Neural Networks 158 (2023) 318–327

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Deep MCANC: A deep learning approach to multi-channel active noise control

Hao Zhang [a,*], DeLiang Wang [a,b]

[a] Department of Computer Science and Engineering, Ohio State University, Columbus, OH 43210-1277, USA
[b] Center for Cognitive and Brain Sciences, Ohio State University, Columbus, OH 43210-1277, USA

ARTICLE INFO

Article history:
Received 13 September 2022
Received in revised form 15 November 2022
Accepted 22 November 2022
Available online 25 November 2022

Keywords:
Active noise control
Deep learning
Multi-channel ANC
Quiet zone
Nonlinear distortions

ABSTRACT

Traditional multi-channel active noise control (MCANC) is based on adaptive filtering and usually uses a separate control unit for each channel. This paper introduces a deep learning based approach for multi-channel active noise control (ANC). The proposed approach, called deep MCANC, encodes optimal control parameters corresponding to different noises and environments, and jointly computes the multiple canceling signals to cancel or attenuate the primary noises captured at error microphones. A convolutional recurrent network (CRN) is employed for complex spectral mapping where the summated power of error signals is used as the loss function for CRN training. Deep MCANC is a fixed-parameter ANC approach and large-scale multi-condition training is employed to achieve robustness against a variety of noises. We explore the performance of deep MCANC with different setups and investigate the impact of factors such as the number of loudspeakers and microphones, and the position of a secondary source, on ANC performance. Experimental results show that deep MCANC is effective for wideband noise reduction and generalizes well to untrained noises. Moreover, the proposed approach is robust against variations in reference signals and works well in the presence of nonlinear distortions.

1. Introduction

Active noise control (ANC) is a noise cancellation methodology based on the principle of destructive superposition of sound waves; more specifically, an unwanted primary noise is canceled by generating an anti-noise with the same amplitude but opposite phase (Kuo & Morgan, 1996). It has attracted increasing attention over the past decades, and has been used in industrial applications such as headphones (Kuo et al., 2006), automobiles (Cheer & Elliott, 2015), airplanes (Wilby, 1996), and medical equipment (Kajikawa et al., 2012). A recent trend extends the control region of ANC to achieve noise cancellation at multiple spatial points or within a spatial zone (Kajikawa et al., 2012; Murao et al., 2017; Pawełczyk, 2008). However, the performance of single-channel ANC is limited when it comes to noise control in three-dimensional space (Elliott et al., 1987). Multi-channel ANC (MCANC) that employs multiple microphones and loudspeakers has 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 Fig. 1. The MCANC system takes the reference signals and error signals, recorded by reference microphones and error microphones, respectively, as inputs to update the weights of controllers so that the canceling signals generated can superpose with the primary noises at error microphones. Conventionally, MCANC is accomplished by optimizing controller weights through adaptive algorithms so that the sum of the error signals is minimized (Liu et al., 2009; Murao et al., 2017; Patel & George, 2020). Adaptive filters such as filtered-x least mean square (FxLMS), fast affine projection, mixed-error approach, and their variable step-size versions, are commonly used ANC algorithms and have been extended to MCANC modules (Bouchard, 2003; Elliott et al., 1987; Kuo & Morgan, 1999; Liu et al., 2009; Murao et al., 2017; Patel & George, 2020).

Active noise control systems can be developed using adaptive as well as fixed-parameter (also known as fixed-coefficient or fixed-filter) techniques (Lam et al., 2021). Standard adaptive MCANC algorithms estimate the $J \times K$ secondary paths (acoustic paths from loudspeakers to error microphones) during an initial stage and use the estimated secondary paths to update controllers. For decentralized MCANC systems, each of the possible feedforward channels requires a separate adaptive filter, resulting in $I \times J$ controllers (Elliott et al., 1987; Patel & George, 2020). To achieve noise control over multiple points or within a spatial zone, more channels need to be added to the system and the computational complexity of MCANC algorithms

* Corresponding author.
E-mail addresses: zhang.6720@osu.edu (H. Zhang), dwang@cse.ohio-state.edu (D. Wang).

https://doi.org/10.1016/j.neunet.2022.11.029
0893-6080/

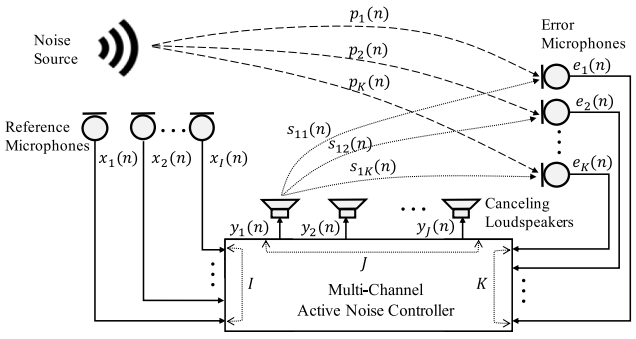

Fig. 1. Diagram of a general $I \times J \times K$ multi-channel active noise control system that consists of I reference microphones, J canceling loudspeakers, and K error microphones. Symbol p_k denotes the primary path from the noise source to the kth error microphone, and s_{jk} denotes the secondary path from the jth canceling loudspeaker to the kth error microphone.

grows accordingly (Patel & George, 2020; Shi et al., 2020). This factor acts as a bottleneck in the real-time implementation of MCANC and numerous efforts have been made to alleviate the complexity (Bouchard, 2003; Lorente et al., 2015; Murao et al., 2017). However, the reduction of computational complexity usually comes at the expense of noise attenuation performance (Shi et al., 2021). In addition, the adaptive nature of these algorithms poses an inherent risk of update divergence in the presence of external disturbances or errors in secondary path modeling (Krukowicz, 2013; Lam et al., 2021; Shi et al., 2017). Moreover, the slow convergence of the adaptive algorithms results in a limited reduction of dynamic noise, and noise reduction is not perceived immediately (Shi et al., 2022).

Fixed-parameter noise control methods utilize pre-trained control filters to attenuate noise instantaneously. They have become an effective alternative to adaptive ANC, as exemplified by noise-canceling headphones. Different from adaptive methods, fixed filters determine their coefficients during offline training and deploy these carefully tuned coefficients in actual operations (Shi et al., 2020, 2018). Tanaka et al. implement an MCANC system with fixed noise control filters to reduce computational complexity (Tanaka et al., 2014). Shi et al. propose a selective fixed-filter active noise control (SFANC) method which selects a pre-trained control filter to attenuate a detected primary noise rapidly (Shi et al., 2020). Later, a modified version of SFANC employs a convolutional neural network (Shi et al., 2022). Fixed-parameter ANC methods are feasible when the application environment does not change rapidly. However, such approaches are usually optimized for a limited range of noise types, resulting in limited generalization performance (Shi et al., 2020).

Many studies assume linear ANC systems. However, nonlinear effects caused by loudspeakers and acoustic paths are commonplace in practical ANC systems (Kukde et al., 2020), such as the gain saturation effect of loudspeakers (Kuo et al., 2004). It has been shown that a small nonlinearity in a secondary path can have a significant impact on the behavior of linear adaptive filters (Costa et al., 2002). Having multiple loudspeakers could therefore adversely impact the overall performance of an MCANC system further.

Deep neural networks have been widely used in recent years to address speech and audio processing tasks, including speaker recognition (Bai & Zhang, 2021), speech separation (Chen & Zhang, 2021; Xian et al., 2021), and noise-robust voice conversion (Du et al., 2022). In a previous study, we first formulated ANC as a supervised learning problem and proposed a deep learning approach, called deep ANC, to address the nonlinear ANC task (Zhang & Wang, 2021a). It has been shown that deep ANC is effective for wideband noise reduction and generalizes well to different noises. Later, Shi et al. proposed a deep learning based SFANC method which employs a convolutional neural network to classify noise types and identify the most suitable fixed control filter for different incoming noise (Shi et al., 2022). More recently, Chen et al. utilized deep learning to address the nonlinearity of the secondary path in an ANC system and proposed a secondary path-decoupled method using two pre-trained convolutional recurrent networks (Chen et al., 2021). Deep learning methods achieve active noise control by training a deep neural network (DNN) offline, and can therefore be viewed as fixed-parameter ANC. Compared to conventional fixed filter methods, deep ANC is capable of attaining nonlinear active noise reduction for a variety of noises through large-scale training.

