A Deep Learning Approach to 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

We formulate active noise control (ANC) as a supervised learning problem and propose a deep learning approach, called deep ANC, to address the nonlinear ANC problem. A convolutional recurrent network (CRN) is trained to estimate the real and imaginary spectrograms of the canceling signal from the reference signal so that the corresponding anti-noise can eliminate or attenuate the primary noise in the ANC system. Large-scale multi-condition training is employed to achieve good generalization and robustness against a variety of noises. The deep ANC method can be trained to achieve active noise cancellation no matter whether the reference signal is noise or noisy speech. In addition, a delay-compensated strategy is introduced to address the potential latency problem of ANC systems. Experimental results show that the proposed method is effective for wide-band noise reduction and generalizes well to untrained noises. Moreover, the proposed method can be trained to achieve ANC within a quiet zone.

Index Terms: Active noise control, deep learning, deep ANC, spatial ANC, nonlinear distortion

1. Introduction

Active noise control is a noise cancellation methodology based on the principle of superposition of acoustic signals. The goal of ANC systems is to generate an anti-noise with the same amplitude and opposite phase of the primary (unwanted) noise in order to cancel or attenuate the primary noise [1]. Traditionally, an active noise controller is implemented using adaptive filters in a recursive way to optimize filter characteristics by minimizing an error signal. Filtered-x least mean square (FxLMS) and its extensions are the most widely used active noise controllers due to their simplicity, robustness and relatively low computational load [2]. However, nonlinear distortions are inevitably introduced to the anti-noise in applications of ANC due to the limited quality of electronic devices such as amplifiers and loudspeakers. LMS based methods are fundamentally linear and fail to identify the underlying filter accurately in the presence of nonlinearities. Even a small nonlinearity can have a significant, negative impact on the FxLMS behavior [3].

Many adaptive nonlinear ANC algorithms have been proposed to address nonlinear distortions. The Volterra expansion [4, 5] and tangential hyperbolic function based FxLMS (THF-FxLMS) [6] have been shown to be effective for modeling mild nonlinearities for nonlinear ANC. Other algorithms such as bilinear FxLMS, filtered-s LMS, and leaky FxLMS have been investigated to address nonlinearity [7]. However, their performance is limited in the presence of strong nonlinearities. Neural networks have also been introduced to address nonlinear ANC [8], considering their ability in handling nonlinear relations. A multilayer perceptron is introduced in [9] for active control of vibrations. The studies in [10] and [11] use functional link artificial neural network to handle the nonlinear effect in ANC. Other nonlinear adaptive models such as radial basis function networks [12], fuzzy neural networks [13], and recurrent neural networks [14] have been developed to further improve the ANC performance. These neural network architectures for nonlinear ANC utilize online adaptation or training to characterize an optimal controller and thus can still be regarded as adaptive algorithms.

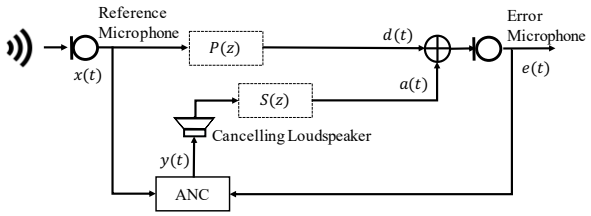

Figure 1: *Diagram of a single-channel feedforward ANC system, where $P(z)$ and $S(z)$ denote the frequency responses of primary path and secondary path, respectively.*

ANC aims to output a canceling signal to eliminate or attenuate the primary noise. In this paper, we propose a new approach, named deep ANC, to address ANC, particularly the nonlinear ANC problems. Deep learning is capable of modeling complex nonlinear relationships and can potentially play an important role in addressing nonlinear ANC problems. Specifically, a convolutional recurrent network (CRN) [15] is trained to estimate the real and imaginary spectrograms of a canceling signal from the reference signal. The subsequent anti-noise is obtained by passing the canceling signal through a loudspeaker and secondary path. Finally, the error signal is used to calculate the loss function for training the CRN model.

