Audio Engineering Society

Conference Paper 5

Presented at the AES International Conference on Machine Learning and Artificial Intelligence for Audio 2025 September 8–10, London, UK

This paper was peer-reviewed as a complete manuscript for presentation at this conference. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Hybrid Learning-based Active Noise Control in Encapsulated Structures

Alkahf Aboutiman[1], Hamid Reza Karimi[1], and Francesco Ripamonti[1]

[1]*Politecnico di Milano, Department of Mechanical Engineering, Milan, Italy*

Correspondence should be addressed to Alkahf Aboutiman (alkahfahmat.aboutiman@polimi.it)

ABSTRACT

Active noise control (ANC) in enclosed environments such as vehicle cabins presents substantial challenges due to complex acoustic paths. To address this, a hybrid ANC strategy is proposed, combining adaptive filtering with deep learning techniques. A finite element model (FEM) was developed to simulate the vibroacoustic dynamics of a closed test structure referred to as the "Noise Box" which reproduces the acoustic conditions inside a car. This model highlights the presence of non-minimum phase characteristics, which complicate control by introducing delays, undershoots, or initial noise amplification before attenuation occurs. To improve control performance under these conditions, a two-dimensional convolutional neural network (2D CNN) is used to select the most appropriate pre-trained control filter based on the spectral features of the incoming noise. Experimental results show that this learning-based hybrid ANC approach achieves superior noise reduction compared to conventional adaptive algorithms.

1 Introduction

Encapsulated environments such as trains, cars, and airplanes can significantly benefit from active noise control (ANC) systems, which provide an advanced solution for reducing unwanted sound [1, 2]. These systems operate by detecting noise through strategically placed sensors and generating anti-noise signals via speakers, resulting in effective noise cancellation in targeted areas. This technology enhances passenger comfort by mitigating the disturbances caused by external and internal sound sources. However, implementing ANC in such environments presents several challenges, particularly in maintaining high performance across a broad frequency spectrum. Addressing these challenges requires sophisticated techniques, including precise acoustic modeling, advanced path identification, and adaptive control algorithms [3, 4].

A fundamental obstacle to effective ANC is the intrinsic behavior of the control environment. In systems exhibiting non-minimum phase characteristics, the initial control signal may unintentionally induce undershoot, phase shift or delay in the control. This phenomenon complicates the convergence of adaptive algorithms and imposes fundamental limitations on overall system performance [5, 6]. Additionally, complex acoustic paths and resonances within the structure can introduce further constraints on the effectiveness of traditional control techniques, necessitating the development of more robust solutions [7].

This study introduces an ANC algorithm specifically

designed for an encapsulated structure referred to as the "noise box," which replicates the acoustic properties of a car interior. The proposed algorithm, SFANC-NLMS, builds upon previous research to enhance noise reduction efficiency in challenging acoustic environments. Inspired by [8], it incorporates a lightweight two-dimensional convolutional neural network (2D CNN) running on a co-processor to select the most appropriate pre-trained control filter for different noise types. Concurrently, the NLMS algorithm continuously updates the filter coefficients at the sampling rate for each noise frame, ensuring real-time adaptability. This hybrid approach enables the system to maintain effective noise suppression despite the complexity of the control environment.

2 Noise Box FEM model

2.1 Model and geometry

In the finite element model, a cavity was selected to simulate the interior of a car, referred to as the "Noise Box." To replicate the sound field within this enclosed structure, modeling both the internal and external fields was avoided due to computational cost constraints. Instead, only the air cavity and the vibrating part of the plate are modeled. The cavity takes the form of a pentagonal prism, with an additional compartment where the plate is located. Its dimensions are $0.825\,m \times 1.018\,m \times 0.645\,m$. The cavity is closed off with a rectangular aluminum plate, measuring $0.825\,m \times 1.018\,m$, and the physical properties of the plate are provided in Table 1.