This study extends deep ANC to the multi-channel domain. The resulting approach, 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). Rather than estimating multiple secondary paths and adaptive controllers individually, the proposed method trains a convolutional recurrent network (CRN) (Tan & Wang, 2019) to encode the optimal control parameters of an MCANC controller and output multiple canceling signals simultaneously so that the corresponding anti-noises match the primary noises at desired locations. As ANC is inherently sensitive to both the magnitude and phase of an anti-noise, we use complex spectral mapping to estimate both magnitude and phase responses of MCANC outputs (Tan & Wang, 2019; Williamson et al., 2016). The performance of MCANC is impacted by factors such as the number of loudspeakers and microphones, and the position of a secondary source. We systematically investigate the effects of these factors on overall performance, in order to guide design choices for MCANC systems. Furthermore, evaluations in the presence of nonlinear distortions and variations in reference signals are carried out to demonstrate the robustness of the proposed method.

Compared to a preliminary version (Zhang & Wang, 2021b), this paper conducts more extensive evaluations and investigates the performance of deep MCANC under different experimental setups. Moreover, we explore the influence of various factors on ANC performance and assess the robustness of the proposed method against variations in reference signals. We also provide new evaluation results on recorded noise signals in realistic environments and discuss the generalization ability of deep MCANC.

The remainder of this paper is organized as follows. Section 2 describes the deep MCANC approach. Section 3 describes experiment settings. Evaluation and comparison results are presented in Section 4. Section 5 provides further discussions and concludes this paper.

2. Deep MCANC

2.1. Signal model

A general $I \times J \times K$ MCANC system is shown in Fig. 1. The primary path $p_k(n)$ and secondary path $s_{jk}(n)$ correspond to the acoustic responses from a noise source and the jth canceling loudspeaker, respectively, to the kth error microphone, where $j = 1, 2, \ldots, J$ and $k = 1, 2, \ldots, K$. The reference noises $\{x_1, x_2, \ldots, x_I\}$ sensed by reference microphones are fed to multi-channel active noise controllers to get the canceling signals $\{y_1, y_2, \ldots, y_J\}$. Assuming $w_{ji}(n)$ is the active noise controller placed between the ith reference microphone and the jth canceling loudspeaker, the canceling signal at the jth loudspeaker is obtained as

$$y_j(n) = \sum_{i=1}^{I} w_{ji}^T(n) x_i(n) \tag{1}$$

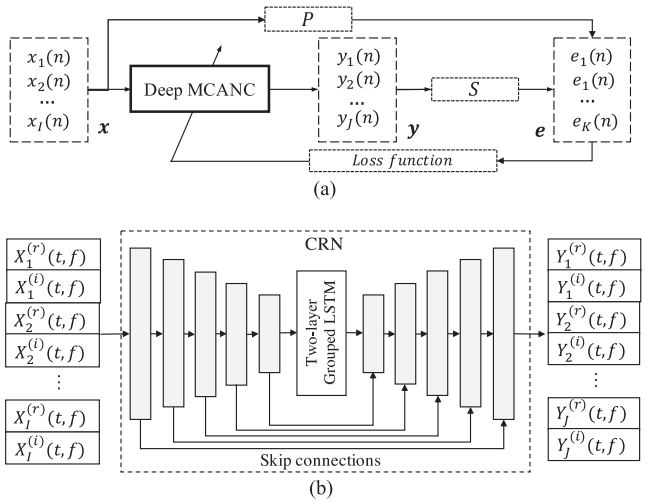

Fig. 2. Diagram of (a) deep MCANC approach, and (b) CRN based deep MCANC. P and S denoted primary and secondary paths. Superscripts (r) and (i) denote real and imaginary parts of signals, respectively.

where n is the time index, and the superscript T indicates transpose. These canceling signals are then passed through the canceling loudspeakers and the secondary paths to generate anti-noises in order to cancel or attenuate primary noises at the error microphones. The anti-noise generated by the jth canceling loudspeaker and received by the kth error microphone can be written as

$$a_{jk}(n) = s_{jk}(n) * f_{\mathrm{LS}}\{y_j(n)\} \tag{2}$$

where $f_{\mathrm{LS}}\{\cdot\}$ denotes the function of loudspeaker, and $*$ denotes convolution.

The error microphones are placed over desired locations to measure the residual noise components, and the error signal sensed by the kth error microphone is given by

$$e_k(n) = d_k(n) - \sum_{j=1}^{J} a_{jk}(n) \tag{3}$$

where

$$d_k(n) = p_k(n) * x(n) \tag{4}$$

is the primary noise received at the kth error microphone, and $x(n)$ denotes the source noise. Note that the anti-noises are subtracted in (3) to achieve noise cancellation.

2.2. Deep learning for MCANC

The proposed method uses deep learning for MCANC. It trains a DNN with large-scale multi-condition training to directly approximate an optimal MCANC controller to minimize the total energy of all the error microphones under different situations. The diagram of deep MCANC is given in Fig. 2. Our goal is to jointly estimate J canceling signals from the I reference signals so that the corresponding anti-noises cancel the primary noises at the K error microphones. In the proposed method, we use reference signals (reference noises) as inputs and set the ideal anti-noises as the training targets. To achieve complete noise cancellation at desired locations, the ideal anti-noises should be the same as the primary noises. During training, the outputs of deep MCANC (canceling signals) are treated as "intermediate products" and the anti-noises are generated by passing these products through the corresponding loudspeakers and secondary paths. The loss function calculated from all the error signals is then used to guide model training.

Table 1
Network details of the CRN architecture, where T denotes the number of time frames in a spectrogram.

Layer name	Input size	Hyperparameters	Output size
conv2d 1	$2I \times T \times 161$	$1 \times 3, (1, 2), 16$	$16 \times T \times 80$
conv2d 2	$16 \times T \times 80$	$1 \times 3, (1, 2), 32$	$32 \times T \times 39$
conv2d 3	$32 \times T \times 39$	$1 \times 3, (1, 2), 64$	$64 \times T \times 19$
conv2d 4	$64 \times T \times 19$	$1 \times 3, (1, 2), 128$	$128 \times T \times 9$
conv2d 5	$128 \times T \times 4$	$1 \times 3, (1, 2), 256$	$256 \times T \times 4$
reshape_1	$256 \times T \times 4$	–	$T \times 1024$
grouped_lstm_1	$T \times 1024$	1024	$T \times 1024$
grouped_lstm_2	$T \times 1024$	1024	$T \times 1024$
reshape_2	$T \times 1024$	–	$256 \times T \times 4$
deconv2d 5	$512 \times T \times 4$	$1 \times 3, (1, 2), 128$	$128 \times T \times 9$
deconv2d 4	$256 \times T \times 9$	$1 \times 3, (1, 2), 64$	$64 \times T \times 19$
deconv2d 3	$128 \times T \times 19$	$1 \times 3, (1, 2), 32$	$32 \times T \times 39$
deconv2d 2	$64 \times T \times 39$	$1 \times 3, (1, 2), 16$	$16 \times T \times 80$
deconv2d 1	$32 \times T \times 80$	$1 \times 3, (1, 2), 2J$	$2J \times T \times 161$

2.3. Features and training targets

The performance of ANC is inherently sensitive to both magnitude and phase of anti-noises. Our MCANC makes use of the real and imaginary spectrograms of reference signals as inputs to estimate the real and imaginary spectrograms of canceling signals, as shown in Fig. 2(b). A reference signal $x_i(n)$ is sampled at 16 kHz, and windowed into 20-ms frames with 10-ms frame shift. Then a 320-point short time Fourier transform (STFT) is applied to each time frame to produce the real and imaginary spectrograms of $x_i(n)$, which are denoted as $X_i^{(r)}(t, f)$ and $X_i^{(i)}(t, f)$, respectively, within a T-F unit at time t and frequency f, superscripts (r) and (i) denote real and imaginary parts of signals.

