eISSN 2300-262X
ISSN 0137-5075

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

ARCHIVES of ACOUSTICS

QUARTERLY

Vol. 51, No. 1, 2026

WARSAW

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

ARCHIVES of ACOUSTICS

QUARTERLY, Vol. 51, No. 1, 2026

Research Papers

Technical Notes

OSA 2025

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

Editorial Office
Institute of Fundamental Technological Research PAS, Pawińskiego 5B, 02-106 Warsaw, Poland
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 LaTeX: Jezierska Katarzyna (Institute of Fundamental Technological Research PAS, Poland)

Archives of Acoustics Vol. 51, No. 1, pp. 3–23 (2026), https://doi.org/10.24423/archacoust.2026.4190

Research Paper

Application of Acoustic Metamaterials in the Design of Muffling Unit of Internal Combustion Engine

Tianshuo Li, Lei Zheng*, Zhengyang Bi, Hongchao Ji

College of Mechanical Engineering, North China University of Science and Technology
Tangshan, Hebei, China

*Corresponding Author: zhenglei1@163.com

Received February 13, 2025; revised October 29, 2025; accepted November 24, 2025; available online January 15, 2026; version of record March 17, 2026; published issue March 27, 2026.

Aiming at the problems of large volume, high exhaust resistance and difficulty in suppressing noise in the 500 Hz to 100 Hz frequency band of traditional internal combustion engine exhaust mufflers, a noise reduction unit design based on acoustic metamaterials is proposed. Based on the equivalent medium theory, an acoustic model with a ring structure and multi-region variable refractive index was established. Phase control is achieved by helically winding the acoustic channel to change the refractive index, and the basic dimensions of the acoustic metamaterial muffling unit are calculated. The sound field distribution, transmission loss and flow field characteristics of the muffling unit are simulated and analyzed. This structure utilizes a multi-layer acoustic channel structure, effectively alleviating the problem of insufficient low-frequency noise elimination caused by the asymmetry of Fano interference. It achieved a transmission loss of over 10 dB within 85 % of the 500 Hz to 1000 Hz frequency band, and still maintained excellent noise reduction performance under high-frequency conditions through multi-level phase control. By connecting multiple units in series, a transmission loss of 10 dB can be achieved within 85 % of the 500 Hz to 1000 Hz frequency band. The exhaust flow field of the muffling unit was simulated and analyzed. Whether used alone or in series with the traditional muffling structure, the exhaust resistance remained within the range of 360 Pa to 370 Pa. Experimental tests show that when the metamaterial muffler unit is used in combination with the traditional muffler, it effectively achieves targeted noise elimination in the 500 Hz to 1000 Hz frequency band, and also demonstrates clear noise reduction capabilities in higher frequency ranges. The high noise suppression characteristics, high gas passage characteristics and compact volume characteristics of this structure provide more potential analysis methods and design schemes for the research and development of internal combustion engine mufflers and noise reduction accessories.

Keywords: acoustic metamaterials, muffler, transmission loss, exhausting resistance.

1. Introduction

Internal combustion engine noise has always been one of the significant sources of urban noise pollution (Wan *et al.*, 2023), among which exhaust noise has always accounted for a considerable proportion in engine noise emissions (Narayan *et al.*, 2021). Mufflers can suppress noise while ensuring exhaust capacity. However, the structural design of the noise reduction units is restricted by multiple performance indicators, such as noise reduction frequency, structural dimensions, acoustic performance, and aerodynamic performance (Zhang *et al.*, 2018).

The traditional structure design of mufflers is generally based on the acoustic wave equation, using reflection and interference methods to achieve the cancellation of sound waves (Araujo, Pestana, 2024; Li *et al.*, 2024; Lee *et al.*, 2020); through the Helmholtz resonance principle to absorb the vibration energy of sound waves

(Bai *et al.*, 2023; Kheybari, Ebrahimi-Nejad, 2019) or based on the small hole injection theory to shift the vibration sound waves to high frequencies and ultra-high frequencies (Chen *et al.*, 2020; Wang *et al.*, 2012). Under high-speed exhaust conditions, these structures will increase the exhaust resistance of the engine. Meanwhile, due to the longer wavelength of low-frequency noise and the fact that high-temperature exhaust conditions make this issue more prominent, this imposes certain limitations on the spatial structure and size design of the muffler when eliminating noise with frequencies below 1000 Hz. However, human hearing has different loudness responses to noise of different frequencies. Noise ranging from 500 Hz to 2000 Hz has a higher loudness. Therefore, noise elimination under the frequency conditions of 500 Hz to 1000 Hz has become a difficult point in the muffler of internal combustion engines. However, these muffler structures will inevitably cause an increase in the exhaust resistance of the energy engine. Martos *et al.* (2023) showed in his study on the exhaust back pressure of Euro 6 diesel engine that when the back pressure was increased from about 4500 Pa to 12 500 Pa, the average effective pressure (BMEP) of the engine under the medium load condition of 2000 rpm decreased by about 10 %. Dalla Nora *et al.* (2016) showed in the study of backpressure sensitive two-stroke GDI engine that the exhaust resistance of the engine has a significant impact on the performance of the engine under the condition of 800 rpm and 2000 rpm. Under the condition of 800 rpm, when the backpressure is increased from 104 Kpa to 120 Kpa. The power and efficiency of the engine have both decreased by 15 % to 20 %.

Especially when multiple muffler structures are connected in series, the contradiction between the acoustic effect of the muffler, its spatial volume and exhaust resistance becomes more prominent (Liao *et al.*, 2024; Su *et al.*, 2020).

Acoustic metamaterials, also known as acoustic artificial structure or Acoustic Metamaterials, applied the principle of optical metamaterials and based on the theory of equivalent media, acoustically artificial designed structural units arranged individually or with periodic array are adopted (Liu *et al.*, 2020), as shown in Fig. 1. Through complex spatial structures or combinations of multiple media, the equivalent acoustic parameters, such as the equivalent volume modulus and the equivalent elastic modulus, are changed to change the sound field (Fang *et al.*, 2006; Ye *et al.*, 2020; Orazbayev, Fleury, 2020; Zeng *et al.*, 2013).

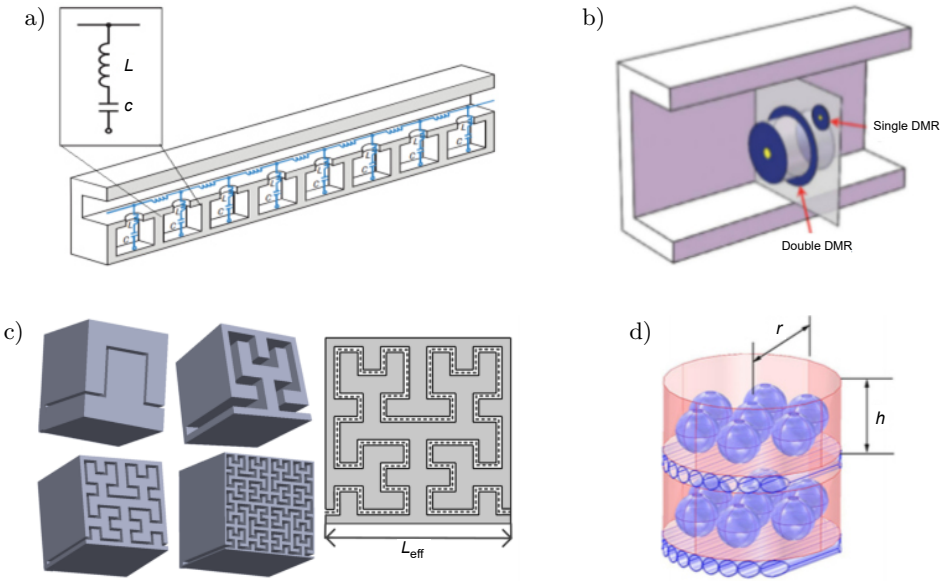


Fig. 1. Various types of acoustic metamaterials: a) local resonance, b) thin film, c) spatially coiled, d) supramolecular.

Acoustic metamaterials can achieve better acoustic effects in the fields of sound field control, sound absorption and sound insulation, especially in the aspects of abnormal reflection refraction of sound waves, compact sound absorption structure, etc. At the same time, it also provides certain ideas in ventilation and sound insulation, especially in the aspect of noise reduction with large porosity, such as: local resonance model (Jiang *et al.*, 2023; Li *et al.*, 2022; Liu *et al.*, 2019; Zhang *et al.*, 2024) and phase regulation model (Guo *et al.*, 2023; Xu *et al.*, 2021).

Mohamad *et al.* (2021) combined acoustics and fluid mechanics to analyze the system back pressure increase behavior caused by Helmholtz resonance near the small hole during the muffling process of the perforated pipe. Under the exhaust gas velocity condition of 40 m/s, the perforated pipe model would lead to a pressure loss of 795 Pa to 889 Pa. Meanwhile, the research shows that the secondary noise caused by the increased resistance will lead to a 10 % increase in the outlet sound pressure level. Sun *et al.* (2020) pointed out that interference based on local resonance only works at resonant frequencies, which means that there is only a narrow operating frequency range near each noise suppression interference frequency. Considering that noise is usually broadband noise, there are still certain challenges in designing acoustic ventilation structures for broadband. The muffler based on the phase control principle has greater advantages in gas permeability.

Zhang *et al.* (2024) proposed a combined acoustic metamaterial (CAM) based on multi-channel Fano resonance. By coupling helical channels with hollow structures, it achieves broadband sound insulation of >13 dB in the 520 Hz to 989 Hz frequency band while maintaining a ventilation rate of 63 %. Experimental verification shows that its peak acoustic transmission loss reaches 62 dB, providing a lightweight solution for traffic noise control.

Ghaffarivardavagh *et al.* (2019) proposed a super-open acoustic metamaterial (UOM) based on the Fano interference effect. By comparing the acoustic impedance of the central ventilation hole (60 % open area) with that of the spiral channel. It achieves a 94 % acoustic energy attenuation at 460 Hz (with a STL peak of 26 dB), providing a highly breathable solution for ventilation and noise reduction structures.

Although the operating frequency range is expanded by inducing multiple resonances, the airflow channel is weakened to a large extent, and the phase-controlled structure can provide greater porosity while supporting broadband noise cancellation. For spiral metamaterials, the plane space of traditional acoustic metamaterials is folded and extended to the 3D space, which greatly improves the space utilization rate. The higher center hole area is obtained, so it is more suitable for the basic structural unit of internal combustion engine exhaust noise reduction. Xu *et al.* (2025) combined the Archimedes spiral with the ventilation function and verified through experiments the balance between broadband noise reduction and high ventilation efficiency of the spiral metamaterials, providing a new geometric configuration reference for the design of ventilation and noise reduction units. Lu *et al.* (2025) combined Fano resonance with 3D expansion, solving the problems of the single shape and difficulty in practical application of traditional Fano-type mufflers, providing a theoretical basis for the modular design and 3D layout of ventilation mufflers. The above research indicates that the acoustic suppression device based on the principle of phase interference essentially does not eliminate sound but blocks its passage, enabling energy consumption at any point before the blocking point. Its high flowability is suitable for engine exhaust passages with high-speed and low-resistance requirements.

It can be seen from this that the medium and low-frequency components in the exhaust noise of internal combustion engines, due to their longer sound wave wavelengths, the traditional noise reduction technology based on pipe length tuning is difficult to meet the compact design requirements of modern power systems. However, acoustic metamaterials have a more compact structure and more efficient noise reduction capabilities, and the phase control model within them can provide a higher gas high pass rate. Therefore, based on the equivalent medium theory of acoustic metamaterials, we take advantage of the phase control structure and, under the premise of maximizing the utilization of space, design a low-resistance muffler with a hollow structure to achieve targeted noise elimination in the 500 Hz to 1000 Hz frequency band of internal combustion engines. This not only compensates for the current noise defect of internal combustion engine scheduling, but also provides an effective idea for wideband noise reduction of pipelines.

2. Structure design of noise reduction unit based on equivalent medium theory

2.1. Basic principles of acoustic metamaterials

The phenomena such as reflection and transmission in homogeneous media can be solved by mathematical equations in the classical propagation theory, but the propagation characteristics are difficult to analyze

when the medium is non-homogeneous. At this time, the inhomogeneous medium is treated as a homogeneous medium under some approximate conditions, which provides a solution to the problem. In this way, the equivalent medium theory is developed, which is also an important theoretical basis for the study of metamaterials. Therefore, the metamaterials of this array arrangement are equivalent to a uniform medium, which is also called metamaterials.

The concept of metamaterials was initially introduced into the field of optics to deal with electromagnetic waves using materials with negative dielectric constants and negative magnetic permeability. For a monochromatic wave in an isotropic medium, the square sum dispersion relation of the refractive index is (Li *et al.*, 2014):

$$k_s^2 = \frac{n_s^2 \omega^2}{c^2}, \tag{1}$$

where k_s is the wave number, n_s is the refractive index, ω is the frequency, c is the speed of light.

Since both electromagnetic wave and acoustic wave satisfy the wave equation, they have a large number of analog characteristics. This theoretical knowledge can be applied to acoustic wave, and acoustic metamaterial characteristics can be theoretically obtained, and the formula of acoustic refractive index can be obtained as follows (Chen *et al.*, 2024):

$$n_s^2 = \frac{\varepsilon}{\mu}, \tag{2}$$

where $\varepsilon$is the dielectric constant (in optics) and also the mass density (in acoustics), μ is the magnetic permeability (in optics) and also volume elastic modulus (in acoustics).

As shown in Fig. 2, in the first quadrant, the wave propagates normally in the pipeline with positive mass density and bulk elastic modulus ($\rho > 0$, $K > 0$). In the second quadrant, when the Helmholtz resonator structure is adopted, the resonant sound wave generated through the resonator is opposite to the incident sound wave, and the local force of the sound wave is opposite to its vibration acceleration direction ($\rho < 0$, $K > 0$), the refractive index is imaginary, and the sound wave propagation is terminated, that is, the sound attenuation is realized. In the fourth quadrant, when the thin film structure is used, when the vibration frequency of the thin film is opposite to the vibration frequency of the sound wave, the volume of the thin film will expand under pressure and shrink under tension ($\rho > 0$, $K < 0$), resulting in the sound wave stopping propagation (Naify *et al.*, 2012). In the third quadrant, the two structures exist at the same time, which presents a double negative characteristic ($\rho > 0$, $K < 0$). At a specific frequency, the sound wave can continue to propagate, but the acoustic refractive index will change, so as to achieve special functions such as acoustic lens (Wang *et al.*, 2024).

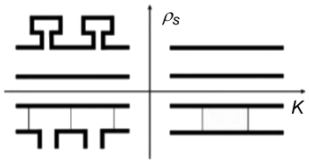


Fig. 2. Quadrant diagram of equivalent acoustic parameters.

From Eq. (1) and Eq. (2), it can be seen that when the dielectric constant or magnetic permeability is negative, the square of the refractive index remains negative, that is, the refractive index is an imaginary number, and the wave propagation stops. When both the dielectric constant and magnetic permeability are negative, the square of the refractive index remains positive, that is, the refractive index is still a real number. The propagation of the wave can continue, but the propagation characteristics will change. It is precisely this feature that provides the theoretical condition for phase control to achieve the muffler of the internal combustion engine displacement (Shaikh *et al.*, 2024).

The concept of acoustic metamaterials is not limited to locally resonant materials or periodic structures. Sound waves in a fluid are longitudinal scalar waves, and unlike electromagnetic waves, acoustic waveguides have no cut-off frequency and can limit their path through narrow channels at will. By limiting the sound path and adjusting the phase of the sound wave, the transmission track with negative dispersion characteristics ($\rho < 0$, $K < 0$)

can be folded and compressed into the subwavelength range (Kumar, Lee, 2020), such as the spatial winding metamaterials, topological acoustic metamaterials, fractal acoustic metamaterials, spiral metamaterials and acoustic metasurface. For the helical spatial acoustic metamaterials, the incident sound wave is confined to the helical subwavelength cross section channel wound in 3D space, resulting in the phase delay of the sound wave propagating along the slender path, thus achieving a higher acoustic refractive index. The phase delay is calculated as

$$\Delta\varphi = 2\pi k l_0, \tag{3}$$

where $\Delta\varphi$ is phase delay and l_0 is the valid channel length.

2.2. Design of equivalent medium theoretical model for acoustic metamaterials

In order to regulate the phase of sound waves in different acoustic transmission areas, the sound attenuation is realized through the strongly coupled dipole vibration. The acoustic model of the acoustic attenuation unit can be equivalent to the theoretical model of the radial double-layer material structure, as shown in Fig. 3. According to the effective medium theory, the central part and the surrounding part are characterized by refractive indices n_1 and n_2, respectively.

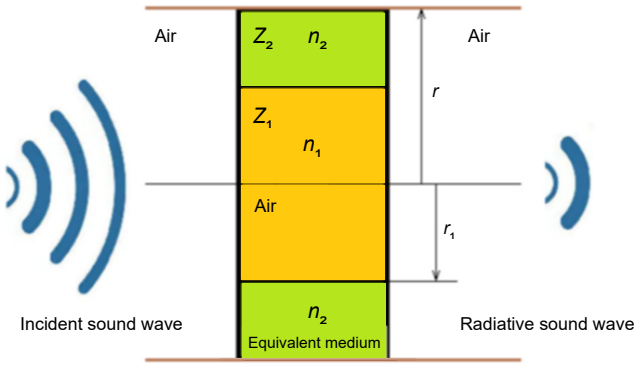


Fig. 3. Equivalent medium model of muffler element.

Since the acoustic metamaterial unit is a subwavelength scale, the vibration of the acoustic metamaterial unit can be characterized by the characteristic mode function and the characteristic frequency by its Green function (Yang *et al.*, 2013):

$$G(h,h') = \sum_n \frac{p_n(h)p_n^*(h')}{(\omega_n^2 - \omega^2)/B_n}, \tag{4}$$

where $p_n(h)$ and $p_n^*(h')$ are the n-order eigenmodular functions and their conjugations at positions h and h', ω_n is the characteristic frequency of the n-order, B_n is the average modulus of the n-order eigenmodulus:

$$\frac{1}{B_n} = \frac{\int p_n(h)\, p_n^*(h)\, \mathrm{d}v}{B_0 s}, \tag{5}$$

where s is the cross-sectional area of the unit, B_0 is the volume modulus of air, $B_0 = \rho_0 c_0$, ρ_0 is the air density, and c_0 is the air sound velocity.

Vibration in acoustic metamaterial units can be expressed through monopole and dipole modes (Yang *et al.*, 2014):

$$\overline{G}_+ = \omega^2 \sum_n \frac{p_n^*(h)\,[p_n(h) + p_n(-h)]}{(\omega_n^2 - \omega^2)/B_n}, \tag{6}$$

$$\overline{G}_- = \omega^2 \sum_n \frac{p_n^*(h)\,[p_n(h) - p_n(-h)]}{(\omega_n^2 - \omega^2)/B_n}, \tag{7}$$

where h is the axial position coordinate of the acoustic metamaterial unit.

Based on the effective medium theory, the acoustic metamaterials unit can be equivalent to a homogenized medium of the same thickness, where the effective wave number $\overline{k}$ satisfies the equation:

$$\left(\frac{\partial^2}{\partial x^2} + \overline{k}_s^2\right)\overline{p}\,(z) = 0. \tag{8}$$

Then the wave number can be determined by the unipolar mode and dipolar mode as (Sun *et al.*, 2020):

$$\overline{k}_s = \frac{1}{h}\operatorname{arccot}\sqrt{-\frac{\overline{G}_+}{\overline{G}_-}}. \tag{9}$$

Effective impedance $\overline{Z}$ can be expressed as

$$\overline{Z} = -\frac{\overline{G}_+}{\omega\cot\left(\overline{k}_s h\right)}. \tag{10}$$

Finally, the transmission coefficient of the acoustic metamaterials unit is obtained (Xiao *et al.*, 2013):

$$T = \frac{4e^{2ih\overline{k}_s}\overline{Z}Z_0}{\left(\overline{Z}+Z_0\right)^2 - e^{4ih\overline{k}_s}\left(\overline{Z}-Z_0\right)^2}, \tag{11}$$

where Z_0 is the acoustic impedance in the air, $i = \sqrt{-1}$.

Correspondingly, the transmission loss equation is

$$\mathrm{TL} = 10\log_{10}\left(\frac{W_i}{W_t}\right) = -20\log_{10}|T|, \tag{12}$$

where W_i is the incident sound power, W_t is the transmitted sound power, T is the transmission coefficient.

From Eq. (11) and Eq. (12), it can be known that we have two methods to achieve noise reduction. The first approach is to increase the impedance of the air or metamaterial to create a significant difference, thereby reducing the transmission coefficient T and increasing the transmission loss TL. This is not easy to achieve under the condition of ventilation. The second approach is to set the real part of the transmission coefficient to zero and retain only the imaginary part. This processing method can also suppress the propagation of sound waves, which is exactly the effect we expect. To this end, we need to adjust the value of $h\overline{k}_s$ in $e^{2ih\overline{k}_s}$ so that there exists a relationship of $2hk = (2n+1)\pi$, which makes the imaginary part of the exponent of the factor $e^{2ih\overline{k}_s}$ equal to $(2n+1)\pi$.

When $\overline{G}_+ = \overline{G}_-$, the effective wave number becomes pure virtual infinite, then $\overline{k} \to \infty$, which indicates that the system has the same monopole and dipole response intensity, and the extreme attenuation of the incident wave is the complete block of the acoustic metamaterials unit. At this point, the impedance expressed by Eq. (10) is infinite, while the real part of the expressed transmission coefficient expressed by Eq. (11) is 0. The transmission loss expressed in Eq. (12) is the imaginary part of the transmission coefficient T, which is a relatively large value at this time. This provides a theoretical basis for sound field elimination for us to achieve noise reduction during the displacement of internal combustion engines.

2.3. Structural design of acoustic metamaterials muffling unit

If we want to achieve noise reduction in the 500 Hz to 1000 Hz frequency band of internal combustion engines through the model shown in Fig. 3, there are still two significant problems. Firstly, we need to find the corresponding material or structure of the equivalent medium in reality. Secondly, we also need to extend the noise reduction frequency of this model to the 500 Hz bandwidth.

2.3.1. Design of equivalent media

Although it is difficult to obtain high refractive index acoustic materials in the outer ring of the equivalent medium model in reality, the acoustic metamaterials wound in space can be utilized to extend the acoustic wave

transmission path, equivalently achieving a sound velocity significantly lower than that in the central region and a nearly zero acoustic refractive index. This enables the regulation of the incident acoustic wave phase and the realization of dipole vibration at the exit with a certain phase difference from the central channel. The spiral spatial winding structure can maximize the utilization of axial space and is more suitable for the structural requirements of exhaust noise reduction.

For the acoustic channel of the central cylindrical tubular structure and the acoustic channel of the outer ring spiral structure, there is an approximate proportional relationship between the acoustic impedance value (Ghaffarivardavagh *et al.*, 2019):

$$\frac{Z_2}{Z_1} = \frac{\pi_1^2}{l\,(r - r_1)}, \tag{13}$$

where Z_2 is the acoustic impedance of the outer loop spiral channel, Z_1 is the acoustic impedance of the central channel, and r_1 is the diameter of the central channel, r is the diameter of the outer ring, and l is the pitch of the spiral structure. When the ratio of Z_2/Z_1 is close to 1, the acoustic vibration response intensity of the two regions can be obtained respectively through impedance balance. To achieve noise reduction over a wider frequency range, we reduced the impedance of part Z_2 (Khodabakhsh *et al.*, 2025; Lin *et al.*, 2024) to make the value of Z_2/Z_1 is 0.6, which also provided the preparatory conditions for us to expand the design of the outer ring structure into multiple channels.

