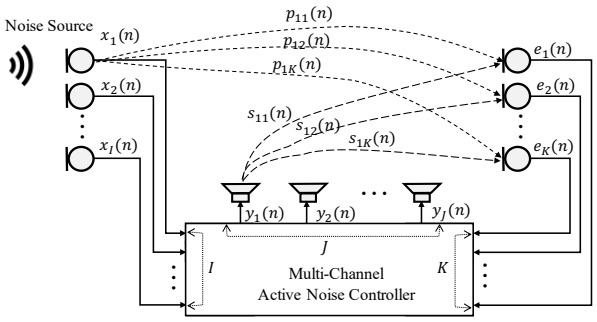

Figure 1: *Diagram of a general $I \times J \times K$ multi-channel active noise control system. p_{ik} denotes the primary path from the i^{th} reference microphone to the k^{th} error microphone. s_{jk} denotes the secondary path from the j^{th} canceling loudspeaker to the k^{th} error microphone.*

In addition, many studies assume that the ANC systems are linear. However, nonlinear effects in sensors and acoustic paths are commonplace in practical ANC systems [14]. The most common nonlinearity in an ANC system is the saturation effect caused by the limited quality of loudspeakers [15]. It has been shown that a small nonlinearity in a secondary path can have a significant impact on behavior of linear adaptive filters [16]. Having multiple loudspeakers could therefore negatively impact the overall performance.

In a previous study, we have formulated ANC as a supervised learning problem and proposed a deep learning approach, called deep ANC, to address nonlinear ANC problem [17]. It has been shown in [17] that deep ANC can be trained to achieve ANC no matter whether the reference signal is noise or noisy speech and generalizes well to different types of noises. Besides, a delay-compensated training strategy is introduced to tackle the processing latency issue of frequency-domain ANC. However, this study deals with only single-channel ANC.

This study extends deep ANC to address MCANC problems. The resulting method, called deep MCANC, is investigated for active noise control at multiple spatial points (multi-point ANC) and within a spatial zone (generating a quiet zone).

http://dx.doi.org/10.21437/Interspeech.2021-1512

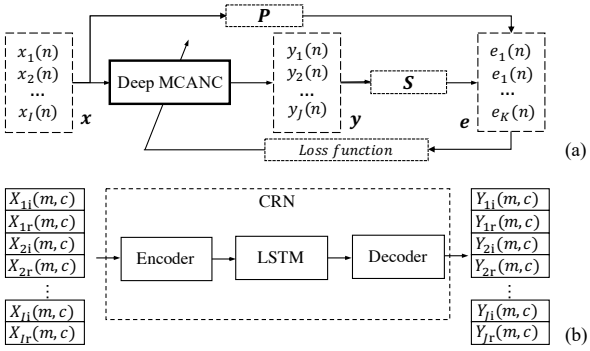

Figure 2: *Diagram of (a) the deep MCANC approach, and (b) CRN based deep MCANC. Subscripts r and i denote real and imaginary spectra of signals, respectively.*

Different from traditional MCANC methods that need to estimate multiple secondary paths and adaptive controllers individually, the proposed method trains a CRN (convolutional recurrent network) [18] to encode the optimal control parameters of an MCANC system and output multiple canceling signals simultaneously. Comparisons with adaptive filter methods under different MCANC setups are carried out, and a systematic investigation is provided on the impact of the number of loudspeakers and error microphones on the overall performance.

The remainder of this paper is organized as follows. Section 2 presents the deep learning approach to MCANC. Experiments and comparisons are given in Section 3. Section 4 concludes the paper.

2. Method description

2.1. Signal model

A general $I \times J \times K$ MCANC system is shown in Figure 1. This system consists of $I \times K$ primary paths and $J \times K$ secondary paths. It uses a set of J canceling signals $\{y_1, y_2, \cdots, y_J\}$ to generate anti-noises and cancel or attenuate primary noises at the error microphones. Ignoring the function of loudspeakers, the anti-noise generated by the j^{th} canceling loudspeaker and received by the k^{th} error microphone can be written as

$$a_{jk}(n) = s_{jk}(n) * y_j(n) \tag{1}$$

where $j = 1, 2, \cdots, J$, $k = 1, 2, \cdots, K$. Let $d_k(n)$ denote the primary noise sensed by the k^{th} error microphone, the corresponding error signal is given by

$$e_k(n) = d_k(n) - \sum_{j=1}^{J} s_{jk}(n) * y_j(n) \tag{2}$$

2.2. Deep learning for MCANC

The proposed method uses supervised learning for MCANC. Rather than estimating the multiple secondary paths and active noise controllers individually, it trains a deep neural network to directly approximate the optimal multi-channel active noise controller so as to minimize the total energy of all the error microphones under different situations. The diagram of deep learning based MCANC is given in Figure 2. The overall approach is to estimate J canceling signals simultaneously from the I reference signals so that the corresponding anti-noises attenuate the primary noises at the K error microphones.