To attenuate the primary noises at the desired locations, we set the ideal anti-noises, i.e., the primary noises, as the training targets. The deep MCANC model is trained to output the real and imaginary spectrograms of the J canceling signals, $Y_j^{(r)}(m, c)$ and $Y_j^{(i)}(m, c)$, $j = 1, 2, \ldots, J$, simultaneously. These outputs are then fed to the inverse Fourier transform to derive waveform signals $y_j(t)$. The anti-noises, which can be regarded as estimates of the training targets, are generated by passing the canceling signals through the corresponding loudspeakers and secondary paths.

2.4. Loss function and learning machine

The objective of deep MCANC is to generate canceling signals that minimize the error signals received at all error microphones. Therefore, the loss function is calculated as the sum of the K error signals:

$$Loss = \frac{\sum_{k=1}^{K} \sum_{n=1}^{L} e_k^2(n)}{KL} \tag{5}$$

where $e_k(n)$ is defined as (3), and L is the length of error signals.

We employ CRN for model training. The CRN has an encoder–decoder architecture, where the encoder and decoder comprise five convolutional layers and five deconvolutional layers, respectively, as shown in Fig. 2(b). A two-layer recurrent network with long short-term memory (LSTM) is inserted between them to account for temporal dynamics of audio signals, and a group strategy is used in LSTM (Gao et al., 2018) with the group number set to two. In our implementation, the encoder is a stack of convolutional layers and pooling layers, which serves to extract high-level features from the raw input. The decoder has the same structure as the encoder but in the reverse order and it ensures that the output of the decoder has the same shape as the input.

The detailed description of the CRN architecture is provided in Table 1. Layer kind and position are shown under layer name. The input and output size of each layer are marked as

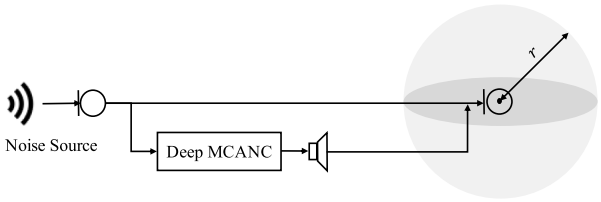

Fig. 3. Deep MCANC for noise attenuation within a sphere with a radius of r to generate a quiet zone.

FeatureMaps × *TimeSteps* × *FreqChannels*. The hyperparameters are specified in the (*kernelSize*, *strides*, *outChannels*) format. We apply zero-padding to the time direction for all the convolutions and the deconvolutions. The number of feature maps in each decoder layer is doubled due to skip connections. Batch normalization is adopted right after each convolution (or deconvolution) and before activation. Exponential linear units (ELUs) are used as the activation function in all convolutional and deconvolutional layers except the output layer. As shown in Fig. 2(b), we utilize skip connections, which feed the output of each encoder layer to the corresponding decoder layer, to improve the flow of information and gradients through the CRN. The number of input and output channels are set to $2I$ and $2J$, respectively.

2.5. Deep MCANC for quiet zone

Besides noise attenuation at multiple spatial points (locations of error microphones), deep MCANC can be trained to achieve ANC within a spatial zone using one or multiple canceling loudspeakers and multi error microphones. The general strategy is illustrated in Fig. 3. The goal is to attenuate noise in a target region. To achieve this, we train the deep MCANC controller on a variety of room impulse responses (RIRs) sampled within a spatial zone during training in an RIR-independent way. 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 simulated and only used during model training. Once the model is trained, these virtual error microphones are no longer used during the inference stage.

In order to achieve stronger noise attenuation within the quiet zone, more virtual error microphones are needed during training stage to sample as many positions within the zone as possible. One idea for efficient model training is to calculate the loss function each time from a randomly sampled subset of the K virtual error microphones (Zhang & Wang, 2021b). The model trained this way saves the amount of computation while still covering all K positions within the quiet zone.

3. Experimental setup

3.1. Experimental settings

To train a noise-independent model, we expose the MCANC model to a large variety of noisy environments in the training stage (Chen et al., 2016) and use 10 000 non-speech environmental sounds (noises) from a sound-effect library (http://www.sound-ideas.com) to create the training set. Babble noise, factory noise, operating room noise (denoted as "oproom"), and speech-shaped noise (denoted as "SSN") from the NOISEX-92 dataset (Varga & Steeneken, 1993), as well as recorded noises in the DEMAND dataset (Thiemann et al., 2013), are used for testing. All the noises are wideband without any low-pass filtering, i.e. they have significant energy across the entire frequency range. The test noises are unseen during training, and hence can evaluate the generalization ability of the proposed method.

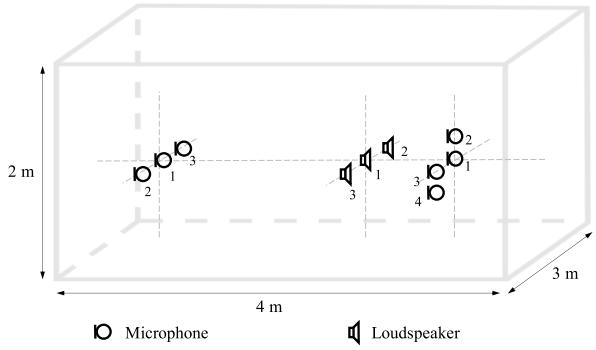

Fig. 4. Illustration of the MCANC experimental setup.

The physical structure of an ANC system is usually modeled as a rectangular enclosure and many studies have shown the effectiveness of ANC systems for noise canceling in enclosed rooms (Cheer, 2012; Parkins et al., 2000; Samarasinghe et al., 2016). In this study, we simulate a rectangular enclosure of size 3 m × 4 m × 2 m (width × length × height) and use the image method (Allen & Berkley, 1979) to generate RIRs for the primary and secondary paths of an MCANC system. The MCANC system with three reference microphones, three canceling loudspeakers, and four error microphones, is shown in Fig. 4 for the multi-point ANC scenario. Our evaluation considers a single noise source located at the position (1.5, 1, 1) m and is recorded by three reference microphones located at (1.5, 1, 1) m, (1.4, 1, 1) m, (1.6, 1, 1) m. In this scenario we can equate the noise and the first reference microphone recording. The three canceling loudspeakers are located at (1.5, 2.5, 1) m, (1.6, 2.5, 1) m, (1.4, 2.5, 1) m, and the four error microphones at (1.5, 3, 1) m, (1.5, 3, 1.1) m, (1.4, 3, 1) m, (1.4, 3, 0.9) m. In the following experiments, an $I \times J \times K$ ANC system refers to the MCANC setup that consists of the first I ($1 \leq I \leq 3$) reference microphones, the first J ($1 \leq J \leq 3$) loudspeakers and the first K ($1 \leq K \leq 4$) error microphones described here. For the quiet zone scenario, we use the same J ($1 \leq J \leq 3$) canceling 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, and 0.25 s, are used for generating training RIRs. With each T60, we generate multiple RIRs for all the primary and secondary paths in the MCANC system. For testing, we use RIRs with reverberation time 0.2 s as the default test RIRs, and the RIRs generated with untrained T60s are used to test the generalization ability of deep MCANC.

We create 20 000 training signals and 100 test signals. Each noise signal is created by randomly cutting a 3-s signal from the 10 000 noise signals. The primary noise at an error microphone is generated by convolving the source noise with the corresponding primary path (see Fig. 1). The estimated anti-noises are generated by passing the canceling signals through a loudspeaker function and then RIRs for the corresponding secondary paths. The CRN model is trained using the AMSGrad optimizer (Reddi et al., 2019) with a learning rate of 0.001 for 30 epochs.

3.2. Evaluation metric

We use normalized mean squared error (NMSE) (Das & Panda, 2004; Tan & Jiang, 2009) to evaluate noise attenuation performance of the proposed method. NMSE is a commonly used metric for ANC evaluations and it is defined as

$$\text{NMSE} = 10 \log_{10} \frac{\sum_{t=1}^{L} e^2(t)}{\sum_{t=1}^{L} d^2(t)} \tag{6}$$

Table 2
Average NMSE (dB) of ANC systems under different setups and with untrained noises.