According to Eq. (3), the acoustic channel length is used to control the phase. According to the trigonometric relationship between the length and height of the helix structure, the acoustic refractive index n_s can be expressed as

$$n_s = \frac{1}{\cos\beta}, \tag{14}$$

where β is spiral angle.

To achieve noise elimination in the frequency range of 500 Hz to 1000 Hz and minimize the volume to the extreme, we have designed a helical structure muffler. When the exhaust pipe temperature is 350°, the sound velocity is approximately 550 m/s. Taking the conventional exhaust pipe inner diameter (internal diameter) of 5 cm as the design standard, the basic dimensions of the acoustic metamaterials muffling unit are obtained as follows: l = 11.2 cm, r_1 = 2.5 cm, r = 5.5 cm, and the spiral angle is 3.6π. The basic dimensions of the acoustic metamaterials muffling unit is shown in Fig. 4. However, the external center diameter of this structure is 8 cm, and the corresponding center noise reduction frequency is 843 Hz. The theoretical noise reduction frequency is between 550 Hz and 1050 Hz.

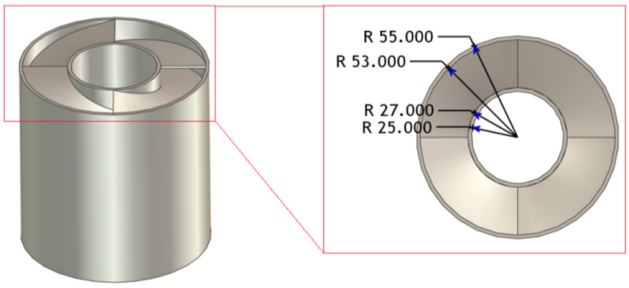


Fig. 4. Basic dimensions of the acoustic metamaterial muffling unit.

2.4. Design of wideband noise reduction

Although incomplete coupling can achieve a wider noise reduction frequency range, due to the asymmetry of Fano interference, the transmission loss in the low-frequency region decays too quickly, and thus an effective noise reduction effect cannot be obtained. To achieve noise reduction in lower frequency bands under the same volume conditions, further, we adopt a multi-channel structure to obtain multiple noise reduction frequency bands, weakening the influence of Fano interference asymmetry. As shown in Fig. 5, the outer ring area is radially divided into multiple acoustic channels to enhance the targeting of the noise reduction frequency.

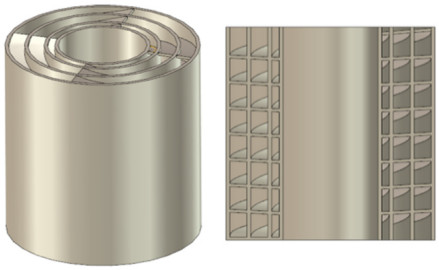


Fig. 5. Acoustic metamaterial muffling unit model.

According to the trigonometric relationship of the spiral development line, different acoustic channels correspond to different spiral outer diameters, respectively corresponding to different spiral angulation β:

$$\cos\beta = \frac{H}{2\pi r}, \tag{15}$$

where H is the spiral height, r is the spiral radius.

By bringing Eq. (15) into the relation between refractive index and spiral radius shown in Eq. (14), the relation between channel radius and acoustic refractive index can be obtained at a fixed height:

$$n_s = \frac{2\pi r}{H}. \tag{16}$$

For the multi-channel phase-controlled noise reduction structure, because each layer acoustic channel corresponds to different acoustic refractive index output and phase difference, multiple noise reduction frequency bands can be superimposed to make the noise reduction frequency band distribution more uniform.

Therefore, without changing the original overall structure, we segmented the outer ring channel and obtained the structures of two-layer and three-layer outer ring channels, respectively. The final size is shown in Fig. 6, which achieves phase shift of multiple channels to enhance the sound suppression ability of the 500 Hz and 1000 Hz frequency bands.

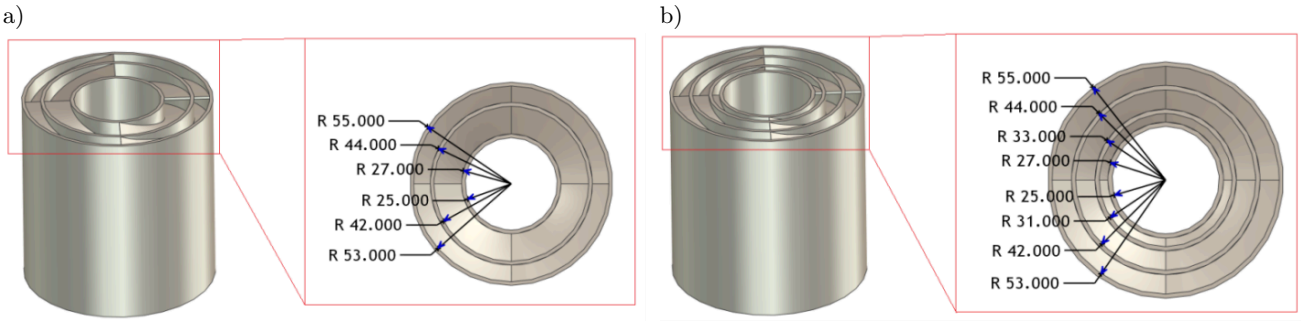


Fig. 6. Different outer ring channel structures: a) two-layer outer loop channel, b) three-layer outer loop channel.

3. Simulation analysis of acoustic performance of muffling unit

In order to more intuitively understand the coupling process and sound field distribution of sound waves generated by phase regulation of different acoustic channels at the exit, an acoustic metamaterial muffler model was established, and a simulation analysis of the sound pressure level and sound field was carried out by Harmonic Acoustics of ANSYS workbench 2020.

3.1. Simulation model, mesh, and parameter settings

The model medium is air, with the sound velocity set at 550 m/s. The acoustic input is set at the inlet end, and the radiation boundary is set at the outlet end. All walls are assumed to be rigid walls. The model is divided

into grids, with the grid type being tetrahedral grids and the grid size being 5 mm. The total number of grid elements is 164 000. The structure and grid of acoustic metamaterials muffling unit model is shown in Fig. 7. The model parameter is shown in Table 1. And monitor the sound field difference in the red wireframe area, as well as detect the sound pressure level difference between the outlet end and the inlet end to obtain the sound transmission loss.

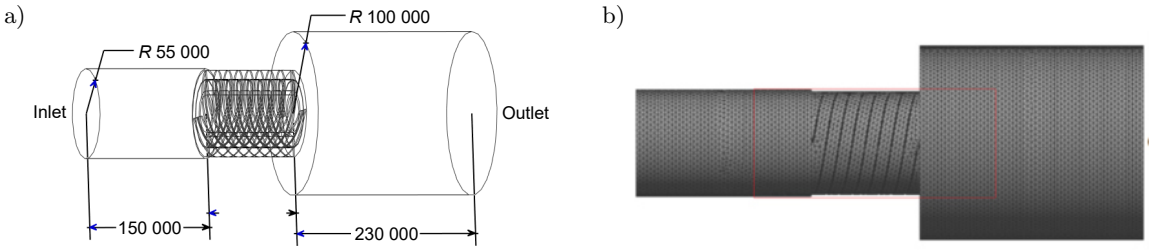


Fig. 7. Structure and grid of acoustic metamaterial muffling unit model: a) structure model, b) grid model.

Table 1. Model parameter.

Medium	Speed of sound	Inlet	Outlet	Wall type	Mesh type	Element size	Element count
Air	550 m/s	Acoustic excitation	Radiation boundary	Rigid	Tetrahedral	5 mm	164 000

3.2. Sound field analysis of noise reduction unit

Figure 8 shows the sound pressure level distribution of the single channel acoustic metamaterial unit and its vicinity at different frequencies. The phase difference between the plane wave at the entrance of the acoustic metamaterials and the acoustic wave vibration at the central through-hole after passing through the helical structure is generated. Since the length of the spiral channel is a fixed value, according to Eq. (3), different phase control results will be produced for waves of different frequencies. When there is a significant phase difference between the radiating wave at the exit of the central channel and the spiral channel, the result is shown in Fig. 8a. The destructive Fano interference generated by this strong coupling phenomenon significantly weakens the radiation of the sound wave in the far field, thus preventing the transmission of the sound wave. However, when there is no obvious phase difference or the phase difference is close to 2π, as shown in Fig. 8b, the sound waves of different acoustic channels can still produce the same direction upward vibration, resulting in the reduction of the noise reduction effect.

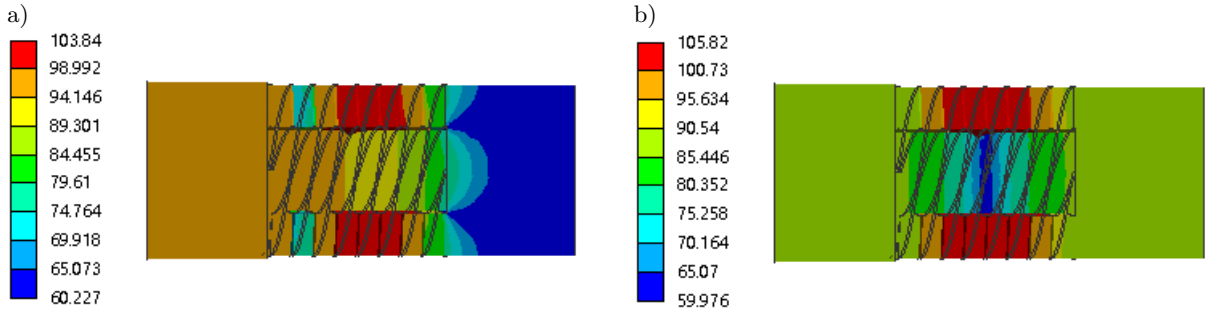


Fig. 8. Sound pressure level nephogram of single spiral channel at different frequencies: a) 760 Hz, b) 520 Hz.

For multi-channel structures, as shown in Fig. 9, different spiral radii have different acoustic refractive indices, resulting in different phase regulation results for plane incident waves at the same frequency, corresponding to different acoustic attenuation frequencies, which also makes the vibration environment more complex. Figure 9a shows the sound field distribution when there is a significant phase difference between the outer ring acoustic channel and the central channel, and when the coupled vibration is significantly different from the vibration of the inner ring acoustic channel. As shown in Fig. 9b, it is the sound field distribution when there is a significant phase difference between the inner ring acoustic channel and the central channel, and the coupled vibration is

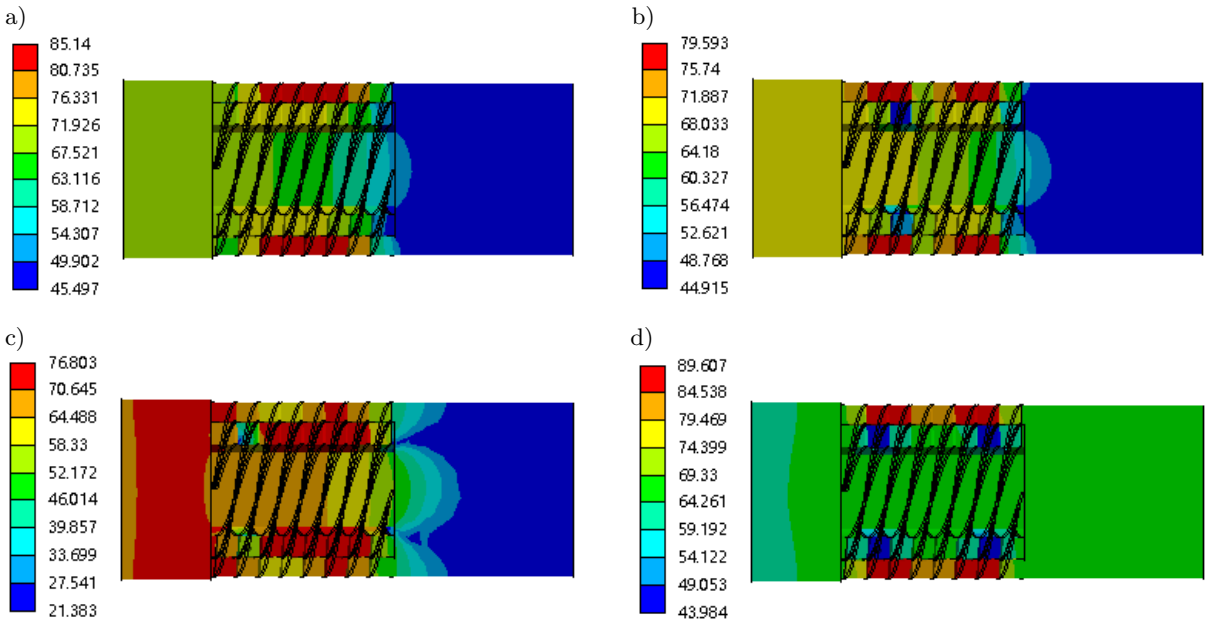


Fig. 9. Sound pressure level nephogram of double helix channel at different frequencies: a) 540 Hz, b) 920 Hz, c) 800 Hz, d) 1020 Hz.

significantly different from the vibration of the outer ring acoustic channel. In addition, there is also the influence of multi-channel coupled vibration, as shown in Fig. 9c, which refers to the sound field distribution when there is a significant phase difference between the inner ring and the outer ring acoustic channel, and when there is a significant phase difference between the coupled vibration and the central acoustic channel. Similarly, coupling vibration caused by small phase difference between the phase control result and the central channel may also occur, which greatly weakens the sound reduction effect, as shown in Fig. 9d, which refers to the resonance generated when the phase difference of each acoustic channel is a 2π multiple, reducing the difference in sound pressure between the outlet and the inlet and greatly reducing the sound reduction effect.

3.3. Transmission loss analysis of muffling unit

For three-channel acoustic metamaterials, the vibration becomes more complex, which greatly increases the difficulty of theoretical calculation and sound field analysis. Therefore, the acoustic simulation of transmission loss of different structure acoustic units is carried out to make a comparative analysis. As shown in Fig. 10a, for the case of a single track, due to the high central muffler frequency and the asymmetry of Fano interference, no obvious muffler below 740 Hz can be formed under the design structure size, which means that to achieve a lower frequency muffler, the length of the structure needs to be significantly increased. In order to match the impedance, it is also necessary to significantly increase the diameter of the structure, resulting in an increase in the size of the noise reduction structure. In the case of dual outer loop channels, as shown in Fig. 10b, although a narrowband passing frequency is increased compared with that of a single channel, the noise attenuation ability between 520 Hz and 730 Hz is significantly enhanced, thus obtaining a lower noise attenuation frequency. However, there are still some differences in the noise reduction intensity between 500 Hz to 740 Hz and 740 Hz to 1000 Hz. Through a large number of simulations, a three-channel sound attenuation structure can be obtained when the impedance of each channel is reasonably matched, and its transmission loss is shown in Fig. 10c. Although a narrowband passing frequency is added again, the transmission loss between 500 Hz and 1000 Hz is relatively uniform, and the width of each passing frequency is also reduced.

It can be seen that the three-channel noise reduction structural unit can obtain a lower noise reduction frequency on a wider noise reduction bandwidth, and then meet the noise reduction target requirements of 500 Hz to 1000 Hz. The dimensions of the acoustic metamaterial muffler were obtained as follows: inner diameter r_1 = 2.5 cm, outer diameter r = 5.5 cm, height l = 11.2 cm, spacer radii were 32 cm and 41 cm, respectively, and

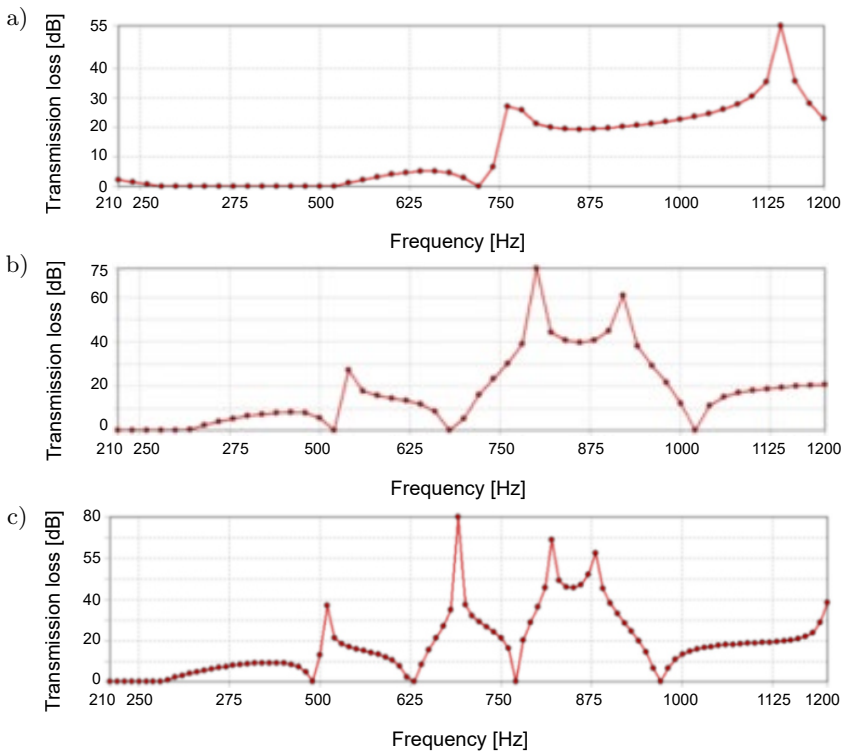


Fig. 10. Transmission loss of different spiral channel muffler chips at 210 Hz to 1200 Hz: a) single helix, b) double helix, c) triple helix.

spiral rotation angle was 3.6π. The structure has a transmission loss of more than 5 dB in the 93 % target noise attenuation frequency bandwidth, a transmission loss of more than 10 dB in the 85 % target noise attenuation frequency bandwidth, and a transmission loss of more than 20 dB in the 30 % target noise attenuation frequency bandwidth.

3.4. Multistage phase difference transfer loss analysis

The acoustic metamaterial noise reduction unit can also achieve multi-stage phase difference control at higher frequencies, so as to obtain the corresponding noise reduction ability, but this also makes the transmission loss curve more complicated. For this purpose, we conducted the test as shown in Fig. 11.

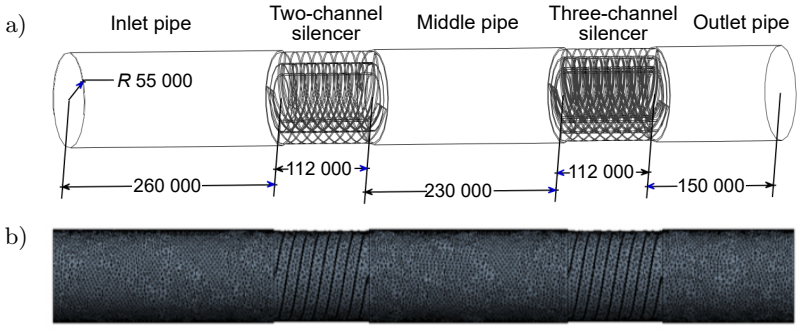


Fig. 11. Structure and grid of acoustic metamaterial muffling unit: a) structure model, b) grid model.

As shown in Fig. 12, when the phase difference of the sound field is $(2n+1)\pi$, the dipole vibration condition as shown in Eq. (11) can still be formed, which significantly increases the transmission loss of the acoustic metamaterial muffling unit, which means that the phase-regulated acoustic metamaterials can maintain effective muffling ability in a wider frequency range. It has obvious advantages over narrow band absorption with the fixed local resonance structure.

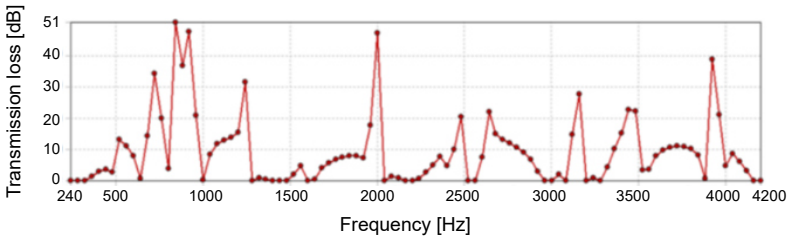


Fig. 12. Transmission loss at 240 Hz to 4200 Hz of three channel muffler chip.

3.5. Transmission loss analysis of combined structure of muffling unit

Through a number of different structures of the noise units in series form 1D linear combination, can also achieve a wider range of noise or in the rated target frequency band to obtain stronger noise reduction effect and more stable transmission loss. As shown in Fig. 13, when the three-channel muffling unit and the two-channel muffling unit are combined in series, due to the difference in the passing frequency of the two structural sounds, the three-channel muffling unit can be blocked by the two-channel muffling unit at the passing frequency and cannot radiate to the external field through a reasonable design. The vibration of the two-channel muffler through the frequency is greatly weakened at the entrance of the two-channel muffler due to the obstruction of the three-channel muffler, and it cannot radiate to the external field. Therefore, in the range of the target muffler frequency band, the passing frequency of the sound wave completely disappears, and a transmission loss of the full frequency band greater than 10 dB is obtained.

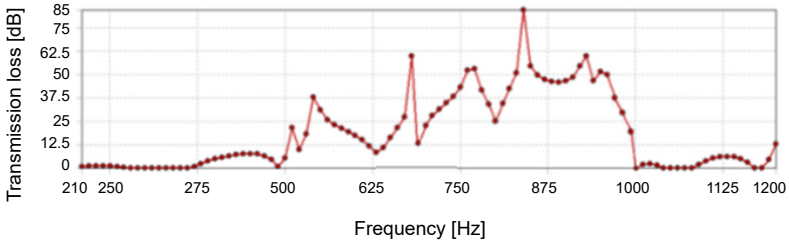


Fig. 13. Transmission loss of three channel and two channel muffling unit combination at 210 Hz to 1200 Hz.

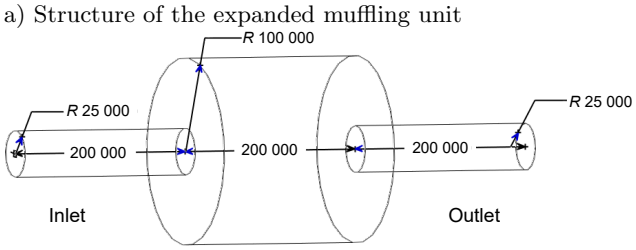
4. Simulation analysis of flow field of muffling unit

4.1. Exhaust resistance analysis of muffling unit

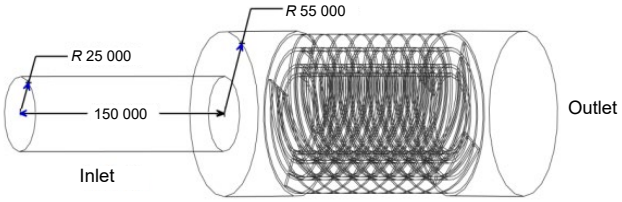
According to the principle of internal combustion engine work, too high exhaust resistance on the one hand, will increase the piston pump gas work loss, on the other hand, will lead to cylinder exhaust pressure rise, residual exhaust gas increase, which will have an adverse effect on the fuel consumption rate and effective power of the internal combustion engine, therefore, the muffler must have good gas passage and small exhaust resistance.