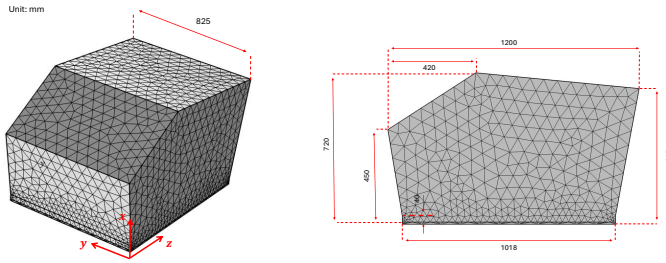
Table 1: Physical properties of the plate material

Property	Value
Thickness, t (m)	0.0004
Density, ρ (kg/m^3)	2700
Modulus of elasticity, E (Pa)	70e9
Poisson's ratio, ν	0.33

Figure 1 shows the dimensions of the noise box used in the model. COMSOL Multiphysics® was used for finite element modeling [9].

2.2 Physics definition

The aluminum plate was modeled in 3D using the Shell physics module, with quadratic shell elements by default. These elements follow the Mindlin-Reissner



(a) (b)

Fig. 1: Geometry of the Noise Box FEM models in (a) 3D and (b) 2D.

theory, and the interpolation method chosen is MITC (Mixed Interpolation of Tensorial Components), which allows for modeling both thin (Kirchhoff theory) and thick (Mindlin theory) shells. Given the thinness of the plate, with a thickness-to-span ratio of $h/a = 0.0048$, much smaller than 1/20 and less than 1/200, Kirchhoff's theory is most appropriate for this study, as suggested in the literature [10, 11].

The shell elements possess three degrees of freedom (DOF) in the x, y, and z directions. For damping, an isotropic structural loss factor $\eta = 0.003$ was selected, based on median values for aluminum plates from [12]. The edges of the plate are fixed to the cavity.

For meshing, triangular elements are applied to the plate with a maximum size of 0.024 m. The air cavity is meshed with tetrahedral elements, where the maximum element size, λ_{max}, is determined by the following equation:

$$\lambda_{max} = \frac{c}{N \cdot f_{max}} \tag{1}$$

Here, $c = 343\,m/s$ is the speed of sound, and $f_{max} = 1000\,Hz$ is the maximum frequency to be solved. N represents the number of mesh elements per wavelength, and it is recommended to use 7 to 10 elements per wavelength [13]. This vibroacoustic problem falls under multiphysics studies, requiring acoustic-structure coupling due to the interaction between the structural plate and the acoustic cavity. The cavity is modeled acoustically using the "Acoustics Frequency Domain" study. The air properties are set by $\rho_0 = 1.2043\,kg/m^3$ and $c = 343.2\,m/s$ at 20°C, assuming static pressure remains constant at 1 atm. The surrounding walls are

considered acoustically rigid, with $\bar{v}_n = 0$. To simulate a reverberation room, impedance is defined by $Z_0 = 100 \times \rho_0 \times c$.

2.3 Computation of the frequency response function

To determine these paths, the frequency response of the Noise Box is extracted for three distinct paths. The disturbance is modeled as a plane wave, applying its corresponding acoustic pressure to the aluminum plate. This pressure is the sum of the incident wave pressure p_{inc} and the reflected pressure p_{refl}, as defined in equations 2, 3, and 4.

$$p_{inc} = Ae^{i\phi} e^{-i(k_x x + k_y y + k_z z)} e^{i\Phi} \tag{2}$$

$$p_{refl} = Ae^{i\phi} e^{-i(-k_x x + k_y y + k_z z)} e^{i\Phi} \tag{3}$$

$$p_{total} = p_{inc} + p_{refl} \tag{4}$$

Here, k_x, k_y, k_z are the wave numbers, while Φ denotes the wave's phase. The angles θ and ϕ, which define the propagation direction, are both equal to $\pi/2$. The frequency response computation begins with a Single-Input-Single-Output (SISO) system. Sensors are modeled as 'Point Probes,' each measuring different physical entities: a reference accelerometer, an error microphone, and a canceling loudspeaker. The finite element model developed in previous section is used to compute the frequency response functions (FRFs) of all control paths in the active noise control (ANC) algorithm. These paths include the reference path, the primary path, and the secondary path in our case.

- **Reference Path:** The frequency response is the ratio of the incident acoustic pressure (Pa) to the plate's normal acceleration (m/s^2), yielding units of $(m/s^2)/Pa$.
- **Primary Path:** The response is the ratio of the incident acoustic pressure to the error microphone's measured pressure (Pa), making it dimensionless.
- **Secondary Path:** The canceling loudspeaker is modeled as a monopole with a volume flow rate of $1\,m^3/s$. Its unit is $Pa/m^3 \cdot s$.