		1 × 1 × 1	1 × 2 × 1	1 × 2 × 2	
Babble	FxLMS	−6.95 (0.3)	−9.12 (0.3)	−7.60 (0.06)	−7.94 (0.06)
	PMI-FxLMS	−7.02 (0.3)	−9.40 (0.3)	−7.85 (0.06)	−8.17 (0.06)
	Deep ANC	−12.08	−16.27	−12.93	−13.34
Factory	FxLMS	−6.62 (0.4)	−9.06 (0.3)	−7.51 (0.07)	−8.11 (0.07)
	PMI-FxLMS	−6.67 (0.4)	−9.29 (0.3)	−7.81 (0.07)	−8.38 (0.07)
	Deep ANC	−11.71	−14.73	−12.26	−12.39
Oproom	FxLMS	−6.93 (0.3)	−7.69 (0.2)	−7.15 (0.08)	−7.24 (0.08)
	PMI-FxLMS	−7.18 (0.3)	−8.38 (0.2)	−7.50 (0.08)	−7.60 (0.08)
	Deep ANC	−11.14	−14.72	−10.84	−11.95
SSN	FxLMS	−6.45 (0.2)	−9.51 (0.2)	−8.48 (0.06)	−9.48 (0.06)
	PMI-FxLMS	−6.49 (0.2)	−10.35 (0.2)	−9.00 (0.06)	−9.87 (0.06)
	Deep ANC	−12.53	−17.10	−12.26	−12.39

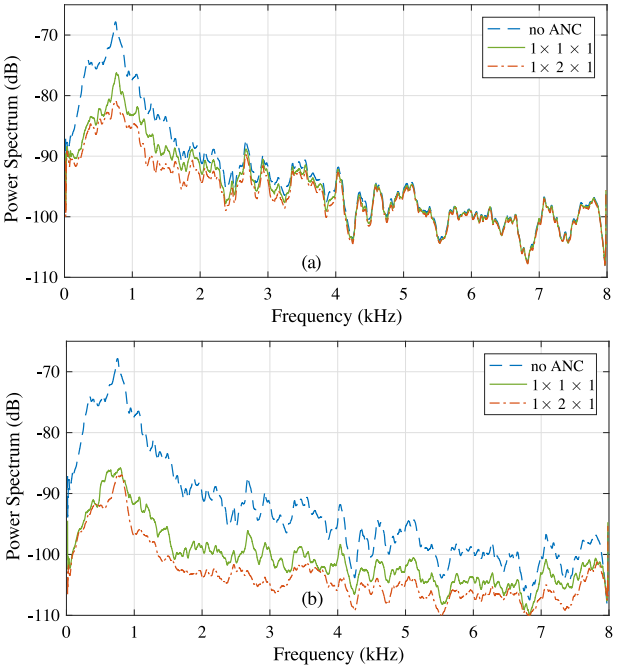

Fig. 5. Power spectrum of signals obtained with SSN noise under 1 × 1 × 1 and 1 × 2 × 1 setups using (a) PMI-FxLMS, and (b) proposed method.

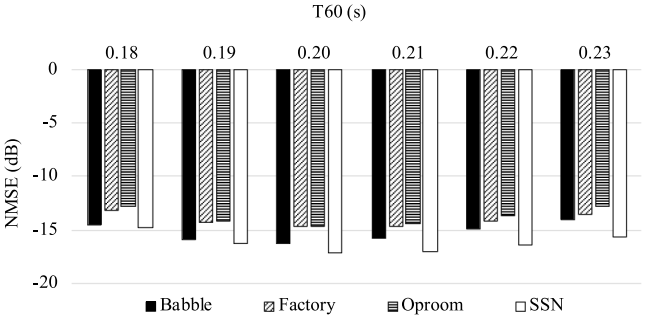

Fig. 6. Average NMSE for 1 × 2 × 1 deep MCANC using untrained RIRs with different T60s. The NMSE results for the trained T60 of 0.2 s are included for comparison purposes.

where L is the length of the signal. NMSE values are typically below zero, with a lower value indicating better noise attenuation.

4. Experimental results and comparisons

4.1. Deep MCANC for multi-point ANC

We first evaluate the performance of the proposed method under 1 × 1 × 1, 1 × 2 × 1, and 1 × 2 × 2 ANC setups, where the single-channel ANC (1 × 1 × 1) can be seen as a special case of MCANC. In the evaluations of this subsection and Sections 4.2–4.4, we use a linear loudspeaker function in (2), and nonlinear functions will be considered in Section 4.5. Two traditional ANC algorithms, FxLMS and post-masking-based FxLMS (PMI-FxLMS) are utilized for comparison. FxLMS is the most commonly used ANC algorithm and PMI-FxLMS is its modified version for faster convergence and better noise attenuation (Shi et al., 2019). The step sizes of FxLMS and PMI-FxLMS are chosen for different noises according to the criteria given in Shi et al. (2019) to ensure stable updating and good noise attenuation. The length of the error signal memory in PMI-FxLMS is set to 10, which is large enough to achieve good performance of ANC. The proposed and comparison methods are tested with four types of untrained noises and the average NMSE of test signals is given in Table 2. The NMSE values of the traditional ANC algorithms correspond to the final steady state results, and the associated step sizes are given inside the parentheses. For the MCANC setup with two error microphones (1 × 2 × 2), we give NMSE results at both error microphones in two separate columns in the table. It can be seen from this table that the proposed deep MCANC consistently outperforms the other methods under different setups and generalizes well to untrained noises.

We provide power spectrum curves in Fig. 5 for SNN for further comparisons. Power spectrum measures signal power as a function of frequency, and illustrates relative noise attenuation achieved at various frequencies. The curves presented in the figure show that deep MCANC outperforms the traditional method of PMI-FxLMS, and using more canceling loudspeakers improves the noise attenuation performance. Moreover, the proposed method is effective for ANC at low- and high-frequencies, while the comparison method is only effective at low frequencies. Traditional ANC is known to be restricted to low frequencies due to factors such as convergence and latency (Kuo & Morgan, 1999; Samarasinghe et al., 2016), and narrow-band or low-pass filtered noises are frequently utilized for performance evaluation. We use wideband noises in this study for evaluation, which partly explains why the amount of noise attenuation for the comparison methods is lower in Table 2 than typically reported in the literature.

Fig. 6 shows the average NMSE of deep MCANC when tested with RIRs generated using different T60 values (both trained and untrained). The NMSE results for the trained T60 of 0.2 s are included as reference. It is seen that the performance of deep MCANC generalizes well to untrained RIRs.

Table 3 provides the results of deep MCANC using different numbers of reference microphones. Using more reference microphones helps to capture the primary noise better and leads to

