

eISSN 2300-262X

ISSN 0137-5075



POLISH ACADEMY OF SCIENCES
INSTITUTE OF FUNDAMENTAL TECHNOLOGICAL RESEARCH
COMMITTEE ON ACOUSTICS

ARCHIVES of ACOUSTICS

QUARTERLY

Vol. 50, No. 2, 2025

WARSAW



ARCHIVES of ACOUSTICS

QUARTERLY, Vol. 50, No. 2, 2025

Research Papers

Najamuddin, U.U. Sheikh, A.Z. Sha'ameri, <i>CAPSE-ViT: A lightweight framework for underwater acoustic vessel classification using coherent spectral estimation and modified vision transformer</i>	161
S. Mahmoud, L. Saleh, I. Chouaib, <i>Experimental results of diver detection in harbor environments using single acoustic vector sensor</i>	173
F. Zeng, Y. Han, H. Yang, D. Yang, F. Zheng, <i>Single vector hydrophone DOA estimation: Leveraging deep learning with CNN-CBAM</i>	187
Y. Tang, C. Fan, Y. Li, C. Chen, Z. Peng, <i>A hybrid method for predicting the bistatic target strength based on the monostatic scattering sound pressure data</i>	199
I.Y. Iliev, H.Z. Karaivanov, <i>Mapping sound pressure levels: A novel approach to determining near-field and far-field regions</i>	209
M. Melnyk, T. Kamisiński, R. Kinasz, R. Vynarovych, <i>An IT system for assessing noise level distribution in historical urban centers</i>	215
X. Han, J. Peng, <i>Multi-label bird species classification using transfer learning network</i>	223
P. Paupy, P. Tabla, D. Huggenberger, F. Elfi, E.N. Morel, J.R. Torga, <i>Determination of the elastic constant of the top plate of a cello in the interaction with the bridge</i>	235
G.S. Paddan, M.J. Howell, <i>Assessment of noise from different types of rounds fired from a saluting gun</i>	243
A. Oktav, A. Ener, F. Özenc, S. Sari, <i>The impact of HVAC systems on speech intelligibility in university classrooms</i>	253
P. Yang, N. Mohamad Usamah, A.A. Saad, A.Z. Ahmad Mazlan, <i>Structural parametric study of a piezoelectric energy harvester for a specific excitation frequency of an electric motor, considering fatigue life</i>	265
S. Sachar, P. Doerffer, P. Flaszynski, J. Kotus, K. Doerffer, J. Grzelak, M. Piotrowicz, <i>Aeroacoustic effect of savonius rotor segmentation</i>	277

Editorial Board

Editor-in-Chief: NOWICKI Andrzej (Institute of Fundamental Technological Research PAS, Poland)

Deputy Editor-in-Chief: GAMBIN Barbara (Institute of Fundamental Technological Research PAS, Poland)

Associate Editors

General linear acoustics and physical acoustics

RDZANEK Wojciech P. (University of Rzeszów, Poland)

SNAKOWSKA Anna (AGH University of Krakow, Poland)

Architectural acoustics

KAMISIŃSKI Tadeusz (AGH University of Krakow, Poland)

MEISSNER Mirosław (Institute of Fundamental Technological Research PAS, Poland)

Musical acoustics and psychological acoustics

MIŚKIEWICZ Andrzej (The Fryderyk Chopin University of Music, Poland)

PREIS Anna (Adam Mickiewicz University, Poland)

Underwater acoustics and nonlinear acoustics

MARSZAL Jacek (Gdańsk University of Technology, Poland)

Speech, computational acoustics, and signal processing

DRGAS Szymon (Poznan University of Technology)

KOCIŃSKI Jędrzej (Adam Mickiewicz University, Poland)

Ultrasonics, transducers, and instrumentation

GAMBIN Barbara (Institute of Fundamental Technological Research PAS, Poland)

OPIELIŃSKI Krzysztof (Wrocław University of Science and Technology, Poland)

TASINKIEWICZ Jurij (Institute of Fundamental Technological Research PAS, Poland)

Sonochemistry

DZIDA Marzena (University of Silesia in Katowice, Poland)

Electroacoustics

ŻERA Jan (Warsaw University of Technology, Poland)

Vibroacoustics, noise control and environmental acoustics

ADAMCZYK Jan Andrzej (Central Institute for Labor Protection – National Research Institute, Poland)

KLEKOT Grzegorz (Warsaw University of Technology, Poland)

KOMPALA Janusz (Central Mining Institute, Poland)

LENIOWSKA Lucyna (University of Rzeszów, Poland)

PIECHOWICZ Janusz (AGH University of Krakow, Poland)

PLEBAN Dariusz (Central Institute for Labor Protection – National Research Institute, Poland)

Journal Managing Editor: JEZIERSKA Eliza (Institute of Fundamental Technological Research PAS, Poland)

Advisory Editorial Board

Chairman: KOZACZKA Eugeniusz (Polish Academy of Sciences, Poland)

BATKO Wojciech (AGH University of Krakow, Poland)

BLAUERT Jens (Ruhr University, Germany)

BRADLEY David (The Pennsylvania State University, USA)

CROCKER Malcolm J. (Auburn University, USA)

DOBRUCKI Andrzej (Wrocław University of Science and Technology, Poland)

HANSEN Colin (University of Adelaide, Australia)

HESS Wolfgang (University of Bonn, Germany)

LEIGHTON Tim G. (University of Southampton, UK)

LEWIN Peter A. (Drexel University, USA)

MAFFEI Luigi (Second University of Naples SUN, Italy)

PUSTELNY Tadeusz (Silesian University of Technology, Poland)

SEREBRYANY Andrey (P.P. Shirshov Institute of Oceanology, Russia)

SUNDBERG Johan (Royal Institute of Technology, Sweden)

ŚLIWIŃSKI Antoni (University of Gdańsk, Poland)

TITTMANN Bernhard R. (The Pennsylvania State University, USA)

TORTOLI Piero (University of Florence, Italy)

VORLÄNDER Michael (Institute of Technical Acoustics, RWTH Aachen, Germany)

Polish Academy of Sciences
Institute of Fundamental Technological Research PAS
Committee on Acoustics PAS

Editorial Board Office

Pawińskiego 5B, 02-106 Warsaw, Poland

phone (48) 22 826 12 81 ext. 206

e-mail: akustyka@ippt.pan.pl <https://acoustics.ippt.pan.pl>

Indexed in BazTech, Science Citation Index-Expanded (Web of Science Core Collection),

ICI Journal Master List, Scopus, PBN – Polska Bibliografia Naukowa,

Directory of Open Access Journals (DOAJ)

Recognised by The International Institute of Acoustics and Vibration (IIAV)

Edition co-sponsored by the Ministry of Science and Higher Education

PUBLISHED IN POLAND

Typesetting in L^AT_EX: JEZIERSKA Katarzyna (Institute of Fundamental Technological Research PAS, Poland)

Research Paper

CAPSE-ViT: A Lightweight Framework for Underwater Acoustic Vessel Classification Using Coherent Spectral Estimation and Modified Vision Transformer

NAJAMUDDIN, Usman Ullah SHEIKH, Ahmad Zuri SHA'AMERI

*Faculty of Electrical Engineering, Universiti Teknologi Malaysia, UTM Skudai
Johor, Malaysia*

*Corresponding Author e-mail: najamuddin@graduate.utm.my

*Received February 24, 2025; revised April 10, 2025; accepted April 10, 2025;
published online May 29, 2025.*

Underwater acoustic target classification has become a key area of research for marine vessel classification, where machine learning (ML) models are leveraged to identify targets automatically. The major challenge is inserting area-specific understanding into ML frameworks to extract features that effectively distinguish between different vessel types. In this study, we propose a model that uses the coherently averaged power spectral estimation (CAPSE) algorithm. Vessel frequency spectra is first computed through the CAPSE analysis, capturing key machinery characteristics. Further, the features are processed via a vision transformer (ViT) network. This method enables the model to learn more complex relationships and patterns within the data, thereby improving the classification performance. This is accomplished by using self-attention mechanisms to capture global dependencies between features, enabling the model to focus on relationships throughout the entire input. The results, evaluated on standard DeepShip and ShipsEar datasets, show that the proposed model achieved a classification accuracy of 97.98% and 99.19% while utilizing just 1.90 million parameters, outperforming other models such as ResNet18 and UATR-Transformer in terms of both accuracy and computational efficiency. This work offers an improvement to the development of efficient marine vessel classification systems for underwater acoustics applications, demonstrating that high performance can be achieved with reduced computational complexity.

Keywords: underwater acoustic targets; CAPSE; vision transformer; CNN; LOFAR gram.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Accurate classification of underwater acoustic targets is crucial in naval defense, underwater surveillance, and environmental monitoring (BJØRNØ, 2017; DOMINGOS *et al.*, 2022; THOMAS *et al.*, 2020). The ability to distinguish between different types of marine vessels based on their acoustic signatures is essential for operations such as threat detection and marine traffic management (MCKENNA *et al.*, 2024). However, the underwater acoustic environment poses unique challenges due to complex propagation effects, ambient noise, and interference from various sources, making this task particularly difficult (ASLAM *et al.*, 2024). Traditional classification methods, while effective in

controlled conditions, become less efficient with high levels of noise and randomness present in real-world underwater scenarios. This creates a need for improved methods that can enhance the quality of the target signature and take advantage of deep learning to achieve more accurate and reliable classification (LUO *et al.*, 2023).

Conventional underwater acoustic target classification has relied on signal processing techniques such as Fourier transforms, wavelet analysis, and mel-frequency cepstral coefficients (MFCC), which are effective for identifying specific features in clean signals (Müller *et al.*, 2024). However, these methods face difficulties when dealing with highly noisy or distorted signals. With the rise of deep learning, convolutional

neural networks (CNNs) have been employed to classify acoustic signals by first transforming them into spectrograms and then treating the problem as an image classification task (ZENG *et al.*, 2019). CNNs leverage spatial hierarchies to capture local features from these spectrograms, but their reliance on local convolutions limits their ability to capture global dependencies in the data (YANG *et al.*, 2024). This shortcoming is particularly problematic for underwater acoustic signals, where the temporal and spectral relationships within the signal are essential for accurate target classification. Local features in a spectrogram refer to specific, small-scale patterns over short time or frequency ranges, such as individual machinery noises (FENG, ZHU, 2022). Global features, in contrast, represent broader patterns across time and frequency, capturing the overall acoustic signature of the source or vessel. The vision transformer (ViT), a recently developed deep learning model, offers an alternative approach by employing a self-attention mechanism that captures both local and global dependencies, making it more suited for tasks that require holistic data analysis (DOSOVITSKIY *et al.*, 2020).

Another major challenge in underwater acoustics is the low signal-to-noise ratio (SNR), making target detection and classification challenging. Earlier research indicates that targets become undetectable when SNR falls below critical levels of -14.4 dB, and with nearly 90% of vessels receiving SNR below 0 dB in ambient noise conditions (SIDDAGANGAIAH *et al.*, 2016). Another significant challenge in underwater acoustics is the low SNR of acoustic signatures resulting from environmental noise, surface reflections, and interference (LAMPERT, O'KEEFE, 2013). Preprocessing techniques, such as the coherently averaged power spectral estimation (CAPSE), are designed to enhance the quality of acoustic signals by averaging power spectra across multiple observations, thus reducing noise and improving the clarity of key signal features (LAN *et al.*, 2020).

The integration of CAPSE and ViT forms the core of this study. We use CAPSE as a preprocessing step to improve tonal signals and minimize noise, highlighting target-specific features through coherent spectral averaging. The processed signals are then converted into low frequency analysis and recording (LOFAR) grams, which are fed into the ViT model for classification. The ViT ability to capture both local and global machinery features dependencies using its attention mechanism is exploited. The proposed method was assessed using DeepShip (IRFAN *et al.*, 2021) and ShipsEar (SANTOS-DOMÍNGUEZ *et al.*, 2016), a publicly available dataset, where it outperformed other methods reported in the literature, delivering higher accuracy and enhanced generalization. This approach highlights the potential for incorporating CAPSE and the modified ViT deep learning method for improving classification performance in noisy underwater.

The rest of the article is arranged as follows: Sec. 2 is an overview of existing studies in literature; Sec. 3 highlights the proposed methodology, dataset preprocessing techniques, and the model parameters employed in the experiments; Sec. 4 showcases the results, emphasizing the advantages of the proposed methodology; finally, Sec. 5 provides a conclusion, summarizing the main contributions and proposed future work.

2. Related works

Research on marine vessel classification using acoustic noise has explored a range of signal processing techniques and machine learning models (BIANCO *et al.*, 2019). Early methods relied heavily on manual interpretation of acoustic signatures by sonar operators, depending entirely on their expertise (DOMINGOS *et al.*, 2022). However, with advancements in computational power and deep learning techniques, automated classification has become an area of increasing interest.

One of the most established techniques is the fast Fourier transform (FFT), which converts time-domain signals into the frequency domain, enabling the identification of spectral components, for understanding the underlying patterns (FENG *et al.*, 2021). However, it is ineffective for representing underwater acoustic signals due to their non-stationary nature. The wavelet transform offers both time and frequency information, providing variable resolution that makes it effective for analyzing non-stationary signals with varying patterns, but it is sensitive to ambient noise (KIM *et al.*, 2021). MFCCs are frequently applied in sound analysis due to their ability to capture the perceptual characteristics of audio signals (LIM *et al.*, 2007). They are also designed to mimic the human ear's sensitivity to different frequencies, making them particularly useful in tasks such as speech and sound classification (SHARMA *et al.*, 2020). The constant-Q transform offers constant resolution across octaves, making it well-suited for logarithmic frequency analysis (SINGH *et al.*, 2021). In addition to these general techniques, LOFAR is a method focused on detecting long-term spectral patterns, which is particularly useful for identifying sustained sounds such as engine noise or other mechanical signals (LI *et al.*, 2023). On the other hand, the detection of envelope modulation on the noise (DEMON) technique is specifically designed to detect modulation spectrum caused by rotating components such as propellers and blades (PARK, JUNG, 2021).

Numerous multi-modal recognition techniques have been investigated for marine vessel classification. In (YUAN *et al.*, 2019), a method was developed that combines both optical images and radiated noise from vessels as input data, allowing for a more comprehensive classification approach. LUO *et al.* (2021) applied a multi-window spectral analysis method to capture a range of in-band frequency features, provid-

ing a more detailed and accurate representation of the acoustic environment. Additionally, SONG *et al.* (2021) significantly improved underwater noise classification by extracting the one-third octave noise spectrum, power spectral density, and MFCC features. These various approaches aim to increase classification accuracy by integrating and leveraging multiple feature sets, enhancing the robustness of the recognition process.

Machine learning techniques, such as support vector machines (SVM) and shallow neural networks (SNN), have long been used in underwater acoustic classification. These techniques rely on efficient feature extraction methods to transform raw acoustic signals into feature vectors, which are subsequently input into the network (DE MOURA, DE SEIXAS, 2016). For example, SHERIN and SUPRIYA (2015), used enhanced SVM classifiers to differentiate types of vessel noise. With advances in deep learning, research has increasingly focused on more complex neural networks. KHISHE and MOHAMMADI (2019) applied MFCC as inputs to a neural network optimized by the salp swarm algorithm and achieved an accuracy of 97.1% (HEGAZY *et al.*, 2020). However, these fully connected networks still face challenges in capturing deep, complex features in multiple-class scenarios due to their relatively simple architecture.

To overcome these limitations, CNNs have been used to map raw waveforms or time-frequency representations directly to vessel types (HU *et al.*, 2021; LUO *et al.*, 2021). CNNs have demonstrated good performance in classifying vessels using acoustic signals. For instance, CAO *et al.* (2019), introduced the CNN combining second-order pooling (SOP) and the constant-Q transform for feature extraction, outperforming traditional classifiers such as VGG-Net and deep belief networks by achieving an accuracy of 96.3%. Custom CNN architectures, such as VesselNet, have also been proposed to enhance the classification of LOFAR spectrograms. CINELLI *et al.* (2018) designed VesselNet specifically for spectrogram classification, using the two-pass split-window filter with resulting in a precision of 88.1% on proprietary dataset (DE CARVALHO *et al.*, 2021).

The transformer architecture has been extensively applied in fields such as natural language processing (NLP) (RAFFEL *et al.*, 2020), computer vision (CV) (DOSOVITSKIY, 2020), and audio classification (NOUMIDA, RAJAN, 2022), consistently demonstrating superior performance. Recently, CHEN *et al.* (2024) introduced Swin transformer for ship-radiated noise classification, combining DEMON spectra and mel-spectrograms through feature fusion and attention mechanisms. The achieved performance on standard dataset was 98.62% and 99.01%. However, its performance with weak acoustic signals due to masking by both self-generated broadband noise, an increase in distance from the receiver, and ambient noise from nat-

ural sources is unknown. This masking effect degrades the clarity and detectability of the vessel's tonal components (IKPEKHA *et al.*, 2018). Similarly, the large size of the deep learning network makes it unsuitable for real-time applications. This paper introduces the ViT, with self-attention, as the classifier. This lightweight transformer architecture significantly reduces training time and resource requirements (CHEN *et al.*, 2024).

3. Methodology

This section details the methodology for underwater acoustic target classification, using CAPSE for signal enhancement and ViT for classification. CAPSE improves spectral clarity by reducing noise, while ViT leverages self-attention mechanisms to capture patterns in the enhanced spectrograms and improve classification accuracy.

3.1. CAPSE

CAPSE is a signal processing technique designed to enhance the detection of sinusoids in noisy environments. Unlike traditional methods such as the periodogram and Welch's method, CAPSE preserves phase coherence across multiple signal segments, resulting in a substantial improvement in SNR (FENG *et al.*, 2021).

For a sinusoidal signal S_0 embedded in noise, the Fourier transform for each segment, S_k can be expressed as

$$S_k(\omega) = S_0(\omega)e^{j\phi_k}, \quad (1)$$

where $\phi_k = \omega_0 kD$ represents the phase difference between the Fourier transforms of the k -th and the 1st segments at frequency ω_0 . CAPSE aims to coherently average the signal across multiple segments K , thus enhancing SNR:

$$\bar{X}(\omega) = (1/K) \sum_{k=0}^{K-1} X_k(\omega)e^{-j\phi_k}. \quad (2)$$

This offset introduces a phase variation across segments, which can be corrected by applying an additional DFT along the segment indices given in Eq. (3), yielding:

$$\hat{X}(\omega_l, \omega_m) = (1/K) \sum_{k=0}^{K-1} X_k(\omega_l)e^{-j\omega_v k}, \quad (3)$$

$$\omega_l = \arg \max_{\omega_m} |X(\omega_l, \omega_v)|^2. \quad (4)$$

CAPSE spectrum is then defined in Eq. (5), where ω_l and ω_m are the indices of angular frequencies, measured in radians per second, and K is the number of segments (LAN *et al.*, 2020):

$$P_{xx}^{\text{CAPSE}}(\omega) = (1/UM) |\hat{X}(\omega_l, \omega_{\delta l})|^2. \quad (5)$$

By maximizing the energy component, CAPSE preserves the most significant spectral information, making it a robust method for tonal detection in noisy environments. Details of the algorithm can be found in (FENG *et al.*, 2021).

3.2. Vision transformer

ViT is employed as the classification model, leveraging its self-attention mechanism to capture both local and global dependencies on data features (DOSOVITSKIY *et al.*, 2020). The acoustic signals pre-processed with CAPSE are transformed into LOFAR grams and treated as 2D images, displaying frequency components along the horizontal axis and temporal progression along the vertical axis. Let LOFAR gram be denoted as $x \in \mathbb{R}^{W \times H}$, where W is the number of frequency bins and H is the number of time steps. The LOFAR gram is divided into non-overlapping patches, each of size $Q \times P$, where Q and P represent the patch dimensions in frequency and time domains, respectively. These patches are then flattened into 1D vectors, creating a sequence of patch embeddings, $X_p = [x_1, x_2, \dots, x_N]$, where $N = 400$ is the total number of patches. Each patch acts as an independent token for the transformer input. Positional encodings $E \in \mathbb{R}^{N \times D}$ are added to the patch embeddings to retain the relative positional information, creating the input sequence, z_0 for the transformer layers,

$$z_0 = [x_1 E, x_2, \dots, x_N E] + E. \quad (6)$$

The core of the ViT model is its multi-head self-attention mechanism, which allows the model to compute attention weights between different patches in the sequence. For each attention head, the input sequence z_0 is transformed into a query (\mathbf{Q}), key (\mathbf{K}), and value (\mathbf{V}) matrices:

$$\mathbf{Q} = z_0 \mathbf{W}_{\mathbf{Q}}, \quad \mathbf{K} = z_0 \mathbf{W}_{\mathbf{K}}, \quad \mathbf{V} = z_0 \mathbf{W}_{\mathbf{V}}, \quad (7)$$

where $\mathbf{W}_{\mathbf{Q}}, \mathbf{W}_{\mathbf{K}}, \mathbf{W}_{\mathbf{V}} \in \mathbb{R}^{D \times D}$ are the learnable weight matrices. The attention score for each patch is computed as (PANG *et al.*, 2023)

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{D}}\right) \mathbf{V}. \quad (8)$$

Multiple attention heads are applied in parallel, enabling the model to focus on different regions of the LOFAR gram simultaneously. The outputs from the attention heads are concatenated and then passed through a feed-forward network for additional processing.

Following the attention module, a series of transformer encoder layers are applied, each containing a multi-head self-attention block and a position-wise feed-forward network. These layers help in progressively learning higher-level representations from the patch sequence. Each encoder layer includes residual connections and layer normalization to stabilize training:

$$z'_l = \text{LayerNorm}(z_{l-1} + \text{MultiHeadAttention}(z_{l-1})), \quad (9)$$

$$z_l = \text{LayerNorm}(z'_l + \text{FeedForward}(z'_l)), \quad (10)$$

where l denotes the current transformer layer.

After passing through several transformer layers, the final sequence representation z_L is obtained from the last encoder block. A class token is appended to the patch sequence during input, and this token's representation at the final layer z_L^{class} is extracted and passed to a classification head. The classification head consists of a fully connected layer followed by a softmax activation function, which produces the class probabilities for the acoustic target:

$$y = \text{softmax}\left(\mathbf{W}_{\text{class}} z_L^{\text{class}}\right), \quad (11)$$

where $\mathbf{W}_{\text{class}}$ is the weight matrix of the classification layer.

Figure 1 presents the process flow, where the classifier takes $50 \times 4000 \times 1$ grayscale LOFAR grams as

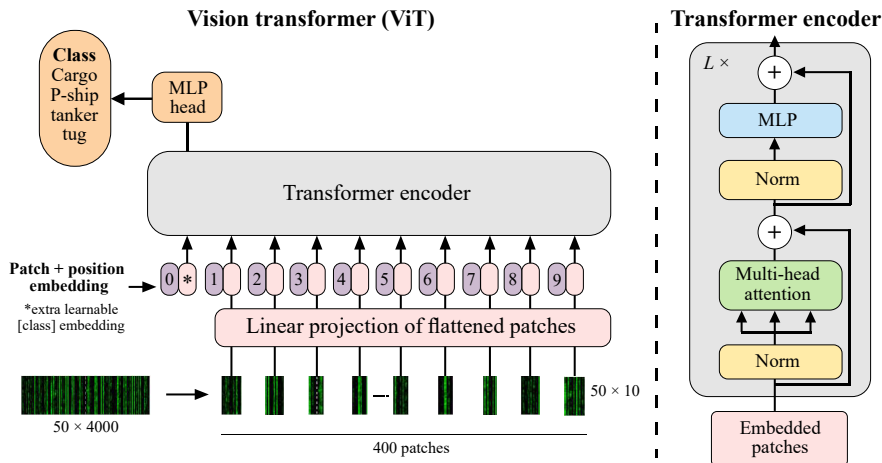


Fig. 1. Modified ViT network architecture with custom size LOFAR gram images (DOSOVITSKIY *et al.*, 2020).

input and is divided into 400 patches of size 50×10 to ensure that each patch spans the full temporal resolution while maintaining fine spectral resolution, allowing the model to preserve tonal shifts caused by Doppler effects or environmental variability within a single patch. The ViT model was trained using stochastic gradient descent with momentum as the optimization method, with an initial learning rate set to 0.001. Training was conducted over a maximum of 10 epochs with a mini-batch size of 64, utilizing a GPU to accelerate the process. A modified ViT network model with description and learnable parameters for each layer is presented in Table 1. The training was conducted on a system featuring an AMD Ryzen 5 3600 processor (6-core), 32 GB RAM, 500 GB SSD storage, and an NVIDIA GTX 1660 SUPER GPU with 6 GB of memory.

Table 1. Number of parameters of each layer of the CNN architecture.

Modules	Layers	Number of parameters
Input processing	Image input	–
	Patch embedding	160 400
	Embedding concatenation	400
	Position embedding	200 400
	Layer norm	800
Feature extractor	Self-attention	641 600
	Layer norm	800
	Encoder block 1	480 400
	Encoder block 2	480 400
	Layer norm	800
Classifier	Head	1604
Total parameters		1 967 604

3.3. Data preprocessing

The generation of LOFAR grams involves a systematic analysis of publicly available DeepShip and ShipsEar datasets. The process starts with loading audio files and configuring the key parameters of the CAPSE algorithm. This includes setting a window size of 16 000 samples with a 50% overlap and a sampling rate of 8 kHz.

For the LOFAR gram, the algorithm processes the audio signal in segments, applying a Hanning window to reduce spectral leakage. A real FFT is performed on each window, normalizing the power in each frequency bin.

The first half of the bins is preserved, followed by applying an FFT to each column of the spectrum. After squaring the magnitudes, the maximum value in each column is stored. The resulting spectrum is saved as a row vector in a (50×4000) matrix in logarithmic scale, and the matrix is saved as PNG images. Figure 2 shows samples of zoomed LOFAR gram

(0 Hz–1600 Hz) images generated for different classes in DeepShip dataset (IRFAN *et al.*, 2021). Here, the spectral components due to machinery are visible against normalized broadband noise.

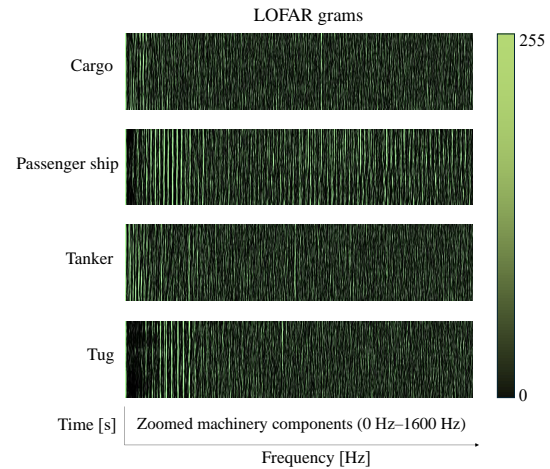


Fig. 2. Sample of LOFAR grams of DeepShip dataset developed using CAPSE algorithm.

Following preprocessing, the dataset was randomly divided into three distinct subsets: 70% of the data was allocated for training, where the model learns patterns and features within the data; 15% was kept for validation, which is used to adjust the weights of the neural networks of the model. Early stopping was used to prevent overfitting by evaluating its performance on unseen data during training; the remaining 15% of the dataset was used for testing the trained model to evaluate its performance metrics.

4. Results and discussion

This section presents and analyzes the results of our proposed method for underwater acoustic target classification. The performance of the model is evaluated on a benchmark dataset, with a focus on classification accuracy, and the advantages of CAPSE enhanced spectral representations and ViT. Comparative results with existing methods are also discussed. Figures 3 and 4 show the accuracy and loss curves for the

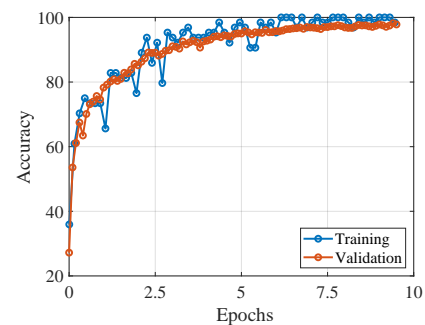


Fig. 3. Accuracy curves for training and validation on DeepShip dataset.

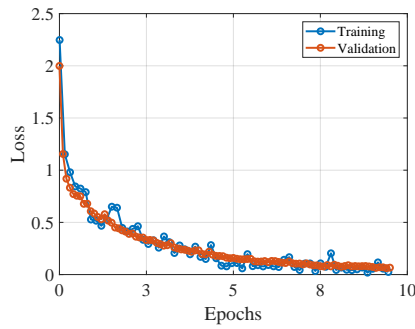


Fig. 4. Loss curves for training and validation on DeepShip dataset.

training and validation process on DeepShip dataset. The network shows rapid accuracy improvement and convergence with minimal overfitting. The loss steadily decreases, indicating stable and effective training.

4.1. Classification performance

Tables 2 and 3 provide an overview of DeepShip and ShipsEar datasets used for evaluating the classification performance, showing the number of samples for each vessel class. Despite the variation in sample sizes, the model demonstrated effective generalization across all classes, maintaining high performance even for classes with fewer samples, such as the tug class in DeepShip or class B from ShipsEar, shown in the confusion matrix in Figs. 5a and 5b.

Table 2. Class description of DeepShip dataset (IRFAN *et al.*, 2021).

Class label	Number of samples
Cargo	4242
Passenger ship	4641
Tanker	4454
Tug	4054

Table 3. Class description of ShipsEar dataset (SANTOS-DOMÍNGUEZ *et al.*, 2016).

Class label	Vessel type	Number of samples
Class A	Mussel boats, dredgers, fishing boats, trawlers, and tugboats	389
Class B	Sailboats, motorboats, and pilot boats	313
Class C	Passenger ferries	842
Class D	Ro-ro vessels and ocean liners	492
Class E	Background noise recordings	229

In terms of classification performance on DeepShip dataset in Fig. 5a, as presented in Table 4, the model achieved excellent results across all vessel types. The cargo class achieved the highest accuracy at 98.90%,

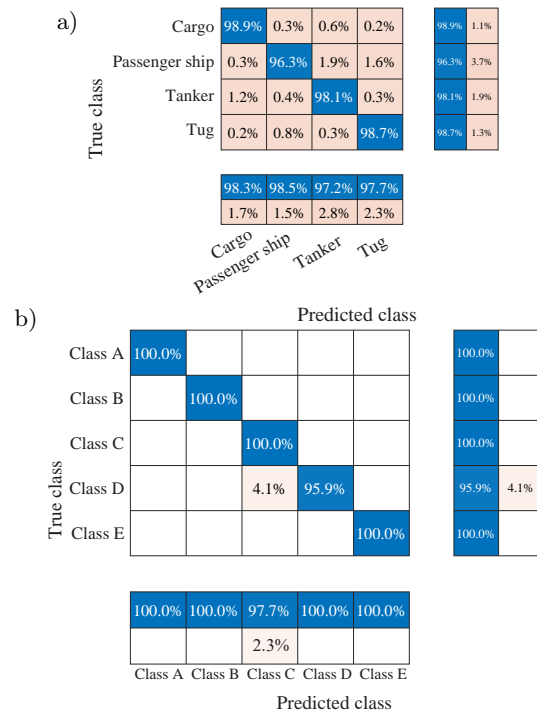


Fig. 5. Confusion matrix for modified ViT network: a) DeepShip dataset; b) ShipsEar dataset.

Table 4. Classification performance on DeepShip dataset.

Label	Accuracy [%]	Precision [%]	Recall [%]	<i>F1</i> -score [%]
Cargo	98.90	98.36	98.90	98.63
Passenger ship	96.26	98.38	96.26	97.31
Tanker	98.05	97.20	98.05	97.62
Tug	98.68	97.98	98.68	98.33
Average	97.98	97.98	97.98	97.97
Std Dev	1.1971	0.5516	1.1971	0.6119

with the tug class following at 98.68%. The passenger ship and tanker classes show slightly lower accuracies of 96.26% and 98.05%, respectively. These small differences indicate that the model is consistent in identifying all vessel types, regardless of their sample size.

The precision, recall, and *F1*-score metrics further demonstrate the model robustness. The *F1*-scores, which balance precision and recall, are consistently high for all classes, ranging from 97.31% for the passenger ship class to 98.63% for the cargo class, highlighting the model ability to maintain high classification performance across diverse acoustic characteristics. The model average *F1*-score of 97.97% across all classes reflects its ability to generalize well to unseen data, making it a reliable tool for underwater acoustic target classification tasks. Although there are some minor performance variations, particularly for the passenger ship class, the overall results confirm the model's effective classification capability, demonstrating robust generalization across all vessel types.

Similarly, the confusion matrix for the ShipsEar dataset in Fig. 5b, shows high classification performance across all vessel classes as shown in Table 5, with an accuracy of (100 %) achieved for classes A, B, and E, indicating that the model can reliably distinguish musel boats, sailboats, and background noise recordings. Class C (passenger ferries) and class D (Ro-ro vessels and ocean liners) exhibit minor confusion, with 2.3% of class C misclassified as class D and 4.1% of class D misclassified as class C. This overlap suggests that these vessel types share similar acoustic characteristics, likely due to comparable propulsion systems or operational behaviours. However, the model maintains over 95% accuracy for all classes, demonstrating strong generalization.

Table 5. Classification performance on ShipsEar dataset.

Label	Accuracy [%]	Precision [%]	Recall [%]	F1-score [%]
Class A	100	100	100	100
Class B	100	100	100	100
Class C	100	96.10	100	98.01
Class D	95.95	100	95.95	97.93
Class E	100	100	100	100
Average	99.19	99.22	99.19	99.19
Std Dev	1.8130	1.7424	1.8130	1.1111

4.2. Features visualization

The t-SNE method is employed to visually analyze the model feature extraction process. High-dimensional features of vessel radiated noise data are projected into a two-dimensional space to observe how well the model separates different vessel classes. The first visualization in Fig. 6 shows the t-SNE plot for the input layer before the ViT model initialization, where the samples are scattered with no clear patterns or groupings. After the model processes the data, the second visualization shown in Fig. 7 presents a clear separation of classes, with most samples correctly clustered into distinct groups according to their labels. Very few instances remain misclassified, potentially

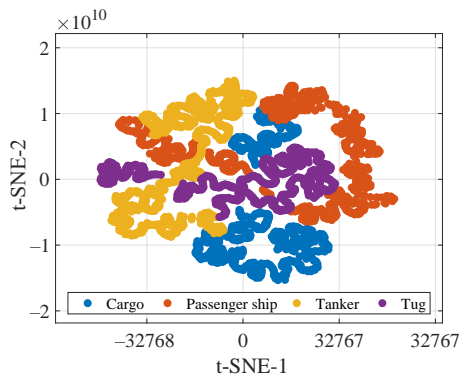


Fig. 6. t-SNE high dimensional features visualization of untrained network.

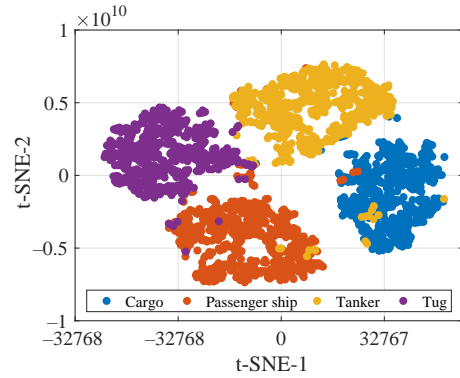


Fig. 7. t-SNE high dimensional visualization after the network is fully trained.

due to a weak or ambiguous target signature. By highlighting these outliers, the model limitations and areas for improvement become apparent. Overall, this visualization confirms that the model effectively learns discriminative features, resulting in well-formed class clusters in the feature space.

4.3. Performance under varying SNR

In real-world maritime scenarios, the underwater acoustic environment is subject to varying levels of environmental noise originating from natural sources such as wind, wave activity, and marine life. Such noise can complicate the accurate classification and interpretation of acoustic signals, highlighting the need to evaluate the performance of classification models under adverse conditions. To systematically examine the robustness and generalization abilities of the classifiers, this study simulated diverse noise environments by injecting Gaussian white noise at multiple SNR levels into the original acoustic signals. Using DeepShip dataset, the power of the signal was computed for each case, and zero-mean noise with a specified power level was generated and added to achieve a targeted SNR value. The objective was to mimic real-world situations where the clarity of received signals is degraded by external noise sources.

The performance of the model was assessed, as depicted in Fig. 8. As the SNR decreases from 20 dB to 0 dB, the overall accuracy declines, demonstrating the negative impact of increasing noise on the model's performance. Notably, the average classification accuracy remains above 50% at 5 dB, indicating a moderate level of robustness in noisy conditions. Among the vessel classes, the cargo class consistently achieves higher accuracy, which can be attributed to its stronger and more distinguishable acoustic signature. In contrast, the model exhibits reduced performance for certain classes such as tankers and passenger ships, whose acoustic characteristics are more susceptible to noise interference. To address this limitation, future work will focus on exploring alternative classifier configura-

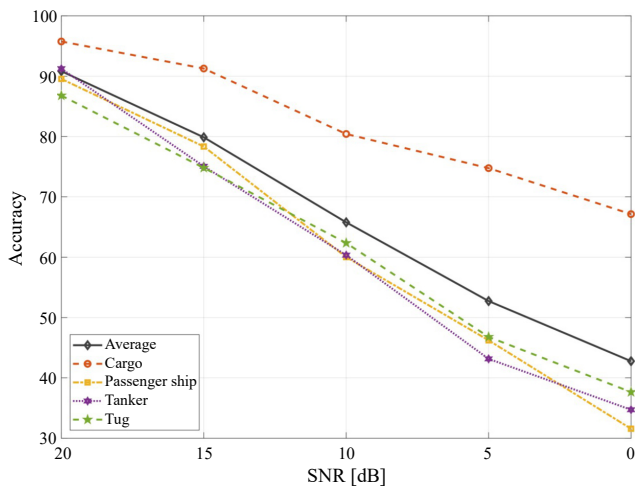


Fig. 8. Performance of classifier under varying SNR conditions.

tions to further enhance robustness and improve class-specific accuracy under challenging acoustic environments.

4.4. Comparison with earlier research

Most of the prior vessel classification research using deep learning techniques commonly employs CNN models, with a majority of these studies relying on non-standard datasets to evaluate the classification performance of their proposed methods. This reliance on diverse datasets makes it challenging to consistently compare the performance of different models.

Table 6 compares the classification accuracy of our model with earlier deep learning-based studies that utilize both the DeepShip and ShipsEar datasets. The comparative analysis demonstrates that our proposed model achieves an accuracy of 97.98% on the DeepShip dataset, which is competitive with the state-of-the-art. It surpasses models such as UATR-transformer, which achieved 95.30%, and significantly outperforms DRA-CNN, which lagged at 89.20%. On the ShipsEar dataset, our model achieves an impressive accuracy of 99.19%, further solidifying its competitiveness. Although the HAUT Fusion model slightly outperforms our method with an accuracy of 99.01% on the DeepShip dataset and 98.62% on the ShipsEar dataset, it does so at a considerable computational cost. The HAUT Fusion model utilizes 30.33 million

parameters, compared to the 1.90 million parameters used by our classifier model. This highlights that our model had a better balance between accuracy and computational efficiency, making it a favorable choice for practical applications.

Moreover, the ResNet18 model, despite achieving commendable accuracies of 96.37% on the DeepShip dataset and 94.30% on the ShipsEar dataset, operates with a significantly larger parameter count of 11.70 million. In contrast, our model maintains high performance while using only 1.90 million parameters, underscoring its efficiency in terms of model complexity and memory requirements. Similarly, DRA-CNN, while achieving better accuracies of 97.10% on the ShipsEar dataset, uses only 0.26 million parameters, but its performance on the DeepShip dataset is considerably lower (89.20%).

These findings suggest that the proposed model provides an effective solution with lower computational demand, making it suitable for deployment in environments with limited resources without sacrificing accuracy. The consistent high performance across both datasets (DeepShip and ShipsEar) further validates the generalizability of our approach.

5. Conclusions and future work

A new framework leveraging deep learning, based on CAPSE as preprocessing and ViT as classifier is designed to enhance the performance of classification of marine vessel based on their radiated noise. The model demonstrates a robust performance, achieving an accuracy of 97.98% while maintaining a significantly lower parameter of 1.9 million compared to other state-of-the-art models. The results highlight the model's efficiency in extracting discriminative features with minimal computational complexity, making it suitable for real-time or resource-constrained environments. Despite marginally lower accuracy compared to HAUT Fusion, our model's efficiency in terms of parameter usage offers a compelling advantage. These findings emphasize the effectiveness of the proposed approach in balancing accuracy and computational cost for passive underwater acoustic target classification tasks.

Future work may explore further optimization of the feature extraction process and the potential integration of additional domain-specific knowledge

Table 6. Classification performance on ShipsEar dataset.

Models	Accuracy [%] DeepShip	Accuracy [%] ShipsEar	Parameters (million)
ResNet18 (HONG <i>et al.</i> , 2021)	96.37	94.30	11.70
DRA-CNN (CHEN <i>et al.</i> , 2021)	89.20	97.10	0.26
UATR-transformer (FENG, ZHU, 2022)	95.30	96.90	2.60
HAUT Fusion (CHEN <i>et al.</i> , 2024)	99.01	98.62	30.33
Proposed method	97.98	99.19	1.90

to enhance performance. Furthermore, due to the limited availability of publicly available datasets and the inadequate class of vessel types that are recorded, it is difficult to fully assess the robustness of the model under various environmental scenarios. To address this issue, we plan to further investigate vessel radiated noise and synthetically generate signals for different scenarios using mathematical modelling.

FUNDINGS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF 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

The data that support the findings of this study are available from the corresponding author upon reasonable request.

AUTHORS' CONTRIBUTION

Study conception and design: Najamuddin, Usman Ullah Sheikh, and Ahmad Zuri Sha'ameri. Data collection, analysis and interpretation of results, and the draft manuscript preparation: Najamuddin. All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors express their gratitude to Universiti Teknologi Malaysia (UTM) for their invaluable support and resources that have facilitated this research.

References

- ASLAM M.A. *et al.* (2024), Underwater sound classification using learning based methods: A review, *Expert Systems with Applications*, **255**(Part 1): 124498, <https://doi.org/10.1016/j.eswa.2024.124498>.
- BIANCO M.J. *et al.* (2019), Machine learning in acoustics: Theory and applications, *The Journal of the Acoustical Society of America*, **146**(5): 3590–3628. <https://doi.org/10.1121/1.5133944>.
- BJØRNØ L. (2017), Underwater acoustic measurements and their applications, [in:] *Applied Underwater Acoustics*, Neighbors T.H., III, Bradley D. [Eds.], pp. 889–947, Elsevier, <https://doi.org/10.1016/B978-0-12-811240-3.00014-X>.
- CAO X., TOGNERI R., ZHANG X., YU Y. (2019), Convolutional neural network with second-order pooling for underwater target classification, *IEEE Sensors Journal*, **19**(8): 3058–3066, <https://doi.org/10.1109/JSEN.2018.2886368>.
- CHEN J., HAN B., MA X., ZHANG J. (2021), Underwater target recognition based on multi-decision LOFAR spectrum enhancement: A deep-learning approach, *Future Internet*, **13**(10): 265, <https://doi.org/10.3390/fi13100265>.
- CHEN L., LUO X., ZHOU H. (2024), A ship-radiated noise classification method based on domain knowledge embedding and attention mechanism, *Engineering Applications of Artificial Intelligence*, **127**(Part B): 107320, <https://doi.org/10.1016/j.engappai.2023.107320>.
- CINELLI L.P., CHAVES G.S., LIMA M.V.S. (2018), Vessel classification through convolutional neural networks using passive sonar spectrogram images, [in:] *Proceedings of the Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2018)*, pp. 21–25, <http://doi.org/10.14209/sbrt.2018.340>.
- DE CARVALHO H.T., AVILA F.R., BISCAINHO L.W.P. (2021), Bayesian restoration of audio degraded by low-frequency pulses modeled via Gaussian process, *IEEE Journal of Selected Topics in Signal Processing*, **15**(1): 90–103, <https://doi.org/10.1109/JSTSP.2020.3033410>.
- DE MOURA N.N., DE SEIXAS J.M. (2016), Novelty detection in passive SONAR systems using support vector machines, *2015 Latin-America Congress on Computational Intelligence (LA-CCI)*, <https://doi.org/10.1109/LA-CCI.2015.7435957>.
- DOMINGOS L.C.F., SANTOS P.E., SKELTON P.S.M., BRINKWORTH R.S.A., SAMMUT K. (2022), A survey of underwater acoustic data classification methods using deep learning for shoreline surveillance, *Sensors*, **22**(6): 2181, <https://doi.org/10.3390/s22062181>.
- DOSOVITSKIY A. *et al.* (2020). An image is worth 16x16 words: Transformers for image recognition at scale, arXiv, <https://doi.org/10.48550/arXiv.2010.11929>.
- FENG S., JIANG K., KONG X. (2021), A line spectrum detector based on improved coherent power spectrum estimation, *Journal of Physics: Conference Series*, **1971**(1): 012006, <https://doi.org/10.1088/1742-6596/1971/1/012006>.
- FENG S., ZHU X. (2022), A transformer-based deep learning network for underwater acoustic target recognition, *IEEE Geoscience and Remote Sensing Letters*, **19**: 1–5, <https://doi.org/10.1109/LGRS.2022.3201396>.
- HEGAZY A.E., MAKHLOUF M.A., EL-TAWEL G.S. (2020), Improved salp swarm algorithm for feature selection, *Journal of King Saud University – Computer and Information Sciences*, **32**(3): 335–344, <https://doi.org/10.1016/j.jksuci.2018.06.003>.
- HONG F., LIU C., GUO L., CHEN F., FENG H. (2021), Underwater acoustic target recognition with ResNet18

- on shipsear dataset, *2021 IEEE 4th International Conference on Electronics Technology (ICET)*, pp. 1240–1244, <https://doi.org/10.1109/ICET51757.2021.9451099>.
16. HU G., WANG K., LIU L. (2021), Underwater acoustic target recognition based on depthwise separable convolution neural networks, *Sensors*, **21**(4): 1429, <https://doi.org/10.3390/s21041429>.
 17. IKPEKHA O.W., ELTAYEB A., PANDYA A., DANIELS S. (2018), Operational noise associated with underwater sound emitting vessels and potential effect of oceanographic conditions: A Dublin Bay port area study, *Journal of Marine Science and Technology*, **23**: 228–235, <https://doi.org/10.1007/s00773-017-0468-4>.
 18. IRFAN M., JIANGBIN Z., ALI S., IQBAL M., MASOOD Z., HAMID U. (2021), DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification, *Expert Systems with Applications*, **183**: 115270, <https://doi.org/10.1016/j.eswa.2021.115270>.
 19. KHISHE M., MOHAMMADI H. (2019), Passive sonar target classification using multi-layer perceptron trained by salp swarm algorithm, *Ocean Engineering*, **181**: 98–108, <https://doi.org/10.1016/j.oceaneng.2019.04.013>.
 20. KIM K.-I., PAK M.-I., CHON B.-P., RI C.-H. (2021), A method for underwater acoustic signal classification using convolutional neural network combined with discrete wavelet transform, *International Journal of Wavelets, Multiresolution and Information Processing*, **19**(04): 2050092, <https://doi.org/10.1142/S0219691320500927>.
 21. LAMPERT T.A., O'KEEFE S.E.M. (2013), On the detection of tracks in spectrogram images, *Pattern Recognition*, **46**(5): 1396–1408, <https://doi.org/10.1016/j.patcog.2012.11.009>.
 22. LAN H., WHITE P.R., LI N., LI J., SUN D. (2020), Coherently averaged power spectral estimate for signal detection, *Signal Processing*, **169**: 107414, <https://doi.org/10.1016/j.sigpro.2019.107414>.
 23. LI X., WANG D., TIAN Y., KONG X. (2023), A method for extracting interference striations in lofargram based on decomposition and clustering, *IET Image Processing*, **17**(6): 1951–1958, <https://doi.org/10.1049/ipr2.12768>.
 24. LIM T., BAE K., HWANG C., LEE H. (2007), Classification of underwater transient signals using MFCC feature vector, *2007 9th International Symposium on Signal Processing and Its Applications, ISSPA 2007, Proceedings*, pp. 1–4, <https://doi.org/10.1109/ISSPA.2007.4555521>.
 25. LUO X., CHEN L., ZHOU H., CAO H. (2023), A survey of underwater acoustic target recognition methods based on machine learning, *Journal of Marine Science and Engineering*, **11**(2): 384, <https://doi.org/10.3390/jmse11020384>.
 26. LUO X., ZHANG M., LIU T., HUANG M., XU X. (2021), An underwater acoustic target recognition method based on spectrograms with different resolutions, *Journal of Marine Science and Engineering*, **9**(11): 1246, <https://doi.org/10.3390/jmse9111246>.
 27. MCKENNA M.F. *et al.* (2024), Understanding vessel noise across a network of marine protected areas, *Environmental Monitoring and Assessment*, **196**(4): 369, <https://doi.org/10.1007/s10661-024-12497-2>.
 28. MÜLLER N., REERMANN J., MEISEN T. (2024), Navigating the depths: A comprehensive survey of deep learning for passive underwater, *IEEE Access*, **12**: 154092–154118, <https://doi.org/10.1109/ACCESS.2024.3480788>.
 29. NOUMIDA A., RAJAN R. (2022), Multi-label bird species classification from audio recordings using attention framework, *Applied Acoustics*, **197**: 108901, <https://doi.org/10.1016/j.apacoust.2022.108901>.
 30. PANG D., WANG H., MA J., LIANG D. (2023), DCTN: A dense parallel network combining CNN and transformer for identifying plant disease in field, *Soft Computing*, **27**(21): 15549–15561, <https://doi.org/10.1007/s00500-023-09071-2>.
 31. PARK J., JUNG D.-J. (2021), Deep convolutional neural network architectures for tonal frequency identification in a lofargram, *International Journal of Control, Automation and Systems*, **19**(2): 1103–1112, <https://doi.org/10.1007/s12555-019-1014-4>.
 32. RAFFEL C. *et al.* (2020), Exploring the limits of transfer learning with a unified text-to-text transformer, *Journal of Machine Learning Research*, **21**(140): 1–67.
 33. SANTOS-DOMÍNGUEZ D., TORRES-GUIJARRO S., CARDENAL-LÓPEZ A., PENA-GIMENEZ A. (2016), ShipsEar: An underwater vessel noise database, *Applied Acoustics*, **113**: 64–69, <https://doi.org/10.1016/j.apacoust.2016.06.008>.
 34. SHARMA G., UMAPATHY K., KRISHNAN S. (2020), Trends in audio signal feature extraction methods, *Applied Acoustics*, **158**: 107020, <https://doi.org/10.1016/j.apacoust.2019.107020>.
 35. SHERIN B.M., SUPRIYA M.H. (2015), Selection and parameter optimization of SVM kernel function for underwater target classification, [in:] *2015 IEEE Underwater Technology (UT)*, pp. 1–5, <https://doi.org/10.1109/UT.2015.7108260>.
 36. SIDDAGANGAIAH S., LI Y., GUO X., CHEN X., ZHANG Q., YANG K., YANG Y. (2016), A complexity-based approach for the detection of weak signals in ocean ambient noise, *Entropy*, **18**(3): 101, <https://doi.org/10.3390/e18030101>.
 37. SINGH P., SAHA G., SAHIDULLAH M. (2021), Non-linear frequency warping using constant-Q transformation for speech emotion recognition, [in:] *2021 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–6, <https://doi.org/10.1109/ICCCI50826.2021.9402569>.

38. SONG G., GUO X., WANG W., REN Q., LI J., MA L. (2021), A machine learning-based underwater noise classification method, *Applied Acoustics*, **184**: 108333, <https://doi.org/10.1016/j.apacoust.2021.108333>.
39. THOMAS M., MARTIN B., KOWARSKI K., GAUDET B., MATWIN S. (2020), Marine mammal species classification using convolutional neural networks and a novel acoustic representation, [in:] *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2019. Lecture Notes in Computer Science*, **11908**: 290–305, https://doi.org/10.1007/978-3-030-46133-1_18.
40. YANG Y., YAO Q., WANG Y. (2024), Underwater acoustic target recognition method based on feature fusion and residual CNN, *IEEE Sensors Journal*, **24**(22): 37342–37357, <https://doi.org/10.1109/JSEN.2024.3464754>.
41. YUAN F., KE X., CHENG E. (2019), Joint representation and recognition for ship-radiated noise based on multimodal deep learning, *Journal of Marine Science and Engineering*, **7**(11): 380, <https://doi.org/10.3390/jmse7110380>.
42. ZENG Y., ZHANG M., HAN F., GONG Y., ZHANG J. (2019), Spectrum analysis and convolutional neural network for automatic modulation recognition, [in:] *IEEE Wireless Communications Letters*, **8**(3): 929–932, <https://doi.org/10.1109/LWC.2019.2900247>.

Research Paper

Experimental Results of Diver Detection in Harbor Environments Using Single Acoustic Vector Sensor

Saier MAHMOUD*, Louay SALEH, Ibrahim CHOUAIB

*Department of Electronic and Mechanical Systems, Higher Institute for Applied Sciences and Technology
Damascus, Syria**Corresponding Author e-mail: saeir.mahmoud@hiast.edu.sy*Received February 5, 2025; revised March 30, 2025; accepted March 31, 2025;
published online May 29, 2025.*

This paper addresses the detection of divers with an open-circuit scuba. An acoustic vector sensor (AVS), which contains four channels, one for the pressure component, and three for orthogonal particle velocity components is proposed to be used. A novel covariance matrix analysis (CMA) method is presented for estimating the signal power using AVS in three-dimensional measurements. This method is based on solving a quartic equation that relates the determinant and trace of the AVS covariance matrix to the reciprocal of the signal-to-noise ratio (SNR) in a three-dimensional isotropic acoustic field with spherical isotropic noise. This method is compared with two traditional methods: the AVS pressure channel power, and the minimum variance distortionless response (MVDR) beamformer, in estimating the acoustic power associated with the diver's breathing. Experimental data from sea trials demonstrate the capability of all three methods to reconstruct the waveform of the acoustic diver signal and highlight the periodic breathing patterns. The diver's breathing rate and corresponding power are estimated using the fast Fourier transform (FFT) of the power signal, therefore serving as a key signature for diver detection. The experiment demonstrates that the CMA method gives better diver detection index compared to the other two methods.

Keywords: scuba diver; acoustic vector sensor (AVS); detection; minimum variance distortionless response (MVDR); signal-to-noise ratio (SNR); covariance matrix analysis (CMA); fast Fourier transform (FFT).



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

An acoustic signal is generated when the diver breathes through an apparatus regulator, with air bubbles discharged from it. This signal is a broad-spectrum (JOHANSSON *et al.*, 2010) ranging from hundreds of hertz to 75 kHz (TU *et al.*, 2020), quasi-periodic (GOROVOY *et al.*, 2015), with a periodicity ranging from 2.44 s to 7.09 s, and exhibits a repetition pattern corresponding to the diver's breathing rate, which typically falls within the range of 0.14 Hz to 0.41 Hz (DONSKOY *et al.*, 2008). This range variability is influenced by several factors such as the diver's age, experience, activity level, and scuba equipment used (DONSKOY *et al.*, 2008). Passive sonar systems utilize this periodic signal to detect the presence of a diver. The respiratory cycle of a diver initiates with the inhalation phase, which corresponds to sound frequencies exceeding 2 kHz, followed by the exhalation

phase, which corresponds to frequencies below 2 kHz (TU *et al.*, 2020). Both signals are useful in detection systems (HARI *et al.*, 2015). The specific frequency band of interest for analysis varies among scientific papers, with some focusing on the band with a high-frequency band above 2 kHz (TU *et al.*, 2020; JIN, XU, 2020; LENNARTSSON *et al.* 2009; LI *et al.*, 2015), others focus on the low-frequency band (KORENBAUM *et al.*, 2016; GOROVOY *et al.*, 2014). The detectability of a diver can be determined by two distinctive indicators: the power of the frequency band and its repetition rate (KORENBAUM *et al.*, 2020), which apply to the two types of scuba, open circuit and closed circuit, while the first emits more acoustic noise (DONSKOY *et al.*, 2008), the detection of a diver with a closed circuit is still challenging.

As the acoustic signal propagates away from the source, two distinct fields are produced: pressure and particle motion (PM). Pressure is a scalar quantity

that can be measured using a hydrophone, whereas PM is a vector quantity that is oriented parallel to the direction of wave propagation in the free far field. PM can be measured directly using inertial sensors such as accelerometers or geophones, or indirectly through a configuration of nearby pressure sensors. In last case, the differential measurements approximated the pressure gradient, providing an estimate of PM acceleration (GRAY *et al.*, 2016; NEDELEC *et al.*, 2021). The direct measurement approach involves challenges related to buoyancy, compliance, suspension, geometry, and flow. In the contrast, the indirect measurement approach faces many issues such as spacing, calibration uncertainty, noise, and flow (GRAY *et al.*, 2016). Integrating the PM sensor with an omnidirectional pressure sensor into a single unit results in what is known as an acoustic vector sensor (AVS) (ROH *et al.*, 2022), which has recently gained attention and has increased usability (YUAN *et al.*, 2022; DONG *et al.*, 2024). This sensor has a wide range of applications in both terrestrial and aquatic environments. In terrestrial environment, it is used for localization, tracking, and speech enhancement (CHEN *et al.*, 2018; CAO *et al.*, 2017), while in aquatic environment – the focus of this paper – it is employed for detection (YUAN *et al.*, 2022), localization (CHEN *et al.*, 2023), and tracking (NAGANANDA, ANAND, 2017).

Detection using a single AVS still requires further research efforts (YUAN *et al.*, 2022). The energy-flux detector is proved to perform as a maximum likelihood ratio detector under isotropic noise conditions (SUN *et al.*, 2003). For horizontal isotropic noise, YUAN *et al.* (2022) introduced a method to estimate the signal power by analyzing the covariance matrix of the 2D-AVS output. Incorporating this estimation into detection has proven to be more effective than traditional energy detectors under nonstationary ambient noise. Furthermore, an adaptive matched filter is proposed with 2D-AVS for passive broadband source detection, demonstrating superior performance against noise and interference (MA *et al.*, 2019). The signal waveform can be optimally estimated using the minimum variance distortion response (MVDR) beamformer, also known as the Capon beamformer or the minimum power distortion response (MPDR) (VAN TREES, 2002; ZHAO *et al.*, 2018). This technique is particularly valuable for direction of arrival (DOA) estimation (ZHAO *et al.*, 2018), which is a topic of growing interest among researchers. Various algorithms have been explored with a single AVS, including the arctan (BEREKETLI *et al.*, 2015), intensity-based (WANG *et al.*, 2014; NEHORAI, PALDI, 1994), velocity-covariance-based (NEHORAI, PALDI, 1994), beamforming (ZHAO *et al.*, 2018; BERKETLI *et al.*, 2015), maximum likelihood (LEVIN *et al.*, 2012), multiple signal classification MUSIC (ZHAO *et al.*, 2018), and estimation of signal parameters via rotational invariance

techniques Esprit (TICHAUSKY *et al.*, 2001; PAULRAJ *et al.*, 1985).

Multiple algorithms are available for detection the diver's acoustic signal. While all of these algorithms contribute to estimating the energy of the breathing frequency, they vary in the way they reconstruct the waveform of the diver's signal. The envelope spectrum was used within a 30 kHz–35 kHz bandwidth, achieving a detection range of up to 25 m (LENNARTSSON *et al.*, 2009). CHUNG *et al.* (2007) utilized a multiband normalized matched filter but a reference signal is needed. TU *et al.* (2020) employed the envelope spectrum detection method within 13 kHz–18 kHz bandwidth and extended the range of detection from 20 m to 40 m by using an adaptive noise subtraction approach. All the aforementioned studies utilize data from a single hydrophone. Conversely, other studies have employed two hydrophones to reconstruct the waveform of the diver's signal through cross-correlation. This cross-correlation analysis determines the time delay between the signals received by the hydrophones, which is used in DOA estimation (KORENBAUM *et al.*, 2020; SUTIN *et al.*, 2013).

This paper focuses on utilizing an AVS to capture the acoustic signals emitted by a diver with an open-circuit scuba. Equations that relate the determinant and trace of the AVS covariance matrix to the reciprocal of signal-to-noise ratio (SNR) are extracted in a three-dimensions isotropic acoustic field with spherical isotropic noise. Solving these equations results in power signal estimation. We name this technique as covariance matrix analysis (CMA). Additionally, the MVDR beamformer is used to estimate the power signal by optimizing the azimuth and elevation angle values in order to maximize the MVDR spectrum. The presence of a diver is estimated through comparing the breathing frequency estimated power with a predefined threshold which is estimated empirically using a recorded data for ambient noise. The detection algorithm is evaluated using data obtained from sea trials.

The structure of this paper is arranged as follows. Section 2 describes the AVS with its mathematical model and covariance matrix. The signal power estimation using the AVS P -channel is presented in Sec. 3. In Sec. 4, the signal power estimation by maximizing the spectrum of MVDR beamformer is presented. In Sec. 5, the power signal estimation using the CMA method is performed. This method based on analysing the covariance matrix of AVS channels: pressure and velocity, which results in a quartic equation. Solving this equation provides an estimation of the signal power. The effectiveness of this method is evaluated using simulated data. The proposed diver detection approach using AVS is explained in Sec. 6 and evaluated in Sec. 7 using sea trial data. Finally, Sec. 7 concludes the paper by summarizing the key findings and implications derived from the study.

2. Acoustic vector sensor model

The AVS sensor has four output channels, one for pressure signal (P -channel) and three orthogonal components for particle velocity signal (\mathbf{V} -channels). So that, the AVS captures more information about the acoustic field comparing to hydrophones. The relationship between the PM velocity (\mathbf{v}), acceleration (\mathbf{a}), and the pressure (p) given by Euler's equation, is described as

$$\frac{d\mathbf{v}}{dt} = \mathbf{a} = -\frac{\nabla p}{\rho}, \quad (1)$$

where t is the time, ∇ is the gradient operation, and ρ is the water density.

The PM vector, which has information about the direction of the signal, can be measured in two ways: 1) using a geophone to measure the PM velocity (\mathbf{v}), or accelerometer to measure the acceleration (\mathbf{a}), this way is named direct measurement type or inertial type; 2) by estimation of the ∇p using multiple spaced hydrophones, this way is named indirect type or gradient pressure type (GRAY *et al.*, 2016; NEDELEC *et al.*, 2021).

Under plane wave conditions, the relationship between the pressure signal and the PM velocity signal is expressed as (ABRAHAM, 2019)

$$\mathbf{v} = -\frac{p}{\rho c} \mathbf{u}, \quad (2)$$

where ρc represents acoustic impedance, while c denotes the speed of sound in water,

$$\mathbf{u} = [\cos \theta \cos \phi \quad \cos \theta \sin \phi \quad \sin \theta]^T$$

is a unit vector oriented from sensor to source, ϕ is the azimuth angle, and θ is the elevation angle (Fig. 1).

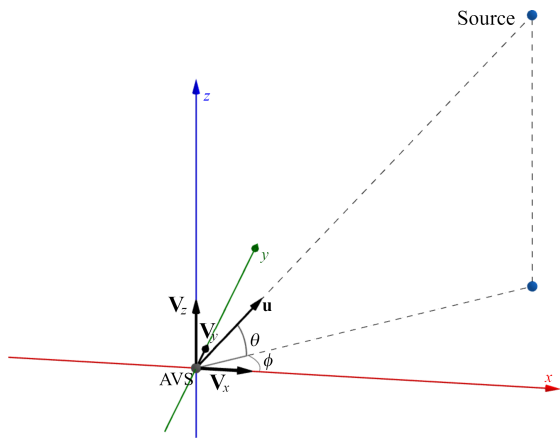


Fig. 1. AVS coordinates.

The AVS output $\mathbf{s}(t)$, after scaling the velocity channels with ρc , is modelled as

$$\begin{aligned} \mathbf{s}(t) &= [p(t) + n_p(t) \quad \mathbf{v}^T(t) + \mathbf{n}_v^T(t)] \\ &= \mathbf{h}(\phi, \theta)p(t) + \mathbf{n}_s(t), \end{aligned} \quad (3)$$

where

$$\mathbf{h}(\phi, \theta) = [1 \quad \mathbf{u}^T]^T, \quad \mathbf{n}_s(t) = [n_p(t) \quad \mathbf{n}_v^T(t)]^T$$

is the ambient-noise, $n_p(t)$ is the pressure noise with power σ_{np}^2 , and $\mathbf{n}_v^T(t)$ is the velocity noise with power σ_{nv}^2 . Under the ambient isotropic noise condition, the relation between σ_{np}^2 and σ_{nv}^2 is given as $\sigma_{np}^2 = 3\sigma_{nv}^2$ (LEVIN *et al.*, 2012).

Assuming $p(t)$ and $\mathbf{n}_s(t)$ are uncorrelated, the covariance matrix of $\mathbf{s}(t)$ is given as

$$\mathbf{R} = \frac{1}{T} \int_T \mathbf{s}(t) * \mathbf{s}^T(t) dt = \mathbf{R}_s + \mathbf{R}_n, \quad (4)$$

where \mathbf{R}_n and \mathbf{R}_s are the covariance matrix of noise and signal. In practice, \mathbf{R} can be estimated from the received signal as following (YUAN *et al.*, 2022; LIU *et al.*, 2019):

$$\widehat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^N \mathbf{s}(n)\mathbf{s}^T(n), \quad (5)$$

where N is the length of snapshot. This matrix is a crucial in array signal processing, as demonstrated in the subsequent sections.

3. Signal power estimation using the AVS P -channel

The AVS contains a pressure sensor that functions as an omnidirectional hydrophone. The sensor's output, $s_p(t)$, is expressed as

$$s_p(t) = p(t) + n_p(t). \quad (6)$$

The power of measurement signal, assuming the noise $n_p(t)$ is uncorrelated with the signal $p(t)$, is defined as

$$\begin{aligned} y^2 &= \frac{1}{T} \int_T s_p^2(t) dt = \frac{1}{T} \int_T p^2(t) dt + \sigma_{np}^2 \\ &= \sigma_s^2 + \sigma_{np}^2 = \sigma_s^2 \left(1 + \frac{1}{\text{SNR}}\right), \end{aligned} \quad (7)$$

where

$$\sigma_s^2 = \int_T p^2(t) dt$$

is the signal power. For a high SNR, the power of pressure signal approaches to signal power.

In practice, this power is estimated by averaging the instantaneous power over a snapshot of length N as

$$y^2 = \frac{1}{N} \sum_{n=1}^N s_p^2(n). \quad (8)$$

This method is used to estimate the power of the acoustic signal emitted by the diver to highlight its periodicity, and is compared to other signal power estimation methods described in the next section.

4. Signal power estimation using MVDR beamformer

Beamforming is a technique employed in array signal processing to steer, shape, and concentrate signals received from sensors toward a desired direction. The shape and width of the beam are adjusted by combining multiple signals through weighting and delays. The AVS features an array of four sensors located at the same point, allowing for the application of beamforming without the need for delay adjustments. The first order beamformer is formed as (LEVIN *et al.*, 2012)

$$y(n) = \mathbf{w}^T \mathbf{s}(n) = \alpha p(n) + (1 - \alpha) \mathbf{u}_{\text{str}}^T \mathbf{v}, \quad (9)$$

where

$$\mathbf{w}^T = [\alpha \quad (1 - \alpha) \mathbf{u}_{\text{str}}^T]^T$$

is the weights with $\alpha \in [0 \ 1]$, and \mathbf{u}_{str} is the look vector.

The parameter α significantly influences the beam pattern, functioning as a monopole for $\alpha = 1$, and a dipole for $\alpha = 0$, while the optimum value of

$$\alpha = \alpha_0 = \frac{\sigma_{nv}^2}{\sigma_{np}^2 + \sigma_{nv}^2}$$

corresponds to the DOA estimator with maximum reliability (LEVIN *et al.*, 2012).