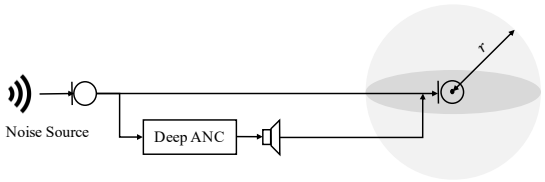

Figure 3: *Deep ANC for noise attenuation within a sphere with a radius of r to generate a quiet zone.*

In the proposed method, we use reference signals as inputs and set the ideal anti-noises as training targets. To achieve complete noise cancellation, the ideal anti-noise received at each error microphone should be the same as the corresponding primary noise. During training, the outputs of deep MCANC, canceling signals, are passed through the secondary paths to generate anti-noises. The loss function calculated from all the error signals is used for model training, as is shown in Figure 2(a). We utilize CRN as the learning machine and the diagram of it is given in Figure 2(b). The CRN is an encoder-decoder architecture, where the encoder and decoder comprise five convolutional layers and five deconvolutional layers, respectively. Between them is a two-layer LSTM with a group strategy [19], where the group number is set to 2. A detailed description of the CRN architecture is provided in [18] except that the number of input and output channels are changed to $2I$ and $2J$, respectively.

2.3. Features, training targets, and loss function

Deep MCANC utilizes the real and imaginary spectrograms of reference signals as inputs to estimate the complex spectrograms of canceling signals. The reference signals $x_i(t)$ sampled at 16 kHz are divided into 20-ms frames with a 10-ms overlap between consecutive frames. Then a 320-point short time Fourier transform is applied to each time frame to produce the real and imaginary spectrograms of $x_i(t)$, which are denoted as $X_{i\mathtt{r}}(m, c)$ and $X_{i\mathtt{i}}(m, c)$, respectively, within a T-F unit at time m and frequency c.

To attenuate the primary noises at error microphones, the ideal anti-noises are used as the training targets. The CRN is trained to output the real and imaginary spectrograms of the canceling signals, ($Y_{j\mathtt{r}}(m, c)$ and $Y_{j\mathtt{i}}(m, c)$), which are sent to the inverse Fourier transform to derive waveform signals $y_j(t)$. The anti-noises are then generated by passing the estimated canceling signals through the secondary paths using (1).

MCANC is trained to minimize the error signals received at all the error microphones and the loss function is defined as

$$\frac{\sum_{k=1}^{K} \sum_{n=1}^{L} e_k^2(n)}{KL} \tag{3}$$

where $e_k(n)$ is defined as (2), L is the length of the error signal.

2.4. Proposed method for generating a quiet zone

Besides achieving noise attenuation at multiple spatial points (locations of error microphones), deep ANC can be trained to achieve ANC within a spatial zone using one or multiple canceling loudspeakers and plenty of error microphones. The general strategy is given in Figure 3. We train the deep ANC in an RIR-independent way by exposing the model to a variety of RIRs sampled within a spatial zone during training. To be specific, we simulate the quiet zone as a sphere with a radius of r and randomly select K points within the sphere as the locations of

the error microphones. We call these error microphones virtual error microphones since they are only used during model training. Once the model is trained, they can be removed during the inference stage.

In order to achieve better noise attenuation within the quiet zone, more virtual error microphones are needed during training stage to cover as many positions within the zone as possible. This, however, leads to an increase in the amount of calculation since more error signals need to be calculated to get the loss function. For efficient training, a modified loss is introduced for this scenario, given as

$$\frac{\sum_{k=1}^{K'}\sum_{n=1}^{L} e_k^2(n)}{K'L} \tag{4}$$

where $K' < K$. Instead of calculating all the K error signals for each training sample, the modified loss only calculates K' of them, i.e., loss function of the modified training strategy is calculated from a randomly selected size-K' subset of $\{e_1(n), e_2(n), \cdots, e_K(n)\}$ each time. The model trained with the modified loss sufficiently saves the amount of calculation while still covers all the K positions within the zone.