In order to understand the fluid mechanics performance of the acoustic metamaterials muffling unit, a comparative analysis was conducted with the most common and simple-structured expanded muffling unit (expansion chamber size: radius 10 mm, length 20 mm) as a reference. As shown in Fig. 14a and Fig. 14b, the flow fields of the traditional expanded muffling unit and the acoustic metamaterial muffling unit are compared and analyzed through the fluent fluid dynamics analysis module in ANSYS Workbench 2020. The single-phase flow model is adopted. The expanded muffler uses a hexahedral unit grid with a grid size of 5 mm and a number of 55 000 grids, with a grid aspect ratio of 2.17. The acoustic muffling unit uses a tetrahedral to polyhedral unit grid with a grid size of 5 mm and a number of 40 400 grids, with a grid aspect ratio of 13.4. The grid models obtained are shown in Fig. 14c and Fig. 14d. The turbulence model adopts the standard $k-\varepsilon$ two-path model; when calculating, the inlet boundary condition is the velocity inlet, and the average inlet velocity is 70 m/s. The outlet boundary condition is a pressure outlet, and the wall boundary condition is an adiabatic non-slip wall. The key parameters for flow field simulation of muffling units is shown in Table 2.

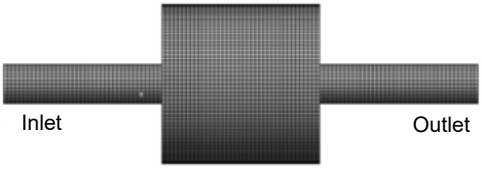
a) Structure of the expanded muffling unit



b) Structure of the acoustic metamaterial muffling unit



c) Grid of the expanded muffling unit



d) Grid of the acoustic metamaterial muffling unit

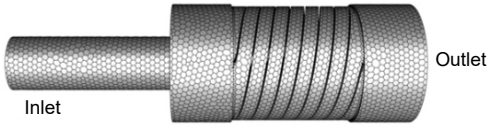


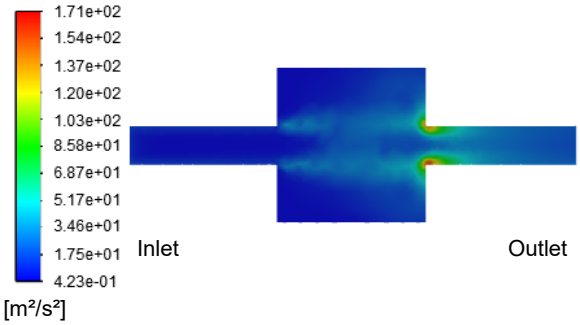
Fig. 14. Structure and grid of expanded muffling unit and acoustic metamaterial muffling unit.

Table 2. Key parameters for flow field simulation of muffling units.

Parameter type	Expanded muffling unit	Acoustic metamaterials muffling unit
Physical model	Single-phase flow model	Single-phase flow model
Mesh type	Hexahedral element mesh	Tetrahedral to polyhedral element mesh
Mesh size	5 mm	5 mm
Number of meshes	55 000	40 400
Mesh aspect ratio	2.17	13.4
Turbulence model	Standard $k-\varepsilon$ two-equation model	Standard $k-\varepsilon$ two-equation model
Inlet boundary condition	Velocity inlet, average velocity 70 m/s	Velocity inlet, average velocity 70 m/s
Outlet boundary condition	Pressure outlet	Pressure outlet
Wall boundary condition	Adiabatic non-slip wall	Adiabatic non-slip wall
Corresponding diagrams	Fig. 14a, Fig. 14c	Fig. 14b, Fig. 14d

As shown in Fig. 15, the cloud diagram of turbulent kinetic energy, when the inlet pipe diameter is the same, the energy loss of the expanding muffling unit mainly occurs at the outlet pipe of the cavity, which causes the change of the direction of the gas velocity vector at the outlet to a certain extent, causing resistance to gas flow. As the expansion length and diameter of the acoustic metamaterial muffler are obviously smaller than that of the expansion type muffler, the turbulence at the exit is not obvious, and the main turbulence is located at the end face of the outer ring. However, since the end face of the outer ring is a spiral ventilation structure, allowing a certain airflow to pass through, the turbulence intensity is significantly lower than that of the expansion type muffler. The energy dissipation of the acoustic metamaterial muffler is obviously lower than that of the expansion type muffler under the condition of pipeline gas flow, indicating that the structure is conducive to reducing the

a)



b)

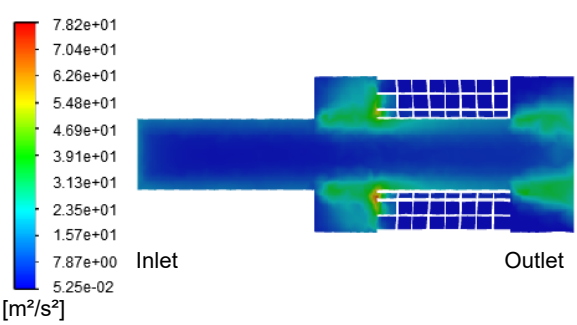


Fig. 15. Turbulent kinetic energy nephogram of different muffling units:
a) expanded muffling unit, b) acoustic metamaterial muffling unit.

transmission resistance. Meanwhile, the lower turbulent kinetic energy indicates that the acoustic metamaterial muffler is not easy to produce strong aerodynamic regeneration noise.

As can be seen from Fig. 16, after the gas in the inlet pipe of the expansion muffler enters the expansion chamber, the speed decreases significantly, which consumes a large amount of gas kinetic energy, resulting in a decrease in the dynamic pressure inside the pipeline. The pressure inside the chamber needs to be re-established to make the gas flow out quickly. This allows the kinetic energy of the gas in the pipeline to be preserved and maintains a high dynamic pressure of the gas, which is conducive to reducing the exhaust resistance.

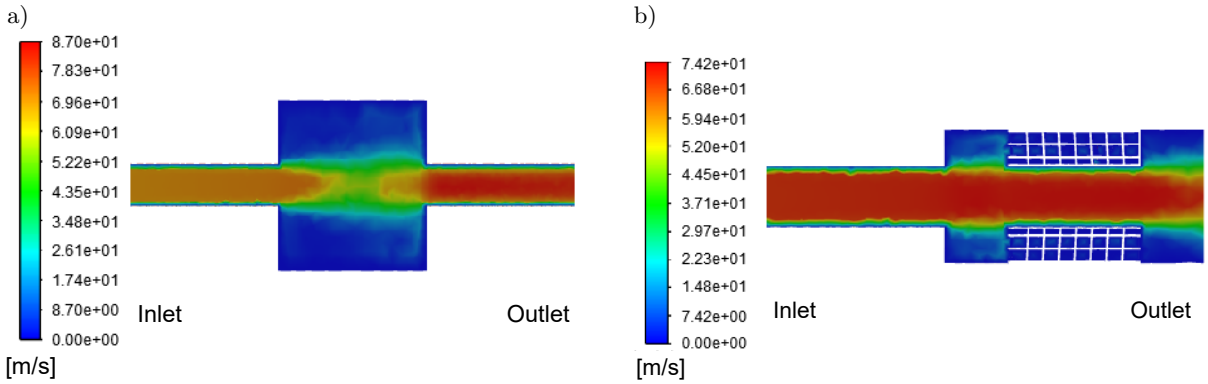


Fig. 16. Velocity nephogram of different muffling units: a) expanded muffling unit, b) acoustic metamaterial muffling unit.

By comparing the pressure loss of the two muffler devices, as shown in Fig. 17, under the condition of 70 m/s, the exhaust resistance of the expanded muffling unit can reach 2280 Pa, while the exhaust resistance of the acoustic metamaterials muffling unit is only 360 Pa. It can be seen that the volume of the acoustic metamaterials muffling unit is not only significantly smaller than that of the traditional muffling unit, but also that of the acoustic metamaterial muffling unit. Moreover, there are obvious advantages in exhaust resistance. It can be seen that while enhancing the noise reduction performance, the metamaterial muffler maintains the high kinetic energy characteristics of the gas in the central channel, avoiding the energy loss caused by the process of first dissipating and then establishing the gas kinetic energy in the expansion muffler, thereby suppressing the increase of equipment resistance and system back pressure. This is conducive to avoiding the increase in residual waste and the reduction of the charge coefficient caused by the poor exhaust of the engine due to the increased demand for noise reduction capacity, and can also reduce the pumping loss of the engine, ultimately achieving an improvement in the power and efficiency of the engine.

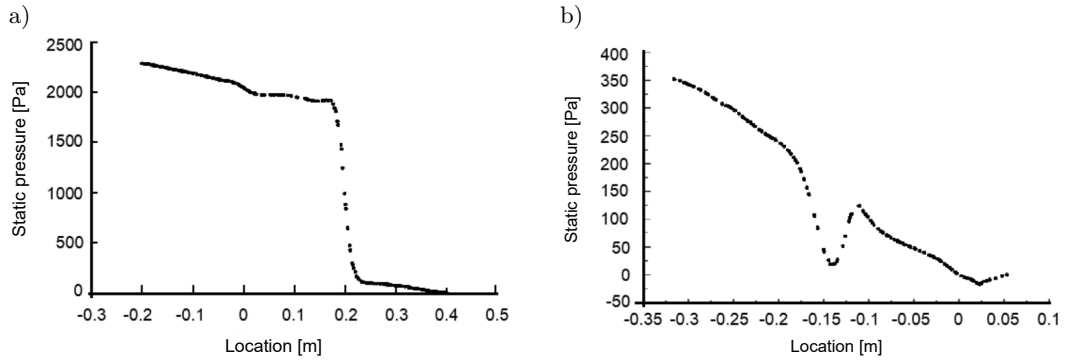


Fig. 17. Pressure loss on axis of different muffling units: a) acoustic metamaterial muffling unit, b) expanded muffling unit.

4.2. Analysis of exhaust resistance in series between traditional muffling unit and acoustic metamaterial muffling unit

For the muffler system, the best muffler effect often comes from the series muffler structure. Since the size parameters of the acoustic metamaterial muffling unit are more than the number of constraint formulas, the struc-

ture design has a multi-degree of freedom dimension relationship, which provides good conditions for the series with the traditional muffling unit. The series can be in the form of two muffler devices closely connected in series or the acoustic metamaterial muffling unit directly installed at the end of the pipeline, in which the exhaust resistance is the sum of the exhaust resistance of the two muffler structures directly installed at the end of the pipeline. Therefore, simulation is only carried out on the connected and series structure, as shown in Fig. 18. Two types of muffling units are connected in series without sacrificing too large volume of muffler. One end of the acoustic metamaterial muffler is connected to the inlet of the expansion muffler, the other end is connected to the air inlet, and the outlet of the expansion muffler is connected to the exhaust outlet to obtain the combined muffler structure.

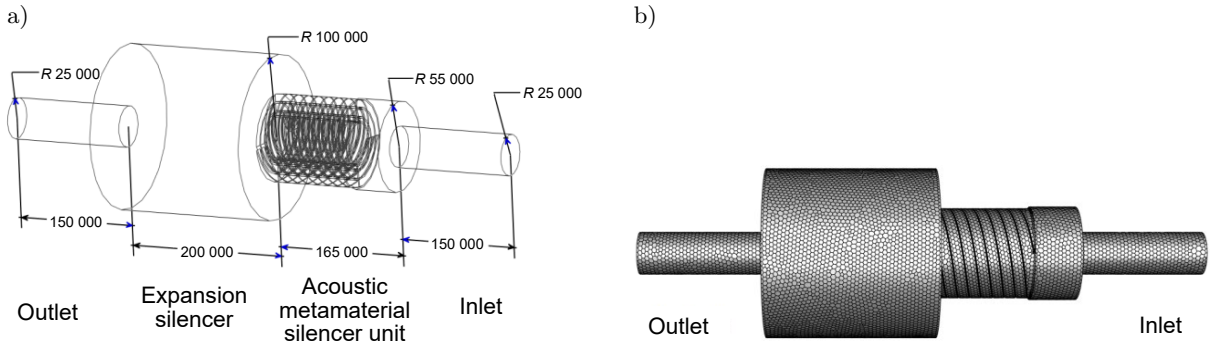


Fig. 18. Structure (a) and grid (b) of combined muffler models.

The velocity and turbulent kinetic energy cloud images are shown in Fig. 19 and Fig. 20. It can be seen that gas maintains a high flow speed in the pipe of the acoustic metamaterial muffler. At the same time, the side wall of the acoustic metamaterial muffler has a certain airflow passage, thus weakening the turbulence of gas entering the expansion type and reducing energy dissipation, so that the flow speed of gas in the axis of the expansion chamber is slightly higher than that of the independent expansion muffler structure. To a certain extent, this is conducive to maintaining the kinetic energy of the gas and reducing the exhaust loss, but as a result, the gas obtains a higher speed at the exit, and also forms a stronger turbulence, resulting in a certain energy dissipation.

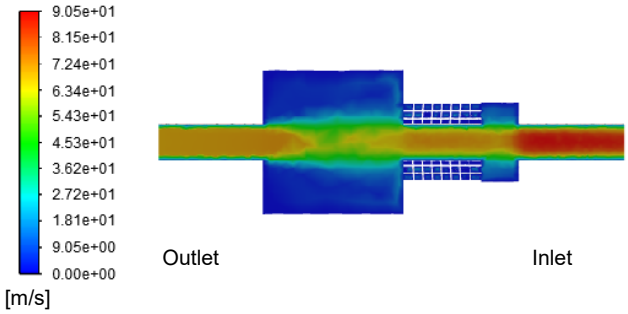


Fig. 19. Velocity nephogram of combined muffler.

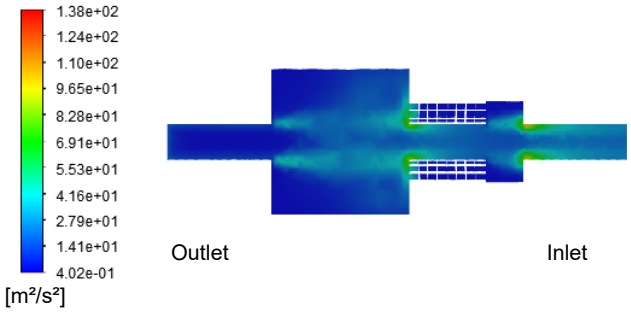


Fig. 20. Cloud diagram of turbulent kinetic energy of combined muffler.

As shown in Fig. 21, the overall curve is consistent with but slightly higher than that of the expansion muffler. At 70 m/s, the exhaust loss is 2650 Pa. Compared with the 2280 Pa exhaust loss of the expansion muffler structure alone, the exhaust resistance of the acoustic metamaterial muffling unit only increases 370 Pa, indicating that in the series structure, the exhaust resistance of the acoustic metamaterial muffling unit only increases 370 Pa. Compared with when used alone, in the case of combined use, the resistance increase caused by the acoustic metamaterial muffler remains linearly rising. Under combined use conditions, there will be no mutual interference of resistance, and only a very small resistance cost is needed to significantly improve the noise reduction performance of the muffler.

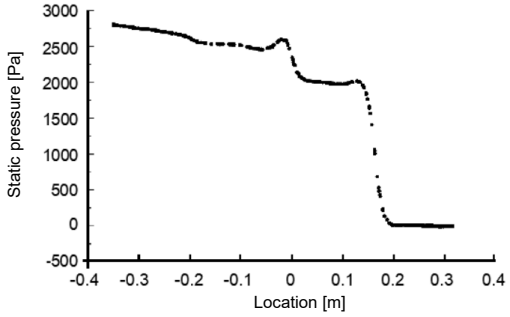


Fig. 21. Pressure loss on the axis of combined muffler.

5. Noise reduction performance test

5.1. Experimental conditions and methods

In order to test the acoustic performance of the acoustic metamaterial muffler, a single cylinder diesel engine model ZS1105 was used as a prototype to compare the noise difference before and after the original muffler was added to the acoustic metamaterial muffler.

Main equipment required for the experiment: one internal combustion engine testing machine, one muffler straight tube, one original muffler, one ceramic 3D printed three-channel acoustic metamaterial muffling unit, as shown in Fig. 22a, and one AWA6228+ multifunctional sound level meter of Qingdao Juchuang. The acoustic metamaterial muffler and the original muffler can be combined in various ways. In this experiment, the acoustic metamaterial muffler is installed directly at the end of the original muffler, as shown in Fig. 22b.

a) b)

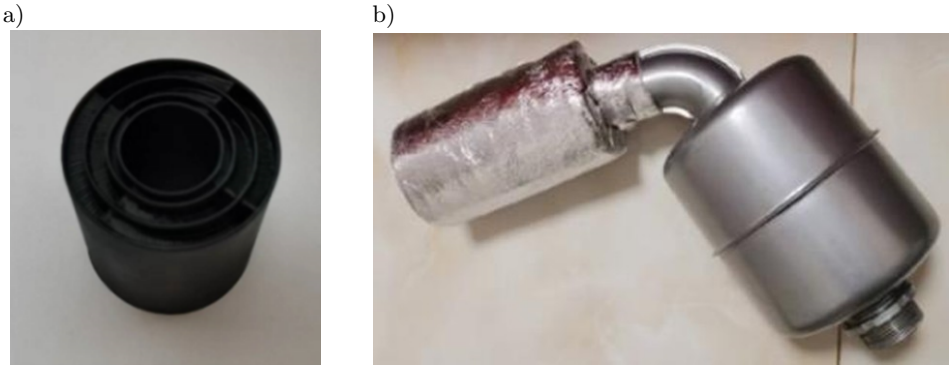


Fig. 22. Acoustic metamaterial muffling unit (a) and its combined structure with original muffler (b).

5.2. Lab process

The noise suppression performance test is carried out according to '7.2 Laboratory Measurement Method' in GB/T 4759-2009 'Internal Combustion Engine exhaust muffler – Measurement Method.' First, in order to reduce the impact of other types of diesel engine noise on the test results, a retaining wall is set up, the diesel engine and

the muffler are placed on both sides of the retaining wall, and then a hole is drilled in the retaining wall and the diesel engine and the muffler are connected through a straight pipe. After the diesel engine ran stably for 15 min, the diesel engine speed was set to 2200 r/min, and the exhaust noise signal was recorded with a sound level meter at 45° at the exhaust pipe outlet and 0.5 m away from the exhaust pipe outlet. The sampling frequency was 1/3 in the frequency range of 31.5 Hz to 4000 Hz, and two experiments were carried out. The measured results were averaged. The acoustic metamaterial muffling units installed in series with the straight pipe without muffler, the original muffler, and the original muffler were respectively tested for noise, as shown in Fig. 23.

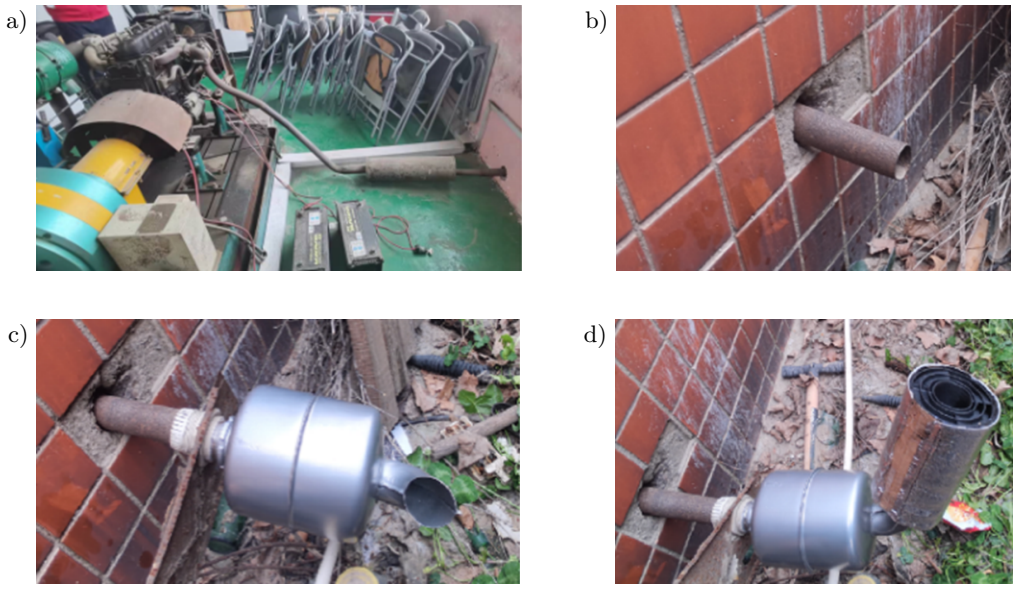


Fig. 23. Experimental test conditions: a) indoor testing machine, b) exhaust straight pipe, c) expanding muffler, d) combined muffler.

5.3. Performance test results

The noise test results are obtained, as shown in Fig. 24, when the speed is 2200 r/m, the 1/3 octave sound pressure level in the range of 25 Hz to 4000 Hz. The noise pressure level at different frequencies shows that the noise reduction effect of the original muffler starts to increase significantly after 1000 Hz, and before 1000 Hz, the transmission loss is obviously low, while after the series Acoustic metamaterial noise reduction unit, the noise reduction effect is not obvious below the frequency of 250 Hz. Between 250 Hz and 500 Hz, the transmission loss

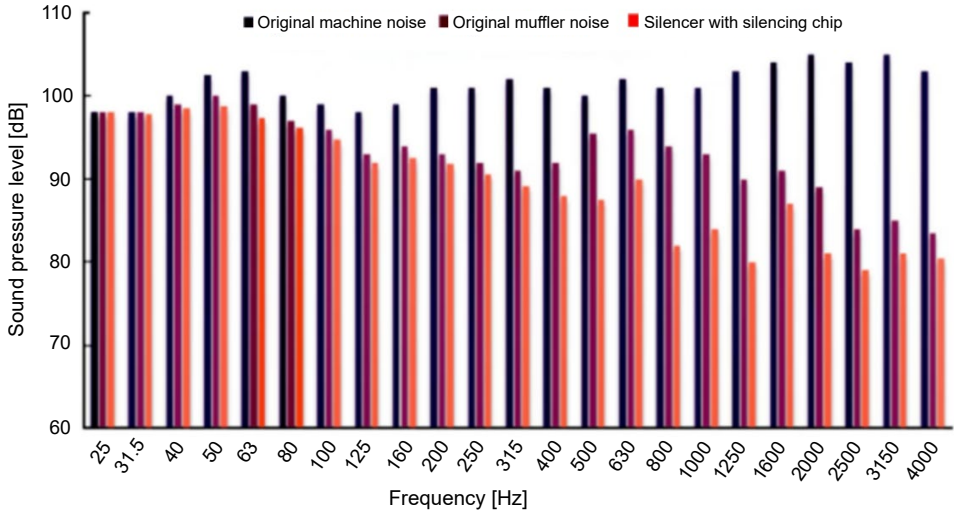


Fig. 24. 1/3 octave sound pressure level under different muffler conditions.

gradually increases, which is caused by the interference of sound waves to a certain extent caused by dipole vibration generated by different acoustic channels. At this stage, although the phase difference between the vibration phase of each channel and that of the central airflow channel is less than $\pi/2$, it gradually expands, showing that the transmission loss gradually increases and the noise reduction ability gradually increases. After 500 Hz, interference conditions are formed, and the transmission loss begins to increase significantly. In the target noise reduction frequency band, the insertion loss of 5 dB to 15 dB is formed. After 1000 Hz, the transmission loss gradually decreases as the phase difference becomes larger. It shows that the multipole vibration caused by the second and above phase difference of sound wave occurs in this frequency stage.

