Interspeech 2022
18-22 September 2022, Incheon, Korea

Attentive Recurrent Network for Low-Latency Active Noise Control

Hao Zhang[1], *Ashutosh Pandey*[1] *and 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, pandey.99, wang.77}@osu.edu

Abstract

Processing latency is a critical issue for active noise control (ANC) due to the causality constraint of ANC systems. This paper addresses low-latency ANC in the deep learning framework (i.e. deep ANC). A time-domain method using an attentive recurrent network is employed to perform deep ANC with smaller frame sizes, thus reducing algorithmic latency of deep ANC. In addition, a delay-compensated training strategy is introduced to perform ANC using predicted noise for several milliseconds. Moreover, we utilize a revised overlap-add method during signal resynthesis to avoid the latency introduced due to overlaps between neighboring time frames. Experimental results show that the proposed strategies are effective for achieving low-latency deep ANC. Combining the proposed strategies is capable of yielding zero, even negative, algorithmic latency without significantly affecting ANC performance.

Index Terms: Active noise control, deep ANC, attentive recurrent network (ARN), algorithmic delay, low-latency

1. Introduction

Active noise control (ANC) is a noise cancellation technology based on the principle of destructive superposition of acoustic signals. It works by generating an anti-noise with the equal amplitude and opposite phase of the primary (unwanted) noise, hence resulting in the cancellation of both when they are superposed at an error microphone [1]. Conventionally, ANC is accomplished by optimizing controller weights using adaptive algorithms so that the error signal is minimized [2]. Filtered-x least mean square (FxLMS) and its extensions are the most widely used active noise controllers and have been implemented in both time and frequency domains [3]. However, these methods are fundamentally linear and do not perform satisfactorily in the presence of nonlinear distortions [4].

Recently, deep learning has been utilized for fixed-parameter ANC [5] considering the capacity of deep neural networks in modeling complex nonlinear relationships [6, 7, 8, 9, 10]. In a previous study, we proposed a deep learning approach, called deep ANC, to address nonlinear ANC [6, 7]. Subsequently, Shi et al. introduced a deep learning based selective fixed-filter active noise control method [9]. More recently, Chen et al. proposed a secondary path-decoupled ANC method using two pre-trained convolutional recurrent networks [10].

A unique constraint of ANC is that it targets noise in physical space unlike, say, noise reduction in mobile communication. Specifically, the error microphone of an ANC system adds primary noise and anti-noise signals arriving at its location acoustically. This leads to the causality constraint of ANC systems; that is, the sum of controller processing time and the secondary path acoustic delay must be no greater than the primary path acoustic delay (the time for noise to propagate along the primary path) [11, 12]. The causality constraint is a dominant factor in the design of ANC systems, and it must be satisfied to perform noise attenuation [13]. However, block-based algorithms such as frequency-domain FxLMS and deep ANC, possess an algorithmic delay determined by the frame size since they are implemented in a frame-by-frame manner [14, 15]. This delay could violate the causality constraint and is considered a major limitation for block-based ANC algorithms.

Building on deep ANC, this paper aims to reduce its algorithmic delay and achieve low-latency deep ANC. The contributions of this paper are summarized below. First, we introduce a time-domain deep ANC utilizing attentive recurrent network (ARN) [16], which enables the implementation of deep ANC with smaller frame sizes. Second, to effectively account for algorithmic latency, a delay-compensated training strategy is proposed to perform ANC using predicted noise. Third, a revised overlap-add (OLA) method is utilized during signal resynthesis to avoid the latency introduced by overlaps between neighboring frames. Finally, we combine the proposed strategies to achieve deep ANC with zero or even negative algorithmic latency, which represents a big stride towards alleviating the causality constraint of ANC and allowing for more flexibility in ANC design.

The remainder of this paper is organized as follows. Section 2 introduces the signal model and deep ANC method. Section 3 describes proposed low-latency deep ANC techniques. Comparisons and experimental results are presented in Section 4. Section 5 concludes the paper.