3. Experimental results

3.1. Experimental setup

To train a noise-independent model, the training set is created by using 10000 non-speech environmental sounds from a sound-effect library (http://www.sound-ideas.com) [20]. Babble noise, factory noise, engine noise, and speech-shaped noise (denoted as SSN) from the NOISEX-92 dataset [21] are used for testing. The testing noises are unseen during training, and hence evaluate the generalization ability of the proposed method. We create 20000 training signals and 100 test signals. Each noise signal is generated by randomly cutting a 3-second-long signal from the original noise.

The primary and secondary paths are simulated as room impulse responses (RIRs) using the image method [22]. Many studies have shown the effectiveness of ANC systems for active noise canceling in enclosed rooms [23, 24, 25]. We simulate a rectangular enclosure of size 3 m × 4 m × 2 m and consider MCANC systems with single reference microphone, J canceling loudspeakers, and K error microphones, $J = 1, 2, 3$, $K = 1, 2, 3$, for multi-point ANC scenario. The reference microphone is located at the position (1.5, 1, 1) m, the three canceling loudspeakers are located at the positions (1.5, 2.5, 1) m, (1.6, 2.5, 1) m, (1.4, 2.5, 1) m, respectively, and the three error microphone are located at the positions (1.5, 3, 1) m, (1.5, 3, 1.1) m, (1.4, 3, 1) m, respectively. In the following experiments, $1 \times m \times n$ ANC system denotes the setup that uses the first m ($m \leq 3$) loudspeakers and first n ($n \leq 3$) error microphones described here. For the quiet zone scenario, we use the same J loudspeakers and set the center of the quiet zone at the position (1.5, 3, 1) m. Five reverberation times (T60s) 0.15 s, 0.175 s, 0.2 s, 0.225 s, 0.25 s are used for generating RIRs. The RIRs with reverberation time 0.2 s are used for testing.

The model is trained using the AMSGrad optimizer [26] with a learning rate of 0.001 for 30 epochs. Performance of MCANC method is evaluated in terms of normalized mean square error (NMSE), which is defined as

$$\text{NMSE} = 10\log_{10}[\textstyle\sum_{n=1}^{L} e^2(n) / \sum_{n=1}^{L} d^2(n)] \tag{5}$$

The value of NMSE is usually below zero and a lower value indicates better noise attenuation.

Table 1: *Performance comparison with different ANC setups and noises.*

		$1 \times 1 \times 1$	$1 \times 2 \times 1$	$1 \times 2 \times 2$	
Babble	FxLMS	-5.83	-7.26	-6.20	-6.38
	PMl-FxLMS	-6.05	-7.42	-6.60	-6.79
	Deep ANC	**-12.08**	**-16.27**	**-12.93**	**-13.34**
Factory	FxLMS	-5.86	-7.32	-6.14	-6.36
	PMl-FxLMS	-6.01	-7.51	-6.54	-6.78
	Deep ANC	**-11.71**	**-14.73**	**-12.26**	**-12.39**

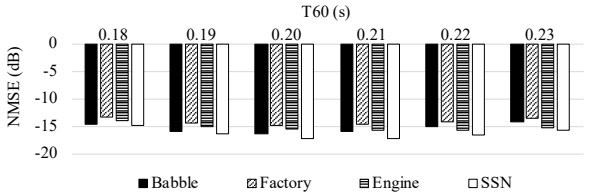

Figure 4: *Average NMSE for deep learning based* $1 \times 2 \times 1$ *ANC system with different noises and untrained RIRs with different T60s. The NMSE results for the trained T60 of 0.2 s are included as a reference.*

Table 2: *Average NMSE of deep learning based MCANC for multi-point ANC with different number of canceling loudspeakers and error microphones.*

Babble			Factory		
$1 \times 1 \times 1$	$1 \times 2 \times 1$	$1 \times 3 \times 1$	$1 \times 1 \times 1$	$1 \times 2 \times 1$	$1 \times 3 \times 1$
-12.08	-16.27	-15.52	11.71	-14.73	-14.20
$1 \times 1 \times 2$	$1 \times 2 \times 2$	$1 \times 3 \times 2$	$1 \times 1 \times 2$	$1 \times 2 \times 2$	$1 \times 3 \times 2$
-8.41	-12.93	-14.55	-7.94	-12.26	-13.53
-9.27	-13.34	-13.78	-8.70	-12.39	-12.55
$1 \times 1 \times 3$	$1 \times 2 \times 3$	$1 \times 3 \times 3$	$1 \times 1 \times 3$	$1 \times 2 \times 3$	$1 \times 3 \times 3$
-8.17	-8.36	-12.62	-7.65	-8.21	-12.00
-8.73	-9.45	-11.69	-8.22	-8.99	-10.93
-8.16	-9.72	-13.34	-7.57	-9.13	-12.11