The final noise is compared with the loudness curve of NR90, as shown in Fig. 25. After the addition of the acoustic metamaterial muffling unit, the overall exhaust noise level is significantly lower than that of the N90 curve after the excessive noise at the stage of 500 Hz to 1000 Hz is eliminated, and the noise reduction of the frequency band with insufficient noise reduction under the original muffler condition is realized. Noise reduction that significantly exceeds the loudness curve can significantly reduce the loudness of noise emissions.

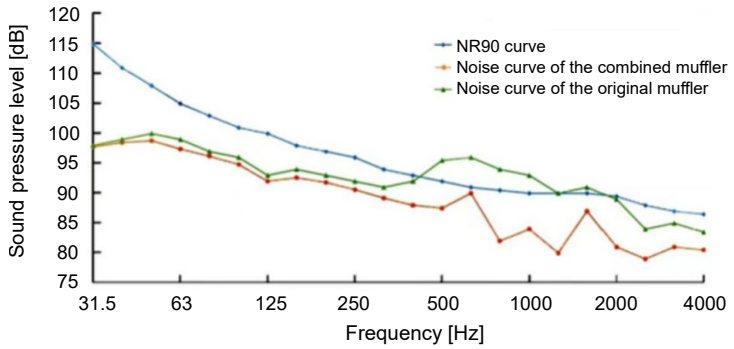


Fig. 25. Comparison between NR90 curve and noise curve after muffling.

It can be seen that the acoustic metamaterials can achieve noise reduction in the frequency band of 500 Hz to 1000 Hz, and the frequency band above 1000 Hz can achieve certain noise reduction ability through the interference of multi-order frequencies. Its advantage is that through structural adjustment and design, higher noise reduction ability can be obtained in the frequency where noise reduction is difficult. When the muffler is used alone, there will be obvious difference in transmission loss at different frequencies, but when combined with other muffler structures, it can carry out targeted muffler without significantly increasing the exhaust resistance, so as to solve the problem of insufficient muffler at some frequencies.

6. Conclusion

In this paper, according to the structural characteristics of internal combustion engine exhaust pipe and the performance requirements of exhaust noise reduction, the acoustic model is established by using the equivalent medium theory, and the phase control of sound field is realized by means of space winding, and the design of multi-channel helical metamaterials noise reduction unit is completed. The mechanism and performance of noise reduction are studied through acoustic simulation and experimental tests, and the following conclusions are drawn:

1. The acoustic metamaterial muffler adopts a spiral coil structure, which has a high space utilization rate, and adopts a multi-spiral channel structure, which reduces the asymmetric influence of Fano interference, and obtains obvious low and medium frequency muffler ability in the compact structure size.
2. The acoustic metamaterial unit realizes the change of acoustic refractive index in the equivalent acoustic region by means of phase regulation, and obtains the multi-pole vibration interference condition at the exit of the acoustic structure, which not only has obvious noise reduction effect at the target noise reduction frequency, but also can obtain obvious noise reduction ability at the high-order phase difference frequency. A stronger ability to eliminate sound without significant frequency can be obtained.

3. The acoustic metamaterial unit has a low exhaust resistance. Under the condition of single use or series use, the exhaust resistance is only 360 Pa to 370 Pa when the exhaust speed is 70 m/s, which has obvious advantages over the traditional noise reduction structure.

In summary, the multi-helical channel phase control acoustic metamaterial muffler has a good performance in terms of structure size, muffler performance, muffler frequency and exhaust resistance, and the structural design is flexible, so it will have a good application prospect in exhaust noise reduction.

Fundings

This work was supported by the National Natural Science Foundation of China (52475354) and the Industry-University Collaborative Education Project of the Ministry of Education (221001414112514).

Conflict of interests

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.

Authors' contributions

All authors contributed equally to this work, reviewed, and approved the final manuscript.

References

1. Araujo E.S., Pestana R.C. (2024), A numerical scheme based on the Taylor expansion and Lie product formula for the second-order acoustic wave equation and its application in seismic migration, *Geophysical Prospecting*, **72**(5): 1745–1763, https://doi.org/10.1111/1365-2478.13481.
2. Bai C., Wu F., Wang J., Jiang J. (2023), A broadband low-frequency muffler based on neural network method and Helmholtz resonator with helical neck, *Journal of Vibration and Control*, **29**(17–18): 3942–3951, https://doi.org/10.1177/10775463221107197.
3. Chen N. *et al.* (2024), Research on prediction of yellow flesh peach firmness using a novel acoustic real-time detection device and Vis/NIR technology, *LWT*, **209**: 116772, https://doi.org/10.1016/J.LWT.2024.116772.
4. Chen Z., Ji Z., Huang H. (2020), Acoustic impedance of perforated plates in the presence of fully developed grazing flow, *Journal of Sound and Vibration*, **485**: 115547, https://doi.org/10.1016/j.jsv.2020.115547.
5. Dalla Nora M., Lanzanova T.D.M., Zhao H. (2016), Effects of valve timing, valve lift and exhaust backpressure on performance and gas exchanging of a two-stroke GDI engine with overhead valves, *Energy Conversion and Management*, **123**: 71–83, https://doi.org/10.1016/j.enconman.2016.05.059.
6. Fang N. *et al.* (2006), Ultrasonic metamaterials with negative modulus, *Nature Materials*, **5**(6): 452–456, https://doi.org/10.1038/nmat1644.
7. Ghaffarivardavagh R., Nikolajczyk J., Anderson S., Zhang X. (2019), Ultra-open acoustic metamaterial muffler based on Fano-like interference, *Physical Review B*, **99**(2): 024302, https://doi.org/10.1103/PhysRevB.99.024302.
8. Guo J., Fang Y., Qu R., Zhang X. (2023), Development and progress in acoustic phase-gradient metamaterials for wavefront modulation, *Materials Today*, **66**: 321–338, https://doi.org/10.1016/j.mattod.2023.04.004.
9. Jiang T., Han Q., Li C. (2023), Topologically tunable local-resonant origami metamaterials for wave transmission and impact mitigation, *Journal of Sound and Vibration*, **548**: 117548, https://doi.org/10.1016/j.jsv.2022.117548.
10. Kheybari M., Ebrahimi-Nejad S. (2019), Locally resonant stop band acoustic metamaterial muffler with tuned resonance frequency range, *Materials Research Express*, **6**(2): 025802, https://doi.org/10.1088/2053-1591/aaed4b.
11. Khodabakhsh S., Movahhedy M.R., Mohammadi K. (2025), A Helmholtz resonator based on spiral neck acoustic metamaterial for noise reduction, *Applied Acoustics*, **240**: 110957, https://doi.org/10.1016/j.apacoust.2025.110957.
12. Kumar S., Lee H.P. (2020), Recent advances in acoustic metamaterials for simultaneous sound attenuation and air ventilation performances, *Crystals*, **10**(8): 686, https://doi.org/10.3390/cryst10080686.

13. Lee J.K., Oh K.S., Lee J.W. (2020), Methods for evaluating in-duct noise attenuation performance in a muffler design problem, *Journal of Sound and Vibration*, **464**: 114982, https://doi.org/10.1016/j.jsv.2019.114982.

14. Li L., Diao Y., Wu H., Jiang W. (2022), Thickness-designable acoustic metamaterial for passive phased arrays, *Applied Acoustics*, **198**: 108942, https://doi.org/10.1016/j.apacoust.2022.108942.

15. Li P., Yao S., Zhou X., Huang G., Hu G. (2014), Effective medium theory of thin-plate acoustic metamaterials, *The Journal of the Acoustical Society of America*, **135**(4): 1844–1852, https://doi.org/10.1121/1.4868400.

16. Li S., Lan Y., Hong L. (2024), A method to analyze the radiation characteristics of a liquid column resonance transducer based on fluid motion, *The Journal of the Acoustical Society of America*, **155**(1): 252–261, https://doi.org/10.1121/10.0024339.

17. Liao D., Cao H., Li S., Cheng W., Yan X., An C. (2024), Study on the control of flow field by resonance acoustic mixing technology for purification of high performance spherical HATO crystals, *Separation and Purification Technology*, **339**: 126688, https://doi.org/10.1016/j.seppur.2024.126688.

18. Lin D.K. *et al.* (2024), An acoustic impedance design method for tubular structures with broadband sound insulations and efficient air ventilation, *Applied Acoustics*, **220**: 109983, https://doi.org/10.1016/j.apacoust.2024.109983.

19. Liu M., Li P., Du Q., Peng P. (2019), Reflected wavefront manipulation by acoustic metasurfaces with anisotropic local resonant units, *EPL*, **125**(5): 54004, https://doi.org/10.1209/0295-5075/125/54004.

20. Liu Y. *et al.* (2020), Three-dimensional fractal structure with double negative and density-near-zero properties on a subwavelength scale, *Materials & Design*, **188**: 108470, https://doi.org/10.1016/j.matdes.2020.108470.

21. Lu T., Li Z., Zhang Y. (2025), Broadband low-transmission study of ventilation metasurfaces based on Archimedean spirals, *Applied Acoustics*, **228**: 110335, https://doi.org/10.1016/j.apacoust.2024.110335.

22. Martos F.J., Soriano J.A., Braic A., Fernández-Yáñez P., Armas O. (2023), A CFD modelling approach for the operation analysis of an exhaust backpressure valve used in a Euro 6 Diesel engine, *Energies*, **16**(10): 4112, https://doi.org/10.3390/en16104112.

23. Mohamad B., Karoly J., Zelentsov A., Amroune S. (2021), Investigation of perforated tube configuration effect on the performance of exhaust mufflers with mean flow based on three-dimensional analysis, *Archives of Acoustics*, **46**(3): 561–566, https://doi.org/10.24425/aoa.2021.138148.

24. Naify C.J., Chang C.M., McKnight G., Nutt S.R. (2012), Scaling of membrane-type locally resonant acoustic metamaterial arrays, *The Journal of the Acoustical Society of America*, **132**(4): 2784–2792, https://doi.org/10.1121/1.4744941.

25. Narayan S., Kaisan U.M., Abubakar S. (2021), A review of vibro-acoustic techniques for control of combustion engine noise, *Journal of Applied Engineering Science*, **19**(4): 880–885, https://doi.org/10.5937/JAES0-29675.

26. Orazbayev B., Fleuly R. (2020), Far-field subwavelength acoustic imaging by deep learning, *Physical Review X*, **10**(3): 031029, https://doi.org/10.1103/PhysRevX.10.031029.

27. Shaikh S. *et al.* (2024), Acoustic propagation and transmission loss analysis in shallow water of Northern Arabian Sea, *Journal of Marine Science and Engineering*, **12**(12): 2256, https://doi.org/10.3390/jmse12122256.

28. Su H., Wu P., Xue J., Zhang Y., Zhang H. (2020), Analysis of flow field characteristics and structure optimization of the split-stream rushing muffler for diesel engine, *Noise Control Engineering Journal*, **68**(1): 101–111, https://doi.org/10.3397/1/37688.

29. Sun M., Fang X., Mao D., Wang X., Li Y. (2020), Broadband acoustic ventilation barriers, *Physical Review Applied*, **13**(4): 044028, https://doi.org/10.1103/PhysRevApplied.13.044028.

30. Wan C. *et al.* (2023), A low-complexity multi-channel active noise control system using local secondary path estimation and clustered control strategy for vehicle interior engine noise, *Mechanical Systems and Signal Processing*, **204**: 110786, https://doi.org/10.1016/j.ymssp.2023.110786.

31. Wang C., Lu Z., Zhang K. (2012), Small-hole arrays of ceramic material manufactured by micro powder injection molding, *The International Journal of Advanced Manufacturing Technology*, **59**(9–12): 969–976, https://doi.org/10.1007/s00170-011-3570-5.

32. Wang C., Zhao H., Wang Y., Zhong J., Yu D., Wen J. (2024), Topology optimization of chiral metamaterials with application to underwater sound insulation, *Applied Mathematics and Mechanics*, **45**(7): 1119–1138, https://doi.org/10.1007/s10483-024-3162-8.

33. Xiao D.W., Cheng J.F., Zhang J.Z. (2013), Blind DOA estimation based on single acoustic vector hydrophone, *Advanced Materials Research*, **2428**(706–708): 678–681, https://doi.org/10.4028/www.scientific.net/AMR.706-708.678.

34. Xu C., Guo H., Chen Y., Dong X., Ye H., Wang Y. (2021), Study on broadband low-frequency sound insulation of multi-channel resonator acoustic metamaterials, *AIP Advances*, **11**(4): 045321, https://doi.org/10.1063/5.0047416.

35. Xu Z., Fang Y., Wang Z. (2025), Enhanced fano-type broadband acoustic ventilated muffler with arbitrary geometrical configurations, *Frontiers in Physics*, **13**: 1586858, https://doi.org/10.3389/fphy.2025.1586858.

36. Yang M., Ma G., Yang Z., Sheng P. (2013), Coupled membranes with doubly negative mass density and bulk modulus, *Physical Review Letters*, **110**(13): 134301, https://doi.org/10.1103/PhysRevLett.110.134301.

37. Yang M., Ma G., Wu Y., Yang Z., Sheng P. (2014), Homogenization scheme for acoustic metamaterials, *Physical Review B*, **89**(6): 064309, https://doi.org/10.1103/PhysRevB.89.064309.

38. Ye Y.-R., Wang X.-P., Chen T.-N., Chen Y.-Y. (2020), Step-by-step structural design methods for adjustable low-frequency sound insulation based on infinite plate-type acoustic metamaterial panel, *Modern Physics Letters B*, **34**(21): 6–18, https://doi.org/10.1142/S0217984920502206.

39. Zeng H.-C., Luo C.-R., Chen H.-J., Zhai S.-L., Ding C.-L., Zhao X.-P. (2013), Flute-model acoustic metamaterials with simultaneously negative bulk modulus and mass density, *Solid State Communications*, **173**: 14–18, https://doi.org/10.1016/j.ssc.2013.08.017.

40. Zhang Y., Wu P., Ma Y., Su H., Xue J. (2018), Analysis on acoustic performance and flow field in the split-stream rushing muffler unit, *Journal of Sound & Vibration*, **430**: 185–195, https://doi.org/10.1016/j.jsv.2018.04.025.

41. Zhang X., Zhao C., Wang P., Chen R. (2024), Combined acoustic metamaterial design based on multi-channel Fano resonance effect, *Journal of Applied Physics*, **136**(1): 013105, https://doi.org/10.1063/5.0211050.

ARCHIVES OF ACOUSTICS Vol. 51, No. 1, pp. 25–36 (2026), https://doi.org/10.24423/archacoust.2026.4290

Research Paper

Analysis on Modal Distribution and Modal Density-Based Crossover Frequency in Cabin-Sized Enclosures

Ziyu WANG[(1)], Liangfen DU[(2)], Guangzheng YU[(1)*]

[(1)] *Acoustic Laboratory, School of Physics and Optoelectronics, South China University of Technology* Guangzhou, China

[(2)] *Department of Building Environment and Energy Engineering, Hong Kong Polytechnic University* Hung Hom, Hong Kong

*Corresponding Author: scgzyu@scut.edu.cn

Received July 30, 2025; revised November 14, 2025; accepted November 24, 2025; available online January 16, 2026; version of record February 10, 2026; published issue March 27, 2026.

In the sound field simulation of cabin-sized enclosures, the Schroeder frequency (SF) is still employed to estimate the crossover frequency (CF) that determines the validity ranges of wave-based and geometrical acoustic methods. However, because cabin-sized enclosures exhibit distinct modal behaviors from typical medium- and large-scale rooms, the validity of SF in such enclosures has not been thoroughly tested. This study introduces the modal density-based crossover frequency (MDCF) to systematically evaluate the applicability of SF in cabin-sized enclosures. The MDCF employs the same dense modal criterion as SF. However, its modal parameters, are derived from the numerical eigenfrequency analysis. This contrasts with the SF formula, where these parameters are determined solely by the room volume and reverberation time. Ten models are constructed for evaluation, grouped into two volume sets: 8 m^3 (cabin-sized) and 80 m^3 (common-sized). Each set comprises five distinct geometrical shapes from rectangular models to simplified vehicle shapes. The results reveal that, for cabin-sized enclosures under low absorption boundary conditions, the MDCF is typically 70 Hz to 150 Hz lower than SF; the discrepancies decrease to 20 Hz to 50 Hz in 80 m^3 rooms. Furthermore, the MDCF varies with room shapes at a constant volume, while the SF remains nearly unchanged. These findings demonstrate that MDCF provides a more reliable CF estimation for rooms with irregular shapes, highlighting the importance of considering accurate modal parameters in the acoustic analysis of cabin-sized models.

Keywords: vehicle acoustics, room acoustic modes, crossover frequency, Schroeder frequency.

Notations

B_{HP} – half-power bandwidth,
$\widetilde{B}_{HP}$ – simplified half-power bandwidth,
$B_{HP}(f_{n,c})$ – n-th half-power bandwidth,
$\mathbf{C}$ – damping matrix,
c_0 – speed of sound,
f_{MDCF} – modal density-based crossover frequency,
f_n – n-th eigenfrequency under rigid boundary (or the real part of n-th eigenfrequency $f_{n,c}$),
$f_{n,c}$ – n-th eigenfrequency
f_{SF} – Schroeder frequency,
f_u – upper limit frequency,
k – wavenumber,
$\mathbf{K}$ – stiffness matrix,
L – total edge length,
L_x, L_y, L_z – room dimensions,
M – modal overlap degree,
$\mathbf{M}$ – mass matrix,
N_f – modal number,
$\widetilde{N}_f$ – simplified modal number,
n_x, n_y, n_z – modal orders along each axis,
p – sound pressure,
S – total surface area,
T_{60} – reverberation time,
V – volume,

Z – boundary acoustic impedance,
α – absorption coefficient,
δ – modal decay factor,
$\widetilde{\delta}$ – simplified modal decay factor,
$\delta_{n,c}$ – n-th modal decay factor,
Δf_n – eigenfrequency spacing,
$\Delta\widetilde{f}_n$ – simplified eigenfrequency spacing,
ρ – air density,
ζ – normalized boundary impedance, $\zeta = \zeta_i + j\zeta_r$.

1. Introduction

In the study of sound field simulation and analysis for cabin-sized enclosures, the crossover frequency (CF) serves as a critical threshold that defines the optimal transition between wave-based and geometrical acoustic methods (Kleiner, Tichy, 2014; Rougier, 2018; Siltanen *et al.*, 2010). Below the CF, the sound field is dominated by a limited set of discrete normal modes, where the wave-based methods such as the finite element method (FEM) can provide accurate predictions (Sakuma *et al.*, 2014). Above the CF, the modal density increases, leading to a more even distribution of sound energy; therefore, geometrical acoustic methods – ray tracing methods – provide accurate predictions with greater computational efficiency (Savioja, Svensson, 2015; Savioja, Xiang, 2019).

In practice, CF is commonly estimated by the classical Schroeder frequency (SF) formula, which has been validated mainly in the medium- and large-scale sound fields (Southern *et al.*, 2011; 2013; Summers *et al.*, 2004). Based on the statistic acoustic theory, the SF is determined by two parameters: the room volume V, which describes the modal distribution behavior, and the reverberation time T_{60}, which characterize the decay rates of all modes (Schroeder, Kuttruff, 1962; Schroeder, 1987; 1996). Both the modal distribution and the modal decay properties serve as critical parameters for the theoretical determination of CF, and SF offers a convenient indicator for estimating CF (Brinkmann *et al.*, 2019).

However, the simplifications and assumptions of modal properties in the derivation of SF impose limitations. The simplified modal distribution formula – originally derived from the distribution of oblique modes in rectangular rooms – approximates the modal distribution of enclosures with different shapes solely through their volumes V (Bolt, 1946; Morse, Uno Ingard, 1968); it not only undercounts modes but also neglects the effects of room geometry on modal distribution (Gunawan, Aditanoyo, 2018; Meissner, 2021). The reverberation time T_{60}, which characterizes the modal decay rate, is typically assumed to be a frequency-independent value – estimated either from an average absorption coefficient in analytical simulation or from the 1 kHz one-third octave band in measurements – to yield a specific SF value (Kuttruff, 2016; Nélisse, Nicolas, 1997). Such a single global parameter T_{60}, however, is insufficient to capture the decay behaviors of individual modes. Because the decay rate of each mode depends on how it interacts with the damping boundaries: modes that encounter more damping surfaces exhibit faster energy decay than those encountering fewer damping surfaces (Bastine *et al.*, 2021; Zheng *et al.*, 2021). For instance, in rectangular rooms, the oblique modes decay more rapidly than the axial and tangential modes, and similar variations are also observed in non-rectangular enclosures.

The limitations from the simplified modal parameters become more pronounced in cabin-sized enclosures. In such enclosures, the number of modes within a given frequency band is limited and their distribution is sparse (Kleiner, Tichy, 2014; Meissner, 2017). Therefore, any underestimation using a volume-based formula can lead to substantial relative errors. Moreover, the premise of reverberation time T_{60} is often invalid in cabin-sized enclosures due to the rapid interaction of sound waves with damping boundaries, which causes faster energy decay (Ferreira *et al.*, 2016; Rindel, 2015). As a result, these simplified modal parameters introduce significant errors into the modal analysis of cabin-sized enclosures. Despite these limitations, the SF remains a reference for estimating the CF in several vehicle acoustics studies (Aretz, Vorländer, 2014a; 2014b; Granier *et al.*, 1996; Pinardi *et al.*, 2021). Hence, it is necessary to reassess the validity of using SF to estimate CF, as its applicability in this context is insufficiently established.

To address these issues, this study analyzes the modal characteristics of cabin-sized enclosures, and proposes the modal density-based crossover frequency (MDCF) for CF estimation using accurate modal parameters. Ten models – comprising two volume sets (8 m^3 and 80 m^3) and five distinct geometries per set – are simulated via the FEM to obtain accurate eigenfrequencies and modal parameters. By comparing the numerically derived modal

spacing and bandwidth with analytical solutions, we quantify the discrepancies between MDCF with SF, and elucidate the mechanisms underlying these differences in cabin-sized enclosures.

The paper is organized as follows: Sec. 2 introduces the research models, including the model shapes, boundary conditions, and numerical simulation setup; Sec. 3 reviews the analytical formulas and numerical simulations of modal parameters, details the derivation process of the SF and MDCF, and presents the relations between them; Sec. 4 compares the results between analytical (SF-based) and numerical (MDCF-based) modal parameters across different geometries and boundary conditions, and further compares the MDCF and SF. Finally, Sec. 5 concludes with the main findings and academic contributions of MDCF in vehicle acoustics studies.

2. Research models

The research models comprise two sets of proportionally scaled models with a volume ratio of 1:10, the small-scale set has a volume of $8\,\mathrm{m}^3$, representing vehicle cabins, while the large-scale set has a volume of $80\,\mathrm{m}^3$, representing ordinary meeting rooms. Each volume set contains five geometrical configurations: three rectangular rooms with dimension ratios ($L_x : L_y : L_z$) as follows – model A (1.00 : 3.00 : 4.00), model B (1.00 : 1.50 : 2.50), and model C (1.00 : 1.20 : 1.45) (Rindel, 2021), and two chamfered models: model D and model E, created by applying diagonal cuts to model B (similar dimensions to actual cabins). Figure 1 provides the exact dimensions and geometries for the five small-scale models ($8\,\mathrm{m}^3$).