2. Active noise control

2.1. Signal model

A typical feedforward ANC system consists of a reference microphone, a canceling loudspeaker, and an error microphone, as is shown in Figure 1 (a). The reference signal $x(n)$ sensed by a reference microphone is fed to the active noise controller to generate a canceling signal $y(n)$. The canceling signal is then passed through a canceling loudspeaker and the secondary path to get an anti-noise $a(n)$ in order to cancel or attenuate the primary noise $d(n)$. The corresponding error signal received at the error microphone is obtained as

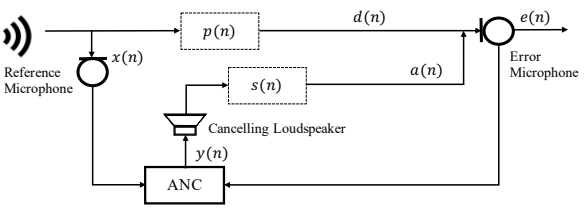
$$\begin{aligned} e(n) &= d(n) - a(n) \\ &= p(n) * x(n) - s(n) * f_{\text{LS}}\{y(n)\} \end{aligned} \tag{1}$$

where n is the time index, $p(n)$ and $s(n)$ denote the primary and secondary path, respectively, symbol $f_{\text{LS}}\{\cdot\}$ denotes the function of loudspeaker, and $*$ denotes linear convolution. Note that the anti-noise is subtracted in (1) to achieve noise cancellation.

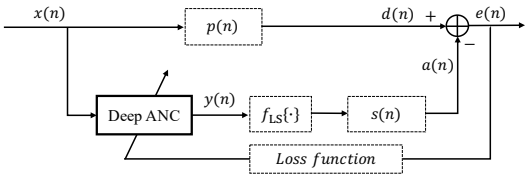
The system causality of ANC is determined by the positions of devices and the control unit's processing latency. With a given ANC setup, the algorithmic latency of a controller should be designed as low as possible to satisfy the causality constraint.

Copyright © 2022 ISCA

10.21437/Interspeech.2022-811

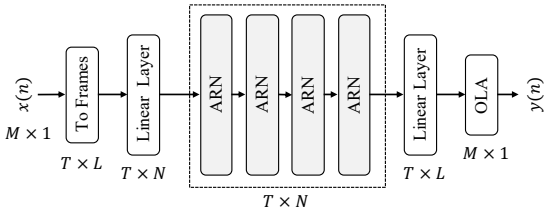


(a) Single-channel feedforward ANC system

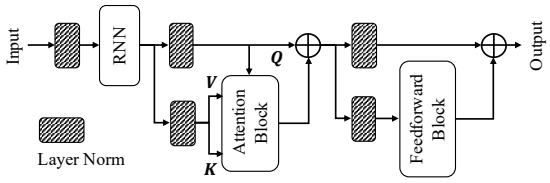


(b) Deep ANC approach

Figure 1: *Diagrams of (a) single-channel feedforward ANC system, and (b) deep ANC approach.*



(a) ARN based time domain ANC



(b) Architecture of ARN

Figure 2: *Diagrams of (a) ARN based time-domain ANC, and (b) ARN architecture.*

2.2. Deep ANC

Unlike traditional ANC methods, which require individually estimating the secondary path and adaptive filter, deep ANC employs supervised learning to train a deep neural network using large-scale multi-condition training to directly approximate an optimal controller to minimize the error signal in a variety of noisy environments [6, 7]. A diagram of the deep ANC approach is given in Figure 1(b). The overall goal is to estimate a canceling signal from the reference signal so that the corresponding anti-noise cancels the primary noise. Deep ANC takes as input a reference signal and sets the ideal anti-noise as the training target. The ideal anti-noise should be the same as the primary noise to accomplish complete noise cancellation. During training, the output of deep ANC is treated as an "intermediate product", and the anti-noise is produced by passing deep ANC output through the loudspeaker and secondary path. The loss function calculated from the error signal is then used to guide model training.