One of the most widely employed beamformers is the MVDR. The objective is to determine the weights \mathbf{w}^T that produces the signal $p(n)$ at the beamformer's output without distortion while minimizing noise power. This problem is mathematically formulated as follows:

$$\begin{aligned} \mathbf{w}_{\text{MVDR}} &= \arg \min_{\mathbf{w}} y^2(n) = \arg \min_{\mathbf{w}} \mathbf{w}^T \mathbf{R} \mathbf{w} \\ &\text{subject to} \quad \mathbf{w}^T \mathbf{h}(\theta, \phi) = 1. \end{aligned} \quad (10)$$

Using the Lagrange multiplier (VAN TREES, 2002), the solution is given as

$$\mathbf{w}_{\text{MVDR}} = \frac{\mathbf{R}^{-1} \mathbf{h}}{\mathbf{h}^T \mathbf{R}^{-1} \mathbf{h}}, \quad (11)$$

the covariance matrix \mathbf{R} should represent the noise covariance matrix in MVDR, and the array measurement covariance matrix in MPDR. When the steering direction aligns with the signal direction, MVDR and MPDR are identical (VAN TREES, 2002).

If the vector \mathbf{h} is known, and the covariance matrix \mathbf{R} is estimated using Eq. (5), the weight vector \mathbf{w}_{MVDR} can be estimated, allowing the estimation of $\hat{p}(n) = y(n) = \mathbf{w}_{\text{MVDR}}^T \mathbf{s}(n)$ in the output of beamformer. And the associated power is:

$$\begin{aligned} y^2(n) &= \mathbf{w}_{\text{MVDR}}^T \mathbf{R} \mathbf{w}_{\text{MVDR}} \\ &= \frac{1}{\mathbf{h}^T(\theta, \phi) \mathbf{R}^{-1} \mathbf{h}(\theta, \phi)} = f(\theta, \phi). \end{aligned} \quad (12)$$

In case where the direction angles are unknown, the MVDR spectrum $f(\theta, \phi)$, which is a measure of power radiating from direction (θ, ϕ) , is computed by changing (θ, ϕ) over all possible values. The solution for the DOA problem is by searching (θ, ϕ) for maxima of $f(\theta, \phi)$. Therefore, the MVDR can be used in addition to DOA estimation, to estimate the max power radiating from this estimated direction. The estimated power can be applied to reconstruct the waveform of AVS signals. Figure 2 presents an example of the spectrum of the MVDR for a single source emitting a linear frequency signal within the band [600–1000] Hz, with -3 dB power, from a position $(\theta, \phi) = (45^\circ, 45^\circ)$, and with an SNR = 5 dB for additive Gaussian noise.

For the experimental data, only the estimated signal power is utilized, as this paper focuses on diver detection rather than the diver's DOA.

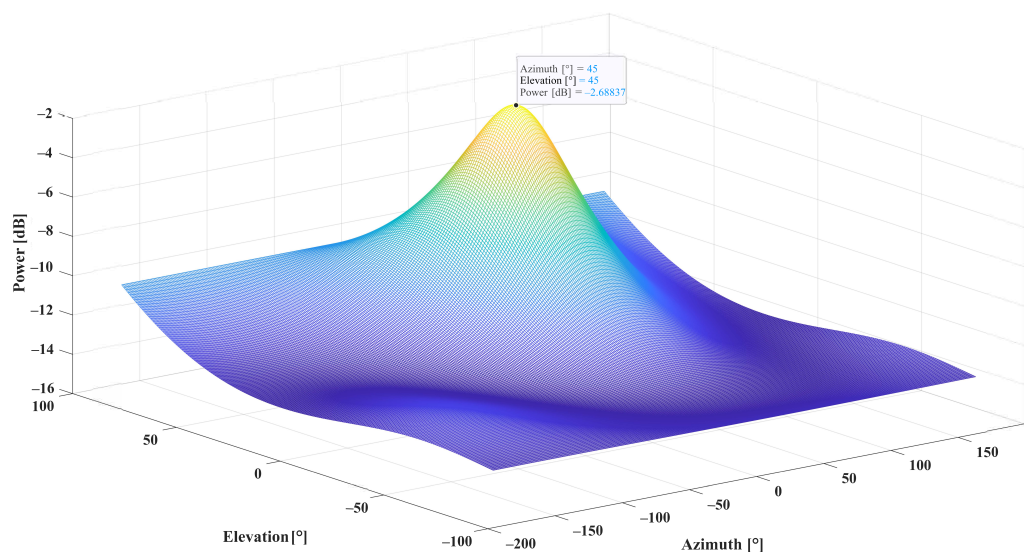


Fig. 2. MDVR spectrum for source radiates signal from $(\theta, \phi) = (45^\circ, 45^\circ)$ angles and SNR = 5 dB.

5. Signal power estimation using the AVS channels through CMA

For 2D-AVS, the signal power was estimated by solving a cubic equation that relates signal and noise power, to the determinant and trace of the AVS covariance matrix. The cubic equation is derived under the assumption of a horizontal isotropic acoustic field (YUAN *et al.*, 2022). In this section, the solution for 3D-AVS with isotropic noise in both horizontal and vertical dimensions is presented.

The covariance matrix of the AVS output is the sum of signal covariance \mathbf{R}_s and the noise covariance \mathbf{R}_n , as shown in the Eq. (4). The matrix \mathbf{R}_s is expressed as follows:

$$\mathbf{R}_s = \mathbf{h}(\theta, \phi) \mathbf{h}^T(\theta, \phi) \sigma_s^2$$

$$= \begin{bmatrix} 1 & \cos \theta \cos \phi & \cos \theta \sin \phi & \sin \theta \\ \cos \theta \cos \phi & \cos^2 \theta \cos^2 \phi & \cos^2 \theta \cos \phi \sin \phi & \cos \theta \sin \theta \cos \phi \\ \cos \theta \sin \phi & \cos^2 \theta \cos \phi \sin \phi & \cos^2 \theta \sin^2 \phi & \cos \theta \sin \theta \sin \phi \\ \sin \theta & \cos \theta \sin \theta \cos \phi & \cos \theta \sin \theta \sin \phi & \sin^2 \theta \end{bmatrix} \sigma_s^2. \quad (13)$$

This matrix has four eigenvalues $\lambda_1 = 2\sigma_s^2$ and $\lambda_2 = \lambda_3 = \lambda_4 = 0$.

The matrix \mathbf{R}_n , under isotropic noise, is expressed as follows:

$$\mathbf{R}_n = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/3 \end{bmatrix} \sigma_{np}^2. \quad (14)$$

As a result, the matrix \mathbf{R} can be expressed as follows:

$$\mathbf{R} = \begin{bmatrix} 1 + \alpha & \cos \theta \cos \phi & \cos \theta \sin \phi & \sin \theta \\ \cos \theta \cos \phi & \cos^2 \theta \cos^2 \phi + \frac{\alpha}{3} & \cos^2 \theta \cos \phi \sin \phi & \cos \theta \sin \theta \cos \phi \\ \cos \theta \sin \phi & \cos^2 \theta \cos \phi \sin \phi & \cos^2 \theta \sin^2 \phi + \frac{\alpha}{3} & \cos \theta \sin \theta \sin \phi \\ \sin \theta & \cos \theta \sin \theta \cos \phi & \cos \theta \sin \theta \sin \phi & \sin^2 \theta + \frac{\alpha}{3} \end{bmatrix} \sigma_s^2, \quad (15)$$

where

$$\alpha = \frac{\sigma_{np}^2}{\sigma_s^2} = \text{SNR}^{-1}.$$

The eigenvalues of \mathbf{R} are

$$\lambda_{1,2} = \frac{1}{3} \left(2\alpha \pm \sqrt{\alpha^2 + 9} + 3 \right) \sigma_s^2 \quad \text{and} \quad \lambda_3 = \lambda_4 = \frac{1}{3} \alpha \sigma_s^2.$$

When there is only a signal $\alpha \rightarrow 0$, the eigenvalues are $\lambda_1 = 2\sigma_s^2$, and $\lambda_2 = \lambda_3 = \lambda_4 = 0$. These correspond to the eigenvalues of \mathbf{R}_s . Conversely, when $\alpha \rightarrow \infty$, indicating the absence of a signal and the presence of noise only, the eigenvalues are given as $\lambda_1 = \sigma_{np}^2$, and $\lambda_2 = \lambda_3 = \lambda_4 = \frac{\sigma_{np}^2}{3}$ that are associated with the eigenvalues of \mathbf{R}_n .

The trace of \mathbf{R} is expressed as

$$\text{trace}(\mathbf{R}) = \sum_{i=1}^4 \lambda_i = 2(\alpha + 1)\sigma_s^2. \quad (16)$$

The determinant of \mathbf{R} is expressed as

$$\det(\mathbf{R}) = \prod_{i=1}^4 \lambda_i = \frac{1}{27} \alpha^3 (\alpha + 4) \sigma_s^8. \quad (17)$$

Substituting σ_s^2 into Eqs. (16) and (17) yields a quartic equation in α , which can be presented as

$$k\alpha^4 + 4k\alpha^3 + 6\alpha^2 + 4\alpha + 1 = 0 \quad (18)$$

where

$$k = 1 - \frac{\text{trace}(\mathbf{R})^4}{432 \det(\mathbf{R})}.$$

This equation has a real positive root corresponding with $k < 0$ as shown in Appendix. By solving Eqs. (16)–(18), the values of α , σ_s^2 , and σ_{np}^2 can be determined. In practice, by estimating the covariance matrix $\hat{\mathbf{R}}$ as indicated in Eq. (5), these parameters can be determined using the following equations:

$$\hat{k} = 1 - \frac{\text{trace}(\hat{\mathbf{R}})^4}{432 \det(\hat{\mathbf{R}})},$$

$$\hat{k} \hat{\alpha}^4 + 4\hat{k} \hat{\alpha}^3 + 6\hat{\alpha}^2 + 4\hat{\alpha} + 1 = 0, \quad (19)$$

$$\hat{\sigma}_s^2 = \frac{\text{trace}(\hat{\mathbf{R}})}{2(\hat{\alpha} + 1)}, \quad \hat{\sigma}_{np}^2 = \alpha \hat{\sigma}_s^2.$$

Figure 2 shows the implementation of this method for a linear frequency signal within the band [600–1000] Hz, with additive Gaussian noise and using 1000 Mont Carlo runs. Figure 3a presents the SNR esti-

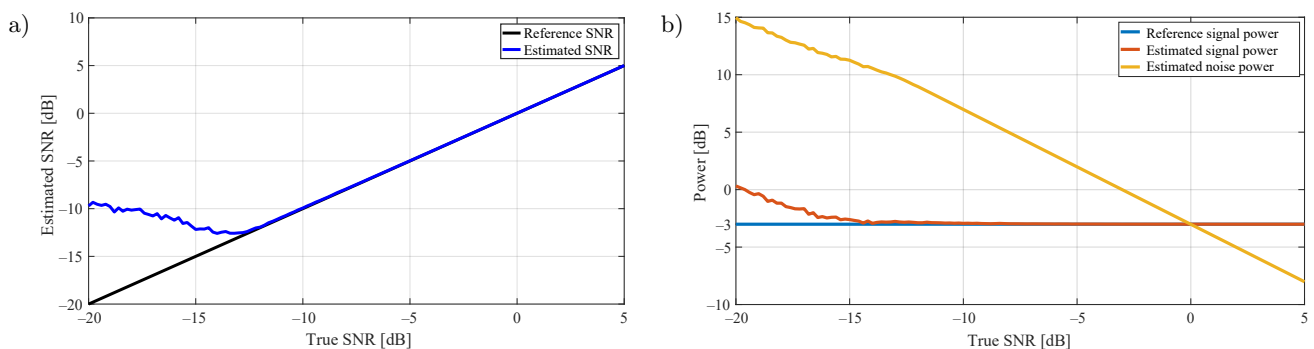


Fig. 3. Estimation using CMA: a) SNR; b) signal and noise power.

mation as the actual SNR varies within the range of $[-20 \ 5]$ dB. The accurate estimation achieved for $\text{SNR} > -12$ dB. While Fig. 3b presents the estimation of noise power and signal power as the SNR changes within the same range. And the true power signal estimation (the true value is -3 dB) is for $\text{SNR} > -12$ dB.

This simulation data demonstrates the technique's capability in estimation the signal power under low SNR. So, this method with the above two method is used to detect the acoustic signal power generated by the diver.

6. Steps for diver detection using AVS

Our method for detecting the diver presence involves analyzing the energy associated with their breathing rate. The process is shown in Fig. 4.

The process has the following steps:

- 1) capture the acoustic signal from the diver using an AVS;
- 2) apply a bandpass filter to increase the SNR;
- 3) reshape the signal to resemble a periodic waveform, employing various techniques including estimated signal power using the AVS pressure channel (P -channel), CMA, and MVDR beamformer;
- 4) compute the power of the diver's acoustic signal power within the frequency band of $0.14 \text{ Hz} - 0.41 \text{ Hz}$, utilizing the fast Fourier transform (FFT);

- 5) compare the computed breath power energy against a predefined threshold.

The proposed methods, AVS pressure channel, CMA, and MVDR, are evaluated and compared in diver detection under sea trials, as discussed in the following section.

7. Experimental results

A passive sonar system (Fig. 5a) was placed 5 m deep on the floor of the marine basin in Tartous harbour. Figure 5 illustrates both the system and the experimental location. The system includes two hydrophones and an AVS. The AVS is of the type VHS-90, with a sensitivity of -180 dB across four channels. The VHS-90 sensor contains three pairs of accelerometers arranged along three orthogonal directions, and six hydrophones connected in parallel to a single output in order to obtain an omnidirectional response. All sensors are encapsulated and coated uniformly with polyurethane material to satisfy the waterproof and sound-permeable requirements.

An expert young diver, with an open-circuit scuba, navigated around the sensor without following a consistent path due to the highly murky water conditions. The movement is roughly drawn in Fig. 6. The diver started his trajectory at point A (18, 5) directed to point B (0, 10). He rested at B for 0.35 min, then pro-

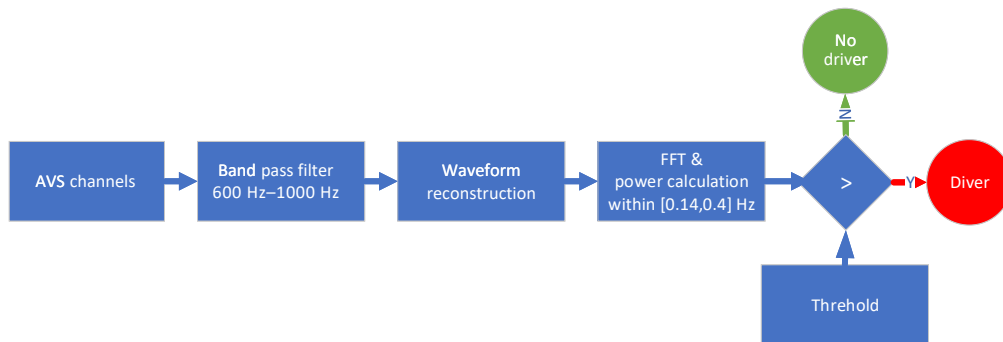


Fig. 4. Diver detection algorithm.

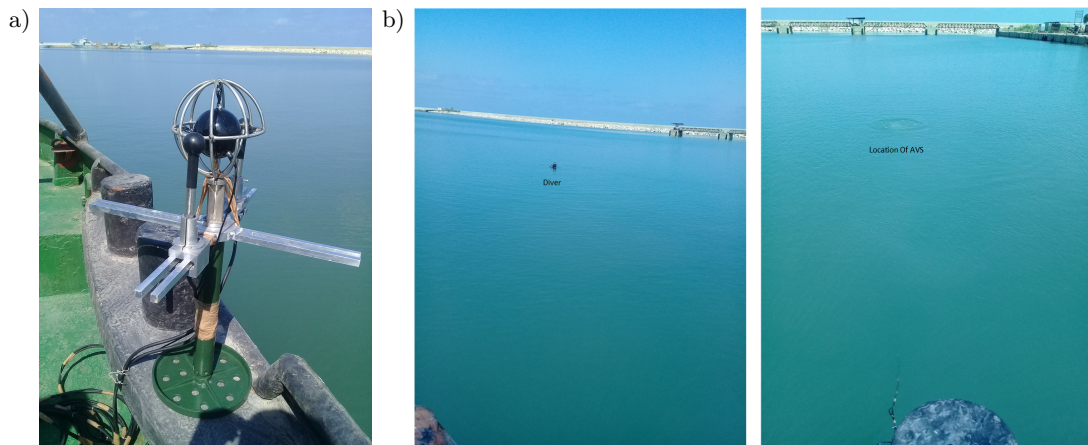


Fig. 5. Our system and experimental location: a) passive sonar system; b) experimental location.

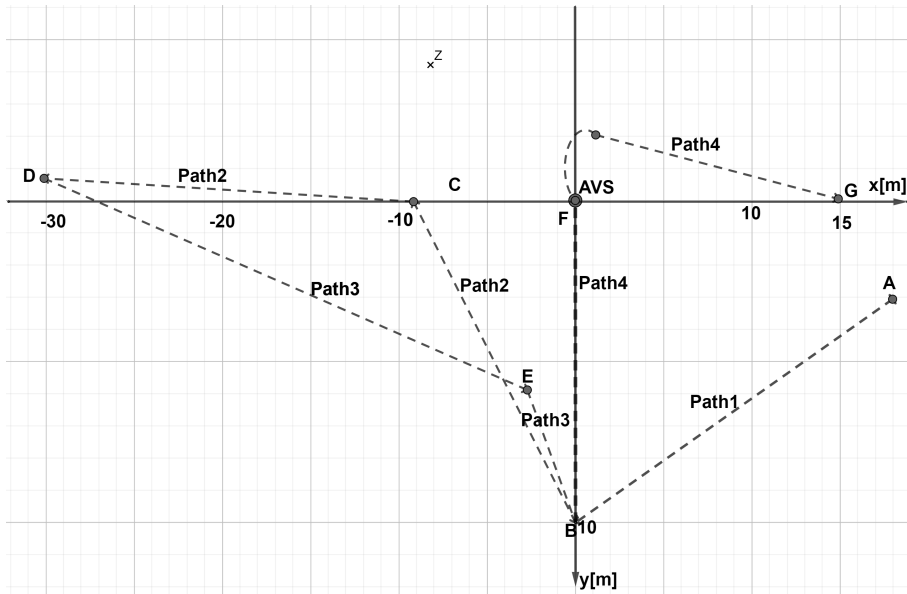


Fig. 6. Approximation of movement of diver around AVS.

ceeded to point C, then D, which is approximately 30 m away from the AVS. The diver has then returned toward point E, before heading back to point B. The diver rested at B 0.5 min, then went toward the AVS, he circled around the sensor and heading back to point G.

The data was captured using an A/D converter at a sampling rate of 44 100 Hz with 24-bit resolution. The MATLAB program was used for processing.

The recorded data was filtered by the 100th-order bandpass FIR filter within the 600 Hz–1000 Hz range in the time domain, corresponding to the exhalation and air bubble signals. Figure 7 illustrates the AVS’s channels over time and their corresponding power (the power is calculated using a sliding window with 200 ms and 50 % overlapping). The figure shows a repeated pattern with period 6.3 s. Each pulse corresponds with one breathing of the diver.

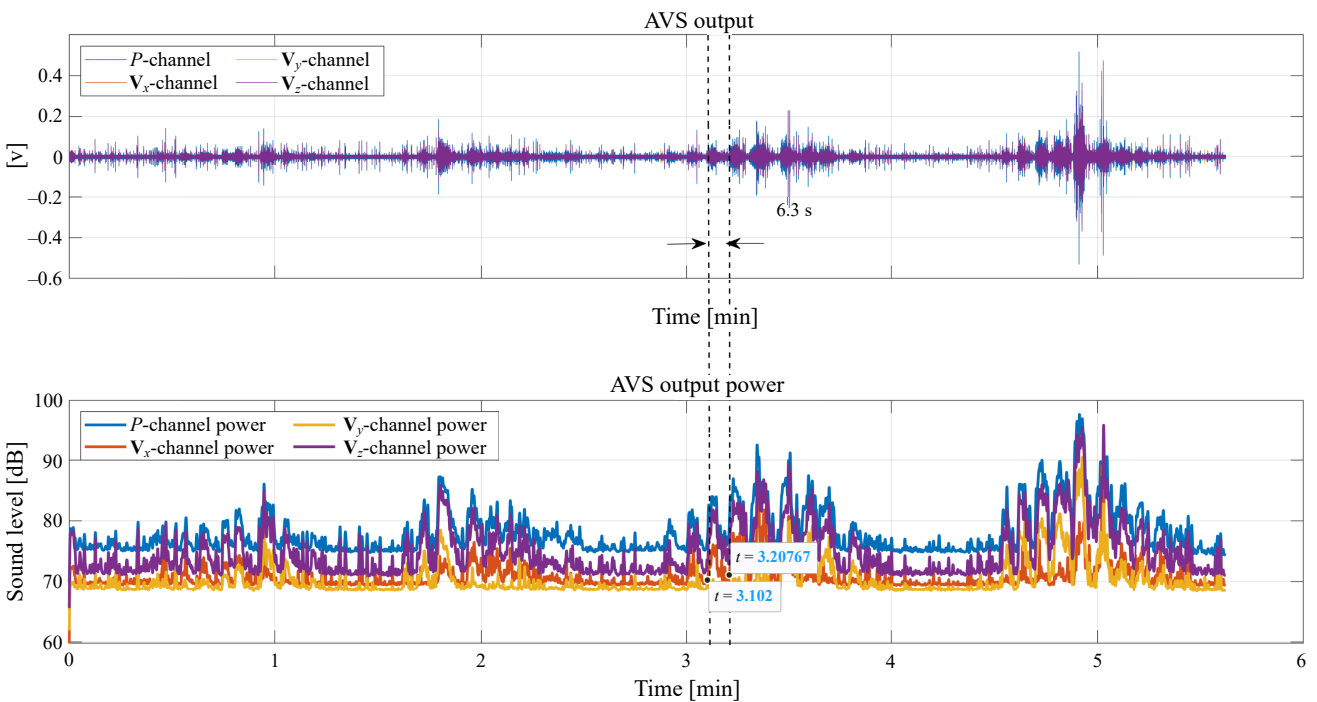


Fig. 7. AVS output signals and corresponding power levels.

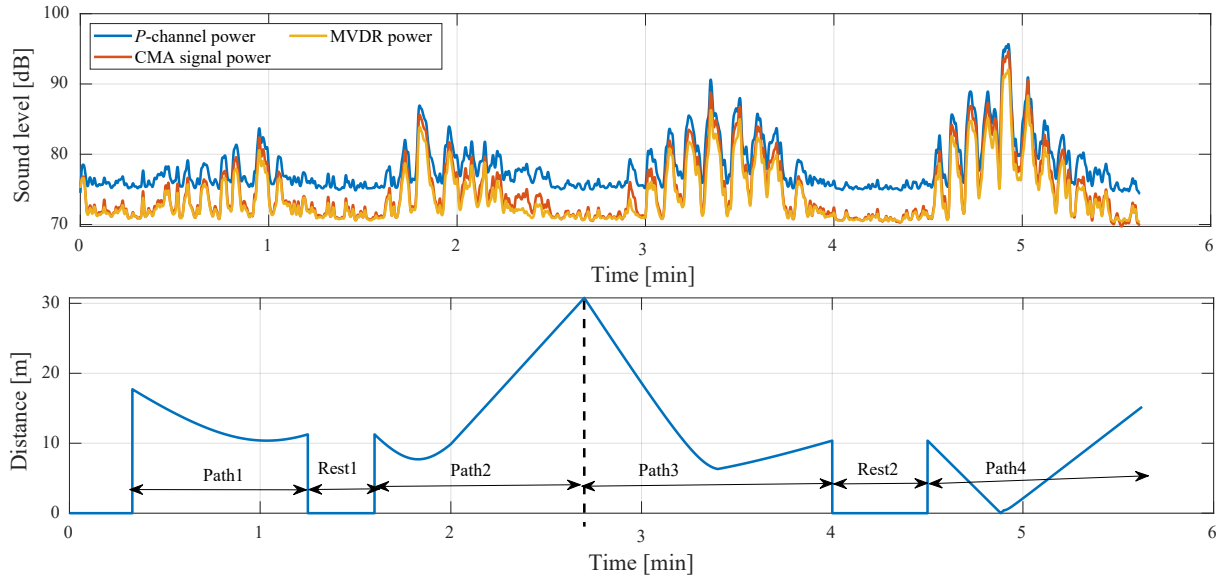


Fig. 8. Signal power estimation and distance between diver and AVS.

The three estimation methods (P -channel power, MVDR, and CMA) were implemented, and the results are presented in Fig. 8, which also shows the distance between the diver and the AVS over time. As illustrated in this figure, the signal power decreases with increasing the distance from diver to AVS, falling below the noise power by the end of Path3. This power decreasing is resulted from attenuation caused by the underwater environment. Furthermore, this figure illustrates that all three methods effectively estimate the waveform of the diver's signal, which exhibits periodic characteristics resulting from repetition in the breathing process. However, the interference between pulses, caused by scattering air bubbles, is evident especially when the diver is close to sensor (Path4), where the acoustic signal from the previous breath persists as the current breath begins.

Utilizing a P -channel for detection is feasible; however, this approach relies only on the pressure sensor that captures the noise from all directions. In contrast, the CMA method depends on the correlation between the pressure and velocity measurements, similar to the MVDR method. However, the MVDR method involves greater computational complexity resulted from searching for maximum power.

Figure 8 illustrates that the CMA and MDVR methods highlight the breathing pulses more effectively than the P -channel method. This is because they reduce noise power when the diver is silent or when SNR is low (e.g., Rest2). However, when the SNR is high (Path4), the estimated powers of all three methods converge to approximate the signal power.

After estimating the power signal which showed the periodicity in the diver's signal, it is necessary to determine the power within the diver breath rate range ($[0.14\text{--}0.42]$ Hz), as this serves as an indicator of the

diver's presence. This can be done by calculation the FFT of estimated power signal over a window of appropriate length that must contain multiple breathing cycles. Figure 9 illustrates FFT using a 13-second window when the diver is not silent. A local maximum at 0.15 Hz is found – indicating a period with 6.7 s – for all the three estimation methods, with advantages in value to CMA first, MVDR, then P -channel. The diver index D_{index} is calculated by summing the squares of FFT values between 0.14 Hz–0.42 Hz as

$$D_{\text{index}} = \frac{1}{M} \sum_{f=0.14}^{f=0.42} |\text{FFT}\{y^2(n)\}(f)|^2,$$

where M is the number of frequencies within the specified range. The results from repeating this process across the entire signals are illustrated in Fig. 10, indicating an increasing in diver index when the diver is actively breathing. By comparing the diver index to the predefined threshold, a decision of diver detection can be made. The CMA method shows a higher diver index value than the MVDR and P -channel methods, giving it an advantage in diver detection. For assertion, the metrics: accuracy, recall, precision, and $F1$ -score are calculated and compared over these methods.

The predefined threshold T_h can be estimated by applying the three methods to recorded ambient noise and selecting the maximum value as D_{index}^N . We then set the threshold $T_h = 1.3D_{\text{index}}^N$ as indicated by TU (2020). The probability of detection P_d can be expressed as

$$P_d = \begin{cases} 1 & D_{\text{index}} \geq 2T_h, \\ (D_{\text{index}} - T_h)/T_h & T_h < D_{\text{index}} < 2T_h, \\ 0 & D_{\text{index}} \leq T_h. \end{cases} \quad (20)$$

This formula was applied to recorded data, and the results are presented in Fig. 10, which shows false alarms

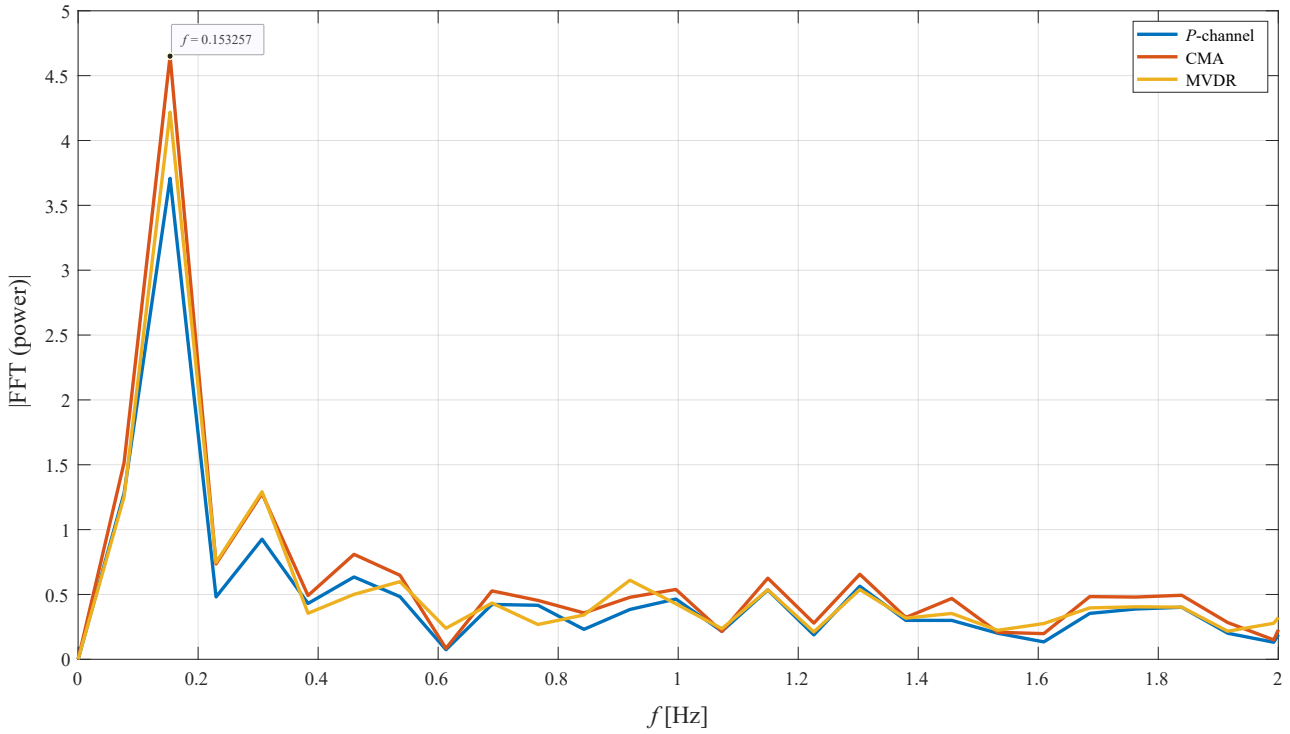


Fig. 9. FFT of window with 13s length where the diver is present.

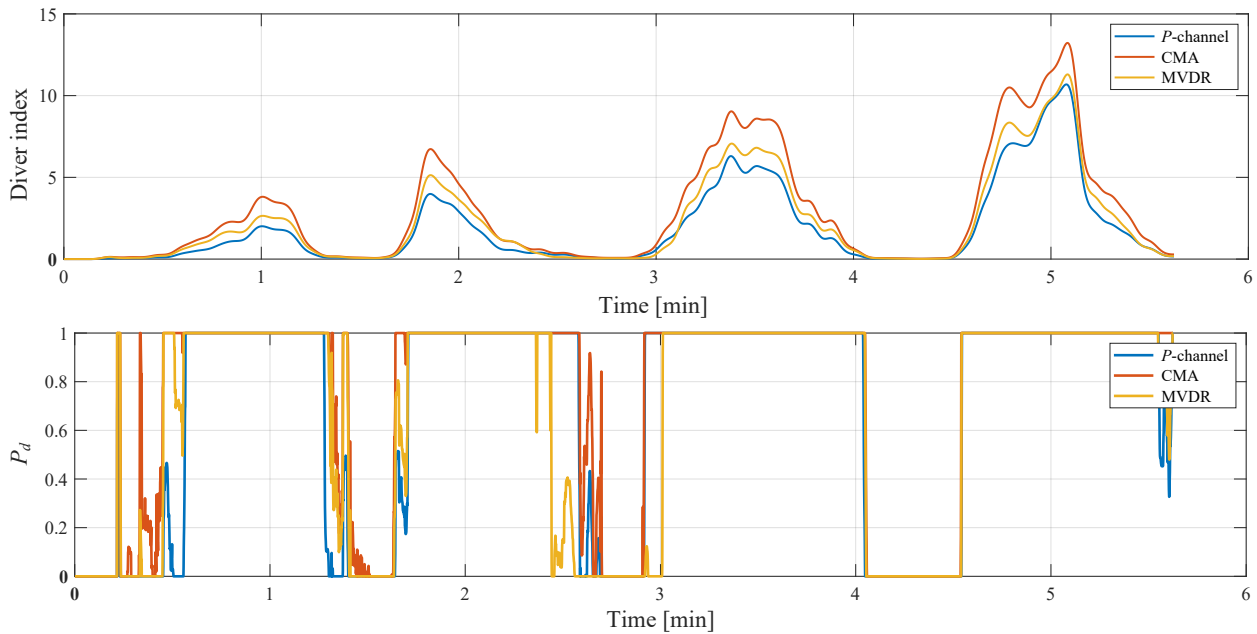


Fig. 10. Calculation of: a) the diver index for journey; b) probability of detection.

in the beginning. Additionally, a low value of P_d is observed when the diver is either silent or far away from the sensor. To determine the detection distance for the three methods, the detection probability for Path3 was calculated and shown in Fig. 11. This figure shows that the distance of detection is approximately 29m for both the P -channel and CMA methods with a preference for CMA, while this distance is about 26 m for the MVDR method.

To compare between the three detection methods, several commonly metrics were calculated as shown in Fig. 12. These metrics were derived based on the confusion matrix provided in Table 1.

The confusion matrix was estimated based on the defined threshold $T_h = 1.3D_{index}^N$. The matrix initially indicates comparable performance among the three methods, with the CMA method showing an advantage in correct classifications (true positives), while

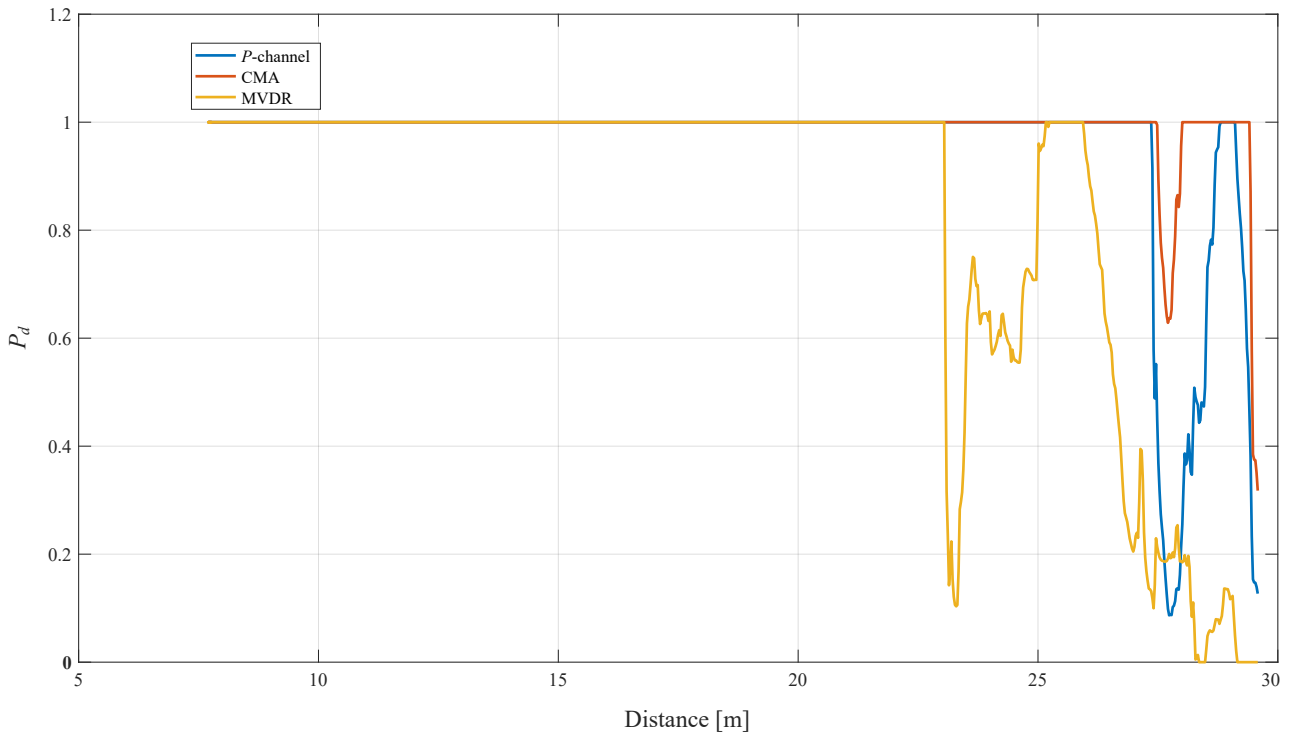


Fig. 11. Detection probability.

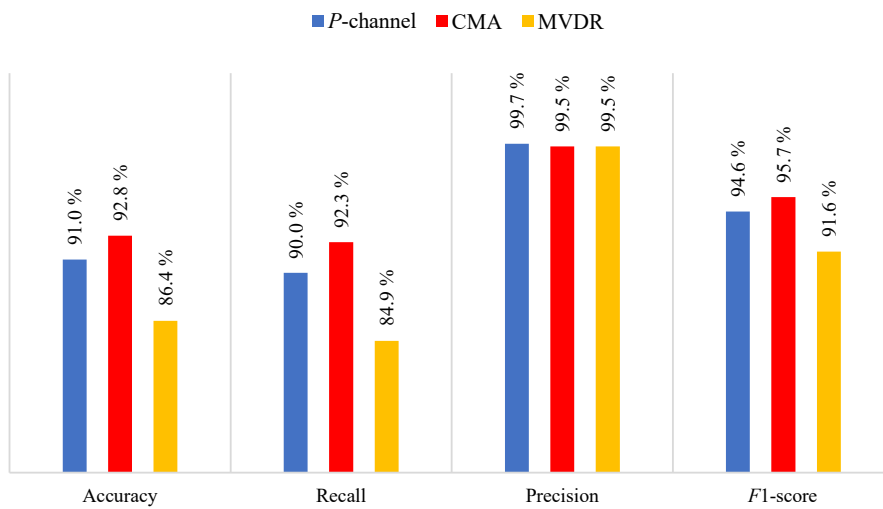


Fig. 12. Comparison of the three methods in detection performance under $T_h = 1.3D_{index}^N$.

Table 1. Confusion matrix under $T_h = 1.3D_{index}^N$.

		P-channel		MVDR		CMA	
		Actual values					
		Positive	Negative	Positive	Negative	Positive	Negative
Predicted values	Positive	4102	11	3868	20	4203	21
	Negative	454	628	688	619	353	618

the *P*-channel method exhibits a lower value of false positives. It is important to note that increasing the threshold value affects the confusion matrix, generally resulting in degraded performance. However, beyond a certain threshold value, the performance of

the MVDR method becomes superior to that of the *P*-channel method.

The results in Fig. 12 illustrate that: the accuracy of detection for the three methods was 86 % to 93 %, the recall was 85 % to 92 %, the precision was 99 % to

100%, and the $F1$ -score was 92% to 96%. These values demonstrate the effectiveness of the three methods in detecting the presence and absence of divers underwater, with advantage to the CMA method.

The experimental data demonstrated the possibility of using the three methods (P -channel, CMA, and MVDR power) in diver detection with a preference for the CMA method.

Therefore, detection is impacted by the method used to represent the periodicity resulting from the breathing process, which affects the diver index value. Furthermore, the long detection window (13 s) is a limitation of using the diver's breath rate as a key signature.

8. Conclusion

This study investigated the feasibility of detecting a diver with an open-circuit scuba by analyzing the acoustic signals emitted during respiration, which are captured by AVS. The AVS focuses on a low-frequency range of [600–1000] Hz, corresponding to the exhalation phase. The detection method involves estimating the signal power associated with the diver's breathing. In this paper, a novel method was presented to estimate the signal power by analyzing the covariance matrix of the 3D-AVS channels. This analysis led to the derivation of a quartic equation that relates the determinant and trace of the AVS covariance matrix to the reciprocal of the SNR. Solving this equation allows for the estimation of both the signal power and the SNR. Simulated data demonstrated the effectiveness of this method in estimating signal power under low SNR (–12 dB). Additionally, the paper presented the MVDR beamformer, which showed potential for estimating signal power along with azimuth and elevation angles using simulated data. The estimated power via CMA, MVDR, and P -channel methods was also compared using trial data, showing a periodicity corresponding to the diver's exhalation. The trial data demonstrated that the CMA method provides a stronger diver detection index compared to MVDR and the P -channel. However, a limitation of these methods is the long detection duration. This issue could potentially be addressed using AI algorithms to identify other features in acoustic diver signals. Investigating this approach will be a focus of our future work.

Appendix. Roots of quartic equation

The existence of one real positive number for the quartic equation (Eq. (18)) is demonstrated, and it can be rewritten as

$$a\alpha^4 + b\alpha^3 + c\alpha^2 + d\alpha + e = 0,$$

$$a = kb = 4kc = 6, \quad d = 4, \quad e = 1.$$

By applying the change of variable $\alpha = \beta - \frac{b}{4a} = \beta - 1$, the depressed quartic equation has been obtained, which has the following form:

$$\beta^4 + p\beta^2 + q\beta + r = 0,$$

where

$$p = \frac{8ac - 3b^2}{8a^2} = 6 \left(\frac{1}{k} - 1 \right) = 6g,$$

$$q = \frac{b^3 - 4abc + 8a^2d}{8a^3} = -8 \left(\frac{1}{k} - 1 \right) = -8g,$$

$$r = \frac{16ab^2c - 64a^2bd - 3b^4 + 256a^3e}{256a^4} = 3 \left(\frac{1}{k} - 1 \right) = 3g,$$

$$g = \left(\frac{1}{k} - 1 \right).$$

The presence of real roots can be determined by evaluating the signs or values of two terms (PRODANOV, 2021) in the following form:

$$\begin{aligned} \delta(p, q, r) &= 256r^3 - 128p^2r^2 + 144pq^2r + 16p^4r - 27q^4 \\ &\quad - 4p^3q^2 = 6912g^3(g+1)^2, \end{aligned}$$

$$L(p, q, r) = 8pr - 9q^2 - 2p^3 = -432g^2(g+1).$$

There are several cases depending on the value of g , as follows:

- 1) $g = 0 (k = 1) \rightarrow \delta = 0, L = 0, p = 0$: rewrite Eq. (18) as $(\alpha + 1)^4 = 0$, the root is $\alpha_i = -1$, this value is invalid because it is a negative number;
- 2) $g = -1 (|k| \rightarrow \infty) \rightarrow \delta = 0, L = 0, p < 0$: rewrite Eq. (18) as $\alpha^3(\alpha + 4) = 0$, there are two real roots: $\alpha_1 = 0$, and $\alpha_2 = -4$. The root α_1 indicates that the power of noise equals to zero, and the root α_2 is invalid;
- 3) $g < 0 (k < 0 \cup k > 1) \rightarrow \delta < 0$: Eq. (18) has two distinct real roots;
- 4) $g > 0 (0 < k < 1) \rightarrow \delta > 0, L < 0$: Eq. (18) has not any real roots.

Thus, the acceptable case is 3, where there are two distinct real roots. To determine the sign of these roots, Descartes' rule of signs is applied, which is based on analyzing the sign changes in the coefficients of the polynomial, as follows:

- $k > 1$: the polynomial $f(\alpha) = k\alpha^4 + 4k\alpha^3 + 6\alpha^2 + 4\alpha + 1$ has no changes in the sign of coefficients. So, the two distinct roots are negative. And this case is invalid;
- $k < 0$: the polynomial $f(\alpha) = k\alpha^4 + 4k\alpha^3 + 6\alpha^2 + 4\alpha + 1$ has one sign change, indicating that the Eq. (14) has one positive real root. Additionally, the polynomial $f(-\alpha) = k\alpha^4 - 4k\alpha^3 + 6\alpha^2 - 4\alpha + 1$ shows three sign changes, meaning it has either three or one negative roots. Since Eq. (18) has two distinct roots, this results one invalid negative real root and one valid positive real root.

As a result, when $k < 0$, Eq. (18) has one positive root, which represents the valid solution for the reciprocal of SNR.

References

- ABRAHAM D.A. (2019), *Underwater Acoustic Signal Processing: Modeling, Detection, and Estimation*, Springer, USA.
- BEREKETLI A., GULDOGAN M.B., KOLCAK T., GUDU T., AVSAR A. (2015), Experimental results for direction of arrival estimation with a single acoustic vector sensor in shallow water, *Journal of Sensors*, **2015**(1): 1–9, <https://doi.org/10.1155/2015/401353>.
- CAO J., LIU J., WANG J., LAI X. (2017), Acoustic vector sensor: Reviews and future perspectives, *IET Signal Processing*, **11**(1): 1–9, <https://doi.org/10.1049/iet-spr.2016.0111>.
- CHEN Y., WANG W., WANG Z., XIA B. (2018), A source counting method using acoustic vector sensor based on sparse modeling of DOA histogram, *IEEE Signal Processing Letters*, **26**(1): 69–73, <https://doi.org/10.1109/LSP.2018.2879547>.
- CHEN Y., ZHANG G., WANG R., RONG H., YANG B. (2023), Acoustic vector sensor multi-source detection based on multimodal fusion, *Sensors*, **23**(3): 1301, <https://doi.org/10.3390/s23031301>.
- CHUNG K.W., LI H., SUTIN A. (2007), A frequency-domain multi-band matched-filter approach to passive diver detection, [in:] *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, pp. 1252–1256, <https://doi.org/10.1109/ACSSC.2007.4487426>.
- DONG H., SUO J., ZHU Z., LI S. (2024), Improved underwater single-vector acoustic DOA estimation via vector convolution preprocessing, *Electronics*, **13**(9): 1796, <https://doi.org/10.3390/electronics13091796>.
- DONSKOY D.M., SEDUNOV N.A., TSIONSKIY M.A. (2008), Variability of SCUBA diver's acoustic emission, [in:] *Optics and Photonics in Global Homeland Security IV*, pp. 272–282, <https://doi.org/10.1117/12.783500>.
- GOROVOY S. *et al.* (2014), A possibility to use respiratory noises for diver detection and monitoring physiologic status, [in:] *Proceedings of Meetings on Acoustics*, **21**(1): 070007, <https://doi.org/10.1121/1.4893767>.
- GOROVOY S. *et al.* (2015), Detecting respiratory noises of diver equipped with rebreather in water, [in:] *Proceedings of Meetings on Acoustics*, **24**(1): 070020, <https://doi.org/10.1121/2.0000171>.
- GRAY M., ROGERS P.H., ZEDDIES D.G. (2016), Acoustic particle motion measurement for bioacousticians?: Principles and pitfalls, [in:] *Fourth International Conference on the Effects of Noise on Aquatic Life*, **27**(1): 010022, <https://doi.org/10.1121/2.0000290>.
- HARI V.N., CHITRE M., TOO Y.M., PALLAYIL V. (2015), Robust passive diver detection in shallow ocean, [in:] *OCEANS 2015 – Genova*, pp. 1–6, <https://doi.org/10.1109/OCEANS-Genova.2015.7271656>.
- JIN B., XU G. (2020), A passive detection method of divers based on deep learning, [in:] *2020 IEEE 3rd International Conference on Electronics Technology (ICET)*, pp. 650–655, <https://doi.org/10.1109/ICET49382.2020.9119556>.
- JOHANSSON A.T., LENNARTSSON R.K., NOLANDER E., PETROVIĆ S. (2010), Improved passive acoustic detection of divers in harbor environments using pre-whitening, [in:] *OCEANS 2010 MTS/IEEE SEATTLE*, pp. 1–6, <https://doi.org/10.1109/OCEANS.2010.5664549>.
- KORENBAUM V., KOSTIV A., GOROVOY S., DOROZHKO V., SHIRYAEV A. (2020), Underwater noises of open-circuit scuba diver, *Archives of Acoustics*, **45**(2): 349–357, <https://doi.org/10.24425/aoa.2020.133155>.
- KORENBAUM V.I. *et al.* (2016), The possibility of passive acoustic monitoring of a scuba diver, *Doklady Earth Sciences*, **466**(2): 187–190, <https://doi.org/10.1134/S1028334X16020136>.
- LENNARTSSON R.K., DALBERG E., PERSSON L., PETROVIĆ S. (2009), Passive acoustic detection and classification of divers in harbor environments, [in:] *OCEANS 2009*, pp. 1–7, <https://doi.org/10.23919/OCEANS.2009.5422407>.
- LEVIN D., HABETS E.A.P., GANNOT S. (2012), Maximum likelihood estimation of direction of arrival using an acoustic vector-sensor, *The Journal of the Acoustical Society of America*, **131**(2): 1240–1248, <https://doi.org/10.1121/1.3676699>.
- LI S., HU B., ZHOU W., ZHAO L. (2015), Experimental study for feature extraction of diver with atmospheric diving suit, [in:] *OCEANS 2015-MTS/IEEE Washington*, pp. 1–5, <https://doi.org/10.23919/OCEANS.2015.7401820>.
- LIU A., YANG D., SHI S., ZHU Z., LI Y. (2019), Augmented subspace MUSIC method for DOA estimation using acoustic vector sensor array, *IET Radar, Sonar and Navigation*, **13**(6): 969–975, <https://doi.org/10.1049/iet-rsn.2018.5440>.
- MA L., GULLIVER T.A., ZHAO A., GE C., BI X. (2019), Underwater broadband source detection using an acoustic vector sensor with an adaptive passive matched filter, *Applied Acoustics*, **148**: 162–174, <https://doi.org/10.1016/j.apacoust.2018.12.023>.
- NAGANANDA K.G., ANAND G.V. (2017), Underwater target tracking with vector sensor array using acoustic field measurements, [in:] *OCEANS 2017 – Aberdeen*, pp. 1–10, <https://doi.org/10.1109/OCEANSE.2017.8084844>.
- NEDELEC S.L. *et al.* (2021), Best practice guide for underwater particle motion measurement for biological applications, Technical report by the University of Exeter for the IOGP Marine Sound and Life Joint Industry Programme, <https://www.researchgate.net/publication/356911609>.
- NEHORAI A., PALDI E. (1994), Acoustic vector-sensor array processing, [in:] *IEEE Transactions on Signal Processing*, **42**(9): 2481–2491, <https://doi.org/10.1109/78.317869>.

25. PAULRAJ A., ROY R., KAILATH T. (1985), Estimation of signal parameters via rotational invariance techniques – Esprit, *Nineteenth Asilomar Conference on Circuits, Systems and Computers*, pp. 83–89, <https://doi.org/10.1109/ACSSC.1985.671426>.
26. PRODANOV E.M. (2021), Classification of the real roots of the quartic equation and their pythagorean tunes, *International Journal of Applied and Computational Mathematics*, **7**(6): 1–14, <https://doi.org/10.1007/s40819-021-01152-w>.
27. ROH T. *et al.* (2022), Fabrication and underwater testing of a vector hydrophone comprising a triaxial piezoelectric accelerometer and spherical hydrophone, *Sensors*, **22**(24): 9796, <https://doi.org/10.3390/s22249796>.
28. SUN G., YANG D., ZHANG L., SHI S. (2003), Maximum likelihood ratio detection and maximum likelihood DOA estimation based on the vector hydrophone [in Chinese], *Acta Acustica*, **28**(1): 66–72, <https://doi.org/10.15949/j.cnki.0371-0025.2003.01.013>.
29. SUTIN A., SALLOUM H., DELORME M., SEDUNOV N., SEDUNOV A., TSIONSKIY M. (2013), Stevens passive acoustic system for surface and underwater threat detection, [in:] *2013 IEEE International Conference on Technologies for Homeland Security (HST)*, pp. 195–200, <https://doi.org/10.1109/THS.2013.6698999>.
30. TICHAVSKY P., WONG K.T., ZOLTOWSKI M.D. (2001), Near-field/far-field azimuth and elevation angle estimation using a single vector hydrophone, [in:] *IEEE Transactions on Signal Processing*, **49**(11): 2498–2510, <https://doi.org/10.1109/78.960397>.
31. TU Q., YUAN F., YANG W., CHENG E. (2020), An approach for diver passive detection based on the established model of breathing sound emission, *Journal of Marine Science and Engineering*, **8**(1): 44, <https://doi.org/10.3390/JMSE8010044>.
32. VAN TREES H.L. (2002), *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, pp. 428–709, John Wiley & Sons, <https://doi.org/10.1002/0471221104>.
33. WANG X., CHEN J., HAN J., JIAO Y. (2014), Optimization for the direction of arrival estimation based on single acoustic pressure gradient vector sensor, *International Journal of Naval Architecture and Ocean Engineering*, **6**(1): 74–86, <https://doi.org/10.2478/ijnaoe-2013-0164>.
34. YUAN M., WANG C., DA L., LI Q. (2022), Signal detection method using a single vector hydrophone in ocean acoustics, *The Journal of the Acoustical Society of America*, **152**(2): 789–798, <https://doi.org/10.1121/10.0013219>.
35. ZHAO A., MA L., HUI J., ZENG C., BI X. (2018), Open-lake experimental investigation of azimuth angle estimation using a single acoustic vector sensor, *Journal of Sensors*, **2018**(1), <https://doi.org/10.1155/2018/4324902>.

Research Paper

**Single Vector Hydrophone DOA Estimation:
Leveraging Deep Learning with CNN-CBAM**

Fanyu ZENG, Yaning HAN, Hongyuan YANG, Dapeng YANG, Fan ZHENG*

*Key Laboratory of Geophysical Exploration Equipment, Ministry of Education,
College of Instrumentation and Electrical Engineering, Jilin University
Changchun, China*

*Corresponding Author e-mail: zhengfan@jlu.edu.cn

*Received November 18, 2024; accepted March 9, 2025;
published online May 5, 2025.*

In recent years, single vector hydrophones have attracted widespread attention in target direction estimation due to their compact design and advantages in complex underwater acoustic environments. However, traditional direction of arrival (DOA) estimation algorithms often struggle to maintain high accuracy in non-stationary noise conditions. This study proposes the novel DOA estimation method based on a convolutional neural network (CNN) and the convolutional block attention module (CBAM). By inputting the covariance matrix of the received signal into the neural network and integrating the CBAM module, this method enhances the model's sensitivity to critical features. The CBAM module leverages channel and spatial attention mechanisms to adaptively focus on essential information, effectively suppressing noise interference and improving directional accuracy. Specifically, CBAM improves the model's focus on subtle directional cues in noisy environments, suppressing irrelevant interference while amplifying essential signal components, which is crucial for an accurate DOA estimation. Experimental results under various signal-to-noise ratio (SNR) conditions validate the method's effectiveness, demonstrating superior noise resistance and estimation precision, providing a robust and efficient solution for underwater acoustic target localization.

Keywords: single vector hydrophone; direction of arrival (DOA); convolutional neural network (CNN); convolutional block attention module (CBAM); noise resistance.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, vector hydrophones have gained a wide range of research and applications in underwater target detection and localization. Compared with traditional scalar hydrophones, vector hydrophones can record acoustic pressure information and partially mitigate isotropic noise. Traditional direction of arrival (DOA) estimation algorithms primarily include high-resolution algorithms based on eigenvalue decomposition, such as multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance techniques (ESPRIT) (TICHAUSKY *et al.*, 2001), which perform well in idealized stationary noise environments. However, nonlinear effects, noise interference, and multipath effects in real marine environments often degrade the algorithm performance.

Advances in deep learning have facilitated the application of various neural networks in the underwater DOA estimation. XIAO *et al.* (2020) proposed a deep unfolding network called DeepFPC, which is based on a fixed-point algorithm, utilizes 1-bit quantization measurements for sparse signal recovery, and has been successfully applied to the DOA estimation, significantly improving estimation accuracy and computational efficiency. In parallel with deep learning advancements, XU *et al.* (2022) developed a block sparse-based dynamic compressed sensing estimator for underwater acoustic communication, addressing challenges like impulsive noise. This method's adaptability to varying underwater channel conditions enhances the potential of neural network-based approaches for improving estimation accuracy in noisy, real-world marine environments. LIU *et al.* (2024) in-

roduced a deep learning-based method for graph similarity computation, which contributes to the field of signal processing and could further enhance DOA estimation techniques. In parallel, XU *et al.* (2019) proposed the M-SIMMUKF algorithm for tracking underwater maneuvering targets, demonstrating its robustness under dynamic and noisy conditions. Moreover, WAJID *et al.* (2020; 2022) explored direction estimation and tracking methods using acoustic vector sensors, highlighting their ability to improve source localization in underwater environments, providing insights that complement deep learning methods for the robust DOA estimation in real-world conditions. LIU *et al.* (2021) proposed the DOA estimation method for underwater acoustic arrays based on a convolutional neural network (CNN), which significantly enhanced the direction estimation accuracy of underwater signals, with strong adaptability and excellent noise resistance. VARANASI *et al.* (2020) combined spherical harmonic decomposition with a deep learning framework to achieve the robust DOA estimation in complex environments, providing an effective solution for signal direction estimation under high-noise conditions. Numerous studies have shown that neural network-based methods can effectively improve the DOA estimation accuracy and adaptability in complex noise environments. For example, YAO *et al.* (2020) proposed a recursive neural network model that achieves the DOA estimation for unknown signal sources through the Toeplitz matrix reconstruction. NIU *et al.* (2017a; 2017b) investigated the performance of three machine learning methods – feedforward neural networks, support vector machines, and random forests – based on vertical arrays for source ranging and validated the feasibility and effectiveness of these methods at different signal-to-noise ratios (SNR). Progress has also been made in machine learning applications for underwater surface and subsurface target resolution in vertical arrays, and direction estimation with horizontal arrays. CHI *et al.* (2019) employed a feedforward neural network combined with early stopping for source ranging, which effectively enhanced the model’s generalization ability, allowing it to maintain strong ranging performance across various environments. CHOI *et al.* (2019) used supervised learning methods to classify surface and submerged vessels in the ocean, significantly improving classification accuracy, demonstrating the potential of their method for practical marine monitoring. OZANICH *et al.* (2020) employed a feedforward neural network for the DOA estimation, demonstrating the efficiency and accuracy of their method in underwater acoustics, further validating the potential of

deep learning in this field. These methods demonstrate excellent performance not only in synthetic data but also show significant potential in practical ocean experiments. The application of neural networks and machine learning expands the possibilities for the DOA estimation with single vector hydrophones, particularly in terms of adaptability and real-time performance.

This paper proposes the CNN-CBAM-based DOA estimation method that uses a normalized covariance matrix as input and incorporates the convolutional block attention module (CBAM) to enhance key feature extraction (WOO *et al.*, 2018) this design is particularly advantageous for the underwater DOA estimation, where capturing subtle directional cues amidst noise is critical. The model is trained on a simulated dataset to improve generalization. Experimental results validate the performance advantages of this approach under varying signal-to-noise conditions, providing an efficient and robust solution for underwater target direction estimation.

2. Vector signal model and data preprocessing

2.1. Single vector hydrophone signal model

Under the far-field plane wave assumption, a single vector hydrophone can measure the sound pressure p and the three velocity components, ν_x , ν_y , and ν_z , at a single point in the sound field. Under ideal conditions, the sensitivities of the sound pressure sensors and velocity sensors are identical, so the received signal for a single vector hydrophone can be represented as

$$\begin{aligned} p(t) &= \sum_{i=1}^N s_i(t) + n_p(t), \\ \nu_x(t) &= \sum_{i=1}^N s_i(t) \cos \theta_i \cos \varphi_i + n_x(t), \\ \nu_y(t) &= \sum_{i=1}^N s_i(t) \sin \theta_i \cos \varphi_i + n_y(t), \\ \nu_z(t) &= \sum_{i=1}^N s_i(t) \sin \varphi_i + n_z(t), \end{aligned} \quad (1)$$

where $s_i(t)$ represents the incident signal from the i -th source; θ_i and φ_i denote the horizontal and pitch angles, respectively; $n_p(t)$, $n_x(t)$, $n_y(t)$, and $n_z(t)$ indicate the noise in the sound pressure and three velocity channels. This expression can be rewritten in a matrix form for further processing:

$$x(t) = A \cdot s(t) + n, \quad (2)$$

$$x(t) = [p(t) \ \nu_x(t) \ \nu_y(t) \ \nu_z(t)]^T, \quad (3)$$

$$A = \begin{bmatrix} 1 & 1 & & 1 & & \\ \cos \theta_1 \cos \varphi_1 & \cos \theta_2 \cos \varphi_2 & \cdots & \cos \theta_i \cos \varphi_i & \cdots & \cos \theta_N \cos \varphi_N \\ \sin \theta_1 \cos \varphi_1 & \sin \theta_2 \cos \varphi_2 & \cdots & \sin \theta_i \cos \varphi_i & \cdots & \sin \theta_N \cos \varphi_N \\ \sin \varphi_1 & \sin \varphi_2 & & \sin \varphi_i & & \sin \varphi_N \end{bmatrix}^T, \quad (4)$$

$$s(t) = [s_1(t) \ s_2(t) \ \dots \ s_i(t) \ \dots \ s_N(t)]^T, \quad (5)$$

$$n = [n_p(t) \ n_x(t) \ n_y(t) \ n_z(t)]^T. \quad (6)$$

2.2. Data preprocessing

Before the received signal $x(t)$ is input into the neural network, it requires preprocessing to enable the neural network to more effectively extract features. First, the covariance matrix \mathbf{Rxx} of the received signal is computed:

$$\mathbf{Rxx} = E [x(t)x^T(t)], \quad (7)$$

followed by normalization of \mathbf{Rxx} :

$$\mathbf{Rxx}_{i,j} = \begin{cases} 2 \times \frac{\mathbf{Rxx}_{i,j} - \min_{\text{val}}}{\max_{\text{val}} - \min_{\text{val}}} - 1, & \mathbf{Rxx}_{i,j} \neq 0, \\ 0, & \mathbf{Rxx}_{i,j} = 0, \end{cases} \quad (8)$$

where $\mathbf{Rxx}_{i,j}$ represents the element at the i -th row and j -th column of the matrix \mathbf{Rxx} ; \min_{val} and \max_{val} are the minimum and maximum elements of the matrix \mathbf{Rxx} , respectively. By normalizing only the non-zero elements, we retain the sparse structure of \mathbf{Rxx} , which is crucial for maintaining the integrity of the signal representation.

3. Network-based direction estimation of a single vector hydrophone

3.1. Convolutional neural network

A CNN is a feedforward deep neural network based on convolutional computations and consists of input, hidden and output layers. The hidden layers include convolutional layers, activation functions, pooling layers, and fully connected layers. The convolutional layer performs convolution operations on input data using kernels of various sizes, with each layer containing multiple kernels. Each kernel consists of weight coefficients and biases and is activated by an activation function. The convolutional layers are connected in sequence to extract higher-dimensional data features through multiple convolution operations. The formula for the convolutional layer is as follows:

$$f_j^{(l)} = g \left(\sum_{i=1}^n w_{ij}^{(l)} \cdot x_i^{(l-1)} + b_j^{(l)} \right), \quad (9)$$

where $f_j^{(l)}$ represents the feature value of the j -th feature in the l -th layer, capturing the output of the convolution operation for this feature; $w_{ij}^{(l)}$ is the weight coefficient connecting the i -th input feature in the $(l-1)$ -th layer to the j -th feature in the current layer; $x_i^{(l-1)}$ denotes the feature value of the i -th input in the $(l-1)$ -th layer, serving as the input to the current layer;

$b_j^{(l)}$ represents the bias term corresponding to the j -th feature in the l -th layer, which offsets the weighted sum of inputs; $g(\cdot)$ is the activation function, unlike a sigmoid function, which is inherently nonlinear, ReLU is piecewise linear but still allows the network to model complex relationships through its composition across multiple layers. Lastly, n is the number of input features in the previous layer. The fully connected layer links the extracted features through neurons and uses ReLU as the activation function:

$$g(x) = \max(0, x). \quad (10)$$

In recent years, CNNs have shown significant potential for improving the DOA estimation accuracy through their feature extraction capabilities. However, underwater acoustic environments present unique challenges; complex noise and interference can hinder CNN’s ability to focus on key features. These challenges necessitate an enhanced model structure that can effectively extract features while dynamically adjusting its focus to prioritize relevant information within high-dimensional data.

The CBAM addresses this issue by introducing an adaptive attention mechanism that refines feature selection based on channel and spatial importance. Integrating CBAM into the CNN architecture enables the model to selectively enhance informative features while suppressing irrelevant background noise. This design is particularly advantageous for the underwater DOA estimation, where capturing subtle directional cues amidst noise is critical. By leveraging the channel and spatial attention, CBAM integration not only enhances directional discrimination but also improves the robustness and accuracy of the DOA estimation process.

3.2. CBAM module

To fully leverage the CNN’s capability for feature extraction from high-dimensional data matrices, this study improves the traditional neural network by introducing CBAM to the CNN structure, thereby enhancing the model’s detail extraction capability for the DOA estimation. The CBAM spatial-channel attention module is illustrated in Fig. 1.

The channel attention module (CAM) adaptively adjusts channel weights in the feature map, enhancing feature selection. For instance, CAM effectively emphasizes subtle directional cues in underwater acoustic data, improving the model’s focus amidst noise interference. First, the input feature map undergoes global average pooling and max pooling along the spatial dimension, resulting in two channel descriptors that represent global average and maximum features. Next, these descriptors pass through a shared fully connected layer sequence, including layers for dimensionality reduction and restoration, with a ReLU ac-

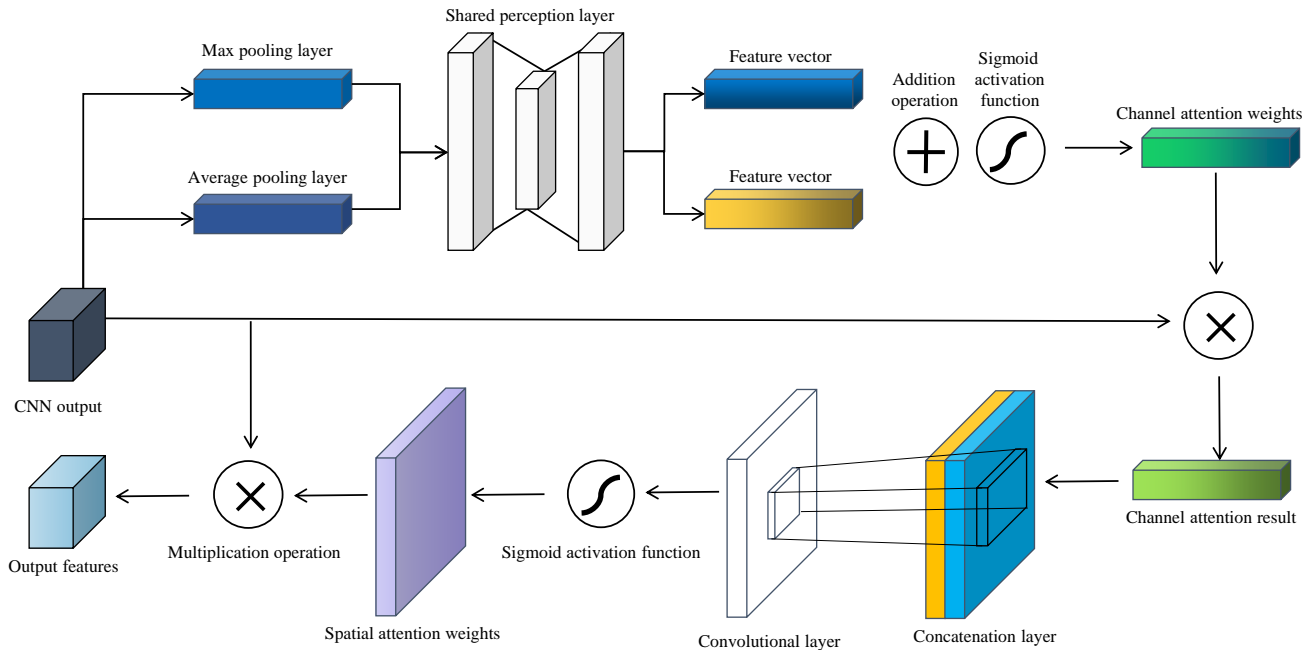


Fig. 1. Working principle of the CBAM attention mechanism in feature enhancement.

tivation function connecting the intermediate layers. Then, the two feature vectors are summed and passed through a Sigmoid activation function to obtain the attention weights for each channel. Finally, these weights are multiplied by the original input feature map on a per-channel basis, completing the channel weighting process. This design allows the network to adaptively allocate attention based on the global features of each channel, thereby effectively enhancing its focus on target features. The Sigmoid activation function is defined as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (11)$$

where x represents the input value, and $\sigma(x)$ represents the output of the Sigmoid function.