To the best of our knowledge, this paper represents the first study to formulate ANC as a supervised learning problem and use deep learning to address it. Our study makes four main contributions. First, complex spectral mapping is employed to estimate both magnitude and phase responses for accurate estimation [16, 17], and large-scale multi-condition training is used to attenuate a variety of noises and cope with the variations in acoustic environments. Second, in addition to attenuating noise from the noise input, we propose to train deep ANC to selectively attenuate the noise components of a noisy speech signal and let the underlying speech pass through. Namely, deep ANC in principle is able to maintain the target signal embedded in noise by selectively canceling the noise components of the noisy signal. Third, we introduce a delay-compensated training strategy to tackle a shortcoming of frequency-domain ANC algorithms: processing latency. Fourth, we expand deep ANC to perform ANC within a small spatial zone (in order to produce a quiet zone). This is a more useful but more challenging task compared to ANC at a given spatial location.

The remainder of this paper is organized as follows. Sec-

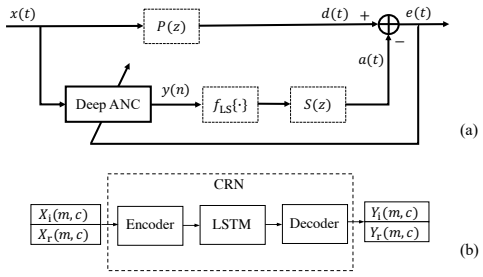

Figure 2: *Diagram of (a) the deep ANC approach, and (b) CRN based deep ANC.*

tion II presents the deep ANC approach. Evaluation metrics and experimental results are shown in Section III. Section IV concludes the paper.

2. Deep ANC

2.1. Signal model

A typical feedforward ANC system is shown in Figure 1, and it consists of a reference microphone, a canceling loudspeaker, and an error microphone. The reference signal $x(t)$ is picked up by a reference microphone. The canceling signal $y(t)$ generated by the ANC is passed through the canceling loudspeaker and the secondary path to get the anti-noise $a(t)$. The corresponding error signal sensed by the error microphone is defined as:

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

where t is the time index, $d(t)$ is the primary signal received at the error microphone, $w(t)$ represents the active noise controller, $f_{\text{LS}}\{\cdot\}$ denotes the transfer function of the loudspeaker, $*$ denotes linear convolution, and the superscript T means transpose. Furthermore, $p(t)$ and $s(t)$ denote the impulse responses of the primary and secondary path, respectively.

Adaptive algorithms alleviate the effect of the secondary path by filtering the reference signal with an estimate of the secondary path $\hat{S}(z)$ before feeding it to the controller [18]. The secondary path is usually estimated during an initial stage with separate procedures and the performance of ANC methods depends largely on the accuracy of $\hat{S}(z)$ estimation.

2.2. Deep learning for active noise control

Different from traditional ANC methods that need to estimate the secondary path and active noise controller individually, deep ANC uses supervised learning and trains a deep neural network to directly approximate an active noise controller to minimize the error signal under different situations. The diagram of deep ANC is shown in Figure 2(a). The overall approach is to estimate a canceling signal from the reference signal so that the corresponding anti-noise attenuates the primary noise. In the proposed method, we use reference signal as the input and set the ideal anti-noise as the training target. To achieve complete noise cancellation, the ideal anti-noise should be the same as the primary noise. During training, the output of deep ANC is passed through the loudspeaker and the secondary path to generate the anti-noise. The loss function calculated from the error signal is used for training the model.

Formulating ANC as a supervised learning problem is nontrivial. There are two conceptual obstacles to such a formulation. First, it is not straightforward to define what the training target should be for a deep neural network (DNN). Although the ideal canceling signal for attenuating a primary noise is known, it cannot be used directly as the desired output of the DNN due to the existence of the loudspeaker and the secondary path (see Figure 2). Second, the primary and secondary paths can be time-varying and the transfer function that the DNN needs to approximate can be different for different acoustic conditions. This seems to imply that a supervised learning model needs to predict a one-to-many mapping, an impossible job. These obstacles may explain why ANC has not been approached from the deep learning standpoint. However, as detailed in the next section, we have access to the ideal anti-noise to supervise DNN training, and the DNN can be trained to estimate, for a given input, some average of the different outputs for different scenarios. With these observations, ANC can be formulated as a deep learning task.

2.3. Feature extraction and training target

The reference signal $x(t)$ is sampled at 16 kHz and divided into 20-ms frames with a 10-ms overlap between consecutive frames. Then a 320-point short time Fourier transform (STFT) is applied to each time frame to produce the real and imaginary spectrograms of $x(t)$, which are denoted as $X_{\text{r}}(m,c)$ and $X_{\text{i}}(m,c)$, respectively, within a T-F unit at time m and frequency c. The proposed CRN based deep ANC takes $X_{\text{r}}(m,c)$ and $X_{\text{i}}(m,c)$ as input features for complex spectral mapping.