3. Low-latency deep ANC

3.1. Algorithmic latency of deep ANC

Deep ANC is a block-based method where signals are processed in a frame-by-frame manner. Specifically, an input signal is first divided into short overlapping blocks of time-domain samples and then transformed into a sequence of frames. Resynthesis of a time-domain signal is achieved by converting frames back to samples and combining neighboring frames using the overlap-add (OLA) method [17]. These procedures incur an algorithmic delay determined by the frame length and frame shift.

An illustration of OLA is given later in Figure 4(a), where a signal with M samples is chunked into T frames with a frame size of L and frame shift of J. Due to overlaps between neighboring frames, to fully synthesize a sample, all frames that contain this sample need to be processed. For example, we have to wait till the end of a frame to generate the first $J-1$ samples of it, which results in a delay in the range of $(L-J, L]$ samples. In this paper, we define the algorithmic latency as the maximum delay, L, for simplicity.

3.2. ARN based time-domain ANC

The most direct way of reducing algorithmic latency of deep ANC is to shorten the frame size. For frequency-domain methods, using a smaller frame size results in a lower frequency resolution and may degrade system performance [18]. We propose to implement deep ANC using ARN in the time-domain, which can be easily extended to be used with smaller frame sizes. Further, we find ARN to be highly effective for the deep ANC task even with smaller frame sizes.

ARN is recently proposed in [19] for efficiently combining attention mechanisms [20] with recurrent neural network (RNN). The processing flow of ARN based time-domain deep ANC is shown in Figure 2(a). A reference noise $x(n)$ with M samples is first chunked into T overlapping frames with a frame size of L and frame shift of J. Following that, a linear layer is used to project these frames to a representation of size N, which is then processed by a four-layered ARN. The output of ARNs is projected back to size L using another linear layer. Finally, OLA is utilized to obtain the waveform of canceling signal $y(n)$.

The architecture of the ARN layer used in this study is shown in Figure 2(b). It comprises an RNN with LSTM, a self-attention block, and a feed forward block. The input to ARN is firstly layer normalized [21] and fed to an RNN. The output of the RNN is then normalized using two parallel layer normalizations, where the fist layer normalized output is used as query (**Q**) and the second one is used as key (**K**) and value (**V**) for the following attention block. The output of the attention block is added to **Q** to form a residual connection. Afterwards, the final output is normalized using two separate layer normalizations, in which the first output is processed using the feedforward block and the second output is added to the output of the feedforward block in a residual way. Details on the underlying attention block and feedforward block can be found in [16].

3.3. Delay-compensated training

Another strategy for reducing latency is to perform ANC using predicted noise, and the resulting strategy is named delay-compensated training. During model training, instead of correctly aligning input and training target, we train the model to

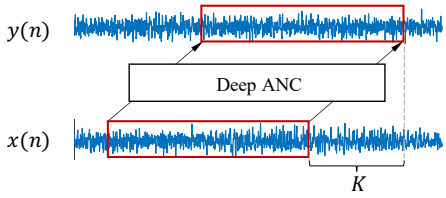
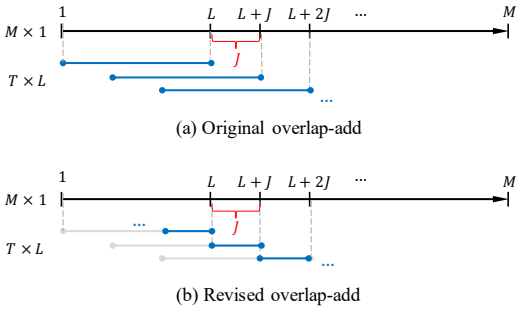