The spatial attention module (SAM) learns the weight distribution in the spatial dimension to highlight key area information, suppressing interference from background or irrelevant regions. First, the input feature map undergoes average pooling and max pooling along the channel dimension to produce two single-channel feature maps, representing spatial average and maximum information, respectively. Next, these two feature maps are concatenated along the channel dimension to form a two-channel feature map. This concatenated feature map is then processed by a convolutional layer with a kernel size of 7×7 , capturing a broader range of spatial dependencies and producing a single-channel spatial attention weight map. Finally, this weight map is passed through a Sigmoid activation function and multiplied element-wise with the original input feature map to complete spatial weighting. Through this approach, the SAM can adaptively

focus on key regions within the feature map, enhancing the model's spatial representation capability.

The CBAM combines channel attention and spatial attention to dynamically adjust the weights of key information within the feature map. Channel attention emphasizes key feature channels to enhance the role of different channel features in the network, while spatial attention focuses on critical regions within the feature map, thus capturing essential information required for accurate direction estimation.

3.3. Network structure

The overall network structure is illustrated in Fig. 2. This network model is a deep learning architecture based on a CNN combined with a CBAM, designed for the DOA estimation. The model includes two convolutional layers: the first layer increases the input feature channels from 1 to 32, and the second layer further increases the channels to 64. In the convolutional layers, ' 3×3 ' specifies the kernel size, and the third number indicates the number of kernels. Each convolutional layer is immediately followed by a CBAM module to enhance channel and spatial attention for the features. After processing by the convolutional layers and CBAM modules, the feature data is flattened and passed to two fully connected layers, containing 128 and 64 neurons, respectively, ultimately outputting two directional estimation values. Through the integration of convolution and attention mechanisms, this network structure can more effectively extract key features, thereby improving the accuracy of the DOA estimation.

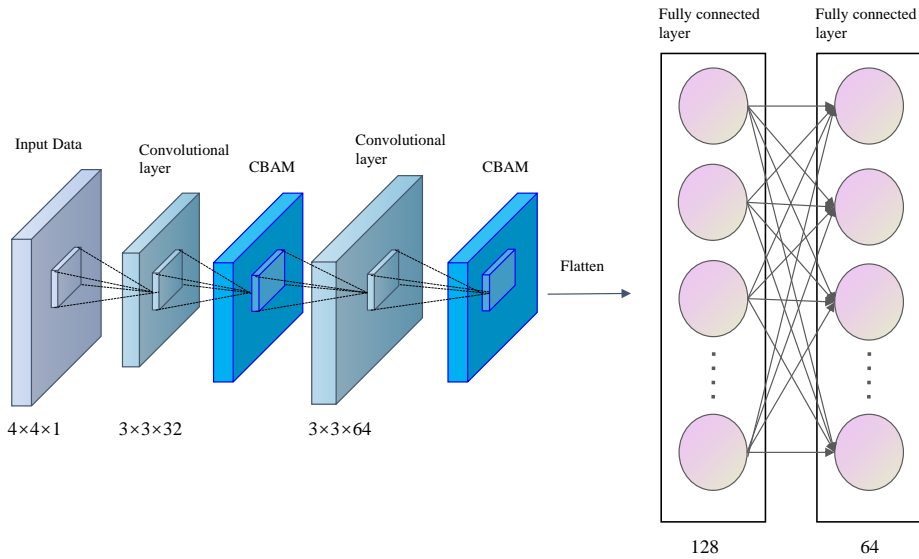


Fig. 2. Network architecture of CNN-CBAM.

3.4. Training process

Monte Carlo simulation is used to generate received signals $x(t)$ without added colored noise according to Eqs. (2)–(6); then, the covariance matrix \mathbf{R}_{xx} of the received signal is computed according to Eq. (7) and normalized to a 4×4 matrix to serve as input for the neural network.

Each sample includes the covariance matrix of noiseless signals generated at the specific azimuth and elevation angles. First, an angle conversion factor, the number of array elements, the sampling frequency, and the time sequence are set up to simulate the basic received signal. Sample angles are randomly generated

within the specified azimuth and elevation ranges, and their corresponding array manifold vectors are calculated and multiplied with the basic signal to obtain the received data. The covariance matrix is then constructed from the received data, and its non-zero elements are normalized by mapping their values to the range $[-1, 1]$, resulting in a normalized covariance matrix. All generated covariance matrices form a dataset for a neural network input, with the array of the sample azimuth and elevation angles serving as output labels for training the deep learning model. The training data consists of noiseless, clean data, with target angles randomly selected between 0° and 359° . Figure 3 shows the time-domain waveforms of the

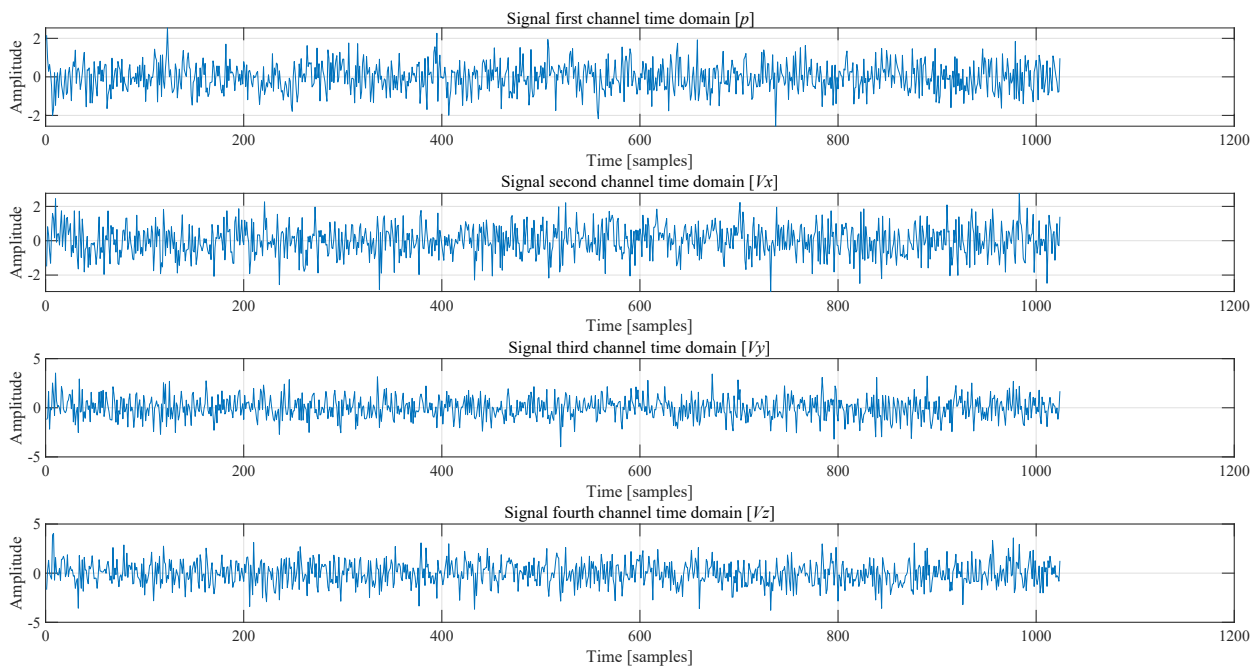


Fig. 3. Time-domain signals for four channels when the SNR is 0 dB.

received signal $x(t)$ for each of the four channels when the SNR is 0.

Training uses the mean squared error (MSE) loss function and the Adam optimizer. The model undergoes training for 180 epochs, with each epoch beginning by initializing the accumulated loss in training mode. Data is loaded in batches to the specified computation device (e.g., GPU), and the model outputs are obtained through forward propagation, with losses calculated between the output and true labels. Loss is backpropagated to update model parameters, and the mean loss for each epoch is accumulated. During training, a StepLR scheduler adjusts the learning rate dynamically every 100 epochs to enhance convergence. At the end of each epoch, the loss and current learning rate are recorded and displayed to monitor training progress. The training loss is shown in Fig. 4.

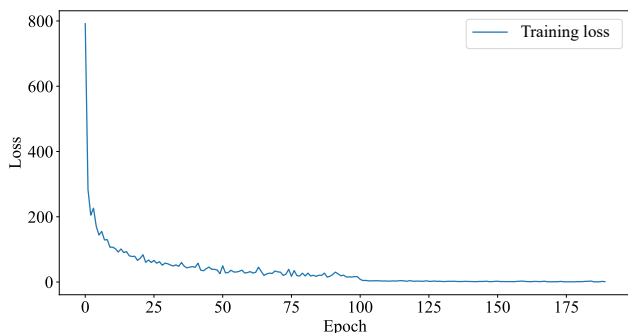


Fig. 4. Training loss variation.

To evaluate the potential overfitting issue, we conducted K -fold cross-validation ($K = 4$) and recorded the training and validation losses. The key parameters used in the validation process are as follows:

- K -value (n_splits): 4, indicating the dataset was divided into 4 subsets for cross-validation;
- batch size: 32, defining the number of samples processed in each iteration;
- number of epochs: 180, representing the total training iterations;
- optimizer: Adam, with a learning rate of 0.001;
- loss function: MSE, used to measure the discrepancy between predicted and true labels.

The results of the K -fold cross-validation are visualized in Fig. 5, which depicts the average training loss (blue line) and average validation loss (red line) over the epochs. The figure shows that both the training and validation losses exhibit a sharp decline in the initial epochs, followed by a gradual stabilization as the number of epochs increases. Notably, the training and validation loss curves remain closely aligned throughout the training process, the convergence of the loss curves to a low and stable level, along with the minimal gap between the training and validation losses, suggests that the model effectively avoids overfitting.

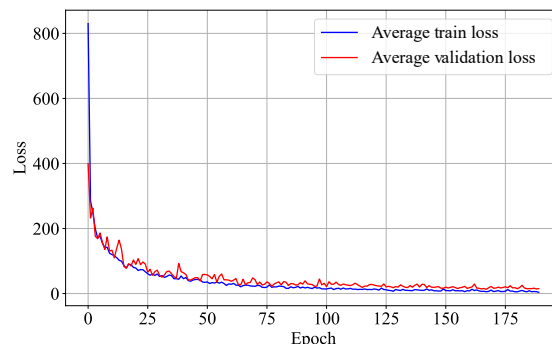


Fig. 5. Training and validation loss curves over epochs in K -fold cross-validation.

This demonstrates the model's ability to fit the data well.

4. Simulation results analysis

The study evaluates the CNN-CBAM model's single-target direction estimation performance across varying SNRs. In the simulation, the target azimuth and elevation angles are randomly selected within the range of 0° to 359° . SNR values are set to -5 dB, 0 dB, 5 dB, 10 dB, and 15 dB in 5 dB increments, with 10 000 data samples generated for each SNR, totaling 50 000 samples. The sampling frequency of the single vector hydrophone is 1 Hz, with each snapshot containing 1024 sample points and one direction estimated per snapshot. The x -axis and y -axis represent the azimuth and elevation angle errors relative to the target's true position, with blue points indicating errors within 10° for both angles. In Fig. 6, the left subfigure shows the histogram of azimuth estimation errors, while the right subfigure shows the histogram of elevation estimation errors. Each subfigure displays the error distributions under SNRs of 15 dB, 10 dB, 5 dB, 0 dB, and -5 dB. As SNR decreases, the error distribution gradually broadens, and errors increase. At higher SNRs, such as 15 dB, 10 dB, and 5 dB azimuth and elevation errors are primarily within 5° . At lower SNRs, like 0 dB and -5 dB, the CNN-CBAM model demonstrates reliable performance, with the majority of estimation errors not exceeding 15° in azimuth and 10° in pitch.

In this simulation, the target azimuth and elevation angles were set to $[45^\circ, 45^\circ]$, with other simulation parameters remaining consistent with previous settings. The target direction was estimated using the weighted histogram method, MUSIC, Capon, the fourth-order cumulant method (GUO *et al.*, 2018), SBL (Liang *et al.*, 2021), and the CNN-CBAM model. Figure 7 illustrates the CNN-CBAM model's estimation results under various SNR conditions, where the x -axis and y -axis represent the azimuth and elevation angle errors relative to the true target position, with blue points

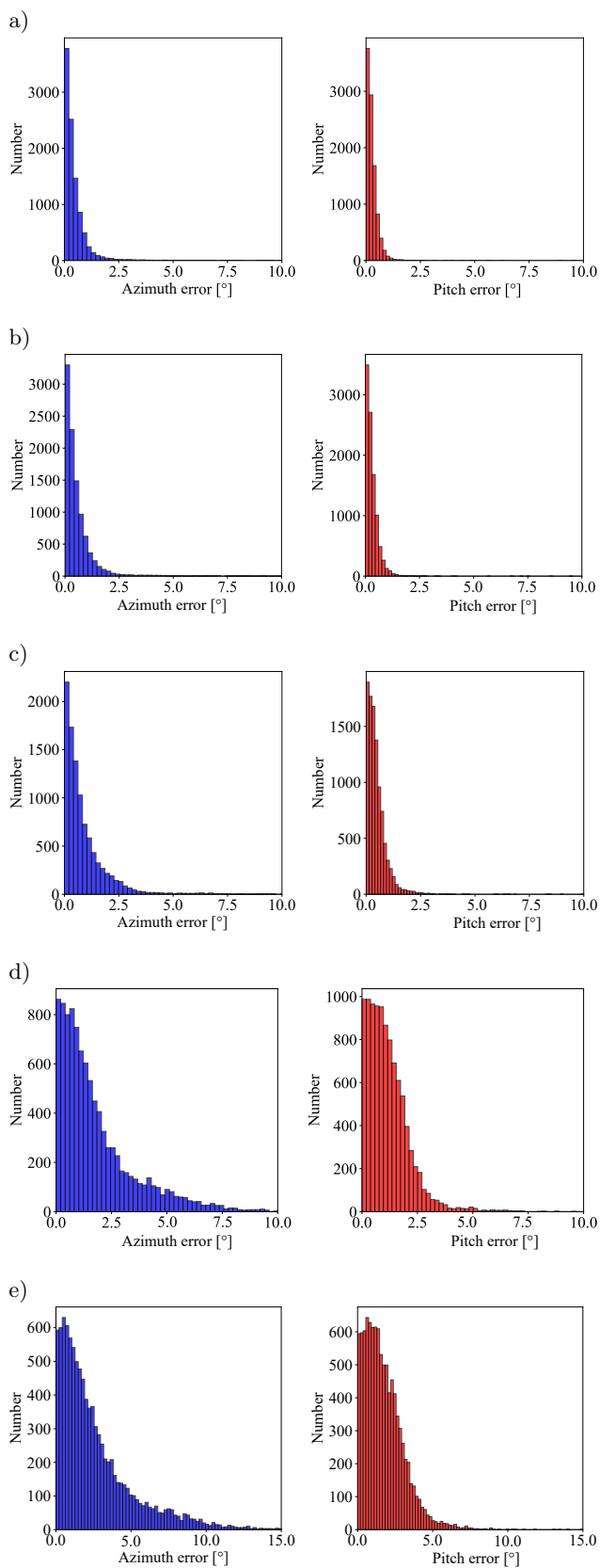


Fig. 6. Random direction estimation results of the CNN-CBAM method under different SNRs: a) SNR = 15 dB; b) SNR = 10 dB; c) SNR = 5 dB; d) SNR = 0 dB; e) SNR = -5 dB.

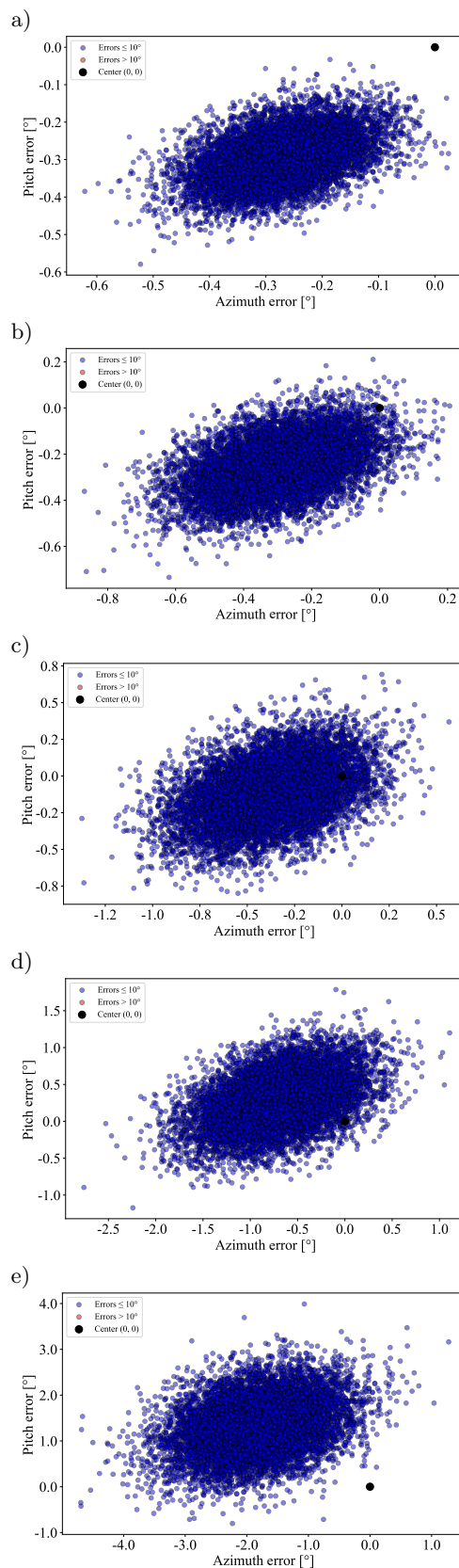


Fig. 7. Estimation results of the CNN-CBAM method for $[45^\circ, 45^\circ]$ under different SNRs: a) SNR = 15 dB; b) SNR = 10 dB; c) SNR = 5 dB; d) SNR = 0 dB; e) SNR = -5 dB.

indicating errors within 10° for both angles. The results show that under SNR conditions of 5 dB, 10 dB, and 15 dB, the CNN-CBAM model achieves effective direction estimation with azimuth and elevation errors not exceeding 2° . Even under SNR conditions of -5 dB and 0 dB, the model successfully estimates the source's direction with errors in both azimuth and elevation within 7° .

The analysis of azimuth and pitch angle error distributions under varying SNR conditions (SNR = 15, 10, 5, 0, -5) reveals a systematic increase in distribution non-uniformity and measurement inaccuracy as SNR decreases. At high SNR (15 dB, 10 dB), the majority of error points are clustered in the southwest direction relative to the origin, indicating high measurement precision with minimal deviation. As SNR reduces to 5 dB, the error distribution shifts predominantly to the west, reflecting a moderate decline in accuracy. In low SNR conditions (0 dB and -5 dB), the error points are concentrated in the northwest direction, demonstrating significant dispersion and the emergence of systematic errors. This directional non-uniformity in the error distribution is attributed to noise interference and system instability, which are exacerbated under low SNR conditions. To quantify this systematic deviation, we introduce the concept of bias (B), defined as the mean difference between the estimated angles (\hat{y}_i) and the true angles (y_i):

$$B = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i), \quad (12)$$

where n is the total number of measurements. This bias term captures the systematic error component, which becomes increasingly significant as SNR decreases, highlighting the need for robust error correction strategies in low SNR environments.

The bias in azimuth and pitch measurements refer to the systematic deviation of the estimated values from their true values. In this study, the bias is quantified as the mean error of azimuth and pitch measurements under different SNR conditions. As shown in Table 1, the mean azimuth errors exhibit a consistent negative bias across all SNR levels, ranging from -0.27° at 15 dB to -1.82° at -5 dB. This indicates that the azimuth estimates are systematically lower than the true values, and the magnitude of this bias increases with decreasing SNR.

Table 1. Biases of azimuth and pitch angles under different SNR conditions.

SNR [dB]	Azimuth error mean [$^\circ$]	Pitch error mean [$^\circ$]
15	-0.27095	-0.29417
10	-0.29320	-0.24869
5	-0.37225	-0.11144
0	-0.68524	0.28555
-5	-1.82368	1.33251

Similarly, the mean pitch errors demonstrate a transition from negative to positive bias as the SNR decreases. At higher SNR levels (e.g., 15 dB), the pitch errors show a negative bias of -0.29° , suggesting that the pitch estimates are slightly lower than the true values. However, as the SNR decreases, the bias shifts towards positive values, reaching 1.33° at -5 dB. This indicates that the pitch estimates become systematically higher than the true values under low SNR conditions.

The observed biases in both azimuth and pitch measurements highlight the influence of SNR on the accuracy of the estimation process. The increasing negative bias in azimuth and the transition from negative to positive bias in pitch suggest that the estimation algorithms may be more susceptible to noise in certain directions or dimensions. These findings emphasize the need for bias correction techniques, particularly in low SNR environments, to improve the accuracy of azimuth and pitch measurements.

Figure 8 shows the estimation results of different methods under a -5 dB SNR. MUSIC, Capon, weighted histogram, and fourth-order methods use a spectral peak search step size of 0.1° , while SBL employs a grid step size of 0.1° . CNN-CBAM and SBL stand out as the most effective methods for the DOA estimation, offering high accuracy. While weighted histogram and fourth-order cumulant methods remain competitive. MUSIC and Capon methods are more sensitive to noise and exhibit higher estimation errors. However, a notable limitation of CNN-CBAM in this scenario is that a significant portion of its estimates do not uniformly distribute around the origin, as observed in the error distribution plot. This deviation indicates that while CNN-CBAM achieves high accuracy in many cases, its estimates can be biased or skewed under low SNR conditions, leading to occasional instability. This limitation highlights the need for further refinement of the method to ensure more consistent and uniform performance across all scenarios. Future work could focus on enhancing the noise resilience of CNN-CBAM by optimizing its attention mechanisms, incorporating additional noise suppression techniques, or integrating it with probabilistic frameworks like those used in SBL to address this issue and improve its robustness in highly noisy environments.

This study adopts the root mean square error (RMSE) as the performance metric for direction estimation, where \hat{y}_i represents the estimated data and y_i represents the actual data:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}. \quad (13)$$

Figure 9 illustrates the RMSE of six DOA estimation methods, weighted histogram, MUSIC, Capon, fourth-order, CNN-CBAM, and SBL, across SNR levels ranging from -5 dB to 15 dB. As SNR increases,

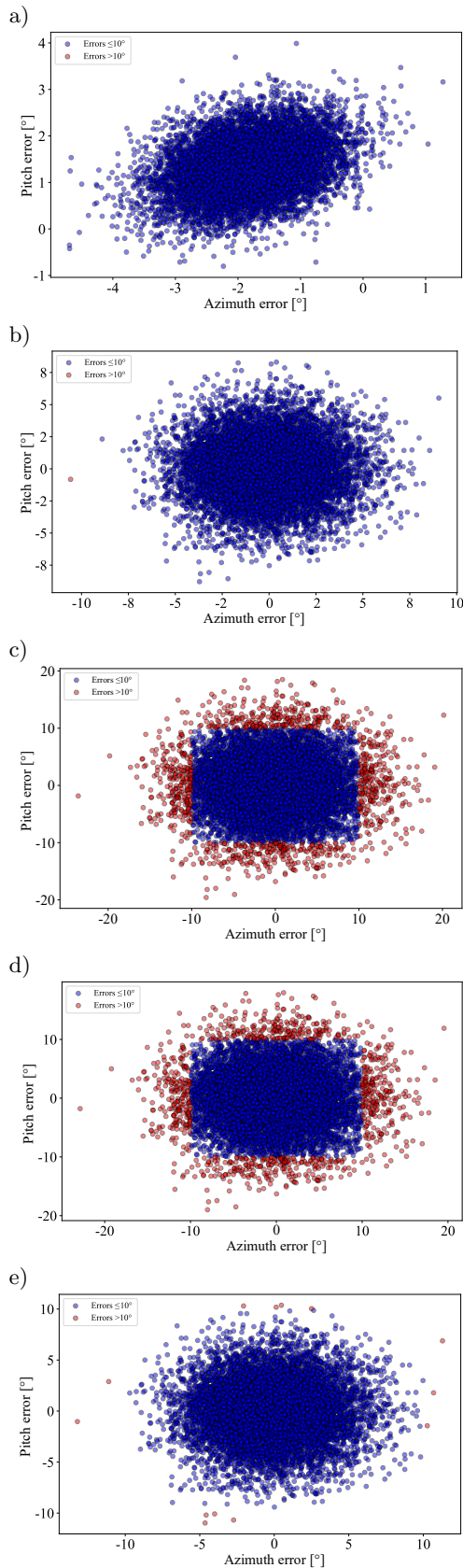


Fig. 8. Estimation results of various methods for $[45^\circ, 45^\circ]$ when the SNR is -5 dB: a) CNN-CBAM; b) weighted histogram; c) MUSIC; d) Capon; e) fourth-order cumulant; f) SBL.

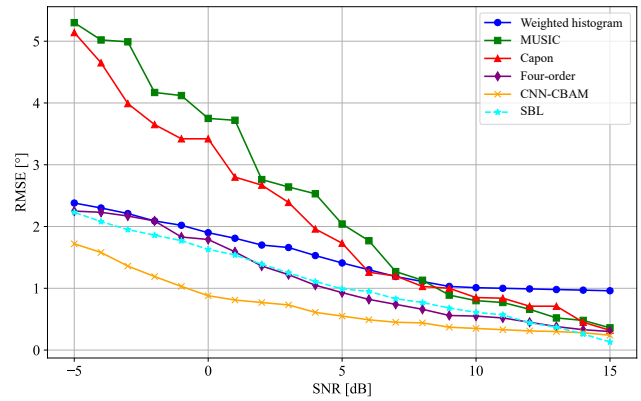


Fig. 9. RMSE of various methods for estimating $[45^\circ, 45^\circ]$ under different SNRs.

the RMSE for all methods decreases, reflecting improved estimation accuracy. CNN-CBAM consistently achieves the lowest RMSE values, demonstrating high accuracy across all SNR conditions, particularly excelling at higher SNR levels. The fourth-order also performs well, closely following CNN-CBAM and SBL, while MUSIC and Capon show moderate performance with higher RMSE values at lower SNR. Overall, CNN-CBAM, the fourth-order and SBL stand out as the most effective methods, offering high accuracy and reliability in the DOA estimation.

Underwater environments are characterized by complex noise conditions, including not only Gaussian noise but also other types of noise such as impulse noise, ambient noise, and biological noise. To evaluate the adaptability of the CNN-CBAM model to such environments, we conducted experiments by adding impulse noise to the data at a SNR of 0 dB, with impulse noise ratios ranging from 0.05 to 0.25. We compared the RMSE of six DOA estimation methods. Figure 10 illustrates the RMSE performance of six DOA estimation methods, the RMSE of all methods generally rises, with MUSIC and Capon showing the most significant degradation in performance. In contrast, CNN-CBAM performs well under low impulse

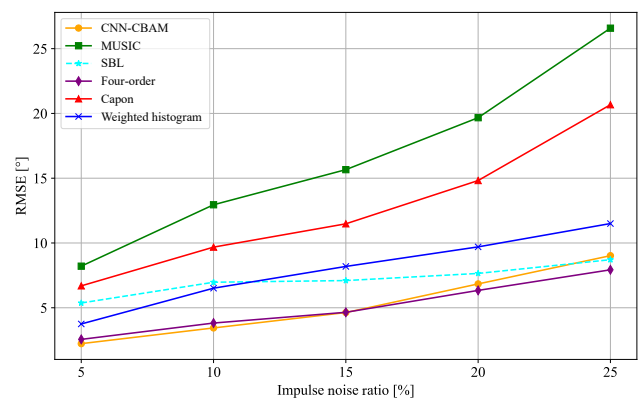


Fig. 10. RMSE of various methods under varying impulse noise ratios.

noise ratios (5% and 10%), achieving lower RMSE values, however, as the noise ratio increases, its performance degrades rapidly, with RMSE rising significantly, highlighting its sensitivity to higher levels of impulse noise, the fourth-order cumulant demonstrates greater resilience, with a slower increase in RMSE. The results suggest that CNN-CBAM, despite its advanced architecture, may require enhancements such as noise suppression techniques or hybrid approaches to improve its performance in environments dominated by impulse noise. Future work could focus on integrating traditional signal processing methods with deep learning models to achieve better adaptability to complex underwater noise conditions.

Table 2 presents the computation time of six DOA estimation methods for single and multiple (10) targets, MUSIC, Capon, weighted histogram, and fourth-order methods use a spectral peak search step size of 1° , while SBL employs a grid step size of 1° , revealing that CNN-CBAM, despite its slightly longer computation time (0.125 s) for single-target estimation compared to Capon (0.136 s) and weighted histogram (0.031 s), demonstrates superior scalability and efficiency for multiple targets, requiring only 0.228 s for 10 targets. This advantage stems from its parallel processing capability, attention mechanism, and optimized framework, which minimize computational overhead in complex scenarios. In contrast, methods like the fourth-order and SBL exhibit significantly longer computation times (6.722 s and 13.279 s, respectively) for multiple targets, making them less practical. Thus, CNN-CBAM emerges as an efficient choice for the real-time DOA estimation, particularly in applications involving continuous estimation.

Table 2. Comparison of methods in processing time.

Method	Time for single target [s]	Time for 10 targets [s]
CNN-CBAM	0.125	0.228
MUSIC	0.226	3.306
Capon	0.136	1.437
Weighted histogram	0.031	0.359
Fourth-order	0.583	6.722
SBL	1.247	13.279

5. Discussions

The proposed CNN-CBAM model represents a significant advancement in the DOA estimation for single vector hydrophones, particularly in complex underwater acoustic environments. By integrating the CBAM into a CNN, the model achieves superior noise resistance and estimation accuracy across a wide range of SNRs. This innovative approach addresses the limitations of traditional methods such as MUSIC and

Capon, which often struggle with non-stationary noise and multipath effects. The model's ability to adaptively focus on critical features through channel and spatial attention mechanisms establishes it as a robust solution for real-time underwater target localization.

However, several challenges remain to be addressed. A notable limitation is the model's performance in multi-source environments or scenarios with overlapping signal sources. While CNNs excel in one-to-one mapping tasks, their performance deteriorates when handling multiple concurrent sources. This degradation is primarily attributed to the inherent complexity of disentangling overlapping signals, which demands more sophisticated feature extraction and separation techniques. Future work should prioritize enhancing the model's capability to handle multi-source scenarios, potentially through the integration of advanced signal separation algorithms or hybrid architectures that combine CNNs with other machine learning paradigms.

The computational efficiency of the CNN-CBAM model is another critical consideration. As demonstrated in Table 2, the model exhibits competitive processing times for single-target estimation and demonstrates superior scalability for continuous estimation. This efficiency is largely due to the parallel processing capabilities of CNNs and the optimized attention mechanisms of CBAM. Nevertheless, computational requirements may escalate significantly in multi-source environments, necessitating further optimization of the network architecture and training process. Future research should explore techniques such as model pruning, quantization, and distributed computing to enhance scalability and reduce computational overhead.

Integrating the CNN-CBAM model into existing underwater acoustic systems presents additional challenges. A key issue is compatibility with legacy hardware and software, which may require substantial modifications to accommodate the deep learning framework. Moreover, the model's reliance on large datasets for training poses logistical challenges in data collection and preprocessing. To address these issues, future work should focus on developing modular and adaptable frameworks that can be seamlessly integrated into existing systems, as well as exploring transfer learning techniques to reduce dependency on extensive training data.

The CNN output, while not always precise, is 'precisely wrong' in the sense that it consistently deviates from the true values in a predictable manner. This systematic bias, particularly evident in low SNR conditions, underscores the need for robust error correction strategies. Future research should investigate methods to mitigate this bias, such as incorporating probabilistic frameworks or ensemble learning techniques, to improve the model's reliability and accuracy. By addressing these challenges and advancing the proposed ap-

proach, the CNN-CBAM model has the potential to significantly enhance the state of the art in the DOA estimation, providing a robust and efficient solution for underwater acoustic target localization in real-world applications.

Future research directions should focus on advancing the multi-source DOA estimation through the integration of signal separation techniques or hybrid architectures, enhancing the model's capability in complex environments. Systematic biases in the model's output, particularly under low SNR conditions, must be addressed through robust error correction strategies to ensure reliable and accurate estimations. Computational efficiency and scalability can be further optimized via techniques such as model pruning and distributed computing, enabling real-time applications. To facilitate seamless integration into existing underwater acoustic systems, modular frameworks should be developed, overcoming compatibility and logistical challenges. Additionally, leveraging transfer learning techniques can reduce dependency on extensive training datasets while improving adaptability to diverse operational scenarios. Furthermore, real-world experiments will be conducted to validate the method's effectiveness in practical underwater environments, ensuring its robustness and applicability in real-world scenarios. By addressing these critical areas, the CNN-CBAM model is poised to significantly advance the state of the art in the DOA estimation, offering a robust and efficient solution for underwater acoustic target localization in real-world applications.

6. Conclusion

This study proposes a CNN-CBAM-based approach for the DOA estimation using a single vector hydrophone, enhancing accuracy in complex underwater environments. By integrating the CBAM with a CNN, the model processes normalized covariance matrices to focus on critical channels and spatial features. Experimental results demonstrate robustness across varying SNRs, with azimuth and elevation errors within 5° at higher SNRs (15 dB, 10 dB, 5 dB) and within 15° in azimuth and 10° in pitch at lower SNRs (0 dB, -5 dB).

The CNN-CBAM model outperforms traditional methods such as MUSIC and Capon in precision and noise resistance, addressing limitations of eigenvalue decomposition-based methods in non-stationary noise and multipath environments. Challenges remain in multi-source environments, where overlapping signals degrade performance. Future work will focus on enhancing the multi-source DOA estimation, optimizing computational efficiency, leveraging transfer learning for practical deployment, and conducting real-world experiments to validate the method's effectiveness. These advancements will solidify the CNN-CBAM

model as a robust solution for real-time underwater target localization.

FUNDINGS

This work was supported by the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (grant no. SML2022SP201).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

References

- CHI J., LI X., WANG H., GAO D., GERSTOFT P. (2019), Sound source ranging using a feed-forward neural network trained with fitting-based early stopping, *The Journal of the Acoustical Society of America*, **146**(3): EL258–EL264, <https://doi.org/10.1121/1.5126115>.
- CHOI J., CHOO Y., LEE K. (2019), Acoustic classification of surface and underwater vessels in the ocean using supervised machine learning, *Sensors*, **19**(16): 3492, <https://doi.org/10.3390/s19163492>.
- GUO Y., HAN J., WANG C. (2018), Multi-acoustic source localization algorithm based on fourth-order moments for single vector hydrophone [in Chinese], *Journal of Sichuan University Natural Science Edition*, **55**: 733.
- LIU B., WANG Z., ZHANG J., WU J., QU G. (2024), DeepSIM: A novel deep learning method for graph similarity computation, *Soft Computing*, **28**: 61–76, <https://doi.org/10.1007/s00500-023-09288-1>.
- LIU Y., CHEN H., WANG B. (2021), DOA estimation based on CNN for underwater acoustic array, *Applied Acoustics*, **172**: 107594, <https://doi.org/10.1016/j.apacoust.2020.107594>.
- LIANG G., SHI Z., QIU L., SUN S., LAN T. (2021), Sparse Bayesian learning based direction-of-arrival estimation under spatially colored noise using acoustic hydrophone arrays, *Journal of Marine Science and Engineering*, **9**(2): 127, <https://doi.org/10.3390/jmse9020127>.
- NIU H., OZANICH E., GERSTOFT P. (2017a), Ship localization in Santa Barbara Channel using machine learning classifiers, *The Journal of the Acoustical Society of America*, **142**(5): EL455–EL460, <https://doi.org/10.1121/1.5010064>.
- NIU H., REEVES E., GERSTOFT P. (2017b), Source localization in an ocean waveguide using supervised machine learning, *The Journal of the Acoustical Society of America*, **142**(3): 1176–1188, <https://doi.org/10.1121/1.5000165>.
- OZANICH E., GERSTOFT P., NIU H. (2020), A feed-forward neural network for direction-of-arrival estimation, *The Journal of the Acoustical Society of Amer-*

- ica*, **147**(3): 2035–2048, <https://doi.org/10.1121/10.0000944>.
10. TICHAVSKY P., WONG K.T., ZOLTOWSKI M.D. (2001), Near-field/far-field azimuth and elevation angle estimation using a single vector hydrophone, *IEEE Transactions on Signal Processing*, **49**(11): 2498–2510, <https://doi.org/10.1109/78.960397>.
 11. VARANASI V., GUPTA H., HEGDE R.M. (2020), A deep learning framework for robust DOA estimation using spherical harmonic decomposition, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **28**: 1248–1259, <https://doi.org/10.1109/TASLP.2020.2984852>.
 12. WAJID M., KUMAR A., BAHL R. (2020), Direction estimation and tracking of coherent sources using a single acoustic vector sensor, *Archives of Acoustics*, **45**(2): 209–219, <https://doi.org/10.24425/aoa.2020.132495>.
 13. WAJID M., KUMAR A., BAHL R. (2022), Microphone-based acoustic vector sensor for direction finding with bias removal, *Archives of Acoustics*, **47**(2): 151–167, <https://doi.org/10.24425/aoa.2022.141646>.
 14. WOO S., PARK J., LEE J.Y. (2018), CBAM: Convolutional block attention module, [in:] *Computer Vision – ECCV 2018*, Ferrari V., Hebert M., Sminchisescu C., Weiss Y. [Eds], Cham: Springer, pp. 3–19, https://doi.org/10.1007/978-3-030-01234-2_1.
 15. XIAO P., LIAO B., DELIGIANNIS N. (2020), DeepFPC: A deep unfolded network for sparse signal recovery from 1-Bit measurements with application to DOA estimation, *Signal Processing*, **176**: 107699, <https://doi.org/10.1016/j.sigpro.2020.107699>.
 16. XU L., CHEN L., LI Y., JIANG W. (2022), A block sparse-based dynamic compressed sensing channel estimator for underwater acoustic communication, *Journal of Marine Science and Engineering*, **10**(4): 536, <https://doi.org/10.3390/jmse10040536>.
 17. XU L., MA Y., YANG Z., GAO T. (2019), Tracking of underwater maneuvering target via M-SIMMUKF algorithm, [in:] *Proceedings of the 6th International Conference on Information Science and Control Engineering (ICISCE)*, pp. 630–634, <https://doi.org/10.1109/ICISCE48695.2019.00131>.
 18. YAO Y., LEI H., HE W. (2020), A-CRNN-based method for coherent DOA estimation with unknown source number, *Sensors*, **20**(8): 2296, <https://doi.org/10.3390/s20082296>.

Research Paper

A Hybrid Method for Predicting the Bistatic Target Strength Based on the Monostatic Scattering Sound Pressure Data

Yuhang TANG⁽¹⁾, Chenyu FAN^{(1)*}, Yinhao LI⁽²⁾,
Changxiong CHEN⁽²⁾, Zilong PENG⁽²⁾

⁽¹⁾ *PLA Unit 92578*
Beijing, China

⁽²⁾ *School of Energy and Power*
Jiangsu University of Science and Technology
Zhenjiang, China

*Corresponding Author e-mail: 15350715833@163.com

Received January 8, 2025; revised May 8, 2025; accepted May 10, 2025;
published online June 16, 2025.

To predict the scattered acoustic field for underwater targets with separate transmission and reception points, a forecasting method based on limited scattered acoustic pressure data is proposed. This method represents the scattered acoustic field as the product of an acoustic scattering transfer function and a source density function. By performing numerical integration, the transfer function is obtained using the model surface grid information as input. An equation system concerning the unknown source density function is then derived using the computed scattering transfer matrix, the principle of acoustic reciprocity, and the geometric properties of the target. The unknown source density function is solved using the least squares method. The scattered field with separate transmission and reception points is then obtained by multiplying the calculated transfer matrix with the estimated source density function. This paper applies the finite element method (FEM) to solve the scattering field for a benchmark model with separate transmission and reception points. Using a subset of the elements as input, predictions of the omnidirectional scattered field were made. The predicted results were subsequently compared with those obtained from FEM simulations. The simulation results demonstrate that the proposed method maintains high computational accuracy and is applicable to the prediction of low-frequency scattered fields from underwater targets with spatially separated source and receiver. Further comparison with the FEM-calculated target strength patterns across varying incident–reception angles reveals a high level of agreement, indicating that accurate bistatic target strength predictions can be achieved with a limited amount of input data.

Keywords: acoustic scattering properties; bistatic scattering sound field; scattering sound field prediction; target strength.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Acoustic waves are the only physical field capable of transmitting information effectively over long distances in the ocean. Devices that use acoustic waves for underwater detection, positioning, navigation, and communication are collectively referred to as sonar (TANG *et al.*, 2018). Currently, sonar technology and

detection methods are undergoing significant transformation, with increasingly complex underwater acoustic environments. Future developments in sonar technology are expected to integrate active and passive systems, multi-band capabilities, and multifunctionality. Key directions include low-frequency, high-power, adaptive array processing, and distributed transmission and reception systems.

The rise of big data and artificial intelligence has also enabled the possibility of coordinated networks and formations of sonar-equipped tools, such as unmanned surface and underwater vehicles (UUVs). However, complex hull structures face unprecedented challenges from these three-dimensional underwater detection networks.

Research methods for studying the echo characteristics of complex hull structures can be categorized into theoretical solutions, numerical methods, and approximation methods. Theoretical solutions are primarily used to analyze the scattering acoustic fields of regular models, such as spheres and infinitely long cylinders. For more complex shapes and materials, numerical methods like the T-matrix method (WATERMAN, 2005), the finite element/boundary element method (FEM/BEM) (ZHOU *et al.*, 2009), and the finite difference time domain (FDTD) method (SCHNEIDER *et al.*, 1998). Numerical methods offer a broad range of applications and are capable of solving target scattering problems under virtually any conditions. For high-frequency, large-scale models, the computational speed of numerical methods is often slow, necessitating the use of approximation methods for more efficient calculations. FAN and ZHOU (2006) proposed a modified planar element method that incorporates considerations for occlusion and secondary scattering effects. PENG *et al.* (2018) presented a method to predict echo characteristics of surface targets using the Kirchhoff approximation. XUE *et al.* (2023) improved the computational efficiency of the patch element method by replacing planar elements with surface elements, thereby enhancing the overall calculation speed of the method.

The bistatic target detection offers advantages such as an excellent, wide detection range, and strong anti-jamming capability, leading many scholars to conduct in-depth studies on the bistatic acoustic scattering characteristics. CHEN *et al.* (2024) proposed and validated a prediction method for underwater acoustic scattering. LIU *et al.* (2012) proposed a modification to the scattering integration region on the target surface, thereby extending the applicability of physical acoustics to arbitrary separation angles. WANG *et al.* (2022) developed a time-domain transformation method based on the Kirchhoff approximation for calculating bistatic acoustic scattering of underwater rigid targets. In the field of bistatic underwater maneuvering target tracking. GUNDERSON *et al.* (2017) discussed the interference and resonance structures present in the frequency responses of the targets, and presented bistatic spectra for a variety of elastic sphere materials. MENG *et al.* (2024) proposed a highlight model for predicting bistatic acoustic scattering characteristics. SCHMIDT (2001) studied bistatic scattering from buried targets in shallow water. AGOUNAD *et al.* (2023) systematically analyzed the guided wave propagation characteristics of cylindrical shells under

bistatic acoustic scattering configurations from both theoretical and experimental perspectives, and proposed a time-frequency analysis-based method to estimate the group velocities of waves propagating in different directions. CHENG *et al.* (2010) established a bistatic scattering strength model for underwater targets under far-field conditions based on the Kirchhoff approximation. PARK *et al.* (2006) studied the bistatic acoustic scattering phenomena of a hemispherical closed cylinder. GU *et al.* (2025) proposed an improved rapid prediction method for solving the full-space bistatic scattering acoustic field of underwater vehicles. ZHANG *et al.* (2011) conducted a thorough review of the existing research, summarizing key advancements and trends in this area. LONG *et al.* (2022) demonstrated through research that, in bistatic systems, detection performance for configurations such as quadrilateral, hexagonal, rhombic, and checkerboard layouts consistently outperforms that of collocated transmission and reception configurations. SCHENCK *et al.* (1995) developed a hybrid method to predict the complete three-dimensional acoustic scattering function from limited data by using computational models and least-squares problems. Currently, research on underwater bistatic scattering acoustic fields primarily focuses on typical configurations such as collocated transmission and reception, as well as forward scattering. There is relatively little research on the relationship between collocated and bistatic scattering fields. In early radar systems, a method known as the separation theorem was used to estimate the bistatic target strength based on the known monostatic target strength. However, the separation theorem is only applicable for small separation angles. There is currently no well-established method for calculating and relating the scattering echoes in bistatic systems.

This paper investigates the transmit-receive separation echo characteristics for large angles and omnidirectional scenarios, focusing on the prediction of transmit-receive separation scattering acoustic fields based on limited data. A prediction method for transmit-receive separation scattering acoustic fields is established, combining limited data, model geometry properties, numerical integration, acoustic reciprocity, and the least squares method. Using the integral formula of the acoustic field and the surface integral equation, the far-field scattering sound pressure is expressed as the product of the unknown sound source density function and the sound scattering transfer matrix. Through numerical integration, the target surface grid model is used as input to obtain the sound scattering transfer matrix. Based on the calculated sound scattering transfer matrix, the principle of acoustic reciprocity, and the geometric properties of the target, a system of equations for the unknown sound source density function is derived. The unknown density function is then solved using the least squares method. The

calculated sound scattering transfer matrix and the unknown sound source density function are multiplied to obtain the transmit-receive separation scattering acoustic field. The FEM is used to solve the transmit-receive separation scattering acoustic field for a Benchmark model. Several elements from the model are used as input to predict the omnidirectional transmit-receive separation scattering acoustic field. A comparison is made between the predicted results and the FEM results. The simulation results show that the method has good computational accuracy and can be used for calculating the transmit-receive separation scattering acoustic fields of complex underwater targets.

2. Theoretical method

The scattered acoustic field can be expressed as the product of the sound transfer function and the equivalent surface source density function of the target (SCHENCK *et al.*, 1995). As shown in Fig. 1, under the plane wave incidence from the direction $\widehat{\mathbf{x}}^{\text{inc}}$, the sound pressure at the receiver point x can be expressed as

$$p^s(x, \widehat{\mathbf{x}}^{\text{inc}}) = \frac{1}{4\pi} \int_S q(\xi, \widehat{\mathbf{x}}^{\text{inc}}) \left[\frac{\partial}{\partial n_\xi} + i \right] \cdot \frac{e^{-ikr(x, \xi)}}{r(x, \xi)} dS(\xi), \quad (1)$$

where q is the unknown source density function; S represents the target surface; $\widehat{\mathbf{x}}^{\text{inc}}$ is the unit vector in the direction of the incident point; ξ denotes a point on the target surface and $\widehat{\mathbf{n}}$ is the unit normal vector at a point on the target surface. Where $R(x)$ is the distance between the target and the receiver point in the far field, and $r(x, \xi)$ is the distance between a point ξ on the target surface and the receiver point in the near field. The target surface S is discretized into N_s surface elements, Eq. (1) is expressed as

$$p^s(x, \widehat{\mathbf{x}}^{\text{inc}}) = \frac{1}{4\pi} \sum_{l=0}^{N_s} q_l(\widehat{\mathbf{x}}^{\text{inc}}) \int_{s_0} \left[\frac{\partial}{\partial n_\xi} + i \right] \cdot \frac{e^{-ikr(x, \xi)}}{r(x, \xi)} dS(\xi), \quad (2)$$

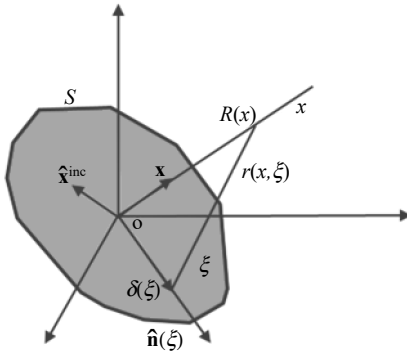


Fig. 1. Schematic diagram of target sound scattering.

when the receiver point x is defined in the far field of the integration surface S , $r(x, \xi) \doteq R(x)$ and divided by the spherical wave propagation factor $e^{-ikR(x)}/R(x)$, the far-field scattered sound pressure is defined as follows:

$$p_{ff}^s(x, \widehat{\mathbf{x}}^{\text{inc}}) = \frac{1}{4\pi} \sum_{l=0}^{N_s} q_l(\widehat{\mathbf{x}}^{\text{inc}}) \cdot \int_S [ik\widehat{\mathbf{x}}\widehat{\mathbf{n}}(\xi) + i] e^{-ik\widehat{\mathbf{x}}\delta(\xi)} dS(\xi). \quad (3)$$

Define the scattering acoustic field matrix \mathbf{S} , where the elements are denoted as $S_{mn} = p_{ff}^s(\widehat{\mathbf{x}}_m, \widehat{\mathbf{x}}_n^{\text{inc}})$. Additionally, define the source density matrix \mathbf{Q} , with the elements represented as $Q_{ln} = q_l(\widehat{\mathbf{x}}_n^{\text{inc}})$, therefore, we have the following relationship:

$$\mathbf{S} = \mathbf{C}^{ff} \mathbf{Q}. \quad (4)$$

The vectors \mathbf{S}_{ex} , \mathbf{Q}_{ex} , and \mathbf{C}_{ex} are expressed as

$$\mathbf{S}_{ex} = [S_{11}, S_{12}, \dots, S_{1n}, S_{21}, S_{22}, \dots, S_{2n}, \dots, S_{m1}, S_{m2}, \dots, S_{mn}]^T, \quad (5)$$

$$\mathbf{Q}_{ex} = [q_{11}, q_{12}, \dots, q_{1n}, q_{21}, q_{22}, \dots, q_{2n}, \dots, q_{m1}, q_{m2}, \dots, q_{mn}]^T, \quad (6)$$

$$\mathbf{C}_{ex} = [\mathbf{C}_e^1, \mathbf{C}_e^2, \dots, \mathbf{C}_e^m]^T, \quad (7)$$

$$\mathbf{C}_e^i = \begin{bmatrix} \mathbf{C}_i & 0 & \dots & 0 \\ 0 & \mathbf{C}_i & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{C}_i \end{bmatrix}, \quad (8)$$

where, the matrix element \mathbf{S} corresponds to the acoustic pressure, \mathbf{C} to the transfer function, and q to the source density function.

The sound scattering matrix \mathbf{S} is an $m \times n$ matrix, the sound scattering transfer matrix \mathbf{C}^{ff} is an $m \times l$ matrix, and the source density matrix \mathbf{Q} is an $l \times n$ matrix. The sound scattering transfer matrix can be computed using the Gaussian-Legendre quadrature method for numerical integration. The sound scattering matrix \mathbf{S} and the source density function matrix \mathbf{Q} can be rewritten as column vectors \mathbf{S}_{ex} and \mathbf{Q}_{ex} , respectively. Simultaneously, the matrix \mathbf{C}^{ff} undergoes a matrix transformation to yield \mathbf{C}_{ex} :

$$\mathbf{S}_{ex} = \mathbf{C}_{ex} \mathbf{Q}_{ex}. \quad (9)$$

According to the principle of acoustic reciprocity, the ratio of the excitation applied at point A to the response at point B is equal to the ratio of the excitation applied at point B to the response at point A. The elements of the sound scattering matrix satisfy:

$$\mathbf{S}_{mn} = \mathbf{S}_{nm}. \quad (10)$$

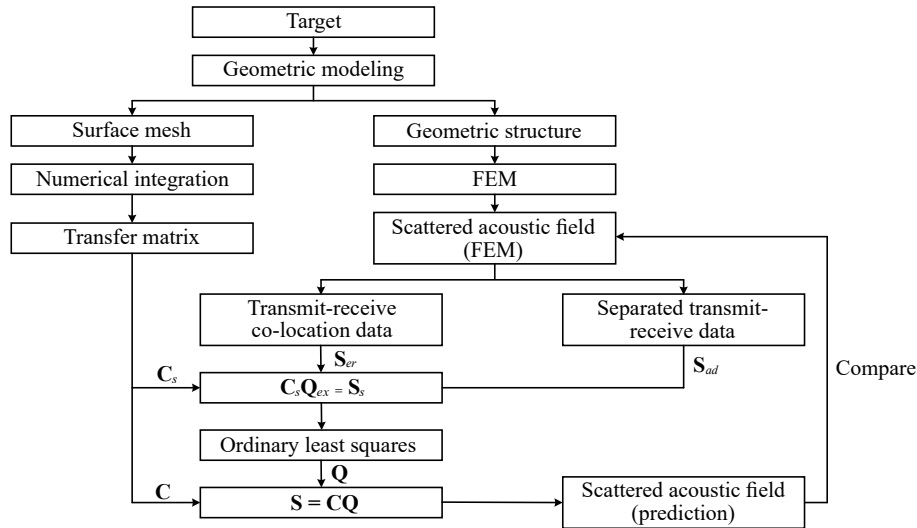


Fig. 2. Flow chart for calculating.

For the bistatic scattering matrix \mathbf{S} , it is a symmetric matrix where the upper triangular elements are equal to the lower triangular elements. By performing elimination while retaining the diagonal elements, $m_1(m_1 + 1)/2$ equations are obtained:

$$\mathbf{S}_{re} = \mathbf{C}_{re} \mathbf{Q}_{re}, \quad (11)$$

where \mathbf{S}_{re} is composed of the monostatic scattered acoustic pressure and a zero vector.

Assuming that the monostatic acoustic scattering matrix \mathbf{S}_{re} is known, a portion of the scattered sound pressure data \mathbf{S}_{ad} from the receive-transmit separated configuration is added to the vector \mathbf{S}_{re} . At the same time, the corresponding row vector in the matrix \mathbf{C}_{ex} that corresponds to \mathbf{S}_{ad} is found. These are then combined to form the matrix \mathbf{C}_{ad} , which is subsequently appended to the matrix \mathbf{C}_{re} . This results in a sound scattering matrix \mathbf{S}_s that contains both monostatic and partial bistatic configurations, as well as the corresponding matrix \mathbf{C}_s is:

$$\mathbf{S}_s = \begin{bmatrix} \mathbf{S}_{re} \\ \mathbf{S}_{ad} \end{bmatrix}, \quad \mathbf{C}_s = \begin{bmatrix} \mathbf{C}_{re} \\ \mathbf{C}_{ad} \end{bmatrix}. \quad (12)$$

The least squares method is used to approximate the solution of Eq. (6), resulting in the source density matrix \mathbf{Q}_{ex} . The obtained column vector \mathbf{Q}_{ex} is then transformed into an $l_1 \times n_1$ matrix \mathbf{Q} . This matrix \mathbf{Q} , along with the calculated bistatic scattering transfer matrix \mathbf{C}_{ml}^{ff} , is substituted into Eq. (3) to obtain the bistatic scattered sound pressure \mathbf{S} . The detailed procedure is illustrated in Fig. 2.

The target strength is calculated using the following expression, assuming an incident acoustic pressure of 1 Pa:

$$TS = 20 \log(\text{abs}(\mathbf{S}_s)). \quad (13)$$

To evaluate how much input data is required to achieve accurate prediction results, the ratio η between

the elements of the input scattered sound pressure data and the predicted scattered sound pressure data can be expressed as

$$\eta = \frac{k(k+1) + 2m_1}{2m_1^2 - k(k+1) - 2m_1} \times 100\%. \quad (14)$$

3. Case studies

The finite element software COMSOL Multiphysics, a multi-physics coupling software, is used to solve the scattering sound field under planar wave incidence. Partial sound scattering data obtained from simulation calculations and the target surface mesh are used as input to predict the separated transmit-receive scattering sound field. Finally, the prediction results are compared and analyzed with the finite element results.

3.1. Cylinder

Figure 3 is a schematic diagram of the triangular mesh model for column targets. The acoustic scatter-

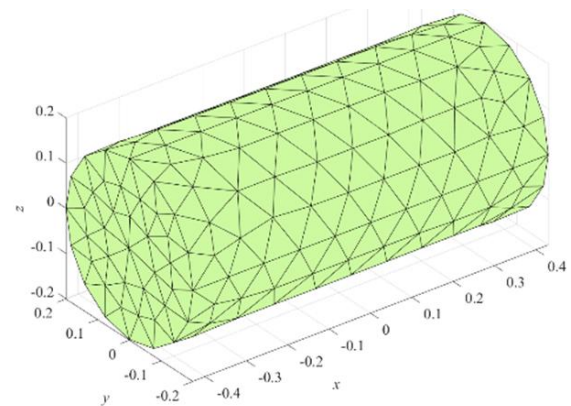


Fig. 3. Cylindrical target surface geometry mesh model.

ing transfer function is calculated according to the grid model C_{ml}^{ff} , the calculation frequency is 100 Hz–1 kHz, the step length is 50 Hz, the incidence angle and receiving angle are 0° – 360° , and the step length is 2° , different quantities of finite element calculation data are taken as input for prediction.

According to the grid model, the calculation conditions for the sound scattering transfer function remain consistent with those described earlier. Predictions are performed using varying amounts of finite element simulation data as input. Bistatic target strength predictions for cylindrical targets are conducted using a rigid cylinder as shown in Fig. 4.

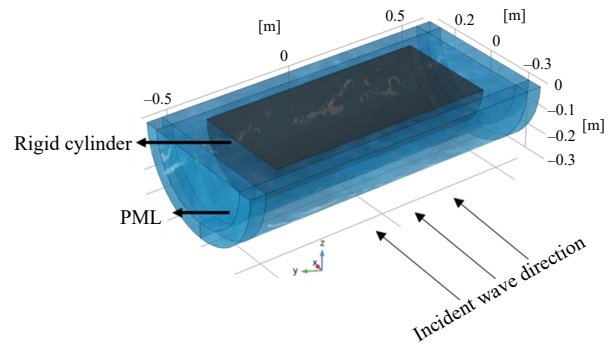


Fig. 4. Schematic diagram of rigid cylinder COMSOL modelling.

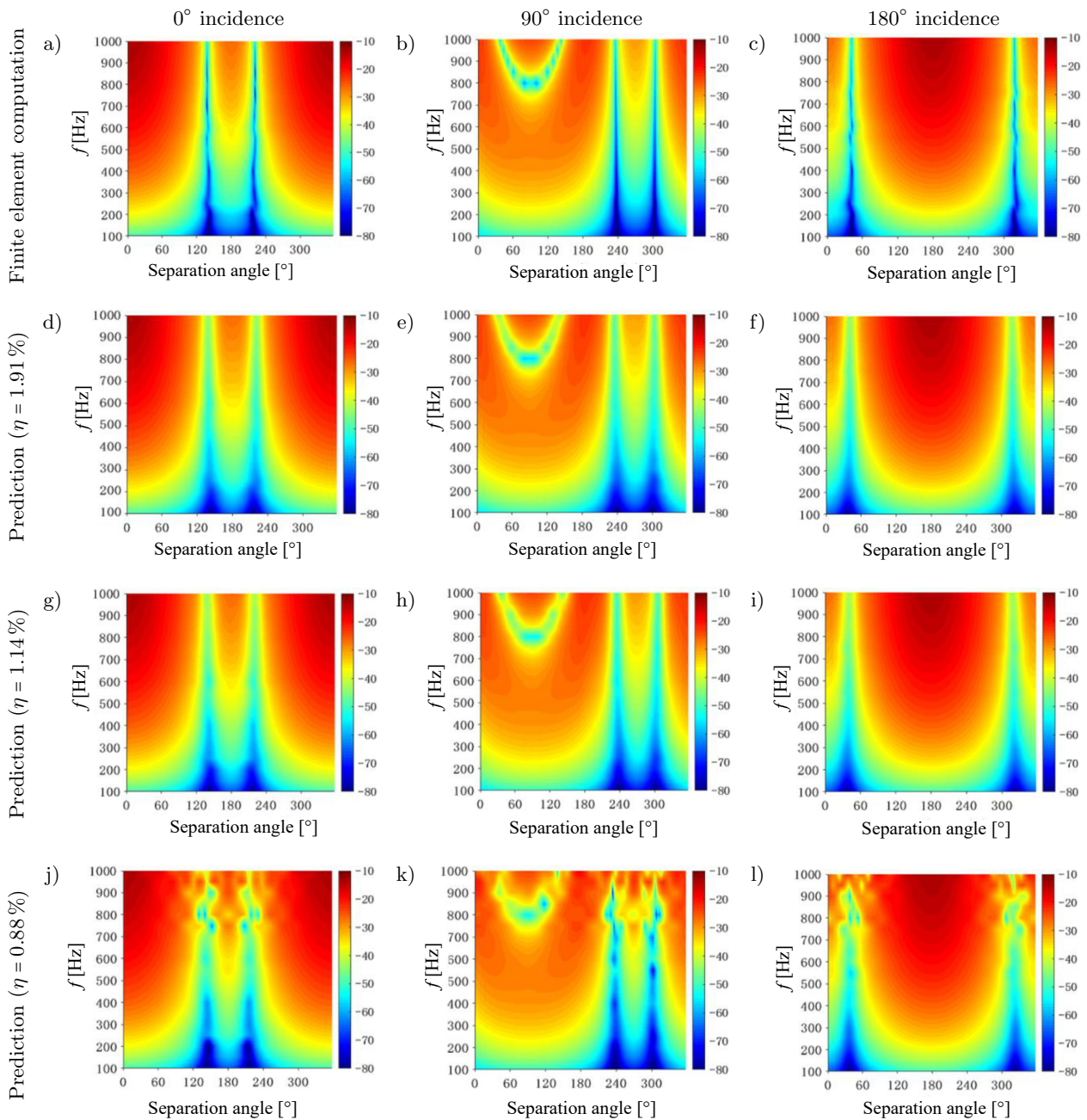


Fig. 5. Comparison of frequency-angle spectra of rigid cylinder predictions and FEM target strength.

The bistatic scattering sound field matrix \mathbf{S} is constructed by sorting out the finite element calculation results, the bistatic scattered sound field is predicted by controlling the ratio η of the input element and the prediction element. Figure 5 shows the comparison results of the target strength frequency-angle spectrum between forecast results and finite element calculation results. Figure 6 compares the target strength predicted by the proposed method and the finite element method under different incident and receiving angles.

3.2. Benchmark model

The applicability of the method discussed in this paper is examined by selecting the Benchmark single-shell model, which has a total length of 62 m. The specific dimensions are shown in Fig. 7.

The underwater vehicle model is subjected to the finite element simulation, where the hull is set as rigid. The incident wave is a harmonic plane wave with unit amplitude in the xOy plane. The calculation conditions are the same as those for the cylinder.

As shown in Fig. 8, the predicted results for the Benchmark scaled model are similar to the prediction trends for the cylindrical model. As the number of input elements increases, the predicted results become more consistent with the analytical solution. However, the prediction accuracy decreases as the frequency increases. Figure 8 presents the bistatic target strength maps for the Benchmark scaled model and the FEM; Figs. 8a–8c show the finite element results, Figs. 8d–8f present the predicted results for $\eta = 3.15\%$; and Figs. 8g–8i show the predicted results for $\eta = 1.91\%$. Figure 9 compares the predicted and finite element computed bistatic target strength maps for the Benchmark scaled model.

4. Conclusion

This paper presents a prediction method for bistatic scattering sound fields based on limited data. The bistatic scattering sound field is expressed as the product of the target's sound scattering transfer function and the source density function. The target surface mesh is used as input to numerically integrate

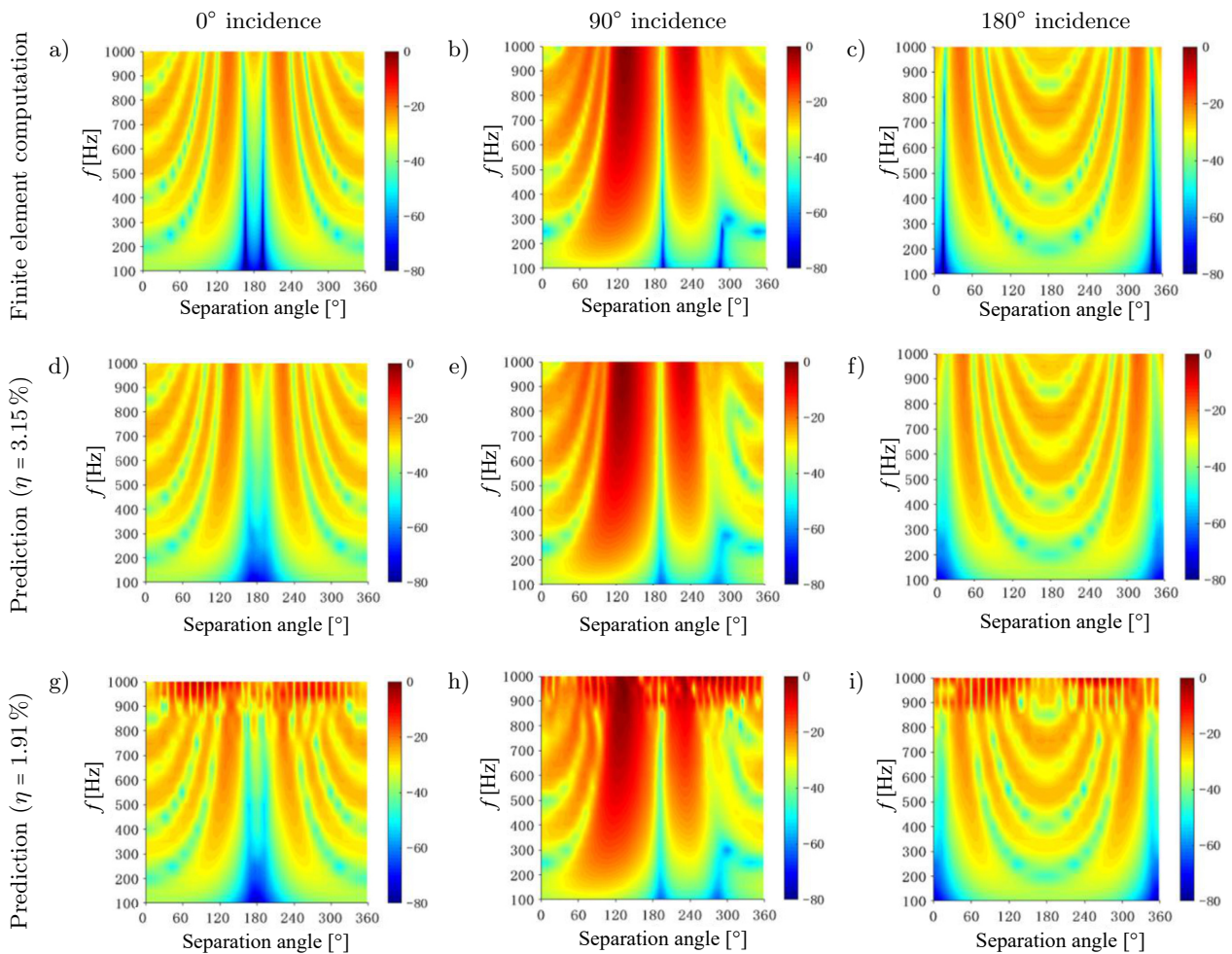


Fig. 8. Comparison of the predicted and finite element target strength frequency-angle spectra for the Benchmark scaled model.