To attenuate the primary noise at the error microphone, the ideal anti-noise (the primary noise) is used as the training target. The CRN is trained to output the real and imaginary spectrograms of the canceling signal $Y_{\text{r}}(m,c)$ and $Y_{\text{i}}(m,c)$. Which are sent to the inverse Fourier transform to derive a waveform signal $y(t)$. The anti-noise is then generated by passing the canceling signal through the loudspeaker and secondary path.

2.4. Two training strategies and their loss functions

Deep ANC can be trained to achieve noise cancellation no matter whether the reference signal is noise or noisy speech by using proper training data and loss functions. Two training strategies are introduced for the deep ANC in this study:

Deep ANC trained with noise: We use noise signal $n(t)$ as the reference signal and train the deep ANC to attenuate the primary noise. The loss function is defined as:

$$Loss1 = [\textstyle\sum_{n=1}^{L} e^2(t)]/L \tag{2}$$

where L is the length of the signal, $e(t)$ is defined in (1). The model trained this way aims to cancel any signals received at the reference microphone and create a relatively silent surrounding.

Deep ANC trained with noisy speech: The deep ANC is trained to cancel surrounding noise while still catching speech signal. The reference signal used to train the deep ANC system is a mixture of noise $n(t)$ and speech $v(t)$, and the corresponding primary signal $d(t)$ is

$$\begin{aligned} d(t) &= p(t) * [v(t) + n(t)] \\ &= p(t) * v(t) + p(t) * n(t) \end{aligned} \tag{3}$$

where $p(t) * n(t)$ and $p(t) * v(t)$ are, respectively, the noise and speech components of the primary signal. In order to attenuate only noise components and let speech pass through, the training target should be set to the noise components and the ideal error signal is equivalent to $p(t) * v(t)$. The loss function is defined as:

$$Loss2 = \{\textstyle\sum_{n=1}^{L} [e(t) - p(t) * v(t)]^2\}/L \tag{4}$$

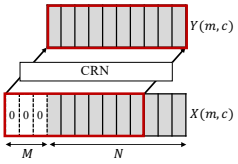

Figure 3: *Diagram of the delay-compensated strategy.*

2.5. Learning machine

The proposed deep ANC employs CRN for complex spectral mapping, as is shown 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 [15, 20]. We employ exponential linear units (ELUs) [21] in all convolutional and deconvolutional layers except the output layer. Linear activation is used in the output layer for spectrogram estimation. The model is trained using the AMSGrad optimizer [22] with a learning rate of 0.001 for 30 epochs.

2.6. Delay-compensated training

The proposed approach uses real and imaginary spectrograms as inputs and outputs and it can be regarded as a frequency-domain ANC algorithm. However, the frequency-domain ANC algorithms usually cause a time delay between the input and output of ANC [23].

A delay-compensated training strategy is proposed for the deep ANC as an alternative to address this problem. The main idea is to train a model to predict the canceling signal a few frames in advance. A diagram of this strategy is shown in Figure 3, where N denotes the total number of frames in an input signal, M denotes the number of frames we want to predict in advance. Specifically, the input signal is first revised by padding M frames of zeros in the front. Then a cut of N frames from the revised input is used as the new input signal to train the model. Since the target signal is kept unchanged, it is equivalent to use the input signal to predict M future frames of the target. In our experiments, the input signal is windowed into 20-ms frames with 10-ms frame shift. Using the delay-compensated training can save $10 \times M$ ms for the system.

3. Experimental results

3.1. Performance metrics

Performance of the proposed method is evaluated in terms of normalized mean square error (NMSE), short-time objective intelligibility (STOI) [24] and perceptual evaluation of speech quality (PESQ) [25]. NMSE is defined as:

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

The value of NMSE is usually below zero and a lower value indicates better noise attenuation. STOI and PESQ are obtained by comparing the error signal $e(t)$ with the speech component of the primary signal, $p(t) * v(t)$. A higher score indicates better intelligibility and quality.

3.2. Experiment setting

To train a noise-independent model, we use 10000 noises from a sound-effect library (http://www.sound-ideas.com) to create the training set [26]. Engine noise, factory noise and babble noise from NOISEX-92 dataset [27] are used for testing. These testing noises are unseen during training and they are used to evaluate the generalization ability of the proposed method.