3.2. Comparison with FxLMS based methods

We first compare the deep learning based MCANC with FxLMS and post-masking-based FxLMS (PMl-FxLMS) [27] under a single-channel ANC setup ($1 \times 1 \times 1$) and two MCANC setups ($1 \times 2 \times 1$ and $1 \times 2 \times 2$). FxLMS is the most commonly used ANC algorithm and PMl-FxLMS is a recently proposed algorithm that modifies FxLMS for faster convergence and better noise attenuation [27]. As is shown in Table 1, the proposed deep learning based method consistently outperforms the other methods under different setups. Figure 4 gives the average NMSE of deep learning based MCANC when tested with different noises and RIRs generated with different T60 values. It shows that the proposed method generalizes well to untrained noises and RIRs.

3.3. Deep MCANC for multi-point ANC

This part explores the impact of the number of loudspeakers and microphones on the deep learning based MCANC for multi-point ANC. We compare nine different ANC setups and present the results in Table 2. For an MCANC system with n error microphones, we measure the average NMSE at each microphone and provide all the n values in the table. It is seen from the table that with a fixed number of error microphones, the noise atten-

Table 3: *Performance of deep learning based MCANC for generating a quiet zone with different training strategies. r (cm) is the distance from testing positions to the center of the zone.*

		$r=0$	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	Ave
Babble	Loss in (3)	-10.65	-10.60	-10.31	-9.62	-9.46	-7.38	-9.67
	Loss in (4)	-10.49	-10.41	-10.15	-9.44	-8.97	-7.22	-9.45
Factory	Loss in (3)	-9.92	-9.89	-9.60	-8.94	-8.81	-6.66	-8.97
	Loss in (4)	-9.89	-9.83	-9.54	-8.84	-8.42	-6.52	-8.84

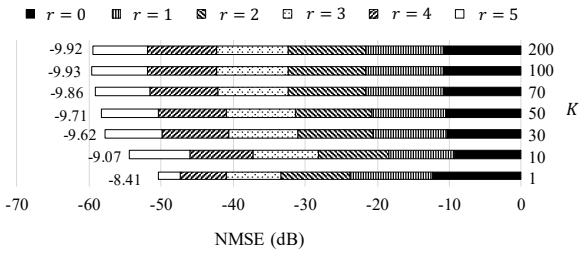

Figure 5: *Performance of deep learning based MCANC for generating a quiet zone with different number of virtual error microphones K and babble noise. The values shown on the left of the bar chat are the average values achieved within the zone.*

Table 4: *Performance of deep learning based MCANC for generating a quiet zone with different number of loudspeakers.*

		$r=0$	$r=1$	$r=2$	$r=3$	$r=4$	$r=5$	Ave
Babble	J=1	-10.49	-10.41	-10.15	-9.44	-8.97	-7.22	-9.45
	J=2	-14.15	-14.16	-13.27	-11.74	-11.53	-9.02	-12.31
	J=3	-15.53	-15.59	-15.51	-14.37	-13.69	-11.59	-14.38
Factory	J=1	-9.89	-9.83	-9.54	-8.84	-8.42	-6.52	-8.84
	J=2	-12.86	-12.87	-12.15	-10.79	-10.60	-8.25	-11.25
	J=3	-14.09	-14.15	-14.08	-13.17	-12.59	-10.87	-13.16

uation performance improves with the increase of the number of canceling loudspeakers. Meanwhile, more canceling loudspeakers are needed to achieve noise cancellation at more spatial points.

3.4. Deep MCANC for generating a quiet zone

This part studies the performance of deep learning based MCANC for generating a quiet zone. We simulate the quiet zone as a sphere and set the radius of it to $r = 5$ cm, which is appropriate, for example, for around one ear of a driver inside a vehicle. For testing, besides the center of the quiet zone ($r = 0$ cm), the performance is also evaluated at locations on spheres of different radius ($r = 1, 2, 3, 4, 5$ cm). Ten different locations are randomly generated for each test case and the average results are reported. Unless otherwise stated, the number of virtual error microphones is set to $K = 100$.