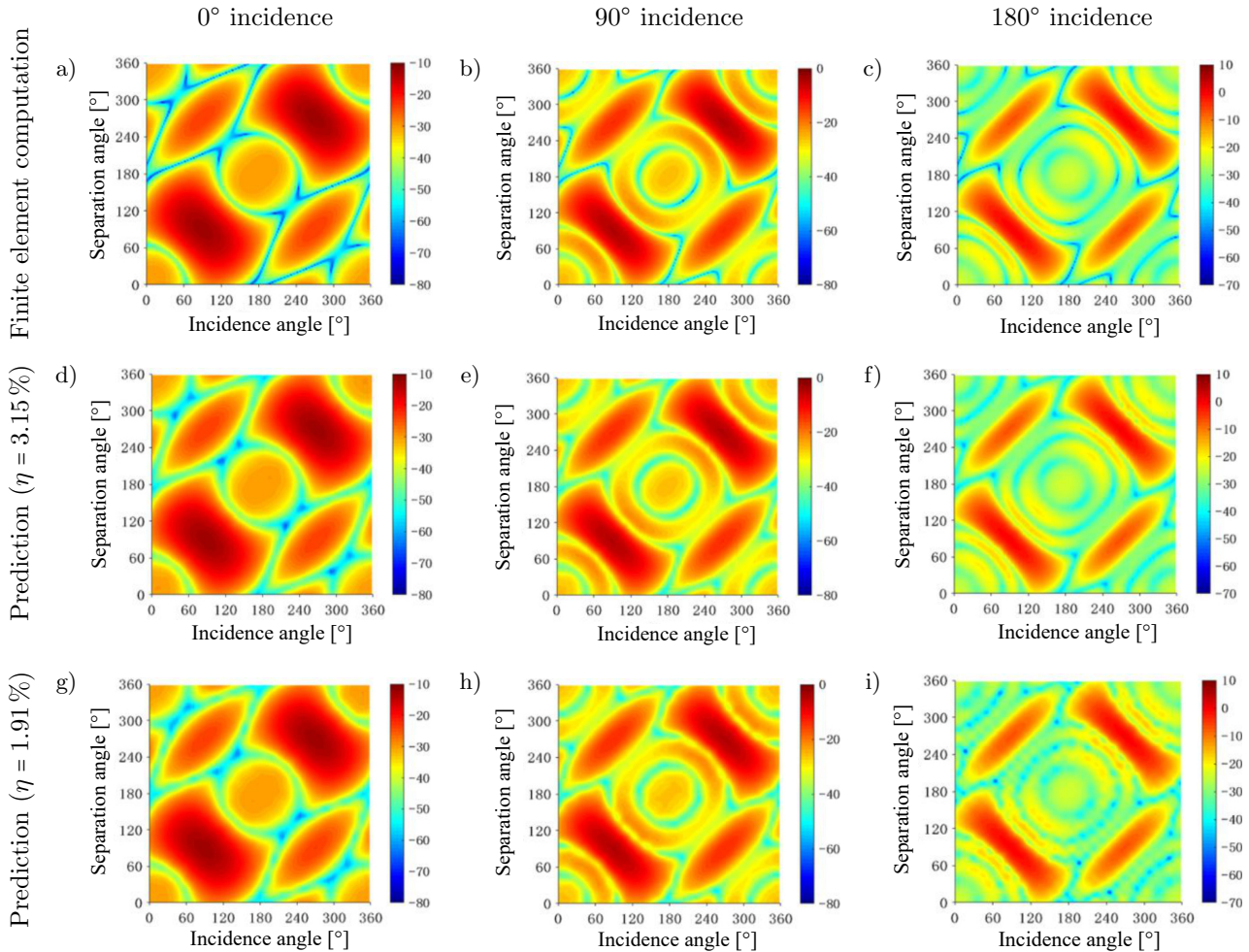


Fig. 9. Comparison of the predicted and finite element computed bistatic target strength maps for the Benchmark scaled model.

and obtain the bistatic target sound scattering transfer function. Using the known co-located transmit-receive scattered sound pressure, n sets of separated transmit-receive scattered sound pressures, and their corresponding sound scattering transfer functions as input, the source density function is solved using the least squares method. This method establishes a connection between single-base and bistatic scattering sound fields, offering significant practical value in acoustic scattering experiments, numerical calculations, and underwater countermeasure applications. The proposed prediction method has the following features:

- 1) This prediction method is applicable to low-frequency bistatic scattering sound field forecasting. In theory, it can be used to calculate the scattering sound field for any complex target model without relying on the internal structure of the target. It only requires the input of the surface geometric mesh and known scattered sound pressure data to predict the bistatic scattering sound field for targets with complex structures. The method demonstrates good computational ef-

ficiency for low-frequency, small target bistatic scattering sound field calculations.

- 2) The prediction accuracy of this method is dependent on both the precision and location of the input elements. The higher the precision of the input elements, the more accurate the prediction. Additionally, the location of the input elements has a significant impact on the prediction accuracy. When the known input elements correspond to the peaks of the scattering sound field, the prediction performs better. For bistatic scattering sound fields, the strongest scattered echoes typically occur in the forward scattering direction of the target, while strong echoes are also found in the backscattering direction, i.e., the co-located transmit-receive direction. Furthermore, scattering echoes from the mirrored reflection directions are also significant. In contrast, the scattering echo in the vertical direction of the incident wave is relatively weak. Therefore, using elements from these directions as known data for prediction results in higher computational efficiency.

- 3) The method proposed in this paper estimates the target strength of any complex shape and structure at any given angle of view based on a small amount of known data. It can be used for the engineering prediction of bistatic target strength for complex underwater targets and is applicable in underwater bistatic experiments. This method only requires the target surface mesh, making the modeling process simple and the calculation speed relatively fast, which holds certain potential application value in numerical calculations and underwater countermeasures. However, this method is suitable for low-frequency bistatic scattering sound field prediction, and the accuracy of the predictions is dependent on the position and precision of the input elements.

FUNDINGS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CONFLICT OF 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.

References

1. AGOUNAD S., DÉCULTOT D., CHATI F., LÉON F., KHANDOUCH Y. (2023), Experimental study of the bistatic acoustic scattering from cylindrical shell, *Mechanical Systems and Signal Processing*, **186**: 109892, <https://doi.org/10.1016/j.ymsp.2022.109892>.
2. CHEN C.X., PENG Z.L., SONG H., XUE Y.Q., ZHOU F.L. (2024), Prediction method of the low-frequency multi-static scattering sound field for underwater spherical targets based on limited data [in Chinese], *Journal of Shanghai Jiaotong University*, **58**(7): 1006–1017.
3. CHENG G., ZHANG M., DU K., YAO W. (2010), Model and experiments for bistatic scattering strength of underwater objects, [in:] *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, **9**: V9-489-V9-492, <https://doi.org/10.1109/ICCASM.2010.5622985>.
4. FAN J., ZHOU L.K. (2006), Graphical acoustics computing method for echo characteristics calculation of underwater targets [in Chinese], *Acta Acustica*, **31**(6): 511–516.
5. GU R., PENG Z., XUE Y., XU C., CHEN C. (2025), An improved fast prediction method for full-space bistatic acoustic scattering of underwater vehicles, *Sensors*, **25**(8): 2612, <https://doi.org/10.3390/s25082612>.
6. GUNDERSON A.M., ESPAÑA A.L., MARSTON P.L. (2017), Spectral analysis of bistatic scattering from underwater elastic cylinders and spheres, *The Journal of the Acoustical Society of America*, **142**(1): 110–115, <https://doi.org/10.1121/1.4990690>.
7. LIU Y.C., ZHANG M.M., WANG Y.L. (2012), Calculation of target bistatic acoustic scattering using physical acoustic method [in Chinese], *Journal of Naval University of Engineering*, **24**(4): 101–103.
8. LONG L.Y., ZHAO H.S., LI D. (2022), Analysis of array configuration design for multi-base sonar buoy systems, *Acoustics and Electronics Engineering*, **2022**(1): 6–9.
9. MENG Z.R., SONG Y., YU F.X., ZHANG C.L., ZHANG Y.Q. (2024), Highlight model of underwater target acoustic scattering in the bistatic system, *Journal of Computers*, **35**(2): 231–249, <http://doi.org/10.53106/199115992024043502015>.
10. PARK S.H. et al. (2006), Bistatic scattering from a hemi-spherically capped cylinder, *The Journal of the Acoustical Society of Korea*, **25**(3E): 115–122.
11. PENG Z.L., WANG B., FAN J. (2018), Simulation and experimental studies on acoustic scattering characteristics of surface targets, *Applied Acoustics*, **137**: 140–147, <https://doi.org/10.1016/j.apacoust.2018.02.014>.
12. SCHENCK H.A., BENTHIE G.W., BARACH D. (1995), A hybrid method for predicting the complete scattering function from limited data, *The Journal of the Acoustical Society of America*, **98**(6): 3469–3481, <https://doi.org/10.1121/1.413779>.
13. SCHMIDT H. (2001), Bistatic scattering from buried targets in shallow water, [in:] *Proceedings, GOATS 2000 Conference*, La Spezia, **2001**: 21–22.
14. SCHNEIDER J.B., WAGNER C.L., BROCHAT S.L. (1998), Implementation of transparent sources embedded in acoustic finite-difference time-domain grids, *The Journal of the Acoustical Society of America*, **103**(1): 136–142, <https://doi.org/10.1121/1.421084>.
15. TANG W.L., FAN J., MA Z.C. (2018), *Acoustic Scattering of Underwater Targets* [in Chinese], Beijing: Science Press.
16. WANG B., WANG W.H., FAN J., ZHAO K.Q., ZHOU F.L., TAN L.W. (2022), Modeling of bistatic scattering from an underwater non-penetrable target using a Kirchhoff approximation method, *Defence Technology*, **18**(7): 1097–1106, <https://doi.org/10.1016/j.dt.2022.04.008>.
17. WATERMAN P.C. (2005), New formulation of acoustic scattering, *The Journal of the Acoustical Society of America*, **45**(6): 1417–1429, <https://doi.org/10.1121/1.1911619>.
18. XUE Y.Q., PENG Z.L., YU Q., ZHANG C., ZHOU F., LIU J. (2023), Analysis of acoustic scattering characteristics of underwater targets based on Kirchhoff approximation and curved triangular mesh [in Chinese], *Acta Armamentarii*, **44**(8): 2424–2431.
19. ZHANG L.L., YANG R.J., YANG C.Y. (2011), Research on underwater maneuvering target tracking technology [in Chinese], *Technical Acoustics*, **30**(1): 68–73, <https://doi.org/10.3969/j.issn1000-3630.2011.01.012>.
20. ZHOU L.K., FAN J., TANG W.L. (2009), Analyzing acoustic scattering of elastic objects using coupled FEM-BEM technique [in Chinese], *Journal of Shanghai Jiaotong University*, **43**(8): 1258–1261.

Research Paper

Mapping Sound Pressure Levels: A Novel Approach to Determining Near-field and Far-field Regions

Iliyan Yordanov ILIEV^{(1)*}, Hristo Zhivomirov KARAIIVANOV⁽²⁾⁽¹⁾ Department of Information Technology, Nikola Vaptsarov Naval Academy
Varna, Bulgaria⁽²⁾ Department of Theory of Electrical Engineering and Measurements, Technical University of Varna
Varna, Bulgaria*Corresponding Author e-mail: i.y.iliev@naval-acad.bgReceived May 30, 2024; revised January 25, 2025; accepted February 5, 2025;
published online April 2, 2025.

The present study focuses on the spatial characteristics of the sound pressure level (SPL) generated by a circular piston (a circular-shaped acoustic transducer or loudspeaker). It presents a short theoretical review to aid understanding the primary sound field characteristic – acoustic pressure – as a function of time, frequency, directivity angle, and distance from the source. The study introduces a simple and practical criterion for determining the near- and far-field boundary along the axis of the circular piston as a function of frequency. This criterion is validated through theoretical analysis and experimental measurements. Overall, the results show the influence of circular piston parameters on the SPL spatial distribution.

Keywords: sound pressure level (SPL); spatial characteristics; near-field; far-field; boundaries; criterion.

Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The study of sound's spatial characteristics plays a critical role in numerous engineering and medical applications (STEFANOWSKA, ZIELIŃSKI, 2024). Over the years, extensive research has been conducted on sound propagation in different environments, such as water (VAN GEEL *et al.*, 2022) and air (ILIEV, ZHIVOMIROV, 2015; KUDRIASHOV, 2017).

Additionally, various experiments have explored sound produced by different sources, including unconventional ones (ÖZTÜRK, TIRYAKIOGLU, 2020).

To better understand the sound field generated by an acoustic transducer, its spatial characteristics must be examined using both conventional and innovative methods (KLIPPEL, BELLMAN, 2016; SHI *et al.*, 2022). In this context, the sound pressure p_a is the physical quantity (parameter) that provides the most detailed information about the structure of the sound field in the space.

Theoretical analysis indicates that the SPL near a circular piston is not uniformly distributed. In classical acoustics, two primary regions are distinguished

around an acoustic source: the near-field and far-field. A third region, known as the very near-field, is primarily the subject of research in vibration mechanics. In some references, the near-field is referred to as the Fresnel zone, while the far-field is called the Fraunhofer zone. Each zone has distinct formulas for determining the sound pressure p_a . This raises questions about the delineation of the boundaries defining various zones and the corresponding SPL variations within each zone. Equally crucial is pinpointing the moment at which the SPL stabilizes, ensuring accurate sound reproduction.

This paper builds upon the author's previous research (ILIEV, ZHIVOMIROV, 2015), which provided a theoretical overview of established mathematical techniques for calculating the SPL generated by a circular piston in both near-field and far-field conditions. Expanding on that foundation, this work proposes a unified framework to define the boundary between the near- and far-field zone of a circular piston.

Understanding the formation of the SPL spatial structure around the acoustic transducer enhances its

design and application precision, especially in monitoring and screening systems.

2. Background

As shown by KINSLER *et al.* (2000) the total acoustic pressure p_a generated by a circular piston at an arbitrary point A (Fig. 1) can be obtained using the well-known Huygens–Fresnel principle (also referred to as the Rayleigh integral):

$$p_{a(\text{entire})}(\theta, f, r, t) = j\rho_s f \nu_m \int_Q \frac{1}{l} e^{j(2\pi ft - \frac{2\pi fl}{c})} dQ, \quad (1)$$

where θ – elevation angle; f – frequency of the sound signal; r – distance between the circular piston and observation point A ; t – time; ρ_s – density of the medium; ν_m – deformation amplitude the transducer surface; Q – surface area of the circular piston; l – distance between the elementary section dQ and observation point A ; c – speed of sound in the medium.

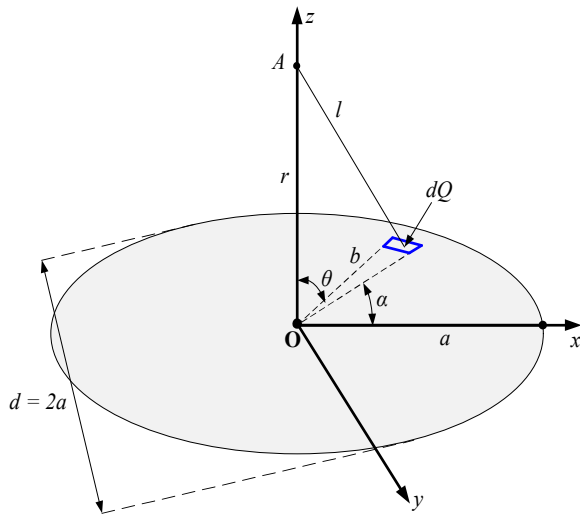


Fig. 1. Determination of the SPL in front of a circular piston.

Equation (1) is not suitable for direct practical implementation. However, the entire acoustic pressure $p_{a(\text{entire})}$, generated by a circular piston, can also be computed using the authors' modified expression (ILIEV, 2014):

$$p_{a(\text{entire})}(\theta, f, r, t) = \rho_s f \nu_m e^{j2\pi ft} \int_0^a b db \int_0^{2\pi} \frac{e^{-\frac{j2\pi f}{c}(\sqrt{r^2+b^2+2rb\sin\theta\cos\alpha})}}{\sqrt{r^2+b^2+2rb\sin\theta\cos\alpha}} d\alpha, \quad (2)$$

where a – radius of the circular transducer; b – radial distance between the elementary section (point source) dQ and the center of the circular piston O ; α – directivity angle (azimuth).

A simplified solution of Eq. (2) allows for calculating the sound pressure in the far-field $p_{a(\text{far})}$ (when $r \gg d$) on the axis of the circular piston (when $\theta = 0$), as follows (KINSLER *et al.*, 2000):

$$p_{a(\text{far})}(0, f, r, t) = \frac{\rho_s f \nu_m}{r} e^{j2\pi ft} \cdot \int_0^a b db \int_0^{2\pi} e^{-\frac{j2\pi fr}{c}} d\alpha. \quad (3)$$

Also, a simplified version of Eq. (2) for the near-field SPL $p_{a(\text{near})}$, on the axis of the circular piston, is given (KINSLER *et al.*, 2000):

$$p_{a(\text{near})}(0, f, r) = 2\rho_s f \nu_m \cdot \left| \sin \left\{ \frac{\pi fr}{c} \left[\sqrt{1 + \left(\frac{a}{r}\right)^2} - 1 \right] \right\} \right|. \quad (4)$$

Equation (4) is graphically represented in Fig. 2, evaluated for various distances and frequencies for a circular piston with $a = 0.08$ m. This provides insight into the spatial structure of the near-field SPL.

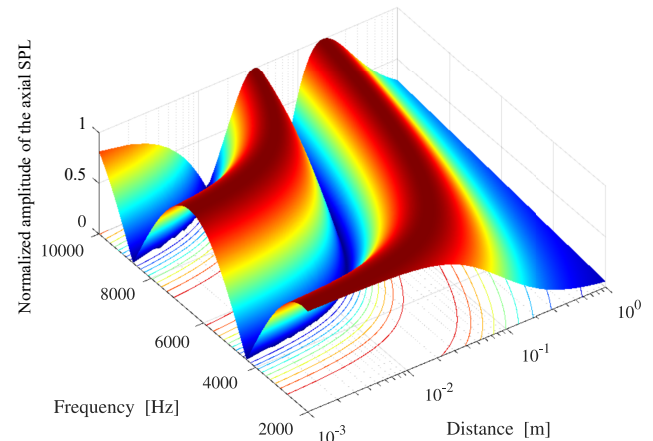


Fig. 2. Axial near-field SPL structure for various distances and frequencies for a circular piston with radius $a = 0.08$ m, evaluated by Eq. (4). The pulsations of the SPL are evident.

Equation (4) reveals that extrema in the axial SPL occur due to the sine function. For a constant sound speed c and a specific circular piston with radius a , the sine function reaches extremum values at distances (LEPENDIN, 1978):

$$r_m = \frac{a^2 f}{mc} - \frac{mc}{4f}, \quad (5)$$

where $m \in \mathbf{N}$. The maxima occur when m is odd, and the minima when m is even. The first SPL maximum (moving toward the circular piston axis) appears for $m = 1$, at a distance:

$$r_{\text{max}(\text{first})} = \frac{a^2 f}{c} - \frac{c}{4f}. \quad (6)$$

This relationship is depicted in Fig. 3. It turns out that it represents the contour of the outer ridge of the SPL structure shown in Fig. 2.

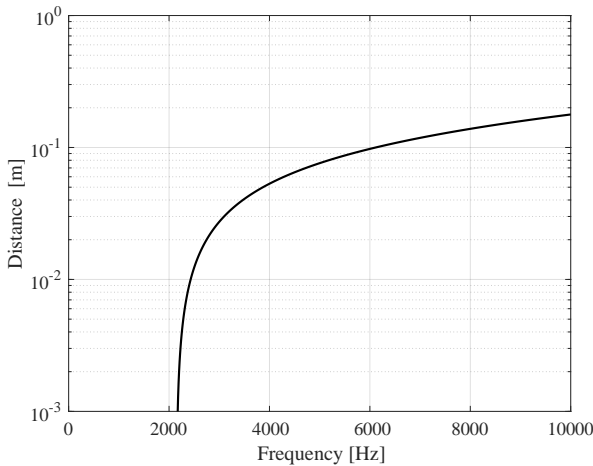


Fig. 3. Distance of the first axial SPL maximum, evaluated by Eq. (6), vs. frequency for a circular piston with $r = 0.08$ m. The plot can also address the problem of identification of the frequency of the first axial SPL maximum in relation to distance (see Eq. (7)).

For distances $r < r_{\max(\text{first})}$ the axial SPL suffers from maxima and minima, and for distances $r > r_{\max(\text{first})}$ the SPL decreases monotonically, approaching an asymptotic dependence $1/r$ (ROSSING, 2017). Therefore, one may consider the distance $r_{\max(\text{first})}$ as a reasonable threshold or dividing line between regions where the SPL is not completely formed and where the SPL becomes asymptotic. It is evident that r_m is frequency-dependent – the higher the frequency, the longer the near-field zone of the transducer.

Moreover, Eq. (6) reveals that the distance $r_{\max(\text{first})}$ has physical meaning only for frequencies $f > \frac{c}{2a}$ (ILIEV, 2014). For those frequencies, the radiation of the circular piston resembles that of a simple source, that is, without extrema.

Consequently, from Eq. (6), one can derive a relationship to determine the frequency of the first maximum:

$$f_{\max(\text{first})} = \frac{c(\sqrt{a^2 + r^2} + r)}{2a^2}. \quad (7)$$

It should be noted that different authors (KLEINER, 2013; GELFAND, 2017) have proposed various expressions (similar to Eq. (6)) to estimate the last maximum, which some consider the upper border distance of the near-field. In (KOZIEŃ, 2012), the hybrid intensity method is used to determine the boundary between the near and far fields. However, these methods have some disadvantages. The first (KLEINER, 2013; GELFAND, 2017) is too rough and lacks accuracy. The second method (KOZIEŃ, 2012) is too complicated and less practical.

Currently, there is no established criterion for delineating all zones, each with its unique characteristics, that influence the distribution and formation of the SPL in front of the circular piston.

3. Boundary between the near- and far-field

Identifying the region where sound pressure is fully developed provides valuable insights into the practical utility of a circular piston. Conversely, employing a circular piston at frequencies and distances where the SPL is not fully formed yields unsatisfactory outcomes.

In this paper, the authors introduce an enhanced method that not only determines the boundary between the fully developed acoustic field and the area where the acoustic field suffers from interferences, but also provide a new interpretation of the zones and their characteristics in front of the circular piston. The latter is based on the normalized difference between the axial SPLs calculated by Eqs. (2) and Eq. (3):

$$\Delta p_a(0, f, r) = \frac{p_{a(\text{entire})} - p_{a(\text{far})}}{p_{a(\text{far})}}. \quad (8)$$

For a given circular piston, the boundary can be identified in terms of distance and frequency when both Eq. (2) and Eq. (3) yield similar results.

The axial overall sound pressure level produced by a circular piston with a radius $a = 0.08$ m, as determined by Eq. (2), is shown in Fig. 4. Additionally, the axial far-field SPL, derived from Eq. (3), is depicted in Fig. 5. Both equations are numerically evaluated in the MathCAD environment and graphically represented in MATLAB[®]. The normalized difference between the entire field SPL and the far-field SPL on the axis, calculated by Eq. (8), is presented in Fig. 6.

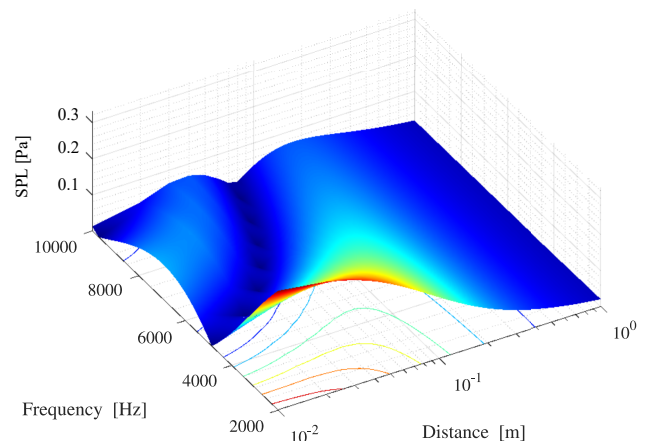


Fig. 4. Axial overall field SPL, evaluated by Eq. (2), for a circular piston with $r = 0.08$ m.

The authors propose that the distance where the SPL difference $\Delta p_a < 30\%$ (using the well-known 3 dB rule) should be considered the dividing line between the area with unstable SPL (i.e., the near-field zone)

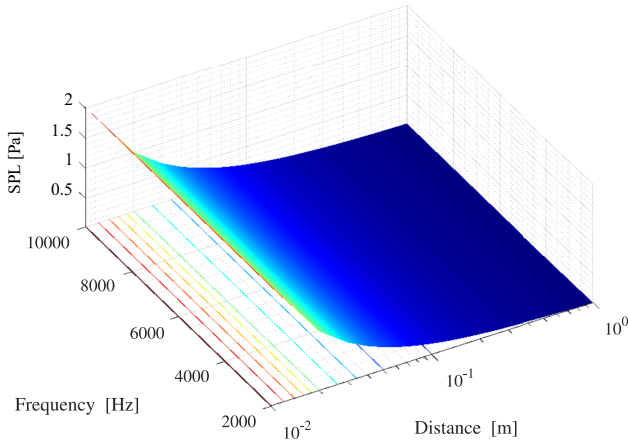


Fig. 5. Axial far-field SPL, evaluated by Eq. (3), for a circular piston with $r = 0.08$ m.

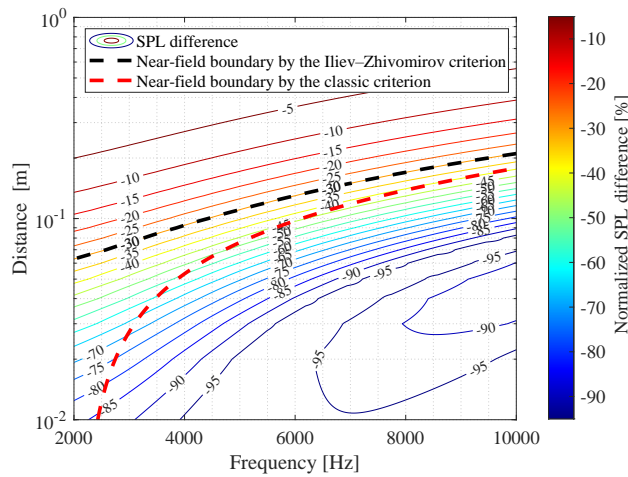


Fig. 6. Normalized SPL difference for a circular piston with $r = 0.08$ m, calculated by Eq. (8). The black dashed line depicts the boundary between the near- and far-field according to the newly proposed criterion. The red dashed line represents the outer ridge of the SPL structure, evaluated by Eq. (4).

and the area with asymptotic SPL (i.e., the far-field zone). It is evident that the boundary calculated by this rule strongly differs from the generally accepted rule-of-thumb, which is based on the outer ridge of the SPL structure evaluated by Eq. (4).

When applying the proposed method to distinguish the dividing line between the interferential area and the asymptotic area of the SPL for a circular piston with a radius $a = 0.08$ m, one observes interferences for frequencies $f > 2147$ Hz up the upper-frequency limit of the circular piston – 10 000 Hz. By applying the 30 % difference criterion between the SPLs calculated by Eq. (2) and Eq. (3), one can determine the distance r for each frequency where the SPL experiences interferences. For example, at $f = 5000$ Hz, using the isobars from Fig. 6, the interference region is found for $r < 0.11$ m (close to $r_{\max(\text{first})}$).

4. Experimental results

An experiment was conducted to validate the theoretical statements and provide a semi-quantitative, intuitive understanding of the proposed method, without claiming accuracy. In future work, more precise results will be obtained using numerical simulations and accurate measurements in a controlled environment.

The object of the experiment is a circular piston (loudspeaker) with a radius $a = 0.08$ m, identical to the loudspeaker with the same radius used for the theoretical analysis. For this loudspeaker, and taking into account the aforementioned factors, one may expect extremes in the SPL to occur between 2147 Hz and the maximum reproducible frequency of this loudspeaker model – 10 000 Hz.

Furthermore, for distances greater than $r > 0.065$ m at $f > 2147$ Hz and $r > 0.21$ m at $f = 10\,000$ Hz, the SPL is expected to decrease monotonically.

The measurements of the axial SPL are carried out using an active sound-level meter, DAQ-system, laptop, and MATLAB[®] software developed by the authors. The measurements are performed using a sine-wave signal with a frequency of $f = 5000$ Hz at eight different distances: $r = \{0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$ m. The results are shown in Fig. 7 and are compared with the theoretical axial SPL in the near-field, calculated by Eq. (4).

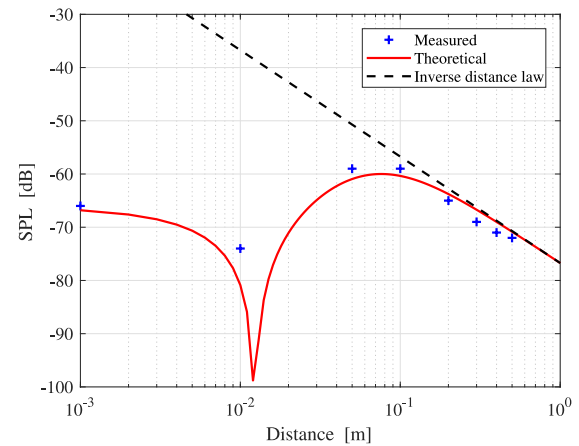


Fig. 7. Axial SPL at $f = 5000$ Hz theoretically estimated (solid red line) and measured at distances $r = \{0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$ m (blue crosses).

There is a well-pronounced overlap between the measured SPL (represented by blue crosses) and the predicted theoretical SPL (solid red line). Noticeable extrema in the axial SPL occur at $r = 0.012$ m (SPL minimum) and $r = 0.076$ m (SPL maximum), which are theoretically predicted by Eq. (5) for this particular operating frequency. For relatively large distances r , the theoretical and the experimental SPLs decrease monotonically, approaching an asymptotic dependence $1/r$ (black dashed line), in agreement with

Eqs. (3) and (4), and the proposed interpretation. In this particular case, the near-field boundary is situated at $r = 0.11$ m, according to Eq. (8). Beyond this point, the SPL decreases strictly monotonically, as expected.

5. Conclusion

The newly proposed method for distinguishing the dividing line between the interference and asymptotic areas of the SPL produced by a circular piston is relatively simple, easy to apply, and practically accurate. If one knows the radius a and frequency f of the circular piston, one can implement Eqs. (2), (3), and (8) to generate a graph similar to those shown in Fig. 6.

From the presented review of the different areas in front of the circular piston, the following conclusions can be drawn:

- the larger the circular piston radius a , the longer the interference area;
- the higher the frequency f , the longer the interference area;
- the higher the sound speed c , the shorter the interference area.

Furthermore, when dealing with complex signals exhibiting a complex and non-stationary spectral composition, a region of pronounced distortions in the sound field emerges within a specific area in front of the circular piston (specifically, between 0 m and 0.21 m in this case). This occurs due to the fact that for certain spectral components, this region functions as a near-field zone. In contrast, for others, it represents a far-field zone.

In this context, it becomes highly significant to define an ‘unconditional far-field zone’, which is determined by the maximum frequency of the considered signal or the maximum reproducible frequency of the circular piston, whichever is lower. Likewise, the concept of an ‘unconditional near-field zone’ needs to be introduced, with its boundary defined by either the psychical characteristics of the circular piston ($f > \frac{c}{2a}$) or by the lowest frequency within the signal spectrum, whichever is higher. The intermediate region between these two zones can be referred to as the ‘transitory zone’.

The results shown in Figs. 2, 3, 4, 5, and 7 can be reproduced by the user using the MATLAB[®] scripts and data given in (Zhivomirov, Iliev, 2024) as supplementary material. This allows for a better understanding of the visualization and the ability to reuse or adapt the code for specific user data.

The conclusions drawn in this paper should be considered in the development of various screening systems (such as sonography, sound localization systems, audio systems, etc.) and in the development of systems using circular pistons in general.

References

1. GELFAND S.A. (2017), *Hearing: An Introduction to Psychological and Physiological Acoustics*, 6th ed., CRC Press, USA, <https://doi.org/10.1201/9781315154718>.
2. ILIEV I. (2014), Polar response of a circular piston, *TEM Journal*, **3**(3): 230–234.
3. ILIEV I., ZHIVOMIROV H. (2015), On the spatial characteristics of a circular piston, *Romanian Journal of Acoustics and Vibration Journal*, **12**: 29–34.
4. KINSLER L.E., FREY A.R., COPPER A.B., SANDERS J.V. (2000), *Fundamentals of Acoustics*, 4th ed., Wiley, USA.
5. KLEINER M. (2013), *Electroacoustics*, Taylor & Francis, USA, <https://doi.org/10.1201/b13859>.
6. KLIPPEL W., BELLMAN C. (2016), Holographic near-field measurement of loudspeaker directivity, *Journal of the Audio Engineering Society*, **141**: 9598.
7. KOZIEŃ M. (2012), Acoustic nearfield and farfield for vibrating piston in geometrical and intensity formulations, *Acta Physica Polonica A*, **121**: 132–135, <http://doi.org/10.12693/APhysPolA.121.A-132>.
8. KUDRIASHOV V. (2017), Improvement of range estimation with microphone array, *Journal of Cybernetics and Information Technologies*, **17**(1): 113–125, <https://doi.org/10.1515/cait-2017-0009>.
9. LEPENDIN L. (1978), *Acoustics* [in Russian], pp. 264–267, Graduate School, Moscow, USSR.
10. ÖZTÜRK H., TIRYAKIOĞLU B. (2020), Radiation of sound from a coaxial duct formed by a semi-infinite rigid outer cylinder and infinite inner cylinder having different linings, *Archives of Acoustics*, **45**(4): 655–662, <https://doi.org/10.24425/aoa.2020.135253>.
11. ROSSING T.D. (2017), *Handbook of Acoustics*, pp. 86–91, Springer, USA.
12. SHI T., BOLTON J.S., THOR W. (2022), Acoustic far-field prediction based on near-field measurements by using several different holography algorithms, *Journal of the Acoustical Society of America*, **151**(3): 2171–2180, <https://doi.org/10.1121/10.0009894>.
13. STEFANOWSKA A., ZIELIŃSKI S. (2024), Spatial sound and emotions: A literature survey on the relationship between spatially rendered audio and listeners’ affective responses, *INTL Journal of Electronics and Telecommunications*, **70**(2): 293–300, <https://doi.org/10.24425/ijet.2024.149544>.
14. van GEEL N.C.F., RISCH D., WITTICH A. (2022), A brief overview of current approaches for underwater sound analysis and reporting, *Marine Pollution Bulletin*, **178**: 113610, <https://doi.org/10.1016/j.marpolbul.2022.113610>.
15. ZHIVOMIROV H., ILIEV I. (2024), Mapping sound pressure level of a circular piston (supplementary material), Mendeley Data, V2, <https://doi.org/10.17632/b89rhwhrrk.2>.

Research Paper

An IT System for Assessing Noise Level Distribution in Historical Urban Centers

Mykhaylo MELNYK⁽¹⁾, Tadeusz KAMISIŃSKI^{(2)*}, Roman KINASZ⁽³⁾, Roman VYNAROVYCH⁽¹⁾

⁽¹⁾ *Department of Computer Aided Design Systems, Lviv Polytechnic National University
Lviv, Ukraine*

⁽²⁾ *Department of Mechanics and Vibroacoustics,
Faculty of Mechanical Engineering and Robotics, AGH University of Krakow
Kraków, Poland*

⁽³⁾ *Department of Civil and Geotechnical Engineering and Geomechanics,
Faculty of Civil Engineering and Resource Management, AGH University of Krakow
Kraków, Poland*

*Corresponding Author e-mail: kamisins@agh.edu.pl

Received December 8, 2024; accepted February 16, 2025;

published online May 5, 2025.

This article explores the challenge of identifying noise-generating factors in traffic flows (TFs) within the constrained spaces and imperfect transport networks of historical cities, using Lviv as a case study. Experimental studies were conducted to measure the equivalent noise levels at different times of the day on selected streets in Lviv. These streets are characterized by dense development, paved surfaces, and a high volume of vehicular and rail traffic. The study identified correlations between noise levels, traffic volumes, and vehicle speeds during daytime and nighttime periods. Notably, vehicle speed was found to have a more significant impact on noise levels than the number of vehicles.

Through the analysis of these findings, empirical mathematical models were developed and validated using the Lagrange interpolation polynomial to predict noise pollution levels on selected streets at specific times. The developed computer system enables quick forecasting of noise levels for a given street while simultaneously provides data to manage TF as a factor affecting noise generation. Crucially, this tool can also assist in calculating the required specifications for acoustic insulation on building façades adjacent to these TFs.

Keywords: IT system; noise measurement; road noise; MATLAB; equivalent noise level.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The continued growth in vehicle numbers and the intensification of traffic flows (TFs) within the constrained urban spaces and flawed transport networks contribute to escalating noise pollution issues in most large cities. Addressing these challenges proves particularly difficult in the city of Lviv, especially in its central, historical district, which is dense with intersections and narrow streets. Elevated noise levels pose a significant public health threat that directly impacts the health of urban residents. Consequently, designing an efficient passenger transport system in such settings

necessitates the development and implementation of sophisticated software models that account not only for TFs but also for all contributing factors.

In an effort to assess and evaluate the environmental impact of transport, experimental studies were proposed and conducted to measure equivalent noise levels at various times of the day, focusing on selected streets in Lviv as representative examples of old historical cities. The chosen research locations are characterized by dense urban development along roads with heavy traffic on deteriorating cobblestone surfaces and tram routes, revealing the compounded challenges of managing urban noise in such environments.

2. Analysis of the latest research and publications

The impact of road noise on urban areas has been the subject of in-depth global research due to its significant effects on public health and quality of life. Across Europe, substantial studies have examined the contributing factors of noise pollution and proposed potential mitigation strategies (BROWN, 2015; OW, GHOSH, 2017; TITU *et al.*, 2022).

Research by OZER *et al.* (2008) and ZAMBON *et al.* (2018) identify urbanization, population growth, and the associated surge in vehicular traffic as primary contributors to rising noise levels, particularly in urban settings. Such noise adversely affects human health, leading to sleep disorders, cardiovascular diseases, and cognitive impairments (BASNER *et al.*, 2014; DZHAMBOV, LERCHER, 2019; HEGEWALD *et al.*, 2020).

Further studies, such as those by PETRESCU *et al.* (2015), recognize road characteristics (notably road surface conditions), along with traffic intensity, speed, and vehicle type as major noise sources in urban environments. Several mitigation strategies, including the implementation of quieter road surfaces, have been explored.

In Ukraine, the study of urban noise pollution caused by vehicular traffic has gained significant importance. Recent research (MIRONOVA *et al.*, 2021; RESHETCHENKO, 2018; LUCHKO, 2010; ADAMENKO *et al.*, 2017) has examined noise pollution levels on the streets of various Ukrainian cities, including Lviv. Notably, an analysis by GRYNCHYSHYN *et al.* (2021) revealed significant noise pollution on the main streets in central Lviv, where levels frequently exceeded permissible standards, especially in areas with narrow streets and aged buildings.

Research by KALYN and SHELEVIY (2016) focused on the issue of noise pollution within the urban ecosystem of Lviv, identifying its sources and key characteristics, evaluating methods for its mitigation, and analyzing noise levels across different city road segments.

Studies by KACHMAR (2013), KACHMAR and LANETS (2020), KACHMAR *et al.* (2018), and ZUBYK and KHODAN (2014) explored the impact of road surface conditions on noise levels in Lviv. These works highlighted that permissible noise standards are often exceeded in urban areas, particularly in central Lviv. The authors argue that while reducing traffic noise using traditional methods is challenging and financially demanding, controlling and decreasing noise at the source is more feasible. In this context, passenger cars, which constitute a major portion of TF, are significant noise contributors. The research conducted in Lviv presents valuable insights for developing information systems aimed at assessing noise levels in populated areas, especially in cities with prolonged road

load times like Lviv. The analysis of recent advancements and research in this field underscores the significance of these experiments and highlights opportunities for further scientific inquiry.

Noise pollution studies are critically important in urban environments, where the population is continually exposed to various sources of noise. Traffic noise, a significant contributor, results from the operation of vehicle engines, wheels, brakes, and aerodynamics (ZUBYK, KHODAN, 2014). The primary metric for evaluating noise levels is the equivalent continuous sound level that transmits the same amount of energy over a specific period as fluctuating noise would during the same period. The equivalent continuous sound level is calculated using the following equation:

$$L_{eq} = 10 \log \left(\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} 10^{0.1L_{A(t)}} dt \right), \quad (1)$$

where $L_{A(t)}$ is the sound level value and t_1 , t_2 are the time periods of measurement.

3. Conducting an experiment

To gain a comprehensive understanding of the fluctuations in noise levels throughout the day, experiments were conducted at various times. These results are intriguing, as they illustrate the onset of increasing noise pollution levels.

The collected data are invaluable for further analysis and practical application. While noise intensity is a significant factor, the distribution and perception of noise and factors influencing these two are equally crucial. Additional data gathered during the measurements, such as the day of the week, time of day, air temperature, pressure, humidity, and wind speed, are pivotal for in-depth analysis. For instance, air temperature and humidity influence the speed of sound, while wind direction can alter noise direction and intensity.

Video recordings enabled the classification vehicles by category, including trams, which aids in identifying noise sources and evaluating their impact on overall noise pollution levels.

This supplementary information can be employed to calibrate computer simulation systems for sound propagation or to test new methods of predicting road noise. Additionally, it enhances the accuracy and reliability of models used to analyze and forecast noise pollution in Lviv. Such a meticulous approach to data collection and analysis contributes to the development of more effective noise control strategies, ultimately improving the quality of life for local residents.

For better data usability, a detailed diagram was created to document the building, with all measurement points clearly marked (Fig. 1).

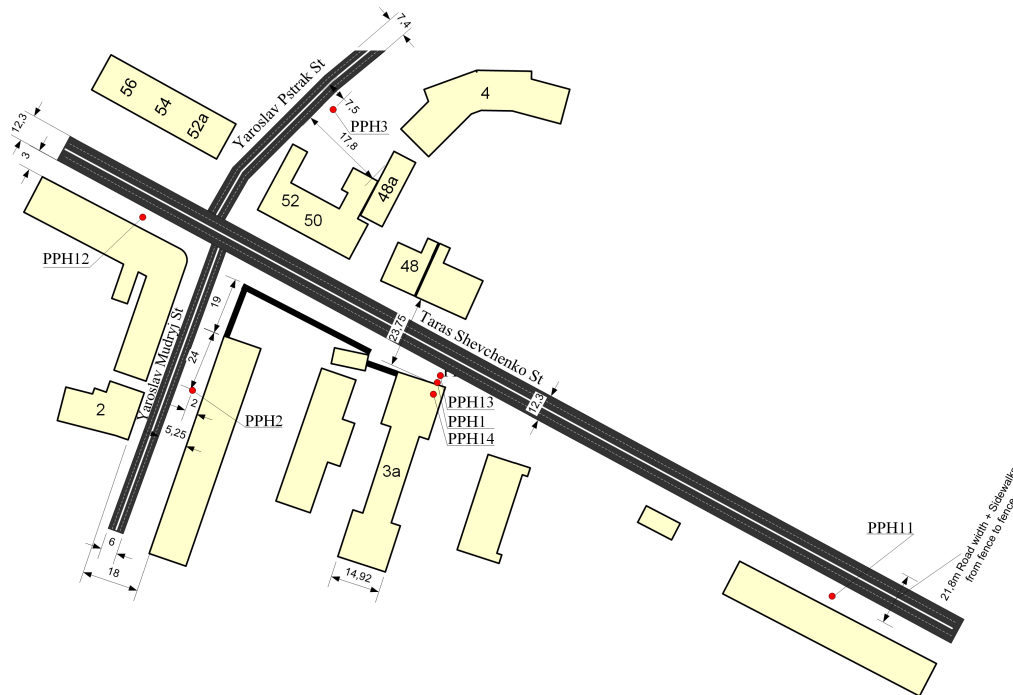


Fig. 1. Placement of measuring points on Shevchenko Street and adjacent streets.

4. Coordinates of measurement points

To facilitate further computational simulations, all points where noise levels were monitored were meticulously marked on the diagram, and their coordinates were accurately recorded using a GPS logger, as summarized in Table 1.

Additionally, extensive experiments were conducted on Taras Shevchenko Street (Fig. 2). The accumulated measurements adequately illustrate the diurnal trend of noise level fluctuations.



Fig. 2. Measurement of the equivalent noise level during the day on Taras Shevchenko Street.

The results are depicted in Fig. 3, where PPH1 represents the equivalent noise level [dB(A)] for Taras

Shevchenko Street, PPH2 for Yaroslav Mudryj Street, and PPH3 for Yaroslav Pstrak Street. Notably, Taras Shevchenko Street exhibited the highest noise levels, which was anticipated due to the presence of trams. The analysis of Taras Shevchenko Street reveals a rapid increase in noise levels until 8:00, peaking at 10:00, followed by a gradual decline. However, this decline is minimal, only about 2.5 dB, during the rush hour from 17:30 to 19:00, after which noise levels rise again and only start to decrease after 21:30. The reduction in noise levels during the evening rush hour might initially appear counterintuitive due to the higher traffic volumes. However, it is the traffic congestion that significantly reduces vehicle speeds, which, particularly on cobblestone surfaces, results in lower noise levels.

An intriguing contrast is observed on Yaroslav Mudryj Street, where the conditions are reversed. Around 15:30, while noise levels are lower on Taras Shevchenko Street, they peak on Yaroslav Mudryj Street. This occurs because, with fewer cars on Taras Shevchenko Street, vehicles can move more swiftly along Yaroslav Mudryj street, generating more noise, and vice versa. When TFs onto Taras Shevchenko Street, Yaroslav Mudryj Street experiences congestion, causing vehicles near the measurement point to move at almost zero speed.

Table 1. Geographic coordinates of measurement points.

Measurement point number	GPS coordinates	Microphone height above road level [m]	Distance to the wall of the house [m]
PPH1	49.8440065, 24.0108481	1.5	2
PPH2	49.843972, 24.010861	1.5	2
PPH3	49.844806, 24.011583	1.5	7.5 (from the center of the road)

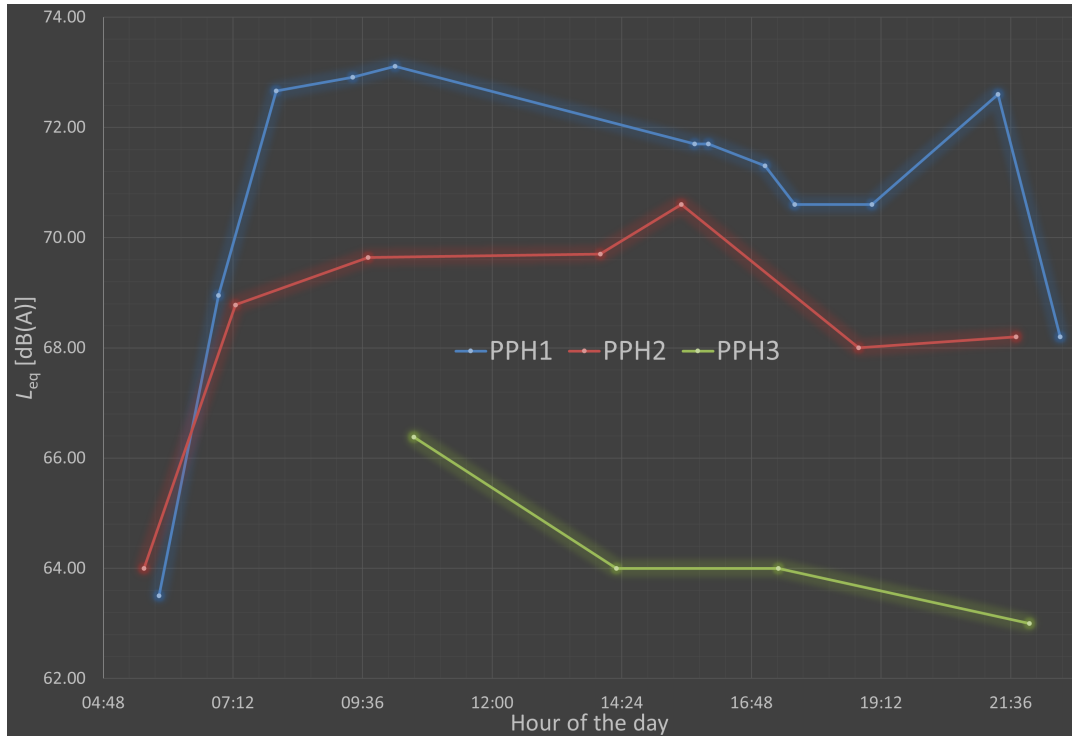


Fig. 3. Graph illustrating the dependence of noise level on the time of day.

Based on the experimental data, the hours with the most significant noise levels were identified: the most congested hour during the day is from 09:40 to 10:40 and at night from 22:00 to 23:00.

On Taras Shevchenko Street, vehicular traffic lacks smooth flow during the day, with vehicles frequently caught in traffic jams and moving at low speeds; conversely, at night, though fewer in number, vehicles tend to move at higher speeds.

5. Mathematical model

5.1. Mathematical model of the equivalent noise level for Taras Shevchenko Street

Using the Lagrange interpolation polynomial, a mathematical model was developed to determine the level of noise pollution on Taras Shevchenko Street in Lviv, as follows:

$$f(t) = \begin{cases} 67.482t + 48.395 & \text{for } 0.23 \leq t \leq 0.30, \\ 8.4245t + 66.235 & \text{for } 0.30 < t \leq 0.40, \\ 0.3349t + 69.505 & \text{for } 0.40 < t \leq 0.58, \\ 771.57t^3 - 1719.9t^2 + 1253.3t - 229.344 & \text{for } 0.58 < t \leq 0.90, \end{cases} \quad (2)$$

where $f(t)$ represents the equivalent noise level [dB(A)] as a function of time, while t denotes the time of day, segmented into 24 hours. For instance, to calculate the noise level at 12:00, $t = \frac{12}{24} = 0.5$.

Observations from Fig. 3 indicate that noise levels vary significantly throughout the day. Consequently, to provide a more precise approximation of the experimental data, the time of day has been segmented into five distinct periods. A separate polynomial is calculated for each interval, specifically chosen to optimally reflect the temporal fluctuations in noise levels.

To evaluate the accuracy of the developed model (Eq. (2)), a graphical representation was constructed as depicted in Fig. 4. It clearly shows that the experimental data, indicated with red markers, align well

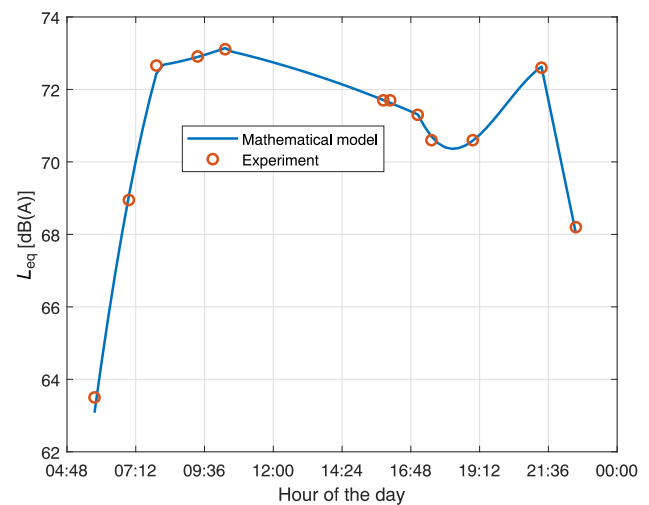


Fig. 4. Comparison of the mathematical model with experimental data of the equivalent noise level during the day for Shevchenko Street.

with the outcomes derived from the Taras Shevchenko Street model, represented by the blue line. This alignment validates the model's accuracy, confirming its suitability for estimating the equivalent noise levels from 5:00 to 23:00.

5.2. Mathematical model for determining the noise level on Yaroslav Mydryj Street

Similar to the previous model, the Lagrange interpolation polynomial was also employed to develop a mathematical model for determining the noise pollution levels on Yaroslav Mudryj Street in Lviv. This approach ensures consistency in modeling techniques across different urban settings, thereby facilitating comparative analyses and refined predictions:

$$f(t) = \begin{cases} -392.5t^2 + 327.7t + 7.0386 & \text{for } 0.24 \leq t \leq 0.33, \\ 20.644t^2 - 10.746t + 73.948 & \text{for } 0.33 < t \leq 0.43, \\ -6.9624t^2 + 1.5893t + 73.692 & \text{for } 0.43 < t \leq 0.71, \\ -1326.3t^3 + 3336.3t^2 - 2773.7t + 833.54 & \text{for } 0.71 < t \leq 0.89, \\ -91.826t + 154.35 & \text{for } 0.89 < t \leq 0.94. \end{cases} \quad (3)$$

The implementation of model (Eq. (3)) within the MATLAB system is illustrated in Fig. 5.

The program code, as depicted in Fig. 5, facilitated verification of the model's accuracy against experimental data. As the graph demonstrates, there is a congruence between the model and the experimental data, affirming the model's validity (Fig. 6).

```
twymSzew=[0.23 0.30 0.40 0.58 0.65 0.78 0.90];
ywymSzew=[64.00 68.78 69.64 69.70 70.60 68.00 68.20];

t=0.23:0.01:0.90;
[r c]=size(t);
for i=1:c
    if (t(i)>=0.23) & (t(i)<=0.30)
        y(i)=67.482.*t(i) + 48.395;
    elseif (t(i)>0.30) & (t(i)<=0.40)
        y(i)=8.4245.*t(i) + 66.235;
    elseif (t(i)>0.40) & (t(i)<=0.58)
        y(i)=0.3349.*t(i) + 69.505;
    elseif (t(i)>0.58) & (t(i)<=0.90)
        y(i)=771.57.*t(i).^3 - 1719.9.*t(i).^2 + 1253.3.*t(i) - 229.344;
    end
end

hp = plot(t,y,'LineWidth',1.5);
hold on
hpwym = plot(twymSzew,ywymSzew,'o','LineWidth',1.5);
ha = get(hp,'Parent');
Data = get(ha,'XTick');
timestr = datestr(Data,15);
set(ha,'XTickLabel',timestr);
```

Fig. 5. Segment of the program for determining the noise level on Yaroslav Mudryj Street during the day.

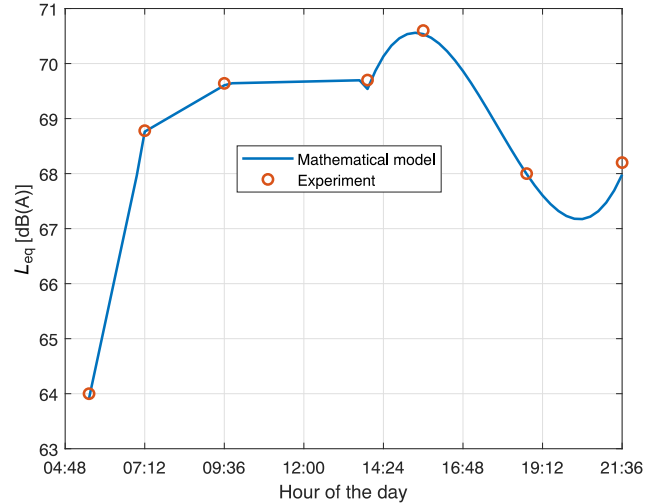


Fig. 6. Comparison of the mathematical model with experimental data of the equivalent noise level during the day for Yaroslav Mudryj Street.

Subsequent to model validation, a user interface was developed. Figure 7 displays this interface, which features a slider for rapidly adjusting the time of day, a drop-down list containing the names of streets for

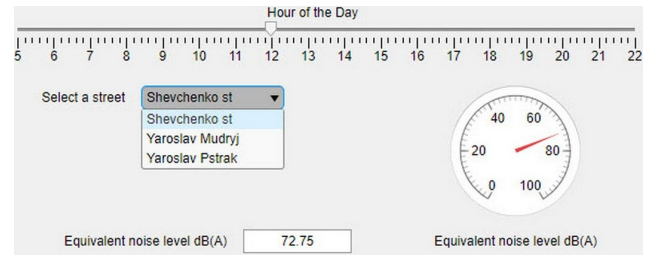


Fig. 7. User interface of the information system for determining the equivalent noise level for selected streets in Lviv.

which the equivalent noise level will be calculated, and a display window that presents the results and a scale for expedient noise level assessment.

6. Conclusions

Experimental studies carried out on Taras Shevchenko Street and its vicinity revealed significant details regarding the acoustic environment. The findings indicate that vehicle speed has a more substantial effect on noise levels than traffic volume and correlates with the fluidity of TF. Furthermore, it was observed that the noise levels on these streets exceed the allowable daytime levels of 55 dB(A) (DBN V.1.1-31:2013, 2013). To mitigate noise pollution, measures such as reducing vehicle speeds or enhancing window insulation on building facades could be implemented to reduce the noise impact on residents.

The developed mathematical models for predicting equivalent noise levels based on the time of day have proven to be both effective and accurate in estimating noise levels on urban streets. The results validate the ability of our information system for the prompt and accurate assessment of noise levels on a designated street at a given time.

The findings and analyses from this research are essential for the further development of sound propagation modeling systems. They provide a foundation for calibrating and refining existing systems, as well as testing innovative methods and approaches within this field. This progress is expected to enhance the effectiveness and accuracy of noise prediction systems, ultimately contributing to reducing noise pollution and enhancing the quality of life for urban residents.

References

- ADAMENKO Y., COMAN M., KUNDELSKA T. (2017), Ecological safety of Ivano-Frankivsk urban system according to acoustical and electromagnetic load factors, *Scientific Bulletin Series D: Mining, Mineral Processing, Non-Ferrous Metallurgy, Geology and Environmental Engineering*, **31**(2): 27–33.
- BASNER M. *et al.* (2014), Auditory and non-auditory effects of noise on health, *The Lancet*, **383**(9925): 1325–1332, [https://doi.org/10.1016/S0140-6736\(13\)61613-X](https://doi.org/10.1016/S0140-6736(13)61613-X).
- BROWN L. (2015), Effects of road traffic noise on health: From burden of disease to effectiveness of interventions, *Procedia Environmental Sciences*, **30**: 3–9, <https://doi.org/10.1016/j.proenv.2015.10.001>.
- DBN V.1.1-31:2013 (2013), Protection of territories, buildings, and structures from noise [in Ukrainian], Ministry of Regional Development, Construction, and Housing and Communal Services of Ukraine.
- DZHAMBOV A.M., LERCHER P. (2019), Road traffic noise exposure and depression/anxiety: An updated systematic review and meta-analysis, *International Journal of Environmental Research and Public Health*, **16**(21): 4134, <https://doi.org/10.3390/ijerph16214134>.
- HEGEWALD J. *et al.* (2020), Traffic noise and mental health: A systematic review and meta-analysis, *International Journal of Environmental Research and Public Health*, **17**(17): 6175, <https://doi.org/10.3390/ijerph17176175>.
- GRYNCHYSHYN N.M., SHUPLAT T.I., ZHORINA O.O. (2021), Noise pollution in the main streets of the central part of Lviv [in Ukrainian], *Bulletin of Lviv State University of Life Safety*, **24**: 6–11, <https://doi.org/10.32447/20784643.24.2021.01>.
- KACHMAR R.Ya. (2013), Assessment of environmental and economic losses from traffic noise in the city of Lviv [in Ukrainian], *Automotive Transport: Research*, **1**(231): 10–13.
- KACHMAR R., LANETS O. (2020), The impact of parameters of traffic flows in the Lviv street-road network on the level of environmental and economic losses [in Ukrainian], *Transport Technologies*, **1**(1): 83–91, <https://doi.org/10.23939/tt2020.01.083>.
- KACHMAR R., LYODA V., POLYAKEVYCH V. (2018), The impact of road pavement in Lviv on the level of noise pollution [in Ukrainian], *Bus Manufacturing and Passenger Transportation in Ukraine: Proceedings of the III All-Ukrainian Scientific and Practical Conference*, pp. 189–191.
- KALYN B.M., SHELEVYI M.I. (2016), Directions for optimizing the noise factor of traffic flows in Lviv [in Ukrainian], *Scientific Bulletin of LNUVMBT named after S.Z. Gzhytskyi*, **18**(2): 104–107.
- LUCHKO I.A. (2010), Results of the study of noise load on streets, roads, and avenues in Kyiv [in Ukrainian], *Bulletin of NTUU “KPI”. Series “Mining”*, **19**: 188–197.
- MIRONOVA N.H., MOROZOV A.V., MOROZOVA T.V., RYBAK V.V. (2021), Study of acoustic load from traffic flow in the city of Khmelnytskyi [in Ukrainian], *Roads and Bridges*, **24**: 193–205.
- OW L.F., GHOSH S. (2017), Urban cities and road traffic noise: Reduction through vegetation, *Applied Acoustics*, **120**: 15–20, <https://doi.org/10.1016/j.apacoust.2017.01.007>.
- OZER S., IRMAK M.A., YILMAZ H. (2008), Determination of roadside noise reduction effectiveness of *Pinus sylvestris* L. and *Populus nigra* L. in Erzurum, Turkey, *Environmental Monitoring and Assessment*, **144**(1–3): 191–197, <https://doi.org/10.1007/s10661-007-9978-6>.
- PETRESCU V., CIUDIN R., CLAUDIU I., CIOCA L.-I., VICTOR N. (2015), Traffic noise pollution in a historical city – Case study project within environmental engineering field of study, *3rd International Engineering and Technology Education Conference & 7th Balkan*

Region Conference on Engineering and Business Education, pp. 1–8.

17. RESHETCHENKO A.I. (2018), Study of the impact of automobile traffic flows on the acoustic environment of urban landscapes [in Ukrainian], *Municipal Economy of Cities. Series “Technical Sciences and Architecture”*, **146**: 180–183.
18. TITU A.M., BOROIU A.A., MIHAILESCU S., POP A.B., BOROIU A. (2022), Assessment of road noise pollution in urban residential areas – A case study in Pitești, Romania, *Applied Sciences*, **12**(4053): 1–14, <https://doi.org/10.3390/app12084053>.
19. ZAMBON G., ROMAN H.E., SMIRAGLIA M., BENOCCI R. (2018), Monitoring and prediction of traffic noise in large urban areas, *Applied Sciences*, **8**(2): 251, <https://doi.org/10.3390/app8020251>.
20. ZUBYK S.V., KHODAN M.M. (2014), Architectural and planning methods for combating urban traffic noise [in Ukrainian], *Scientific Bulletin of NLTU of Ukraine*, **24**(11): 185–191.

Research Paper

Multi-label Bird Species Classification Using Transfer Learning Network

Xue HAN, Jianxin PENG*

*School of Physics and Optoelectronics, South China University of Technology
Guangzhou, China**Corresponding Author e-mail: phjxpeng@163.com*Received September 20, 2024; revised April 3, 2025; accepted April 28, 2025;
published online June 9, 2025.*

Bird sounds collected in the field usually include multiple birds of different species vocalizing at the same time, and the overlapping bird sounds pose challenges for species recognition. Extracting effective acoustic features is critical to multi-label bird species classification task. This work has extended an efficient transfer learning technique for labelling and classifying multiple bird species from audio recordings, further laying the foundation for conservation plans. A synthetic dataset was created by randomly mixing original single-species bird audio recordings from the Cornell Macaulay Library. The final dataset consists of 28 000 audio clips, each 5 s long, containing overlapping vocalizations of two or three bird species among 11 different species. Several pre-trained convolutional neural networks (CNNs), including InceptionV3, ResNet50, VGG16, and VGG19, were evaluated for extracting deep features from audio signals represented as mel spectrograms. The long short-term memory network (LSTM) was further employed to extract temporal features. A multi-label bird species classification was investigated. The absolute matching rate, accuracy, recall, precision, and $F1$ -score of the InceptionV3+LSTM model for multi-label bird species classification are 98.25 %, 99.32 %, 99.41 %, 99.90 %, and 99.57 %, respectively, with the minimum Hamming loss of 0.0062. The results show that the proposed method has excellent performance and can be used for multi-label bird species classification.

Keywords: transfer learning; multi-label bird species classification; InceptionV3; LSTM.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Field recordings of bird sounds typically contain vocalizations from multiple bird species occurring simultaneously, known as the ‘dawn chorus’, a phenomenon common in natural habitats. However, relatively few studies have addressed the challenge of multi-label bird species classification in these realistic acoustic environments. Early studies primarily relied on classical acoustic features and traditional machine learning approaches. For example, BRIGGS *et al.* (2012) manually segmented overlapping bird sounds recorded from the H.J. Andrews Experimental Forest (548 audio clips, each containing 1–5 species) and utilized multi-instance multi-label K-nearest neighbor (MIML-KNN), achieving an accuracy of 96.1 %. LENG and DAT TRAN (2014) combined spectral features, MFCC, and linear predictive coding (LPC) extracted from NIPS4B dataset (687 audio clips, containing multiple bird species per clip) and trained ensemble clas-

sifiers, obtaining an AUC of 91.74 %. LIU (2016) introduced a transfer learning feature mapping method based on MFCC and Gaussian mixture models (GMM) for multi-label bird sound classification. The method was evaluated on NIPS4B and an artificial dataset (constructed by mixing xeno-canto bird audio), achieving the Hamming loss of 0.1024.

With the advancement of deep learning, recent approaches have increasingly utilized convolutional neural networks (CNNs) to automatically learn acoustic features. SPRENGEL *et al.* (2016) proposed a CNN approach trained on the BirdCLEF 2016 dataset, achieving a mean average precision (MAP) score of 0.686 for identifying the dominant bird species in audio recordings, surpassing previous state-of-the-art results. BRAVO SANCHEZ *et al.* (2021) used the CNN framework SincNet on NIPS4Bplus bird recordings, achieving an accuracy of 73.56 %. NOUMIDA and RAJAN (2022) proposed a hierarchical attention-based bidirectional gated recurrent unit (BiGRU) model with

MFCC, trained on the xeno-canto dataset, achieving an $F1$ -score of 0.85. [ABDUL KAREEM and RAJAN \(2023\)](#) fused MFCC-RNN and mel spectrogram-CNN methods, obtaining an $F1$ -score of 0.75 on the xeno-canto dataset.

Although CNNs effectively extract local features from spectrograms, they often neglect long-term temporal dependencies in acoustic data. Integrating CNNs with recurrent neural networks (RNNs), such as long short-term memory (LSTM), addresses this limitation by capturing sequential acoustic patterns ([SAINATH et al., 2015](#); [NISHIKIMI et al., 2021](#); [LIU et al., 2021](#)).

Transfer learning allows models to leverage the knowledge learned from large-scale datasets and tasks, significantly reducing training parameters and accelerating learning processes ([WEISS et al., 2016](#)). Transfer learning is very helpful when there is insufficient data to fully train a model, such as recognizing uncommon bird species ([HUANG, BASANTA, 2021](#)). [GUNAWAN et al. \(2021\)](#) applied a transfer learning technique to avoid overfitting when classifying endangered species, such as the small footed owl in Indonesia. Deep CNN models, including VGG ([SIMONYAN, ZISSERMAN, 2014](#)), ResNet ([HE et al., 2016](#)), and Inception networks ([SZEGEDY et al., 2016; 2017](#)), have shown superior performance on image classification tasks, making them ideal candidates for transfer learning. [SEVILLA and GLOTIN \(2017\)](#) successfully adapted the Inception-v4 network to bird sound classification, achieving the highest accuracy 71.4% on the Bird-CLEF 2017 dataset. Transfer learning has also proven effective in various multi-label classification tasks, including autonomous driving ([LI et al., 2021](#)), natural language sentiment analysis ([TAO, FANG, 2020](#)), and transformer-based models across multiple domains ([GÓMEZ-GÓMEZ et al., 2023](#)).

In this study, inspired by previous works, we utilize transfer learning models including VGG16, VGG19, InceptionV3, and ResNet50 to extract deep acoustic features from mel spectrograms of bird sounds. An LSTM network is integrated to capture temporal

dependencies across frames. We specifically focus on multi-label classification tasks involving simultaneous vocalizations of two or three bird species, using synthetic datasets created from Cornell’s Macaulay Library recordings (28 000 audio clips, each lasting 5 s). The feature extraction capability, classification accuracy, and generalization performance of these integrated models are comprehensively analyzed.