The physical structure of an ANC system is usually modeled as a rectangular enclosure to show the effectiveness of ANC systems for noise canceling [28–30]. In our experiments, we simulate a rectangular enclosure of size 3 m $\times$ 4 m $\times$ 2 m and use the image method [31] to generate impulse responses (IRs) of primary and secondary paths. The reference microphone is located at the position (1.5, 1, 1) m, the canceling loudspeaker is located at the position (1.5, 2.5, 1) m and the error microphone is located 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 the IRs and the length of the them is set to 512. The IRs with reverberation time 0.2 s are used for testing.

Saturation effects produced by the loudspeakers are the most important nonlinearity in ANC systems [3, 32]. In the related studies of nonlinear ANC, the loudspeaker saturation is represented by the scaled error function (SEF) [6, 33]:

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

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 becomes a hard limiter as η^2 tends to zero. To investigate the robustness of the proposed method against nonlinear distortions, four loudspeaker transfer functions are used during training stage, which are $\eta^2 = 0.1$, $\eta^2 = 1$, $\eta^2 = 10$, and $\eta^2 = \infty$ (linear). For testing, we use both trained and untrained loudspeaker transfer functions.

The deep ANC is trained to handle cases when the reference signal is either noise or noisy speech. To achieve this, we generate 20000 training signals and 100 test signals for each case. Each noise signal is created by randomly cut a 6-second-long signal from the 10000 noises. The speech signal used to generate the noisy speech is obtained from the TIMIT dataset [34] by randomly choosing 200 speakers (100 male speakers and 100 female speakers). To create a noisy speech, utterances from a randomly selected speaker are mixed with a random noise cut from the 10000 noises at a signal-to-noise ratio (SNR) randomly chosen from [5, 10, 15, 20] dB. The signals received at the error microphone are simulated by using randomly selected loudspeaker transfer function and IRs during training.

3.3. Performance of deep ANC trained with noise

We first evaluate the performance of deep ANC system trained with noise, which is denoted as CRN_n. The proposed method is compared with FxLMS [2] and THF-FxLMS [6] in a linear system ($\eta^2 = \infty$) and two nonlinear systems ($\eta^2 = 0.5$, $\eta^2 = 0.1$). The filter length of the comparison methods is set to 512 (equals to the length of IRs) and the step sizes of them are set to the values given in [35, 36] to ensure stable updating process and good noise attenuation. Table 1 shows the average NMSE of 100 testing signals, where CRN_n-1, and CRN_n-2 denote the models trained with the delay-compensated strategy to predict 1 and 2 frames in advance, respectively. It is seen from this table that the performance of FxLMS is decreased when it comes to nonlinear systems. THF-FxLMS models the secondary path as a nonlinear model and it achieves good noise attenuation in different nonlinear systems. The deep ANC system outperforms comparison algorithms and generalizes well to

Table 1: *Performance comparison with respect to different noises and nonlinear distortions.*

Noise type	Engine			Factory			Babble		
η^2	∞	0.5	0.1	∞	0.5	0.1	∞	0.5	0.1
FxLMS	-6.78	-5.26	-4.54	-5.88	-4.73	-1.67	-6.04	-4.32	-3.37
THF-FxLMS	–	-6.70	-6.55	–	-5.86	-5.75	–	-6.02	-5.97
CRN_n	**-11.07**	**-10.98**	**-10.60**	**-9.58**	**-9.50**	**-9.17**	**-9.49**	**-9.45**	**-9.27**
CRN_n-1	-9.60	-9.53	-9.25	-8.47	-8.42	-8.19	-8.80	-8.76	8.62
CRN_n-2	-7.93	-7.89	-7.72	-6.97	-6.94	-6.81	-7.00	-7.00	-6.89

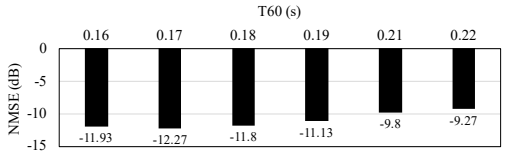

Figure 4: *Average NMSE for CRN_n with engine noise,* $\eta^2 = 0.1$*, and untrained IRs generated with different T60s.*