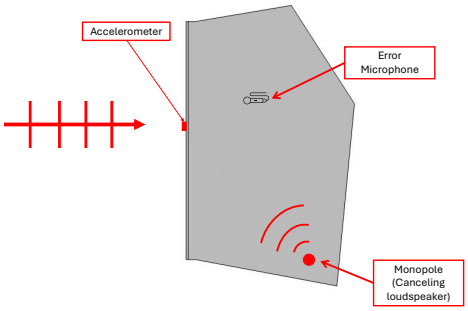


Fig. 2: A plane wave diagram represents the external disturbance entering the cavity, while an internal monopole simulates the canceling loudspeaker for the secondary path.

This setup allows for the analysis of the system's frequency response in noise control applications. Figure 2 presents the method used to compute theses FRFs.

Accurate FRF computation is crucial for deriving the state-space model required for control implementation. The transfer function is estimated using a residue-pole representation:

$$H(s) = \sum_n \frac{C_n}{s - A_n} + D \tag{5}$$

where $s = i\omega f$, C_n are residues, A_n are poles, and D is the direct term. The coefficients are approximated to match the computed frequency response using the AAA (Adaptive Antoulas-Anderson) algorithm [14]. The state-space model is first derived in the continuous domain and then discretized using the Zero-Order Hold (ZOH) method, assuming constant input between sampling intervals. This conversion, performed at 16 kHz, enables digital control design [15]. The resulting FRFs are shown in Figures 3, 4, and 5.

3 The SFANC-NLMS Algorithm

3.1 Presentation of the algorithm

The Selective Fixed-Filter Active Noise Control (SFANC) method enhances noise cancellation by pre-training a database of control filters and selecting the most suitable one in real-time [16]. It operates in two

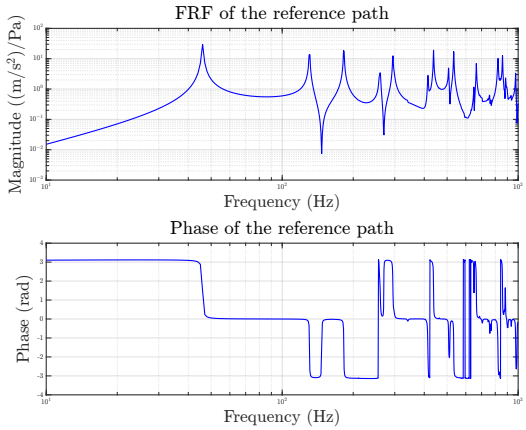


Fig. 3: Frequency response of the reference path

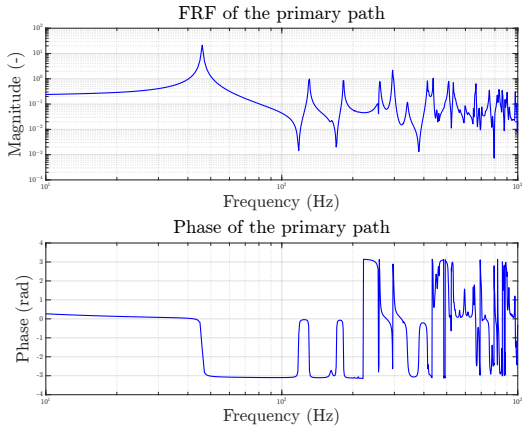


Fig. 4: Frequency response of the primary path

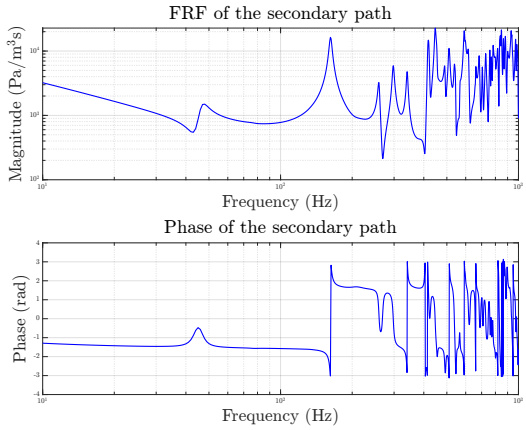


Fig. 5: Frequency response of the secondary path

stages: offline pre-training and online control. During pre-training, filtered white noise samples covering various frequency ranges (20 Hz–1000 Hz) are used to generate optimal control filters via the Normalized Least Mean Square (NLMS) algorithm, as shown in Table 2. These filters are stored in a database.