2. Method

2.1. Transfer learning models

The core idea of transfer learning is to leverage knowledge from the source domain to improve performance in a related target task. In this study, we utilize four pre-trained convolutional neural network architectures – VGG16, VGG19, InceptionV3, and ResNet50 – originally trained on the ImageNet dataset ([DENG et al., 2009](#)). Each architecture has distinct characteristics that influence its performance on multi-label bird species classification tasks.

2.1.1. Classification model based on pre-trained network VGG16/VGG19

VGG is a classic image classification network based on the ImageNet database. Its characteristic is to use a convolutional layer with a smaller kernel (3×3) instead of a convolutional layer with a larger kernel. On the one hand, it can reduce parameters, and on the other hand, it is equivalent to performing more nonlinear mapping, increasing the network’s expressive power ([SIMONYAN, ZISSERMAN, 2014](#)). The VGG16 pre-trained network framework is shown in Fig. 1. The VGG19 model adds three additional convolutional layers on top of VGG16: one $3 \times 3 \times 256$ and two $3 \times 3 \times 512$. VGG19 has a deeper network than VGG16, and increasing the depth can effectively improve performance. The part before the fully connected layer is commonly referred to as the feature extraction layer.

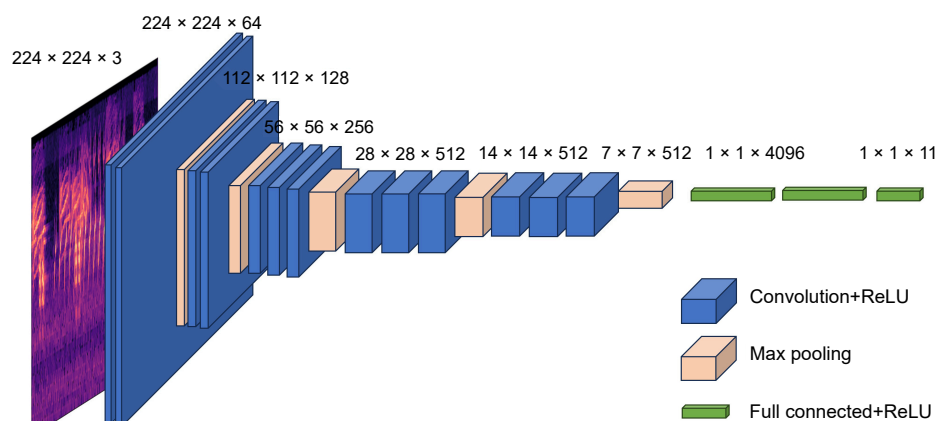


Fig. 1. VGG16 network structure diagram (based on ([SIMONYAN, ZISSERMAN, 2014](#))).

2.1.2. Classification model based on pre-trained network InceptionV3

The method of increasing the number of convolutional layers to enhance the learning ability of the network is not always feasible, because after the network reaches a certain depth, increasing the number of network layers will cause the problem of random gradient disappearance and explosion, and also lead to a decrease in accuracy. Moreover, complex networks can also bring high computational costs. The Inception module decomposes large convolutions into multiple small convolutions, where multiple small convolution kernels simultaneously convolve the image and aggregate information at different scales, as shown in Fig. 2 (SZEGEDY *et al.* 2016). This can significantly reduce network parameters without losing features. The key to the InceptionV3 network is to use Inception modules and two asymmetric decomposition structures to construct different types of Inception module groups.

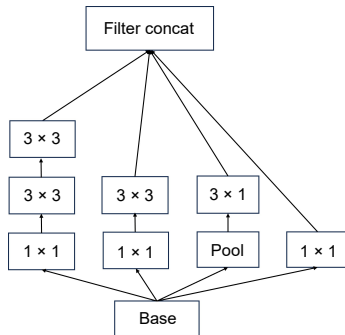


Fig. 2. Each 5×5 convolution in the Inception module is replaced by two 3×3 convolutions (based on (SZEGEDY *et al.*, 2016)).

2.1.3. Classification model based on pre-trained network ResNet50

In response to the problem of gradient disappearance, HE *et al.* (2016) proposed a residual structure that not only solves the gradient problem, but also improves its feature expression ability with the increase of network layers, thereby improving classification performance. Figure 3 shows a residual structure in ResNet50, which includes cross layer connections that allow input to be directly passed across layers and then added to the convolutional result. This helps the model converge towards the equal mapping

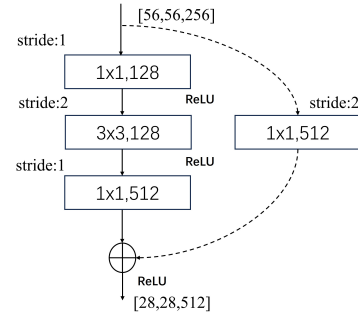


Fig. 3. Residual structure diagram (HE *et al.*, 2016).

direction, ensuring that the final accuracy is not affected by the depth of the model. Table 1 shows the ResNet50 model structure and network parameters.

Table 1. Model training parameter statistics.

Models	LSTM input size	LSTM trainable parameters	FC parameters	Total trainable parameters
VGG16/VGG19	3 584	3 933 184	2 827	3 936 011
InceptionV3	16 384	17 040 384	2 827	17 043 211
ResNet50	14 336	14 943 232	2 827	14 946 059

We chose these models due to their distinct advantages: VGG models provide strong representational power, InceptionV3 excels at multi-scale feature extraction with fewer parameters, and ResNet50 effectively manages training of very deep networks. A comparative analysis of these models is summarized in Table 2.

2.2. Pre-trained convolutional neural network fused with LSTM

The feature sequence extracted by the pre-trained CNN cannot be directly fed into the LSTM network. Taking the VGG16 model as an example, the final extracted feature map has dimensions of $7 \times 7 \times 512$ ($H \times W \times C$). The original mel spectrogram of size 224×224 is compressed spatially to 7×7 , and the number of channels deepens from 3 (original RGB channels) to 512. Through convolution and pooling operations, the spatial structural information in the mel spectrogram is transformed into deep feature representations, where each channel corresponds to a particular response pattern, such as edges, textures, and colors.

Table 2. CNN networks comparison.

CNN networks	Depth (layers)	Parameters, complexity	Feature extraction strategy
VGG16/19	16/19	High parameters, high computational complexity	Small 3×3 convolutions, captures fine-grained local features
InceptionV3	48	Moderate parameters, efficient computation due to parallel modules	Multi-scale feature extraction via parallel convolutions (Inception modules)
ResNet50	50	High parameters but efficient training	Deep residual structure enabling hierarchical feature abstraction

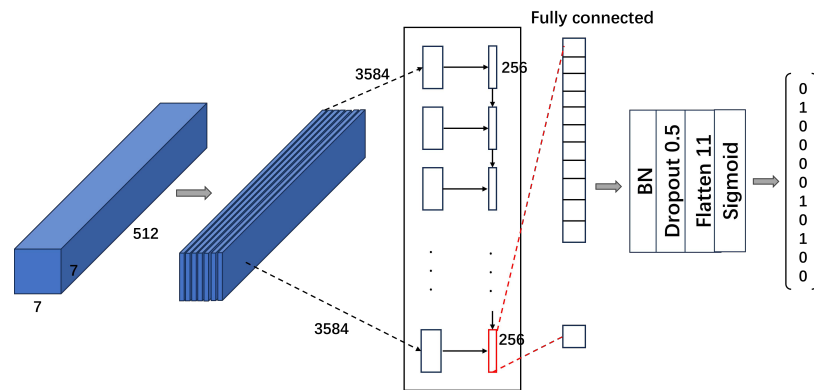


Fig. 4. Feature sequence obtained from pre-trained VGG16 input with LSTM.

To make these features suitable for temporal modeling by LSTM, the feature maps are reshaped based on the spatial dimension (e.g., width dimension) into a sequential format. Specifically, we expand the feature maps along the width dimension (W) into a feature sequence of size 7×3584 , where 7 corresponds to the number of time steps, and 3584 (i.e., 7×512) represents the features at each time step. Figure 4 illustrates the detailed procedure of feeding these feature sequences input into the LSTM network. Similarly, for other pre-trained CNN architectures, the extracted feature map dimensions differ slightly: InceptionV3 produces a feature sequence of $8 \times 8 \times 2048$, and ResNet50 yields a feature sequence of $7 \times 7 \times 2048$. These feature maps are processed in the same way as described above to prepare sequential data suitable for input into the LSTM network.

The number of neurons in the output layer of the LSTM network is 11 for classifying bird species. Sigmoid is used as the activation function for the output layer, ensuring that each neuron outputs a probability of 1. When the probability value is greater than 0.5, the predicted value output by the neuron is 1, indicating that the bird species corresponding to the neuron exists in the audio sample. The fusion model also adds a batch normalization (BN) layer to avoid gradient disappearance. Dropout layers are introduced to reduce overfitting by randomly dropping units during training, thereby improving the generalization performance of the network. The parameters in the feature extraction modules of VGG16, VGG19, InceptionV3, and ResNet50 are frozen, and the remaining parts of the network are trained using our dataset in this work. To help assess model complexity, we provide the number of trainable parameters (excluding frozen CNN layers) for each configuration, as shown in Table 1.

3. Experimental settings

3.1. Dataset construction

Bird sound recordings of 11 bird species used for this work are collected from the Macaulay Library at

Cornell University¹. Table 3 shows the selected original audio of each bird. These recordings are in MP3 format, with a sampling frequency of 44 100 Hz and a bit rate of 128 000 bps. Each recording is annotated as a single species of bird vocalization. To create multi-label bird species samples, we randomly selected audio segments from the original single-species recordings and combined them digitally using audio mixing software ‘Adobe Audition’. Before mixing, the audio segments were normalized to the same volume level to prevent any single recording from dominating due to volume differences. Each resulting 5-second segment contains clearly annotated overlapping vocalizations from either 2 or 3 different bird species. The detailed information is shown in Table 4.

Figures 5 and 6 represent mel spectrograms of syllable overlapping for 2 and 3 bird species, respectively. The mel spectrograms were constructed using the Librosa library with a sample rate of 44 100 Hz, the FFT window length (n_{fft}) of 1024 samples, a hop length of 512 samples (50% overlap), and 128 mel filter banks. A visual inspection reveals that as the number of overlapping bird species increases, the spectral complexity and signal interference also become more pronounced. For instance, in Figs. 5a and 5c, the individual vocal patterns of each species are relatively separable, often occurring in distinct frequency bands or time intervals. However, in Figs. 6b and 6c, the spectrograms show significantly denser and more continuous activity across both frequency and time, making it more difficult to visually or algorithmically disentangle individual species. This suggests that classification tasks involving three or more overlapping bird species are inherently more challenging due to increased spectral overlap, which can obscure characteristic frequency patterns and temporal features.

The labels of each audio segment are manually reviewed and verified to ensure the presence of corresponding bird sounds. The ratio of training set to testing set is divided into 3:1.

¹Specifically retrieved via the search interface at: <https://search.macaulaylibrary.org/catalog>

Table 3. Audio file information of 11 bird species.

Bird species	Number of downloaded audio files	Audio file name
Downy Woodpecker	12	ML107289, ML433684551, ML320270011, ML216529601, ML288560951, ML89889581, ML259178751, ML94232, ML282354581, ML249048571, ML218533941, ML539363
Northern Flicker	9	ML60535251, ML47981841, ML6891, ML84808, ML224667, ML176938031, ML6802, ML63072, ML299493831
Black-capped Chickadee	10	ML381756441, ML202239, ML227931651, ML228999, ML359860121, ML244530591, ML442275881, ML315584611, ML9334271, ML217850561
White-breasted Nuthatch	9	ML51757711, ML196990751, ML304498191, ML105313481, ML313785451, ML120214, ML169318341, ML88195851, ML245567141
Northern Cardinal	7	ML101113031, ML94284, ML94286, ML94285, ML325248201, ML434987071, ML24184651
House Finch	9	ML369617771, ML44967, ML110958961, ML331732541, ML161496541, ML22938, ML56843, ML22941, ML12932
Pine Siskin	5	ML156434831, ML22902731, ML176160, ML89549511, ML219631251
Western Backyard Birds	6	ML481585181, ML203884811, ML425203981, ML279795071, ML2425203911, ML168880461
Steller's Jay	8	ML35291431, ML202130641, ML44859, ML410551461, ML42204, ML119017701, ML192457, ML90747421
Evening Grosbeak	5	ML148939381, ML160442941, ML129191951, ML77259, ML227584
Blue Jay	13	ML166281501, ML177463211, ML345934681, ML107392, ML264268971, ML260458751, ML421603721, ML539887, ML219634, ML13448, ML359246651, ML223790721, ML20432

Table 4. Detailed information of multi-label dataset.

Category	Number of training samples	Number of test samples	In total
2	11 550	3 850	15 400
3	9 450	3 150	12 600
Total	21 000	7 000	28 000

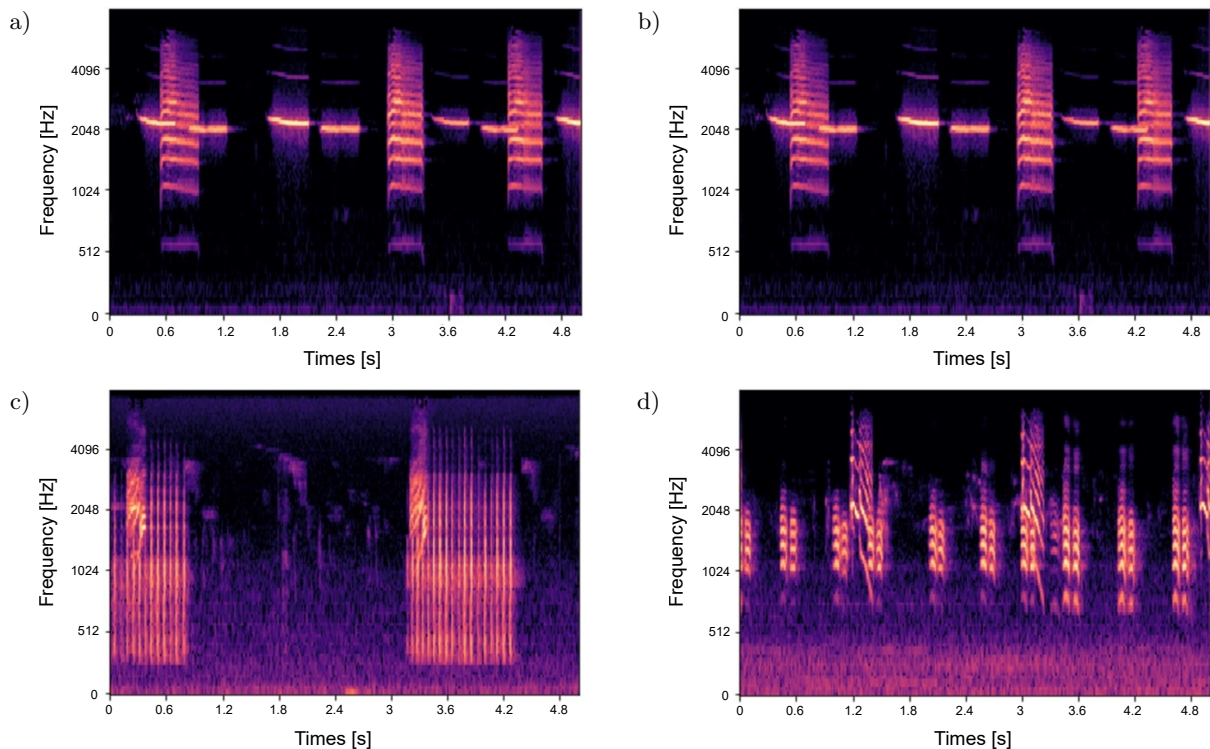


Fig. 5. Mel spectrogram of mixed audio of syllables overlapping between two species of birds: a) Black capped Chickadee–Blue Jay; b) Pine Siskin–Steller's Jay; c) Downy Woodpecker–House Finch; d) Northern Flicker–White-breasted Nuthatch.

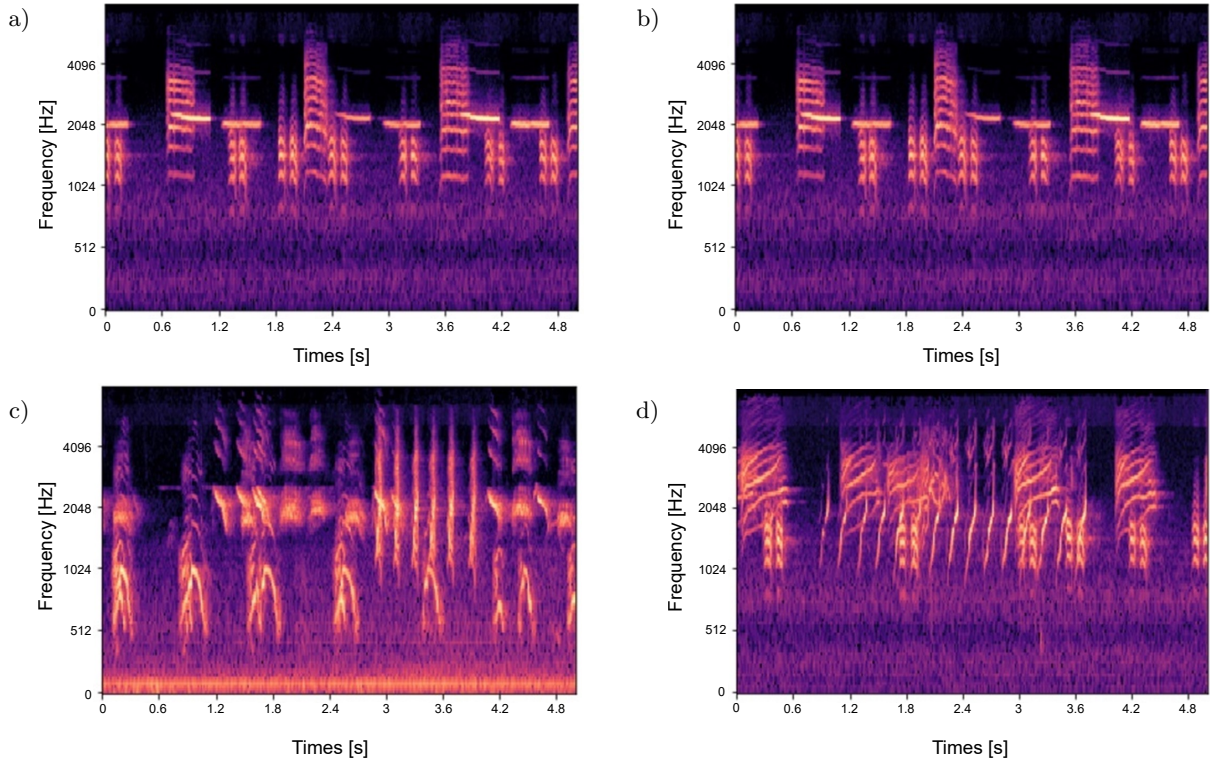


Fig. 6. Mel spectrogram of mixed audio of syllables overlapping between three species of birds: a) Black capped Chickadee–White-breasted Nuthatch–Blue Jay; b) Downy Woodpecker–Northern Flicker–White-breasted Nuthatch; c) Western Backyard Birds–Steller’s Jay–Evening Grosbeak; d) White-breasted Nuthatch–Northern Cardinal–Pine Siskin.

3.2. Objective evaluation

Unlike single label classification tasks, a sample in a multi-label classification task can have multiple labels. Firstly, without considering partially correct evaluation metrics, the sample can only be predicted correctly if the predicted label is exactly the same as the true label (PANIRI *et al.*, 2020). This evaluation metric is called the exact match ratio (ZHANG *et al.*, 2016) and the calculation formula is as follows:

$$\text{Exact match ratio} = \frac{1}{n} \sum_{i=1}^n I(\widehat{Y}_i = Y_i), \quad (1)$$

where I is the indicator function. When Y_i is completely equivalent to \widehat{Y}_i , I is 1, otherwise it is 0; \widehat{Y}_i is the predicted label set for sample i , Y_i is the ground truth label set; n represents the total number of samples. It can be seen that this evaluation metric is very strict for the classification model. In addition, only some labels that are correctly predicted can also be used to evaluate the performance of classification models. The commonly used performance metrics include accuracy, recall, precision, and $F1$ -score (GODBOLE, SARAWAGI, 2004). Accuracy is defined as the proportion of correctly predicted labels to the union of predicted and true labels for each sample, averaged across all samples:

$$\text{Accuracy} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \cup \widehat{Y}_i|}{|Y_i \cap \widehat{Y}_i|}, \quad (2)$$

where $|Y_i \cap \widehat{Y}_i|$ is number of correctly predicted labels for sample i , $|Y_i \cup \widehat{Y}_i|$ is total number of unique labels in the prediction and true labels for sample i . Recall measures the proportion of correctly predicted labels out of all true labels for each sample, averaged over all samples:

$$\text{Recall} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \cap \widehat{Y}_i|}{|Y_i|}, \quad (3)$$

where $|Y_i|$ is total number of true labels for sample i . Precision is defined as the proportion of correctly predicted labels out of all predicted labels for each sample, averaged over all samples:

$$\text{Precision} = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i \cap \widehat{Y}_i|}{|\widehat{Y}_i|}, \quad (4)$$

where $|\widehat{Y}_i|$ is total number of labels predicted for sample i . The $F1$ -score for each sample is the harmonic mean of precision and recall, averaged across all samples:

$$F1\text{-score} = \frac{1}{n} \sum_{i=1}^n 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}. \quad (5)$$

In addition, performance metrics also include the Hamming loss (SOROWER, 2010). The Hamming loss evaluates the fraction of labels that are incorrectly predicted across all samples:

$$\text{Hamming loss} = \frac{1}{kn} \sum_{i=1}^n \sum_{j=1}^k I(y_{ij} \neq \widehat{y}_{ij}), \quad (6)$$

where k is the total number of labels, y_{ij} is the true value of the j -th label for sample i , and \hat{y}_{ij} is the predicted value of the j -th label for a sample i . The smaller the value of the Hamming loss, the better the performance of the classification model.

3.3. Implementation details

The hardware environment for experiments is a server with Inter I9-7920X CPU and NVIDIA GTX RTX1080Ti GPU, and the operating system is Ubuntu 16.04. All experimental models are built based on the PyTorch deep learning framework, with the PyTorch version number 1.9.1. During the training process, MultiLabelSoftMarginLoss (CHENG *et al.*, 2021) is used as the loss function, and the stochastic gradient descent (SDG) is used to update the network parameters. Momentum is set to 0.9, the learning rate is set to e-4, epoch is set to 300, and batch size is set to 32.

4. Result

4.1. Classification results of different transfer learning models

The results of multi-label bird species classification under different transfer learning models are shown in Table 5. According to Table 5, the InceptionV3 model has the best classification performance, with exact match ratio, accuracy, recall, precision, and $F1$ -score of 93.04%, 97.30%, 97.50%, 99.75%, and 98.30%, respectively, and the Hamming loss of 0.026. The InceptionV3 model uses decomposition convolution, which decomposes large convolution factors into small convolutions and asymmetric convolutions, effectively reducing parameters and avoiding overfitting. The Inception modules use multiple branches to extract high-order features with different levels of abstraction, enriching the network's expressive power (SZEGEDY *et al.*, 2016). The absolute matching rate of the VGG19 model is 4.23% lower than that of the VGG16 model, which proves that blindly adding convolutional layers will not improve the classification performance and will lead to overfitting of the model. The exact match ratio, accuracy, recall, precision, and $F1$ -score of the ResNet50

Table 5. Multi-label bird species classification results of four transfer learning models.

	VGG16	VGG19	InceptionV3	ResNet50
Exact match ratio [%]	87.87	83.64	93.04	85.62
Accuracy [%]	95.20	93.63	97.30	94.40
Recall [%]	95.69	94.41	97.50	95.22
Precision [%]	99.41	99.00	99.75	99.02
$F1$ -score [%]	96.97	95.97	98.30	96.46
Hamming loss	0.045	0.059	0.026	0.051

model are 85.62%, 94.40%, 95.22%, 99.02%, and 96.46%, respectively, with the Hamming loss of 0.051. The classification performance of the ResNet50 model is better than VGG19, but inferior to VGG16, indicating that the residual structure has to some extent alleviated the overfitting phenomenon of the model. The CNN structure affects the results of multi-label bird sounds classification.

Figures 7 and 8, respectively, depict the variation curves of exact match ratio and the Hamming loss for four pre-trained models. As shown in Fig. 7, the exact match ratio of ResNet50 did not significantly improve after the 50th epoch, while the exact match ratio of InceptionV3 continued to increase, approaching saturation approximately after the 100th epoch. From Fig. 8, it can be seen that after the 50th epoch, the Hamming loss of InceptionV3 is significantly lower than VGG16, VGG19, and ResNet50. Overall, the VGG16 and VGG19 models are not suitable for multi-label bird species classification.

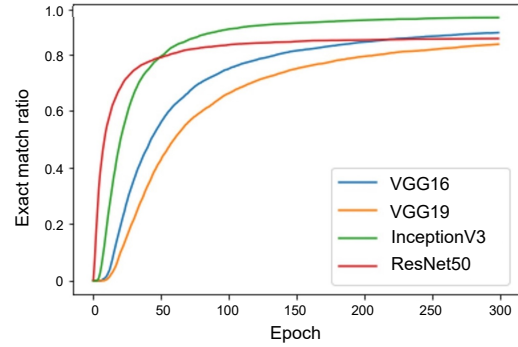


Fig. 7. Exact match ratio curves of four transfer learning models.

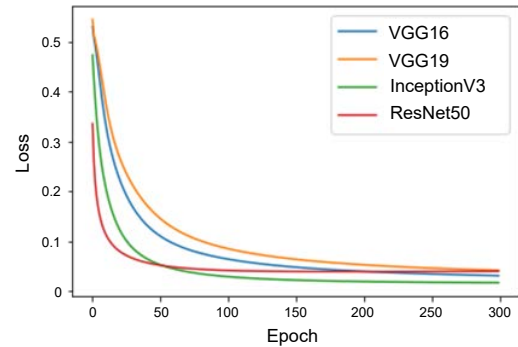


Fig. 8. Hamming loss curves of four transfer learning models.

4.2. Classification results of four transfer learning models fused with LSTM

Table 6 shows the classification results of four transfer learning models fused with LSTM. Comparing Tables 5 and 6, it can be seen that after fusing LSTM, the classification performance of VGG16, VGG19, InceptionV3, and ResNet50 has all improved significantly. The exact match ratio of ResNet50 in-

creased the most, by 12.89 %, VGG19 by 6.42 %, InceptionV3 by 5.21 %, and VGG16 by 3.73 %. The LSTM network can learn the time series characteristics in feature sequences, and the time series of vocalizations of different bird species vary. InceptionV3+LSTM has the best classification performance among all models, with exact match ratio, accuracy, recall, precision, and $F1$ -score of 98.25 %, 99.32 %, 99.42 %, 99.90 %, and 99.57 %, respectively. Moreover, the Hamming loss also reaches a minimum of 0.0062. The Hamming loss of InceptionV3+LSTM is reduced by 0.0198 compared to InceptionV3, indicating that the prediction error and missing error of multi labels are minimized. This is because the CNN-LSTM model can learn complicated patterns from data more rapidly and correctly than the CNN model alone. However, the precision of InceptionV3+LSTM is slightly lower by 0.05 % than ResNet50+LSTM. The exact match ratio, accuracy, recall, precision, and $F1$ -score of the ResNet50+LSTM model are 6.55 %, 2.46 %, 1.99 %, 0.52 %, and 1.52 % higher than those of VGG16+LSTM, respectively. This indicates that the feature sequence obtained by ResNet50 contains more time series characteristics. The classification performance of VGG19 fusion LSTM network has been improved, but it is still lower than the other three transfer learning models.

Table 6. Multi-label bird species classification results of four transfer learning models fused with LSTM.

	VGG16 + LSTM	VGG19 + LSTM	InceptionV3 + LSTM	ResNet50 + LSTM
Exact match ratio [%]	91.60	90.06	98.25	98.15
Accuracy [%]	96.78	96.21	99.32	99.24
Recall [%]	97.28	96.79	99.41	99.27
Precision [%]	99.43	99.32	99.90	99.95
$F1$ -score [%]	97.99	97.62	99.57	99.51
Hamming loss	0.029	0.034	0.0062	0.0073

Figures 9 and 10, respectively, depict the exact match ratio and the Hamming loss variation curves of four transfer learning models fused with

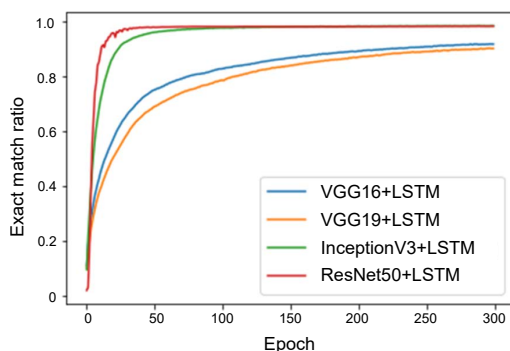


Fig. 9. Exact match ratio curves of four transfer learning models fused with LSTM.

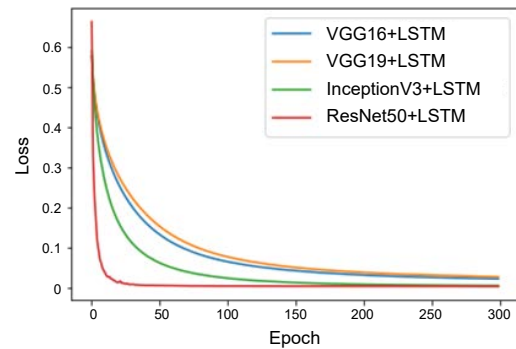


Fig. 10. Hamming loss curves of four transfer learning models fused with LSTM.

LSTM. As shown in Fig. 9, the exact match ratio of the VGG16+LSTM and VGG19+LSTM classification models is significantly lower than InceptionV3+LSTM and ResNet50+LSTM. Before the 70th epoch, the exact match ratio of ResNet50+LSTM was slightly higher than InceptionV3+LSTM. After the 70th epoch, there was no significant difference between the two fusion models. As shown in Fig. 10, before the 200th epoch, InceptionV3+LSTM had the higher Hamming loss than ResNet50+LSTM. After the 200th epoch, there was no significant difference between the two fusion models. Considering the performance of the model and computing resources, we chose InceptionV3+LSTM as the multi-label bird species classification model.

4.3. Classification confusion matrices of InceptionV3+LSTM

To further analyze the performance of the InceptionV3+LSTM model in multi-label bird species classification, we present the confusion matrices for each label in Table 7. These confusion matrices provide a detailed view of the model's prediction accuracy for each individual label. For label 0, the high number of true positives (1332) and true negatives (5661) indicates that the model performs well in identifying this label. However, label 2 has a relatively higher number of false positives (8) and false negatives (11) compared to other labels, which may suggest that this label is more challenging for the model to classify accurately. Label 4 has a moderate number of false positives (4) but a higher number of false negatives (23), indicating that the model may have difficulty in correctly identifying this label.

Overall, the confusion matrices demonstrate that the InceptionV3+LSTM model has a high accuracy in classifying most labels, with only a few exceptions where misclassifications occur. This suggests that the model is capable of effectively learning the features and temporal patterns in the bird sound data, leading to accurate multi-label classification results.

Table 7. Confusion matrix for different labels.

Confusion matrix for label 0		
Label 0: real	Predict	
	0	1
0	5661	5
1	2	1332
Confusion matrix for label 1		
Label 1: real	Predict	
	0	1
0	5082	3
1	2	1913
Confusion matrix for label 2		
Label 2: real	Predict	
	0	1
0	5154	8
1	11	1827
Confusion matrix for label 3		
Label 3: real	Predict	
	0	1
0	5284	0
1	4	1712
Confusion matrix for label 4		
Label 4: real	Predict	
	0	1
0	5318	4
1	23	1655
Confusion matrix for label 5		
Label 5: real	Predict	
	0	1
0	5379	0
1	16	1605
Confusion matrix for label 6		
Label 6: real	Predict	
	0	1
0	5440	0
1	6	1554
Confusion matrix for label 7		
Label 7: real	Predict	
	0	1
0	5543	0
1	0	1457
Confusion matrix for label 8		
Label 8: real	Predict	
	0	1
0	5620	0
1	13	1367
Confusion matrix for label 9		
Label 9: real	Predict	
	0	1
0	5647	0
1	20	1333

5. Discussion

Table 8 shows a comparison of the relevant studies with the present study in terms of method and perfor-

mance. In multi-label bird species classification tasks, syllable overlap can limit manual feature extraction (LIU, 2016; BRIGGS *et al.*, 2012; NOUMIDA, RAJAN, 2022; LENG, DAN TRAN, 2014; ABDUL KAREEM, RAJAN, 2023), because syllable segmentation is a crucial step. The accuracy of any classifier that relies on segmentation is sensitive to the quality of the segmentation (FAGERLUND, 2004). A recent study has shown that deep learning is an effective method for classifying birds based on their sounds, such as processing large amounts of audio data, which allows it to detect subtle differences between bird sounds (MICHAUD *et al.*, 2023). Researchers usually increase the number of convolutional layers to extract more detailed features from the audio raw waveform (BRAVO SANCHEZ *et al.*, 2021) and mel spectrograms (ABDUL KAREEM, RAJAN, 2023). These methods result in more training parameters and the need for sufficient data to train model parameters. To reduce the number of trainable parameters and address the issue of data availability for the deep convolutional network to be effectively trained, a transfer learning approach is adopted in this study. For multi-label bird species classification, we employed the ImageNet-trained InceptionV3 convolution network. However, the CNN model ignores the temporal dependence of bird sounds. The pre-trained InceptionV3 is further fused with LSTM to extract time series characteristics from the feature sequence. Table 7 demonstrates that the proposed multi-label bird species classification method based on pre-trained InceptionV3 fused with LSTM has excellent performance.

The experimental results demonstrate the superiority of the InceptionV3+LSTM model in multi-label bird species classification, particularly in challenging cases involving overlapping syllables from two or three bird species. This confirms that combining convolutional feature extraction with temporal modeling via LSTM yields significant benefits over CNNs alone.

However, despite the strong performance, the proposed method relies on a large number of parameters and pre-trained models trained on image datasets (ImageNet), which may not optimally capture the characteristics of audio spectrograms. Furthermore, although we simulate overlapping bird calls, the synthetic nature of the dataset may not fully capture the complexities of real-world soundscapes such as environmental noise or unpredictable call patterns.

Compared to previous studies listed in Table 8, our method achieves state-of-the-art performance, yet direct comparisons remain difficult due to the diversity of datasets, label types, and evaluation metrics. Future work could benefit from the establishment of standardized multi-label bird audio benchmarks and the integration of more audio-specialized architectures, such as attention-based transformers or audio foundation models.

Table 8. Comparative analysis with other methods.

Reference work	Method	Dataset	Performance
LIU (2016)	Based on MFCC feature transfer	NIPS4B and xeno-canto	Hamming loss 0.1024
BRIGGS <i>et al.</i> (2012)	Spectral features with MIML-KNN	H.J. Andrews Experimental Forest	Accuracy 96.1 %
NOUMIDA, RAJAN (2022)	MFCC with BiGRU	xeno-canto	<i>F1</i> -score 0.85
LENG, DAT TRAN (2014)	Spectral features, MFCC and LPC with ensemble model	NIPS4B	AUC 91.74 %
BRAVO SANCHEZ <i>et al.</i> (2021)	SincNet	NIPS4Bplus	Accuracy 73.56 %, AUC 74.85 %, Precision 74.81 %, Recall 73.56 %
ABDUL KAREEM, RAJAN (2023)	Fused the MFCC-RNN and mel spectrogram-CNN	xeno-canto	<i>F1</i> -score 0.75
Proposed method in this work	Pre-trained InceptionV3 with LSTM	Cornell Macaulay Library	Exact match ratio 98.25 %, Accuracy 99.32 %, Recall 99.42 %, Precision 99.90 %, <i>F1</i> -score 99.57 %, Hamming loss 0.0062

6. Conclusions

In recent years, research on multi-label bird sound classification has been limited, particularly for realistic scenarios where two or three bird species vocalize simultaneously within the same audio segment. Moreover, most existing works rely on researcher-constructed datasets due to the lack of publicly available multi-label bird datasets, which makes performance comparison and method validation challenging. To address these issues, this study proposed a multi-label bird species classification model based on a transfer learning architecture fused with LSTM. Specifically, our method focuses on the realistic challenge of identifying two or three bird species vocalizing simultaneously within the same 5-second audio clip. Traditional syllable segmentation methods often struggle in such overlapping scenarios. By applying pre-trained convolutional neural networks to extract deep acoustic features from mel spectrograms, and further integrating LSTM to capture temporal dependencies, the proposed model effectively addresses this challenge. Experimental results demonstrate that the InceptionV3+LSTM fusion model achieves outstanding performance in multi-label classification, with an exact match ratio of 98.25 %, accuracy of 99.32 %, recall of 99.42 %, precision of 99.90 %, *F1*-score of 99.57 %, and the minimum Hamming loss of 0.0062.

CONFLICT OF 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.

References

1. ABDUL KAREEM N., RAJAN R. (2023), Multi-label bird species classification using sequential aggregation strategy from audio recordings, *Computing and Informatics*, **42**(5): 1255–1280, <https://doi.org/10.31577/cai.2023.5-1255>.
2. BRAVO SANCHEZ F.J., HOSSAIN M.R., ENGLISH N.B., MOORE S.T. (2021), Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture, *Scientific Reports*, **11**: 15733, <https://doi.org/10.1038/s41598-021-95076-6>.
3. BRIGGS F. *et al.* (2012), Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach, *The Journal of the Acoustical Society of America*, **131**(6): 4640–4650, <https://doi.org/10.1121/1.4707424>.
4. CHENG Y., MA M., LI X., ZHOU Y. (2021), Multi-label classification of fundus images based on graph convolutional network, *BMC Medical Informatics and Decision Making*, **21**: 82, <https://doi.org/10.1186/s12911-021-01424-x>.
5. DENG J., DONG W., SOCHER R., LI L.J., LI K., LI F.F. (2009), ImageNet: A large-scale hierarchical image database, [in:] *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, <https://doi.org/10.1109/CVPR.2009.5206848>.
6. FAGERLUND S. (2004), *Automatic recognition of bird species by their sounds*, MSc. Thesis, Helsinki University of Technology.
7. GODBOLE S., SARAWAGI S. (2004), Discriminative Methods for Multi-labeled Classification, [in:] *Advances in Knowledge Discovery and Data Mining. PAKDD 2004. Lecture Notes in Computer Science*, Dai H., Srikant R., Zhang C. [Eds.], **3056**: 22–30, https://doi.org/10.1007/978-3-540-24775-3_5.
8. GÓMEZ-GÓMEZ J., VIDAÑA-VILA E., SEVILLANO X. (2023), Western Mediterranean Wetland Birds dataset:

- A new annotated dataset for acoustic bird species classification, *Ecological Informatics*, **75**: 102014, <https://doi.org/10.1016/j.ecoinf.2023.102014>.
9. GUNAWAN K.W., HIDAYAT A.A., CENGGORO T.W., PARDAMEAN B. (2021), A transfer learning strategy for owl sound classification by using image classification model with audio spectrogram, *International Journal on Electrical Engineering and Informatics*, **13**(3): 546–553, <https://doi.org/10.15676/ijeei.2021.13.3.3>.
 10. HE K., ZHANG X., REN S., SUN J. (2016), Deep residual learning for image recognition, [in:] *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
 11. HUANG Y.-P., BASANTA H. (2021), Recognition of endemic bird species using deep learning models, *IEEE Access*, **9**: 102975–102984, <https://doi.org/10.1109/ACCESS.2021.3098532>.
 12. LENG Y.R., DAT TRAN H. (2014), Multi-label bird classification using an ensemble classifier with simple features, *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, pp. 1–5, <https://doi.org/10.1109/APSIPA.2014.7041649>.
 13. LI G., JI Z.F., CHANG Y.L., LI S., QU X.D., CAO D.P. (2021), ML-ANet: A transfer learning approach using adaptation network for multi-label image classification in autonomous driving, *Chinese Journal of Mechanical Engineering*, **34**: 78, <https://doi.org/10.1186/s10033-021-00598-9>.
 14. LIU A. *et al.* (2021), Residual recurrent CRNN for end-to-end optical music recognition on monophonic scores, arXiv, <http://arxiv.org/abs/2010.13418>.
 15. LIU H.T. (2016), *A study on multi-label transfer learning algorithm and application in the bird sounds recognition*, Msc. Thesis, Nanjing Forestry University.
 16. MICHAUD F., SUEUR J., LE CESNE M., HAUPERT S. (2023), Unsupervised classification to improve the quality of a bird song recording dataset, *Ecological Informatics*, **74**: 101952, <https://doi.org/10.1016/j.ecoinf.2022.101952>.
 17. NISHIKIMI R., NAKAMURA E., GOTO M., YOSHII K. (2021), Audio-to-score singing transcription based on a CRNN-HSMM hybrid model, *APSIPA Transactions on Signal and Information Processing*, **10**(1): e7, <https://doi.org/10.1017/ATSIP.2021.4>.
 18. NOUMIDA A., RAJAN R. (2022), Multi-label bird species classification from audio recordings using attention framework, *Applied Acoustics*, **197**: 108901, <https://doi.org/10.1016/j.apacoust.2022.108901>.
 19. PANIRI M., DOWLATSHAHI M.B., NEZAMABADI-POUR H. (2020), MLACO: A multi-label feature selection algorithm based on ant colony optimization, *Knowledge-Based Systems*, **192**: 105285, <https://doi.org/10.1016/j.knsys.2019.105285>.
 20. SAINATH T.N., VINYALS O., SENIOR A., SAK H. (2015), Convolutional, long short-term memory, fully connected deep neural networks, [in:] *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4580–4584, <https://doi.org/10.1109/ICASSP.2015.7178838>.
 21. SEVILLA A., GLOTIN H. (2017), Audio bird classification with Inception-v4 extended with time and time-frequency attention mechanisms, *Working Notes of CLEF 2017 – Conference and Labs of the Evaluation Forum*, Cappellato L., Ferro N., Goeuriot L., Mandl T. [Eds.], **1866**, https://ceur-ws.org/Vol-1866/paper_177.pdf.
 22. SIMONYAN K., ZISSERMAN A. (2014), Very deep convolutional networks for large-scale image recognition, arXiv, <http://arxiv.org/abs/1409.1556>.
 23. SOROWER M.S. (2010), A literature survey on algorithms for multi-label learning.
 24. SPRENGEL E., JAGGI M., KILCHER Y., HOFMANN T. (2016), Audio based bird species identification using deep learning techniques, *Working Notes of CLEF 2016 – Conference and Labs of the Evaluation forum*, Balog K., Cappellato L., Ferro N., Macdonald C. [Eds.], **1609**, <https://ceur-ws.org/Vol-1609/16090547.pdf>.
 25. SZEGEDY C., IOFFE S., VANHOUCKE V., ALEMI A. (2017), Inception-v4, Inception-ResNet and the impact of residual connections on learning, [in:] *Proceedings of the AAAI Conference on Artificial Intelligence*, **31**(1), <https://doi.org/10.1609/aaai.v31i1.11231>.
 26. SZEGEDY C., VANHOUCKE V., IOFFE S., SHLENS J., WOJNA Z. (2016), Rethinking the Inception Architecture for Computer Vision, [in:] *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818–2826, <https://doi.org/10.1109/CVPR.2016.308>.
 27. TAO J., FANG X. (2020), Toward multi-label sentiment analysis: a transfer learning based approach, *Journal of Big Data*, **7**: 1, <https://doi.org/10.1186/s40537-019-0278-0>.
 28. WEISS K., KHOSHGOFTAAR T.M., WANG D. (2016), A survey of transfer learning, *Journal of Big Data*, **3**: 9, <https://doi.org/10.1186/s40537-016-0043-6>.
 29. ZHANG L., TOWSEY M., XIE J., ZHANG J., ROE P. (2016), Using multi-label classification for acoustic pattern detection and assisting bird species surveys, *Applied Acoustics*, **110**: 91–98, <https://doi.org/10.1016/j.apacoust.2016.03.027>.

Research Paper

Determination of the Elastic Constant of the Top Plate of a Cello in the Interaction with the Bridge

Pablo PAUPY^{(1)*}, Pablo TABLA⁽²⁾, Dario HUGGENBERGER⁽¹⁾, Federico ELFI⁽¹⁾,
Eneas N. MOREL^{(2),(3)}, Jorge R. TORGA^{(2),(3)}

⁽¹⁾ *Grupo de Vibraciones, Facultad Regional Delta, Universidad Tecnológica Nacional
Campana, Buenos Aires, Argentina*

⁽²⁾ *Grupo de Fotónica Aplicada, Facultad Regional Delta, Universidad Tecnológica Nacional
Campana, Buenos Aires, Argentina*

⁽³⁾ *Consejo Nacional de Investigaciones Científicas y Técnicas
CABA, Buenos Aires, Argentina*

*Corresponding Author e-mail: ppaupy@frd.utn.edu.ar

*Received July 13, 2024; revised January 29, 2025; accepted February 6, 2025;
published online April 7, 2025.*

This paper aims to determine the equivalent static elastic constant of a cello's top plate in the interaction with the bridge. Experimental results detailing this constant are presented based on measuring the deformation and forces caused by a system of calibrated springs in similar conditions to that obtained when these forces are produced by the action of the strings. Subsequent tests are conducted following an intervention by a luthier to adjust the sound post, with the aim of assessing the impact on the elastic constants.

Keywords: elasticity; organology; cello bridge; musical acoustics; boundary conditions; interferometry.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The main function of the cello bridge is to transform the vibrations generated in the strings into vibrations of the top plate. The strings transmit this vibration to the bridge at the points of contact between the two elements, and the bridge transmits it to the top plate through the two supporting feet. The way the bridge performs this transmission and what its modes of vibration are like, are central issues in the final characteristics of the instrument and have been studied by several authors. MINNAERT and VLAM (1937) published a pioneering work, where the normal, flexural, and torsional modes of the bridge were studied. Later, in 1963, Steinkopf introduced basic mechanical models of the bridge to obtain the frequency response (CREMER, 1984). BISSINGER (2006) carried out studies of various violin bridges and a detailed analysis of the bridge as a filter in the transmission of vibrations to the top plate of the instrument. Among the works with

experimental results, REINICKE and CREMER (1970) can be mentioned as one of the first to use optical interferometric techniques to measure vibrations in the instrument. Subsequently, JANSSON *et al.* (1994) made interesting contributions to the experimental analysis of violin body vibrations and their effects on the frequency response. In (JANSSON, 2004) the importance of the shape and dimensions of the bridge base and foot on the characteristics of the coupling with the top plate of the instrument is determined, demonstrating a correlation between the quality of the violin and the shape of the bridge. Other researchers have analysed the frequency response of the bridge by studying the admittance or mobility variables of the system (BOUTILLON, WEINREICH, 1999; ELIE *et al.*, 2013; MALVERMI *et al.*, 2021). Simplified shapes of the instrument or some of its parts were modelled and the frequency response was measured or simulated, taking the bridge as a simple mass-spring system or considering more complex models (WOODHOUSE, 2005; 2014).

In a separate line of research, the impact of the violin and cello bridge shapes on their static and vibrational characteristics was examined through parametric modelling and simulations utilizing the finite element methods. An important point in these works is the boundary conditions to which the bridge is subjected in contact with the top plate of the instrument during normal operation. These conditions ranged from a fixed point to more realistic systems incorporating combinations of springs and dampers. Several models were considered where each foot of the bridge had up to three translational springs and three rotational springs (KABALA *et al.*, 2018). LODETTI'S *et al.* (2023) work served as a basis for a subsequent analysis on a cello.

The primary objective of this paper was to measure the static elastic constant of the instrument's top plate at the contact points with each of the bridge's support feet. The proposal aimed to measure the value of the elastic constant in the primary direction of displacement, which is perpendicular to the instrument's top plate, understanding this parameter is crucial as it plays a significant role in determining the boundary conditions that the bridge experiences at the contact points with the top plate.

Another goal of this study was to compare the values of the elastic constants measured at the aforementioned contact points, in two different scenarios: in the first, when the instrument is received from the factory; and in the second, after an adjustment made by a professional luthier. Industrially manufactured cellos are often assembled in music stores, and such assembly frequently requires additional adjustments by professional luthiers. These adjustments significantly influence the elastic conditions between the bridge and top plate. This step is crucial to ensure that the structural and acoustic characteristics of the instrument meet higher standards. As the measurements show, the values of the elastic constant after the luthier's work are considerably lower than before the adjustment, particularly at the treble foot (next to the sound post).

The experimental work was conducted on a setup that include the entire instrument under controlled laboratory conditions, with measurements of the elastic constant taken to assess the variation at different points along the top plate. This variation is due to the structural asymmetry present in both the bridge and the instrument's top plate. The asymmetry was influenced by the curvature of the upper part of the bridge and the stiffness disparity resulting from the positioning of the sound post near the treble support ('A' string 220 Hz), and the bass-bar close to the bass support ('C' string 65.4 Hz).

2. Experimental setup

To obtain the values of the elastic constants, an experimental setup was implemented to measure the

deformation of the top plate as a function of the force applied at a series of selected points on the instrument top plate. The deformation of a spring of known elastic constant was used to measure the force and a low coherence optical interferometer was used to measure the deformation distances. With this experimental scheme, deformation measurements were carried out under different conditions and results were obtained for the elastic constants at the selected points. From these measurements it was possible to verify a linear behaviour between the displacement and the force in the range of values of both magnitudes to which the instrument is subjected during its execution.

The strings used for this experiment, replacing the original ones, are Jargar[®] Classic Medium strings. The manufacturer gives the nominal tension of each string (C: 13.8 kg, G: 13.4 kg, D: 14.0 kg, A: 17.9 kg) under normal tuning conditions, with the tension values of the treble strings being higher than those of the bass strings.

The points where the top plate deformation was measured are close to the bridge support points. To ensure clarity, they are identified with numbers 1 to 6. Point 1 is located next to the footrest corresponding to the treble string (note A3 220 Hz) on the tailpiece side of the instrument. Point 2 is located on the right side of the bridge, always seen from the tailpiece side; point 3 on the side of the fingerboard opposite to the location of point 1 and so on in correlative order for points 4, 5, and 6 on the foot corresponding to the bass string. Figure 1 shows the location of these points on the top plate of the instrument.

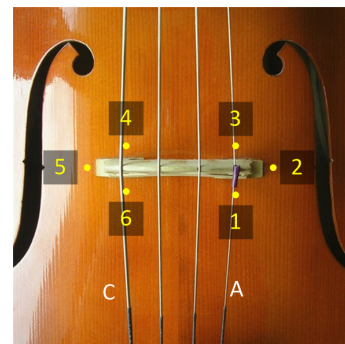


Fig. 1. Strain measurement points indicated on the top plate of the instrument.

Both, the experimental device for measuring force and deformation and the instrument used as sample, were mounted on an anti-vibration table that allows to isolate the system from mechanical noise and external vibrations. A structure was designed and constructed to secure the instrument to the table, providing support for the components used to apply and measure the compression force on the instrument's soundboard, as well as for the corresponding deformation measurement system.

2.1. Instrument fixing system

Clamping the instrument on the test table was carried out with a device specifically designed with three purposes. The first was to ensure that the instrument is statically fixed to the table so that the deformations measured are due exclusively to the deformation of the top plate and not to movements of the instrument with respect to the measuring point. The second was to use this structure for the location of the devices intended to exert and measure the force on the top plate and the corresponding deformation measuring device. The third was to preserve the instrument and develop a method of attachment and mounting that would not damage its structure.

The points where the straps are attached to the instrument's C's were used as a link to the fixation structure as they were considered to be structurally stronger. A mechanical clamp element was designed and constructed with separate upper and lower jaws. The lower part is screwed to the measuring table and the upper part presses the instrument top plate as shown in Fig. 2. Both the lower and upper-part press on wooden blocks covered with plush that are in direct contact with the instrument. In this way, the instrument is supported by this clamp system and no other parts are in direct contact with the measuring table. The points set out in the four blocks are the only elements linking the instrument to the measuring table. Figure 2 shows the instrument placed on the measuring table and the top plate compression system in the measuring situation.



Fig. 2. Cello mounted on the measuring table by means of the clamping frame.

2.2. Device for exerting and measuring forces

To exert and measure the compression force on the top plate of the instrument, two identical mechanical devices were designed and built to replace the original bridge and were placed at the same support points. In this configuration the cello is stripped of the bridge, tailpiece, and strings. Each of these devices has a support point on the top plate of the instrument, a cylinder with a piston associated to a calibrated thread of 1 mm of advance per turn, and an internal spring of known elastic constant (k_R) as it is showed in Fig. 3.

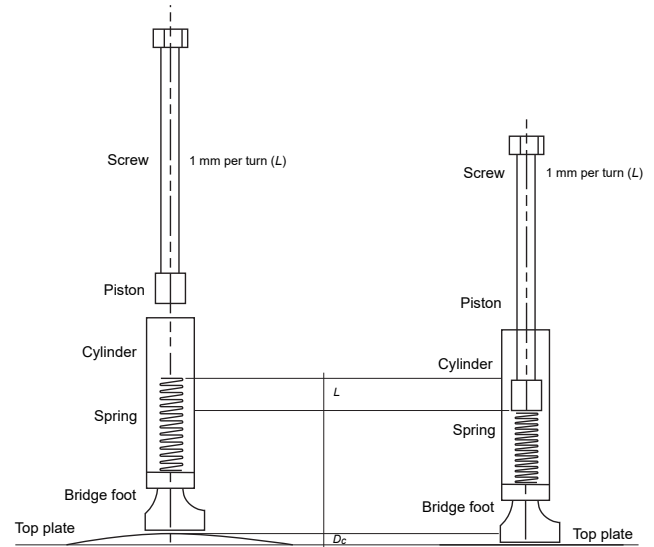


Fig. 3. Device diagrams to exert and measure the compression force on the top plate of the instrument.

To exert a controlled force, the piston is rotated a certain number of revolutions (n) compressing the spring at a distance L . At the other end of the spring, the support point exerts the same force on the instrument top plate. As a result, it deforms and moves a distance D_C , assuming the same behaviour that it has in the tuning and playing process. It is convenient to define the rate of deformation of the top plate per revolution (A_{pr}) of the cylinder as

$$A_{pr} = \frac{D_C}{n}. \quad (1)$$

Considering that the calibrated thread pitch is one millimetre per revolution (10^{-3} m/rev) with an error that we estimate at 1%; the distance L can be expressed as

$$L \pm \Delta L = 10^{-3} \frac{\text{m}}{\text{rev}} * n \pm \left(10^{-5} \frac{\text{m}}{\text{rev}} * n + 10^{-3} \frac{\text{m}}{\text{rev}} \Delta n \right), \quad (2)$$

where Δn , the error in the number of turns, was obtained from the relation that defines A_{pr} :

$$\Delta n = \frac{\Delta D_C}{A_{pr}} + \frac{D_C}{A_{pr}^2} \Delta A_{pr}. \quad (3)$$

The force exerted by the spring (F_R) at both ends can be expressed as

$$F_R \pm \Delta F_R = (k_R \pm \Delta k_R) \cdot ((L \pm \Delta L) - (D_C \pm \Delta D_C)), \quad (4)$$

where k_R , the elastic constant of the spring, was obtained by measuring the stretch produced by a series of standard weights in the range of values comparable to the stretch measured in tuning.

The average of these measurements, taken as the final value with a relative error of 1.6 %, was

$$k_R = (6100 \pm 100) \frac{\text{N}}{\text{m}}.$$

2.3. Deformation measurement with an interferometer

Deformation measurements of the top plate surface were made using frequency domain optical coherent tomography (FD-OCT) with a Fizeau-type fiber optic configuration (VAKHTIN *et al.*, 2003), as shown in Fig. 4a. This configuration allows measurements at points close to the feet of the bridge and allows simultaneous measurement of the deformation at two points. The interference between the optical reflection generated at the end of each optical fibre (F_a and F_b) and the reflections on the cello top plate allow to determine the distances between the end of the fiber and the top plate (D_a and D_b) as shown in Fig. 4b. D_a represents the distance measurements taken at points 1 and 6, while D_b represents the same measurements for points 4 and 3. The distances at points 2 and 5 were calculated as the average of the distances measured at 1 and 3, and 4 and 6, respectively.

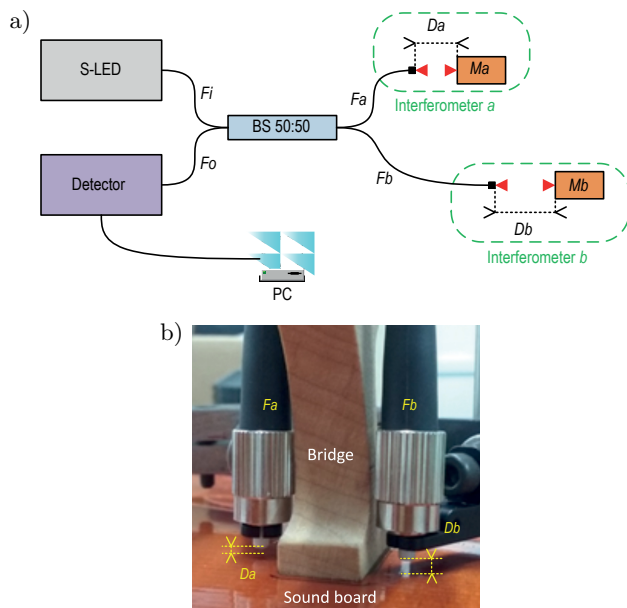


Fig. 4. a) Interferometer schematic. The light emitted by the S-LED source is coupled to the beam splitter (BS 50:50) whose outputs, fibers F_a and F_b , illuminates the cello top plate. D_a and D_b are the distance between each fiber and the top plate. Beams reflected are coupled back into the fibers and sent to the detector through the fiber F_o ; b) image of the optically instrumented cello.

3. Results

The process of measuring and determining the elastic constants of the top plate was carried out in three stages.

In the first stage, only the nominal displacements of the top plate produced by the bridge pressure under normal tuning conditions were measured. Deformation measurements were obtained at points 1, 3, 4, and 6 with the instrument mounted with tailpiece and strings. The deformation values at points 2 and 5 were obtained as the average of the values measured at 1 and 3, and 4 and 6, respectively. Measurements under tuning conditions were made by allowing the instrument to rest and noting that after about eight minutes the stabilisation of the deformation was essentially definitive within the instrumental resolution. A measurement immediately after compression and a second measurement after eight minutes of rest were taken as the norm.

In a second stage, the bridge and strings were replaced by a calibrated spring system. The forces applied with this system corresponded to deformation values equal to those obtained in the first stage. In this way, the values of force and deformation were obtained simultaneously in similar conditions to those achieved in the normal use of the instrument.

In a third stage the same measurements and results are presented as in the second stage but after the luthier has adjusted the instrument. The objective is to compare the values of the elastic constant before and after adjustment.

3.1. First stage – Measurement of deformation imposed by the tuning of strings (before luthier's work)

Table 1 presents the measurements of top plate deformation (D_C) produced during the tuning process in points 1, 3, 4, and 6.

The values labelled 'free' represent the measurements of distances (D_a or D_b) taken before tuning. In this state, the strings were completely slack, positioned between the pegs and the tailpiece, and resting on the bridge without any tension. This condition was maintained for more than eight minutes to ensure the top plate was free of residual tension. Measurements were then taken at points 1, 3, 4, and 6 near the bridge supports, establishing a baseline reference position. This static state served as the zero point for quantifying deformation caused by string tension after tuning.

The values in the column 'tuning 1' corresponds to the first measurement of the same distances (D_a or D_b) taken immediately after tuning the four strings. Eight minutes later a second measurement of the same distances was taken, referenced as 'tuning 2', and adjusted the four strings again with the micro tuning screw to obtain 'tuning 3', which was considered as the final measurement of distances resulting from tuning. This procedure was performed three times (measures 1, 2, and 3), at each of the points of the top plate.

Table 1. Absolute displacements and deformation at points 1, 3, 4, and 6 during tuning process.

Type	Free [μm]	Tuning 1 [μm]	Tuning 2 [μm]	Tuning 3 [μm]	D_{C-T1} [μm]	D_{C-T2} [μm]	D_{C-T3} [μm]
a) absolute displacements and deformation at point 1 during tuning process							
Measure 1	789	1252	1245	1256	463	456	467
Measure 2	797	1255	1252	1260	458	455	463
Measure 3	798	1248	1245	1272	450	447	474
Average	794.7	1251.7	1247.3	1262.7	457.0	452.6	468.0
Error	2.8	2.0	2.3	4.9	3.9	2.9	3.3
b) absolute displacements and deformation at point 3 during tuning process							
Measure 1	712	1394	1389	1402	682	677	690
Measure 2	717	1396	1394	1409	679	677	692
Measure 3	716	1414	1411	1419	696	695	703
Average	715.0	1401.3	1398.0	1410.0	685.7	683.0	695.0
Error	1.6	6.6	6.9	5.1	5.4	6.2	4.2
c) absolute displacements and deformation at point 4 during tuning process							
Measure 1	1132	1790 1787	1803	658	655	671	
Measure 2	1134	1796	1791	1807	662	657	673
Measure 3	1137	1807	1805	1816	670	668	679
Average	1134.3	1797.7	1794.3	1808.7	663.3	660.0	674.3
Error	2.6	5.1	5.6	4.0	3.6	4.1	2.5
d) absolute displacements and deformation at point 6 during tuning process							
Measure 1	1240	1888	1877	1891	648	637	651
Measure 2	1249	1897	1889	1906	648	640	657
Measure 3	1249	1886	1880	1907	637	631	658
Average	1246.0	1890.3	1882.0	1901.3	644.3	636.0	655.3
Error	3.1	3.5	3.7	5.3	3.8	2.7	2.3

The values of deformation referenced as ' D_{C-T1} ', ' D_{C-T2} ', and ' D_{C-T3} ' were obtained as the difference between 'tunings 1, 2, 3' and 'free' measurements, respectively. The average was taken as the representative value. The final error was obtained from the standard statistical error of the three deformations as indicated in the last column of the table, in the same way as all the errors indicated in the last row. The final tuning deformation result, utilized in subsequent calculations, is underlined.

3.2. Second stage – Measurement of deformation imposed by the spring compression system (before luthier's work)

At this stage, the bridge and strings were replaced by a system of calibrated springs to apply and measure

forces on the instrument. The deformation values (D_C) obtained in the previous stage were used as a reference to determine the necessary force that must be applied with the spring system in order to generate similar deformations. To apply the required force in each point, it is necessary to determine the value of the relative rate of deformation per revolution (A_{pr}) and the number of revolutions (n) of the cylinder of the spring device. To obtain these values each spring was compressed three times simultaneously at the points 1, 3, 4, and 6 and left to rest for eight minutes after each adjustment to make the measurements comparable to those made under tuning conditions. Table 2 summarises the three measurements at each point indicating the compression in microns and the number of revolutions between brackets. The average rate per revolution values is reported in the same table.

Table 2. Deformation measurements (D_C), number of revolutions (n), and the rate of advance per revolution (A_{pr}) in each series of measurements.

	Point 1	Point 3	Point 4	Point 6
Measure 1 – D_C (n)	483 μm (14 rev)	720 μm (19.5 rev)	721 μm (9.5 rev)	674 μm (9.5 rev)
Measure 2 – D_C (n)	482 μm (14 rev)	694 μm (20 rev)	683 μm (9 rev)	675 μm (9.5 rev)
Measure 3 – D_C (n)	478 μm (14 rev)	724 μm (20.5 rev)	675 μm (9 rev)	656 μm (9 rev)
Measure 1 – A_{pr}	34.50 $\mu\text{m}/\text{rev}$	36.92 $\mu\text{m}/\text{rev}$	75.89 $\mu\text{m}/\text{rev}$	70.95 $\mu\text{m}/\text{rev}$
Measure 2 – A_{pr}	34.43 $\mu\text{m}/\text{rev}$	34.70 $\mu\text{m}/\text{rev}$	72.56 $\mu\text{m}/\text{rev}$	75.00 $\mu\text{m}/\text{rev}$
Measure 3 – A_{pr}	34.14 $\mu\text{m}/\text{rev}$	35.32 $\mu\text{m}/\text{rev}$	75.00 $\mu\text{m}/\text{rev}$	72.89 $\mu\text{m}/\text{rev}$
Average – A_{pr}	(34.36 \pm 0.11) $\mu\text{m}/\text{rev}$	(34.65 \pm 0.66) $\mu\text{m}/\text{rev}$	(74.48 \pm 0.99) $\mu\text{m}/\text{rev}$	(74.95 \pm 1.65) $\mu\text{m}/\text{rev}$

The equivalent elastic constant of the top plate (k_T) can be defined as

$$F_T = k_T D_C. \quad (5)$$

It is assumed that the elastic force exerted by the top plate ' F_T ' is the same as that exerted by the calibrated spring ' F_R ', then:

$$F_R = F_T = k_T D_C = k_R(L - D_C). \quad (6)$$

So, it is possible to obtain the equivalent elastic constant of the top plate and its error as

$$k_T \pm \Delta k_T = \frac{F_R}{D_C} \pm \left(\frac{\Delta F_R}{D_C} + \frac{F_R}{D_C^2} \Delta D_C \right). \quad (7)$$

Finally, we have the results summarised in Table 3.

We estimate the value of the equivalent elastic constant of the top plate under the support of the foot as the average between the values measured at points 1 and 3 for the *La* (A) and for the *Do* (C) string:

$$k_{T(A)} = (171 \pm 12) \text{ kN/m},$$

$$k_{T(C)} = (75 \pm 5) \text{ kN/m},$$

with relative errors in the order of 8% and 7%, respectively.

Table 3. Synthesis of results at points 1, 3, 4, and 6 for top plate deformation (D_C), relative advance per turn (A_{pr}), estimated number of revolutions to equalize deformations (n), displacement length of the calibration spring (L), force exerted by the spring (F_R), equivalent elastic constant of the top plate (k_T), and its relative error (ER).

	Point 1	Point 3	Point 4	Point 6
D_C [μm]	468.0 \pm 3.3	695.0 \pm 4.2	674.3 \pm 2.5	655.3 \pm 2.3
A_{pr} [$\mu\text{m}/\text{rev}$]	34.36 \pm 0.11	34.65 \pm 0.66	74.48 \pm 0.99	74.95 \pm 1.65
n [rev]	13.62 \pm 0.14	20.06 \pm 0.51	9.05 \pm 0.16	8.74 \pm 0.23
L [mm]	13.62 \pm 0.27	20.06 \pm 0.71	9.05 \pm 0.25	8.74 \pm 0.32
F_R [N]	80.2 \pm 3.0	118.1 \pm 6.2	51.1 \pm 2.4	49.3 \pm 2.7
k_T [kN/m]	171.4 \pm 7.6	169.9 \pm 9.9	75.8 \pm 3.9	75.2 \pm 4.4
ER [%]	4.4	5.8	5.2	5.9

Table 4. Mean deformation with standard and percentage error after luthier's intervention.

	Point 1	Point 3	Point 4	Point 6
Average – D_C	513.0 \pm 4.3 μm	729.7 \pm 6.5 μm	910.7 \pm 5.0 μm	859.7 \pm 12.7 μm

Table 5. Deformation measurements (D_C), number of revolutions (n), and the rate of advance per revolution (A_{pr}) in each series of measurements after the luthier's intervention.