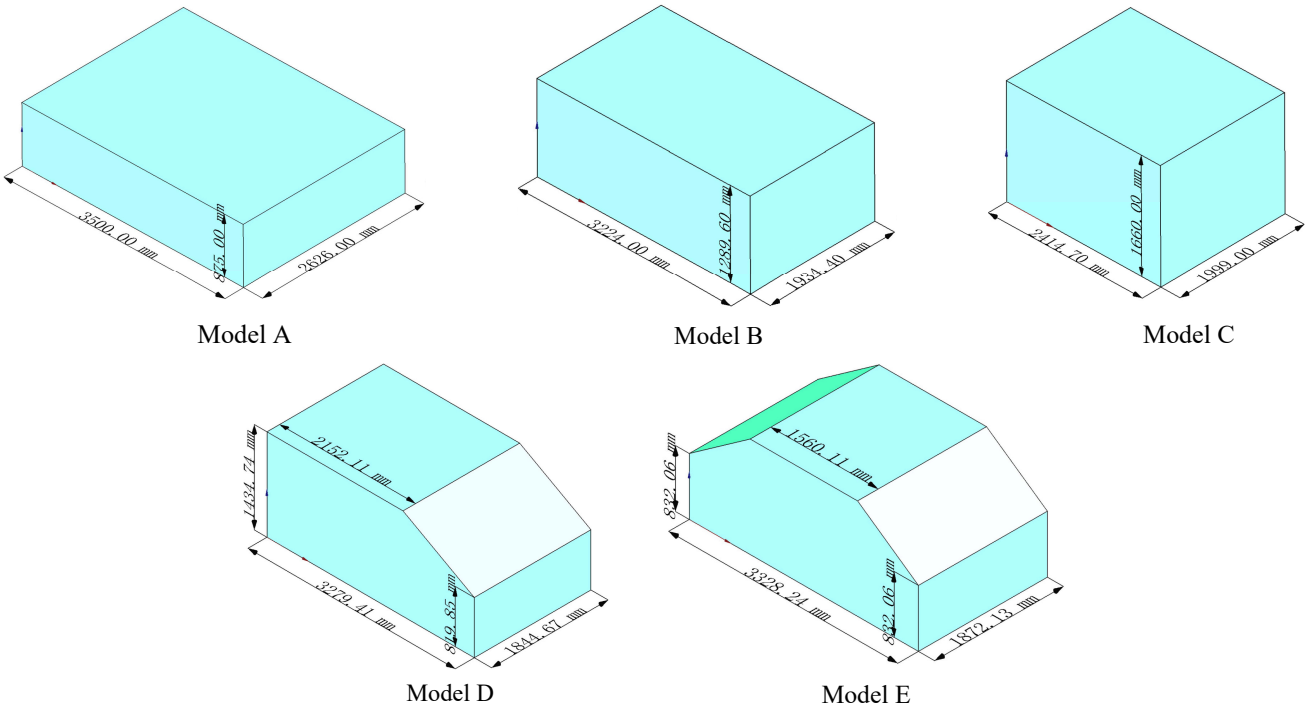


Fig. 1. Dimensions and geometries of five small-scale models. Rectangular models: model A (1.00 : 3.00 : 4.00), model B (1.00 : 1.50 : 2.50), model C (1.00 : 1.20 : 1.45). Non-rectangular models: model D (one oblique plane), model E (two oblique planes). Dimensions are in millimeters [mm].

All interior surfaces are modeled as isotropic and homogeneous boundaries, with a uniform absorption coefficient a ranging from 0.05 to 0.30. Boundary losses are implemented by locally reacting boundaries with Robin boundary conditions (Zheng *et al.*, 2021).

Acoustic eigenmodes are obtained by solving the linearized Helmholtz equation under rigid or damping boundary conditions via the FEM, the calculation process is detailed in the next section. The computational mesh element size is set to $h \leq \frac{\lambda_{\min}}{5}$, where $\lambda_{\min} = 0.0686\,\mathrm{m}$ corresponds to the wavelength at 1 kHz, ensuring a minimum of five elements per wavelength for simulation accuracy.

3. Theory and methods

Modal distribution and modal decay are fundamental properties of normal modes. This section first outlines the theoretical framework of modal theory and the analytical formulas for modal parameters in rectangular rooms, followed by the derivation of the SF. Based on this framework, the corresponding parameters for non-rectangular rooms are obtained through the numerical eigenfrequency analysis. Subsequently, the MDCF is introduced as an alternative estimation method for estimating the CF, and its differences from the SF formula are compared and discussed.

3.1. Analytical solution of modal parameters and derivation of the Schroeder frequency

For a rectangular room with rigid boundaries, the eigenfrequency formula is given by (Kuttruff, 2016):

$$f_n = \frac{c_0}{2}\sqrt{\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2}, \tag{1}$$

where c_0 is the speed of sound in air, L_x, L_y, L_z are the room dimensions, and n_x, n_y, n_z are non-negative integers representing the modal orders along each axis. Modes are classified according to their propagation direction: oblique modes (none of the n_x, n_y, n_z are zero, representing 3D volume propagation), tangential modes (one of the n_x, n_y, n_z is zero, indicating 2D surface propagation), and axial modes (two of the n_x, n_y, n_z are zero, corresponding to 1D axial propagation).

The modal number below an upper frequency f_u in rectangular rooms is given by (Maa, 1939):

$$N_f(f_u) = \frac{4\pi}{3}V\left(\frac{f_u}{c_0}\right)^3 + \frac{\pi}{4}S\left(\frac{f_u}{c_0}\right)^2 + \frac{L}{8}\frac{f_u}{c_0}, \tag{2}$$

where V, S, and L represent the room volume, total surface area, and sum of the room edge lengths, respectively. Since the first volume-governed term dominates at high frequencies, Eq. (2) is commonly simplified as (Weyl, 1911):

$$\widetilde{N}_f(f_u) = \frac{4\pi}{3}V\left(\frac{f_u}{c_0}\right)^3. \tag{3}$$

The eigenfrequency spacing, defined as the difference between consecutive eigenfrequencies, can also be expressed as the number of modes within per unit frequency range. Based on Eq. (3), the average eigenfrequency spacing is expressed as

$$\Delta\widetilde{f}_n = \left(\frac{\mathrm{d}\widetilde{N}_f(f_u)}{\mathrm{d}f_u}\right)^{-1} = \frac{c_0^3}{4\pi V f_u^2}. \tag{4}$$

Under non-rigid boundary conditions, modal energy decays over time. The decay rate is typically quantified by the decay factor δ, following the relation of $E(t) = E_0 e^{-2\delta t}$. The decay factor can be derived from the time required for the sound level decreases by 60 dB, known as T_{60}. Thus, the decay factor is given by (Kuttruff, 2016):

$$\widetilde{\delta} = \frac{3\ln(10)}{T_{60}}. \tag{5}$$

In the frequency region, the half-power bandwidth – defined as the frequency range where energy decays from peak to half-peak – is related to the decay factor by (Kuttruff, 2016):

$$B_{HP} = \frac{\delta}{\pi}. \tag{6}$$

By substituting the Eq. (5) into Eq. (6), the half-power bandwidth can be expressed as

$$\widetilde{B}_{HP} = \frac{3\ln(10)}{\pi \cdot T_{60}}. \tag{7}$$

The modal overlap degree is defined as the ratio of half-power bandwidth and eigenfrequency spacing, can be calculated as (Dance, Van Buuren, 2013):

$$M = \frac{B_{HP}}{\Delta f_n}. \tag{8}$$

According to Schroeder's study, when the modal overlap degree reaches or exceeds 3 – that is, when the eigenfrequencies spacing is less than one-third of the bandwidth – it represents the dense modal distribution (Schroeder, 1996). This criterion can be reformulated as the condition where the half-power bandwidth equals three times of eigenfrequency spacing, expressed as

$$B_{HP} = 3 \times \Delta f_n. \tag{9}$$

By substituting the expressions for eigenfrequency spacing from Eq. (4) and the half-power bandwidth from Eq. (7), the Schroeder frequency is expressed as

$$f_{\mathrm{SF}} = 2065.8 \sqrt{\frac{T_{60}}{V}} \approx 2000 \sqrt{\frac{T_{60}}{V}}. \tag{10}$$

Therefore, the SF provides a general formula, based on the room volume V and decay parameter T_{60}, to estimate the CF, which marks the transition from the discrete to dense modal distribution.

3.2. Numerical modal analysis and the definition of modal density-based crossover frequency

However, in a non-rectangular enclosure with complex boundary conditions, the analytical eigenfrequency formula – Eq. (1) – is no longer applicable. In practice, the FEM and other numerical simulation methods can provide a viable and efficient way to predict the eigenfrequency, as obtaining exact eigenvalues by measurements is often impractical. The FEM solves for the eigenfunctions by discretizing the continuous governing equations (Sakuma *et al.*, 2014). All models in this study are analyzed using the FEM to ensure comparability.

The governing equation (the Helmholtz equation) in the acoustic eigenmode analysis takes the form:

$$\nabla^2 p + k^2 p = 0, \tag{11}$$

where $k = \frac{\omega}{c_0}$ is the wavenumber. Through the Galerkin weighted residual method, the continuous Eq. (11) is discretized into a generalized matrix eigenproblem, with eigenfunctions Φ and corresponding eigenvalues k:

$$\left[\mathbf{K} - k^2 \mathbf{M}\right] \Phi = 0, \tag{12}$$

where $K_{ij} = \int_{\Omega} \nabla D_i \cdot \nabla D_j \, \mathrm{d}\Omega$ is the element of the stiffness matrix $\mathbf{K}$, and $M_{ij} = \int_{\Omega} D_i D_j \, \mathrm{d}\Omega$ is an element of the mass matrix $\mathbf{M}$; D_i and D_j represent shape functions for nodes i and j, and Ω is the computational volume domain.

Based on the numerical eigenfrequency solutions, the eigenfrequency spacing is calculated as the difference between consecutive eigenfrequencies. Compared to Eq. (4), this approach is fundamental and yields randomly fluctuating results rather than an averaged trend:

$$\Delta f_n = f_n - f_{n-1}. \tag{13}$$

Under damping boundary conditions with a specific acoustic impedance Z, the governing equation is modified as

$$\left(\mathbf{K} - k^2 \mathbf{M} + \mathbf{C}\right) \Phi = 0, \tag{14}$$

where $C_{ij} = \frac{i\omega\rho}{Z} \int_{S} D_i D_j \, \mathrm{d}S$ is an element of the damping matrix $\mathbf{C}$, ρ is the ambient air density, and S is the computational surface domain. Consequently, the n-th eigenfrequency is expressed as

$$f_{n,c} = f_n - i\frac{\delta_{n,c}}{2\pi}, \tag{15}$$

where the imaginary part represents the mode-specific decay factor $\delta_{n,c}$, which varies with the modal propagation direction and boundary conditions. Substituting $\delta_{n,c}$ into Eq. (6) yields the mode-specific half-power bandwidth:

$$B_{HP}\left(f_{n,c}\right) = \frac{\delta_{n,c}}{\pi}. \tag{16}$$

Applying the same criterion: 'three modes within a half-power bandwidth B_{HP},' the MDCF is defined from the numerical eigenfrequency solutions. As illustrated in Fig. 2, for each eigenfrequency $f_{n,c}^{(i)}$, the number of modes $N_f^{(i)}$ within its corresponding half-power bandwidth $B_{HP}(f_{n,c}^{(i)})$ is counted. The MDCF is identified as the lowest eigenfrequency for which the half-power bandwidth contains three or more modes, and all subsequent frequencies also satisfy this condition. The half-power bandwidth and its center frequency correspond to the imaginary and real parts of the complex eigenfrequency in Eq. (15).

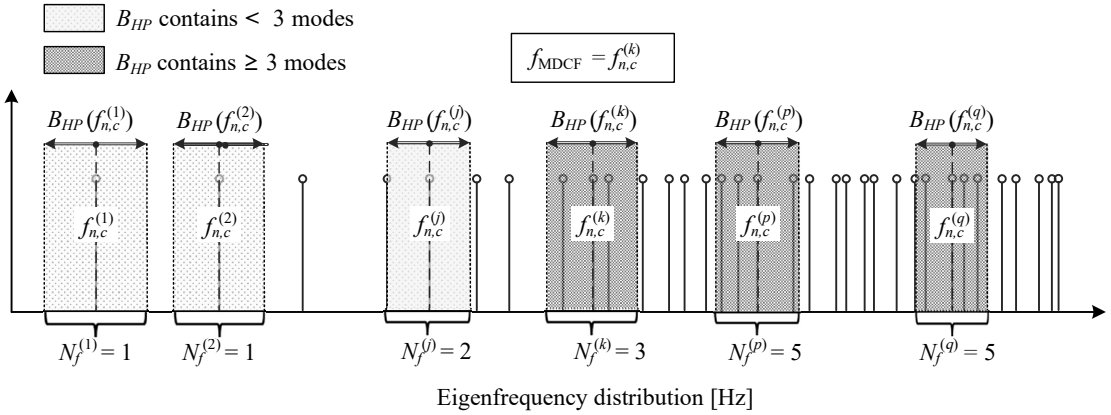


Fig. 2. Definition of modal density-based crossover frequency: the MDCF is the lowest eigenfrequency whose half-power bandwidth contains three or more modes, with all subsequent frequencies also meeting this criterion.

As shown in Fig. 2, light gray blocks represent bands that do not satisfy the dense modal criterion, such as $B_{HP}\left(f_{n,c}^{(1)}\right)$; while dark gray blocks satisfy this criterion. Since the $B_{HP}\left(f_{n,c}^{(k)}\right)$ centered at the eigenfrequency $f_{n,c}^{(k)}$ contains three modes of $N_f^{(k)}$, and all subsequent frequency bands also satisfy this condition, it follows that $f_{\mathrm{MDCF}} = f_{n,c}^{(k)}$.

3.3. Comparison of Schroeder frequency and modal density-based crossover frequency

Both of SF and MDCF employ the 'three mode within a B_{HP}' criterion, but they differ in the employed parameters and derivation process. Table 1 summarizes the modal parameters used in the derivation of SF and MDCF.

Table 1. Modal parameters employed in derivation process of SF and MDCF.

Parameter	Schroeder frequency	Modal density-based crossover frequency
Eigenfrequency spacing	$\Delta\widetilde{f}_n = \frac{c_0^3}{4\pi V f_u^2}$, simplified formulas (rectangular shape)	$\Delta f_n = f_n - f_{n-1}$, numerical results (any geometrical shapes)
Half-power bandwidth	$\widetilde{B}_{HP} = \frac{3\ln(10)}{\pi \cdot T_{60}}$, with global decay T_{60}	$B_{HP}\left(f_{n,c}\right) = \frac{\delta_{n,c}}{\pi}$, mode-dependent

4. Results

The modal parameters obtained from numerical simulations and analytical formulas are compared in the following section.

4.1. Eigenmode distributions under rigid walls

Figure 3 presents the comparison of eigenfrequency spacing under the rigid boundary condition. The numerical results Δf_n from Eq. (13) in 8 m^3 and 80 m^3 rooms are shown as green- and blue-marked lines, respectively; while analytical results $\Delta \widetilde{f}_n$ from Eq. (4) are presented by black dashed and dotted lines. The gray shaded areas indicate the discrepancies between the two methods.

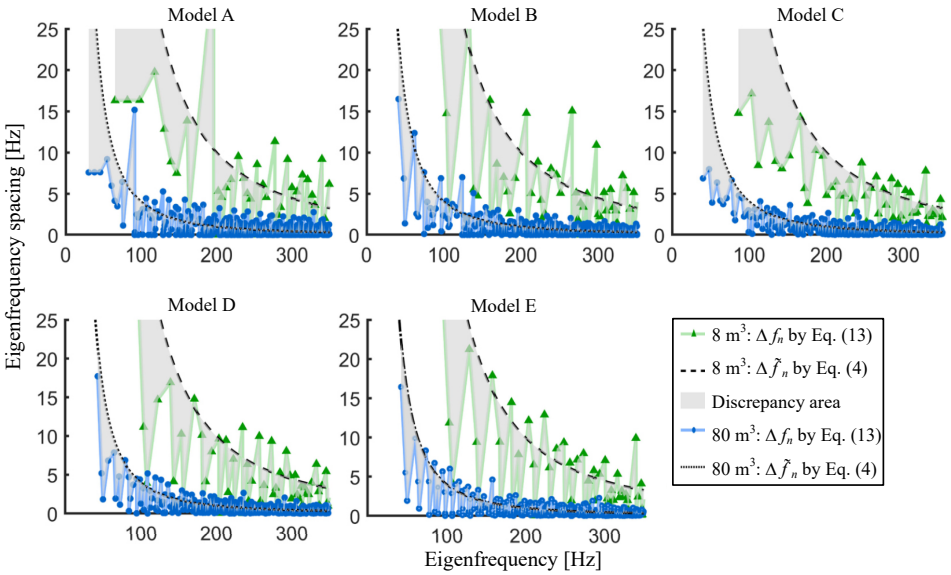


Fig. 3. Eigenfrequency spacing from numerical results Δf_n from Eq. (13) and analytical results $\Delta \widetilde{f}_n$ from Eq. (4) in 8 m^3 rooms (green-marked lines, black-dashed lines) and 80 m^3 rooms (blue-marked lines, black-dotted lines). Gray regions denote discrepancies between them.

Overall, discrepancies are observed across all models, and are more pronounced in small-scale rooms (8 m^3). In the low-frequency range, the eigenfrequency spacings obtained from the analytical results are higher than that from the numerical results, and gradually consistent. For example, below 200 Hz, the average difference between the two methods range from 10 Hz to 20 Hz in 8 m^3 rooms; and from 3 Hz to 10 Hz in 80 m^3 rooms.

Regarding the shape-dependent discrepancies in 8 m^3 rooms, model A exhibits the largest deviation, with differences up to 20 Hz. Models B and E also show substantial discrepancies, with maximum deviations of 10 Hz to 15 Hz. Models C and D display smaller differences, with deviations around 10 Hz. These results indicate that the analytical method becomes less accurate in small-scale enclosures, ultimately affecting the analysis of modal overlap and crossover frequency.

4.2. Half-power bandwidth under damping boundaries

Figure 4 compares the half-power bandwidths under a uniform absorption boundary condition (α = 0.20). The numerical bandwidths $B_{HP}(f_{n,c})$ from Eq. (16) for 8 m^3 and 80 m^3 rooms are presented by green- and blue-marked lines, respectively; and analytical results $\widetilde{B}_{HP}$ from Eq. (7) are shown as black dashed and dotted lines. The gray areas quantify the deviations between the two methods.

Across all models, the bandwidths obtained numerically are consistently higher than those from the analytical formula, and this discrepancy is more pronounced in 8 m^3 rooms. For instance, in 8 m^3 rooms, deviations exceed 10 Hz at certain eigenfrequencies; whereas the discrepancy is narrower in 80 m^3 rooms. Moreover, the numerical results exhibit frequency-dependent variations: at low frequencies, they show a scattered distribution due to varying modal damping; whereas at high frequencies, they stabilize at a higher value owing to the dominance of oblique modes. In contrast, the analytical results remain constant across the frequency range.

As for the shape-dependent discrepancies in 8 m^3 rooms, model C has the largest deviation (nearly 15 Hz), followed by models B and E (up to 10 Hz), and model D (slightly less than 10 Hz). Model A exhibits the smallest

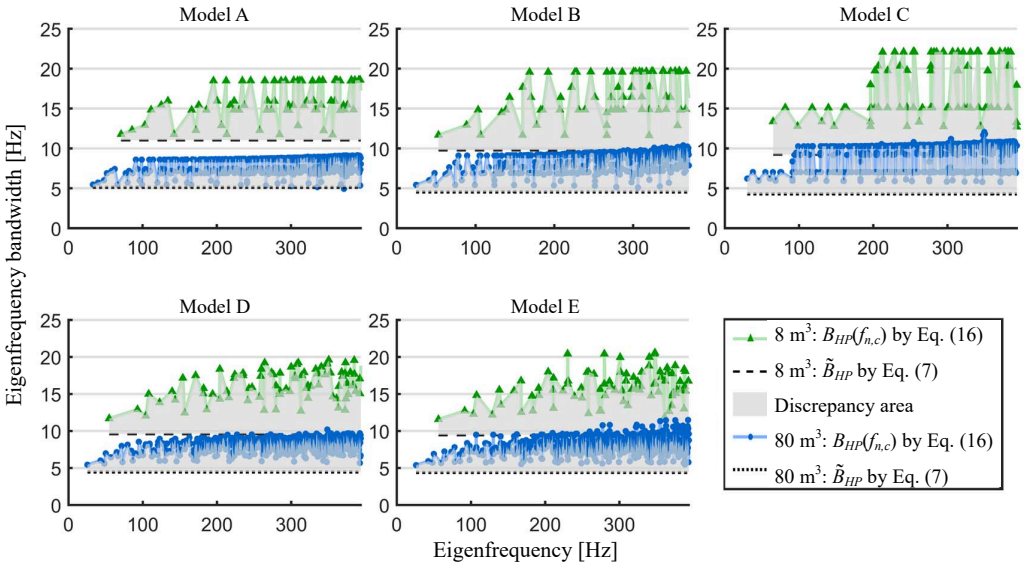


Fig. 4. Half-power bandwidth of numerical results $B_{HP}(f_{n,c})$ from Eq. (16) and analytical results $\widetilde{B}_{HP}$ from Eq. (7) in 8 m^3 rooms (green-marked lines, black-dashed lines), and 80 m^3 rooms (blue-marked lines, black-dotted lines). The boundary absorption coefficient is 0.20, and T_{60} in analytical $\widetilde{B}_{HP}$ is obtained by Eyring T_{60} formula. Gray regions denote discrepancies between them.

discrepancy. These results demonstrate that the analytical formula underestimates the modal damping, and that the discrepancies between numerical and analytical results among different model shapes are more pronounced in small-scale enclosures.

4.3. Modal overlap degree

The ratio of the half-power bandwidth to eigenfrequency spacing is the modal overlap degree, which determines the final crossover frequency. Figure 5 compares the modal overlap degree under the boundary condition of a uniform absorption $\alpha = 0.20$. Green and blue lines represent $B_{HP}(f_{n,c})/\Delta f_n$ obtained by numerical results for 8 m^3 and 80 m^3 rooms, respectively; while black-dashed and dotted lines show $\widetilde{B}_{HP}/\Delta\widetilde{f}_n$ obtained by the analytical formula. Gray areas indicate where a modal overlap degree reach or exceed three. The vertical solid and dashed lines indicate the MDCF and SF for 8 m^3 rooms (green) and 80 m^3 rooms (blue). Above the MDCF and SF, the modal overlap degree exceeds three.

This analysis is critical for determining the MDCF and SF. Overall, the modal overlap degree from numerical results is consistently higher than that from analytical results under the current boundary condition; correspondingly, MDCF is generally lower than the SF in all models except model C, where a sudden drop in the modal bandwidth containing fewer than three modes results in a higher MDCF.

Notably, the difference between MDCF and SF is more pronounced in 8 m^3 rooms. For example, model D exhibits the largest gap, around 140 Hz (MDCF ≈ 220 Hz and SF ≈ 360 Hz), followed by model E at approximate 100 Hz, model B is around 60 Hz, and model A is around 40 Hz. In contrast, 80 m^3 rooms show narrower gaps between MDCF and SF, around 15 Hz to 40 Hz. Based on this methodology, results for other boundary conditions could be derived similarly and are presented in the next section.