Figure 3: *Illustration of using deep ANC to predict K samples in advance.*



(a) Original overlap-add

(b) Revised overlap-add

Figure 4: *Illustrations of (a) OLA, and (b) revised OLA. The corresponding algorithmic latency are L and J, respectively.*

predict the target signal K samples in advance, as is shown in Figure 3. By doing prediction, the proposed strategy can cancel primary noise with K/f_s ms time advance, where fs denotes the sampling frequency, thus compensating for the delay introduced by the controller. The delay-compensated training proposed in this paper is a generalized version of that introduced in [7], where the latter one predicts noise on frame level while the proposed one on sample level.

Doing prediction is the roughest way of reducing algorithmic latency, and it undoubtedly degrades ANC performance to some extent. However, such a strategy is necessary for ANC tasks since it can intrinsically alleviate the causality constraint of ANC systems.

3.4. Revised overlap-add for signal resynthesis

Part of the algorithmic latency of block-based methods originates from the overlap-add procedure. Having overlaps between neighboring frames benefits from averaging/pooling of multiple frames and results in a smooth estimation. Considering that noise signals are relatively stationary and easy to estimate, we propose to revise the OLA method by setting the overlapped samples to zeros during signal resynthesis. The revised OLA, as is shown in Figure 4(b), reduces the algorithmic latency from frame size L to frame shift J. Using revised OLA reduces the algorithmic latency of deep ANC at the cost of acceptable ANC performance degradation.

4. Experiments

4.1. Experiment setting

To train a noise-independent model, we create a training set using 10000 non-speech environmental sounds from a sound-effect library (http://www.sound-ideas.com) [22]. Babble noise, engine noise, speech-shaped noise (denoted as SSN), and factory noise from NOISEX-92 dataset [23] are used for testing.

Table 1: *Average NMSE (dB) of traditional ANC methods and deep ANC models in situations with untrained noises and loudspeaker nonlinearity $\eta^2 = 0.5$. Algorithmic latency of each deep ANC model is provided inside the parentheses.*

		Linear			Nonlinear		
		Babble	Factory	SSN	Babble	Factory	SSN
FxLMS		-6.04	-5.88	-5.95	-4.32	-4.73	-4.38
THF-FxLMS		–	–	–	-6.02	-5.86	-5.98
CRN	20 ms - 10 ms (20 ms)	-10.58	-10.66	-11.36	-10.54	-10.57	-11.30
	16 ms - 8 ms (16 ms)	-10.43	-10.02	-10.76	-10.39	-9.95	-10.70
	4 ms - 2 ms (4 ms)	-9.51	-9.07	-10.25	-9.48	-9.00	-10.21
ARN	20 ms - 10 ms (20 ms)	-11.32	-11.24	-11.74	-11.29	-11.18	-11.70
	16 ms - 8 ms (16 ms)	-11.57	-11.72	-12.20	-11.55	-11.66	-12.16
	4 ms - 2 ms (4 ms)	-11.57	-11.49	-11.68	-11.56	-11.46	-11.68

The test noises are unseen during training, and hence can evaluate the generalization ability of the proposed method.

Many studies have shown the effectiveness of ANC systems for noise canceling in enclosed rooms [24, 25]. We follow the setup given in [6, 7] and simulate a rectangular enclosure of size 3 m × 4 m × 2 m (width × length × height) to carry out experiments. The primary and secondary paths are simulated as room impulse responses (RIRs) using the image method [26]. The reference microphone is located at the position (1.5, 1, 1) m, the canceling loudspeaker is located at (1.5, 2.5, 1) m, and the error microphone at (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 training RIRs. The RIRs with reverberation time 0.2 s are used for testing.