Table 2: Frequency ranges for each control filter C_i.

Control filter (C_i)	Frequency Range (Hz)
C_1	20 – 1000
C_2	20 – 500
C_3	500 – 1000
C_4	20 – 250
C_5	250 – 500
C_6	500 – 750
C_7	750 – 1000

During online operation, the system extracts noise characteristics and selects the optimal filter from the database, eliminating the need for continuous adaptation. This reduces computational complexity and enhances response time. SFANC-NLMS combines SFANC with adaptive filtering to refine the selected filter in real time. A co-processor uses a convolutional neural network (CNN) to analyze noise and choose the best filter, while an adaptive controller fine-tunes it dynamically. This hybrid approach ensures efficient and responsive noise reduction, making it suitable for real-time applications in dynamic environments. Figure 6 presents the algorithm diagram.

The refinement process is handled by the Normalized Least Mean Square (NLMS) algorithm, which updates the adaptive filter coefficients to minimize the residual error between the desired signal and the output. The NLMS algorithm uses the reference signal $r(n)$, and the error signal $e(n)$ to adjust the filter $w(n)$ through a normalized gradient descent. This update rule ensures stability even when the input signal power varies, making it effective in dynamic acoustic environments. The key equations governing the NLMS update are:

$$e(n) = d(n) - s(n) * \left(w^T(n) r(n)\right) \tag{6}$$

$$y(n) = w^T(n) r(n) \tag{7}$$

$$w(n+1) = w(n) + \mu\, e(n)\, \frac{r(n)}{r^T(n) r(n) + \varepsilon} \tag{8}$$

$$0 < \mu < \frac{1}{\lambda_{\max} D_s} \tag{9}$$

where μ is the step size, ε is a small regularization constant, D_s is the group delay of the secondary path, and $\lambda_{\max}$ is the maximum eigenvalue of the autocorrelation matrix of the reference signal. $d(n)$ is the primary noise and $s(n)$ is the secondary path impulse response.

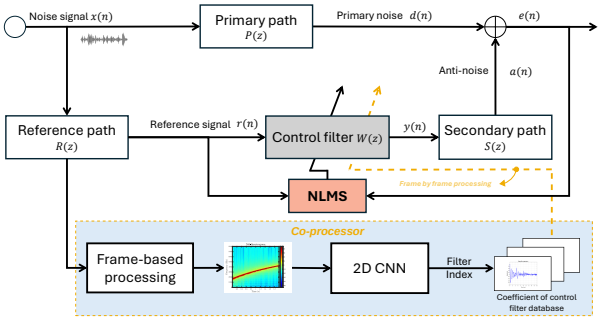


Fig. 6: Block diagram of the hybrid SFANC-NLMS algorithm. For each frame, the 2D CNN running on a co-processor selects the most appropriate pre-trained control filter. The NLMS algorithm then refines and updates the coefficients of the selected filter in real time at the sampling rate.

3.2 2D Convolutional Neural Network Architecture

The model architecture used is a modified version of a 2D CNN presented in [17]. The architecture is optimized to reduce the number of parameters by adjusting the convolutional layers and removing the Softmax layer in favor of direct classification. The detailed architecture of the 2D CNN is provided in Table 3. The input to the network consists of Mel spectrograms computed from audio signals with a sampling rate of 16 kHz. Each spectrogram is generated using a Fast Fourier Transform (FFT) size of 1024, a hop length of 512 , and a number of Mel filter banks set to 64. This results in spectrograms with a fixed frequency resolution of 64 Mel bands. The temporal resolution depends on the duration of the input audio. For a 1-second signal, the spectrogram has a width of 31 time frames. The network begins with a convolutional layer with a 5×5 kernel, extracting basic spectral patterns and producing 16 feature maps. This is followed by a ReLU activation and a 3×3 max pooling operation. The second convolutional block consists of two 3×3 convolutional layers, increasing the feature maps to 32 while preserving spatial resolution. A 6×6 max pooling operation further reduces the dimensionality. The extracted features are then flattened, batch-normalized, and passed through a fully connected layer of 15 neurons. Finally, a second fully connected layer produces the final classification output.