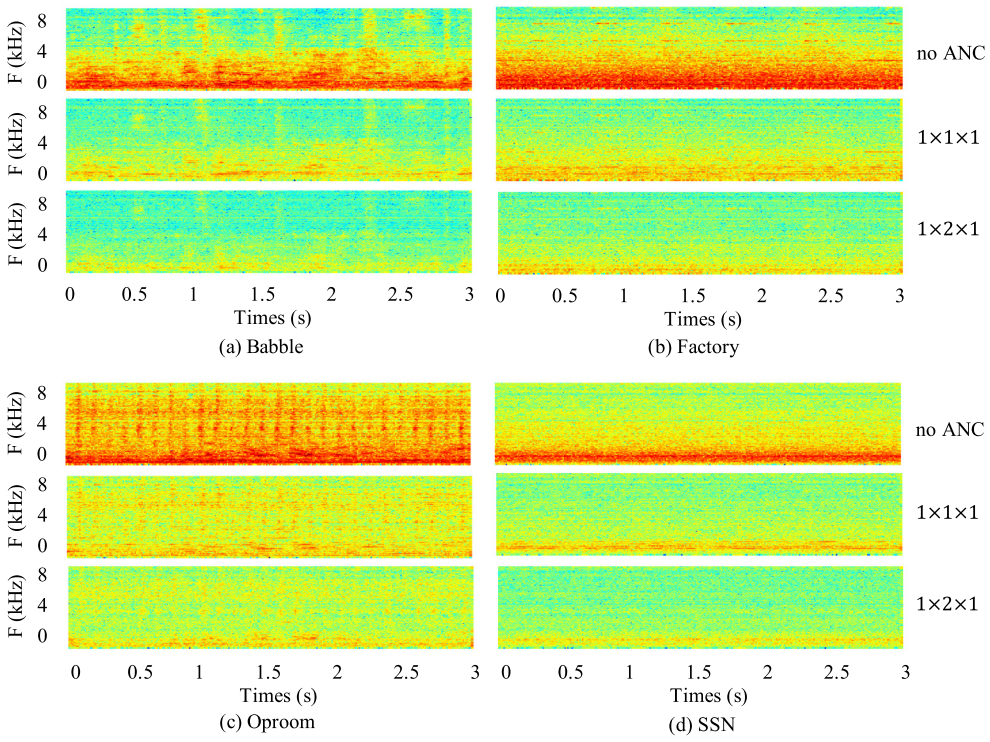

Fig. 7. Spectrograms of test results for different noises. The first, second and third row of each panel illustrates the output of no ANC, deep single-channel ANC, and deep MCANC ($1 \times 2 \times 1$), respectively.

Table 3
Average NMSE (dB) of deep MCANC using different numbers of reference microphones.

	Babble	Factory	Oproom	SSN
$1 \times 1 \times 1$	−12.08	−11.71	−11.14	−12.53
$2 \times 1 \times 1$	−12.57	−12.39	−11.51	−13.29
$3 \times 1 \times 1$	−12.93	−12.59	−11.77	−13.61

a little better noise attenuation performance. As our evaluations consider point-source noises, which are relatively easy to capture, MCANC with a single reference microphone will be taken as the default setting in the following experiments.

The spectrograms of the proposed method tested using different noises are shown in Fig. 7, where the first row of each panel shows the spectrogram of primary noise (no ANC), and the second and third rows show the residual noises (error signals) obtained using $1 \times 1 \times 1$ and $1 \times 2 \times 1$ setups, respectively. It can be seen that the proposed method is capable of achieving wideband noise attenuation. Using two canceling loudspeakers helps to improve the noise attenuation performance over the single loudspeaker setup.

To further explore the impact of the number of canceling loudspeakers and error microphones on multi-point ANC performance, we evaluate the performance of $1 \times J \times K$ MCANC setups with different combinations of J and K values, where J varies from 1 to 3 and K from 1 to 4. The results tested using different noises are shown in Table 4. For an MCANC system with K error microphones, we measure the average NMSE obtained at each microphone and provide all K values in different rows of the table. These results show that, with a fixed number of error microphones, noise attenuation performance usually improves as the number of canceling loudspeakers increases.

4.2. Deep MCANC for quiet zone

This subsection evaluates the performance of deep MCANC for generating a quiet zone. Conventional ANC methods require multiple canceling loudspeakers and error microphones to generate a quiet zone, and the error microphones need to be placed in the target zone, or near the zone with the help of remote sensing techniques (Kajikawa et al., 2012). The proposed method can achieve a quiet zone using a single canceling loudspeaker and does not need to have any physical error microphones during the inference stage (see also Zhang & Wang, 2021a). In this study, we simulate the quiet zone as a sphere and set its radius to $r = 5$ cm; such a zone size is appropriate, for example, for a driver's ears 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 radii ($r = 1, 2, 3, 4, 5$ cm). Specifically, for "$r = d$", we place an error microphone at 10 random positions on the sphere of radius d cm and use the corresponding RIRs to create test signals. Unless otherwise stated, the default number of virtual error microphones is set to $K = 100$.

We start with assessing the impact of the number of virtual error microphones on the quiet zone performance. A $1 \times 1 \times K$ MCANC system with a single canceling loudspeaker and K virtual error microphones is used for generating the quiet zone, where K varies from 1 to 200. The results are provided in Fig. 8 with the average value for each K shown below corresponding bars. Note that the case with $K = 1$, which is the single-channel ANC case, is trained by putting one error microphone at the center of the sphere. It achieves better NMSE than other cases at $r = 0$ cm. But its performance drops significantly as test positions move away from the center of the zone. Using more virtual error microphones achieves better overall noise attenuation within the zone, while the improvement plateaus when K reaches 100.

Table 4
Average NMSE (dB) of deep MCANC for multi-point noise cancellation using different numbers of canceling loudspeakers and error microphones.

Babble			Factory			Oproom			SSN		
1 × 1 × 1	1 × 2 × 1	1 × 3 × 1	1 × 1 × 1	1 × 2 × 1	1 × 3 × 1	1 × 1 × 1	1 × 2 × 1	1 × 3 × 1	1 × 1 × 1	1 × 2 × 1	1 × 3 × 1
−12.08	−16.27	−15.52	−11.71	−14.73	−14.20	−11.14	−14.72	−13.94	−12.53	−17.10	−16.64
1 × 1 × 2	1 × 2 × 2	1 × 3 × 2	1 × 1 × 2	1 × 2 × 2	1 × 3 × 2	1 × 1 × 2	1 × 2 × 2	1 × 3 × 2	1 × 1 × 2	1 × 2 × 2	1 × 3 × 2
−8.41	−12.93	−14.55	−7.94	−12.26	−13.53	−7.47	−10.84	−12.49	−8.19	−12.58	−13.11
−9.27	−13.34	−13.78	−8.70	−12.39	−12.55	−8.27	−11.95	−12.48	−9.66	−13.56	−14.45
1 × 1 × 3	1 × 2 × 3	1 × 3 × 3	1 × 1 × 3	1 × 2 × 3	1 × 3 × 3	1 × 1 × 3	1 × 2 × 3	1 × 3 × 3	1 × 1 × 3	1 × 2 × 3	1 × 3 × 3
−8.17	−8.36	−12.62	−7.65	−8.21	−12.00	−6.99	−7.54	−10.43	−8.08	−8.42	−12.83
−8.73	−9.45	−11.69	−8.22	−8.99	−10.93	−7.88	−8.38	−10.88	−8.47	−10.24	−11.63
−8.16	−9.72	−13.34	−7.57	−9.13	−12.11	−7.31	−8.92	−11.89	−8.74	−9.90	−14.09
1 × 1 × 4	1 × 2 × 4	1 × 3 × 4	1 × 1 × 4	1 × 2 × 4	1 × 3 × 4	1 × 1 × 4	1 × 2 × 4	1 × 3 × 4	1 × 1 × 4	1 × 2 × 4	1 × 3 × 4
−7.13	−9.28	−11.50	−6.59	−9.05	−10.70	−6.08	−8.06	−9.09	−6.89	−9.13	−11.65
−9.26	−9.89	−11.41	−8.69	−9.38	−10.63	−8.19	−8.87	−9.76	−9.36	−9.93	−11.92
−8.18	−9.80	−11.97	−7.56	−8.91	−10.89	−6.96	−8.56	−9.93	−8.62	−9.70	−11.87
−8.50	−9.49	−11.49	−7.42	−8.59	−10.07	−7.44	−8.22	−9.83	−7.23	−8.34	−10.69

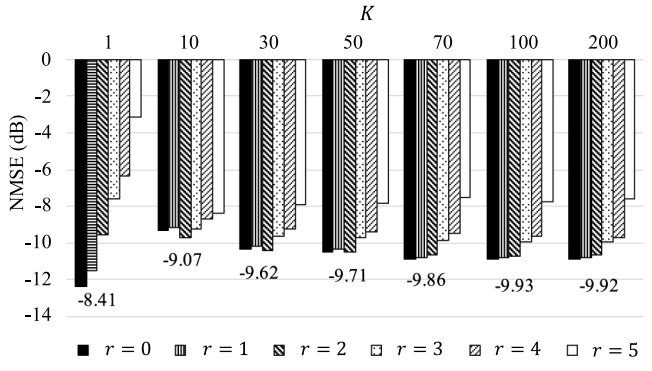

Fig. 8. Average NMSE of deep MCANC for generating quiet zone in babble noise using a $1 \times 1 \times K$ setup with different numbers of virtual error microphones (K), where r (cm) is the distance of a test position to the center of the zone. The value shown below each cluster is the average NMSE within the quiet zone.