4.4. Modal density-based crossover frequency

Based on the foregoing analysis, the MDCF can be determined. Figure 6 compares MDCF with SF in 8 m^3 and 80 m^3 rooms for the boundary absorption coefficients in the range of [0.05, 0.30]. The MDCF is represented by marked lines and SF is represented by dashed lines.

The results show that the MDCF is generally lower than the SF in most cases. Notably, larger discrepancies are observed in 8 m^3 rooms than 80 m^3 rooms. For instance, at $\alpha = 0.20$, the differences of MDCF and SF exhibit

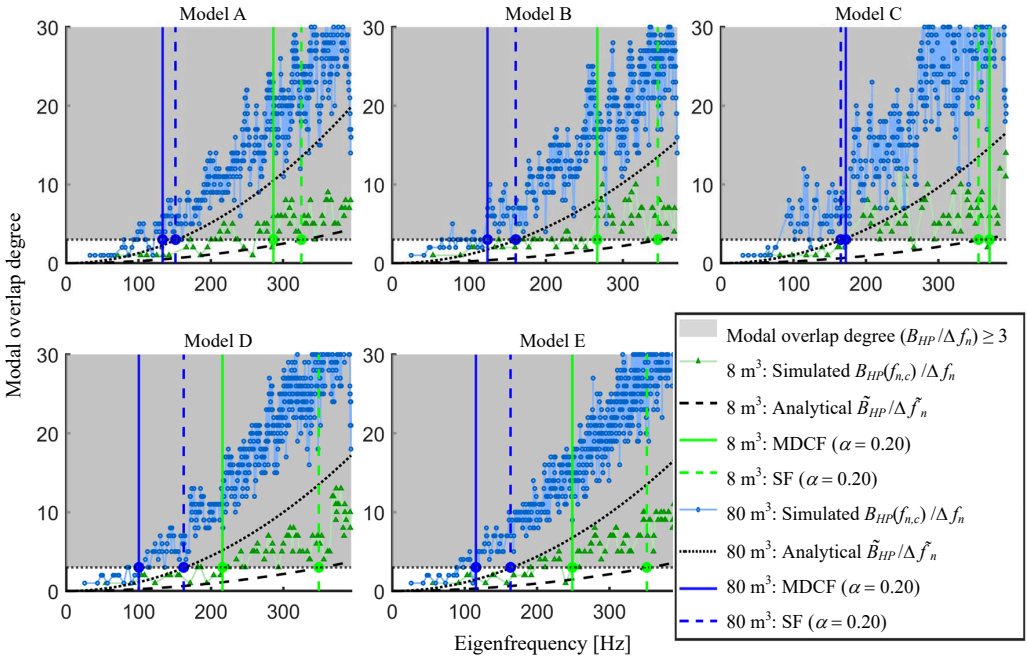


Fig. 5. Modal overlap degree of numerical modal parameters $(B_{HP}(f_{n,c})/\Delta f_n)$ and analytical modal parameters $(\widetilde{B}_{HP}/\Delta\widetilde{f}_n)$ in 8 m^3 rooms (green-marked lines, black-dashed lines) and 80 m^3 rooms (blue-marked lines, black-dotted lines). The boundary absorption coefficient is 0.20. The vertical lines represent the MDCF (solid lines) and SF (dashed lines) in 8 m^3 rooms (green) and 80 m^3 rooms (blue), indicating the frequency beyond which the bandwidth contains more than three modes.

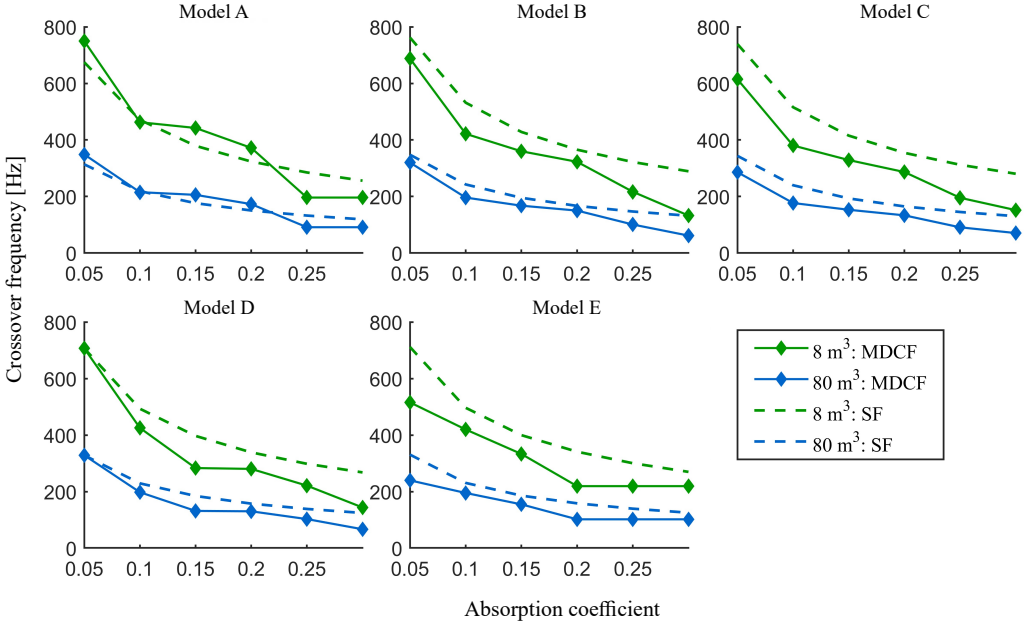


Fig. 6. MDCF and SF in 8 m^3 (green-solid and dashed lines) and 80 m^3 (blue-solid and dashed lines) rooms, when boundary absorption coefficient is from 0.05 to 0.30.

an average of 55 Hz and the peak of 145 Hz in 8 m^3 rooms, whereas in 80 m^3 rooms, the average gap is 25 Hz, and the peak gap is 50 Hz. As analyzed above, discrepancies in the eigenfrequency spacing and half-power bandwidth between numerical and analytical results lead to differences between the MDCF and the SF. These include: the MDCF accounts for actual modal distribution, considering all simulated eigenmodes, and the SF relies on volume-related eigenmodes that overestimate eigenfrequency spacing, particularly in small-scale rooms; and the numerical B_{HP} is consistently wider than analytical predictions, with this discrepancy amplified in smaller enclosures. These

findings explain the mechanisms underlying the discrepancies between the SF and the MDCF when estimating the CF, and quantitatively specify these differences.

5. Conclusion

This study investigates modal characteristics of cabin-sized enclosures by analyzing two key parameters that determine the CF – modal distribution and modal decay – using both numerical simulations and analytical formulations. The MDCF, derived from numerical simulations, is proposed and compared with the SF. Based on ten models with two volumes and five different geometrical shapes, the results show that the MDCF varies with room shapes at a constant volume. Furthermore, the MDCF is lower than SF approximately 70 Hz to 150 Hz in cabin-sized rooms; while the differences decrease from 20 Hz to 50 Hz in larger rooms. Consequently, MDCF provides a more reliable CF estimation for irregularly shaped rooms, and highlights the necessity of incorporating accurate modal parameters in the modal analysis of cabin-sized enclosures. Additionally, the difference in modal decay rates across modal types is analyzed, providing theoretical insights relevant to low-frequency equalization. Future work should include experimental validation to deepen the understanding of modal properties in cabin-sized enclosures.

Fundings

This work was supported by the National Natural Science Foundation of China (grant no. 12474465) and the Natural Science Foundation of Guangdong Province (grant no. 2024A1515011446).

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.

Authors' contributions

Ziyu Wang conceptualized the study, performed the data analysis and interpretation, wrote the original draft. Liangfen Du contributed to data interpretation and review the original draft. Guangzheng Yu contributed to conceptualize the study and review the original draft. All authors reviewed and approved the final manuscript.

Data availability

The data that support the findings of this study are available within the article.

References

1. Aretz M., Vorländer M. (2014a), Combined wave and ray based room acoustic simulations of audio systems in car passenger compartments, Part I: Boundary and source data, *Applied Acoustics*, **76**: 82–99, https://doi.org/10.1016/j.apacoust.2013.07.021.
2. Aretz M., Vorländer M. (2014b), Combined wave and ray based room acoustic simulations of audio systems in car passenger compartments, Part II: Comparison of simulations and measurements, *Applied Acoustics*, **76**: 52–65, https://doi.org/10.1016/j.apacoust.2013.07.020.
3. Bastine A., Abhayapala T.D., Zhang J.A. (2021), Power response and modal decay estimation of room reflections from spherical microphone array measurements using Eigenbeam spatial correlation model, *Applied Sciences*, **11**(16): 7688, https://doi.org/10.3390/app11167688.
4. Bolt R.H. (1946), Note on normal frequency statistics for rectangular rooms, *The Journal of the Acoustical Society of America*, **18**(1): 130–133, https://doi.org/10.1121/1.1916349.

5. Brinkmann F., Aspöck L., Ackermann D., Lepa S., Vorländer M., Weinzierl S. (2019), A round robin on room acoustical simulation and auralization, *The Journal of the Acoustical Society of America*, **145**(4): 2746–2760, https://doi.org/10.1121/1.5096178.
6. Dance S.M., Van Buuren G. (2013), Effects of damping on the low-frequency acoustics of listening rooms based on an analytical model, *Journal of Sound and Vibration*, **332**(25): 6891–6904, https://doi.org/10.1016/j.jsv.2013.07.011.
7. Ferreira T.S., Magalhães P.A., Moura F.L., Ferreira T.S. (2016), The effect of the cavity damping on vehicular evaluation using the finite element method, *Archives of Acoustics*, **41**(1): 87–97, https://doi.org/10.1515/aoa-2016-0009.
8. Granier E., Kleiner M., Dalenbäck B.-I., Svensson P. (1996), Experimental auralization of car audio installations, *Journal of the Audio Engineering Society*, **44**(10): 835–849.
9. Gunawan H., Aditanoyo T. (2018), Study of the effect of splaying wall to modify acoustic modes distribution in small room, *Journal of Physics: Conference Series*, **1075**(1): 012049, https://doi.org/10.1088/1742-6596/1075/1/012049.
10. Kleiner M., Tichy J. (2014), *Acoustics of Small Rooms*, CRC Press, https://doi.org/10.1201/b16866.
11. Kuttruff H. (2016), *Room Acoustics*, 6th ed., CRC Press, https://doi.org/10.1201/9781315372150.
12. Maa D. (1939), Distribution of eigentones in a rectangular chamber at low frequency range, *The Journal of the Acoustical Society of America*, **10**(3): 235–238, https://doi.org/10.1121/1.1915981.
13. Meissner M. (2017), Acoustics of small rectangular rooms: Analytical and numerical determination of reverberation parameters, *Applied Acoustics*, **120**: 111–119, https://doi.org/10.1016/j.apacoust.2017.01.020.
14. Meissner M. (2021), Application of modal expansion method for sound prediction in enclosed spaces subjected to boundary excitation, *Journal of Sound and Vibration*, **500**: 116041, https://doi.org/10.1016/j.jsv.2021.116041.
15. Morse P.M., Uno Ingard U. (1968), *Theoretical Acoustics*, Princeton University Press, Princeton, New Jersey.
16. Nélisse H., Nicolas J. (1997), Characterization of a diffuse field in a reverberant room, *The Journal of the Acoustical Society of America*, **101**(6): 3517–3524, https://doi.org/10.1121/1.418313.
17. Pinardi D., Riabova K., Binelli M., Farina A., Park J.-S. (2021), Geometrical acoustics simulations for ambisonics auralization of a car sound system at high frequency, [in:] *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, https://doi.org/10.1109/I3DA48870.2021.9610977.
18. Rindel J.H. (2015), Modal energy analysis of nearly rectangular rooms at low frequencies, *Acta Acustica United With Acustica*, **101**(6): 1211–1221, https://doi.org/10.3813/aaa.918914.
19. Rindel J.H. (2021), Preferred dimension ratios of small rectangular rooms, *JASA Express Letters*, **1**(2): 021601, https://doi.org/10.1121/10.0003450.
20. Rougier C. (2018), *Influence of crossover frequency on a hybrid acoustic model for room impulse response synthesis*, M.Sc. Thesis, https://api.semanticscholar.org/CorpusID:115525011.
21. Sakuma T., Sakamoto S., Otsuru T. (2014), *Computational Simulation in Architectural and Environmental Acoustics: Methods and Applications of Wave-Based Computation*, Springer, Tokyo.
22. Savioja L., Svensson U.P. (2015), Overview of geometrical room acoustic modeling techniques, *The Journal of the Acoustical Society of America*, **138**(2): 708–730, https://doi.org/10.1121/1.4926438.
23. Savioja L., Xiang N. (2019), Introduction to the special issue on room acoustic modeling and auralization, *The Journal of the Acoustical Society of America*, **145**(4): 2597–2600, https://doi.org/10.1121/1.5099017.
24. Schroeder M.R. (1987), Statistical parameters of the frequency response curves of large rooms, *Journal of the Audio Engineering Society*, **35**(5): 299–306.
25. Schroeder M.R. (1996), The "Schroeder frequency" revisited, *The Journal of the Acoustical Society of America*, **99**(5): 3240–3241, https://doi.org/10.1121/1.414868.
26. Schroeder M.R., Kuttruff K.H. (1962), On frequency response curves in rooms. Comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima, *The Journal of the Acoustical Society of America*, **34**(1): 76–80, https://doi.org/10.1121/1.1909022.
27. Siltanen S., Lokki T., Savioja L. (2010), Rays or waves? Understanding the strengths and weaknesses of computational room acoustics modeling techniques, [in:] *Proceedings of the International Symposium on Room Acoustics, ISRA 2010*.

28. Southern A., Siltanen S., Murphy D.T., Savioja L. (2013), Room impulse response synthesis and validation using a hybrid acoustic model, *IEEE Transactions on Audio, Speech, and Language Processing*, **21**(9): 1940–1952, https://doi.org/10.1109/TASL.2013.2263139.

29. Southern A., Siltanen S., Savioja L. (2011), Spatial room impulse responses with a hybrid modelling method, [in:] *Audio Engineering Society 130th Convention*.

30. Summers J.E., Takahashi K., Shimizu Y., Yamakawa T. (2004), Assessing the accuracy of auralizations computed using a hybrid geometrical-acoustics and wave-acoustics method, *The Journal of the Acoustical Society of America*, **115**(5_Supplement): 2514–2515, https://doi.org/10.1121/1.4809339.

31. Weyl H. (1911), On the asymptotic distribution of eigenvalues [in German], *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, **1911**: 110–117, http://eudml.org/doc/58792.

32. Zheng C., Zhao W., Gao H., Du L., Zhang Y., Bi C. (2021), Sensitivity analysis of acoustic eigenfrequencies by using a boundary element method, *The Journal of the Acoustical Society of America*, **149**(3): 2027–2039, https://doi.org/10.1121/10.0003622.

ARCHIVES OF ACOUSTICS Vol. 51, No. 1, pp. 37–55 (2026), https://doi.org/10.24423/archacoust.2026.4229

Research Paper

TimeGAN and Coordinated Attention Prototype Network Based Prediction Model for Infrasound Signal

Quanbo LU[(1)]*, Xiaojuan HUANG[(2)], Rao LI[(1)], Mei LI[(3)], Dong ZHU[(4)]

[(1)] *College of Communication Engineering, Chongqing Polytechnic University of Electronic Technology*
Chongqing, China

[(2)] *School of Mechanical Engineering, Chongqing Three Gorges University*
Chongqing, China

[(3)] *School of Information Engineering, China University of Geosciences*
Beijing, China

[(4)] *Sevnce Robotics Co, Ltd*
Chongqing, China

*Corresponding Author: luquanbo@cqcet.edu.cn

Received April 8, 2025; revised November 2, 2025; accepted November 13, 2025; available online January 15, 2026; version of record February 6, 2026; published issue March 27, 2026.

Due to the complexity of the infrasound environment and the high costs associated with data collection, frequent acquisition of infrasound data is often impractical, resulting in a limited amount of labeled data. To address the challenge of low classification prediction accuracy caused by data scarcity, this paper proposes an infrasound prediction model based on a time-series generative adversarial network (TimeGAN) and coordinated attention prototype network (CAPN) (TimeGAN-CAPN). The model begins by introducing TimeGAN, where the generative network is trained using a combination of unsupervised and supervised learning. This approach enables the network to operate within the latent space of temporal features and generate time-series data that closely aligns with the distribution of the original data. These generated samples are then combined with the original data to form an augmented dataset. Subsequently, the augmented data is input into the CAPN, which enhances the sample size per class, allowing for more precise class prototypes and improving the prediction accuracy of the model. Furthermore, the quality and diversity of the data generated by TimeGAN are quantitatively and qualitatively assessed using maximum mean discrepancy (MMD) and t-distributed stochastic neighbor embedding (t-SNE), facilitating a comparison and verification of the generated data's performance. Experimental results show that TimeGAN-CAPN significantly outperforms the CAPN model in classification tasks with limited infrasound data, achieving an increase in accuracy of 7.15 %. This demonstrates that the proposed method is highly effective for predicting infrasound-related disasters, particularly in scenarios with small sample sizes.

Keywords: infrasound signal, time-series generative adversarial network, coordinated attention prototype network, maximum mean discrepancy.

1. Introduction

Infrasound (≤20 Hz) refers to sound waves with frequencies below the human hearing range, and is characterized by long propagation distances and strong penetration ability (SOVILLA *et al.*, 2025; LU, LI, 2023). Many extreme events, such as earthquakes, tsunamis, and explosions, generate infrasound waves. Globally, infrasound monitoring has been widely applied in the prediction and prevention of natural disasters. Infrasound event detec-

tion is the foundation of infrasound monitoring, with its main goal being to extract infrasound events from a large amount of background noise and determine the event's scope (Dong *et al.*, 2024). Event detection is significant for subsequent research such as event classification and localization. Therefore, improving the effectiveness of event detection has become a key issue in the field of infrasound research.

The importance of infrasound event detection algorithms in infrasound monitoring has led to rapid advancements in their technological research. Many scholars have conducted studies on infrasound event detection methods, and new methods continue to be introduced. Baeza Moyano, González Lezcano *et al.* (2022) explored the potential health impacts of infrasound and advocates for improvements in housing conditions to mitigate these effects. Watson *et al.* (2022) reviewed the advancements in volcano infrasound research and outlines future directions for further investigation and application in volcanic monitoring. Friedrich *et al.* (2023) examined how infrasound affects the perception of low-frequency sounds and its potential influence on human perception and response. Hupe *et al.* (2022) discussed the use of infrasound data products from the International Monitoring System for atmospheric studies and various civilian applications. Macpherson *et al.* (2023) explored the use of local infrasound to estimate seismic velocity and earthquake magnitudes, offering a new approach for seismic monitoring. Listowski *et al.* (2022) investigated the use of infrasound for remotely monitoring Mediterranean hurricanes, highlighting its potential for early detection and tracking. Zajamsek *et al.* (2023) explored how infrasound influences the detectability of amplitude-modulated tonal noise, focusing on its impact on human perception. Wilson *et al.* (2023) presented findings from a long-term microphone array deployment in Oklahoma, analyzing infrasound and low-audible acoustic detections for various environmental and geophysical applications. Yang *et al.* (2025) examined the correlation between gas desorption processes and infrasound signals, investigating the underlying mechanisms that link the two phenomena. However, the above methods do not consider the prediction of infrasound signals in small sample scenarios.

To address the challenge of low classification prediction accuracy caused by the scarcity of labeled infrasound signal samples, this paper proposes an infrasound prediction model based on coordinated attention prototype network (TimeGAN-CAPN). The model first expands the temporal infrasound data using time-series generative adversarial network (TimeGAN), then combines the generated data with the original dataset to train the prediction model, thereby enhancing its performance. Subsequently, drawing on the principles of metric learning, a coordinated attention mechanism is integrated into the traditional prototype network to extract more discriminative feature information, facilitating the accurate construction of metric prototypes for various types of infrasound. Inspired by the biological binocular system, a deep mutual learning framework is introduced to integrate convolutional neural networks with coordinated attention prototype network (CAPN), further improving the model's prediction accuracy. Experimental results demonstrate that the proposed method outperforms other approaches in classification performance, significantly enhancing disaster early warning rates and advancing the practical application of infrasound detection algorithms.

The structure of this paper is as follows: Sec. 2 provides a brief overview of the basic theories behind TimeGAN, CAPN, and TimeGAN-CAPN, which are used in this study; Sec. 3 presents a performance comparison of different methods through experiments; finally, conclusions are drawn in Sec. 4.

2. Methods

2.1. TimeGAN

Yoon *et al.* (2019) proposed the TimeGAN by combining the flexibility of unsupervised learning with the strong control of the training process in supervised learning. Its training process is essentially a process of solving the min-max problem of a binary function. The model consists of two networks: the reconstruction network and the embedding network; and two generative models: the discriminator and the generator. It uses three different loss functions: the generation loss function, the supervised loss function, and the unsupervised loss function to train the network.

The Time-GAN model uses gradient descent for parameter optimization, with the generator typically taking random noise and vectors as input. The loss function is expressed as follows (Sharma *et al.*, 2024):

$$L_G(Z) = E_{Z\sim P_Z(Z)}\left[\log\left(1 - D(G(Z))\right)\right], \tag{1}$$

where $L_G(\cdot)$ is the generator's loss function, $E(\cdot)$ is the embedding network's expected loss, $G(\cdot)$ is the generator function, $D(\cdot)$ is the discriminator function, $P_Z(\cdot)$ is the noise data distribution, Z is the random variable for noise input.

The input variables for the discriminator are synthetic data and real data to be distinguished, and the loss function is expressed as (Vuletić *et al.*, 2024):

$$L_D(x) = E_{X\sim P_i(x)}\left[\log D(x)\right] + E_{X\sim P_G(x)}\left[\log\left(1 - D(x)\right)\right], \tag{2}$$

where $L_D(\cdot)$ is the real data variable, i is the fake data variable, $P_i(\cdot)$ is the real data distribution, x is the input random variable.

2.2. CAPN

The CAPN model is shown in Fig. 1. It consists of two parallel views: in the global view, a convolutional neural network (CNN) is used to capture inter-class relationships, while in the local view, a prototype network with a coordinated attention mechanism focuses more on matching details (Jiang *et al.*, 2025). The two views are then aggregated through a deep mutual learning framework, implicitly exploring useful knowledge from each other. The training process aims to find the best hyperparameter settings and leverage prior knowledge to better train specific test tasks. Finally, during the testing process, the collaborative features from both views are combined to perform classification tasks, thereby improving the accuracy of few-shot classification prediction. The model is mainly divided into three parts: the global view, the local view, and cross-view mutual learning.