3.3 Training

To train the 2D CNN model, we generated a dataset of 80,000 Mel spectrograms for training and 2,000 for validation and testing. The spectrograms were computed from broadband noise recordings sampled at 16 kHz, filtered through the previously described reference path using randomly varying frequency bands. Each spectrogram corresponds to a 1-second noise segment, resulting in an input size of $1 \times 64 \times 31$. The model was trained using the Adam optimizer with a learning rate of 0.0001, a batch size of 250, and for 30 epochs. The training objective was to minimize the CrossEntropyLoss function. To prevent overfitting, L2 regularization with a factor of 10^{-4} was applied to all trainable parameters. Figure 7 shows the evolution of loss and accuracy during training.

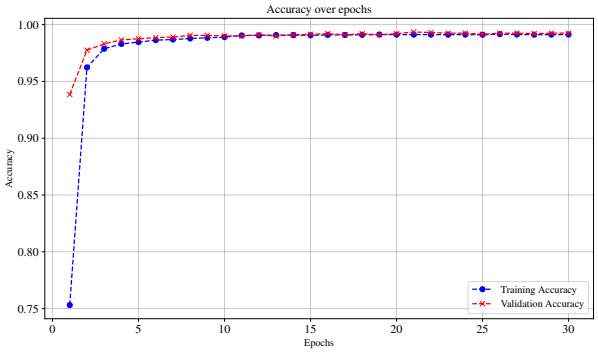
The performance of the proposed 2D CNN was compared against several other 2D network architectures, including the 2D CNN introduced in [17], the ShuffleNet model presented in [18], MobileNet model proposed in [19], and DenseNet presented in [20]. Table 4 summarizes the comparison of these networks. Based on the results, the proposed 2D CNN stands out as the most promising architecture, considering both its accuracy and the number of parameters it requires. To enable this comparison, all architectures were retrained using the constructed dataset.

4 Results

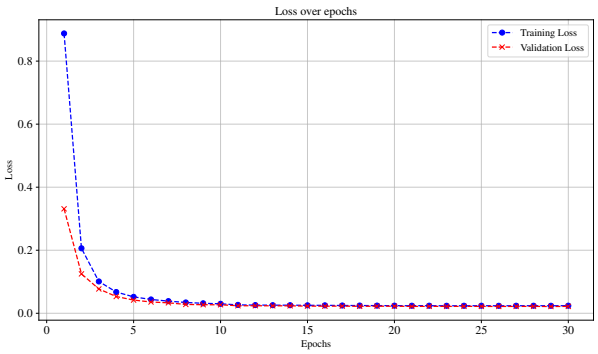
In this section, numerical simulations are performed using a single-channel ANC framework. The system operates at a sampling frequency of 16 kHz, with the control filter consisting of 8192 coefficients. The state-space coefficients introduced in Section 2.3 define the

Table 3: Summary of the proposed 2D CNN network architecture. The input consists of Mel spectrograms with 64 Mel bands and a variable number of time frames, depending on the duration of the input signal. For a 1-second signal.

Layer Index	Layer Type	Configuration	Output Shape
1	Convolutional Layer	Kernel: 5×5, Channels: $1 \rightarrow 16$, Stride: 1, Padding: 2	$[16, 64, 31]$
2	Activation (ReLU)	-	$[16, 64, 31]$
3	Max Pooling	Kernel: 3×3, Stride: 3	$[16, 21, 10]$
4	Convolutional Layer	Kernel: 3×3, Channels: $16 \rightarrow 32$, Stride: 1, Padding: 1	$[32, 21, 10]$
5	Activation (ReLU)	-	$[32, 21, 10]$
6	Convolutional Layer	Kernel: 3×3, Channels: $32 \rightarrow 32$, Stride: 1, Padding: 1	$[32, 21, 10]$
7	Max Pooling	Kernel: 6×6, Stride: 3	$[32, 5, 3]$
8	Flatten	-	$[384]$
9	Batch Normalization	Features: 384	$[384]$
10	Fully Connected Layer	Input: 384, Output: 15	$[15]$
11	Fully Connected Layer	Input: 15, Output: 7	$[7]$



(a) Training Accuracy



(b) Training Loss

Fig. 7: Accuracy and loss curves for the training of the proposed 2D CNN.

reference, primary, and secondary paths. Moreover, a total of seven pre-trained control filters are employed.