	Point 1	Point 3	Point 4	Point 6
Measure 1 – D_C (n)	574 μm (13 rev)	736 μm (14 rev)	931 μm (11 rev)	888 μm (11 rev)
Measure 2 – D_C (n)	530 μm (13 rev)	760.9 μm (14 rev)	969 μm (11.5 rev)	855 μm (11 rev)
Measure 3 – D_C (n)	522 μm (13 rev)	750 μm (14.5 rev)	926 μm (11 rev)	848 μm (11 rev)
Measure 1 – A_{pr}	44.16 $\mu\text{m}/\text{rev}$	52.57 $\mu\text{m}/\text{rev}$	84.64 $\mu\text{m}/\text{rev}$	80.73 $\mu\text{m}/\text{rev}$
Measure 2 – A_{pr}	40.77 $\mu\text{m}/\text{rev}$	54.35 $\mu\text{m}/\text{rev}$	84.26 $\mu\text{m}/\text{rev}$	77.73 $\mu\text{m}/\text{rev}$
Measure 3 – A_{pr}	40.15 $\mu\text{m}/\text{rev}$	51.72 $\mu\text{m}/\text{rev}$	84.18 $\mu\text{m}/\text{rev}$	77.09 $\mu\text{m}/\text{rev}$
Average – A_{pr}	(41.69 \pm 1.27) $\mu\text{m}/\text{rev}$	(52.88 \pm 0.79) $\mu\text{m}/\text{rev}$	(84.36 \pm 0.15) $\mu\text{m}/\text{rev}$	(78.52 \pm 1.14) $\mu\text{m}/\text{rev}$

3.3. Third stage – Measurement of deformation imposed by the spring compression system (after luthier's work)

In this section, we present the same measurement and calculus procedure as in the second stage but after the work done by a professional luthier adjusting the position and length of the sound post.

The appreciation of the luthier as soon as he worked on the instrument was that the sound post was 'too rigid' and that he had to shorten its length and position. After luthier's work we measured the static elastic constants again noticing that the values obtained were considerably smaller than before, especially in the treble foot (next to the sound post).

Table 4 shows the final deformation measurements of the strings during the tuning process after an intervention of a luthier consisting in the adjustment of the sound post location.

Table 5 repeats the results presented in Table 2 after luthier's intervention.

Table 6 presents a summary of the results obtained after the luthier's intervention. The last row shows a new magnitude, the stiffness reduction (Sr), defined as the percentage relative difference in k_T before (second stage) and after (third stage) luthier's intervention. It is clear that the most remarkable reductions occur at points 1 (18%) and 3 (36%), where the sound post has the greatest influence.

Table 6. Synthesis of results at points 1, 3, 4, and 6 for top plate deformation (D_C), relative advance per turn (A_{pr}), estimated number of revolutions to equalize deformations (n), displacement length of the calibration spring (L), force exerted by the spring (F_R), equivalent elastic constant of the top plate (k_T), relative error (ER), and stiffness reduction (Sr) after the intervention of the luthier.

	Point 1	Point 3	Point 4	Point 6
D_C [μm]	513.0 \pm 4.3	729.7 \pm 6.5	610.7 \pm 5.0	859.7 \pm 12.7
A_{pr} [$\mu\text{m}/\text{rev}$]	41.69 \pm 1.27	52.88 \pm 0.79	84.36 \pm 0.15	78.52 \pm 1.14
n [rev]	12.31 \pm 0.48	13.80 \pm 0.33	7.24 \pm 0.08	10.95 \pm 0.16
L [mm]	12.31 \pm 0.60	13.80 \pm 0.47	7.24 \pm 0.15	10.95 \pm 0.27
F_R [n]	72.0 \pm 4.9	79.7 \pm 4.3	40.4 \pm 1.7	61.6 \pm 2.8
k_T [kN/m]	140.4 \pm 10.8	109.2 \pm 7.3	66.2 \pm 3.4	71.7 \pm 4.32
ER [%]	7.7	6.7	5.1	6.0
Sr [%]	18	36	13	5

We estimate the value of the equivalent elastic constant of the top plate under the foot support corresponding to the string *La* (A) and *Do* (C):

$$k_{T(A)} = (125 \pm 15) \text{ kN/m},$$

$$k_{T(C)} = (69 \pm 6) \text{ kN/m},$$

with relative errors in the order of 12 % and 9 %, respectively.

4. Conclusions

An experimental system was developed to measure the force and displacement of the cello's top plate at points near the bridge supports and estimate its static elastic constant. Key findings and contributions are summarized as follows:

- measurement methodology:
 - a) top plate deformation was measured using low-coherence interferometry, a non-contact optical technique with sub-micron resolution;
 - b) the force was applied and measured using a calibrated spring system mounted on a custom-designed device, ensuring minimal spurious deformations and compatibility with standard instrument fixtures;
- phase 2 results:
 - a) before the luthier intervention, the measured static elastic constant (k_T) varied significantly across the different points [kN/m]:
 - * point 1: $k_T = 171.4 \pm 7.6$,
 - * point 3: $k_T = 169.9 \pm 9.9$,
 - * point 4: $k_T = 75.8 \pm 3.9$,
 - * point 6: $k_T = 75.2 \pm 4.4$;
 - b) error (ER) values ranged from 4.4 % to 5.9 %;
- phase 3 results:
 - a) after professional adjustment of the sound post by luthier (who identified excessive rigidity and corrected its position and

length), significant reductions in k_T were observed, particularly at point 3 [kN/m]:

- * point 1: $k_T = 140.4 \pm 10.8$ (\downarrow 18 %),

- * point 3: $k_T = 109.2 \pm 7.3$ (\downarrow 36 %),

- * point 4: $k_T = 66.2 \pm 3.4$ (\downarrow 13 %),

- * point 6: $k_T = 71.7 \pm 4.3$ (\downarrow 5 %);

b) ER values increased slightly (6.0 % to 7.7 %);

– impact of the luthier's intervention:

- a) the most pronounced stiffness reductions occurred near the treble foot up to 36 % in point 3 (string A, adjacent to the sound post), emphasizing the importance of proper sound post positioning and adjustment for achieving optimal elastic properties.

These findings provide valuable insights for cello setup optimization and serve as a reference for modelling the bridge using the finite difference methods. Future work will focus on refining measurement techniques, extending the analysis to additional instruments, and correlating elastic properties with acoustic performance.

This work highlights the significant contribution that the intuition and expertise of luthiers can make to scientific research aimed at understanding the functioning of musical instruments. Collaborative work, combined with the use of new techniques and scientific methods, offers the potential to provide objective insights into subjective aspects, thereby fostering the generation of new knowledge.

The results obtained using low-coherence interferometry suggest that this technique is highly suitable for measuring deformation at different points of the top plate. This paves the way for future studies focusing on other elements of the instrument.

ACKNOWLEDGMENTS

We would like to thank luthier Fabián Santillán for his kindness and willingness to collaborate on the adjustments of the cello.

References

1. BISSINGER G. (2006), The violin bridge as filter, *The Journal of the Acoustical Society of America*, **120**(1): 482–491, <https://doi.org/10.1121/1.2207576>.
2. BOUTILLON X., WEINREICH G. (1999), Three-dimensional mechanical admittance: Theory and new measurement method applied to the violin bridge, *The Journal of the Acoustical Society of America*, **105**(6): 3524–3533, <https://doi.org/10.1121/1.424677>.
3. CREMER L. (1984), *The Physics of the Violin*, The MIT Press, England.
4. ELIE B., GAUTIER F., DAVID B. (2013), Analysis of bridge mobility of violins, [in:] *Proceedings of the Stockholm Music Acoustics Conference 2013*, pp. 54–59, <https://hal.science/hal-01060528> (access: 3.06.2024).
5. JANSSON E., MOLIN N., SALDNER H. (1994), On eigenmodes of the violin – Electronic holography and admittance measurements, *The Journal of the Acoustical Society of America*, **95**(2): 1100–1105, <https://doi.org/10.1121/1.408470>.
6. JANSSON E.V. (2004), Violin frequency response – Bridge mobility and bridge feet distance, *Applied Acoustics*, **65**(12): 1197–1205, <https://doi.org/10.1016/j.apacoust.2004.04.007>.
7. KABALA A., NIEWCZYK B., GAPIŃSKI B. (2018), Violin bridge vibration – FEM, *Vibrations in Physical Systems*, **29**: 2018021, https://vibsys.put.poznan.pl/_journal/2018-29/articles/vibsys_2018021.pdf (access: 3.06.2024).
8. LODETTI L., GONZALEZ S., ANTONACCI F., SARTI A. (2023), Stiffening cello bridges with design, *Applied Sciences*, **13**(2): 928, <https://doi.org/10.3390/app13020928>.
9. MALVERMI R. *et al.* (2021), Feature-based representation for violin bridge admittances, arXiv, <https://doi.org/10.48550/arXiv.2103.14895>.
10. MINNAERT M., VLAM C.C. (1937), The vibrations of the violin bridge, *Physica*, **4**(5): 361–372, [https://doi.org/10.1016/S0031-8914\(37\)80138-X](https://doi.org/10.1016/S0031-8914(37)80138-X).
11. REINICKE W., CREMER L. (1970), Application of holographic interferometry to vibrations of the bodies of string instruments, *The Journal of the Acoustical Society of America*, **47**(4B): 131–132, <https://doi.org/10.1121/1.1912237>.
12. VAKHTIN A.B., KANE D.J., WOOD W.R., PETERSON K.A. (2003), Common-path interferometer for frequency-domain optical coherence tomography, *Applied Optics*, **42**(34): 6953–6958, <https://doi.org/10.1364/AO.42.006953>.
13. WOODHOUSE J. (2005), On the “bridge hill” of the violin, *Acta Acustica united with Acustica*, **91**(1): 155–165.
14. WOODHOUSE J. (2014), The acoustics of the violin: A review, *Reports on Progress in Physics*, **77**(11): 115901, <https://doi.org/10.1088/0034-4885/77/11/115901>.

Research Paper

Assessment of Noise from Different Types of Rounds Fired from a Saluting Gun

Gurmail S. PADDAN, Matt J. HOWELL

Institute of Naval Medicine
Hampshire, United Kingdom

*Corresponding Author e-mail: Gurmail.Paddan472@mod.gov.uk

*Received October 16, 2024; revised April 18, 2025; accepted May 25, 2025;
published online June 16, 2025.*

The difference in sound pressure levels between two types of rounds fired from a saluting gun has been investigated; the rounds being identified as ‘current’ and ‘new’. A 3-pounder saluting gun mounted on a concrete floor based at HMNB Portsmouth, UK, was used in the survey. Sound pressure levels were measured at the two people responsible for operating the gun: the firer and the loader. Twelve current rounds and 24 new rounds were fired during the survey. The new rounds showed a greater variation in peak sound pressure levels between rounds (interquartile range of 2.1 dB, firer’s location) compared with the current rounds (interquartile range of 1.1 dB, firer’s location). The highest *C*-weighted peak sound pressure levels for the firer were 173.1 dB for the current round compared with 166.8 dB for the new round. The corresponding highest *C*-weighted peak sound pressure levels for the loader were 170.6 dB and 163.0 dB, respectively. The difference between median peak sound pressure levels was 8.8 dB for the firer and 9.8 dB for the loader. Similar differences were measured in sound exposure levels between the two types of rounds. Frequency data presented can be used for assessing the suitability of appropriate hearing protectors. Mitigation measures are proposed for further reducing noise exposure of the operators.

Keywords: peak sound pressure level; 3-pounder round; saluting gun; blank ammunition.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The firing of a saluting gun serves a very specific purpose – this shows a sign of respect or welcome, celebratory or for remembrance. That is, the firing marks a commemoration of an event that occurred in the past, or of a current special occasion. One such important and notable occasion was on 9 September 2022 with the passing of Her Majesty the Queen Elizabeth II (the late monarch of the United Kingdom) whereby 96 rounds (or salutes) were fired during the Death Gun Salute from Hyde Park, London. The number of rounds corresponding to the age of Her Majesty upon death. Similar events with saluting guns and the same number of rounds took place at other locations such as Belfast (Northern Ireland), Cardiff (Wales), Edinburgh (Scotland), and Gibraltar. To mark the occasion, 117 rounds were fired from the saluting guns at Portsmouth Naval Base, UK: 96 to mark Her Majesty’s

age (upon her demise) and an additional 21 shots as a mark of respect. Other military bases, including Devonport Naval Base Plymouth, UK fired 96 shots to mark the occasion.

Indeed, there are many ceremonial royal occasions within United Kingdom when saluting guns are fired, these include Accession Day, His Majesty the King’s Birthday, His Majesty King Charles III’s official birthday, Her Majesty the Queen’s Birthday, The State Opening of Parliament, and meetings of visiting Heads of State and the Sovereign. The number of gun salutes vary but could include 21, 41, or 61 firings. The salutes are fired in 10-second intervals until all salutes are complete.

There are other guns which are fired on daily basis. One such example is the ‘One O’clock Gun’ fired from Edinburgh Castle every day at 13:00 to announce the time so that ships could synchronise their chronometric timepieces. Apart from a few exceptions, the gun

has been fired every day since 1861. Noting that the speed of sound is approximately $343 \text{ m} \cdot \text{s}^{-1}$, *Time gun-maps* (n.d.) were produced to account for the delay in the sound being heard at different distances from the gun. All this, however, is of a historic nature and only serves a symbolic purpose in this technological age. Unsuccessful attempts were made to quieten the noise from the gun on health and safety grounds (SHERIDAN, 2024).

Only a few published studies were found dealing with the sound pressure levels emanating from the firing of saluting or ceremonial guns and cannons. Primetake, a manufacturer of ammunition for use in ceremonial saluting guns, states that their 25-pounder cartridge (88 mm bore size) with full charge (454 g of gunpowder) would produce a sound pressure level of 160 dB at a distance of 20 m measured at an angle of 90° to the muzzle. The sound pressure level would reduce to 155 dB with a half charge (227 g). However, no data are provided for the locations of the gun operators who would be stationed behind the cannon.

The use of cannons is not restricted for just celebrating royal occasions in the UK. Various types of cannons are regularly used at some universities in the USA (and other countries) to celebrate and commence activities. Peak sound pressure levels of 174 dB have been reported at the operators' locations from ceremonial cannons (FLAMME *et al.*, 2019). Based on the findings, it was stated that the sound pressure levels could be 'potentially hazardous' and that double hearing protection (earplugs and earmuffs) should be worn by those involved in the firings.

The effect and purpose of firing a saluting gun should serve both as an auditory and a visual signal to those in the local vicinity; although the visual effect would only be observable when in line-of-sight with the saluting gun. The event of firing should not cause any harm or injury to any person involved in the firing or observing the firing. The hazards from this activity might include the noise generated, the flash and emission of any gases. It is considered that these three hazards would not pose a significant risk to any observer as there would normally be an area cordoned off around the gun. However, those involved with the firing, nominally the firer and the loader, would necessarily be exposed to these hazards due to the close proximity with the gun. The main premise with regards to noise exposure should be that the firing of the salutes must not affect the hearing of those observing the firing (spectators) and operating the gun.

The noise generated during firing of saluting guns exposes the operators to undesirable sound pressure levels which pose a risk to their hearing. An assessment of the noise from firing of saluting guns based at HMNB Portsmouth showed that the operators might be exposed to *C*-weighted peak sound pressure levels around 178 dB (PADDAN, HOWELL, 2019). The peak

sound pressure produced during firing was considered to be a potential noise risk and this necessitated the need for control measures. One of the main mitigation measures to reduce noise exposure of the operators included recommending the use of a lower charge in the ammunition used in the saluting guns with the aim of maintaining the visual display (flash), but reducing the noise produced (bang). Therefore, based on the recommendation, an opportunity arose to assess the noise from saluting guns firing ammunition with two different types of charge; that is, using the current charge and a new improved lower charge. The data have been assessed using the available guidance specified in (Statutory Instruments, 2005; Directive 2003/10/EC, 2003). Furthermore, the suitability of hearing protection has been evaluated using the procedure in (Ministry of Defence, 2015) using maximum sound pressure levels.

2. Equipment and procedure

2.1. Saluting gun and ammunition

Noise measurements were made during the firing of the 3-pounder saluting guns mounted on concrete platforms at the South Railway jetty at HMNB Portsmouth, Hampshire, UK (Fig. 1). The three land-based guns were aimed towards the water between Portsmouth and Gosport. One of the guns, shown as the gun on the right in Fig. 1 identified as the 'northern gun', was used during the survey. The block number (serial number) of the gun was 538. The gun was manufactured by the Royal Naval Armament Depot, Plymouth. Other identifying marks included 704.R.N.A.D. PLY.1960.



Fig. 1. Saluting guns at HMNB Portsmouth, UK. The right-most gun was fired during the survey.

Two types of 3-pounder blank ammunition identified as 'current' and 'new' were fired during the survey. Further details are shown in Table 1. The survey was conducted such that 24 rounds of the new ammunition were fired first followed by 12 rounds of the current ammunition. There was a gap of about 7 s between successive rounds.

Table 1. Two types of 3-pounder blank ammunition used in the survey.

Type	Identification	No. of round fired	NSN*	Charge	Casing
L1A1 with primer No. 20 L22-A1	Current	12	1305-99-701-2144	11 oz (312 g)	Brass
L6-A1	New	24	1310-98-209-0397	90 g Pyrodex	Plastic

*NSN is the NATO Stock Number, essentially a part number to uniquely identify equipment used by NATO military services.

The gun was operated by two people identified as the ‘firer’ and the ‘loader’. Both people were kneeling with the firer immediately behind the gun and the loader slightly to the right side of the gun. The firer was facing in the direction of the barrel with his head approximately 1.2 m above the concrete mounting base of the gun. The loader faced backwards to the gun and his head was about 1.4 m above the ground (he had the same stance as the firer but was on the plinth). (Although these two workers are identified as the ‘firer’ and the ‘loader’, firing a round from the gun involved the following operations: the loader pushes down a lever (seen on the right of the gun in Fig. 1) thus opening access to the barrel, the firer inserts the ammunition into the barrel, the loader pulls up the lever thus closing access to the barrel, the firer pulls the trigger to fire the gun and discharge the round.) Both operators were fully protected with suitable clothing and appropriate gloves including anti-flash hoods over their faces. They wore double hearing protection comprising earplugs (E-A-R soft FX) and earmuffs (Peltor Comtac XPI).

Meteorological data (taken from [Ventusky](#)) were recorded which showed the environmental conditions on the day to be a temperature of 9 °C, wind speeds of up to 18 km · h⁻¹, relative humidity of 70 % and no precipitation.

2.2. Noise measurements

Audio recordings were made near the left ears (about 15 cm away) of the firer and the loader, in accordance with UK Health and Safety Executive guidance on measuring peak noise ([Health and Safety Executive, 2021](#)). The microphones for measuring sound pressure levels were mounted on tripods at a 90° incidence (microphone diaphragm parallel to the sound) to the barrel of the gun. Measurements were made using 1/4" high pressure microphones (40BH, GRAS, Denmark) connected to microphone preamplifiers (26AC, GRAS, Denmark). The microphones were fitted with spherical foam windscreens approximately 65 mm in diameter with a 5-mm hole (WQ-1099, Brüel & Kjær, Denmark). The preamplifiers were connected to the input channels of a microphone power supply (12AA, GRAS, Denmark) and then to the data acquisition and analysis system (DATS Tetrad, Prosig, United Kingdom). The output from the Tetrad acquisition system was connected to a laptop running DATS for Windows software (v4.10.01) where 24-bit rate time histories were

acquired simultaneously at a sampling rate of 100 000 samples per second.

Calibration of the complete recording system was carried out using a class 1 pistonphone (42AC, GRAS, Denmark), which gave a sinusoidal calibration tone of 134 dB at a frequency of 250 Hz. The calibration procedure was repeated following the measurement of noise from the gun which showed the equipment to be stable over the measurement period.

2.3. Procedure

The 24 rounds of the new ammunition were fired first followed by the 12 rounds of the current ammunition. There was a 7-second gap between the successive rounds to ensure that there was no interference between the sounds from consecutive rounds. A single computer sound file was saved for each type of round: two time-history waveform files were acquired. Each file was separated into 24 segments for the new ammunition and 12 segments for the current ammunition (each of 0.9 s (±0.1 s) duration) to show the individual rounds in preparation for analysis.

2.4. Analysis of recordings

The human ear does not respond equally at all frequencies, therefore the *A*-weighting is applied to the audible frequency range to represent the reduction in sensitivity to the low frequencies. The *C*-weighting filter is suitable for assessing peak sound pressure levels. Long-term noise-induced hearing loss from moderate to loud noise is highly correlated with the noise exposure in dB(A). The mechanism of instant damage to the ear for extremely loud noise is different and is related to peak *C*-weighted sound pressure levels, L_{Cpeak} ([Health and Safety Executive, 2021](#)). The equivalent continuous sound pressure levels (that is, the time-averaged noise levels), L_{Aeq} and L_{Ceq} , are generally calculated for the measured time histories. However, when assessing noise from single events, such as from weapons fire, the L_{Aeq} and L_{Ceq} will be dependent on the period of measurement. In such a case, the *A*- and *C*-weighted sound exposure levels are calculated, L_{AE} and L_{CE} ([Health and Safety Executive, 2021](#)), indicating the total energy of the signal normalised to a 1-s period. If the measurement comprises a single round, then the L_{AE} is a measure of noise dose per round; the daily noise exposure can be calculated from the number of rounds and the L_{AE} per

round. Maximum A - and C -weighted sound pressure levels, the $L_{AF\max}$ and $L_{CF\max}$, were also measured with the time weighting ‘ F ’ (with a time constant corresponding to 0.125 s) which replicates the ‘fast’ meter response of older analogue sound level meters. The time-domain data were processed using a script written in Python[®] Programming Language. (The results from the Python[®] script were the same as those when compared with commercially available analysis software including DATS Prosig, *HVLab* (ISVR, University of Southampton) and Brüel & Kjær BZ-5503 ‘measurement partner suite’ software.)

The different parameters calculated were selected so that the sound exposure levels (L_{AE} and L_{CE}) could provide the energy within the waveform signals; and the maximum sound pressure levels ($L_{AF\max}$ and $L_{CF\max}$) could be used for an estimation of the suitability of hearing protection in attenuating the peak noise (Ministry of Defence, 2015).

3. Results

Figure 2 shows example time histories of one current and one new round fired from the saluting

gun as measured at the firer’s location. Each time history is of 0.1-second duration encompassing the complete waveform for each round. Four, possibly five, positive peaks are seen in the first 0.006 s of the waveform for the round using current ammunition. The peak sound pressure measured for the current round was 11642 Pa corresponding to 175.3 dB, i.e., unweighted (or dB(Z)); this corresponds to the direct sound transmitted from the muzzle of the cannon. The corresponding C -weighted peak sound pressure level is 172.0 dB; this being lower as the C -weighting will reduce the value. The highest peak (minimum) sound pressure level was -7198 Pa (171.1 dB, Z -weighted). The sound pressure had reduced to below 10% of the peak after about 0.015 s from the start of the waveform (this is referred to as the B -duration; (Department of Defence, 1997)). (The B -duration is defined as the time interval between the peak (either negative or positive) and the last point on the waveform where the value of the pressure has reduced to 10% of the peak value with succeeding values remaining below 10% of the peak value (COLES *et al.*, 1968).) The shape of the waveform for the new round is similar to that for the current ammunition, but the values are lower.

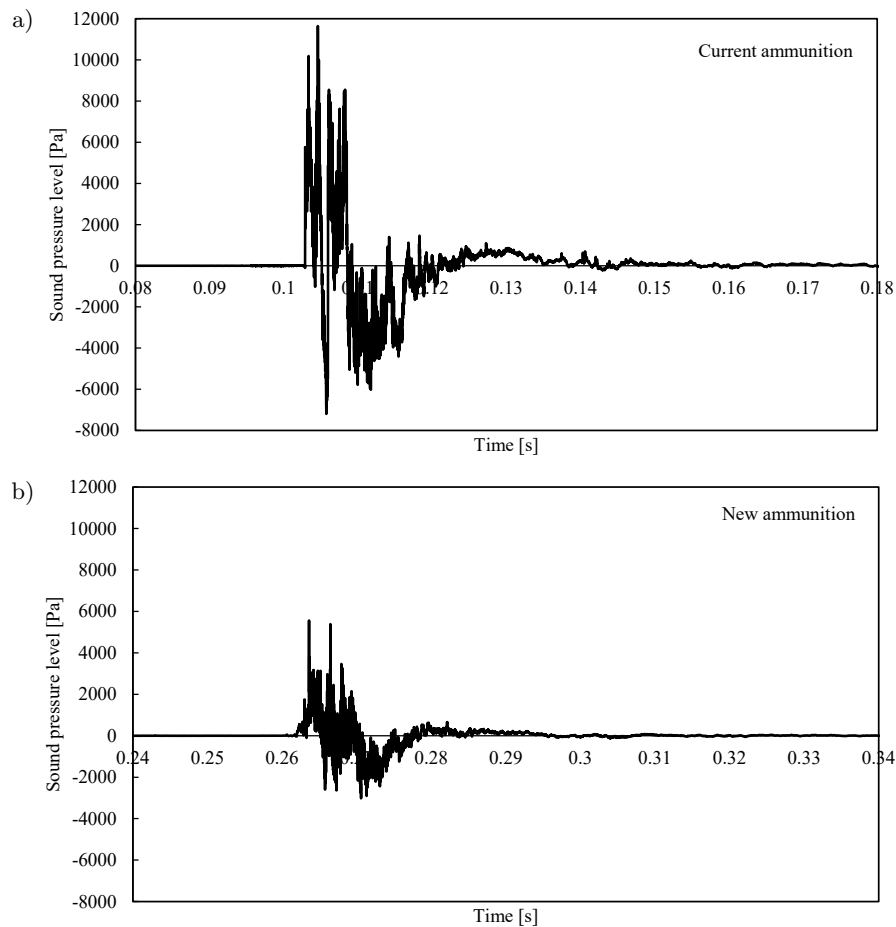


Fig. 2. Sound pressure levels measured near the head of the firer during firing of current and new rounds from the saluting gun.

Table 2. Different parameters for the sound pressure levels measured during firing of current (12 off) and new (24 off) rounds from the saluting gun.

Operator	Ammunition	Parameter	L_{Cpeak} [dB]	L_{AE} [dB]	L_{CE} [dB]	L_{AFmax} [dB]	L_{CFmax} [dB]
Firer	Current	Minimum	170.4	138.7	144.3	147.2	152.6
		Maximum	173.1	142.3	145.9	150.8	154.3
		Median	171.7	141.3	145.4	149.8	153.7
		Interquartile	1.1	0.8	0.4	0.8	0.4
	New	Minimum	158.8	128.2	131.2	136.7	139.6
		Maximum	166.8	135.4	140.7	144.0	149.1
		Median	162.9	131.8	136.1	140.3	144.5
		Interquartile	2.1	1.8	2.3	1.8	2.4
Loader	Current	Minimum	168.1	133.5	141.7	141.9	150.0
		Maximum	170.6	137.3	143.3	145.8	151.7
		Median	169.7	136.6	142.8	145.0	151.2
		Interquartile	1.1	0.7	0.2	0.7	0.2
	New	Minimum	154.2	122.7	128.7	131.1	137.1
		Maximum	163.0	129.4	137.7	138.0	146.2
		Median	159.9	126.8	134.3	135.2	142.8
		Interquartile	3.1	2.8	2.0	2.9	2.0

The peak sound pressure for the new round was 5559 Pa (unweighted (Z) value of 168.9 dB, C -weighted value of 164.1 dB). The B -duration for the new round was 0.019 s. It is noted that although B -durations are presented for the different waveforms, these are not used in the assessment procedures. Both waveforms show many peaks and troughs of unknown origins, possibly relating to reflections from the various structures around the gun (such as the other guns, and solid concrete bases as shown in Fig. 1).

Table 2 presents the sound pressure levels measured for the firer and loader locations with the current and new ammunition rounds. Various parameters based on the number of rounds fired are also presented; 12 rounds of the current type and 24 rounds of the new type. The highest C -weighted peak sound pressure level measured with the current type of round

was 173.1 dB, and that with the new type of round was 166.8 dB; both measurements being at the firer’s location.

3.1. Peak sound pressure level, L_{Cpeak}

Figure 3 shows the peak sound pressure levels, L_{Cpeak} , measured at the firer’s and loader’s locations during the firing of both current and new rounds. The variation between individual rounds can be seen. It is clear that the current rounds, shown as red and green circles, show higher values compared with the new rounds, shown as blue and yellow circles. The greatest difference between the rounds is seen for measurements at the loader with new rounds: round 6 shows a C -weighted peak of 154.2 dB compared with 163.0 dB for round 4 – a difference of 8.8 dB. The new rounds showed higher variation (interquartile range)

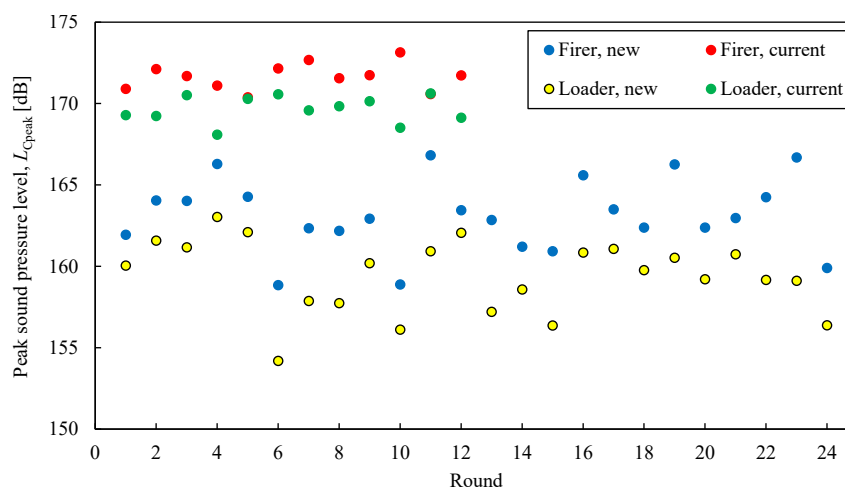


Fig. 3. Peak sound pressure levels (L_{Cpeak}) measured at the firer and the loader during firing of current and new rounds from the saluting gun.

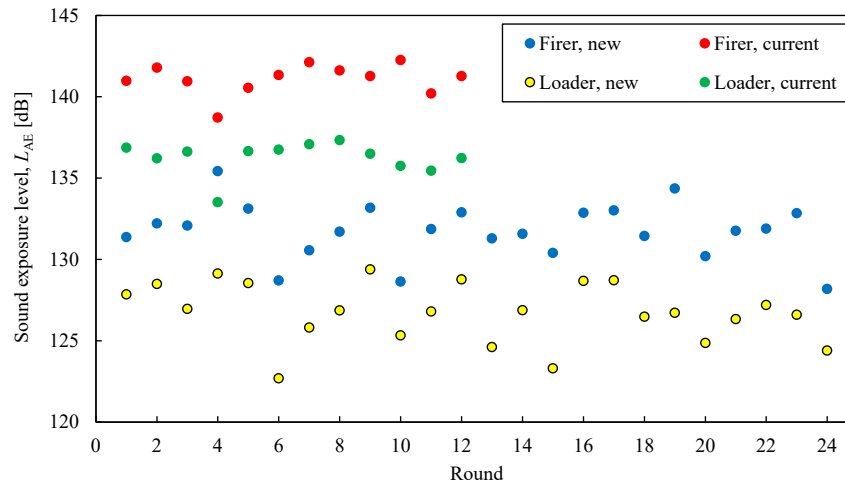


Fig. 4. Sound exposure levels (L_{AE}) measured at the firer and the loader during firing of current and new rounds from the saluting gun.

in peak sound pressure levels compared with the current rounds (see Table 2). Based on the median peak sound pressure levels, the L_{Cpeak} at the firer and the loader for the current ammunition was 8.8 dB and 9.8 dB higher, respectively, compared with the new rounds. Higher L_{Cpeak} values were measured for the firer compared with the loader ($p < 0.01$ for both types of rounds, Student's t -test): 2.0 dB higher for the current rounds and 3.0 dB higher for the new rounds (see median values in Table 2).

3.2. Sound exposure level, L_{AE} and L_{CE}

A - and C -weighted sound exposure levels (L_{AE} and L_{CE}), which provided the noise level normalised to a one-second period, were calculated for each round. For transient noise, such as that associated with weaponry, sound exposure level is a more convenient and preferred measure compared with the equivalent continuous sound pressure level (L_{Aeq}) as it effectively shows the total noise energy per round fired. Table 2 shows the L_{AE} and L_{CE} values for the two types of rounds and for the two operators. Sound exposure levels (L_{AE}) measured for the different combinations of round and location are shown in Fig. 4. The differences in the L_{AE} values are quite clear: the current rounds show higher values compared with the new rounds (difference between median values of 9.5 dB for the firer and 9.8 dB for the loader), and the firer was exposed to higher values compared with the loader (difference between median values of 4.7 dB for the current rounds and 5.0 dB for the new rounds).

The frequency spectra of the different combinations of rounds and measurement locations for the firings from the saluting gun are shown in Fig. 5. The data show the mean and range (minimum and maximum) sound exposure levels (L_{ZE}) corresponding to the 12 current rounds and 24 new rounds fired from the gun.

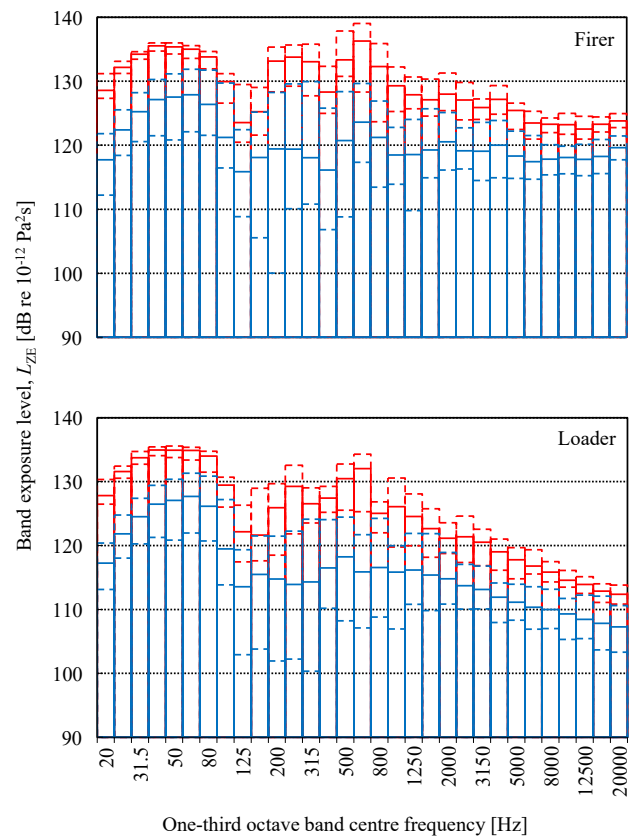


Fig. 5. One-third octave band sound exposure levels L_{ZE} (mean (solid line) and range (dashed line) corresponding to 12 current and 24 new rounds) measured for the current (red line) and new (blue line) rounds.

Without exception, the mean sound exposure levels at the one-third octave frequency bands for current rounds were higher than the new rounds fired from the gun. When averaged over the frequency bands from 20 Hz to 20000 Hz, the current rounds produced sound exposure levels (L_{ZE}) approximately 8.7 dB higher for

Table 3. Median values for the maximum fast sound pressure levels and sound exposure levels measured during firing of rounds from the saluting gun.

Location, round	$L_{AF\max}$ [dB]	$L_{CF\max}$ [dB]	L_{AE} [dB]	L_{CE} [dB]	$L_{CF\max} - L_{AF\max}$ [dB]	$L_{CE} - L_{AE}$ [dB]
Firer, current	149.8	153.7	141.3	145.4	4.0	4.2
Firer, new	140.3	144.5	131.8	136.1	4.1	4.1
Loader, current	145.0	151.2	136.6	142.8	6.2	6.3
Loader, new	135.2	142.8	126.8	134.3	7.0	7.0

the firer and 8.6 dB higher for the loader than the new rounds, respectively. Most of the sound energy was present in frequencies below about 100 Hz.

3.3. Maximum sound levels, $L_{AF\max}$ and $L_{CF\max}$

Various parameters about the sound levels measured during firing of different types of round and location are shown in Table 2. The parameters $L_{C\text{peak}}$, L_{AE} , and L_{CE} do not give an indication of the frequency content present within the sound. These either show a single value within the signal ($L_{C\text{peak}}$) or the total energy present in the signal (L_{AE} and L_{CE}). The difference between the A - and C -weighted maximum sound levels, $L_{AF\max}$ and $L_{CF\max}$, can, however, give a broad indication of the frequency content of the signal. (Although, the spectra shown in Fig. 5 would give a better measure of the differences at the various frequencies.) The differences between the A - and C -weighted levels for both sound exposure levels (L_E) and maximum sound levels ($L_{F\max}$) were therefore calculated for each of the current and new rounds fired from the cannon; median values are shown in Table 3. Table 3 also shows the median differences between A - and C -weighted levels for the two measurement locations (that is, the firer and the loader). Various findings become evident from the data. The differences between C - and A -weighted values are lower for the firer's location than the loader's position; this being the case for both the current and the new rounds. This would indicate that the sound measured at the firer's location contained a higher proportion of high-frequency energy compared with the loader's location; this is confirmed by the spectra shown in Fig. 5. There appears to be only a small difference between the maximum sound levels ($L_{CF\max} - L_{AF\max}$) and the exposure levels ($L_{CE} - L_{AE}$) when comparing data for the current rounds (0.2 dB for the firer and 0.1 dB for the loader). Overall, it is seen that there are only very marginal differences when using either the sound exposure levels or the maximum sound levels.

4. Discussion

The noise measurements for the saluting gun being used with the two types of rounds show that, at the firer's location, the C -weighted peak sound pressure level, $L_{C\text{peak}}$, for the new type of round (162.9 dB) was

about 8.8 dB lower than the current round (171.7 dB). The difference was higher (9.8 dB) for the loader's location. The corresponding differences between sound exposure levels, L_{AE} , were 9.5 dB for the firer and 9.8 dB for the loader's position. This represents a significant decrease in the exposure of the two operators during firing of the gun with the new rounds compared with the current rounds. It could be argued that one of the drawbacks of the lower noise levels is that people in the local vicinity are not able to hear the noises made in a celebratory nature. This consequence, however, should be considered as acceptable and that the health of the operators would (should) outweigh this concern.

Table 2 shows that the variation (interquartile range) in the different sound parameters was greater for the new rounds compared with the current rounds. For example, for measurements made at the loader's position, parameters L_{CE} and $L_{CF\max}$ show that there was 10 times as much variation between rounds for the new ammunition (2.0 dB) compared with the current ammunition (0.2 dB). This large variation for the new rounds is also seen in the unweighted sound exposure levels L_{ZE} calculated in the one-third octave bands as shown in Fig. 5. (It is noted that twice as many of the new type of rounds (24 rounds) were fired compared with the current type (12 rounds).) It is assumed that this difference in noise between successive rounds would be no cause for concern due to the purely commemorative nature of the activity. The source of the variation is, however, not known. One of the factors that might influence this variation could be the specification, or tolerance, used in the manufacture of the rounds. If that were to be the case, then this might imply that the new rounds were manufactured to a lower standard compared with the current rounds. However, such inferences might cause differences of opinion with the manufacturers of such items. A similar discussion has been presented regarding the slightly higher variation (though inconclusive) in sound pressure levels measured when firing blank rounds from three military rifles compared with live rounds (PADDAN, HOWELL, 2022). It was opined that the blank rounds, which are used for educational and training purposes, might be manufactured to less stringent standards compared with the tighter tolerances used when producing live rounds. All this discussion is based on speculation; this topic, however, requires evidence.

Two frequency weightings, A and C , appropriate for exposure to noise were used in the assessment of sounds produced by the rounds fired from the saluting gun (further details about the frequency weightings for sound level meters are shown in (British Standard Institution, 2013)). The C -weighting filter is more appropriate for assessing peak sound pressure levels normally found in impulse or impact signals. Long-term damage to hearing from moderate to loud noise is related to the noise exposure calculated using the A -weighting. It is noted that the A -weighting amplifies the sound pressure levels over the frequency range 1000 Hz to 6000 Hz. Sound exposure levels (L_{ZE}) for one-third octave frequency bands are shown in Fig. 5 for the two types of rounds. Most of the energy in the impulse signals for the new rounds occurred at frequencies below about 100 Hz.

The suitability of hearing protection that can be used to protect the operators from the impulse noise can be determined using the frequency data presented in Fig. 5. Further details about the process involved are given in (Health and Safety Executive; 2021). A method specifically for determining suitable hearing protection for use with military weapons is given and mandated in (Ministry of Defence, 2015). The method is based on using the difference in the maximum sound levels; this is calculated as $L_{CF\max} - L_{AF\max}$. This is used to give a broad indication of the frequencies present in the noise signal. It is seen from Table 3 that the same information can be gleaned from the sound exposure levels: L_{AE} and L_{CE} . The difference between L_{CE} and L_{AE} can provide the same information as the maximum sound levels (that is, $L_{CF\max} - L_{AF\max}$). The development of Defence Standard 00-027 (Ministry of Defence, 2015) is discussed elsewhere (PADDAN, HOWELL, 2025).

The suitability of the hearing protection worn by the firer and the loader can be made using the attenuation properties of the hearing protection combination of earplugs (E-A-R soft FX) and earmuffs (Peltor Comtac XPI), and details about the noise produced. The attenuation properties of this combination of hearing protection were 42 dB for high (H) frequencies, 44 dB for medium (M) frequencies, and 42 dB for low (L) frequencies; the SNR (single number rating) was 45 dB (INSPEC, 2017). For the assessment, the

highest $L_{C\text{peak}}$ of 173.1 dB can be used corresponding to the firer's noise exposure while using the current ammunition (see Table 2). The difference between the two parameters, $L_{CF\max}$ and $L_{AF\max}$, is required to determine the suitability of hearing protection as specified in (Ministry of Defence, 2015); these are shown in Table 3. The guidance (Ministry of Defence, 2015) states that this difference would then be used to calculate a 'modified sound attenuation value, d_m ' as shown in Table 4. Table 3 shows this difference to be between 4.0 and 7.0 depending on the operator (firer or loader) and type of ammunition (current or new). The effective sound pressure level at the ear, $L'_{C\text{peak}}$, is calculated:

$$L'_{C\text{peak}} = L_{C\text{peak}} - d_m. \quad (1)$$

Table 4. Modified sound attenuation values for different impulse or impact noises (adapted from Ministry of Defence, 2015); H = high, M = medium, L = low.

$L_{CF\max} - L_{AF\max}$ [dB]	Modified sound attenuation value, d_m [dB]
≤ 0	H
>0 to 1	M
>1 to 3	M - 5
>3 to 5	L or M - 5 if a lower value
>5 to 10	L - 5
>10	Conditional use of L - 5

Table 5 shows the effective sound pressure level at the ear, $L'_{C\text{peak}}$, based on the attenuation properties of the E-A-R soft FX earplugs worn in combination with the Peltor Comtac XPI earmuffs. The C -weighted peak sound pressure at the ear would be below the exposure limit value, corresponding to 140 dB, specified in (Statutory Instruments, 2005; Directive 2003/10/EC, 2003).

The sound pressure levels measured during firing of the saluting gun with the two types of rounds can be compared and assessed using the guidance in (Statutory Instruments, 2005; Directive 2003/10/EC, 2003). These documents specify the following:

- 1) the lower exposure action values (LEAV) are:
 - a daily or weekly personal noise exposure of 80 dB (A -weighted);
 - a peak sound pressure of 135 dB (C -weighted);

Table 5. Effective sound pressure level at the ear for the loader and firer when using current and new ammunition. Hearing protector (E-A-R soft FX earplugs worn in combination with Peltor Comtac XPI earmuffs) – H = 42 dB, M = 44 dB, L = 42 dB.

Location, round	$L_{CF\max} - L_{AF\max}$ [dB]	Modified sound attenuation value, d_m [dB]	Effective sound pressure level at the ear, $L'_{C\text{peak}}$ [dB]
Firer, current	4.0	M - 5 = 39	134.1
Firer, new	4.1	M - 5 = 39	134.1
Loader, current	6.2	L - 5 = 37	136.1
Loader, new	7.0	L - 5 = 37	136.1

- 2) the upper exposure action values (UEAV) are:
- a daily or weekly personal noise exposure of 85 dB (*A*-weighted);
 - a peak sound pressure of 137 dB (*C*-weighted);
- 3) the exposure limit values (ELV) are:
- a daily or weekly personal noise exposure of 87 dB (*A*-weighted);
 - a peak sound pressure of 140 dB (*C*-weighted).

It is noted that the ‘daily personal noise exposure’ is standardised to an 8-hour period and the ‘weekly personal noise exposure’ is standardised to five 8-hour working periods (40 h per week). Data in Table 2 show peak sound pressure levels, while data in Table 5 correspond to the peak sound pressure levels *after* taking into account the attenuation provided by the hearing protection. These data have been compared with the peak sound pressure level (peak sound pressure level at the ear $L'_{C_{peak}} = 140$ dB) as shown earlier. An assessment with respect to the ‘daily personal noise exposure’ would require a measurement of L_{Aeq} over an 8-hour period. The sound exposure level, L_{AE} , which is a measurement standardised for a 1-second period, can be used. This comparison can also be carried out by calculating the exposure levels in terms of sound exposure values: the LEAV would be L_{AE} 124.6 dB, UEAV would be L_{AE} 129.6 dB, and the ELV would be L_{AE} 131.6 dB. It is seen from Table 2 that, apart from the loader’s position during firing of the new rounds, the median sound exposure levels for the other combinations (operator and type of round) exceeded the UEAV (L_{AE} 129.6 dB). However, it must be emphasised that an assessment with respect to the ‘daily personal noise exposure’ can only be carried out if the peak sound pressure level has not been exceeded.

It is clear from the data that the new rounds produce lower sound pressure levels compared with the current rounds, and that the firer is exposed to higher levels compared with the loader. Further mitigation measures should be considered in reducing noise exposure of the two operators. Some of the measures proposed might not be practicable or feasible but should nonetheless be considered. Also, some of the suggestions might be deemed as being controversial or unpalatable to the operators (or those in command), but these difficult questions should be addressed. There would need to be a fine balance between ceremony and safety. Some of the measures could include the following:

- In theory, sound pressure levels decrease by 6 dB with doubling of distance from the noise source. That is, increasing the distance between the gun and the operators would result in a significant decrease in exposure. One such measure could be to

automate the firing of the gun or to operate the firing mechanism from a distance.

- The firer is exposed to higher sound pressure levels compared with the loader. Rotating (sharing) the jobs carried out by the two operators could be considered such that the firer and the loader alternate their duties between successive rounds. This would ensure that no single operator is continuously exposed to high noise levels. It is noted that this suggestion might impinge on the firing rate (the period between) of successive rounds. The standard period between successive salutes is nominally 10 s.
- The frequency spectra for the two types of rounds show a dominance in sound pressure levels for frequencies below about 100 Hz. Hearing protectors are generally less effective at attenuating low frequencies compared with high frequencies. If possible (and available), careful selection of hearing protection could involve showing preference to those protectors likely to offer greater protection over low frequencies.
- A barrier or screen could be placed between the muzzle of the gun and the two operators. This could take the form of a solid plate attached to the gun thus creating an ‘acoustic shadow’ around the operators. Depending on the design of such a screen or enclosure, this might be expected to reduce the sound pressure level by a few decibels. The barrier would alter the noise received by the operators and might influence the choice of hearing protection worn by the operators.
- The number of rounds, or salutes, is dictated by ceremony and tradition. Maybe fewer salutes could be conducted to mark ‘regular’ occasions. This would not discourage increasing the number of salutes, as necessary, to mark extraordinary events; one such event being the 96 salutes to mark the passing of Her Majesty the Queen Elizabeth II.

5. Conclusions

Noise measurements were made from the firing of two types of rounds from a saluting gun: the current round and a new round. Measurements were made at the firer’s and the loader’s positions. Twelve rounds of the current ammunition and 24 rounds of the new ammunition were fired during the assessment. The highest *C*-weighted peak sound pressure levels for the firer and the loader were 173.1 dB and 170.6 dB, respectively for the current round and, 166.8 dB and 163.0 dB, respectively, for the new round. Lower peak sound pressure levels were measured when new rounds were fired compared with the current rounds: the difference in median peak sound pressure levels were

8.8 dB and 9.8 dB for the firer and the loader, respectively. Sound exposure levels, a parameter which is normalised to a period of 1 s, again were higher for current rounds (*A*-weighted median values of 141.3 dB and 136.6 dB for the firer and loader, respectively) compared with new rounds (*A*-weighted median values of 131.8 dB and 126.8 dB for the firer and loader, respectively).

ACKNOWLEDGMENTS

The authors would like to thank Mr. Z. Francis-Cox, formerly of the Institute of Naval Medicine, for assisting with the current study.

DISCLAIMER

The contents of the work, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect the policy of UK Ministry of Defence.

References

1. British Standard Institution (2013), *Electroacoustics. Sound level meters – Specifications* (BS EN Standard No. 61672-1:2013), <https://knowledge.bsigroup.com/products/electroacoustics-sound-level-meters-specifications-1>.
2. COLES R.R.A., GARINTHER G.R., HODGE D.C., RICE C.G. (1968), Hazardous exposure to impulse noise, *The Journal of the Acoustical Society of America*, **43**(2): 336–343, <https://doi.org/10.1121/1.1910785>.
3. Department of Defense (1997), *Noise limits* (Standard No. MIL-STD-1474D).
4. Directive 2003/10/EC (2003), Directive 2003/10/EC of The European Parliament and of the Council of 6 February 2003 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise), <https://www.legislation.gov.uk/eudr/2003/10/introduction> (access: 27.05.2024).
5. FLAMME G., MURPHY W.J., KARDOUS C.A., BYRNE D.C. (2019), Impulsive exposures from ceremonial and signal cannons, *The Journal of Acoustical Society of America*, **145**(3): 1815, <https://doi.org/10.1121/1.5101634>.
6. Health and Safety Executive (2021), *Controlling noise at work. The control of noise at work regulations 2005 – Guidance on regulations*, 3rd ed., TSO, <https://www.hse.gov.uk/pubns/books/1108.htm>.
7. INSPEC (2017), *Test report* (Report No. 1.17.03.83), INSPEC International Limited.
8. Ministry of Defence (2015), *The measurement of impulse noise from military weapons, explosives and pyrotechnics; and selection of hearing protection* (Defence Standard 00-027).
9. PADDAN G.S., HOWELL M.J. (2019), *Assessment of noise exposure during firing of saluting guns at HMNB Portsmouth*, INM Report 2019.027, Institute of Naval Medicine, Gosport.
10. PADDAN G.S., HOWELL M.J. (2022), A comparison of noise produced by blank and live rounds fired from military rifles, *Applied Acoustics*, **189**: 108599, <https://doi.org/10.1016/j.apacoust.2021.108599>.
11. PADDAN G.S., HOWELL M.J. (2025), Measurement and analysis of noise produced by ‘flash-bang’ hand grenade distraction devices, *Applied Acoustics*, **237**: 110755, <https://doi.org/10.1016/j.apacoust.2025.110755>.
12. SHERIDAN D. (2024), Edinburgh Castle gun salute to be made quieter after ‘health and safety killjoys’ complain, *The Telegraph*, <https://www.telegraph.co.uk/news/2024/04/01/edinburgh-castle-gun-made-quieter-amid-health-safety-rules/> (access: 27.05.2024).
13. Statutory Instruments (2005), Statutory Instruments 2005 No. 1643. Health and safety. The control of noise at work regulations 2005, <https://www.legislation.gov.uk/uksi/2005/1643/contents> (access: 27.05.2024).
14. *Time gun-maps of 1861* (n.d.), Archive.org, https://web.archive.org/web/20090212115320/http://www.edinphoto.org.uk/0_maps_2/0_map-edinburgh.time-gun-1861-_map_notes.htm (access: 27.05.2024).

Research Paper

The Impact of HVAC Systems on Speech Intelligibility in University Classrooms

Akın OKTAV^{(1),(2)*}, Arda ENER⁽²⁾, Faruk ÖZENÇ⁽²⁾, Sila SARI⁽²⁾

⁽¹⁾ *Department of Mechanical Engineering, Alanya Alaaddin Keykubat University
Antalya, Türkiye*

⁽²⁾ *Vibration and Acoustics Laboratory, Alanya Alaaddin Keykubat University
Antalya, Türkiye*

*Corresponding Author e-mail: akin.oktav@alanya.edu.tr

*Received November 15, 2024; revised February 3, 2025; accepted April 4, 2025;
published online May 16, 2025.*

Good speech intelligibility in university classrooms is crucial to the learning process, ensuring that students can clearly hear all conversations taking place in the classroom. While it is well known that speech intelligibility depends on the geometrical characteristics of a space and the properties of its surfaces, other factors need also to be considered. Among the most important are: the heating, ventilation, and air conditioning (HVAC) systems used in classrooms. Fan noise from HVAC systems increases the background noise level (BNL), negatively affecting speech intelligibility. In addition, the movement of air caused by these systems alters room acoustic variables. Although this dynamic situation is often overlooked in the early design stages, HVAC systems are often active during lectures and influence acoustics variables, especially the speech transmission index (STI). In this study, the impact of HVAC systems on the STI was measured in five different unoccupied classrooms in the Rafet Kayış Faculty of Engineering at Alanya Alaaddin Keykubat University. The results were evaluated according to relevant standards. The results of these evaluations offer insights for researchers, architects, and engineers working in the field of acoustics.

Keywords: speech intelligibility; speech transmission index (STI); room acoustic variables; room impulse response (RIR); acoustic performance.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The act of learning is a process influenced by and related to a number of factors, including the environment, infrastructure, the student, and the instructor. The acoustic performance of a space is an important factor that affects students' learning outcomes. Educational institutions should provide well-designed and appropriate spaces in order to improve the quality of education. Speech intelligibility plays an important role in educational settings by directly affecting the quality of communication between students and instructors. Several studies suggest that poor room acoustic performance have a negative impact on speech intelligibility and affects verbal communication be-

tween students and instructors (YANG, MAK, 2018; CHOI, 2020; ENGEL *et al.*, 2020; KAWATA *et al.*, 2023; DI LORETO *et al.*, 2023).

For effective communication, it is not enough to simply hear what the instructor says, as hearing and understanding what is said without loss is an important component of communication. In the field of education, an area where communication is actively used, the quality of communication (information exchange) between instructors and students is closely related to speech intelligibility. Acoustic conditions have been shown to directly affect students' ability to understand speech, often leading to inefficient communication in the classroom (RABELO *et al.*, 2014). When speech intelligibility is inadequate, instructors

may have to raise their voice, students may have problems to maintain focus, and key information may not be conveyed accurately. This situation negatively affects communication between instructors and students and undermines the realization of the learning outcomes intended for the course.

In the courses offered at Alanya Alaaddin Keykubat University (ALKU) Rafet Kayış Faculty of Engineering, lecturers support the educational process by using various visual communication tools, such as slides, graphics, and videos, to teach course material more effectively. While visual aids are commonly used, the primary communication remains verbal. The language of instruction in most departments at this faculty is English. For students whose first language is not English, challenges related to pronunciation, vocabulary, and grammatical structure can hinder proper understanding. These linguistic barriers make it even more difficult for non-native speakers to follow the courses. Non-native speakers require a 4 dB to 5 dB improvement in the signal-to-noise ratio (SNR) to achieve equivalent level of speech intelligibility as native speakers (International Organization for Standardization [ISO], 2003).

Speech intelligibility is related to several objective acoustic metrics such as reverberation time (RT), background noise level (BNL), useful energy (first 50 ms), the early-to-total energy ratio (D50), and the SNR. Some studies have shown that RT has a significant effect on speech transmission index (STI). PAYTON and SHRESTHA (2008) showed that STI measures the extent to which speech envelope modulations are preserved in degraded listening environments. Recently, Chinese speech intelligibility scores has been examined in university classrooms using a hybrid method (HUANG *et al.*, 2023). The results show that to achieve better speech intelligibility, RT at all frequencies should be shorter, and it is better when an RT is flat at low frequencies. The STI is an objective metric that correlates well with the intelligibility of speech, a subjective metric, which is degraded by additive noise and reverberation.

While it is common to categorize students into age groups, typically under 12 and over 12 years old, there are additional factors to be considered in university classrooms (MINELLI *et al.*, 2022). In many universities around the world, students are taught in a second language, different from their mother tongue in education settings. It is important to take this language barrier into account when assessing speech intelligibility. Moreover, there is not enough information on how the speech intelligibility parameter is affected by HVAC systems, particularly through their impact on BNL and SNR. It has been shown that these systems, which are actively used in the educational process, negatively affect SNR ratio, which is one of the acoustic parameters, as well as speech intelligibil-

ity (DI LORETO *et al.*, 2023; ZHU *et al.*, 2024). Considering the possible effects of various factors, studies on speech intelligibility in university classrooms remain relatively scarce.

In this study, acoustic measurements were conducted in five different spaces, both with HVAC systems inactive and operating at different fan speed levels. From the raw acoustic data measured, parameters, including: RT, center time (Ts), D50, STI, strength (G), and SNR, were obtained across seven octave bands (125 Hz–8000 Hz). The results presented are compared for the five different spaces. The acoustic performance of each space is evaluated according to relevant standards. The results of these evaluations are then utilized to draw conclusions of interest to researchers, architects and engineers working in the field of acoustics.

2. Methods and data

2.1. Descriptors of room acoustics for speech intelligibility

The RT, specifically T30, is defined according to Schroeder curves (SC) obtained from the impulse response ($g(t)$) (ROSSING *et al.*, 2014):

$$T_{30} = 2[t(\text{SC} = -35 \text{ dB}) - t(\text{SC} = -5 \text{ dB})], \quad (1)$$

where

$$\text{SC} = 10 \log_{10} \left(\frac{\int_0^t g^2(\tau) d\tau}{\int_0^\infty g^2(\tau) d\tau} \right). \quad (2)$$

Ts is the ratio of early energy to late energy, defined as

$$T_s = \frac{\int_0^\infty \tau g^2(\tau) d\tau}{\int_0^\infty g^2(\tau) d\tau}. \quad (3)$$

D50 is the ratio of useful energy (the first 50 ms) to total energy, and it is expressed as

$$D_{50} = \frac{\int_0^{50\text{ms}} g^2(\tau) d\tau}{\int_0^\infty g^2(\tau) d\tau} \leq 1. \quad (4)$$

G is the energy gain in a reverberant room compared to a free field with a 10 m distance, where theoretically no reverberation occurs, and it is defined as

$$G = 10 \log_{10} \left(\frac{\int_0^\infty g^2(\tau) d\tau}{\int_0^{t_{\text{dir}}} g_{10\text{m}}^2(\tau) d\tau} \right), \quad (5)$$

where t_{dir} refers to the direct sound. SNR is

$$SNR(\omega) = 10 \log_{10} \left(\frac{m(\omega)}{1 - m(\omega)} \right), \quad (6)$$

where $m(\omega)$, the complex modulation transfer function, is given by International Electrotechnical Commission [IEC] (2020):

$$m(\omega) = \frac{\int_0^{\infty} g^2(\tau) e^{-j\omega\tau} d\tau}{\int_0^{\infty} g^2(\tau) d\tau}. \quad (7)$$

STI is given by MEJDI et al. (2019):

$$STI = \min \left(1.0, \sum_{k=q}^7 \alpha_k MTI_k - \sum_{k=q}^6 \beta_k \sqrt{MTI_k MTI_{k+1}} \right), \quad (8)$$

where MTI is the modulation transfer index; 6 and 7 are octave bands, α and β are weighting and redundancy factors, respectively; $q = 1$ for male speakers and $q = 2$ for female speakers, which correspond to 125 Hz and 250 Hz, respectively.

2.2. Spaces

Acoustic measurements were conducted in five educational spaces, each with different dimensions, at ALKU Rafet Kayış Engineering Faculty. Three of these spaces – classrooms A203, D107, and T206 – are used for theoretical courses. The other two spaces where measurements were conducted serve as laboratories: A208 is a computer laboratory and D110 is the Vibration and Acoustics Laboratory. Due to concerns about variations in background noise and for operational reasons, acoustic data were collected during the summer, when there were no students at the university.

The selected spaces represent different types of educational rooms within the faculty. The dimensions of each space are shown on the scaled plans shown in Figs. 1–5. The number of seats in each space is as follows: 99 seats in A203, 64 seats in D107, 60 seats in T206, 58 seats in A208, and space and seats for

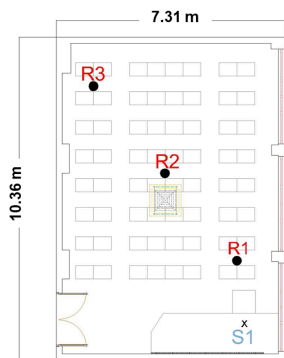


Fig. 1. Scaled plan of classroom D107.

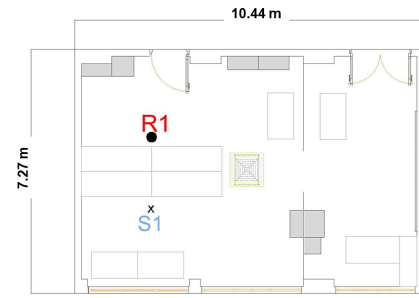


Fig. 2. Scaled plan of Vibration and Acoustics Laboratory, D110.

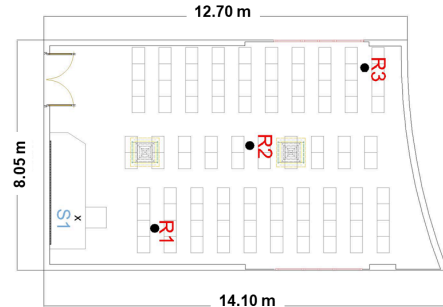


Fig. 3. Scaled plan of classroom A203.

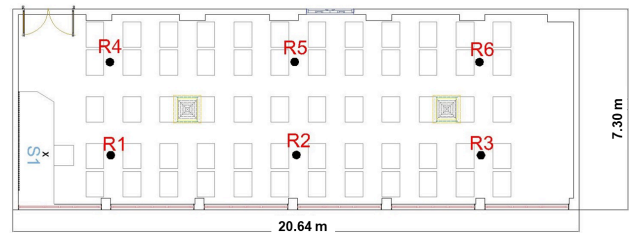


Fig. 4. Scaled plan of classroom T206.

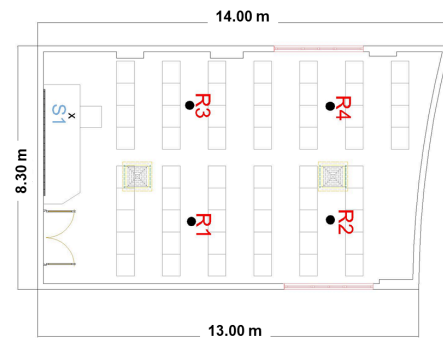


Fig. 5. Scaled plan of classroom A208.

8 researchers in D110. Each space, except D110, is equipped with a lectern located 2 meters in front of the whiteboard. The HVAC system employed in these spaces is a Daikin FXFQ125 round flow cassette model. There are two units installed in A203, A208, and T206, while one unit is located in D107 and D110. The locations of the HVAC systems are shown in Figs. 1–5, as well. The HVAC system can operate at three different fan levels (L, H, HH), with corresponding fan flow rates of 33.0 m³/min, 26.5 m³/min, and 19.9 m³/min, respectively.

The materials used in the design of the spaces are as follows. Stoneware tiles were selected for the floor covering. The side walls are constructed with brick, finished with basic plaster and thin plaster layers. The windows have double-glazed window systems, while solid wood doors are used throughout. To improve acoustics, micro-perforated acoustic panels were installed on the ceiling. The tables and chairs are made of wood, and ceramic-enameled whiteboards are used as writing surfaces. In addition, roll curtains made of polyester material were installed to cover the windows.

2.3. Measurements

The indirect method used in this study to determine acoustic parameters consists of measuring the response of an enclosed space to an impulse signal. The Sinus Qohm QS12 sound source is suitable for measurements between 50 Hz–16 000 Hz with a level of 122 dB across a uniform broadband spectrum. The required omnidirectionality for the measurements is in compliance with the relevant standards (ISO, 2014; 2021). In addition, the sound source meets directivity values as it meets the maximum permissible deviation values in the octave bands of pink noise specified by the relevant standard (ISO, 2009). During the measurements, the height of the sound source was set to 1.5 m from the ground.

Exponential sine sweeps (ESS) are used for impulse stimulation in the measurements due to their ability to separate harmonic distortion and yield higher impulse-to-noise ratios under typical test conditions (MENG *et al.*, 2008; GUIDORZI *et al.*, 2015; ANTONIADOU *et al.*, 2018). A Focusrite Scarlett 18i20 external sound card is used as the audio interface to transmit the sound source and microphone signals to the computer. The response of the space is recorded using omnidirectional GRAS 46AE microphones. The software used to record the raw audio signals and process the data is Dirac v7. Prior to measurements, the microphones are calibrated.

The BNL values of the classrooms were determined prior to the measurement survey. According to the relevant standard (American National Standard, 2010), the BNL values in classrooms are expected to be lower than 35 dBA. This standard value applies to unoccupied classrooms and includes environmental noise

and HVAC-related noise. The measured BNL values (L_{A90}) for classrooms D107, D110, A203, T206, and A208 are presented in Table 1. Upon analyzing the values in the table, it is evident that in most cases, the BNL values exceed the recommended standard value. These values are shown in bold in the table. The measurements were conducted during the summer period when there were no students on campus and environmental noise levels were minimal. It is clear that the most important contributor to the elevated BNL values is the HVAC system. As the fan speed increases, the noise level also increases, which leads to higher BNL values. As will be explained in the next section, HVAC noise also has a significant impact on the SNR. HVAC noise reduces the SNR, which, in turn, has a negative impact on speech intelligibility.

Classroom D110 is an actively used research laboratory. There are two uninterruptible power supplies (UPS) running 24 hours a day in this space. As shown in the values in Table 1, these devices increase the background noise and have a negative impact on speech intelligibility.

The sound source, representing the instructor, is located behind the lectern. A distance of at least 1 m was maintained between the sound source and the side walls. Microphones, representing the students, were positioned 1.2 m above the ground and at least 1 m away from the walls in accordance with ISO (2009). The distance between the sound source and the microphone is an important variable for speech intelligibility. Therefore, microphones were positioned at different distances from the sound source. The locations of microphones and the sound source are shown in Figs. 1–5 for the studied spaces. During the measurement survey, the temperature and relative humidity were continuously monitored and recorded.

The spaces were stimulated with a 21.8 s ESS signal, in the frequency range 20 Hz–20 000 Hz. During the measurements, the polyester roll curtains, windows and doors were closed. The measurements were repeated with the HVAC system off and operating at three different fan levels, as indicated in Table 1. To account for measurement uncertainty, all measurements were repeated three times, and the average of the processed values was considered for analysis. The photographs taken during the measurement survey are shown in Figs. 6–10.

Table 1. Measured BNLs in dBA.

Space	No fan 0 m ³ /min	Fan level I (L), 19.9 m ³ /min	Fan level II (H), 26.5 m ³ /min	Fan level III (HH), 33.0 m ³ /min
D107	29.0	34.7	41.3	47.2
D110	40.2	41.2	43.3	47.4
A203	29.7	38.5	44.0	49.5
T206	31.0	36.5	42.7	47.9
A208	33.6	37.3	43.0	48.2



Fig. 6. Experimental study conducted in classroom D107, microphone position: R2, temperature: 26.5 °C.



Fig. 7. Experimental study conducted in laboratory D110, microphone position: R1, temperature: 25.0 °C.



Fig. 8. Experimental study conducted in classroom A203, microphone position: R3, temperature: 26.5 °C.



Fig. 9. Experimental study conducted in classroom T206, microphone position: R1, temperature: 26.5 °C.



Fig. 10. Experimental study conducted in laboratory A208, microphone position: R2, temperature: 26.5 °C.

3. Results and discussion

The intelligibility of speech in an enclosed space depends on the BNL, the distance between the speaker and the listener, the directivity of the speech, the signal strength of the speech, the sound spectrum of the speech, and reverberation characteristics of the space. While the audio signal of speech spans a wide range of frequencies across 7-octave bands, the 500 Hz–4000 Hz range is critical for speech intelligibility. According to the relevant standard (IEC, 2020), STI is calculated as the weighted sum of the MTI, one for each octave frequency band in the 7-octave band; each MTI value is obtained from modulation transfer function (MTF) values over 14 different modulation frequencies (ELLIOTT, THEUNISSEN, 2009).