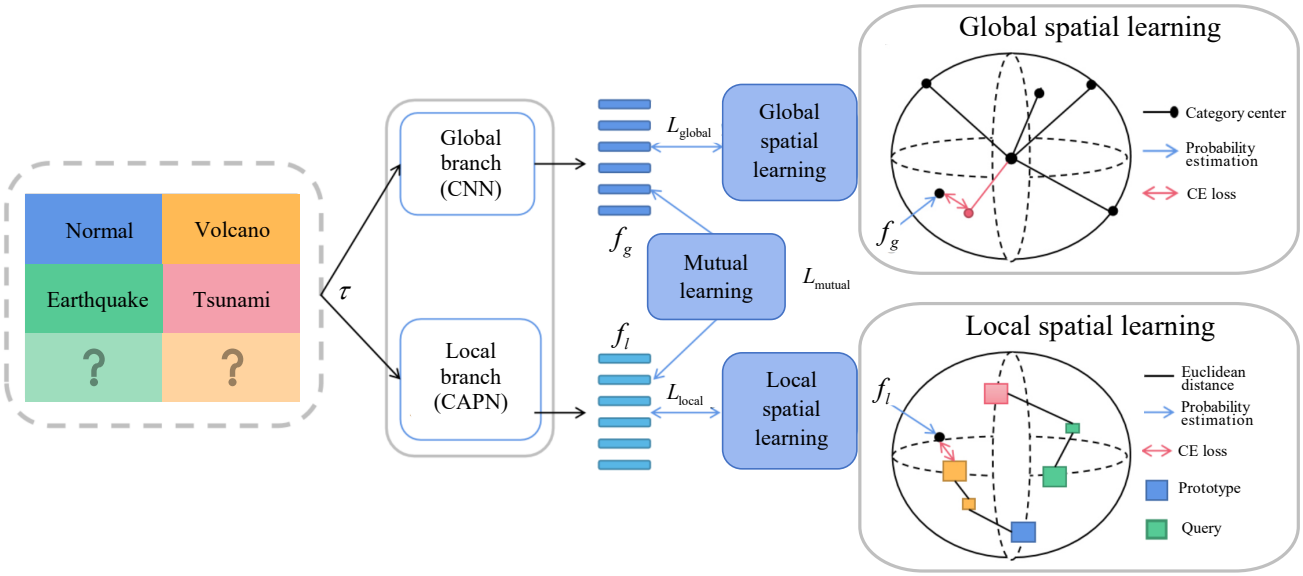


Fig. 1. Structure of CAPN.

2.2.1. Global view

In the global view, a 1D convolutional network is used for training (Zhang *et al.*, 2024). Specifically, for a given task Γ, a global learner $A_{\varnothing}^{G}$ is trained to map each data sample x_i in the Γ set to a high-dimensional space. The probability distribution of x_i is expressed as follows:

$$P\left(y_i = y \,|\, x_i\right) = \sigma\left(A_{\varnothing}^{G}(x_i)\right), \tag{3}$$

where σ is the softmax activation function. It serves to combine the extracted features for nonlinear activation, outputting the probability distribution of each class, which is then used for classification.

The loss function is calculated using cross-entropy, i.e., the negative logarithm of the probability $P(y_i = y \,|\, x_i)$ (Tang *et al.*, 2023). Therefore, the loss function for the global view is as follows:

$$L_{\text{global}} = E_{(x_i,y_i)\epsilon T} \cdot \sum_{i=1}^{N} y_i \log P(y_i = y \,|\, x_i). \tag{4}$$

2.2.2. Local view

In the local view, a prototype network is used to match each query sample with the class prototype from the support set in the embedding space. Therefore, in the local view, a prototype network with a coordinated attention mechanism is applied.

This model consists of three parts: feature embedding, prototype generation, and feature distance-based classification. The structure is shown in Fig. 2. The first step is to use a feature embedding module with an attention mechanism for feature embedding. Samples of the support set and query set are passed through the convolutional layers. By adding the coordinated attention mechanism, both spatial and channel information are extracted, and by embedding position information in the channel attention, accurate position details and long-range dependencies are captured. This allows the feature embedding to focus more on useful local feature information, enhancing the feature representation ability of the feature embedding network. The second step is to compute the class prototype features by averaging the feature maps of samples from the same class. The mean feature serves as the class prototype feature. The third step is to measure the distance between the category prototype features learned by the feature embedding network and the query sample features using a selected distance metric, such as Euclidean distance. According to the principle that similar samples are close and dissimilar samples are far apart, the closest prototype to the query set output is selected as the predicted result, and the network is trained until it meets the required model and label prototypes. Classification prediction is then performed using the saved optimal coordinated attention prototype network.

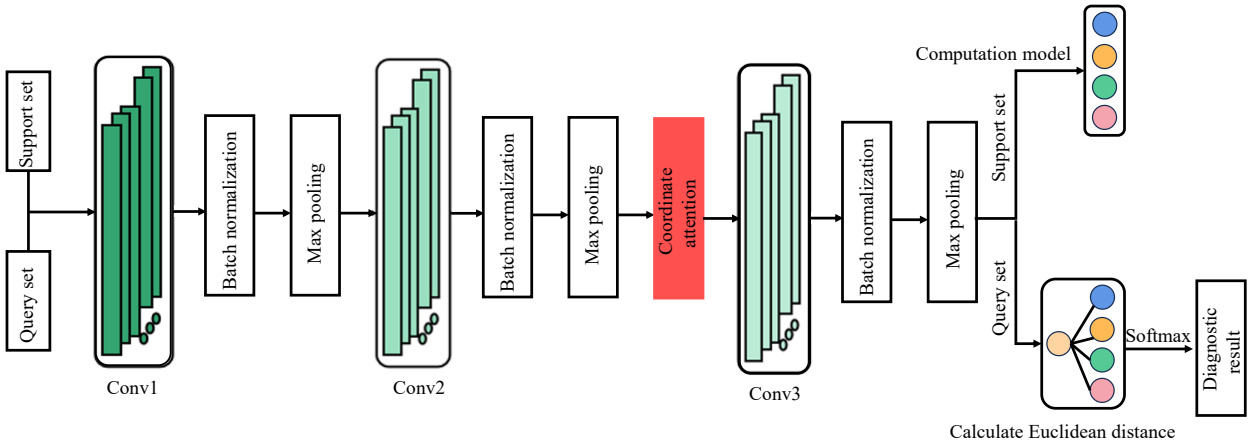


Fig. 2. Structure of CAPN for the classification prediction.

2.2.3. Cross-view mutual learning

In addition to learning within each individual view, the global and local views also mutually promote each other through cross-view interaction in the deep mutual learning network. Specifically, for each view, in addition to completing its own training task, the view also minimizes the imitation loss from the other view. This imitation loss uses the Kullback–Leibler divergence to quantify the match between the prediction probabilities of the two networks, which aids in implicit knowledge transfer (Ji *et al.*, 2020). The mutual loss is shown in Eq. (5), which includes two sub-items:

$$L_{\text{mutual}} = D_{KL}(F_1 \| F_g) + D_{KL}(F_g \| F_1), \tag{5}$$

$$D_{KL}(F_1 \| F_g) = F_1 \log \frac{F_1}{F_g}, \tag{6}$$

$$D_{KL}(F_g \| F_1) = F_g \log \frac{F_g}{F_1}, \tag{7}$$

where $F(\,)$ represents the feature distribution computed by $\sigma(A_{\varnothing}(x))$, and the interaction problem is considered from the perspective of feature distribution consistency. Learning in Euclidean space focuses on relative relationships rather than hard constraints such as mean square error. This is because overly strong supervision signals are not conducive to preserving the specificity of both views (Ruddick *et al.*, 2024).

Thus, the final loss function of the model is given by

$$L_{\text{total}} = \alpha L_{\text{global}} + \beta \widehat{L}_{11} + \gamma L_{\text{mutual}}, \tag{8}$$

where α, β, and γ are the weighting factors. The optimal loss weights α, β, and γ are determined through a systematic hyperparameter tuning process using grid search combined with cross-validation. This paper explores various combinations of α, β, and γ within a pre-defined range, informed by prior work on similar models and the characteristics of infrasound data. The model's performance is evaluated based on classification accuracy and loss using a validation set, with the aim of balancing the three loss terms – reconstruction loss, supervised loss, and unsupervised loss – while avoiding overfitting. After multiple iterations, the values that resulted in the highest overall performance are selected, ensuring the model effectively captured both temporal and discriminative features of the infrasound signals.

2.3. Proposed approach

The TimeGAN-CAPN infrasound prediction model proposed in this paper is illustrated in Fig. 3. It consists of three main components: data preprocessing, data generation, and infrasound prediction. In the data preprocessing

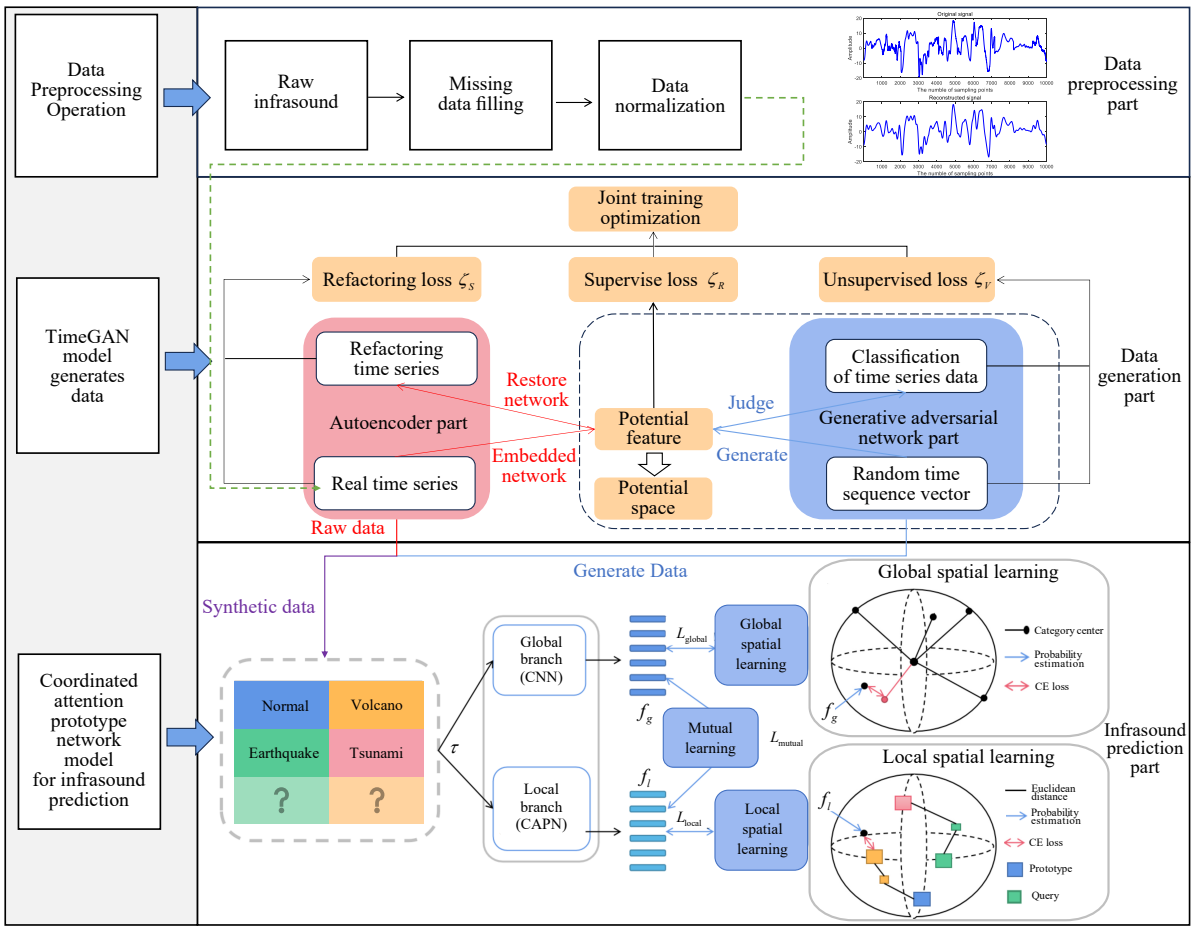


Fig. 3. Framework of the proposed approach.

phase, missing values in the sensor-collected data are imputed using the nearest neighbor interpolation method. The data is then normalized using min-max normalization, ensuring consistent dimensions and complete features, which enhances its usability. In the data generation phase, a TimeGAN model is constructed, and the collected data samples are fed into the generative model. Through an adversarial process between the generator and discriminator in the latent space, the loss function is computed to update the model parameters, ultimately generating high-quality infrasound data samples. These generated samples are then combined with the original data to form an augmented dataset. In the infrasound prediction phase, the synthesized dataset is split into a training set and a test set. The training set is used to train the CAPN-based infrasound prediction model. The final model is then applied to infrasound prediction tasks for early disaster detection.

2.3.1. Data preprocessing

The collected infrasound dataset contains valuable infrasound characteristics but is presented in various forms, lacking uniformity, which makes it unsuitable for direct use in machine learning models. Consequently, data preprocessing is essential to extract useful parameters and convert the infrasound data into a standardized format that can be effectively recognized by learning algorithms. Initially, missing values are imputed to ensure the completeness of the dataset. Following this, the input vectors are normalized to standardize the units, thereby preventing issues such as disproportionately large feature weights that could lead to increased model training time or gradient explosion problems.

In the experiments, the nearest neighbor interpolation method is used to fill missing data. The nearest neighbor interpolation method uses the previous and next values of the missing data (Sehar *et al.*, 2025). Let the value at time t_1 be x_1, at time t_2 be x_2, and at time t_3 be x_3. The missing value x_2 can be expressed as follows:

$$\frac{x_2 - x_1}{x_3 - x_1} = \frac{y_2 - y_1}{y_3 - y_1}. \tag{9}$$

For normalization, the min-max normalization method is used to map the results to the interval $[0, 1]$ (Mitropoulos *et al.*, 2022), as shown in Eq. (10). The original data x is normalized to x^*, where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values in the original data, respectively:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{10}$$

2.3.2. Data generation

In the data generation phase, the TimeGAN model is composed of four primary components: the embedding network, the recovery network, the sequence generator, and the sequence discriminator. The embedding and recovery networks fall under the autoencoder category, while the sequence generator and discriminator are part of the generative adversarial network framework. As a result, TimeGAN involves joint training of both the autoencoding and adversarial components. In the autoencoding section, the embedding network maps high-dimensional data into a lower-dimensional vector, or latent space, to capture essential feature information. The recovery network then reconstructs the data from this latent space back to its original dimensionality, minimizing the reconstruction loss L_R to optimize the representation of the latent space. Following the principle that the dynamics of complex systems are often driven by a smaller set of lower-dimensional factors, the adversarial component trains the sequence generator and discriminator within the latent space produced by the embedding network. This approach alleviates the challenges associated with high-dimensional data during the adversarial training process.

The embedding and recovery functions achieve the mapping from the feature space to the latent space, enabling the adversarial network to learn the potential time characteristics of the data through low-dimensional representations. Let H_S represent the latent vector space containing time-related feature S, and similarly, let H_X represent the latent vector space for the static feature X. The role of the embedding function e is to encode real-time sequences into the latent space, defined as $S \times \prod_t X \to H_S \times \prod_t H_X$. This function uses a recurrent neural

network (RNN) to perform the mapping, encoding both static and temporal features into low-dimensional latent vectors $h_S, h_{1:T} = e(S, X_{1:T})$ which are easier for the network to learn (Ruddick *et al.*, 2024). The embedding function is expressed as

$$\begin{cases} h_S = e_S(S), \\ h_t = e_X(h_S, h_{t-1}, X_t), \end{cases} \tag{11}$$

where $e_S : S \to H_S$ is the embedding function for static features, aimed at converting static features S into low-dimensional static features h_S through mapping, and $e_X : H_S \times H_X \times X \to H_X$ is the RNN-based embedding function for temporal features, aiming to map temporal features X_t into low-dimensional static features h_t. It follows causal ordering, meaning each step's output depends only on the preceding information.

The recovery function γ performs decoding, defined as $H_S \times \prod_t H_X \to S \times \prod_t X$. It uses a feedforward neural network (FNN) (Ruddick *et al.*, 2024) to restore the low-dimensional latent code back into high-dimensional static and temporal features $\widetilde{S}\widetilde{X}_{1:T} = \gamma(h_S, h_{1:T})$:

$$\begin{cases} \widetilde{S} = \gamma_S(h_S), \\ \widetilde{X}_t = \gamma_X(h_t), \end{cases} \tag{12}$$

where $\gamma_S : H_S \to S$ is the recovery function for static features, which is the inverse mapping of h_S, and similarly, $\gamma_X : H_X \to X$ represents the recovery network for temporal feature embeddings, which is the inverse mapping of h_t.

In the autoencoding part, the embedding function maps high-dimensional static and temporal features into a low-dimensional latent space, and the recovery function maps them back to high-dimensional features. Therefore, the embedding function and recovery function are reversible mappings existing between the feature space and latent space. They can accurately represent the high-dimensional reconstructed data $\widetilde{S}, \widetilde{X}_{1:T}$ using high-dimensional original data $S, X_{1:T}$ and low-dimensional latent vectors $h_S, h_{1:T}$. The reconstruction loss L_R of the autoencoder part is shown in Eq. (13), which represents the autoencoder's understanding of the intrinsic patterns in the input data (Ji *et al.*, 2020). By optimizing the reconstruction, the autoencoder can generate higher-quality low-dimensional latent representations:

$$L_R = E_{S, X_{1:T} \sim P}\left[\left\|S - \widetilde{S}\right\|_2 + \sum_t \left\|X_t - \widetilde{X}_t\right\|_2\right]. \tag{13}$$

During TimeGAN's training, two types of data are input into the sequence generator (Ji *et al.*, 2020). In the open-loop mode, the low-dimensional data $\widetilde{h}_S, \widetilde{h}_{1:T}$ generated by the generator is input into the sequence generator to obtain the next generated vector $\widehat{h}_t$. Then, by optimizing the unsupervised loss L_U, the probability of correctly classifying the real data $h_S, h_{1:T}$ and generated data $\widetilde{h}_S, \widetilde{h}_{1:T}$ is increased:

$$L_U = E_{S, X_{1:T} \sim P}\left[\log y_S + \sum_t \log y_t\right] + E_{S, X_{1:T} \sim P}\left[\log\left(1 - \widetilde{y}_S\right) + \sum_t \widetilde{\left(1 - \log y_t\right)}\right]. \tag{14}$$

Due to insufficient adversarial feedback from the sequence discriminator, the sequence generator does not fully capture the conditional distribution of the time steps in the real data (Tang *et al.*, 2023). Therefore, TimeGAN introduces a supervised loss to further constrain the model and alternates training in the closed-loop mode. The low-dimensional temporal latent sequence $h_{1:t-1}$ encoded by the embedding network is input into the sequence generator to obtain the latent vector for the next time step. Then, the supervised loss is optimized using the maximum likelihood method, which reflects the similarity between the data generated by the sequence generator and the data encoded by the autoencoder. This loss measures the difference between distributions $p(H_t | H_S, H_{1,t-1})$ and $\widehat{p}(H_t | H_S, H_{1,t-1})$. The supervised loss L_S obtained using the maximum likelihood method:

$$L_S = E_{S, X_{1:T} \sim P}\left[\sum_t \left\|h_t - g_X\left(h_S, h_{t-1}, z_t\right)\right\|_2\right]. \tag{15}$$

At each training step, the difference between the next latent vector from the embedding function and the next latent vector synthesized by the sequence generator needs to be evaluated. Although the unsupervised loss L_U can guide the sequence generator to create real sequences, the supervised loss L_S ensures that it generates smooth transitions.

2.3.3. Infrasound prediction

The specific process of the coordinated attention mechanism is shown in Fig. 4, which includes two steps: information embedding and attention generation.

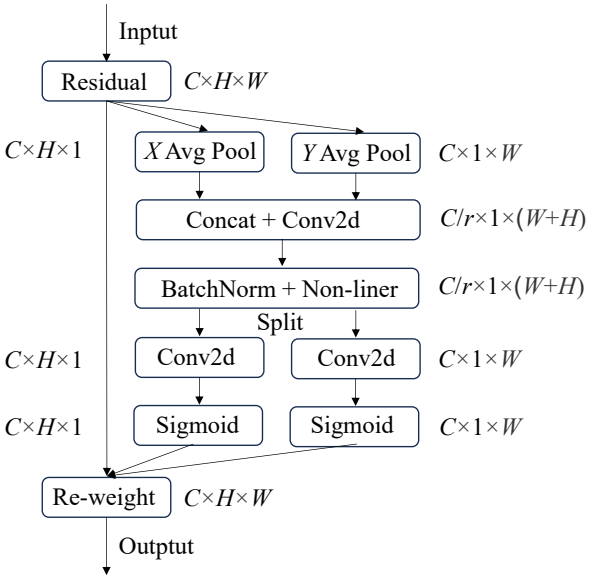


Fig. 4. Schematic of the coordinated attention generation process (Hou *et al.*, 2021).

The information embedding component plays a crucial role in enhancing the attention module's ability to capture a higher-quality global receptive field while preserving the accuracy of positional encoding (Tang *et al.*, 2023). In traditional channel attention, global pooling is commonly used to encode spatial information (Zhang *et al.*, 2024). However, this approach often compresses global spatial data into channels, making it challenging to retain precise positional information (Jiang *et al.*, 2025). To address this issue, the information embedding operation decomposes the global average pooling step by pooling separately along both the horizontal and vertical axes of the input features. This technique aggregates features from both spatial directions, resulting in two feature maps that retain directional information. As shown in Fig. 5, by performing transformations along both directions, long-range dependencies along one spatial axis and positional information along the other are captured by the attention module, enabling the network to more effectively localize key targets.

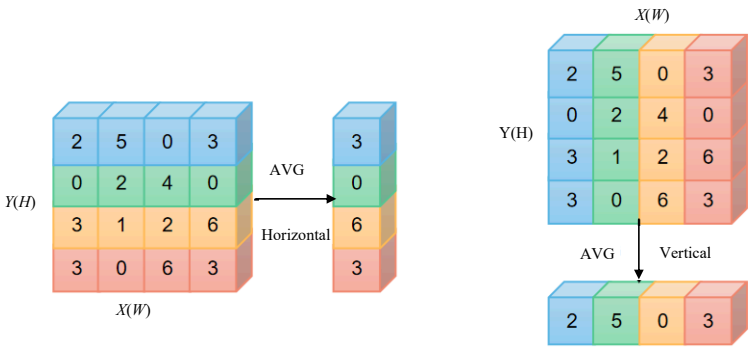


Fig. 5. Schematic of the coordinated attention information embedding operation.

Specifically, in the horizontal direction, the global average pooling operation uses a pooling kernel of size $H \times 1$ to compress the input feature X dimensions from $H \times W \times C$ to $H \times 1 \times C$:

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \le a \le W} x_C(h,a) z_C^h \in R^{C*H*1}. \tag{16}$$

In the vertical direction, the global average pooling operation uses a pooling kernel of $1 \times W$ size to compress the input feature dimensions from $H \times W \times C$ to $H \times 1 \times C$:

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \le b \le H} x_C(b,w) z_C^w \in R^{C*1*W}. \tag{17}$$

The attention generation operation aims to fully utilize the positional information encoded in the embedding operation and capture the regions of interest and relationships between channels. Specifically, the feature maps from the two directions, Z_C^h and Z_C^w, are concatenated along the channel dimension, and then convolution operations are applied using a shared convolutional transformation function F_1, obtaining intermediate feature maps f, $f \in R^{\frac{C}{r} \times (H+W)}$, which encode both horizontal and vertical directions:

$$f = \delta\left(F_1\left(\left[z^h, z^w\right]\right)\right), \qquad f \in R^{\frac{C}{r}*1*(H+W)}, \tag{18}$$

where δ is the non-linear activation function ReLU, $[\cdot, \cdot]$ represents the concatenation along the spatial dimension.