Table 4: Test accuracy and parameter comparison for various networks.

Network	Test Accuracy	Parameters
Proposed method	99.35%	20,959
2D CNN [17]	96.30%	25,855
ShuffleNet [18]	98.80%	246,110
MobileNet [19]	97.45%	2,457,721
DenseNet [20]	99.40%	6,954,759

As detailed in Table 2, these filters are obtained by training on seven distinct white noise signals, each constrained to a specific frequency range. The NLMS algorithm is used to compute the optimal control filters for these training noises, with a step size set to 0.1. To assess the performance of the algorithm, broadband noise is utilized as the disturbance signal. Stationary noise is modeled as band-limited random noise spanning 20 to 450 Hz. The SFANC-NLMS method is benchmarked against both the NLMS and SFANC algorithms.

The stationary noise is generated as a zero-mean, unit-variance Gaussian process. This signal undergoes band-pass filtering between 20 Hz and 450 Hz, using a filter designed with a Hamming window. The window characteristics determine the passband ripple and stopband attenuation, ensuring precise frequency selection. The sampling frequency is maintained at 16 kHz across all simulations, while the control filter length remains fixed at 8192 taps. The simulations are conducted on 10-second-long signals. The results obtained under

these conditions are illustrated in Figure 8, depicting the time-domain response of the controlled signals.

Examining the results depicted in Figure 8 for the different algorithms, we observe that for broadband noise, SFANC exhibits superior performance compared to NLMS. This enhanced effectiveness significantly influences the overall control achieved with SFANC-NLMS. Notably, SFANC-NLMS achieves a lower error amplitude due to its adaptive structure. Additionally, in the case of SFANC, active control is initiated after approximately 1 second. This delay allows the convolutional neural network (CNN) to analyze the characteristics of the primary noise and determine the most suitable pre-trained controller. Conversely, SFANC-NLMS applies control immediately from the start, leveraging its adaptive capabilities to adjust in real time.

Figure 9 presents the Power Spectral Density (PSD) of each controlled signal. The lowest levels are achieved by the signal controlled with SFANC-NLMS (represented by the red curve). To evaluate its effectiveness, we compare it with the PSD of the signal controlled by SFANC (the yellow curve). These results indicate that SFANC-NLMS consistently outperforms all other algorithms at each resonance frequency of the system. Additionally, at anti-resonance frequencies, such as around 110 Hz or 370 Hz, the signal levels controlled by SFANC and SFANC-NLMS exceed that of the primary noise. This phenomenon, known as the waterbed effect, describes a trade-off in control systems, where reducing energy in certain frequency bands inevitably causes an increase in others. This effect is particularly pronounced in non-minimum phase systems, where the presence of unstable zeros imposes fundamental limitations on the achievable performance of control algorithms.

Figure 10 illustrates the noise reduction achieved over time. At the beginning of the process, the noise reduction remains limited due to the initial response delay induced by the non-minimum phase (NMP) characteristics of the system and the time required by SFANC to initiate control. However, after approximately 2 seconds of computation, optimal control is established.Beyond this point, SFANC initially achieves a greater noise reduction compared to SFANC-NLMS. This discrepancy may stem from variations in the controller coefficients, which could momentarily mislead the adaptive algorithm in minimizing the error. Nevertheless, for the remainder of the signal, SFANC-NLMS provides superior noise reduction. On average, the noise reduction achieved by each algorithm is as follows: 1.66 dB for NLMS, 4.02 dB for SFANC, and 5 dB for SFANC-NLMS.