The SNR values in 7-octave bands are shown in Fig. 11 for the various spaces. BNL can be neglected if the SNR exceeds 15 dB in each octave frequency band of interest (in this case, the 7-octave band). However, the strength of the sound source may need to be increased for this to occur. In practice, this means that the instructor needs to raise their voice. Average vocal effort levels are usually measured in anechoic chambers for classification (CUSHING *et al.*, 2011). Average vocal effort levels in anechoic conditions, measured at 1 m, are presented in Table 2. During the measurement survey, the generated sound was adjusted to be at least 15 dB above the BNL.

Table 2. Average vocal effort and sound level in dBA.

	Normal	Raised	Loud	Shouting
Male	58	67	76	89
Female	56	64	70	82

The SNR results clearly show the negative impact of HVAC systems: as the fan speed increases the SNR values decrease, which, in turn, affects speech intelligibility. The variation in SNR is sensitive to frequency, and although the trends are similar, the geometry of

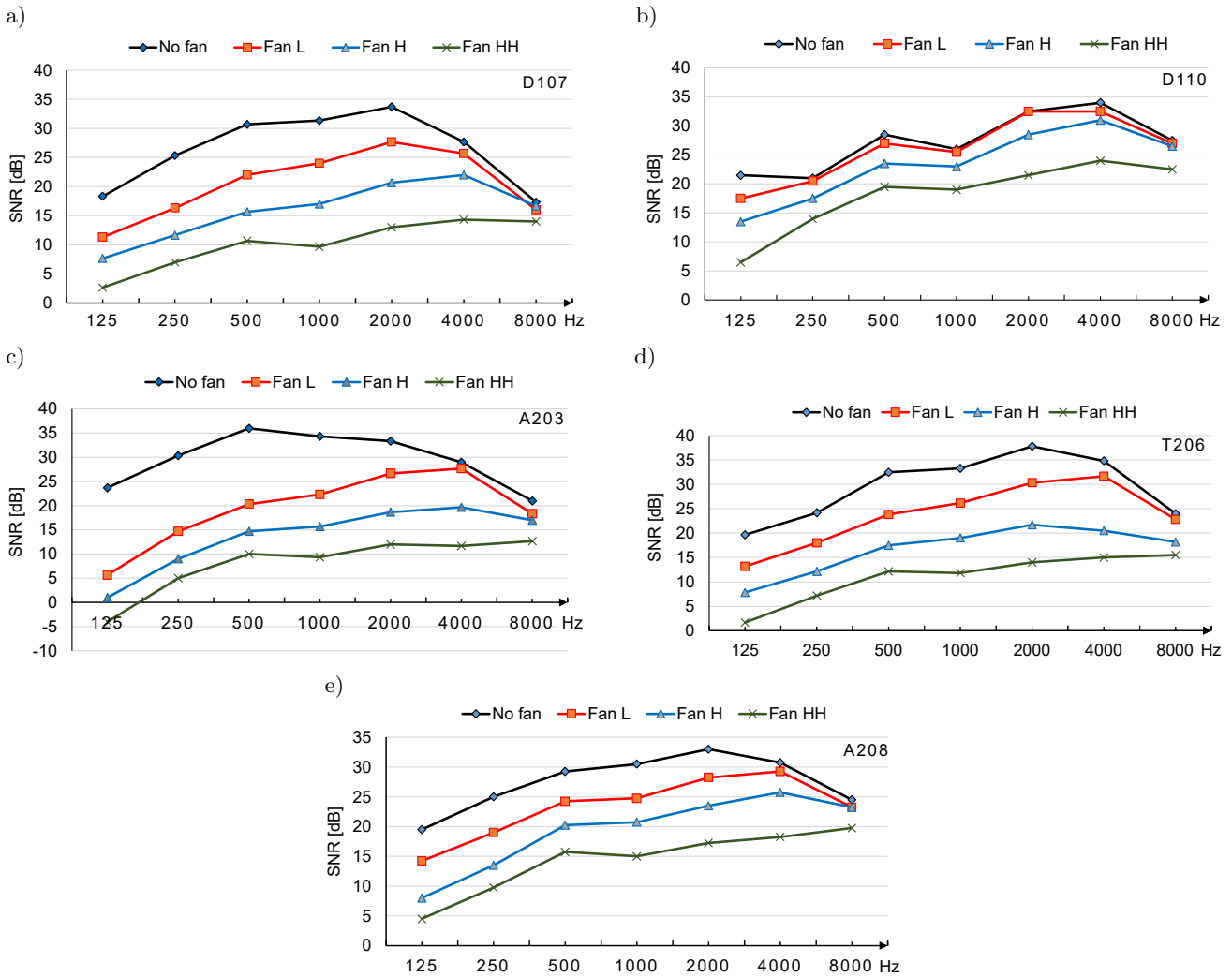


Fig. 11. Variation of SNR in the spaces: a) D107; b) D110; c) A203; d) T206; e) A208.

the space also affects the changes. One way to compensate for the drop in SNR values is for the instructor to raise their voice. The values presented in Table 2 give an idea of the vocal effort required. SNR values in the spaces tend to decrease after 4000 Hz. However, the frequency range 500 Hz–4000 Hz is decisive for speech intelligibility and, within this range, the 1000 Hz and 2000 Hz bands are critical.

The acoustic parameters measured in the spaces are presented in Tables 3–7. The acoustic parameters and their units are as follows: SNR [dB], T30 [s], G [dB], T_s [ms], D50 (unitless [0–1]), and STI (unitless [0–1]).

Speech intelligibility depends on the speaker’s voice reaching the listener directly, as well as the effects of reverberation and background noise. Reverberation and background noise have a distorting effect on the sound that reaches the listener directly. In terms of objective measures, reverberation can be quantified by T30 and background noise by BNL. Since SNR is the ratio of speech to BNL, it is a key factor in determining intelligibility. The focus of the presented work is

primarily on the impact of HVAC systems on speech intelligibility, which can be related to SNR. The average SNR values at 500 Hz and 1000 Hz SNR measured in the spaces, plotted against the blowing flow rate of the HVAC used are shown in Fig. 12. The results show that SNR values tend to decrease as the fan blowing speed of the HVAC systems increases, which, in turn, negatively affects speech intelligibility.

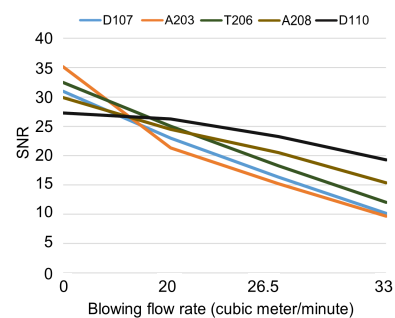


Fig. 12. Variation of SNR in the spaces depending on the impact of HVAC systems.

Table 3. Acoustic parameters measured at D107 for four different fan settings.

Fan	Parameter	Frequency							
			125	250	500	1000	2000	4000	8000
No fan	SNR		18.33	25.33	30.66	31.33	33.66	27.66	17.33
	T30		1.56	1.13	0.94	0.83	0.86	0.81	0.65
	<i>G</i>		23.28	21.16	20.27	19.96	20.26	19.78	18.68
	Ts		107.47	68.57	56.13	50.47	50.53	52.13	36.70
	D50		0.46	0.58	0.64	0.65	0.64	0.61	0.74
	STI	0.66							
Fan L	SNR		11.33	16.33	22.00	24.00	27.67	25.67	16.00
	T30		1.43	1.13	0.87	0.82	0.86	0.80	0.64
	<i>G</i>		23.12	21.17	20.14	20.15	20.24	19.78	18.70
	Ts		107.83	69.50	54.47	50.13	51.23	52.87	37.23
	D50		0.46	0.56	0.67	0.65	0.63	0.60	0.73
	STI	0.65							
Fan H	SNR		7.67	11.67	15.67	17.00	20.67	22.00	16.67
	T30		1.45	1.10	0.87	0.83	0.86	0.80	0.65
	<i>G</i>		22.36	20.01	19.09	19.18	19.12	18.26	17.85
	Ts		105.60	68.90	54.00	51.03	50.40	52.87	36.83
	D50		0.47	0.56	0.67	0.64	0.64	0.62	0.74
	STI	0.65							
Fan HH	SNR		2.67	7.00	10.67	9.67	13.00	14.33	14.00
	T30		1.35	0.98	0.82	0.78	0.83	0.80	0.64
	<i>G</i>		23.21	21.23	20.24	19.97	20.05	19.66	18.71
	Ts		109.33	67.67	54.43	50.53	51.53	52.53	37.17
	D50		0.46	0.57	0.66	0.65	0.63	0.62	0.73
	STI	0.63							

Table 4. Acoustic parameters measured at D110 for four different fan settings.

Fan	Parameter	Frequency							
			125	250	500	1000	2000	4000	8000
No fan	SNR		21.50	21.00	28.50	26.00	32.50	34.00	27.50
	T30		0.77	0.77	0.70	0.66	0.67	0.63	0.54
	<i>G</i>		24.81	22.98	22.76	21.54	22.12	21.53	20.76
	Ts		49.25	36.20	50.50	43.10	40.00	39.65	31.50
	D50		0.80	0.84	0.66	0.69	0.71	0.72	0.79
	STI	0.69							
Fan L	SNR		17.50	20.50	27.00	25.50	32.50	32.50	27.00
	T30		0.70	0.76	0.71	0.68	0.65	0.62	0.54
	<i>G</i>		24.30	23.60	22.83	21.88	22.28	21.72	20.71
	Ts		53.60	39.10	47.70	40.45	40.10	38.70	31.45
	D50		0.73	0.82	0.69	0.69	0.71	0.73	0.79
	STI	0.69							
Fan H	SNR		13.50	17.50	23.50	23.00	28.50	31.00	26.50
	T30		0.70	0.73	0.75	0.65	0.66	0.63	0.54
	<i>G</i>		24.21	23.57	22.91	21.93	22.33	21.60	20.68
	Ts		54.05	39.05	48.80	41.35	40.05	39.85	31.00
	D50		0.73	0.81	0.67	0.69	0.71	0.72	0.80
	STI	0.68							
Fan HH	SNR		6.50	14.00	19.50	19.00	21.50	24.00	22.50
	T30		0.83	0.78	0.69	0.66	0.67	0.62	0.52
	<i>G</i>		24.37	23.59	22.92	21.88	22.25	21.40	20.73
	Ts		57.30	39.60	48.10	41.70	38.50	40.15	31.20
	D50		0.71	0.81	0.68	0.69	0.73	0.72	0.79
	STI	0.63							

Table 5. Acoustic parameters measured at A203 for four different fan settings.

Fan	Parameter	Frequency							
			125	250	500	1000	2000	4000	8000
No fan	SNR		23.67	30.33	36.00	34.33	33.33	29.00	21.00
	T30		1.32	1.07	0.87	0.98	1.12	1.04	0.84
	G		20.24	20.01	18.93	19.47	19.93	19.67	18.22
	Ts		98.90	70.63	57.90	64.33	71.23	65.40	50.90
	D50		0.46	0.56	0.64	0.59	0.52	0.57	0.64
	STI	0.61							
Fan L	SNR		5.67	14.67	20.33	22.33	26.67	27.67	18.33
	T30		1.26	1.05	0.89	0.99	1.10	1.06	0.85
	G		20.35	19.94	18.75	19.45	19.84	19.52	18.15
	Ts		101.53	71.13	58.40	64.87	71.87	66.37	51.17
	D50		0.45	0.56	0.63	0.59	0.52	0.56	0.64
	STI	0.60							
Fan H	SNR		1.00	9.00	14.67	15.67	18.67	19.67	17.00
	T30		1.25	0.94	0.85	0.97	1.06	1.04	0.85
	G		20.54	19.93	18.75	19.35	20.12	19.72	18.39
	Ts		101.87	72.77	59.30	65.17	71.10	66.57	52.27
	D50		0.45	0.56	0.62	0.58	0.53	0.56	0.63
	STI	0.60							
Fan HH	SNR		-4.00	5.00	10.00	9.33	12.00	11.67	12.67
	T30		1.02	0.91	0.87	0.98	1.07	1.05	0.83
	G		20.60	20.14	18.73	19.28	19.98	19.55	18.32
	Ts		105.50	71.77	59.27	64.60	70.67	65.77	52.50
	D50		0.44	0.55	0.64	0.59	0.54	0.57	0.62
	STI	0.57							

Table 6. Acoustic parameters measured at T206 for four different fan settings.

Fan	Parameter	Frequency							
			125	250	500	1000	2000	4000	8000
No fan	SNR		19.67	24.17	32.5	33.30	37.83	34.83	24.00
	T30		2.23	1.48	1.12	1.15	1.23	1.17	0.91
	G		20.79	18.30	17.06	17.48	18.14	17.72	16.23
	Ts		116.58	77.08	77.88	71.15	74.70	69.35	50.87
	D50		0.50	0.56	0.50	0.56	0.54	0.55	0.65
	STI	0.60							
Fan L	SNR		13.17	18.00	23.83	26.17	30.33	31.67	22.83
	T30		2.28	1.46	1.12	1.15	1.23	1.17	0.90
	G		20.84	18.75	17.04	17.49	18.17	17.71	16.27
	Ts		117.03	77.70	77.32	70.70	74.83	69.13	50.90
	D50		0.50	0.56	0.50	0.56	0.54	0.56	0.65
	STI	0.60							
Fan H	SNR		7.83	12.17	17.50	19.00	21.17	20.50	18.17
	T30		2.04	1.43	1.10	1.15	1.22	1.15	0.90
	G		20.78	18.79	17.09	17.45	18.14	17.69	16.34
	Ts		119.07	78.45	78.12	72.25	74.42	69.72	51.02
	D50		0.50	0.55	0.50	0.55	0.54	0.54	0.65
	STI	0.59							
Fan HH	SNR		1.67	7.17	12.17	11.83	14.00	15.00	15.50
	T30		1.79	1.46	1.11	1.12	1.18	1.13	0.88
	G		20.84	18.82	16.99	17.46	18.03	17.65	16.28
	Ts		119.95	78.00	78.45	71.25	75.22	69.88	51.63
	D50		0.49	0.56	0.50	0.56	0.53	0.54	0.64
	STI	0.58							

Table 7. Acoustic parameters measured at A208 for four different fan settings.

Fan	Parameter	Frequency							
			125	250	500	1000	2000	4000	8000
No fan	SNR		19.50	25.00	29.25	30.50	33.00	30.75	24.50
	T30		1.31	0.79	0.79	0.82	0.86	0.84	0.71
	G		19.86	17.50	17.55	17.82	18.06	17.06	17.21
	T_s		90.58	64.33	63.83	58.28	59.63	58.30	44.33
	D50		0.52	0.57	0.52	0.59	0.58	0.58	0.69
	STI	0.64							
Fan L	SNR		14.25	19.00	24.25	24.75	28.25	29.25	23.25
	T30		1.17	0.79	0.79	0.79	0.85	0.85	0.71
	G		19.91	17.47	17.47	17.57	18.01	17.91	17.09
	T_s		90.63	64.70	64.35	59.30	59.95	58.43	45.68
	D50		0.52	0.57	0.52	0.58	0.57	0.58	0.67
	STI	0.63							
Fan H	SNR		8.00	13.50	20.25	20.75	23.50	25.75	23.25
	T30		1.23	0.80	0.78	0.80	0.86	0.83	0.71
	G		19.96	17.43	17.61	17.66	18.04	17.90	17.13
	T_s		90.18	65.45	64.55	59.03	59.88	58.30	44.70
	D50		0.52	0.56	0.52	0.59	0.58	0.59	0.68
	STI	0.63							
Fan HH	SNR		4.50	9.75	15.75	15.00	17.25	18.25	19.75
	T30		1.05	0.73	0.77	0.79	0.83	0.83	0.70
	G		20.10	17.52	17.56	17.63	18.11	17.92	17.21
	T_s		91.35	64.80	63.75	58.53	59.50	57.95	43.63
	D50		0.52	0.57	0.52	0.59	0.57	0.58	0.69
	STI	0.62							

The variation in T30 due to the fan blowing speed is also examined. The average T30 values at 500 Hz and 1000 Hz T30, measured in the spaces against the blowing flow rate of the HVAC used, are shown in Fig. 13. The results show that variations in T30 can be neglected if the just noticeable difference (JND) is taken as 5% relative, according to (ISO, 2009).

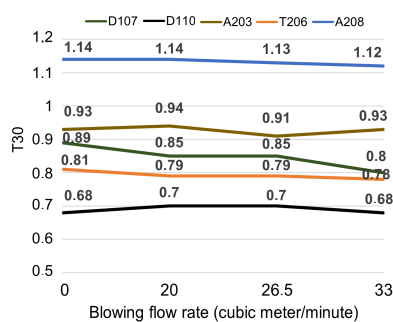


Fig. 13. Variation of T30 in the spaces depending on the impact of HVAC systems.

In line with the above discussion, objective measures for speech intelligibility include STI (HOUTGAST, STEENEKEN, 1985), ALC (articulation loss of consonants) (PEUTZ, 1972), and U50 (useful-to-detrimental ratio) (LOCHNER, BURGER, 1964). Among these, STI is studied in this work to quantify speech intelligibility. The variation in STI in the spaces, depending on

the impact of HVAC systems, is shown in Fig. 14. The ranking presented in the figure is based on the IEC (2020) standard. The results from the measurement survey indicate that the STI values for all five spaces can be categorized as ‘good’, even though the values tend to decrease with increasing fan speed. Note that the JND for STI is 0.03.

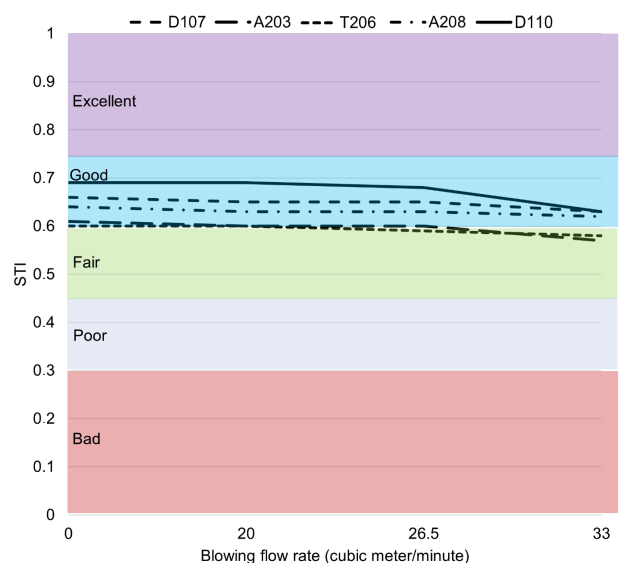


Fig. 14. Evaluation of STI in the spaces depending on the impact of HVAC systems.

On the other hand, it should be noted that some departments in the faculty offer courses in a second language. While the 0.6 threshold is considered ‘good’ for native English students, recent research suggests that this may not be the case for non-native speakers (ISO, 2003; MINELLI *et al.*, 2022). For students whose first language is not English but who use English as a daily second language, an STI value of 0.68 and above can only be considered ‘good’. For students with an intermediate level of proficiency and those with low level of their second language use, an STI value of 0.86 and above can be considered ‘good’.

4. Conclusion

Background noise, RT, and the distance between the listener and speaker all affect STI values. Since SNR is the ratio of speech compared to BNL, it can be considered a key factor in speech intelligibility. The impact of HVAC systems on SNR is significant and negatively affects speech intelligibility, as reflected in the reported STI values.

As highlighted in previous studies (LONGONI *et al.*, 2016; RAZALI *et al.*, 2024), the presence of HVAC systems increases BNL, decreases SNR and deteriorates speech intelligibility. Recommended values for speech intelligibility are T30 between 0.7 s and 1.2 s and D50 >0.5 in rooms (MASOVIC, 2021). For classrooms, optimum T30 values are <0.6 s for students under 12 years old and <0.8 s for students aged 12 and above (Building Bulletin, 2015; MINELLI *et al.*, 2022). The studied spaces partially meet these optimal ranges.

The results show that SNR values tend to decrease as the fan blowing speed of HVAC systems increases, which negatively affects speech intelligibility.

Another concern is whether HVAC systems alter acoustic parameters that affect speech intelligibility, such as RT. The results of the study indicate that the impact of HVAC systems on the reverberation characteristics of space is negligible when the JND value is considered. On the other hand, it is clear that as the size of the space decreases, the RT also decreases, leading to a better STI value.

Previous research (ASTOLFI *et al.*, 2012; MURGIA *et al.*, 2023) suggests that speech intelligibility for learners aged 12 and under requires an STI value of 0.65 and above, while for learners aged 12 and above, STI values of 0.6 and above are considered acceptable. For university students, it can be concluded that if STI values of 0.6 and above are achieved, there is no cause for concern in terms of speech intelligibility. The relevant standard (ISO, 2003) categorizes STI values between 0.6 and 0.75 as ‘good’. However, it is worth noting that for non-native speakers of English, the STI value must fall between 0.68–0.86, depending on their level of English proficiency, to be considered ‘good’.

FUNDINGS

This study was supported by the Scientific and Technological Research Council of Türkiye (TUBITAK) under grant no. 123M884. The authors thank TUBITAK for their support.

References

1. American National Standard (2010), *Acoustical performance criteria, design requirements, and guidelines for schools, Part 1: Permanent Schools* (ANSI/ASA S12.60-2010/Part 1), <https://webstore.ansi.org/standards/asa/ansiasas1260part2010r2020?source=blog> (access: 15.07.2024).
2. ANTONIADOU S., PAPADAKIS N.M., STAVROULAKIS G.E. (2018), Measuring acoustic parameters with ESS and MLS methods: Effect of artificially varying background noise, [in:] *Euronoise 2018 – Conference Proceedings*, https://www.euronoise2018.eu/docs/papers/132_Euro_noise2018.pdf (access: 22.07.2024).
3. ASTOLFI A., BOTTALICO P., BARBATO G. (2012), Subjective and objective speech intelligibility investigations in primary school classrooms, *The Journal of the Acoustical Society of America*, **131**(1): 247–257, <https://doi.org/10.1121/1.3662060>.
4. Building Bulletin (2015), *BB93: Acoustic design of schools – Performance standards*, Department for Education, Education Funding Agency, England, https://assets.publishing.service.gov.uk/media/5a8170d3e5274a2e8ab54012/BB93_February_2015.pdf (access: 14.08.2024).
5. CHOI Y.-J. (2020), Evaluation of acoustical conditions for speech communication in active university classrooms, *Applied Acoustics*, **159**: 107089, <https://doi.org/10.1016/j.apacoust.2019.107089>.
6. CUSHING I.R., LI F.F., COX T.J., WORRALL K., JACKSON T. (2011), Vocal effort levels in anechoic conditions, *Applied Acoustics*, **72**(9): 695–701, <https://doi.org/10.1016/j.apacoust.2011.02.011>.
7. DI LORETO S., CANTARINI M., SQUARTINI S., LORI V., SERPILLI F., DI PERNA C. (2023), Assessment of speech intelligibility in scholar classrooms by measurements and prediction methods, *Building Acoustics*, **30**(2): 165–202, <https://doi.org/10.1177/1351010X231158190>.
8. ELLIOTT T.M., THEUNISSEN F.E. (2009), The modulation transfer function for speech intelligibility, *PLoS Computational Biology*, **5**(3): e1000302, <https://doi.org/10.1371/journal.pcbi.1000302>.
9. ENGEL M., HERRMANN J., ZANNIN P. (2020), Assessment of the sound quality of classrooms through speech transmission index (STI), sound definition (D50) and reverberation time (RT), *Forum Acusticum*.
10. GUIDORZI P.A., BARBARESI L.U., D’ORAZIO D.A., GARAI M.A. (2015), Impulse responses measured with MLS or Swept-Sine Signals applied to architectural acoustics: An in-depth analysis of the two methods and

- some case studies of measurements inside theaters, *Energy Procedia*, **78**: 1611–1616, <https://doi.org/10.1016/j.egypro.2015.11.236>.
11. HOUTGAST T., STEENEKEN H. J. (1985), A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria, *The Journal of the Acoustical Society of America*, **77**(3): 1069–1077, <https://doi.org/10.1121/1.392224>.
 12. HUANG W., PENG J., XIE T. (2023), Study on Chinese speech intelligibility under different low-frequency characteristics of reverberation time using a hybrid method, *Archives of Acoustics*, **48**(2): 151–157, <https://doi.org/10.24425/aoa.2023.145229>.
 13. International Organization for Standardization (2003), *Ergonomics – Assessment of speech communications* (ISO Standard No. 9921:2003), <https://www.iso.org/standard/33589.html> (access: 05.09.2024).
 14. International Organization for Standardization (2009), *Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces* (ISO Standard No. 3382-1:2009), <https://www.iso.org/standard/40979.html> (access: 05.09.2024).
 15. International Organization for Standardization (2014), *Acoustics – Field measurement of sound insulation in buildings and of building elements – Part 1: Airborne sound insulation* (ISO Standard No. 16283-1:2014), <https://www.iso.org/standard/55997.html> (access: 05.09.2024).
 16. International Organization for Standardization (2021), *Acoustics – Laboratory measurement of sound insulation of building elements – Part 1: Application rules for specific products* (ISO Standard No. 10140-1:2021), <https://www.iso.org/standard/73910.html> (access: 05.09.2024).
 17. International Electrotechnical Commission (2020), *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index* (IEC Standard No. 60268-16:2020), <https://webstore.iec.ch/en/publication/26771> (access: 20.08.2024).
 18. KAWATA M., TSURUTA-HAMAMURA M., HASEGAWA H. (2023), Assessment of speech transmission index and reverberation time in standardized English as a foreign language test room, *Applied Acoustics*, **202**: 109093, <https://doi.org/10.1016/j.apacoust.2022.109093>.
 19. LOCHNER J.P.A., BURGER J.F. (1964), The influence of reflections on auditorium acoustics, *Journal of Sound and Vibration*, **1**(4): 426–454, [https://doi.org/10.1016/0022-460X\(64\)90057-4](https://doi.org/10.1016/0022-460X(64)90057-4).
 20. LONGONI H.C. et al. (2016), Speech transmission index variation due to ventilating and air-conditioning system in university classrooms, [in:] *Proceedings of Meetings on Acoustics*, **28**: 015024, <https://doi.org/10.1121/2.0000470>.
 21. MASOVIC D. (2021), Room acoustics, arXiv, <https://doi.org/10.48550/arXiv.2111.01900>.
 22. MEJDI A., GARDNER B., MUSSER C. (2019), Prediction of the speech transmission quality in the presence of background noise using the ray tracing technique, [in:] *26th International Congress on Sound and Vibration (ICSV)*, https://www.academia.edu/64361351/Prediction_of_the_Speech_Transmission_Quality_in_the_Presence_of_Background_Noise_Using_the_Ray_Tracing_Technique (access: 15.11.2024).
 23. MENG Q., SEN D., WANG S., HAYES L. (2008), Impulse response measurement with sine sweeps and amplitude modulation schemes, [in:] *2nd International Conference on Signal Processing and Communication Systems*, <https://doi.org/10.1109/ICSPCS.2008.4813749>.
 24. MINELLI G., PUGLISI G.E., ASTOLFI A. (2022), Acoustical parameters for learning in classroom: A review, *Building and Environment*, **208**: 108582, <https://doi.org/10.1016/j.buildenv.2021.108582>.
 25. MURGIA S., WEBSTER J., CUTIVA L.C.C., BOTTALICO P. (2023), Systematic review of literature on speech intelligibility and classroom acoustics in elementary schools, *Language, Speech, and Hearing Services in Schools*, **54**(1): 322–335, <https://doi.org/10.1044/2022.LSHSS-21-00181>.
 26. PAYTON K.L., SHRESTHA M. (2008), Evaluation of short-time speech-based intelligibility metrics, [in:] *The 9th International Congress on Noise as a Public Health Problem (ICBEN)*, <https://www.researchgate.net/publication/256579565> (access: 15.11.2024).
 27. PEUTZ V.M.A. (1972), Articulation loss of consonants as a criterion for speech transmission in rooms, [in:] *Audio Engineering Society Convention 2ce. Audio Engineering Society*, <https://aes2.org/publications/elibrary-page/?id=1821> (access: 15.11.2024).
 28. RABELO A.T.V., SANTOS J.N., OLIVEIRA R.C., MAGALHÃES M.D.C. (2014), Effect of classroom acoustics on the speech intelligibility of students, *CODAS*, **26**(5): 360–366, <https://doi.org/10.1590/2317-1782/20142014026>.
 29. RAZALI A.W., DIN N.B.C., YAHYA M.N., SULAIMAN R. (2024), Conceptual framework of acoustic comfort design enablers for a classroom: A systematic review, *Journal of Building Engineering*, **95**: 110160, <https://doi.org/10.1016/j.jobbe.2024.110160>.
 30. ROSSING T.D. [Ed.] (2014), *Springer Handbook of Acoustics*, 2nd ed., Springer, Germany.
 31. YANG D., MAK C.M. (2018), An investigation of speech intelligibility for second language students in classrooms, *Applied Acoustics*, **134**: 54–59, <https://doi.org/10.1016/j.apacoust.2018.01.003>.
 32. ZHU P., TAO W., MO F., LU X., ZHANG H. (2024), Experimental comparisons of speech transmission index prediction methods, *Applied Acoustics*, **220**: 109985, <https://doi.org/10.1016/j.apacoust.2024.109985>.

Research Paper

Structural Parametric Study of a Piezoelectric Energy Harvester for a Specific Excitation Frequency of an Electric Motor, Considering Fatigue Life

Ping YANG^{(1),(2)}, Nabil MOHAMAD USAMAH⁽¹⁾, Abdullah Aziz SAAD⁽¹⁾,
Ahmad Zhafran AHMAD MAZLAN^{(1)*} 

⁽¹⁾ School of Mechanical Engineering, Engineering Campus, Universiti Sains Malaysia
Penang, Malaysia

⁽²⁾ College of Intelligent Manufacturing and Automobile, Chongqing Technology and Business College
Chongqing, China

*Corresponding Author e-mail: zhafran@usm.my

Received May 8, 2024; revised January 9, 2025; accepted April 9, 2025;
published online June 6, 2025.

Over the past decade, extensive research has been conducted in the field of piezoelectric energy harvesting, which drives advancements in novel designs and techniques. In this study, the vibration of an electric motor is characterized, and a piezoelectric energy harvester (PEH) beam with a natural frequency of 50 Hz is designed and fabricated. Then, the electromechanical characteristics of the PEH are simulated and tested through both finite element simulation and experiment. The validated simulation model can accurately predict the vibration characteristics of the PEH, which can be utilized for design improvements. Based on this beam structure, three sets of PEHs with different sizes are designed. A study of the output voltage and fatigue life of these different PEH sizes is conducted, and the relationship between the electromechanical coupling effect and its varying values is discussed. Based on the results, design schemes ①~⑥ demonstrate advantages in terms of output voltage efficiency and fatigue strength, making them suitable for various environments and application purposes. This study establishes an efficient method for analyzing the structural parametric performance of piezoelectric cantilever beams, which paves the way for future research on fatigue-based structural design guidelines for PEH in electric motors.

Keywords: piezoelectric energy harvester (PEH); fatigue; output voltage; vibration.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Intelligent devices in the current digital era require constant monitoring to ensure long-term operation and functionality, establish operational profiles, and anticipate specific malfunctions. Electrical devices generate minor wideband vibrations during regular operation. Before malfunctions occur, oscillations that cause variations in displacement, frequency, and other related parameters can be distinguished from steady-state operations. Multiple sensors are needed to monitor the devices during operation, and it has become common to seek an appropriate power source from these elec-

trical devices. At present, the most promising method for powering wireless sensors is to capture the vibration energy produced by the device itself. The vibration frequencies generated by these devices are generally in the power spectrum and harmonic waves. This study focuses on the optimal design of a piezoelectric energy harvester (PEH) for 50 Hz vibrations excited by an electric motor, with the objective of increasing energy output while also improving the structure's durability.

Piezoelectric materials have the ability to convert mechanical vibration energy into electrical energy and have been widely investigated for energy harvesting

due to their high energy density (LI, LEE, 2022). Piezoelectric energy harvesting is a burgeoning topic that has been investigated for use in construction, machinery, automobiles, and aerospace, owing to the significant amount of wasted vibration energy produced by these systems. These devices are frequently utilized as a power source for self-powered wireless sensors and employed in situations where external power is unavailable and using batteries is not a practical solution. The initial application of PEH was presented by PANDA *et al.* (2022), where a lead zirconate titanate (PZT) patch was attached to the root of a cantilever beam.

Piezoelectric elements may experience considerable deformation due to the cantilever beam's ability to modify the frequency and increase the vibration amplitude when a mass block is applied. This configuration is currently the most widely used because of its ease of implementation, high efficiency, and well-developed technology. Extensive research has been carried out over the past two decades on the design, functionality, and application of cantilever beam piezoelectric harvesters. ZHANG *et al.* (2022) indicated that vibration-based PEHs may produce maximum power when operated at the resonant frequency, and the power output drops dramatically as the natural frequency of the PEH deviates from the vibration frequency. According to RAFIQUE and BONELLO (2010), it was found that the resonant frequency of the cantilever beam is negatively correlated with its length while its thickness is positively correlated with the resonance frequency. When the vibration characteristics of the PEH match the vibration frequency of the surrounding environment, the harvester will undergo multi-modal resonance, resulting in significant deformation and causing large stress and strain at the end of the cantilever beam. The harvester's lifespan may be impacted over time by fatigue damage due to the prolonged concentration of stress and strain. Additionally, the output voltage is proportional to its length and inversely proportional to its width and thickness (WANG *et al.*, 2019).

To match the low-frequency vibrations of the surroundings and increase the PEH output voltage, the cantilever beam's length and mass must also be increased, or its width and thickness need to be decreased. However, ROUNDY *et al.* (2005) concluded that as the aspect ratio increases (as natural frequencies differ across structures), the strain at the root of the cantilever beam changes drastically, causing the overall deformation of the piezoelectric layer to increase. As a result, the piezoelectric layer experiences an increase in stress, which has a substantial effect on its fatigue life.

Relevant research has indicated that fatigue life and output voltage are incompatible and cannot be optimized by a single variable. Current research has de-

voted considerable effort to optimizing the structure of energy harvesters based on structural strain and material strength to improve energy harvesting conversion efficiency and power output. Few studies address the contradictions between fatigue damage and output voltage caused by structural parameters. However, there are almost no references studying the coupling effects of multiple variables on output voltage and fatigue life. A few commercially PEHs are available, but they are limited by a narrow bandwidth of operational frequency and concerns regarding the reliability and durability of the structure.

Although numerous studies have investigated the structural parametric and fatigue-related issues of PEH, most have focused on the impact of a single variable on either output voltage or fatigue life. There are not many studies that systematically examine and discuss the relationships among structural size, fatigue damage, and output voltage. Moreover, most research adopts the approach of matching external vibration excitation to the natural frequency of the structure, rather than adjusting the structure's inherent frequency to match a specific external vibration source. However, modifying a single variable can cause changes in the natural frequency of the structure, limiting the practical value of comparing the performance of structures with different natural frequencies in a consistent vibration environment in terms of energy absorption. The most important aspect of designing a PEH beam to match a specific surrounding frequency is to examine the combine influence of multiple variables. These relationships do not have a simple linearly increasing or decreasing relationship, and a practical approach is required to optimize appropriate structural dimensions.

To address this gap, the main aim of this study is to design PEHs integrated into an electric motor by evaluating their energy output and fatigue life, using PEHs with similar natural frequencies but different sizes through laboratory testing and numerical simulation. Then, the optimal design of PEH beams is determined to achieve an optimal balance between fatigue damage and maximum output voltage. This work is organized into six major sections. First, the background, and recent advances in PEHs are reviewed. The second section introduces the electro-mechanical model and the energy harvesting system. In Sec. 3, the dynamic characteristics of the motor and PEH beams with a natural frequency of 50 Hz of natural frequencies are measured and developed, and their electro-mechanical characteristics are then examined using both ANSYS software and experimental validation. In Sec. 4, three sets of PEHs are theoretically modeled, their output voltages and maximum stress are simulated, and five beams are selected for further analysis. In Sec. 5, the fatigue life of the selected beams is calculated, and the design strategies under different usage

environments are discussed. The conclusion summarizes the major findings and the research value of this work.

2. Electro-mechanical model of PEH

Currently, the most commonly used PEH structure is the cantilever beam type (SEZER, KOÇ, 2021), mainly composed of a piezoelectric layer, a beam, and a mass block, as shown in Fig. 1.

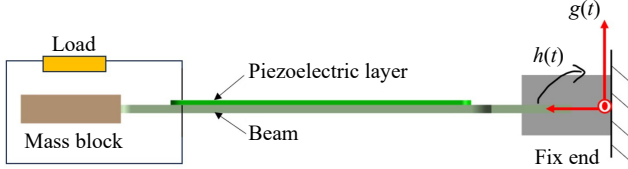


Fig. 1. Piezoelectric cantilever beam model.

Assume that the length of the piezoelectric cantilever beam is L , width is b , thickness of the piezoelectric layer is H , thickness of the elastic substrate layer is D , and the mass block weight is M . The system is subjected to translational excitation $g(t)$ and a small rotational excitation $h(t)$ at the base. Due to the large aspect ratio of the beam, the effects of shear deformation and rotational inertia of the beam are neglected, and the beam is treated as an equivalent Euler–Bernoulli beam. The free vibration equation of the piezoelectric beam can be expressed as (FENG *et al.*, 2023)

$$E \frac{\partial^4 w(x, t)}{\partial x^4} + m \frac{\partial^2 w(x, t)}{\partial t^2} = 0, \quad (1)$$

where E represents the bending stiffness of the piezoelectric beam, m is the mass per unit length of the piezoelectric beam, and $w(x, t)$ represents the transverse displacement at position x along the beam's neutral axis relative to the fixed inertial system. The transverse vibration displacement of the piezoelectric cantilever beam can be expressed as

$$w(x, t) = w_b(x, t) + w_{\text{rel}}(x, t), \quad (2)$$

where $w_b(x, t)$ is the base excitation displacement, and $w_{\text{rel}}(x, t)$ is the displacement relative to the fixed end. Considering the electromechanical coupling effect, the dynamic equation of the piezoelectric cantilever beam can be further written as

$$\begin{aligned} E \frac{\partial^4 w_{\text{rel}}(x, t)}{\partial x^4} + c_s I \frac{\partial^5 w_{\text{rel}}(x, t)}{\partial x^4 \partial t} \\ + c_a \frac{w_{\text{rel}}(x, t)}{\partial t} + m \frac{\partial^2 w_{\text{rel}}(x, t)}{\partial t^2} + \xi V_0(t) \\ = -[m + M\delta(x - L)] \frac{\partial^2 w_b(x, t)}{\partial t^2}, \end{aligned} \quad (3)$$

where c_s is the strain damping coefficient, c_a is the air viscoelastic damping coefficient, and ξ is the electromechanical coupling coefficient. Using the modal superposition method, the solution to Eq. (3) can be written as

$$w_{\text{rel}}(x, t) = \sum_{r=1}^{\infty} \varphi_r(x) q_r(t), \quad (4)$$

where $\varphi_r(x)$ is the normalized mode shape function for the r -th mode of the piezoelectric cantilever beam, and $q_r(t)$ is the corresponding generalized modal coordinate. Solving Eq. (4) yields the relative vibration response of the beam:

$$\begin{aligned} w_{\text{rel}}(x, t) = \alpha e^{j\omega t} \sum_{r=1}^{\infty} \frac{\omega^2 G_0 [m\gamma_r^\omega + M\varphi_r(L) - X_r V_m]}{\omega_r^2 - \omega^2 + j2\zeta_r \omega_r \omega} \\ \cdot \left[\cos \frac{\lambda_r}{L} x - \cosh \frac{\lambda_r}{L} x \right. \\ \left. + K \left(\sin \frac{\lambda_r}{L} x - \sinh \frac{\lambda_r}{L} x \right) \right], \end{aligned} \quad (5)$$

where

$$\gamma_r^\omega = \int_0^L \varphi_r(x) dx, \quad (6)$$

G_0 is the base excitation displacement amplitude, ω is the excitation frequency, j is the imaginary unit, λ_r is the dimensionless frequency of the r -th mode, α is the modal amplitude constant, ω_r is the undamped natural frequency of the r -th mode, X_r is the modal electromechanical coupling coefficient, ζ_r is the mechanical damping ratio, V_m is the voltage amplitude, and the expression for K is:

$$K = \frac{\sin \lambda_r - \sinh \lambda_r + \lambda_r \frac{M}{mL} (\cos \lambda_r - \cosh \lambda_r)}{\cos \lambda_r + \cosh \lambda_r - \lambda_r \frac{M}{mL} (\sin \lambda_r - \sinh \lambda_r)}. \quad (7)$$

According to Kirchhoff's law and the electrical response equation, the expression for the power output of the piezoelectric cantilever beam can be written as

$$P(t) = \frac{V^2(t)}{R} = \frac{1}{R} \left(\frac{\sum_{r=1}^{\infty} \mu_r \frac{j\omega^3 [m\gamma_r^\omega + M\varphi_r(L)]}{\omega_r^2 - \omega^2 + j2\zeta_r \omega_r \omega}}{\sum_{r=1}^{\infty} \frac{j\omega X_r \mu_r}{\omega_r^2 - \omega^2 + j2\zeta_r \omega_r \omega} + \frac{j\omega\psi + 1}{\psi}} G_0 e^{j\omega t} \right)^2, \quad (8)$$

where

$$\psi = \frac{R\epsilon_{33}^s bL}{H}, \quad (9)$$

$$\mu_r = \frac{-d_{31} Y_p e H}{\epsilon_{33}^s L} \frac{d\varphi_r(x)}{dx} \Big|_{x=L}, \quad (10)$$

$$\epsilon_{33}^s = \epsilon_{33}^t - d_{31}^2 Y_p, \quad (11)$$

where Y_p is the elastic modulus of the piezoelectric layer, d_{31} is the piezoelectric constant, e is the distance

from the middle layer of the piezoelectric cantilever beam to the neutral axis, and ϵ_{33}^t represents the elastic compliance constant.

Solving Eqs. (5) and (8) reveals that the geometric dimensions and material parameters of the piezoelectric cantilever beam have a significant impact on its electromechanical response. This observation provides a theoretical foundation for subsequent finite element modeling and simulation analysis (BAO *et al.*, 2021; CHEN *et al.*, 2020).

3. Structural design of PEH

3.1. Dynamic test and characteristics of electric motor

The majority of PEHs are designed for general-purpose applications and are evaluated under simplified harmonic excitations. However, this approach is still far from being ready to be used in real-world applications.

It is well established that vibration-based PEHs generate maximum power when operated at their resonant frequency. A straightforward and effective solution to address this is to broaden the bandwidth of the beam's resonant frequency. The natural frequencies of the motor (shown in Fig. 2a) and the system's dynamic properties under operating conditions must be first tested and analyzed before conducting the PEH design and experiment. These parameters are critical data for conducting the structure design of the PEH (LU, 2018).

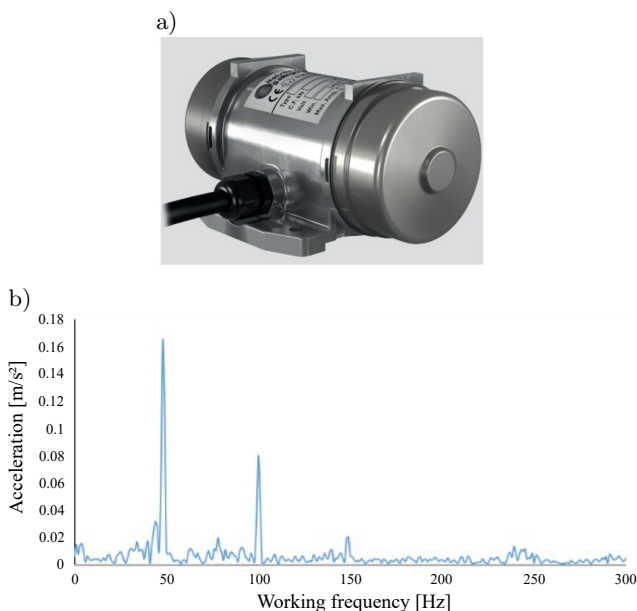


Fig. 2. a) Electric motor (Gaomeng, 2020); b) acceleration frequency spectrum under working condition.

The resonance between the PEH and the motor's natural frequency is not considered here because the

motor's modal test results, conducted earlier, indicated a very high natural frequency. The acceleration of the motor under working conditions (maximum speed) was measured, and the corresponding frequency spectrum is shown in Fig. 2b. From the figure, it can be observed that the motor exhibits several vibration peaks during operation. The peak acceleration response, which is generated by electromagnetic excitation, occurs at 50 Hz, 100 Hz, and 150 Hz – multiples of 5 Hz. These vibration frequencies generated during the operation of electrical devices, such as motors, are at the main operating frequency (50 Hz) and its harmonics. Hence, studying energy harvesting from the vibration frequencies of motors offers significant applicability.

3.2. Finite element analysis

Finite element analysis (FEA) is one of the reliable methods for studying the performance of any designed system prior to its prototype development. Studies by ZHU *et al.* (2010) and AUGUSTYN *et al.* (2014) conducted parametric studies using FEA to determine the ideal system configuration. Meanwhile, studies by ABDELKEFI *et al.* (2014) and AVVARI *et al.* (2017) employed FEA to validate the experimental and analytical models of PEHs to guide further simulation. This work utilizes FEA to guide preliminary structural design and the subsequent structural parametrization, along with dynamic characteristics analysis.

3.2.1. Parameter selection and analysis

Material selection and optimal structural design are crucial for energy harvesting. One important limitation of existing energy harvesting techniques is that the power output performance is subjected to the resonant frequencies of ambient vibrations, which are often random and broadband. To address this issue, researchers have focused on developing efficient PEHs using novel piezoelectric materials by adjusting the natural frequency of the harvester to match the desired vibration frequency. In terms of materials, piezoelectric materials can be categorized into natural and synthetic types. Synthetic materials are further subdivided into ceramic and polymer-based categories. The widely used piezoelectric material is PZT, known for its excellent piezoelectric properties and high dielectric constant. However, it is prone to fatigue fracture under high-frequency cyclic vibration, leading to certain limitations in its application (NIASAR *et al.*, 2020). Thus, polymer-based transducers, such as polyvinylidene difluoride (PVDF) and macro fiber composite (MFC), have gained popularity in recent years due to their flexibility, durability, and resistance to humidity (SHI *et al.*, 2017).

In this study, a laminated structure consisting of a piezoceramic plate, electrodes, and polymer materials, referred to as P-876 DuraAct patch trans-

ducer, is selected (ABDUL SATAR *et al.*, 2022). This material has the advantages of high piezoelectric coefficients and dielectric constants and can be applied onto curved surfaces or used for integration into structures. The operation mode of d_{31} is chosen, with the polarization of the electric potential perpendicular to the stress direction. In terms of the structure, a classic rectangular cantilever beam PEH is adopted, in which the piezoelectric patch is attached to the base near the fixed end, and a tip mass is attached at the free end to decrease the natural frequency.

The structural parametric study can be divided into size, shape, and topology. Size parametric study involves adjusting structural dimension parameters, such as cross-sectional area and thickness, to improve structural performance while maintaining the basic shape and topological configuration. This parametric study involves applying mathematical modeling, simulation, and specific algorithms to find the optimal combination of structural parameters that meet performance, cost, and weight constraints. Given the type and basic dimensions of the piezoelectric sensor have been determined, the research focuses on the effects of beam length and width on output voltage to achieve maximum energy production. The preliminary design of the cantilever harvester base uses an aluminum alloy with a width equal to that of the piezoelectric patch, and a thickness of 0.2 mm (ABDUL SATAR *et al.*, 2022). As discussed in Sec. 2, the experiment results indicate that the acceleration response is mainly attributed to the working frequency of the motor. Therefore, the fundamental frequency of the PEH should be tuned to match the motor's working frequency of 50 Hz. The structural design and parametric study of the PEH will be carried out using the motor's operating frequency of 50 Hz as a target operating frequency, considering both the fatigue life and energy output of the PEH beams.


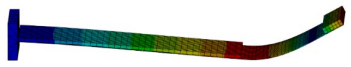

3.2.2. Validation of PEH

Using the piezoelectric coupling analysis module in ANSYS, the SOLID5 element, specifically designed for piezoelectric analysis, is chosen for the piezoelectric layer, while SOLID45 elements are selected for the metal layer and end mass block. Voltage degrees of freedom coupling are applied to the piezoelectric

layer as electrical constraints, and zero displacement constraints are applied to the fixed end as mechanical constraints for modal analysis. The structural natural frequencies and mode shapes can be determined using this analysis to further ascertain its size and inherent characteristics.

The FE modeling and the first three-order modes of the PEH are presented in Table 1. The first-order mode, at 50.05 Hz, aligns precisely with the frequency at which the motor operates. The second-order natural frequency is 125.90 Hz, which is much higher than the first-order mode. The red region of the vibration mode indicates the maximum deformation, while the blue region represents the minimum deformation. The first modal shape exhibits bending deformation, as shown in Table 1, where the maximum displacement occurs at the tip mass block and gradually decreases towards the fixed end. The second mode shape involves shear deformation, with the maximum deformation occurring at positions along the edge of the piezoelectric element, away from the fixed end. The third mode shape is mainly associated with twisting deformation around the horizontal axis. Unlike the bending mode, in the case of torsional vibration, the displacement at the midpoint of the cantilever beam is minimal, while the displacements on both sides in the width direction are larger.

Table 1. Modal frequency and mode shapes.

Modal order	Modal frequency [Hz]	Mode shapes
1	50.05	
2	125.90	
3	198.45	

The first-order mode of the PEH is the fundamental mode and the most appropriate for obtaining the maximum output, making it suitable for the d_{31} working mode. Hence, only the first-order mode is considered in this paper. Table 2 presents the material parameters of the PEH, along with the structural dimensions.

Table 2. Material parameters.

Parameter	Symbol	Piezoelectric patch	Substrate layer (aluminum alloy)	Mass (structure steel)
Density [kg/m^3]	ρ	7500	8920	7850
Young's modulus [GPa]	E_m	56	71.7	210
Volume [mm]	$L \times W \times t$	$60 \times 35 \times 0.8$	$90 \times 35 \times 0.2$	$35 \times 5 \times 2.5$
Poisson's ratio	ν	0.36	0.33	0.31
Piezoelectric constant [$\text{C} \cdot \text{m}^{-2}$]	d_{31}	2.74×10^{-10}	–	–
Dielectric constant [$\text{nF} \cdot \text{m}^{-1}$]	ϵ^t	3.01×10^{-8}	–	–

4. Fabrication and experimental verification of PEH

The surface of the piezoelectric patch is coated with a silver electrode, and external wires are led out from here. The piezoelectric patch is bonded to an aluminum substrate using epoxy resin. Furthermore, the PEH is clamped in a rigid holder and mounted on an electromagnetic shaker to provide the base excitation, as illustrated in Fig 3.

An algorithm to generate a sinusoidal voltage signal is constructed using LabVIEW software, and this signal is supplied through an output module (NI cDAQ 9263), which is then amplified by a power amplifier to drive the shaker (Sentek BTM-100-M). The harvested voltage from the PEH and the measured acceleration are then fed back to LabVIEW software for data monitoring and analysis through the input module (NI cDAQ 9234). The control algorithm for the harvested voltage and output acceleration of the PEH, in terms of time domain and frequency response, is designed in LabVIEW software for further analysis and data validation.

The PEH experimental setup is shown in Fig. 4. The energy harvesting system operates as a coupled field of mechanical and piezoelectricity (electromechanical system). In the experiment, a signal of frequency sweep from 0 Hz to 100 Hz was supplied to

the PEH using an open-circuit sinusoidal excitation, with the input acceleration load recorded at 3 m/s^2 . The resulting tip acceleration-frequency and voltage-frequency responses from this experiment are shown in Fig. 5. It can be observed that when an external excitation frequency approaches the natural frequency of the cantilever beam, the vibration acceleration and output voltage of the PEH reach their peaks. The curve of voltage peaks on both sides exhibits a symmetrical decreasing trend, which indicates that the fabricated piezoelectric oscillator has a clear natural frequency of 50 Hz.

The harmonic response analysis simulation is then carried out on the PEH model under the same excitation levels using ANSYS software. The comparison between the simulation and experimental voltages with open circuits is also illustrated in Fig. 5. This figure indicates that the peak voltage and natural frequency obtained from the simulation are nearly identical to experimental measurements with a 98.5% correlation. The voltage drops sharply when the frequency deviates from the resonant frequency.

It should be noted that the peak voltage obtained from the FEA simulation is slightly higher than the experimental values because of the ideal boundary and loading conditions of FEA. Nevertheless, the FEA modeling results agree well with the experimental values, with the frequency differences at peak voltage values

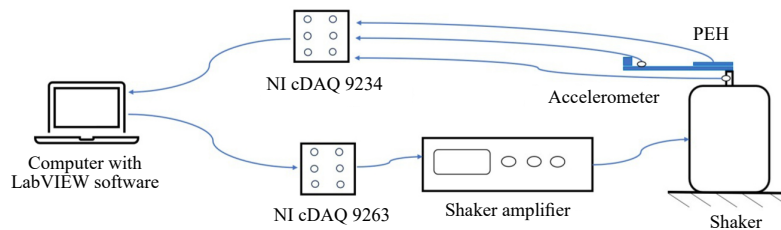


Fig. 3. Diagram of PEH experiment.

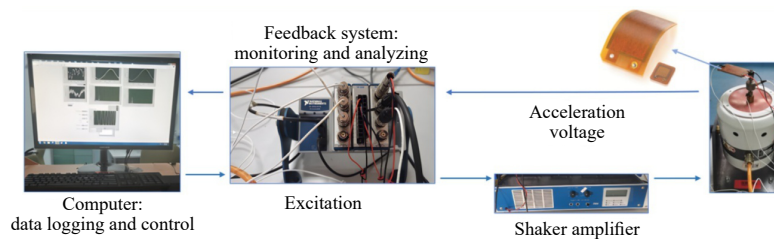


Fig. 4. Experimental setup for PEH experiment.

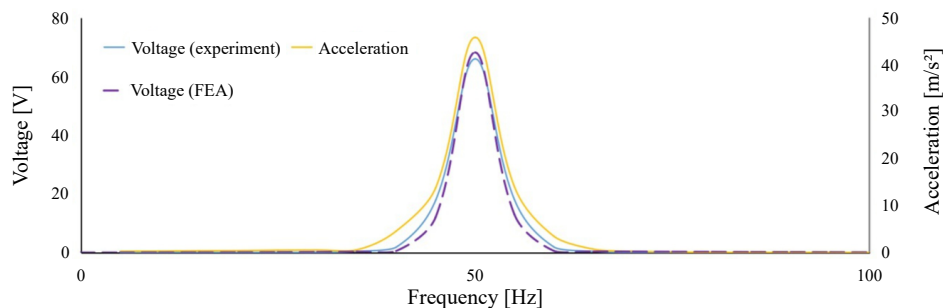


Fig. 5. Comparison of output voltage and acceleration in the frequency-domain.

falling within a range of 0.5 Hz to 1 Hz. Therefore, the FEA model is valid for the fatigue-based structural design study of the PEH.

5. Fatigue-based mechanism and parametric study

5.1. Design of structural parameters

The natural frequency and output voltage of piezoelectric cantilever beams are influenced by several factors. Therefore, it is necessary to perform a statistical analysis of various influencing factors and choose the optimal solution that achieves a natural frequency of 50 Hz, a large output voltage, and acceptable fatigue reliability. By varying the height (t) of the tip mass, a batch of piezoelectric cantilever beams with equal areas but different aspect ratios are designed based on the results from Sec. 3, targeting a natural frequency of 50 Hz. The width and length are rounded to one decimal place to account for constraints during physical experiment.

In order to accurately ascertain the specific size and inherent characteristics, modal analysis of PEHs with the same sectional area but different shapes was conducted using ANSYS software. The voltage-frequency and stress-frequency responses of the PEHs were measured under harmonic base excitation with an acceleration of 3 m/s^2 , as this work aimed to evaluate how the harvesters of different sizes performed when subjected

to identical loading conditions. The output voltage was used as the key parameter to assess the performance of the harvester, while the maximum stress in the PEH beam provided further insight into the behavior of the beam. In this study, only the stress and fatigue life of the piezoelectric patch were considered, as the fatigue strength and bending strength of the matrix material are noticeably higher than those of the piezoelectric patch.

As expected, the fundamental frequency of the designed PEHs with different aspect ratios is close to 50 Hz, as shown in the PEH voltage-frequency responses in Fig. 6a. Meanwhile, the variations in output voltage and stress versus the beam length are shown in Fig. 6b. The different sizes of PEHs combined with their natural frequency, output characteristics, and stress are summarized in Table 3. From Table 3 and Fig. 6, we can observe that both the output voltage and maximum stress exhibit an upward trend as the aspect ratio of the beams increases, with the stress increases being more noticeable. A turning point is reached at a 85 mm length and a 4.3 mm mass height, where the output voltage starts to drop, and the stress increases. The reason may be that the tip mass has an amplifying effect on the output voltage, as supported by pertinent data in (LI, LEE, 2022), and the maximum coupling effect occurring between the beam length and mass height. Subsequently, with the height of the tip mass fixed at 4.3 mm, a set of PEHs with variable lengths and widths will be constructed with a 50 Hz natural frequency as the target.

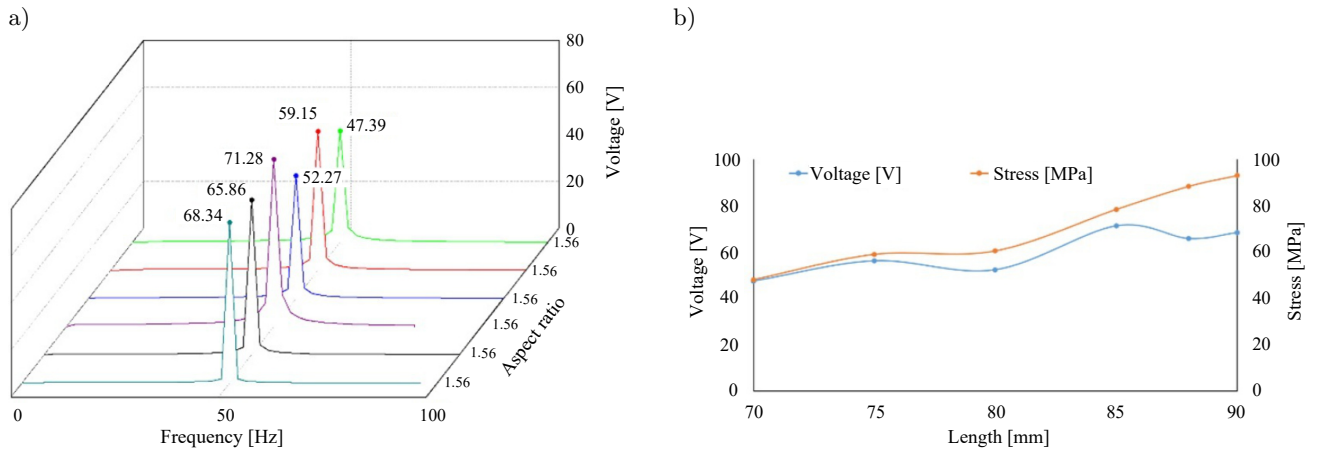


Fig. 6. a) Voltage-frequency response of different beam aspect ratios; b) voltage and stress induced by different beam lengths.

Table 3. Dimensions and performance parameters of Fig. 6.

Aspect ratio (L/W)	$L \times W \times t$ [mm]	Frequency [Hz]	Voltage [V]	Stress [MPa]
1.56	$70 \times 45 \times 9.2$	50.2	47.39	48.62
1.79	$75 \times 42 \times 7.2$	50.3	59.15	58.91
2.05	$80 \times 39 \times 5.6$	50.2	52.27	60.39
2.30	$85 \times 37 \times 4.3$	50.3	71.28	78.28
2.45	$88 \times 35.8 \times 3.8$ ②	50.3	65.86	88.23
2.57	$90 \times 35 \times 1.4$ ①	50.1	68.34	92.69

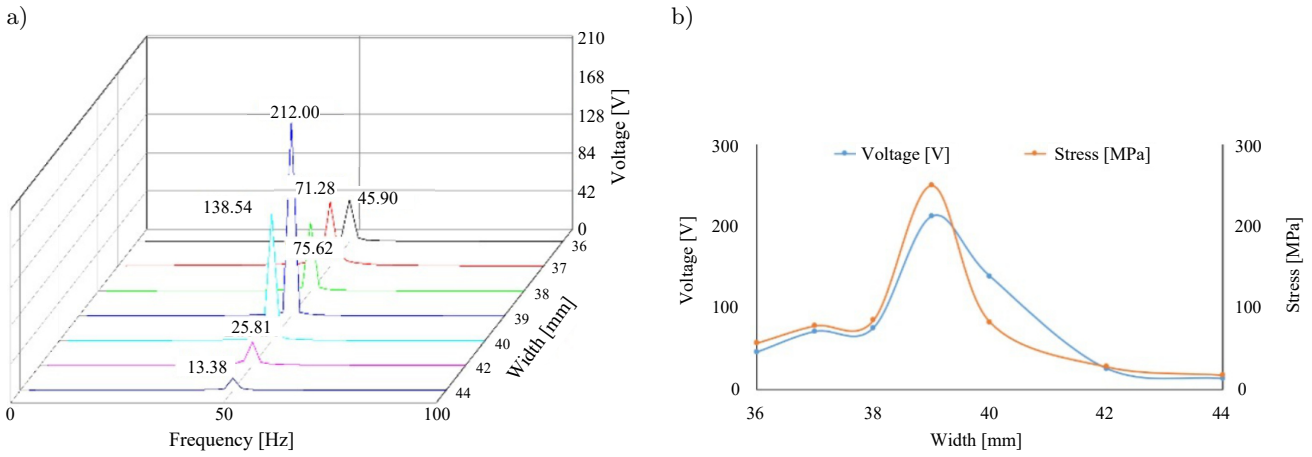


Fig. 7. a) Voltage-frequency response of different beam widths; b) voltage and stress induced by different beam widths.

Table 4. Dimensions and performance parameters of Fig. 7.

$L \times W \times t$ [mm]	Frequency [Hz]	Voltage [V]	Stress [MPa]
$84.6 \times 36 \times 4.3$	50.2	45.90	57.17
$85 \times 37 \times 4.3$ ④	50.3	71.28	78.28
$85.2 \times 38 \times 4.3$	50.3	75.62	85.43
$85.5 \times 39 \times 4.3$	50.1	212.00	249.71
$86.2 \times 40 \times 4.3$ ③	50.1	138.54	82.66
$86.6 \times 42 \times 4.3$	49.9	25.81	28.14
$89.8 \times 44 \times 4.3$	49.8	13.38	17.92

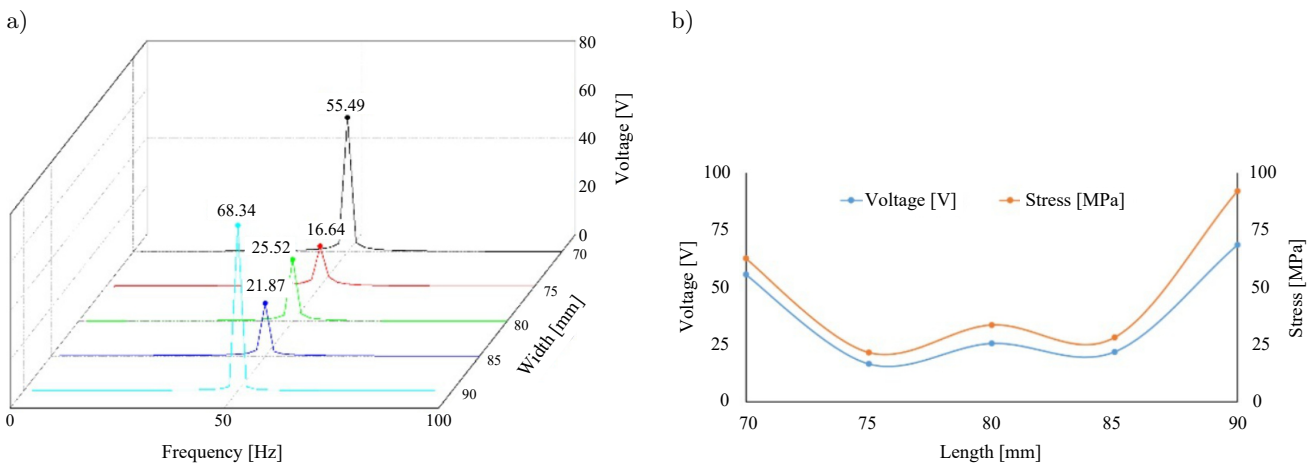


Fig. 8. a) Voltage-frequency response of different beam lengths; b) voltage and stress induced by different beam lengths.

Table 5. Dimensions and performance parameters of Fig. 8.

$L \times W \times t$ [mm]	Frequency [Hz]	Voltage [V]	Stress [MPa]
$90 \times 35 \times 1.4$	50.1	68.34	92.69
$85 \times 35 \times 1.9$	49.8	21.87	28.34
$80 \times 35 \times 3.2$	50.2	25.52	33.52
$75 \times 35 \times 4.8$	50.4	16.64	21.56
$70 \times 35 \times 5.9$ ⑤	50.1	55.49	62.51

The voltage-frequency responses of the designed PEHs with different beam widths are shown in Fig. 7a, and the variations in output voltage and stress versus beam width, using the same tip mass, are shown in Fig. 7b. The specific dimensions and related characteristics of the beam are shown in Table 4. From Table 4 and Fig. 7, we can conclude that as the length and width increase, both the voltage and stress initially grow and subsequently drop. The voltage is at its maximum when the width is 39 mm, while the stress is also relatively significant, even exceeding the material's yield limit of 108 MPa (WU, 2013). The output voltage is relatively high when the width is 40 mm, but the absolute value of the stress is lower than the voltage, which meets the PEH selection criteria. Finally, by using a fixed width of 35 mm (the minimum width, equal to the width of the PEH), with the beam length and mass height as variables, a set of PEHs has been investigated, as shown in the voltage-frequency responses in Fig. 8a, and the variations in voltage and stress versus beam length are shown in Fig. 8b. Specific dimensions and related characteristics are listed in Table 5. From these figures and table, the results indicate that with the increase of the beam length and the decrease in the mass height, both output voltage and stress initially decrease and then increase. When the beam is at its minimum length of 70 mm, the output characteristics are significantly influenced by the mass block; on the other hand, when the mass height is at its minimum of 1.4 mm, the output characteristics are primarily influenced by the beam length.

Based on the foregoing analysis, five beams, labeled ① to ⑤, are chosen for further fatigue life analysis. These beams are chosen based on the criteria of having a higher output voltage and a maximum stress smaller than the material's yield strength limit.

5.2. Prediction of fatigue life

The main methods for fatigue life analysis include the nominal stress method and local stress-strain methods (ZHANG *et al.*, 2014). The former involves estimating life based on the $S-N$ curve using fatigue cumulative damage theory, which is appropriate for high-cycle fatigue failures with cycles greater than 10^4 . The latter calculates the crack initiation life using the local stress-strain approach and predicts crack propagation life using fracture mechanics, making it suitable for low-cycle fatigue failures with cycles less than 10^4 (SALAZAR *et al.*, 2021). Both methods are based on the fatigue characteristics of materials and fatigue cumulative damage theory. Since piezoelectric transducers are subject to high-cycle fatigue, the nominal stress method approach is used in this study.

The fundamental principle of performing fatigue analysis using ANSYS software is to use FEA to obtain the stress and strain distribution of a structure under external periodic loads. These results are then combined with the material fatigue performance curve, and fatigue theory is applied to calculate the fatigue life distribution of the components. This process aids in predicting potential failures of structural components due to fatigue over long durations of operation. Stress-life curves for piezoelectric patch materials are not available in the ANSYS material library. Therefore, it is necessary to set up the fatigue properties of piezoelectric patch materials by generating stress-life curves using parameters provided by the manufacturer, such as elastic modulus and fatigue limit, which allows for the simulation and analysis of the fatigue behavior of piezoelectric patch transducer.