The saturation nonlinearity of loudspeaker is simulated using the scaled error function (SEF) [27, 28]:

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

where y is the input to the loudspeaker, η^2 defines the strength of nonlinearity. The SEF becomes linear as η^2 tends to infinity, and a hard limiter as it tends to zero. To investigate the robustness of the proposed method against nonlinear distortions, four loudspeaker functions are used during training: $\eta^2 = 0.1$ (severe nonlinearity), $\eta^2 = 1$ (moderate nonlinearity), $\eta^2 = 10$ (soft nonlinearity), and $\eta^2 = \infty$ (linear).

We create 20000 training signals and 100 test signals. Each noise signal is created by randomly cutting a 3-second signal from the original noise signals. The ARN model is trained using the Adam optimizer [29] with a learning rate of 0.0001 for 30 epochs. Normalized mean square error (NMSE) is used for ANC evaluation and it is defined as

$$\text{NMSE} = 10\log_{10}[\textstyle\sum_{n=1}^{M} e^2(n) / \sum_{n=1}^{M} d^2(n)] \tag{3}$$

NMSE values are typically below zero, with a lower value indicating better noise attenuation.

4.2. Deep ANC with shorter frame sizes

We first evaluate the performance of deep ANC using different frame sizes, and the results are provided in Table 1. The frame shift is set to half of the frame size, and the corresponding algorithmic latency is shown inside the parentheses. The proposed ARN based deep ANC is compared with a CRN based one introduced in [7]. ARN and CRN have similar trainable parameters while the former is implemented in time-domain and the latter in frequency-domain. Two traditional ANC algorithms, FxLMS and tangential hyperbolic function based FxLMS (THF-FxLMS) [28] are utilized for comparison. FxLMS is the most commonly used ANC algorithm and

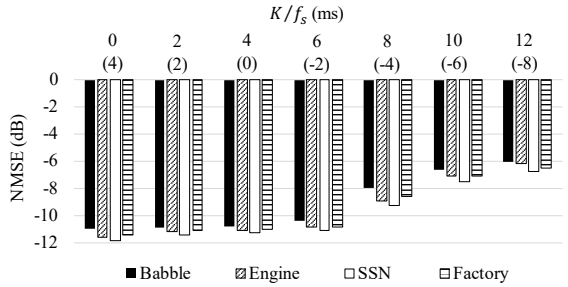


Figure 5: *Performance of ARN based deep ANC with delay-compensated training to attenuate noise with a time advance of different milliseconds. Value inside the parentheses provides algorithmic latency (in ms) of the corresponding model.*

THF-FxLMS is its modified version for handling nonlinear distortions. The step sizes of FxLMS and THF-FxLMS are chosen carefully for different noises according to the criteria given in [28, 30] to ensure stable updating and good noise attenuation.

All ANC methods are tested with untrained noises in both linear and nonlinear situations. It can be seen that ARN based time-domain ANC consistently outperforms comparison methods in both linear and nonlinear situations, and generalizes well to untrained noises. For CRN based frequency-domain method, shortening frame size results in lower ANC performance, e.g., its noise attenuation level drops by more than 1 dB when frame size is reduced from 20 ms to 4 ms. In general, ARN based time-domain method is more proper for low-latency deep ANC since reducing frame size does not affect ANC performance.

4.3. Deep ANC with delay-compensated training

This subsection investigates the performance of deep ANC with delay-compensated training. The frame size and frame shift of ARN are set to 4 ms and 2 ms, respectively. We vary the value of K and train ARN based deep ANC to cancel primary noise with a time advance of different lengths (K/fs, in ms). The predicting results are shown in Figure 5 with the algorithmic delay of each model provided inside the parentheses. The overall noise attenuation performance drops with the increase of predicting length. Predicting 6 ms in advance achieves comparable NMSE values compared to that without doing any prediction while the algorithmic latency is reduced to -2 ms. Therefore, the proposed delay-compensated training strategy can effectively reduce the algorithmic latency of deep ANC to some extent at the cost of slight performance drop.