We first compare the performance of $1 \times 1 \times 100$ ANC systems trained with the loss function given in (3), and the modified loss in (4) (K' is set to 20% of K). The results given in Table 3 show that the two training strategies achieve comparable results. We will use the modified loss in the following experiments considering that the modified loss function is more efficient for model training.

Figure 5 presents the performance of the proposed method for generating the same quiet zone but with different K. It is used to show the impact of the number of virtual microphones on the quiet zone performance. Noted that the case with $K = 1$, which is actually a single-channel ANC system, is trained by putting one error microphone at the center of the sphere. Case "$K = 1$" achieves better NMSE than other cases at $r = 0$ cm. But its performance drops significantly when tested on positions that are farther from the center point of the zone. Using more virtual microphones achieves better noise attenuation within the zone but the improvement becomes insignificant when $K > 100$ for the quiet zone of a radius of 5 cm.

Table 4 shows the quiet zone performance with different number of canceling loudspeakers. Similar to the multi-point ANC scenario, having more canceling loudspeakers helps improve the overall performance in the quiet zone scenario. The average noise attenuation tested with babble noise is improved to -12.31 dB when using 2 loudspeakers, and the performance is further improved by 2.07 dB when $J = 3$.

Table 5: *Performance of deep learning based MCANC in the present of loudspeaker saturation nonlinearity.*

	Babble			Factory		
η^2	∞	0.1	0.5	∞	0.1	0.5
$1 \times 2 \times 1$	-16.43	-16.46	-16.43	-15.54	-15.57	-15.54
$1 \times 1 \times 100$	-9.64	-9.57	-9.64	-8.90	-8.66	-8.87

3.5. Nonlinear MCANC

This part studies the performance of deep MCANC in the presence of nonlinear distortions. We follow the setup given in [17, 28, 29] and simulate the loudspeaker saturation nonlinearity using the scaled error function (SEF) [30]

$$f_{\text{SEF}}(y) = \int_0^y e^{-\frac{z^2}{2\eta^2}} dz \tag{6}$$

The loudspeaker signal is generated by passing a canceling signal through the SEF function, η^2 defines the strength of nonlinearity. The SEF becomes linear as η^2 tends to infinity and becomes a hard limiter as η^2 tends to zero. We retrain the deep ANC models using four loudspeaker functions, which are $\eta^2 = 0.1$ (severe nonlinearity), $\eta^2 = 1$ (moderate nonlinearity), $\eta^2 = 10$ (soft nonlinearity), and $\eta^2 = \infty$ (linear). A randomly selected loudspeaker function is used for each input signal during training. For testing, we use both trained and untrained loudspeaker functions, the results are given in Table 5. The results for the quiet zone case ($1 \times 1 \times 100$) are the average NMSE achieved within the zone. It can be seen that deep learning based MCANC models can be trained to handle both linear and nonlinear cases and the performance generalizes well to untrained saturation effect nonlinearity ($\eta^2 = 0.5$).

4. Conclusion

This paper proposes a deep MCANC method. We have evaluated the performance of the proposed method for noise attenuation in two scenarios: multi-point ANC and generating a quiet zone, and examined the influence of the number of canceling loudspeakers and error microphones on the overall performance. Experimental results under multiple setups show that the proposed deep learning method is effective for MCANC and the active noise control performance generalizes well to untrained noises, RIRs, and saturation nonlinearities.

5. Acknowledgements

This research was supported in part by an NIDCD grant (R01 DC012048) and the Ohio Supercomputer Center.

6. References

[1] S. M. Kuo and D. R. Morgan, *Active noise control systems*. Wiley, New York, 1996, vol. 4.

[2] S. M. Kuo, S. Mitra, and W. S. Gan, "Active noise control system for headphone applications," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 2, pp. 331–335, 2006.

[3] J. Cheer and S. J. Elliott, "Multichannel control systems for the attenuation of interior road noise in vehicles," *Mechanical Systems and Signal Processing*, vol. 60, pp. 753–769, 2015.

[4] J. F. Wilby, "Aircraft interior noise," *Journal of Sound and Vibration*, vol. 190, no. 3, pp. 545–564, 1996.

[5] Y. Kajikawa, W. S. Gan, and S. M. Kuo, "Recent advances on active noise control: open issues and innovative applications," *APSIPA Transactions on Signal and Information Processing*, vol. 1, 2012.

[6] M. PAWEŁCZYK, "Active noise control-a review of control-related problems," *Archives of Acoustics*, vol. 33, no. 4, pp. 509–520, 2008.