The analysis is conducted by applying the identical constant excitation equal to the initial resonant frequency to five PEHs. After obtaining the required parameters and calculating the stress distribution and fatigue life, the simulation results indicate that the maximum stress occurs at the root of the cantilever beam. Stress concentration and early fatigue failure occur at both ends of the cantilever beam, as illustrated in the fatigue life contour map for the piezoelectric patch layer in Fig. 9a and 9b. The blue areas represent the maximum fatigue life, while the red areas indicate the minimum fatigue life that the PEH can withstand. The yellow circles in Fig. 9c illustrate the specific locations on the PEH that are prone to fatigue damage, based on the simulation. This conclusion can

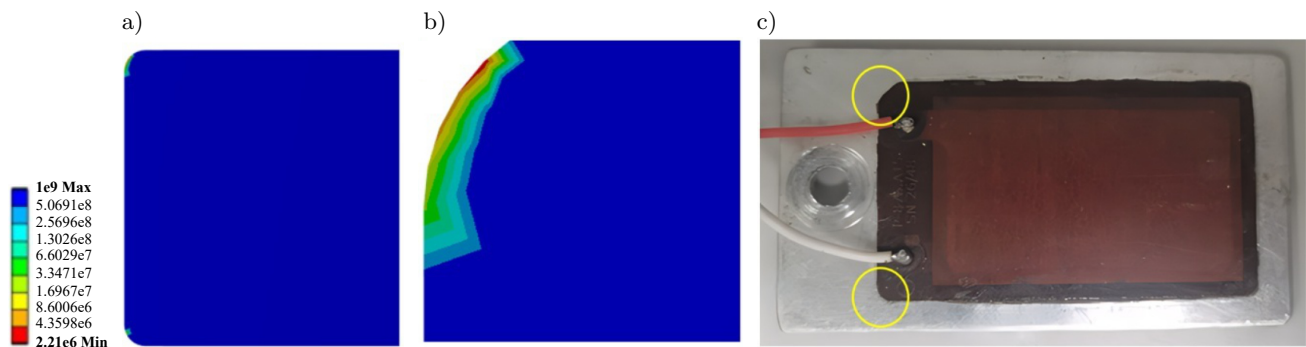


Fig. 9. a)–b) FE simulation results; c) actual PEH beam under study.

Table 6. Fatigue life of beams.

Serial number	Beam	Output voltage	Fatigue life
①	$90 \times 35 \times 1.4$	68.34	1.76×10^7
②	$88 \times 35.8 \times 3.8$	65.86	1×10^9
③	$86.2 \times 40 \times 4.3$	138.54	2.21×10^6
④	$85 \times 37 \times 4.3$	71.28	1.03×10^7
⑤	$70 \times 35 \times 5.9$	55.49	1×10^9

serve as a reference for future structural parameters and optimization studies.

For typical fatigue analysis, if a specimen does not fail after 10^7 stress cycles, it is assumed that the specimen will not fail even after an infinite number of stress cycles (WANG *et al.*, 2021). As shown in Table 6, it can be observed that under the same external loads, beams ①, ②, ④, and ⑤ have the fatigue life exceeding 10^7 cycles, indicating that fatigue failure is unlikely to occur. For a certain natural frequency of the beams, both output voltage and fatigue life are influenced by the combined effects of the mass block and the structural length-to-width ratio, and this influence is non-linear. In general, the larger the output voltage, the shorter the lifespan. Therefore, when designing a PEH for a specific frequency, it is critical to consider the coupling effects of the mass block and aspect ratio, while balancing the trade-off between output voltage and fatigue life to select the best solution that fits the actual vibration environment. In terms of the motor usage environment, that is for indoor electrical devices (such as machine tools), design schemes ① and ④ are recommended. They maximize output voltage while meeting the fatigue life requirements. For outdoor application, with an uncertain vibration environment (such as automotive), design scheme ② is recommended to ensure the cantilever beam has a sufficient safety margin to withstand extreme external loads. If space is limited, design scheme ⑤ can be considered. If the output power of the chosen beam is insufficient to meet certain requirements, increasing the surface area of the piezoelectric layer can be an alternative solution. Option ③ is suitable for situations with high output power demands and a low requirement for safety margin.

6. Conclusion

Long-term vibration loading at a constant amplitude of excitation in a PEH can induce fatigue damage at the root of the cantilever beam in the long run. In this study, three sets of PEHs with distinct fixed and variable values were designed for an electric motor with a 50 Hz vibration frequency. After comprehensively considering the output voltage and fatigue life of the piezoelectric cantilever beam, suitable design solutions for various environments and operational conditions were investigated. The conclusions are summarized as follows:

- The dimensions of the beam and the weight of mass will result in the optimal solution for output voltage and fatigue life at the maximum coupling effect position for a PEH with specific natural frequency of 50 Hz. When designing the structure of the PEH beam, output voltage should not be the sole design objective; instead, comprehensive consideration of both the output voltage and the fatigue life of the piezoelectric layer is necessary.
- Five optimized design schemes with a frequency of 50 Hz were selected in this study, and these are recommended for application in different operational environments and for varying design purposes.
- The structural parametric study of the PEH with a specific natural frequency of 50 Hz, along with the determination and analysis of the optimal scheme, provides a simple and effective method for the structural optimization design of PEH beams.

Based on these conclusions, it is necessary to conduct further research to determine the optimal parameters for the elastic beam in PEH. The fundamental correlations between the parameters and structural properties, as obtained in this study, are shown in Table 7, which will be useful as a reference for future PEH design research.

Table 7. Parameters correlation analysis in designing PEH.

Parameters	Length	Width	Mass
Relationship with structural frequency	Negative correlation	Positive correlation	Negative correlation
Relationship with output voltage	Positive correlation	Negative correlation	Positive correlation

FUNDING

The authors would like to thank the Ministry of Higher Education in Malaysia for the Fundamental Research Grant Scheme (FRGS) under project code FRGS/1/2021/TK0/USM/03/6, and the Science and Technology Research Program of Chongqing Municipal Education Commission (grant no. KJQN 202404004).

References

1. ABDELKEFI A., BARSALLO N., TANG L., YANG Y., HAJJ M.R. (2014), Modeling, validation, and performance of low-frequency piezoelectric energy harvesters,

- Journal of Intelligent Material Systems and Structures*, **25**(12): 1429–1444, <https://doi.org/10.1177/1045389X13507638>.
2. ABDUL SATAR M.H., MURAD A.F., AHMAD MAZLAN A.Z. (2022), Characterization of piezoelectric patch material with hysteresis, saturation, creep, and vibration nonlinearity effects and its application to the active vibration suppression for cantilever beam, *Journal of Vibration and Control*, **28**(3–4): 476–489, <https://doi.org/10.1177/1077546320980571>.
 3. AUGUSTYN E., KOZIEŃ M.S., PRĄCIK M. (2014), FEM analysis of active reduction of torsional vibrations of clamped-free beam by piezoelectric elements for separated modes, *Archives of Acoustics*, **39**(4): 639–644, <https://doi.org/10.2478/aoa-2014-0069>.
 4. AVVARI P.V., YANG Y., SOH C.K. (2017), Long-term fatigue behavior of a cantilever piezoelectric energy harvester, *Journal of Intelligent Material Systems and Structures*, **28**(9): 1188–1210, <https://doi.org/10.1177/1045389X16667552>.
 5. BAO B., WANG Q., WU N., ZHOU S. (2021), Hand-held piezoelectric energy harvesting structure: Design, dynamic analysis, and experimental validation, *Measurement*, **174**: 109011, <https://doi.org/10.1016/j.measurement.2021.109011>.
 6. CHEN C., SHARAFI A., SUN J.Q. (2020), A high density piezoelectric energy harvesting device from highway traffic – Design analysis and laboratory validation, *Applied Energy*, **269**: 115073, <https://doi.org/10.1016/j.apenergy.2020.115073>.
 7. FENG Y., LIU W., GAO M., CHEN H., LU Q. (2023), Analysis of electromechanical response and fatigue life of piezoelectric cantilever beam [in Chinese], *Failure Analysis and Prevention*, **18**(3): 184–190, <http://doi.org/10.3969/j.issn.1673-6214.2023.03.007>.
 8. Gaomeng (n.d.), Shanghai GOM Testing & Technical Co., Ltd., <https://www.gaomengce.com> (access: 1.05.2024).
 9. LI T., LEE P.S. (2022), Piezoelectric energy harvesting technology: From materials, structures, to applications, *Small Structures*, **3**(3): 2100128, <https://doi.org/10.1002/sstr.202100128>.
 10. LU Q. (2018), *Structure design and electro-mechanical performance analysis of vibration piezoelectric composite energy harvester*, Ph.D. Thesis, Harbin Institute of Technology.
 11. NIASAR E.H.A., DAHMARDEH M., GOOGARCHIN H.S. (2020), Optimization of a piezoelectric energy harvester considering electrical fatigue, *Journal of Intelligent Material Systems and Structures*, **31**(12): 1443–1454, <https://doi.org/10.1177/1045389X20923086>.
 12. PANDA S. *et al.* (2022), Piezoelectric energy harvesting systems for biomedical applications, *Nano Energy*, **100**: 107514, <https://doi.org/10.1016/j.nanoen.2022.107514>.
 13. RAFIQUE S., BONELLO P. (2010), Experimental validation of a distributed parameter piezoelectric bimorph cantilever energy harvester, *Smart Materials and Structures*, **19**(9): 094008, <https://doi.org/10.1088/0964-1726/19/9/094008>.
 14. ROUNDY S. *et al.* (2005), Improving power output for vibration-based energy scavengers, *IEEE Pervasive Computing*, **4**(1): 28–36, <https://doi.org/10.1109/MPRV.2005.14>.
 15. SALAZAR R., LARKIN K., ABDELKEFI A. (2021), Piezoelectric property degradation and cracking impacts on the lifetime performance of energy harvesters, *Mechanical Systems and Signal Processing*, **156**: 107697, <https://doi.org/10.1016/j.ymssp.2021.107697>.
 16. SEZER N., KOÇ M. (2021), A comprehensive review on the state-of-the-art of piezoelectric energy harvesting, *Nano Energy*, **80**: 105567, <https://doi.org/10.1016/j.nanoen.2020.105567>.
 17. SHI Y., HALLETT S.R., ZHU M. (2017), Energy harvesting behaviour for aircraft composites structures using macro-fibre composite: Part I – Integration and experiment, *Composite Structures*, **160**: 1279–1286, <https://doi.org/10.1016/j.compstruct.2016.11.037>.
 18. WANG J., QIN X., LIU Z., DING G., CAI G. (2021), Experimental study on fatigue degradation of piezoelectric energy harvesters under equivalent traffic load conditions, *International Journal of Fatigue*, **150**: 106320, <https://doi.org/10.1016/j.ijfatigue.2021.106320>.
 19. WANG S.L., SONG J., ZHAO X.Y. (2019), Research on energy recovery device of electric vehicle seat based on piezoelectric effect, [in:] *2019 13th Symposium on Piezoelectricity, Acoustic Waves and Device Applications (SPAWDA)*, <https://doi.org/10.1109/SPAWDA.2019.8681810>.
 20. WU X. (2013), *The fatigue analysis and experimental research of piezoelectric pump with single chip vibrator*, Master Thesis, University of Jilin.
 21. ZHANG M., MENG Q., WANG H. (2014), Fatigue analysis for cantilever piezoelectric vibration energy harvester.
 22. ZHANG Q., LIU Z., JIANG X., PENG Y., ZHU C., LI Z. (2022), Experimental investigation on performance improvement of cantilever piezoelectric energy harvesters via escapement mechanism from extremely low-frequency excitations, *Sustainable Energy Technologies and Assessments*, **53**(Part B): 102591, <https://doi.org/10.1016/j.seta.2022.102591>.
 23. ZHU M., WORTHINGTON E., TIWARI A. (2010), Design study of piezoelectric energy-harvesting devices for generation of higher electrical power using a coupled piezoelectric-circuit finite element method, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **57**(2): 427–437, <https://doi.org/10.1109/TUFFC.2010.1423>.

Research Paper

Aeroacoustic Effect of Savonius Rotor Segmentation

Shivangi SACHAR^{(1)*}, Piotr DOERFFER⁽¹⁾, Paweł FLASZYŃSKI⁽¹⁾, Józef KOTUS^{(2)a},
Krzysztof DOERFFER⁽²⁾, Joanna GRZELAK⁽²⁾, Michał PIOTROWICZ⁽¹⁾

⁽¹⁾ *Institute of Fluid Flow Machinery, Polish Academy of Sciences*
Gdańsk, Poland

⁽²⁾ *Gdańsk University of Technology*

^a *Faculty of Electronics, Telecommunications and Informatics, Audio Acoustics Laboratory*
Gdańsk, Poland

*Corresponding Author e-mail: shivangi_sachar@hotmail.com

Received February 24, 2025; revised April 10, 2025; accepted April 10, 2025;
published online June 6, 2025.

Switching to renewable energy has been accelerated in recent decades due to the depleting fossil fuel reserves and the need to mitigate environmental and climate degradation. Wind power, especially in urban areas, has seen a significant growth. A critical consideration in the urban wind turbine installation is the noise impact on residents. This study investigates the noise generated by wind turbines under different operational conditions, comparing single-segment and five-segment rotor designs. Various acoustic analyses were conducted, including broadband analysis with weighting curves Z, A, C, and G, a narrowband analysis using $1/12$ octave bands, and broadband calculations of sound quality indicators such as sharpness, roughness, and fluctuation strength (FS). The FS was also examined in the Bark scale frequency domain. The study linked the acoustic analysis with the rotor efficiency related to power production. The findings indicate that five-segment rotors generate less acoustic energy due to phase shifts, enhancing dissipation rates, and acoustic energy decreases with the increasing load, peaking when rotors are free at high revolutions per minute (RPM). While single-segment rotors show higher efficiency, they produce more noise. In contrast, five-segment rotors offer a better sound quality, making them preferable despite a lower efficiency. This research provides essential insights into designing urban wind turbines that balance efficiency and noise, crucial for sustainable energy solutions.

Keywords: wind power; wind turbines; rotor efficiency; Savonius rotor segmentation.



Copyright © 2025 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Noise from wind turbines is a significant concern in energy production, linked to various health issues including irritation, elevated blood pressure, and sleep disturbances (ABBASI *et al.*, 2019; ANJUM, KUMARI, 2022; ZARE *et al.*, 2020; KOTUS, KOSTEK, 2008). High-frequency noise can cause headaches, fatigue, and immune system suppression (MÜNZEL *et al.*, 2018; SZYCHOWSKA *et al.*, 2018; ANJUM, KUMARI, 2022), while residents near wind farms often experience annoyance, impacting their daily tasks (PAWLACZYK-ŁUSZCZYŃSKA *et al.*, 2014; HAFKE-DYS *et al.*, 2016). Previous research indicates that individuals within a 500 m radius of wind energy plants exhibit significant responses to the turbine noise, with

annoyance extending to 1900 m (VAN DEN BERG, 2004). However, further investigations are needed due to limited empirical data linking public annoyance directly to the wind turbine noise.

Wind turbine noise can be classified based on frequency as tonal or broadband, originating from aerodynamic or mechanical sources (HANSEN, HANSEN, 2020). Environmental factors such as temperature, humidity, and obstacles can affect noise propagation, with the turbine blade motion particularly noticeable at night (DESHMUKH *et al.*, 2019; NGUYEN *et al.*, 2020). To mitigate noise pollution, governmental guidelines specify maximum permissible noise levels, necessitating adherence to local factors during turbine deployment (DAVY *et al.*, 2018; GALLO *et al.*, 2016).

Horizontal axis wind turbines (HAWTs) dominate wind energy extraction but face limitations such as size and maintenance requirements (DAVY *et al.*, 2018; GALLO *et al.*, 2016). In contrast, vertical axis wind turbines (VAWTs) offer reduced mechanical noise transmission and diminished aerodynamic noise levels (MÖLLERSTRÖM *et al.*, 2014; GRAHAM, PEARSON, 2022). However, research on VAWTs remains somewhat limited, prompting a need for comprehensive evaluations to optimize efficiency while minimizing noise disruption (DUMITRESCU *et al.*, 2010; IIDA *et al.*, 2004).

The Savonius VAWT (SVAWT) emerges as a promising option due to its reduced noise emission and suitability for urban environments (AKWA *et al.*, 2012; DOERFFER *et al.*, 2021). Despite producing less noise than HAWTs, the trade-off for noise reduction leads to a decreased energy output (OERLEMANS, FUGLSANG, 2012; GHASEMIAN, NEJAT, 2015). Enhancements such as rotor segmentation have been proposed to improve efficiency (KACPRZAK *et al.*, 2013; MAHMOUD *et al.*, 2012; SUN *et al.*, 2012; DRISS *et al.*, 2014).

Research in wind energy has shifted towards urban integration and regulatory compliance, driven by environmental concerns (New World Wind, n.d.; IBIS Power, n.d.). Nevertheless, a detailed analysis of the VAWT power production and acoustics remains lacking (AKWA *et al.*, 2012; SACHAR *et al.*, 2023; JEONG *et al.*, 2014). Thus, evaluating Savonius VAWTs for urban environments, considering noise-efficiency trade-offs, is crucial for public acceptance (New World Wind, n.d.; IBIS Power, n.d.). This study aims to address this gap by conducting experiments on a standard Savonius rotor, exploring the impact of segmentation and assessing the sound quality (FASTL *et al.*, 2007).

2. Wind turbine model and experimental setup

Various configurations of the Savonius rotor were devised and tested for the research detailed in this paper, with wind velocities ranging from 7 m/s to 12 m/s. The experimentation was conducted within a dedicated test section, depicted in Fig. 1, established at the Institute of Fluid Flow Machinery Polish Academy of Sciences (IMP PAN). This facility comprises three inlet fans capable of operating at adjustable speeds. A wind

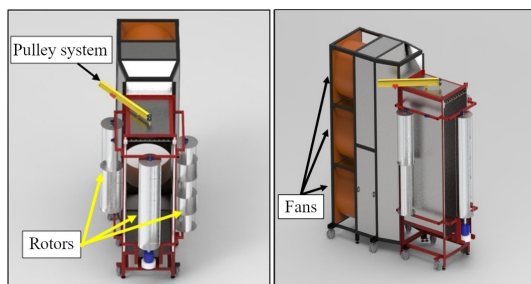


Fig. 1. Test section used in experiments.

velocity of 8.5 m/s was maintained for the investigations reported herein.

The calibration of the test stand posed a challenge in achieving uniformity in the outlet stream. Three streams from propulsion fans were uniformly distributed by changing the local density of meshes at the screen between fans and the test section. An adequate location of one-, two-, and three-layers of meshes allowed obtaining a satisfactory uniformity of the outlet stream. On top of that a honeycomb straightener was inserted to reduce the transversal fluctuations in the air stream. The velocity was measured at the measurement points forming a grid with a spacing of 100 mm to confirm the uniformity of the outlet stream. This was obtained with the help of a support line holding nine Prandtl probes across the stream as shown in Fig. 3. It was measured that the standard deviation of the outlet velocity was around 10%. This uniformity of the incoming stream was satisfactory in such a simple test stand. As shown in Fig. 3, a Prandtl probe is mounted in the middle of the outlet box to provide a reference stream velocity.

The test section can operate one rotor at any given time, while it can accommodate the simultaneous mounting of three distinct rotors for storage purposes (Fig. 1). Additionally, a pulley system facilitates the manipulation of these rotors. The three rotor configurations utilized in this study include single-, double-, and five-segment variants, each with a height of 2.2 m and a diameter (D) of 0.5 m. An integral feature of the test section is its loading apparatus, enabling the modification of the load exerted on the tested rotors. This functionality allows for a comparative analysis under various loading conditions, with concurrent noise measurements captured via microphones. Specifically, this study focuses on a comparison between single and five-segment rotor configurations. The single-segment rotor mirrors the conventional Savonius rotor design, comprising ‘S’-shaped buckets spanning between two end plates at a height denoted as ‘ H ’. In contrast, the five-segment rotor configuration consists of five rotors of identical height stacked atop one another, each rotated at 90° relative to the preceding segment orientation. This arrangement is illustrated in Fig. 2.

The measurement scenario consisted of quantifying the performance of the two wind turbines at their complete operational TSR range. The noise emissions were mainly focused on four cases:

- stopped rotor at different wind speeds;
- no-load condition/free run: the load is being withdrawn and the rotor is at its maximum speed of rotation at any given wind speed;
- loaded condition: rotor working under different load conditions;
- the background noise consisted only of the test section in the lab (without a rotor).

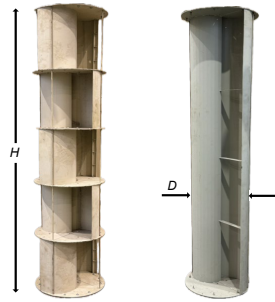


Fig. 2. a) Five-segment rotor; b) single-segment rotor.

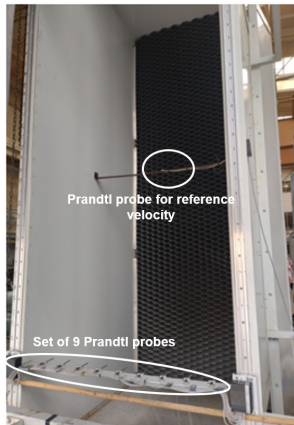


Fig. 3. Test section stream outlet.

The experiment involved both rotors, with a microphone positioned as depicted in Fig. 4. The microphone location is indicated by the red ellipse. It is placed at the same location for all the tests carried out. The microphone is located in the middle of the test section outlet in the vertical direction. In the horizontal direction, it is placed outside the stream leaving the test section. It is at a distance of 0.28 m away from the test section outlet plane, making sure that the microphone is sheltered from the air jet.

At the jet outlet the static pressure is the same as the ambient pressure, therefore one may assume that the free shear layer generated downstream of the test section wall should follow the wall direction. Such a free shear layer starts at the test section outlet and expands with the angle of $\pm 7^\circ$ due to turbulent mixing. This is indicated in Fig. 5a showing that the shear layer is far away from the microphone location.

The presence of the rotor close to the outlet plane may impose further deflection of this shear layer. In order to make sure that the rotor does not enforce the interaction of the shear layer with microphone, the numerical simulations for the investigated configuration have been carried out. This simulations have been done for the two limiting cases: no-load (TSR = 1.8) and maximum-load (TSR = 0.9).

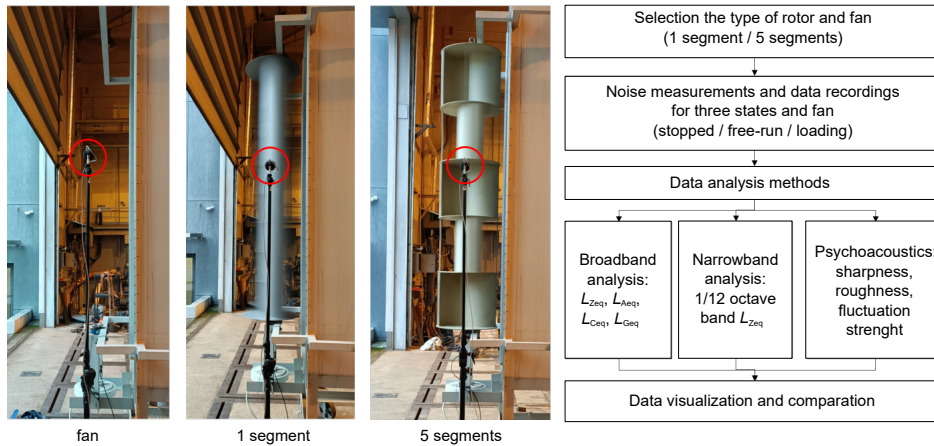


Fig. 4. Noise measurement setup and block diagram of the measurement procedure.

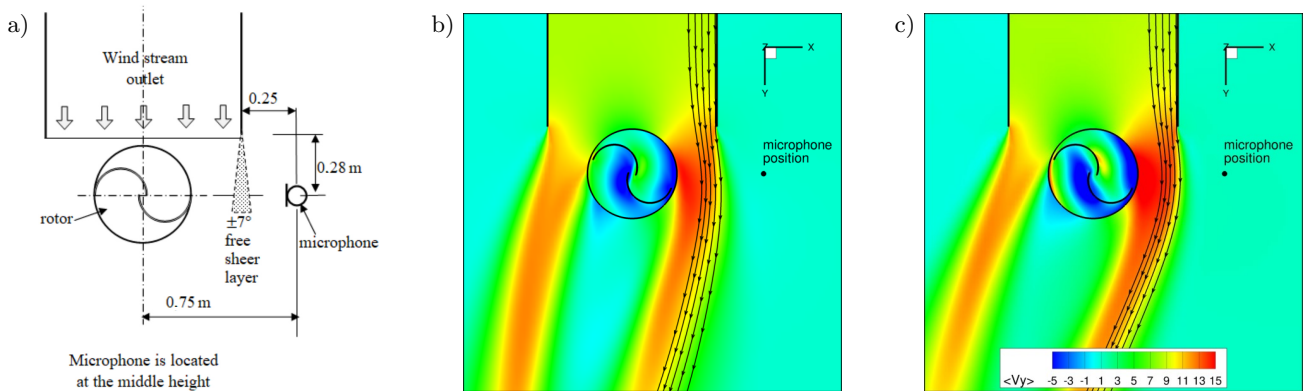


Fig. 5. Acoustic measurement setup: a) test section outlet configuration; b) CFD of the outlet flow TSR = 0.9; c) CFD of the outlet flow TSR = 1.8.

Numerical simulations were performed using the FINE/Open NUMECA/CADENCE software, which employs a finite volume method with a central differencing scheme enhanced by artificial dissipation for spatial discretization, and a Runge–Kutta type time integration scheme. The time step in each case corresponds to a rotor rotation of 2° . A full multigrid strategy was applied to enhance the convergence rate. The calculations were carried out under the assumption of a two-dimensional, incompressible, and turbulent flow (at the mid-span of the wind tunnel), using the $k-\omega$ SST turbulence model. Boundary conditions were defined based on the values measured on the experimental setup. To allow for the full development of the flow, simulations were conducted over 20 full revolutions. The computations were then continued for an additional 10 revolutions, over which the results were time-averaged. The results of the numerical simulations, presented in Figs. 5b and 5c, were obtained based on experience with unsteady flow configurations including aeroacoustic effects (FLASZYNSKI *et al.*, 2021; GRZELAK *et al.*, 2024; SURESH *et al.*, 2024). In Figs. 5b and 5c, the time-averaged velocity (over 10 rotations) is shown. These results indicate that the free shear layer developing downstream of the wind tunnel wall is far from the microphone location. The induced air flow around the microphone is of very low velocity, confirming that the jet downstream of the wind tunnel does not affect microphone recordings.

The objective of the tests was to assess the overall noise generated by the rotor under various loading conditions. Subsequently, the same procedure was repeated, with each rotor type subjected to identical operating conditions. Noise measurements were taken using a sampling method, employing a class 1 Sound Level Meter SVAN 979 with a high quality omnidirectional G.R.A.S. $1/2''$ measuring prepolarized free-field microphone, type 40AE and the SV 17 preamplifier. The sound level meter was calibrated before and after the measurements using the SV 36 acoustic calibrator. The raw sound data was recorded during the noise measurements for further analysis using the SVANPC++ software, version 3.4.4. Measurements were repeated three times for each case $N = 3$, with each sample recorded for T equal to 60 s. The measuring microphone was positioned using a tripod, halfway up the rotor, outside the mainstream from the test section exit, at a distance of 0.7 m from the rotor axis. A windscreens was used for the microphone.

Simultaneously, measurements were taken to determine the rotor performance characteristics and establish the relationship between the coefficient of power (C_p), TSR, and revolutions per minute (RPM). These experiments were conducted under conditions consistent with the noise measurements, with the fan frequency set to 35 Hz to produce a wind velocity of 8.5 m/s.

It is essential to note that the relative noise production discussed in this study between single and five-segment rotors does not fully represent the noise levels produced by rotors individually in real-world conditions, where factors such as inlet turbulence, temperature, humidity, and shear may significantly influence the noise levels.

The recorded acoustic data was analyzed using several methods, as shown in Fig. 4. First, the broadband analysis was performed. In this case the equivalent continuous sound pressure level (L_{eq}) for four weightings curves (Z, A, C, and G) was determined. Next, the analysis of the sound pressure level (SPL) was extended using a narrow band analysis. This analysis was performed using a $1/12$ octave band analyzer. The $1/12$ octave band equivalent SPL with the weighting curve Z was determined to evaluate the actual noise production between the considered rotors. The main aim of this analysis was to evaluate the actual noise production between the considered rotors. Those analyses were performed using the SVANPC++ software, version 3.4.4 and the Pulse Reflex software. The L_{eq} , for a given measurement period (T), was calculated using the selected weighting filter: Z, A, C, or G.

The indicator $L_{Z_{eq,Avg.}}(f)$ was computed for each $1/12$ octave band spectrum, and the average value was calculated as follows:

$$L_{Z_{eq,Avg.}}(f) = 10 \log_{10} \left(\frac{1}{N} \sum_{i=1}^N 10^{L_{Z_{eq,T}(f)/10}} \right) [\text{dB}], \quad (1)$$

where f is the frequency of the $1/12$ octave band, 180 filters were used for the frequency range from 0.71 Hz up to 22000 Hz, N is the number of recorded samples.

The third method was a psychoacoustic analysis. In this case selected metrics related to the sound quality such as sharpness, roughness, and fluctuation strength (FS) were determined. Sharpness is a hearing sensation related to frequency and independent of loudness. Sharpness corresponds to the sensation of a sharp, painful, high-frequency sound and is a comparison of the amount of high frequency energy to the total energy (FASTL, ZWICKER, 2007). Roughness is a complex effect which quantifies the subjective perception of the rapid (15 Hz–300 Hz) amplitude modulation of a sound. The unit of measure is *asper*. One *asper* is defined as the roughness produced by a 1000 Hz tone of 60 dB which is a 100% amplitude modulated at 70 Hz (FASTL, ZWICKER, 2007; COX, n.d.). The FS is similar in principle to roughness except that it quantifies a subjective perception of a slower (up to 20 Hz) amplitude modulation of a sound. The sensation of the FS persists up to 20 Hz, then at this point, the sensation of roughness takes over (FASTL, ZWICKER, 2007; COX, n.d.). The main purpose of this analysis was to extend the knowledge about phenomena related to the perception of the sound emitted by the examined rotors by human beings. A psychoacoustic

analysis was conducted using the Pulse Reflex software, version 21.0.0.567, provided by Bruel & Kjaer (FASTL, ZWICKER, 2007).

For all types of the performed analysis, three independent recordings (independently for each rotor type and considered conditions), with T set to 60 s, were analyzed. Finally, the average value was calculated and presented.

3. Results and analysis

The findings from experiments conducted on a single- and five-segment rotor installed in the test section, as illustrated in Fig. 4, and operated at a wind speed of 8.5 m/s, are depicted in Fig. 6 (see Subsec. 3.2). This wind velocity, generated by the fans, corresponds to a frequency of 35 Hz and is denoted by a dashed vertical line in the plots. It is essential to acknowledge that the acoustic emissions from the fans were notably high. Therefore, the comparison between the two rotors should focus solely on the difference in the SPL, rather than absolute values.

The noise measurements were conducted under the three following conditions:

- rotor under the maximum-loading condition, where the single-segment rotor operated at 300 RPM (equivalent to a frequency of 5 Hz), and the five-segment rotor operated at 250 RPM (equivalent to a frequency of 4.17 Hz);
- rotor without loading, with the five-segment rotor operating at 500 RPM (equivalent to a frequency of approximately 8.5 Hz) and the single-segment rotor operating at 575 RPM (equivalent to a frequency of 9.55 Hz);
- test section with the rotor in a stationary (non-rotating) position.

The aforementioned conditions were applied to:

- the test section with the five-segment rotor;
- the test section with the single-segment rotor;
- the test section noise without a rotor (fan noise).

The obtained measurement data was used to perform 4 types of analysis. First, a broadband analysis was performed.

The effect of the rotor presence and the background noise of the test stand, generated mainly by the propulsion fans, is presented in Fig. 6a. The rotor type plays a minor role on the generated noise. Nevertheless, its

implementation contributes significantly to the noise generated by the propulsion fans. However, this contribution takes place below the frequency of $f < 130$ Hz only.

The detailed results are shown in Subsec. 3.1. Next, the narrowband analysis is detailed in Subsec. 3.2. In Subsec. 3.3., the acoustic analysis was extended by the selected sound quality measures. Finally, a relation between the noise produced and the rotor efficiency is detailed in Subsec. 3.4.

3.1. Broadband analysis

The main aim of the broadband analysis was to show the total level of noise using three types of frequency weighting curves: A, C, Z. For better understanding of the considered phenomena, the broadband analysis was extended by a low frequency analysis, including infrasound, using the weighting curve G. The average equivalent SPL was calculated for this purpose for all the considered cases: stopped rotor, rotor with load (maximum-loading conditions), rotor without load (free-run). The fan noise (the noise level without a rotor) was used as a reference for the cases discussed, as seen in Fig. 6. The average values were calculated using three samples of equivalent sound pressure levels. Each sample value was calculated for 60-second sound recordings. The obtained results are shown in Table 1.

The detailed assessment of the noise level emitted by a rotor in the presented measuring setup is difficult for the following reasons: the used fan system produced a very high level of background noise; the noise emission from the rotating rotor is closely related to the wind speed; and the rotor is not a wind-independent source of the noise. For these reasons, the traditional method of evaluating acoustic emissions – measuring the noise immision and the acoustic background, then calculating the noise emission as the energy difference between them – is not appropriate. In the measurement system used, the rotor acts as a background noise modulator and, in selected situations, causes an increase in the noise level. The way in which the background noise is modulated is directly related to the design of the rotor and the way of its rotation, depending on the load. Taking this fact into account, it was proposed to assess the rotor noise using simply the difference between the noise level with the rotor (in a particular state) and the fan

Table 1. Comparison of sound levels for one-segment and five-segment rotors.

Measurement case	One-segment rotor				Five-segment rotor			
	L_{Aeq}	L_{Ceq}	L_{Zeq}	L_{Geq}	L_{Aeq}	L_{Ceq}	L_{Zeq}	L_{Geq}
Fan noise	95.7	101.9	102.0	82.8	95.7	101.9	102.0	82.8
Stopped rotor	96.2	102.1	102.4	88.2	95.7	101.9	102.2	89.8
Free run	99.0	105.3	109.3	117.1	95.7	102.3	103.3	98.5
Rotor with load	96.1	103.0	111.7	110.0	95.7	101.9	104.1	96.7

noise level. Based on this assumption, the difference levels were calculated for both the considered types of rotors. These calculations were done for two cases.

First, the difference between the emission and the background noise levels was determined. The result was marked using symbol Δ . The obtained results are shown in Table 2. Next, the difference between the rotors was calculated. The one-segment rotor values were used as a reference for this purpose. The results are shown in Table 3. The noise of the stopped rotor results from a static disturbance of the airflow. The noise level exceeding the background acoustic level results from the turbulence introduced by the rotor structure in the air stream because the rotor is stopped. For the single-segment rotor, higher noise levels were recorded for each indicator type (see Table 1). For the free-run mode for the single-segment rotor, the noise level was 3 dB higher compared to the acoustic background for curves A and C and by 7.3 dB for curve Z (see Table 2). For a five-segment rotor, the change in the acoustic background level is very small. An increase in the background noise level of 2.1 dB was noticeable only for the load mode and for the curve Z analysis. The greatest differences for the curve G were noticed between the fan noise and the added rotor. Placing the stopped rotor in the air flow increased the noise level by over 5 dB for the single-segment rotor and by over 7 dB for the five-segment rotor. This situation can be noticed in figures presented in Subsec. 3.2. In the free-run conditions for a single-segment rotor, the L_{Geq} level increased by over 34 dB compared to the level of the fan itself. The noise level for the five-segment rotor in the free-run conditions was 18.6 dB lower than for the single-segment rotor. In the load conditions, the noise level for the single-segment rotor decreased by 7.1 dB compared to the free-run conditions. In the case of the five-segment rotor, the difference between the noise level for the free-run and the load conditions was 1.8 dB. The noise level for the single-segment rotor in the load phase was over 13 dB higher compared to the noise level generated by the five-segment rotor.

An additional analysis involved determining the differences in the noise level between the single-segment rotor and the five-segment rotor. Even when the rotors were not rotating, the noise level was (slightly) higher for the single-segment rotor than for the five-segment rotor. The noise level for the single-segment rotor in the free-run mode was 3 dB higher than the acoustic background (for curves A and C) and

6 dB for curve Z. The difference between curves A, C, and Z is affected by the strong attenuation introduced by curves A and C for low frequency. The noise level for a single-segment rotor in the load mode was slightly higher for A- and C-corrected indicators and 7.6 dB higher for the curve Z. This means that the differences between the rotors occur mainly in the low-frequency range, with higher levels recorded for the single-segment rotor. Based on the research, it has been found that a single-segment rotor causes a greater increase in the background noise level than a five-segment rotor, and a single-segment rotor causes a measurable increase in the acoustic background level, especially for the free run mode and weighting curves A and C. A loaded single-segment rotor causes only a slight increase in the acoustic background level for weighting curves A and C. The greatest increase in the noise level was recorded for the correction curve Z, especially for the loaded single-segment rotor. The five-segment rotor did not cause any significant change in the acoustic background level for the weighting curves A and C. A slight increase in the level was noted only for the load mode and the curve Z. In practice, this means that a five-segment rotor will be less noticeable (in the acoustic field) than a single-segment rotor. Its design ensures that the acoustic background level does not change during its operation.

Based on the broadband analysis, it was found for the curve G that the greatest differences in the noise emission, regardless of the rotor type, occurred for the indicator corrected by the weighting curve G. This clearly indicates a low-frequency type of emission. More detailed considerations regarding the distribution of acoustic energy in the frequency domain are presented in Subsec. 3.2. Quantitatively, it has been shown that low-frequency noise emissions, expressed by the indicator L_{Geq} , for a single-segment rotor are many times higher than for a five-segment rotor. The single-segment rotor in the free-run phase, under the given test conditions, emitted energy nearly 73 times higher than the five-segment rotor. In the load conditions, the acoustic energy emitted by the single-segment rotor was 21 times higher than the acoustic energy emitted by the five-segment rotor. Therefore, the working conditions had a significant impact on the level of emissions. Low-frequency emissions were the highest in the free-run phase. The results obtained have considerable practical significance. First of all, they indicate that the problem of noise emissions con-

Table 2. Noise difference analysis results calculated in relation to fan noise.

Measurement case	One-segment rotor				Five-segment rotor			
	ΔL_{Aeq}	ΔL_{Ceq}	ΔL_{Zeq}	ΔL_{Geq}	ΔL_{Aeq}	ΔL_{Ceq}	ΔL_{Zeq}	ΔL_{Geq}
Stopped rotor	0.6	0.2	0.4	5.4	0.0	0.0	0.2	7.1
Free-run	3.3	3.4	7.3	34.3	0.0	0.4	1.3	15.7
Rotor with load	0.4	1.1	9.7	27.2	0.0	0.0	2.1	13.9

cerns low frequencies. The design of the rotor significantly affects noise emissions. The five-segment rotor is characterized by significantly lower acoustic energy emissions, regardless of the operating conditions. The biggest differences occurred in the free-run phase. To extend the analysis, calculations were performed using sound quality measures. Indicators sensitive to signal modulation, such as roughness and the FS, are of particular importance. The results of this study are shown in Subsec. 3.3.

Table 3. Noise difference analysis between rotors*.

Measurement case	Indicator type			
	ΔL_{Aeq}	ΔL_{Ceq}	ΔL_{Zeq}	ΔL_{Geq}
Stopped rotor	0.6	0.2	0.2	-1.6
Free-run	3.3	3.1	6.0	18.6
Rotor with load	0.4	1.1	7.6	13.3

*positive value: the single-segment rotor produced a greater SPL; negative value: the five-segment rotor produced a greater SPL.

3.2. Narrowband analysis

In this section, the narrowband analysis results are presented. As mentioned before, the narrowband analysis was performed using a constant percentage bandwidth analyzer with a $1/12$ octave band resolution. The results are presented in two groups of figures: first separated for the considered state (Fig. 6), second separated for the considered rotors (Fig. 7).

As depicted in Fig. 6a, both rotors exhibit similar noise level outputs when the rotor is stationary, allowing for a direct comparison under both no-load and loaded conditions (Figs. 6b and 6c). Notably, the angular position of the stopped rotor relative to the wind direction did not influence the results. The RPM of the rotors were standardized as the reference value for each scenario, with the corresponding frequencies and recorded sound energy levels [dB].

Distinct differences in the acoustic energy distribution among the considered rotors are observable, particularly at low frequencies, notably between 1 Hz and 51 Hz, as illustrated in Figs. 6b and 6c. A significant disparity in both frequency and acoustic energy production within this range is evident (as was mentioned in the previous section, especially for the weighting curve G). Consequently, the collective SPL for this frequency band was computed. An analysis of the entire frequency spectrum reveals discrepancies not only in the low-frequency range but also around 1000 Hz. Specifically, the single-segment rotor generated higher sound levels than the five-segment rotor, particularly under no-load conditions (Fig. 6c). For loaded conditions, differences are primarily observed in the low-frequency range, leading to analysis limited to frequencies between 1 Hz to 51 Hz (Fig. 6b).

The initial peak corresponds to a weak impulse frequency of approximately 4.8 Hz when maximum-

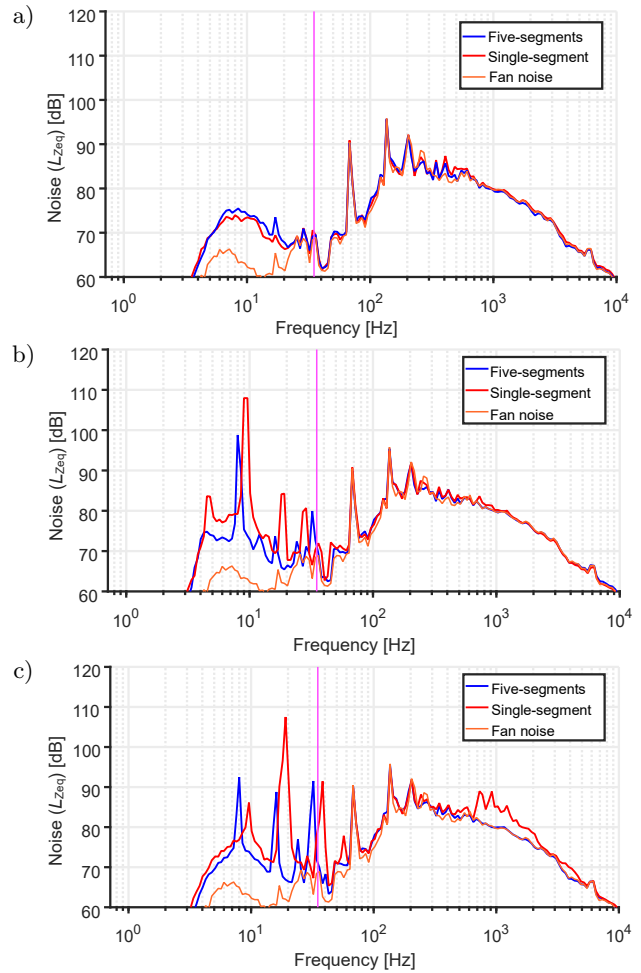


Fig. 6. $1/12$ octave bands spectrum for examined rotors presented separately for considered conditions: a) stopped rotor; b) rotor with load; c) rotor without load (free-run).

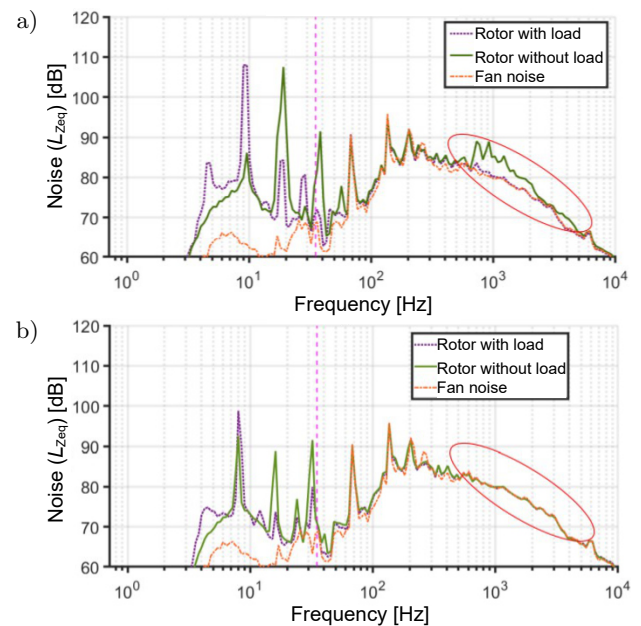


Fig. 7. $1/12$ octave bands spectrum for considered conditions presented separately for examined rotors: a) single-segment rotor; b) five-segment rotor.

loading is applied, aligning with the complete rotor RPM of 286 (4.78 Hz). This phenomenon is also evident in unloaded conditions. Additionally, the effect of two blades is apparent in both scenarios, corresponding to an RPM value of 572 (9.55 Hz), indicating a doubling of the frequency relative to the RPM. Notably, low-frequency, narrow-band rotational components typically occur at the blade passage frequency (the rotational speed multiplied by the number of blades) and its integer multiples, consistent with the findings presented in (PEARSON, GRAHAM, 2014). Mathematically, the acoustic energy for all sub-bands needs to be summed from 1 Hz to 51 Hz. Specifically, 75 sub-bands are present from 0.71 Hz up to 51 Hz. A unified value for this frequency range was computed. Initially, the SPL is converted to a relative exposure, representing the square of the acoustic pressure divided by the square acoustic reference pressure, for each considered sub-band using the equation:

$$Ex = 10^{(l/10)},$$

where Ex denotes the relative exposure and l represents the SPL expressed [dB]. Subsequently, these values are aggregated. Finally, the sum of the relative exposure is transformed to the SPL using logarithms via the equation:

$$L = 10 * \log_{10}(\text{sum}(Ex)).$$

The discrepancy between the rotors was subsequently determined: this difference = -1 dB for stationary rotors, indicating a nearly identical SPL for these conditions; for rotors under no-loading, this difference = 11.2, signifying that the single-segment rotor generated an SPL 11.2 dB higher than the five-segment rotor. Hence, it can also be inferred that the acoustic energy produced by the single-segment rotor is 13.3 times greater than that of the five-segment rotor. A comparable difference of 11.4 is observed for the loaded rotor. In this instance, the acoustic energy produced by the single-segment rotor is 13.8 times greater than that of its five-segment counterpart.

As previously mentioned, an elevation in the noise produced by the single-segment rotor is noted, between approximately 500 Hz and 1500 Hz. This region is delineated by a red ellipse in Fig. 7a and is absent in the case of its five-segment counterpart. Additionally, it was observed that noise peaks in the case of a five-segment rotor transitioned towards broadband frequencies and attenuated at a faster rate. This phenomenon arises from the heightened mixing of complex vortices formed in the case of a five-segment rotor compared to two rotors under similar testing conditions, disrupting the coherence patterns. This effect, akin to the addition of serrations in aircraft wings, serves to distribute noise-generating vortices, consequently reducing noise levels (MOREAU,

DOOLAN, 2013; VAN DER VELDEN, OERLEMANS, 2017; MATHEW *et al.*, 2016; OERLEMANS *et al.*, 2009). This observation elucidates the acoustic energy distribution presented in Fig. 7, where identical frequencies are recorded with varying sound levels for each scenario.

Upon evaluation of both cases, it becomes apparent that the unloaded rotor generates more noise compared to its loaded counterpart, as depicted in Fig. 7. It is evident from the plots and tables that the five-segment rotor yields a lower sound energy compared to its single-segment counterpart, a conclusion further validated by calculating the sound intensity difference between the two cases.

3.3. Psychoacoustics

For the next part of the research, psycho-acoustics related to sound quality was analyzed based on the studies presented in (SAHAI, 2016; BLAUERT, 2005). Figure 8 shows a variation in the different sound parameters under consideration. The FS, representing the modulation in sound, is at its peak value for a frequency of about 4 Hz.

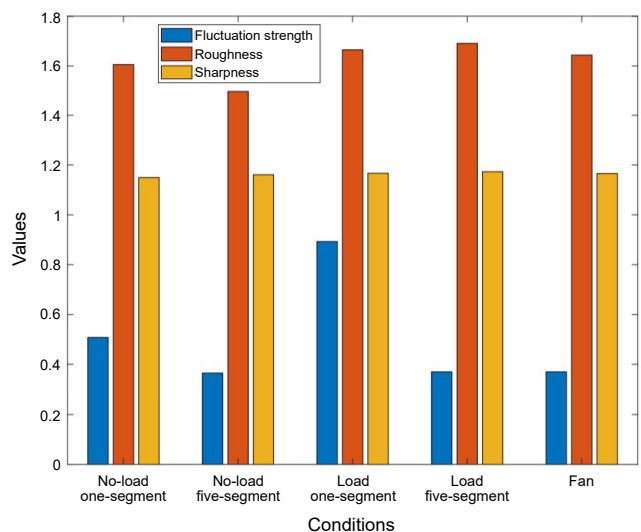


Fig. 8. Information about sound quality for single and five-segment rotor under different operational conditions, the units are as follows: FS [vacil], roughness [asper], sharpness [acum].

A higher FS draws more attention to it and can thus cause more irritation. It is evident from Fig. 6b that a peak related to the BPF (blade passing frequency) takes place about this frequency for the case when the rotor is under maximum-loading. This is also seen in Fig. 8. Another factor under consideration is the sound roughness which reaches a maximum at a frequency of about 70 Hz. As anticipated from the power spectrum plots, sharpness, which is related to how pleasant one feels when hearing the noise, is mainly dominated by higher frequencies and is similar in all the cases.

A broadband analysis of selected SQ parameters showed that it was the FS that was the metrics distinguishing the examined rotor types. Differences in the FS parameter values occurred both in the free-run (slight differences) and the load mode (larger changes). An extended analysis of this indicator in the frequency domain was performed to more precisely check the properties of the tested rotors in terms of diversity in the FS parameter. Due to the fact that SQ measures refer to perceptual phenomena, the FS spectrum is presented on the Bark scale. The analyses were performed for a free run and load phases. The obtained results are shown in Fig. 9. A higher value of the FS indicator (related to the five-segment rotor) in the free run mode for a single-segment rotor occurred for the frequency bands covering the Bark scale from 5 to 8 (the widest range), 11–13, and 17–19. A slight increase in the FS index was noted only for the 3-Bark band in the load phase of the five-segment rotor. It can be summarized that the FS in this rotor type does not depend on the type of rotation and that this rotor type does not induce this type of sensation. In the case of a single-segment rotor, a drastic increase in the FS value was observed in the load conditions. Moreover, a significant increase in the number of bands in which an increase in the FS took place was also observed. The lower range was expanded by a band of Bark 3. The two higher sub-ranges were merged into one range from 10 to 17 Barks. This means that the impression of the modulation strength expanded in the frequency domain, taking on a broadband character and its strength increased significantly.

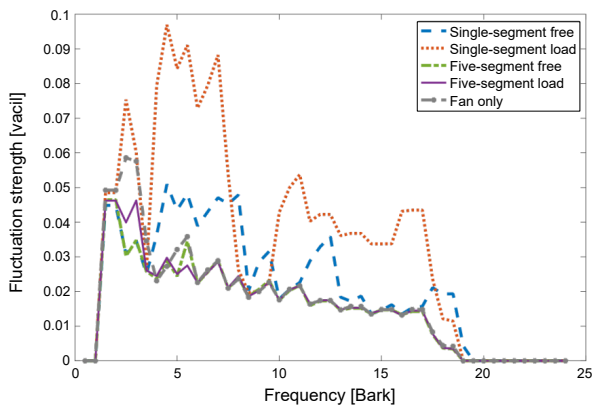


Fig. 9. FS spectrum on the Bark scale for examined rotors under considered conditions.

A single-segment rotor modulates background noise much more effectively. This results directly from its design. The wide blade of the wing captures a large volume of air at a given moment (which is the center of acoustic wave propagation), causing instantaneous rapid changes in the sound pressure (resulting from the rotor rotational speed), which is audible as a noise modulator (a sound similar to a flag flapping or flap-

ping in the wind). This was confirmed by a higher value of the FS index for a single-segment rotor. The modulation of the background noise in a five-segment rotor is very low, and the obtained indicators do not differ significantly from the values obtained for the background noise. Less disturbance of the acoustic background for a five-segment rotor comes at the cost of a lower energy efficiency of this structure. A single-segment rotor is characterized by greater efficiency and more noticeable changes in the acoustic background. The broadband SQ results presented for the fan only, especially the FS, obtained for the fan noise (without rotors), have similar values as the stopped rotors and the five-segment rotor with load.

The frequency domain analysis of the fan noise (without rotors) indicated that the FS distribution was similar to the five-segment rotor (for both conditions: free run and load) and was even greater for Bark 2–4 than for this rotor. It means that the five-segment rotor efficiently reduced the modulation of the sound.

3.4. Rotor efficiency and noise

The final finding, which explores the impact of segmentation on efficiency alongside noise, is illustrated in Fig. 10. The performance variation across the entire operational range of TSR has been depicted. Notably, the single-segment rotor exhibits superior C_p values throughout the TSR range under examination. It should also be noted that the maximum efficiency can be seen at a TSR value of about 1 for both cases, corresponding to a rotor rotational speed of 300 RPM. The maximum rotational speed in the no-load case takes place just above 500 RPM. A single-segment rotor achieves a maximum efficiency of 33.15%, whereas a five-segment rotor attains a maximum efficiency of 26.45%. A relative decrease of 20% in efficiency is observed for a segmented rotor compared to a conventional Savonius rotor.

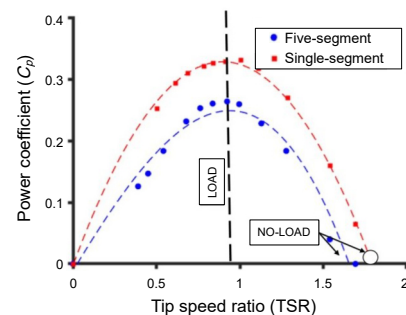


Fig. 10. Performance characteristics with respect to TSR.

Additional plates are inserted in the case of a segmented rotor (Fig. 2). These plates introduce additional surfaces at which boundary layers are formed, inducing viscous losses in the rotor. In addition, these boundary layers are inserted into the highly curved

blades, which generate a pressure gradient normal to the blade surface. Such a pressure gradient generates a transverse flow within the boundary layer, becoming a source of 3D flow structures. These two reasons become a source of increased losses at a segmented rotor.

Comparing these results to the SPL analysis presented in Subsec. 3.2, it should be recalled that the SPL in stopped rotors ($\text{RPM} = 0$) is nearly the same. In the case of maximum-loading rotors, the single-segment rotor produces a much higher SPL than the five-segment rotor. It should be taken into account that the introduction of additional plates in the segmented rotor becomes a source of streamwise vorticity, which is intersecting blade span-wise vortices. This vortex disintegration mechanism reduces noise but often increases losses. This may also shift the emitted noise from low to higher frequencies. The differences occur mainly in the low-frequency region (1 Hz–51 Hz). Thus, it can be concluded that an increase in power production is redeemed at the expense of an increase in noise emissions. The contribution of noise emissions from a single-segment in the free-run condition of the rotor is still higher and includes contributions from higher frequencies (around 1000 Hz), as well.

4. Conclusions

The presented study indicates that the noise produced by a wind turbine is influenced by its operational parameters. This research involves a comparative analysis between a single-segment rotor and a five-segment rotor, offering insights into the most favorable operating conditions and the rotor type that minimizes the noise while maximizing the efficiency. It can be inferred from the results obtained that:

- a five-segment rotor produces less overall acoustic energy as compared to that produced by a single-segment rotor. The changing phase of the segments breaks the coherence of the flow and shifts the noise energy content to higher frequency spectra, where the dissipation rate is faster;
- the amount of acoustic energy produced decreases as the load increases, peaking when the rotor is spinning without any external-load;
- the greatest differences in noise emissions between the considered types of rotors concern low frequency noise.
- under the specified test conditions at a wind speed of 8.5 m/s, a single-segment rotor demonstrates a greater efficiency compared to a five-segment rotor, achieving a maximum performance efficiency of 32 % as opposed to 26.5 % for the five-segment rotor;
- following the conclusions, the sound quality of a five-segment rotor makes it a preferable choice.

Very practical conclusions can be drawn from the analysis: if the lowest possible noise level is a critical parameter in a given area, it is recommended that multi-segment structures are used. However, if the priority is to obtain the highest possible efficiency of the system, then the first choice will be to use single-segment systems.

The analysis of the sound quality measures indicated that the metrics such as roughness and sharpness are similar for both rotors. The significant differences between the examined rotors were noticed in the FS. The greater value of these metrics occurred for the single-segment rotor. The SQ metric values in the case of the five-segment rotor were similar in all conditions (free-run and with load) to the values obtained for the fan noise. Further analysis is required in a quieter environment to maintain the efficiency of the single-segment rotor with low noise levels. An experimental campaign has been planned in collaboration with the Technical University of Delft to localize the noise sources using a microphone array in contrast to a single microphone used for this study. This will help in optimizing the design and reaching our final aim of more production of power with less noise disturbance.

FUNDING

The project received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 860101-zEPHYR and was supported by CI TASK (Gdańsk, Poland).

References

1. ABBASI M. *et al.* (2019), Assessment of role of job components and individual parameters on the raised blood pressure in a noisy industry, *Archives of Acoustics*, **44**(3): 575–584, <https://doi.org/10.24425/aoa.2019.129272>.
2. AKWA J.V., VIELMO H.A., PETRY A.P. (2012), A review on the performance of Savonius wind turbines, *Renewable and Sustainable Energy Reviews*, **16**(5): 3054–3064, <https://doi.org/10.1016/j.rser.2012.02.056>.
3. ANJUM S., KUMARI A. (2022), Evaluation of noise pollution in Bengaluru City, India during COVID-19 pandemic, *Archives of Acoustics*, **47**(2): 131–140, <https://doi.org/10.24425/aoa.2022.141644>.
4. BLAUERT J. (2005), *Communication Acoustics*, Springer, <https://doi.org/10.1007/b139075>.
5. COX T. (n.d.), Roughness – Fluctuation Strength, <https://hub.salford.ac.uk/sirc-acoustics/psychoacoustics/sound-quality-making-products-sound-better/an-introduction-to-sound-quality-testing/roughness-fluctuation-strength/> (access: 27.03.2024).
6. DAVY J.L., BURGEMEISTER K., HILLMAN D. (2018), Wind turbine sound limits: Current status and recommendations based on mitigating noise annoyance,

- Applied Acoustics*, **140**: 288–295, <https://doi.org/10.1016/j.apacoust.2018.06.009>.
7. DESHMUKH S., BHATTACHARYA S., JAIN A., PAUL A.R. (2019), Wind turbine noise and its mitigation techniques: A review, *Energy Procedia*, **160**: 633–640, <https://doi.org/10.1016/j.egypro.2019.02.215>.
 8. DOERFFER K., TELEGA J., DOERFFER P., HERCEL P., TOMPOROWSKI A. (2021), Dependence of power characteristics on Savonius rotor segmentation, *Energies*, **14**(10): 2912, <https://doi.org/10.3390/en14102912>.
 9. DRISS Z., MLAYEH O., DRISS D., MAALLOU M., ABID M.S. (2014), Numerical simulation and experimental validation of the turbulent flow around a small incurved Savonius wind rotor, *Energy*, **74**: 506–517, <https://doi.org/10.1016/j.energy.2014.07.016>.
 10. DUMITRESCU H., CARDOS V., DUMITRACHE A., FRUNZULICĂ F. (2010), Low-frequency noise prediction of vertical axis wind turbines, [in:] *Proceedings of the Romanian Academy – Series A: Mathematics, Physics, Technical Sciences, Information Science*, **11**(1): 47–54.
 11. FASTL H., ZWICKER E. (2007), *Psychoacoustics: Facts and Models*, Springer Berlin, Heidelberg, 3rd ed., pp. 203–238, <https://doi.org/10.1007/978-3-540-68888-4>.
 12. FLASZYNSKI P., PIOTROWICZ M., BACCI T. (2021), Clocking and potential effects in combustor–turbine stator interactions, *Aerospace*, **8**(10): 285, <https://doi.org/10.3390/aerospace8100285>.
 13. GALLO P., FREDIANELLI L., PALAZZUOLI D., LICITRA G., FIDECARO F. (2016), A procedure for the assessment of wind turbine noise, *Applied Acoustics*, **114**: 213–217, <https://doi.org/10.1016/j.apacoust.2016.07.025>.
 14. GHASEMIAN M., NEJAT A. (2015), Aerodynamic noise prediction of a horizontal axis wind turbine using improved delayed detached Eddy simulation and acoustic analogy, *Energy Conversion and Management*, **99**: 210–220, <https://doi.org/10.1016/j.enconman.2015.04.011>.
 15. GRAHAM W., PEARSON C. (2022), Noise from a model-scale vertical-axis wind turbine, *AIAA Journal*, **60**(1): 224–235, <https://doi.org/10.2514/1.J060531>.
 16. GRZELAK J., CARRILLO L. G., NAKIELSKI J., PIOTROWICZ M., DOERFFER K. (2024), An innovative floating system with a Savonius rotor as a horizontal-axis wind turbine, *Polish Maritime Research*, **31**(2): 13–19, <https://doi.org/10.2478/pomr-2024-0017>.
 17. HAFKE-DYS H.Z., PREIS A., KACZMAREK T., BINIAKOWSKI A., KLEKA P. (2016), Noise annoyance caused by amplitude modulated sounds resembling the main characteristics of temporal wind turbine noise, *Archives of Acoustics*, **41**(2): 221–232, <https://doi.org/10.1515/aoa-2016-0022>.
 18. HANSEN C., HANSEN K. (2020), Recent advances in wind turbine noise research, *Acoustics*, **2**(1): 172–207, <https://doi.org/10.3390/acoustics2010013>.
 19. IBIS Power (n.d.), Complicated sustainability issues in real estate, solved beautifully and easily, <https://www.ibispower.eu/powerbest> (access: 27.03.2024).
 20. IIDA A., MIZUNO A., FUKUDOME K. (2004), Numerical simulation of aerodynamic noise radiated from vertical axis wind turbines, [in:] *Proceedings of the 18 International Congress on Acoustics*.
 21. JEONG M.-S., KIM S.-W., LEE I., YOO S.-J., PARK K.C. (2014), Investigation of wake effects on aeroelastic responses of horizontal-axis wind-turbines, *AIAA Journal*, **52**(6): 1133–1144, <https://doi.org/10.2514/1.J051899>.
 22. KACPRZAK K., LISKIEWICZ G., SOB CZAK K. (2013), Numerical investigation of conventional and modified Savonius wind turbines, *Renewable Energy*, **60**: 578–585, <https://doi.org/10.1016/j.renene.2013.06.009>.
 23. KOTUS J., KOSTEK B. (2008), The noise-induced harmful effect assessment based on the properties of the human hearing system, *Archives of Acoustics*, **33**(4): 435–440.
 24. MAHMOUD N., EL-HAROUN A.A., WAHBA E., NASEF M.H. (2012), An experimental study on improvement of Savonius rotor performance, *Alexandria Engineering Journal*, **51**(1): 19–25, <https://doi.org/10.1016/j.aej.2012.07.003>.
 25. MATHEW J., SINGH A., MADSEN J., ARCE LEÓN C. (2016), Serration design methodology for wind turbine noise reduction, *Journal of Physics: Conference Series*, **753**(2): 022019, <https://doi.org/10.1088/1742-6596/753/2/022019>.
 26. MÖLLERSTRÖM E., LARSSON S., OTTERMO F., HYLANDER J., BÄÄTH L. (2014), Noise propagation from a vertical axis wind turbine, [in:] *Proceedings of 43rd International Congress on Noise Control Engineering: Internoise 2014*.
 27. MOREAU D.J., DOOLAN C.J. (2013), Noise-reduction mechanism of a flat-plate serrated trailing edge, *AIAA Journal*, **51**(10): 2513–2522, <https://doi.org/10.2514/1.J052436>.
 28. MÜNDEL T., SCHMIDT F.P., STEVEN S., HERZOG J., DAIBER A., SØRENSEN M. (2018), Environmental noise and the cardiovascular system, *Journal of the American College of Cardiology*, **71**(6): 688–697, <https://doi.org/10.1016/j.jacc.2017.12.015>.
 29. New World Wind (n.d.), WindTree, <https://www.newworldwind.com/windtree> (accessed: 27.03.2024).
 30. NGUYEN D.P., HANSEN K., ZAJAMSEK B., CATCHESIDE P. (2020), Evaluation of wind farm noise amplitude modulation synthesis quality, *Applied Acoustics*, **166**: 107349, <https://doi.org/10.1016/j.apacoust.2020.107349>.
 31. OERLEMANS S., FUGLSANG P. (2012), Low-noise wind turbine design, Siemens AG, <https://www.ewea.org/events/workshops/wp-content/uploads/2012/12/EWEA-Noise-Workshop-Oxford-2012-1-1-Stefan-Oerlemans.pdf> (access: 27.03.2024).
 32. OERLEMANS S., FISHER M., MAEDER T., KÖGLER K. (2009), Reduction of wind turbine noise using optimized airfoils and trailing-edge serrations, *AIAA Journal*, **47**(6): 1470–1481, <https://doi.org/10.2514/1.38888>.

33. PAWLACZYK-ŁUSZCZYŃSKA M., DUDAREWICZ A., ZABOROWSKI K., ZAMOJSKA-DANISZEWSKA M., WASZKOWSKA M. (2014), Annoyance related to wind turbine noise, *Archives of Acoustics*, **39**(1): 89–102, <https://doi.org/10.2478/aoa-2014-0010>.
34. PEARSON C.E., GRAHAM W.R. (2014), Experimental characterization of vertical-axis wind turbine noise, *The Journal of the Acoustical Society of America*, **137**(1): EL111–EL116, <https://doi.org/10.1121/1.4904915>.
35. SACHAR S., DOERFFER P., FLASZYNSKI P., KOTUS J., DOERFFER K., GRZELAK J. (2023), Correlation between the generated noise and effectiveness for a vertical axis Savonius type rotor, [in:] *AIAA SCITECH 2023 Forum*, <https://doi.org/10.2514/6.2023-0611>.
36. SAHAI A.K. (2016), Consideration of aircraft noise annoyance during conceptual aircraft design, Ph.D. Thesis, RWTH Aachen University.
37. SUN X., LUO D., HUANG D., WU G. (2012), Numerical study on coupling effects among multiple Savonius turbines, *Journal of Renewable and Sustainable Energy*, **4**(5): 053107, <https://doi.org/10.1063/1.4754438>.
38. SURESH T., FLASZYNSKI P., CARPIO A.R., KUROWSKI M., PIOTROWICZ M., SZULC O. (2024), Aeroacoustic effect of boundary layer separation control by rod vortex generators on the DU96-W-180 airfoil, *Journal of Fluids and Structures*, **127**: 104133, <https://doi.org/10.1016/j.jfluidstructs.2024.104133>.
39. SZYCHOWSKA M., HAFKE-DYS H., PREIS A., KOCIŃSKI J., KLEKA P. (2018), The influence of audio-visual interactions on the annoyance ratings for wind turbines, *Applied Acoustics*, **129**: 190–203, <https://doi.org/10.1016/j.apacoust.2017.08.003>.
40. VAN DEN BERG G. (2004), Effects of the wind profile at night on wind turbine sound, *Journal of Sound and Vibration*, **277**(4–5): 955–970, <https://doi.org/10.1016/j.jsv.2003.09.050>.
41. VAN DER VELDEN W.C., OERLEMANS S. (2017), Numerical study on combed teeth serrations for wind turbine noise reduction, [in:] *23rd AIAA/CEAS Aeroacoustics Conference*, <https://doi.org/10.2514/6.2017-4172>.
42. ZARE S., GHOTBIRAVANDI M.R., ELAHISHIRVAN H., AHSAEED M.G., ROSTAMI M., ESMAEILI R. (2020), Modeling and predicting the changes in hearing loss of workers with the use of a neural network data mining algorithm: A field study, *Archives of Acoustics*, **45**(2): 303–311, <https://doi.org/10.24425/aoa.2020.133150>.