4.4. Deep ANC with revised overlap-add

Deep ANC with revised OLA is evaluated in this part, and the NMSE values are provided in Table 2. For both CRN and ARN based deep ANC methods, using revised OLA leads to acceptable performance drop while achieving lower algorithmic latency, and ARN based method consistently outperforms CRN based method. The results of deep ANC models with the same algorithmic latency but using delay-compensated training are provided for comparison. It is seen that using revised OLA achieves better performance than doing prediction for reducing the same amount of algorithmic latency, especially for CRN based method.

Table 2: *Average NMSE (dB) of deep ANC models with revised OLA. The results of models with the same algorithmic latency but using predicted noise are provided for comparison.*

		Babble	Engine	SSN	Factory
CRN 20 ms - 10 ms	No prediction (20 ms)	-10.58	-12.87	-11.36	-10.66
	Predict 10 ms (10 ms)	-8.76	-9.69	-9.10	-8.57
	Revised OLA (10 ms)	-9.85	-11.13	-10.37	-9.86
ARN 20 ms - 10 ms	No prediction (20 ms)	-11.32	-12.67	-11.74	-11.24
	Predict 10 ms (10 ms)	-10.57	-12.14	-11.56	-10.81
	Revised OLA (10 ms)	-10.91	-12.15	-11.62	-10.98
ARN 4 ms - 2 ms	No prediction (4 ms)	-11.57	-11.96	-11.68	-11.49
	Predict 2 ms (2 ms)	-10.89	-11.18	-11.46	-11.12
	Revised OLA (2 ms)	-10.91	-11.30	-11.62	-11.22

Table 3: *Average NMSE (dB) of deep ANC models trained using different strategies to achieve zero or even negative algorithmic latency.*

ARN	Babble	Engine	SSN	Factory
20 ms - 10 ms, predict 20 ms (0 ms)	-7.62	-8.44	-8.51	-7.88
20 ms - 10 ms, revised OLA + predict 10 ms (0 ms)	-7.75	-8.46	-8.52	-7.98
20 ms - 5 ms, revised OLA + predict 5 ms (0 ms)	-10.04	-10.30	-10.85	-10.28
4 ms - 2 ms, predict 4 ms (0 ms)	-10.80	-11.12	-11.23	-11.00
4 ms - 2 ms, revised OLA + predict 2 ms (0 ms)	-10.85	-11.23	-11.58	-11.16
4 ms - 2 ms, revised OLA + predict 4 ms (-2 ms)	-10.62	-11.05	-11.28	-10.93

4.5. Low-latency deep ANC

Table 3 shows our investigations of combining the proposed strategies to achieve deep ANC with zero and even negative algorithmic latency. For purely predicting based cases, using smaller frame sizes is preferred to achieve zero algorithmic latency since the total samples need to be predicted are substantially lower. If the frame length is fixed, a shorter frame shift is desired. This is because the algorithmic latency is reduced to the length of frame shift with the help of revised OLA, using smaller frame shifts requires predicting fewer samples to achieve zero latency, which is easier than using larger ones. Smaller frame shifts, on the other hand, results in more overlaps between input frames, which is helpful for estimation. In general, ARN based time-domain ANC is effective for low-latency deep ANC. Combining ARN with the proposed strategies leads to zero or even negative algorithmic latency without affecting ANC performance much.

5. Conclusion

This study focuses on low-latency deep ANC. We have proposed a time-domain deep ANC method using an attentive recurrent network with smaller frame sizes to reduce algorithmic latency. Augmented with delay-compensated training and revised OLA, the algorithmic latency of deep ANC is substantially reduced, which goes a long way towards alleviating the causality constraint of ANC systems and facilitating ANC design. The performance of low-latency deep ANC using different strategies has been evaluated, and the combinations of these strategies lead to zero and even negative algorithmic latency. Future work includes exploring the device implementation of deep ANC and extending the proposed low-latency strategies to other tasks like speech enhancement and speaker separation.