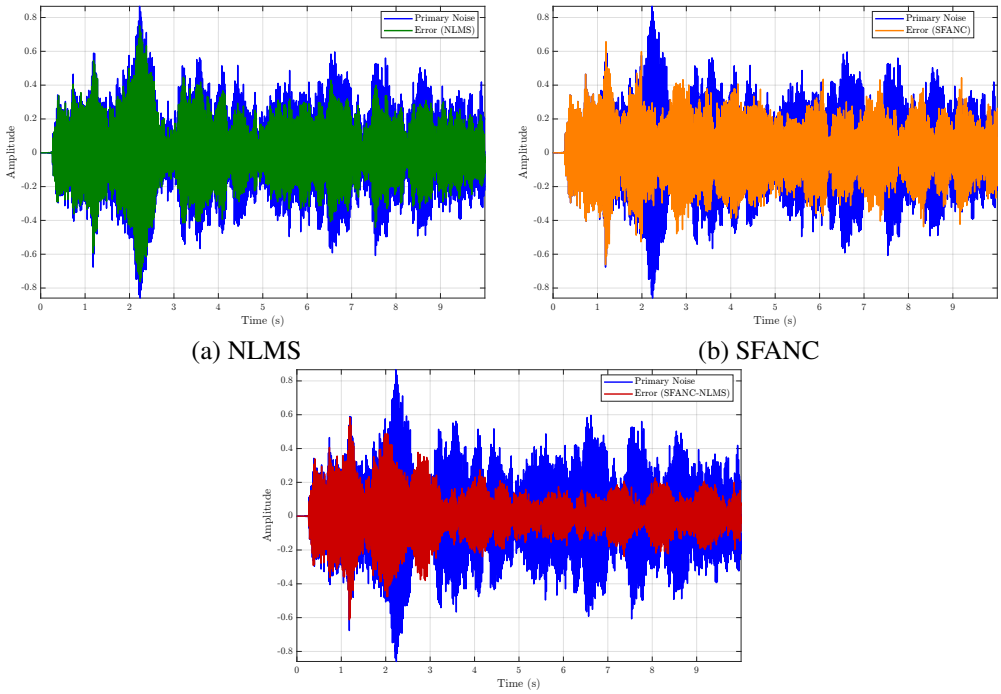
5 Conclusion

This study presents an experimental framework aimed at characterizing the frequency response function of an encapsulated structure, referred to as the Noise Box, which emulates the acoustic properties of a car interior. Given its non-minimum phase behavior, this environment poses significant challenges for active noise control (ANC). The acquired frequency response data were leveraged to develop and implement a novel control strategy. The proposed SFANC-NLMS algorithm integrates deep learning into the ANC framework. Its approach relies on constructing a dataset of multiple pre-trained control filters and employing a 2D convolutional neural network (CNN) to dynamically select the most appropriate controller based on the characteristics of the incoming noise. Additionally, an adaptive algorithm is incorporated to refine the coefficients of the chosen pre-trained controller, ensuring optimal performance in real time. This dual mechanism enhances efficiency beyond conventional adaptive control strategies while mitigating the limitations of SFANC when the primary path undergoes variations over time. A key advantage of the proposed approach is the elimination of secondary path estimation, which traditionally introduces discrepancies between the modeled and actual system dynamics. By bypassing this constraint, SFANC-NLMS reduces computational complexity and demonstrates that, in this specific setup, directly employing NLMS instead of FxNLMS can yield competitive results.

Looking ahead, while deep learning currently plays a supporting role in optimizing the selection of control filters, future advancements could push its involvement further—potentially enabling a neural network to directly synthesize the anti-noise signal. Such a development would be particularly valuable for addressing nonlinear distortions associated with the acoustic environment and electro-acoustic interactions induced by the materials used in noise control implementations.

References

[1] Fuller, C., Hansen, C., and Snyder, S., “Active control of sound radiation from a vibrating rectangular panel by sound sources and vibration inputs:



(a) NLMS (b) SFANC

(c) SFANC-NLMS

Fig. 8: Graphics of the primary noise and error signals obtained using different ANC algorithms.

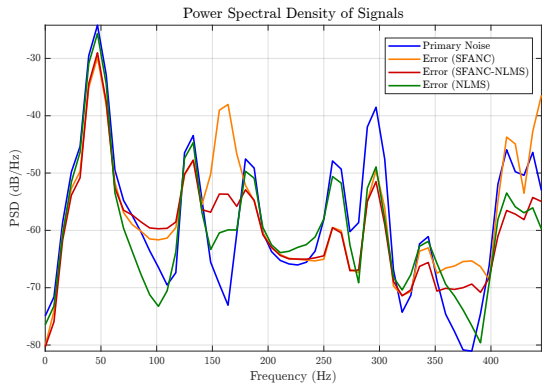


Fig. 9: Power Spectral Density (PSD) of the primary noise and the controlled noise achieved by NLMS, SFANC, and SFANC-NLMS