Table 5
Average NMSE (dB) of deep MCANC for generating quiet zone using different numbers of canceling loudspeakers.

		$r = 0$	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	Ave
Babble	J = 1	−10.83	−10.72	−10.52	−9.77	−9.48	−7.55	−9.81
	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
Oproom	J = 1	−9.20	−8.92	−8.49	−8.38	−7.92	−7.04	−8.33
	J = 2	−11.78	−11.04	−10.50	−10.45	−9.62	−8.29	−10.28
	J = 3	−12.29	−11.50	−11.29	−10.66	−10.65	−9.40	−10.96
SSN	J = 1	−10.04	−9.99	−9.67	−8.88	−8.37	−6.43	−8.90
	J = 2	−14.63	−14.55	−13.51	−11.55	−11.27	−8.58	−12.35
	J = 3	−16.45	−16.19	−16.40	−15.10	−14.37	−11.90	−15.07

Table 5 shows the MCANC performance for generating a quiet zone using different numbers of canceling loudspeakers. Similar to the trend observed in the multi-point ANC scenario, more canceling loudspeakers improve the overall performance of quiet zone generation. This is because, as the number of loudspeakers increases, more canceling signals are generated for canceling or attenuating the primary noise within the fixed zone. For example, the average NMSE with babble noise is improved to −12.31 dB from −9.81 dB by using 2 loudspeakers instead of 1, and the NMSE is further improved by 2.07 dB when 3 loudspeakers are used. With more than 3 loudspeakers, we expect NMSE performance to further improve, although the amount of improvement for each added loudspeaker will diminish.

Table 6
Average NMSE (dB) of deep MCANC for generating quiet zone with different canceling loudspeaker positions.

		Near	Medium	Far
1 × 1 × 100	$r = 0$	−15.78	−11.33	−12.56
	$r = 1$	−15.16	−11.56	−12.52
	$r = 2$	−12.83	−10.06	−11.98
	$r = 3$	−11.38	−9.21	−11.28
	$r = 4$	−10.04	−8.77	−10.88
	$r = 5$	−8.61	−8.00	−10.01
	Average	−12.30	−9.75	−11.54
	Variance	6.40	1.86	0.95

4.3. Position of canceling loudspeaker

The position of a canceling loudspeaker (the secondary source) not only influences the system causality but also determines the amount of noise reduction. To explore the influence of loudspeaker position on ANC performance, we consider three 1 × 1 × 100 MCANC setups with different loudspeaker positions for generating a quiet zone. The distance between the canceling loudspeaker and the center of the quiet zone is set to 0.1 m, 0.5 m, and 1.9 m, respectively. The corresponding positions of the canceling loudspeaker are (2.9, 1.5, 1) m, (2.5, 1.5, 1) m, and (1.1, 1.5, 1) m, which are denoted as "Near", "Medium" and "Far" cases, respectively. To save training time, a smaller dataset with 5000 babble noises is used for model training, and the performance is tested using 100 untrained babble noises. Table 6 shows the NMSE results. It can be seen that the "Near" case achieves better noise cancellation at locations that are close to the center of the quiet zone (locations with smaller r) than the other two cases. Moving the canceling loudspeaker away from the quiet zone (the "Medium" case) results in weaker noise attenuation. Further increasing the distance between canceling loudspeaker and the quiet zone, the "Far" case slightly improves the performance compared to the results obtained in the "Medium" case.

The above observations can be understood from two perspectives. The performance of deep MCANC for generating a quiet zone relies on the magnitude and phase of anti-noises received within the spatial zone. From the reverberation point of view, placing the secondary source nearer to the quiet zone results in higher direct-to-reverberant-ratio (DRR) in anti-noise. Since the noise attenuation level is highly related to the direct part of anti-noise, higher DRR would lead to better noise attenuation. From

Table 7
Robustness of deep MCANC to variations in reference signals.

		1 × 1 × 1	1 × 2 × 1	1 × 1 × 100	1 × 2 × 100
Noisechange	Babble → Factory	−11.83	−15.48	−9.47	−11.87
	Babble → Oproom	−11.58	−15.41	−9.19	−10.90
	Babble → SSN	−12.33	−16.82	−9.39	−12.36
Noisemixture	Babble + Factory	−11.33	−14.85	−9.12	−11.43
	Babble + Oproom	−10.99	−14.60	−8.95	−10.64
	Babble + SSN	−12.41	−17.01	−9.40	−12.42

Table 8
Average NMSE (dB) of deep learning based MCANC in the present of nonlinear distortions.

η^2	1 × 2 × 1			1 × 1 × 100		
	∞	0.1	0.5	∞	0.1	0.5
Babble	−16.43	−16.39	−16.46	−9.68	−9.57	−9.64
Factory	−15.54	−15.46	−15.57	−8.90	−8.66	−8.87
Oproom	−15.77	−15.73	−15.79	−8.46	−8.41	−8.47
SSN	−17.74	−17.79	−17.82	−9.04	−8.97	−9.03

Table 9
Average NMSE (dB) of deep MCANC tested on recorded noises in six realistic environments.

	1 × 1 × 1	1 × 2 × 1	1 × 1 × 100	1 × 2 × 100
NRIVER	−10.92	−13.05	−7.22	−9.10
OMEETING	−12.08	−14.00	−9.37	−10.93
DLIVING	−11.68	−11.93	−9.34	−9.85
PRESTO	−12.76	−15.10	−8.52	−10.15
SPSQUARE	−10.81	−11.08	−8.44	−9.35
TMETRO	−11.30	−11.91	−9.10	−9.91

the wave propagation point of view, given a zone of fixed size, the waves inside a zone that is farther from the sound source have smaller intensity variations within the zone due to the inverse square law of wave propagation (Hartmann, 1998). This makes noise attenuation within this zone more "even". To show this, we calculate the variance of NMSE obtained within the zone and present it in the last row of Table 6. It is seen that the variance of NMSE obtained in the "Far" case is much smaller than that obtained in the "Near" case.

4.4. Robustness test

In ANC applications, variations may occur in reference signals due to the change of acoustic environment. Here, we evaluate deep MCANC models in situations with noise type change and multiple noises occurring simultaneously in reference signals to show the robustness of the proposed method against these variations. For noise type change, we generate test signals by changing the noise type in the reference signal from babble noise to a different noise after 1.5 s, the middle point of a reference signal. For situations with multiple noises, the reference signal is generated as a mixture of babble noise and another noise. The test results are given in Table 7, and they demonstrate the strong robustness of deep MCANC against such variations in reference signals.

4.5. Nonlinear MCANC

The performance of deep MCANC in the presence of nonlinear distortions is studied in this part. We follow the setup given in Agerkvist (2007), Zhang and Wang (2020) and Ghasemi et al. (2016) and simulate the saturation nonlinearity of loudspeaker using the scaled error function (SEF) (Klippel, 2006). That is

$$f_{\text{LS}}(y) = f_{\text{SEF}}(y) = \int_0^y e^{-\frac{z^2}{2\eta^2}} dz \tag{7}$$

where y is the input to the loudspeaker, and 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.

The deep MCANC models are trained using four loudspeaker functions: $\eta^2 = 0.1$ (severe nonlinearity), $\eta^2 = 1$ (moderate nonlinearity), $\eta^2 = 10$ (soft nonlinearity), and $\eta^2 = \infty$ (linear). During training, we randomly select a loudspeaker function for each input signal, and generate the loudspeaker signal by passing a canceling signal through the loudspeaker function. For testing, both trained and untrained ($\eta^2 = 0.5$) loudspeaker functions are used. The results are given in Table 8. These results show that deep MCANC models can be trained to achieve noise attenuation in both linear and nonlinear situations.