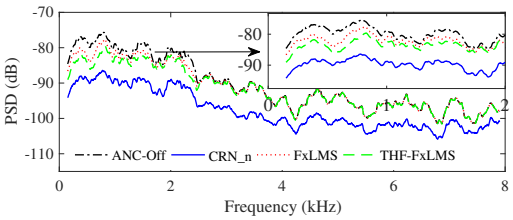

Figure 5: *Power spectrum density of ANC methods with engine noise and loudspeaker nonlinearity* η^2*=0.1.*

untrained noises and untrained nonlinearity ($\eta^2 = 0.5$). In addition, using delay-compensated strategy obtains acceptable noise attenuation while the overall performance is dropped slightly. The results in Figure 4 show that the performance of deep ANC generalizes well to untrained IRs generated with different T60s. We plot the power spectrum density (PSD) curves in Figure 5 for further comparison. It can be seen that the proposed method can achieve wide-band noise reduction while comparison methods are only effective for noise attenuation at lower frequencies.

3.4. Performance of deep ANC trained with noisy speech

This part studies the performance of the deep ANC in situations when the reference signal is a noisy speech. The model trained this way is denoted as CRN_ns. Comparison results when tested with engine noise and nonlinear system with $\eta^2 = 0.1$ are given in Table 2, where CRN_ns-1 and CRN_ns-2 denote the CRN_ns model trained with the delay-compensated strategy to predict 1, and 2 frames in advance, respectively. "Unprocessed" denotes the results when there is no ANC. The second column in Table 2 shows the NMSE values when tested with noise signals. It can be seen that the performance of CRN_ns is comparable to that of CRN_n when tested in the noise only situation even though the former model is trained with noisy speech. For situations with noisy speech, traditional methods and CRN_n focus on minimizing error signal (attenuating reference signal), and therefore distort the speech component as reflected by substantially lower STOI and PESQ values than unprocessed noisy speech. CRN_ns can improve STOI and PESQ values and the performance of CRN_ns-1 and CRN_ns-2 is comparable to that of CRN_ns with a small decrease in terms of PESQ values. The CRN_n performs best in terms of NMSE in both the noise only and noisy speech situations. This is to say, besides attenuating noise, the CRN_n model is capable of attenuating noisy speech.

Table 2: *Performance comparison for the reference signal is noisy speech with loudspeaker nonlinearity* $\eta^2 = 0.1$ *and engine noise.*

	Noise	SNR = 5 dB			SNR = 15 dB		
$\eta^2 = 0.1$	NMSE	STOI	PESQ	NMSE	STOI	PESQ	NMSE
Unprocessed	–	0.79	1.95	0	0.94	2.61	0
FxLMS	-4.54	0.71	1.84	-2.30	0.71	1.90	-0.36
THF-FxLMS	-6.55	0.69	1.73	-3.89	0.74	1.92	-1.58
CRN_n	**-10.60**	0.72	1.71	**-8.75**	0.83	2.02	**-8.66**
CRN_ns	-10.00	0.84	**2.26**	–	**0.96**	**3.00**	–
CRN_ns-1	-8.31	**0.85**	2.24	–	**0.96**	2.95	–
CRN_ns-2	-7.04	0.84	2.14	–	**0.96**	2.87	–

Table 3: *Performance of deep ANC for generating a quiet zone with engine noise and loudspeaker nonlinearity* $\eta^2 = 0.1$.

$\eta^2 = 0.1$	CRN_n	CRN_ns	
Reference signal	Noise	Noisy speech (SNR = 15 dB)	
	NMSE	STOI	PESQ
Unprocessed	–	0.94	2.63
$r = 0$	-9.44	0.96	2.88
$r = 2$	-9.49	0.96	2.90
$0 \leq r \leq 5$	-8.32	0.96	2.88

3.5. Quiet zone

Besides achieving noise attenuation at a spatial location, a more challenging task would be to achieve ANC within a small spatial zone [37]. To produce a quiet zone, we first simulate the spatial zone as a sphere with a radius of 5 cm. Then we randomly select 100 points inside the sphere as the locations of error microphone and generate 100 pairs of IRs as primary and secondary paths by using the image method [31]. 20000 training signals for noise only and noisy speech situations are created with these 100 pairs of IRs and the CRN_n and CRN_ns models are retrained with these signal. Three test sets, with 100 signals in each set, are generated to evaluate the performance of these models. The results are given in Table 3, where "$r = 0$" denotes the case where the error microphone is placed at the center point of the sphere, "$r = 2$" denotes the case with the microphone placed within the sphere and 2 m away from the center point. For the case "$0 \leq r \leq 5$", we randomly place the error microphone at 10 different points within the sphere and use the corresponding 10 pairs of IRs for testing. Generally speaking, the ANC models trained this way would achieve noise attenuation at any point within this sphere and generate a quiet zone.