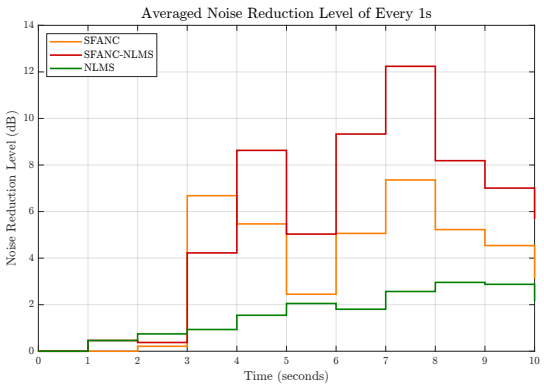


Fig. 10: Average noise reduction each second of the controlled noise using NLMS, SFANC, and SFANC-NLMS.

an experimental comparison," *Journal of Sound and Vibration*, 145(2), pp. 195–215, 1991.

[2] Misol, M., Algermissen, S., and Monner, H. P., "Experimental investigation of different active noise control concepts applied to a passenger car equipped with an active windshield," *Journal of Sound and Vibration*, 331(10), pp. 2209–2219, 2012.

[3] Belgacem, W., Berry, A., and Masson, P., "Active vibration control on a quarter-car for cancellation

of road noise disturbance," *Journal of Sound and Vibration*, 331(14), pp. 3240–3254, 2012.

[4] Aboutiman, A., Rachman, Z., Oberman, T., Aletta, F., Kang, J., Karimi, H. R., and Ripamonti, F., "Subjective perception analysis of active noise control algorithms in an encapsulated structure: An experimental study," *Applied Acoustics*, 239, p. 110823, 2025, ISSN 0003-682X, doi: https://doi.org/10.1016/j.apacoust.2025.110823.

[5] Milani, A. A., Kannan, G., and Panahi, I. M., "On maximum achievable noise reduction in ANC systems," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 349–352, IEEE, 2010.

[6] Wu, L., Qiu, X., and Guo, Y., "A generalized leaky FxLMS algorithm for tuning the waterbed effect of feedback active noise control systems," *Mechanical Systems and Signal Processing*, 106, pp. 13–23, 2018.

[7] Aboutiman, A., Shams, R., Karimi, H. R., Ripamonti, F., and Pawełczyk, M., "Active noise control in encapsulated structures with non-minimum phase characteristics using a Kalman filter approach," *Journal of Sound and Vibration*, 615, p. 119187, 2025, ISSN 0022-460X, doi:https://doi.org/10.1016/j.jsv.2025.119187.

[8] Luo, Z., Shi, D., and Gan, W.-S., "A hybrid sfancfxnlms algorithm for active noise control based on deep learning," *IEEE Signal Processing Letters*, 29, pp. 1102–1106, 2022.

[9] Liu, L., *Design, simulation and testing of a coupled plate-cavity system targeted for vehicle interior noise analysis and control*, Ph.D. thesis, Politecnico di Milano, 2021.

[10] Cremer, L. and Heckl, M., *Structure-borne sound: structural vibrations and sound radiation at audio frequencies*, Springer Science & Business Media, 2013.

[11] Chapelle, D. and Bathe, K.-J., *The finite element analysis of shells-fundamentals*, Springer Science & Business Media, 2010.

[12] Lenk, P. and Coult, G., "Damping of glass structures and components," in *Challenging Glass Conference Proceedings*, volume 2, pp. 341–350, 2010.

[13] Astley, R., Macaulay, G., and Coyette, J., "Mapped wave envelope elements for acoustical radiation and scattering," *Journal of Sound and Vibration*, 170(1), pp. 97–118, 1994.

[14] Nakatsukasa, Y., Sète, O., and Trefethen, L. N., "The AAA algorithm for rational approximation," *SIAM Journal on Scientific Computing*, 40(3), pp. A1494–A1522, 2018.

[15] Moir, T. J. and Moir, T. J., *Rudiments of Signal Processing and Systems*, Springer, 2022.

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

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

[18] Luo, Z., Shi, D., Ji, J., Shen, X., and Gan, W.-S., "Real-time implementation and explainable AI analysis of delayless CNN-based selective fixed-filter active noise control," *Mechanical Systems and Signal Processing*, 214, p. 111364, 2024.

[19] Adapa, S., "Urban sound tagging using convolutional neural networks," *arXiv preprint arXiv:1909.12699*, 2019.

[20] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q., "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.