4.6. Noise attenuation performance under real recordings

Finally, nonstationary noises from the DEMAND corpus (Thiemann et al., 2013) are used to test the performance of deep ANC in realistic conditions. All recordings of the DEMAND corpus are five-minute long and recorded with a 16-channel array. We randomly select one channel of the recordings for testing.

The DEMAND dataset has six categories of noises. We choose one noise from each category to represent distinct environments. These six nonstationary noises is described as follows.

- The "Nature" category: The NRIVER noise, recorded besides a creek of running water.
- The "Office" category: The OMEETING noise, recorded in a meeting room.
- The "Domestic" category: The DLIVING noise, recorded inside a living room.
- The "Public" category: The PRESTO noise, recorded in a university restaurant at lunchtime.
- The "Street" category: The SPSQUARE noise, recorded in a public town square with many tourists.
- The "Transportation" category: The TMETRO noise, recorded in a subway.

The NMSE results are given in Table 9. These results show that the proposed deep MCANC works well for recorded noises under different realistic environments.

5. Discussion

This paper demonstrates the utility of deep learning for solving MCANC problems. For quiet zone generation, we investigate the feasibility of using fewer virtual error microphones for efficient model training. In comparison to deep MCANC trained with a large number of virtual error microphones, training a model using a subset of the virtual microphones, say 20%, achieves comparable quiet zone performance while requiring less computation.

Compared with traditional methods, deep MCANC has the following advantages. First, deep MCANC is capable of canceling different types of noise and effective for wideband noise attenuation. Second, rather than training a separate ANC for each

channel, the proposed method approaches MCANC by training a single controller to generate multi-channel outputs. Third, deep MCANC can be trained to generate a quiet zone using a single canceling loudspeaker and does not need physical error microphones during the inference stage. Fourth, the proposed method can inherently deal with nonlinear distortions. Moreover, as we have shown previously, deep ANC is flexible in terms of training target, e.g., it can be trained to selectively attenuate the noise components of a noisy speech signal and let the underlying speech pass through Zhang and Wang (2021a). This advantage holds for deep MCANC.

Two main issues are worth discussing, which are generalization ability and processing latency. Limited generalization ability is a common problem for fixed-parameter ANC methods. To address this, we employ large-scale multi-condition training to make deep MCANC robust against a variety of noises. Other methods like selective fixed-parameter ANC (Ranjan et al., 2016; Shi et al., 2020, 2018, 2022) have also been proposed to improve generalization ability. These methods work by pre-training multiple fixed-coefficient sets for different environments and selecting a proper one for noise attenuation during the inference stage. In addition, using a hybrid approach that combines fixed-parameter and adaptive ANC could be explored for further improving the generalization ability of deep MCANC.

The processing latency is an important issue due to the causality constraint of ANC systems, whereby the signal of the canceling loudspeaker has to be generated before the actual noise propagates to the canceling loudspeaker. Deep MCANC is a frequency domain method, which incurs an algorithmic delay related to the frame length and frame shift of STFT. This kind of delay is considered as a shortcoming for frequency-domain ANC algorithms and many approaches have been proposed to overcome this (Bendel et al., 2001; Kim et al., 1994; Kuo et al., 2008; Yang et al., 2018). We have proposed a delay-compensated training strategy in Zhang and Wang (2021a) to tackle this latency problem and it can be utilized in deep MCANC as well. In addition, implementing deep MCANC using low-latency time-domain methods (Luo & Mesgarani, 2019; Pandey & Wang, 2019, 2021) could alleviate the latency problem.

In terms of computational complexity, the proposed deep MCANC system contains about 8.8 million (M) trainable parameters and takes 12.72 M floating-point fused multiply-adds per time frame. As expected, the DNN model requires more computations than traditional adaptive algorithms. Model compression and efficient implementation will be addressed in future studies to enable deep MCANC in real-world applications.

The MCANC formulation in this paper is focused on the situation with a single noise source. However, our formulation can be straightforwardly extended to the situation with multiple noise sources at different locations. In this case, the number of primary paths will be increased to $M \times K$, where M is the number of noise sources. The corresponding error signal sensed by the kth error microphone will be expanded to (cf. Eq. (3))

$$e_k(n) = \sum_{m=1}^{M} d_{mk}(n) - \sum_{j=1}^{J} a_{jk}(n) \tag{8}$$

and Eq. (4) will be modified to

$$d_{mk}(n) = p_{mk}(n) * s_m(n) \tag{9}$$

where $s_m(n)$ is the mth noise source, and $d_{mk}(n)$ is the corresponding primary noise at error microphone k. The overall procedure of using deep MCANC for noise attenuation with single and multiple noise sources would otherwise be the same. It is worth noting that attenuating multiple noise sources jointly usually places a separate reference microphone near each source; for example, for road noise cancellation in the in-car environment, a sensor is installed at each wheel to capture the road generated noise (Sano et al., 2001).

6. Conclusion

In this paper, we have introduced a deep MCANC approach to multi-channel active noise control. The proposed approach trains a CRN model to estimate multiple canceling signals simultaneously from reference signals so that the corresponding anti-noises cancel or attenuate the primary noises. We have evaluated the performance of deep MCANC for multi-point ANC and quiet zone generation. The impact of factors such as the number of canceling loudspeakers and error microphones, and the position of a canceling loudspeaker on MCANC performance has been investigated. Extensive experimental results show the effectiveness of deep MCANC for noise attenuation in various scenarios. In addition, the proposed method is robust against untrained noises and works well in the presence of loudspeaker nonlinearity. Future work includes exploring time-domain methods and considering practical issues such as computational complexity.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported in part by an National Science Foundation, USA grant (ECCS-1808932) and the Ohio Supercomputer Center, USA.

References

Agerkvist, F. (2007). Modelling loudspeaker non-linearities. In *AES conference:32nd international conference: DSP for loudspeakers*. Audio Engineering Society.

Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, *65*, 943–950.

Bai, Z., & Zhang, X.-L. (2021). Speaker recognition based on deep learning: An overview. *Neural Networks*, *140*, 65–99.

Bendel, Y., Burshtein, D., Shalvi, O., & Weinstein, E. (2001). Delayless frequency domain acoustic echo cancellation. *IEEE Transactions on Speech and Audio Processing*, *9*, 589–597.

Bouchard, M. (2003). Multichannel affine and fast affine projection algorithms for active noise control and acoustic equalization systems. *IEEE Transactions on Speech and Audio Processing*, *11*, 54–60.

Cheer, J. (2012). *Active control of the acoustic environment in an automobile cabin* (Ph.D. thesis), University of Southampton.

Cheer, J., & Elliott, S. J. (2015). Multichannel control systems for the attenuation of interior road noise in vehicles. *Mechanical Systems and Signal Processing*, *60*, 753–769.

Chen, D., Cheng, L., Yao, D., Li, J., & Yan, Y. (2021). A secondary path-decoupled active noise control algorithm based on deep learning. *IEEE Signal Processing Letters*.

Chen, J., Wang, Y., Yoho, S. E., Wang, D. L., & Healy, E. W. (2016). Large-scale training to increase speech intelligibility for hearing-impaired listeners in novel noises. *Journal of the Acoustical Society of America*, *139*, 2604–2612.

Chen, H., & Zhang, P. (2021). A dual-stream deep attractor network with multi-domain learning for speech dereverberation and separation. *Neural Networks*, *141*, 238–248.

Costa, M. H., Bermudez, J. C. M., & Bershad, N. J. (2002). Stochastic analysis of the filtered-x LMS algorithm in systems with nonlinear secondary paths. *IEEE Transactions on Signal Processing*, *50*, 1327–1342.

Das, D. P., & Panda, G. (2004). Active mitigation of nonlinear noise processes using a novel filtered-s LMS algorithm. *IEEE Transactions on Speech and Audio Processing*, *12*, 313–322.

Du, H., Xie, L., & Li, H. (2022). Noise-robust voice conversion with domain adversarial training. *Neural Networks*, *148*, 74–84.

Elliott, S., Stothers, I., & Nelson, P. (1987). A multiple error LMS algorithm and its application to the active control of sound and vibration. *IEEE Transactions on Audio, Speech, and Language Processing*, *35*, 1423–1434.