[7] T. Murao, C. Shi, W. S. Gan, and M. Nishimura, "Mixed-error approach for multi-channel active noise control of open windows," *Applied Acoustics*, vol. 127, pp. 305–315, 2017.

[8] S. Elliott, I. Stothers, and P. Nelson, "A multiple error LMS algorithm and its application to the active control of sound and vibration," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1423–1434, 1987.

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

[10] M. Bouchard, "Multichannel affine and fast affine projection algorithms for active noise control and acoustic equalization systems," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 1, pp. 54–60, 2003.

[11] V. Patel and N. V. George, "Multi-channel spline adaptive filters for non-linear active noise control," *Applied Acoustics*, vol. 161, p. 107142, 2020.

[12] J. Lorente, M. Ferrer, M. de Diego, and A. Gonzalez, "The frequency partitioned block modified filtered-x NLMS with orthogonal correction factors for multichannel active noise control," *Digital Signal Processing*, vol. 43, pp. 47–58, 2015.

[13] D. Shi, B. Lam, W. S. Gan, and S. Wen, "Block coordinate descent based algorithm for computational complexity reduction in multichannel active noise control system," *Mechanical Systems and Signal Processing*, vol. 151, p. 107346, 2021.

[14] R. Kukde, M. S. Manikandan, and G. Panda, "Incremental learning based adaptive filter for nonlinear distributed active noise control system," *IEEE Open Journal of Signal Processing*, vol. 1, pp. 1–13, 2020.

[15] S. M. Kuo, H. T. Wu, F. K. Chen, and M. R. Gunnala, "Saturation effects in active noise control systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 6, pp. 1163–1171, 2004.

[16] M. H. Costa, J. C. M. Bermudez, and N. J. Bershad, "Stochastic analysis of the filtered-x LMS algorithm in systems with nonlinear secondary paths," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1327–1342, 2002.

[17] H. Zhang and D. L. Wang, "A deep learning approach to active noise control," *Proc. Interspeech 2020*, pp. 1141–1145, 2020.

[18] K. Tan and D. L. Wang, "Learning complex spectral mapping with gated convolutional recurrent networks for monaural speech enhancement," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 380–390, 2019.

[19] F. Gao, L. Wu, L. Zhao, T. Qin, X. Cheng, and T. Liu, "Efficient sequence learning with group recurrent networks," in *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, 2018, pp. 799–808.

[20] J. Chen, Y. Wang, S. E. Yoho, D. L. Wang, and E. W. Healy, "Large-scale training to increase speech intelligibility for hearing-impaired listeners in novel noises," *The Journal of the Acoustical Society of America*, vol. 139, no. 5, pp. 2604–2612, 2016.

[21] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech communication*, vol. 12, no. 3, pp. 247–251, 1993.

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

[23] J. W. Parkins, S. D. Sommerfeldt, and J. Tichy, "Narrowband and broadband active control in an enclosure using the acoustic energy density," *The Journal of the Acoustical Society of America*, vol. 108, no. 1, pp. 192–203, 2000.

[24] J. Cheer, "Active control of the acoustic environment in an automobile cabin," Ph.D. dissertation, University of Southampton, 2012.

[25] P. N. Samarasinghe, W. Zhang, and T. D. Abhayapala, "Recent advances in active noise control inside automobile cabins: Toward quieter cars," *IEEE Signal Processing Magazine*, vol. 33, no. 6, pp. 61–73, 2016.

[26] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of adam and beyond," *arXiv preprint arXiv:1904.09237*, 2019.

[27] C. Shi, N. Jiang, R. Xie, and H. Li, "A simulation investigation of modified FxLMS algorithms for feedforward active noise control," in *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2019, pp. 1833–1837.

[28] F. Agerkvist, "Modelling loudspeaker non-linearities," in *Audio Engineering Society Conference: 32nd International Conference: DSP For Loudspeakers*. Audio Engineering Society, 2007.

[29] S. Ghasemi, R. Kamil, and M. H. Marhaban, "Nonlinear THF-FxLMS algorithm for active noise control with loudspeaker nonlinearity," *Asian Journal of Control*, vol. 18, no. 2, pp. 502–513, 2016.

[30] W. Klippel, "Tutorial: Loudspeaker nonlinearitiescauses, parameters, symptoms," *Journal of the Audio Engineering Society*, vol. 54, no. 10, pp. 907–939, 2006.