6. Acknowledgements

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

7. References

[1] G. C. Goodwin, E. I. Silva, and D. E. Quevedo, "Analysis and design of networked control systems using the additive noise model methodology," *Asian Journal of Control*, vol. 12, no. 4, pp. 443–459, 2010.

[2] 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.

[3] 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.

[4] 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.

[5] B. Lam, W. S. Gan, D. Shi, M. Nishimura, and S. Elliott, "Ten questions concerning active noise control in the built environment," *Build. Environ.*, vol. 200, p. 107928, 2021.

[6] H. Zhang and D. L. Wang, "A deep learning approach to active noise control," in *Proceedings of the 2020 Conference of the International Speech Communication Association*, 2020, pp. 1141–1145.

[7] ——, "Deep ANC: A deep learning approach to active noise control," *Neural Netw.*, vol. 141, pp. 1–10, 2021.

[8] ——, "A deep learning method to multi-channel active noise control," *Proc. Interspeech 2021*, pp. 681–685, 2021.

[9] D. Shi, B. Lam, K. Ooi, X. Shen, and W. S. Gan, "Selective fixed-filter active noise control based on convolutional neural network," *Signal Process.*, vol. 190, p. 108317, 2022.

[10] D. Chen, L. Cheng, D. Yao, J. Li, and Y. Yan, "A secondary path-decoupled active noise control algorithm based on deep learning," *IEEE Signal Process. Lett.*, 2021.

[11] X. Kong and S. M. Kuo, "Study of causality constraint on feedforward active noise control systems," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 2, pp. 183–186, 1999.

[12] L. Zhang and X. Qiu, "Causality study on a feedforward active noise control headset with different noise coming directions in free field," *Applied Acoustics*, vol. 80, pp. 36–44, 2014.

[13] D. Shi, W.-S. Gan, B. Lam, R. Hasegawa, and Y. Kajikawa, "Feedforward multichannel virtual-sensing active control of noise through an aperture: Analysis on causality and sensor-actuator constraints," *The Journal of the Acoustical Society of America*, vol. 147, no. 1, pp. 32–48, 2020.

[14] M. Parviainen, P. Pertilä, T. Virtanen, and P. Grosche, "Time-frequency masking strategies for single-channel low-latency speech enhancement using neural networks," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2018, pp. 51–55.

[15] G. Naithani, G. Parascandolo, T. Barker, N. H. Pontoppidan, and T. Virtanen, "Low-latency sound source separation using deep neural networks," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2016, pp. 272–276.

[16] A. Pandey and D. L. Wang, "Self-attending RNN for speech enhancement to improve cross-corpus generalization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing, in press*, 2022.

[17] R. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 99–102, 1980.

[18] S. U. Wood and J. Rouat, "Unsupervised low latency speech enhancement with RT-GCC-NMF," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, pp. 332–346, 2019.

[19] S. Merity, "Single headed attention RNN: Stop thinking with your head," *arXiv preprint arXiv:1911.11423*, 2019.

[20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[21] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[22] 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," *J. Acoust. Soc. Am.*, vol. 139, pp. 2604–2612, 2016.

[23] 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 Commun.*, vol. 12, pp. 247–251, 1993.

[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] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, 1979.

[27] W. Klippel, "Tutorial: Loudspeaker nonlinearitiescauses, parameters, symptoms," *AES: J. Audio Eng. Soc.*, vol. 54, pp. 907–939, 2006.

[28] S. Ghasemi, R. Kamil, and M. H. Marhaban, "Nonlinear THF-FxLMS algorithm for active noise control with loudspeaker nonlinearity," *Asian J. Control*, vol. 18, pp. 502–513, 2016.

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

[30] W. Chen and Z. Zhang, "Nonlinear adaptive learning control for unknown time-varying parameters and unknown time-varying delays," *Asian Journal of Control*, vol. 13, no. 6, pp. 903–913, 2011.