Gao, F., Wu, L., Zhao, L., Qin, T., Cheng, X., & Liu, T. (2018). Efficient sequence learning with group recurrent networks. In *Proc. HLT-NAACL, Vol. 1* (pp. 799–808).

Ghasemi, S., Kamil, R., & Marhaban, M. H. (2016). Nonlinear THF-FxLMS algorithm for active noise control with loudspeaker nonlinearity. *Asian Journal of Control*, *18*, 502–513.

Hartmann, W. M. (1998). *Signals, sound, and sensation*. New York: Springer.

Kajikawa, Y., Gan, W. S., & Kuo, S. M. (2012). Recent advances on active noise control: open issues and innovative applications. *APSIPA Transactions on Signal and Information Processing*, *1*.

Kim, I.-S., Na, H.-S., Kim, K.-J., & Park, Y. (1994). Constraint filtered-x and filtered-u least-mean-square algorithms for the active control of noise in ducts. *Journal of the Acoustical Society of America*, *95*, 3379–3389.

Klippel, W. (2006). Tutorial: Loudspeaker nonlinearities—Causes, parameters, symptoms. *AES: Journal of the Audio Engineering Society*, *54*, 907–939.

Krukowicz, T. (2013). Neural fixed-parameter active noise controller for variable frequency tonal noise. *Neurocomputing*, *121*, 387–391.

Kukde, R., Manikandan, M. S., & Panda, G. (2020). Incremental learning based adaptive filter for nonlinear distributed active noise control system. *IEEE Open Journal of Signal Processing*, *1*, 1–13.

Kuo, S. M., Mitra, S., & Gan, W. S. (2006). Active noise control system for headphone applications. *IEEE Transactions on Control Systems Technology*, *14*, 331–335.

Kuo, S. M., & Morgan, D. R. (1996). *Active noise control systems*. New York: Wiley.

Kuo, S. M., & Morgan, D. R. (1999). Active noise control: a tutorial review. *Proceedings of the IEEE*, *87*, 943–973.

Kuo, S. M., Wu, H. T., Chen, F. K., & Gunnala, M. R. (2004). Saturation effects in active noise control systems. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *51*, 1163–1171.

Kuo, S. M., Yenduri, R. K., & Gupta, A. (2008). Frequency-domain delayless active sound quality control algorithm. *Journal of Sound and Vibration*, *318*, 715–724.

Lam, B., Gan, W. S., Shi, D., Nishimura, M., & Elliott, S. (2021). Ten questions concerning active noise control in the built environment. *Building and Environment*, *200*, Article 107928.

Liu, L., Gujjula, S., & Kuo, S. M. (2009). Multi-channel real time active noise control system for infant incubators. In *Proc. EMBS* (pp. 935–938). IEEE.

Lorente, J., Ferrer, M., de Diego, M., & Gonzalez, A. (2015). The frequency partitioned block modified filtered-x NLMS with orthogonal correction factors for multichannel active noise control. *Digital Signal Processing*, *43*, 47–58.

Luo, Y., & Mesgarani, N. (2019). Conv-tasnet: Surpassing ideal time–frequency magnitude masking for speech separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *27*, 1256–1266.

Murao, T., Shi, C., Gan, W. S., & Nishimura, M. (2017). Mixed-error approach for multi-channel active noise control of open windows. *Applied Acoustics*, *127*, 305–315.

Pandey, A., & Wang, D. L. (2019). A new framework for CNN-based speech enhancement in the time domain. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *27*, 1179–1188.

Pandey, A., & Wang, D. L. (2021). Dense CNN with self-attention for time-domain speech enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 1270–1279.

Parkins, J. W., Sommerfeldt, S. D., & Tichy, J. (2000). Narrowband and broadband active control in an enclosure using the acoustic energy density. *Journal of the Acoustical Society of America*, *108*, 192–203.

Patel, V., & George, N. V. (2020). Multi-channel spline adaptive filters for non-linear active noise control. *Applied Acoustics*, *161*, Article 107142.

Pawełczyk, M. (2008). Active noise control-a review of control-related problems. *Archives of Acoustics*, *33*, 509–520.

Ranjan, R., Murao, T., Lam, B., & Gan, W. S. (2016). Selective active noise control system for open windows using sound classification. In *Proc. INTER-NOISE* (pp. 1921–1931).

Reddi, S. J., Kale, S., & Kumar, S. (2019). On the convergence of adam and beyond. arXiv preprint arXiv:1904.09237.

Samarasinghe, P. N., Zhang, W., & Abhayapala, T. D. (2016). Recent advances in active noise control inside automobile cabins: Toward quieter cars. *IEEE Signal Processing Magazine*, *33*, 61–73.

Sano, H., Inoue, T., Takahashi, A., Terai, K., & Nakamura, Y. (2001). Active control system for low-frequency road noise combined with an audio system. *IEEE Transactions on Speech and Audio Processing*, *9*(7), 755–763.

Shi, D., Gan, W. S., Lam, B., & Wen, S. (2020). Feedforward selective fixed-filter active noise control: Algorithm and implementation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *28*, 1479–1492.

Shi, C., Jiang, N., Xie, R., & Li, H. (2019). A simulation investigation of modified FxLMS algorithms for feedforward active noise control. In *Proc. APSIPA ASC* (pp. 1833–1837). IEEE.

Shi, D. Y., Lam, B., & Gan, W. S. (2018). A novel selective active noise control algorithm to overcome practical implementation issue. In *Proc. IEEE int. conf. acoust., speech signal process.* (pp. 1130–1134).

Shi, D., Lam, B., Gan, W. S., & Wen, S. (2021). Block coordinate descent based algorithm for computational complexity reduction in multichannel active noise control system. *Mechanical Systems and Signal Processing*, *151*, Article 107346.

Shi, D., Lam, B., Ooi, K., Shen, X., & Gan, W. S. (2022). Selective fixed-filter active noise control based on convolutional neural network. *Signal Processing*, *190*, Article 108317.

Shi, D., Shi, C., & Gan, W. S. (2017). Effect of the audio amplifier's distortion on feedforward active noise control. In *Proc. APSIPA ASC* (pp. 469–473). IEEE.

Tan, L., & Jiang, J. (2009). Adaptive second-order volterra filtered-X RLS algorithms with sequential and partial updates for nonlinear active noise control. In *Proc. ICIEA* (pp. 1625–1630). IEEE.

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

Tanaka, K., Shi, C., & Kajikawa, Y. (2014). Multi-channel active noise control using parametric array loudspeakers. In *Proc. APSIPA* (pp. 1–6). IEEE.

Thiemann, J., Ito, N., & Vincent, E. (2013). The diverse environments multi-channel acoustic noise database (demand): A database of multichannel environmental noise recordings. In *Proc. Meet. Acoust. ICA2013, Vol. 19*. Acoustical Society of America, Article 035081.

Varga, A., & Steeneken, H. J. (1993). 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*, *12*, 247–251.

Wilby, J. F. (1996). Aircraft interior noise. *Journal of Sound and Vibration*, *190*, 545–564.

Williamson, D. S., Wang, Y., & Wang, D. L. (2016). Complex ratio masking for monaural speech separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *24*, 483–492.

Xian, Y., Sun, Y., Wang, W., & Naqvi, S. M. (2021). Convolutional fusion network for monaural speech enhancement. *Neural Networks*, *143*, 97–107.

Yang, F., Cao, Y., Wu, M., Albu, F., & Yang, J. (2018). Frequency-domain filtered-x LMS algorithms for active noise control: A review and new insights. *Applied Sciences*, *8*, 2313.

Zhang, H., & Wang, D. L. (2020). A deep learning approach to active noise control. In *Proc. Interspeech* (pp. 1141–1145).

Zhang, H., & Wang, D. L. (2021a). Deep ANC: A deep learning approach to active noise control. *Neural Networks*, *141*, 1–10.

Zhang, H., & Wang, D. L. (2021b). A deep learning method to multi-channel active noise control. In *Proc. Interspeech* (pp. 681–685).