4. Conclusion

In this paper, we have proposed a deep learning based approach to address the ANC problem with nonlinear distortions. The proposed deep ANC approach can be trained to not only cancel noise, but also selectively attenuate the noise components of noisy speech. We have also introduced a delay-compensated training strategy and investigated the proposed approach for spatial ANC. Systematic evaluations show the effectiveness and robustness of deep ANC for noise attenuation in noise only and noisy speech situations, and the trained DNN model generalizes well to different noises and acoustic environments. With this first successful demonstration, we anticipate that subsequent work will establish deep learning as a main approach to ANC with elevated performance.

5. Acknowledgements

This research was supported in part by two NIDCD grants (R01 DC012048 and R01 DC015521) and the Ohio Supercomputer Center.

6. 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] 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.

[4] L. Tan and J. Jiang, "Adaptive volterra filters for active control of nonlinear noise processes," *IEEE Transactions on signal processing*, vol. 49, no. 8, pp. 1667–1676, 2001.

[5] K. Lashkari, "A novel volterra-wiener model for equalization of loudspeaker distortions," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 5. IEEE, 2006, pp. V–V.

[6] 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.

[7] M. A. Sahib and R. Kamil, "Comparison of performance and computational complexity of nonlinear active noise control algorithms," *ISRN Mechanical Engineering*, vol. 2011, 2011.

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

[9] S. D. Snyder and N. Tanaka, "Active control of vibration using a neural network," *IEEE Transactions on Neural Networks*, vol. 6, no. 4, pp. 819–828, 1995.

[10] G. Panda and D. P. Das, "Functional link artificial neural network for active control of nonlinear noise processes," in *International Workshop on Acoustic Echo and Noise Control*, vol. 2003. Citeseer, 2003, pp. 163–6.

[11] T. Krukowicz, "Active noise control algorithm based on a neural network and nonlinear input-output system identification model," *Archives of Acoustics*, vol. 35, no. 2, pp. 191–202, 2010.

[12] M. Tokhi and R. Wood, "Active noise control using radial basis function networks," *Control Engineering Practice*, vol. 5, no. 9, pp. 1311–1322, 1997.

[13] Q.-Z. Zhang, W.-S. Gan, and Y.-l. Zhou, "Adaptive recurrent fuzzy neural networks for active noise control," *Journal of Sound and Vibration*, vol. 296, no. 4-5, pp. 935–948, 2006.

[14] R. T. Bambang, "Adjoint EKF learning in recurrent neural networks for nonlinear active noise control," *Applied Soft Computing*, vol. 8, no. 4, pp. 1498–1504, 2008.

[15] K. Tan and D. 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.

[16] D. S. Williamson, Y. Wang, and D. Wang, "Complex ratio masking for joint enhancement of magnitude and phase," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 5220–5224.

[17] S.-W. Fu, T.-y. Hu, Y. Tsao, and X. Lu, "Complex spectrogram enhancement by convolutional neural network with multi-metrics learning," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2017, pp. 1–6.

[18] 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.

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

[20] H. Zhang, K. Tan, and D. Wang, "Deep learning for joint acoustic echo and noise cancellation with nonlinear distortions," *Proc. Interspeech 2019*, pp. 4255–4259, 2019.

[21] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

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

[23] 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," *Applied Sciences*, vol. 8, no. 11, p. 2313, 2018.

[24] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.

[25] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, vol. 2. IEEE, 2001, pp. 749–752.

[26] 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.

[27] 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.

[28] S. D. Sommerfeldt, J. W. Parkins, and Y. C. Park, "Global active noise control in rectangular enclosures," in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, vol. 1995, no. 5. Institute of Noise Control Engineering, 1995, pp. 477–488.

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

[30] 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.

[31] 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.

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

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

[34] L. F. Lamel, R. H. Kassel, and S. Seneff, "Speech database development: Design and analysis of the acoustic-phonetic corpus," in *Speech Input/Output Assessment and Speech Databases*, 1989.

[35] 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.

[36] D. Huang and J.-X. Xu, "Discrete-time adaptive control for nonlinear systems with periodic parameters: A lifting approach," *Asian Journal of Control*, vol. 14, no. 2, pp. 373–383, 2012.

[37] 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.