Then, the intermediate feature map f is split along the spatial dimensions into two feature maps f^h and f^w, $f^h \in R^{\frac{c}{r} \times H}$, $f^w \in R^{\frac{c}{r} \times W}$. Each feature map is upsampled using convolution operations F_h and F_w, obtaining two directional attention weights g^h and g^w, $g^h \in R^{C \times H \times 1}$, $g^w \in R^{C \times 1 \times W}$, as follows:

$$g^h = \sigma\left(F_h\left(f^h\right)\right), \tag{19}$$

$$g^w = \sigma\left(F_w\left(f^w\right)\right). \tag{20}$$

Finally, the attention weights g^h and g^w are multiplied with the original features x_C to obtain the scaled features y_C:

$$y_C(a,b) = x_C(a,b) \times g_C^h(a) \times g_C^w(b). \tag{21}$$

2.4. Data set

This study utilizes infrasound data provided by the international monitoring system (IMS) with support from the Comprehensive Nuclear-Test-Ban Treaty Beijing National Data Center. A total of 611 infrasound data sets are collected from six distinct infrasound sensor arrays located globally. These data sets are categorized into three types of infrasound events: earthquake, tsunami, and volcano. All infrasound recordings have a sampling frequency of 20 Hz. Table 1 presents the details of the infrasound data collected from various regions, while Fig. 6 illustrates the geographical distribution of the infrasound stations.

Table 1. Information of infrasound data.

Event type	Data source (IMS station code)	Geographic coordinate	Number of signals	Total	Sampling frequency [Hz]
Earthquake	I14CL	(−33.65, −78.79)	74	203	20
	I30JP	(35.31, 140.31)	124		20
	I59US	(19.59, −155.89)	6		20
Tsunami	I10CA	(50.20, −96.03)	4	218	20
	I22FR	(−22.18, 166.85)	53		20
	I30JP	(35.31, 140.31)	113		20
	I52GB	(−7.38, 72.48)	66		20
Volcano	I30JP	(35.31, 140.31)	189	189	20

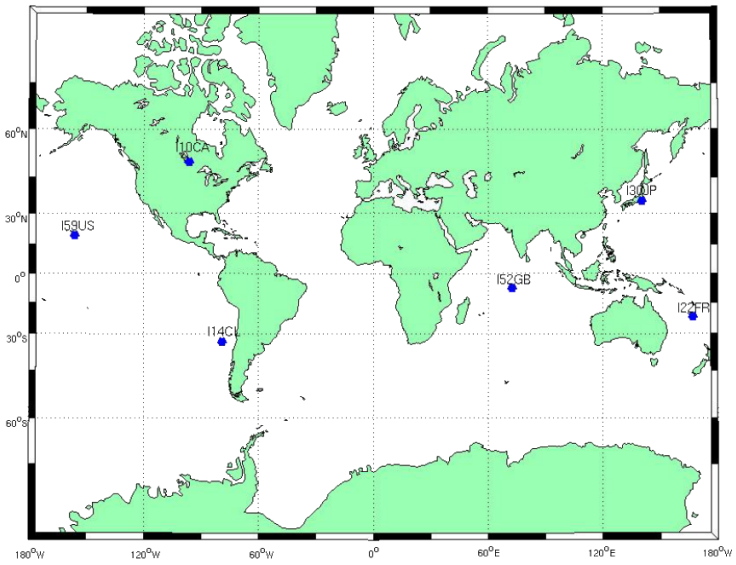


Fig. 6. Map of the infrasound station.

3. Experiments

3.1. Experiments setup

The operating system used in this study is Windows 11, with CUDA 10.0 and cuDNN 7.4 for accelerated training. The hardware used includes an NVIDIA Quadro P4000 (8 GB memory). The network development framework is TensorFlow 1.14, and programming is done in Python. The CPU used is an Intel(R) Core(TM) i5-11320H CPU @ 3.20 GHz, 2.5 GHz. As described in Subsec. 2.2, the key parameters of the CAPN are summarized in Table 2. The simulation validation focuses on applying the infrasound signal data to assess the feature learning performance of the proposed CAPN model. Each infrasound signal consists of 10 400 data points. The dataset is divided into training and testing samples. The input map size for the CAPN model is $128 \times 128 \times 1$. The number of iterations is set to 60.

Table 2. Parameter of CAPN.

Number of layer	Layer type	Kernel size	Filters
1	Convolution 1	12×12	4
2	Maxpooling 1	5×5	–
3	Convolution 2	7×7	4
4	Maxpooling 2	5×5	–
5	Convolution 3	5×5	8
6	Maxpooling 3	5×5	–
7	Flatten	–	–
8	Fully-connected	–	–
9	Softmax	–	–

3.2. Data preprocessing

The infrasound data collected in this study are smoothed to effectively eliminate noise. Figure 7a displays the original infrasound signal, which contains substantial noise. To reduce computational complexity, a moving average filtering method is applied for smoothing, with the resulting signal shown in Fig. 7b. Details of the moving average filtering method can be found in (Mitropoulos *et al.*, 2022). A total of 70 % of the smoothed data are used as the training set, while the remaining 30 % are allocated as the testing set. Finally, data standardization and normalization are performed using Eq. (9) and Eq. (10).

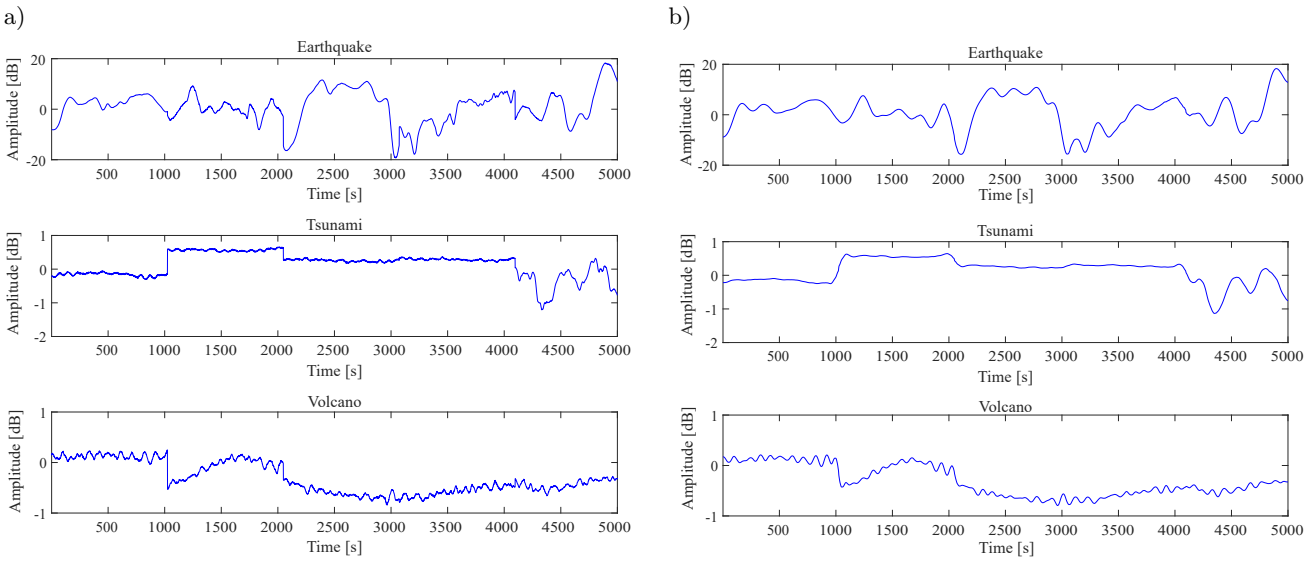


Fig. 7. Original (a) and preprocessing (b) signal.

3.3. Data generation

To evaluate the quality of the data generated by the model, both fidelity and diversity are taken into account. Fidelity refers to the degree to which the generated samples resemble the real data, while diversity ensures that the generated samples do not exhibit excessive similarity to each other. Thus, the performance of the generated model is assessed from both qualitative and quantitative perspectives.

3.3.1. Discriminator score

The performance of the generative model is quantitatively assessed from the quality and diversity of the generated samples. In this study, the maximum mean discrepancy (MMD) metric is used to evaluate the generative model based on the difference in sample distributions. MMD is used to measure the distance between two distributions in a Hilbert space. Thus, for the generative model, this metric can measure the distance between the original data distribution P_o and the generated data distribution P_g. The smaller the MMD distance, the more similar the distributions of the original and generated data are, indicating higher quality of the generated samples and better model performance (Wang *et al.*, 2021).

When calculating the MMD distance, the Gaussian kernel function $K(x, y)$ is used to map the two samples into a real number:

$$K(x, y) = \exp\left(-\|x - y\|^2\right). \tag{22}$$

The MMD distance $D_{\mathrm{MMD}}(P_o, P_g)$ is expressed as follows:

$$D_{\mathrm{MMD}}(P_o, P_g) = E_{x,x'\sim P_o}\left[K(x, x')\right] - 2E_{x\sim P_o, y\sim P_g}\left[K(x, y)\right] + E_{y,y'\sim P_g}\left[K(y, y')\right]. \tag{23}$$

In this experiment, both the original data and generated data have four state types: normal, earthquake, volcano, and tsunami. Therefore, the distribution of the original data is denoted as $P_{oi}, i = (1, 2, 3, 4)$, and the distribution of the generated data is denoted as $P_{gi}, i = (1, 2, 3, 4)$. As shown in Fig. 8, the MMD distance between the original and generated data is calculated for infrasound data after applying several models, including GAN, LSTM, GRU, transformer, WGAN, DCGAN, and TimeGAN. Compared to the other generative adversarial network models, TimeGAN exhibits the smallest MMD distance between the original and generated data. Notably, for the volcano data, the MMD distance is 1.364 times smaller for TimeGAN than for GAN, indicating that the distribution of TimeGAN-generated data closely matches the original data distribution, resulting in superior

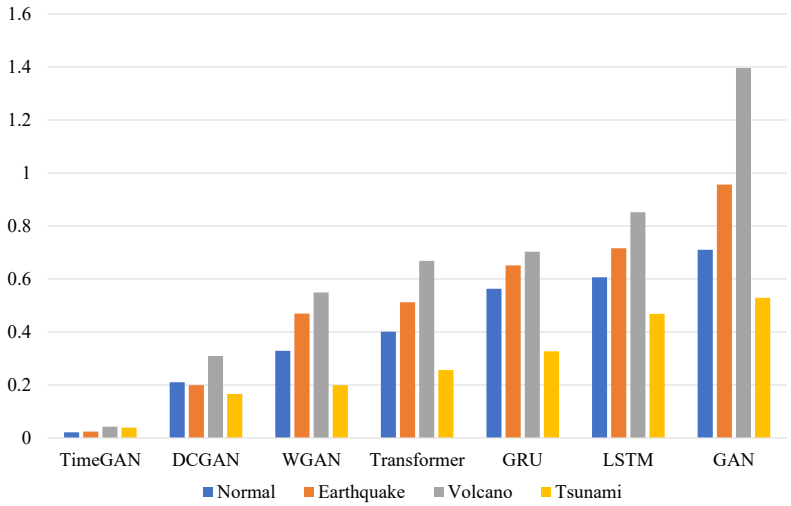


Fig. 8. MMD distances between generated data from different models and real data.

model performance. In contrast, the GAN model shows the largest discrepancy between the generated and original data, making it the least effective model. For the tsunami data, all four generative models show relatively high performance due to the distinct infrasound characteristics. However, for volcano data, where infrasound features are less pronounced, all models exhibit the largest MMD distance, suggesting a greater challenge in accurately modeling such data. Therefore, it can be concluded that TimeGAN generates relatively high-quality samples, outperforming the other models in terms of data fidelity.

In addition, the MMD metric is also used to evaluate the diversity of the generated samples, albeit with a slightly different focus. Here, the goal is to measure the variability between the sample distributions within the generated data. Specifically, the MMD distance between the distributions of individual samples is calculated, and the mean of these distances is taken as the internal MMD distance of the generated data. A higher value of this distance indicates greater variability between the samples, reflecting higher diversity in the generated data and superior performance of the GAN.

Let the number of distribution samples be 1, and let P_i and P_j represent the source distributions of two different samples (Wang *et al.*, 2021). The MMD distance is given by

$$D_{\mathrm{MMD}}(P_i, P_j) = 2 - 2E_{x_i \sim P_i, x_j \sim P_j}\left[K(x_i, x_j)\right] i \neq j. \tag{24}$$

Then, the internal MMD distance of the generated data, which measures the diversity within the samples, is given by

$$D_r = \frac{1}{1 + 2 + \ldots + N - 1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} D_{\mathrm{MMD}}(P_i, P_j). \tag{25}$$

In the experiment, the internal MMD distances of the generated data from three classical generative adversarial networks and the proposed TimeGAN model are calculated, with the results presented in Fig. 9. The analysis reveals that, among the seven generative models, TimeGAN produces data with the largest internal MMD distance, indicating that it generates data with higher diversity. This is particularly evident in the tsunami data. In contrast, GAN and WGAN perform the worst in generating diverse samples. For example, in the tsunami case, the internal MMD distance of the data generated by GAN is 0.0024 smaller than that of TimeGAN, WGAN is 0.002 smaller, and LSTM is 0.0018 smaller than TimeGAN. These results demonstrate that TimeGAN outperforms other models in terms of generating diverse and varied infrasound samples.

A comprehensive analysis of the MMD metric reveals that, in most cases, GAN, WGAN, and LSTM models only offer rough approximations of the original data, with limited quality and diversity in the generated samples. In contrast, TimeGAN, DCGAN, transformer, and GRU models generate data with higher quality and greater

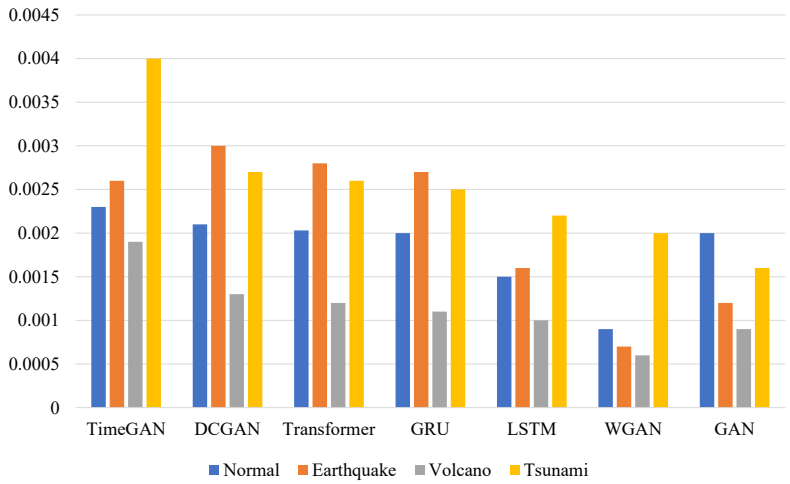


Fig. 9. Internal MMD distances of generated data from different models.

diversity. To further validate the effectiveness of TimeGAN in infrasound prediction, a detailed performance comparison is conducted between these models.

3.3.2. Visualization

In the previous MMD analysis, TimeGAN and DCGAN demonstrated superior performance, and thus, these two models are the focus of further analysis. To qualitatively evaluate the effectiveness of the proposed method, the t-SNE and PCA techniques are applied to visualize the distribution of the generated and original samples in a 2D space. Figure 10 presents the results of the PCA and t-SNE visualizations, where red points represent the feature distribution of the real infrasound data, and blue points represent the feature distribution of the generated infrasound data. The closer the two sets of points are to each other, the better the model's performance, indicating that the distribution of the generated samples closely matches that of the real data.

From the analysis, the feature distribution of the data generated by TimeGAN closely aligns with the feature profile of the original data, demonstrating a high degree of similarity. In contrast, DCGAN fails to generate certain features that are present in the original data, resulting in a mismatch between the feature distribution of the generated and original data. When augmenting time-dependent data, TimeGAN significantly outperforms DCGAN, showcasing its superior performance in capturing temporal dynamics.

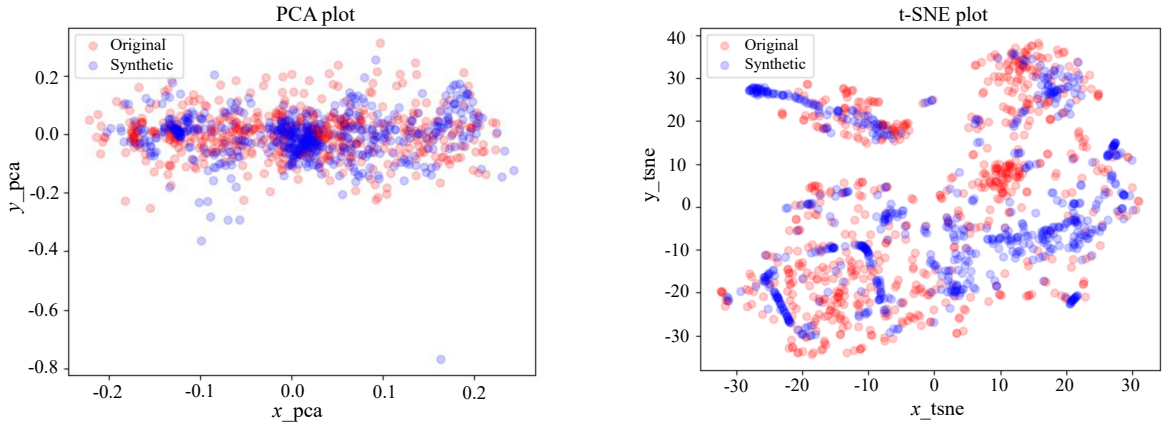
3.4. Infrasound prediction

Considering that the model is designed for infrasound prediction tasks with small sample sizes, the generated samples must be as effective as the real data samples. These samples should serve to augment the dataset, ensuring that each sample contains sufficient information to improve the performance of the infrasound prediction model. The generated data is combined with the original data, with the amount of synthetic data being twice the size of the original dataset. Additionally, three consecutive data points are grouped together to form a single sample. The synthesized dataset is then used to train the diagnostic model, which is subsequently tested on a separate test set. The infrasound prediction accuracy after training is used as an evaluation metric to assess whether the inclusion of the generated data enhances the model's predictive capability, particularly in small sample scenarios.

The results shown in Fig. 11 represent the accuracy after 60 iterations of the model. From the figure, it can be observed that the data generated by the classical GAN-CAPN and LSTM-CAPN model reduces the accuracy by 5.76 % and 3.64 % compared to CAPN, which decreases the performance of the infrasound prediction model. Apart from the GAN-CAPN and LSTM-CAPN model, datasets that were not augmented with a generative

a) PCA dimensionality reduction result of DCGAN

b) t-SNE dimensionality reduction result of DCGAN



c) PCA dimensionality reduction result of TimeGAN

d) t-SNE dimensionality reduction result of TimeGAN

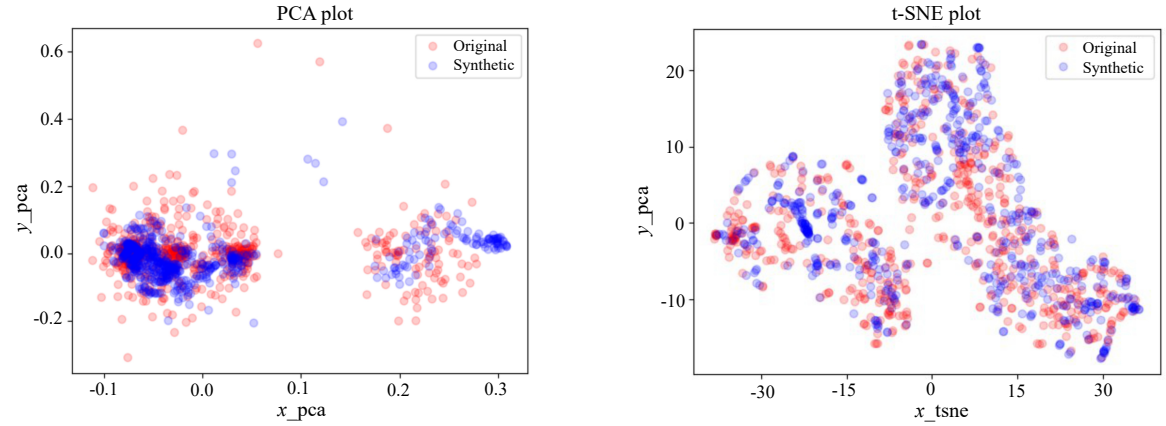


Fig. 10. WGAN and TimeGAN data visualization.

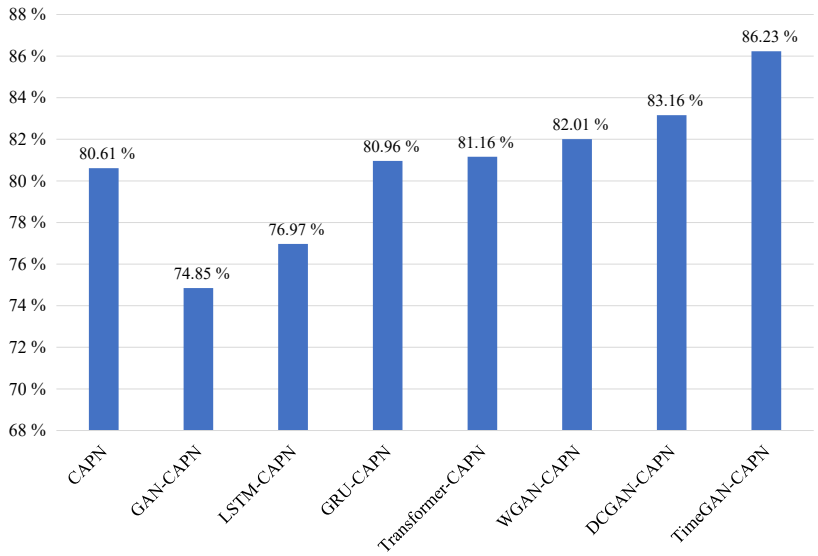


Fig. 11. Comparison of diagnostic accuracy after sample augmentation.

model perform poorly in infrasound prediction after training. Compared to the DCGAN-CAPN, WGAN-CAPN, Transformer-CAPN, and GRU-CAPN models, the TimeGAN-CAPN model generates higher-quality data by considering the internal temporal correlations in the data, effectively addressing the issue of insufficient information in the samples. The prediction accuracy improved by 5.62 % compared to when no augmentation was performed.

Therefore, using the TimeGAN model to augment the infrasound data and inputting the augmented data into the CAPN infrasound prediction model can significantly improve the infrasound prediction accuracy.

Table 3 shows a comparison of the classification performance of eight methods for infrasound signals. The experimental results indicate that TimeGAN-CAPN achieves the best overall classification precision, reaching 84.62 %. In addition, TimeGAN-CAPN also significantly outperforms the other seven classification methods in terms of $F1$-score and recall, with values of 86.02 % and 87.26 %, respectively.

Table 3. Comparison of classification results for four types of infrasound events by different classification networks [%].

Method	$F1$-score	Recall	Precision
CAPN	80.07	79.62	79.16
GAN-CAPN	74.39	73.68	73.07
LSTM-CAPN	76.27	75.89	75.25
GRU-CAPN	80.29	79.68	78.97
Transformer-CAPN	80.76	80.92	79.13
WGAN-CAPN	81.96	82.17	80.86
DCGAN-CAPN	83.11	84.06	81.71
TimeGAN-CAPN	86.02	87.26	84.62

Further analysis shows that, compared to other classification networks, TimeGAN-CAPN exhibits higher classification accuracy and more stable classification performance in infrasound classification. To provide a comprehensive evaluation of its performance, Fig. 12 presents the ROC curve for different methods. From the figure, it is evident that the AUC value of TimeGAN-CAPN reaches 0.8451, significantly higher than the other seven networks, further validating its superior performance in the infrasound signal classification task.

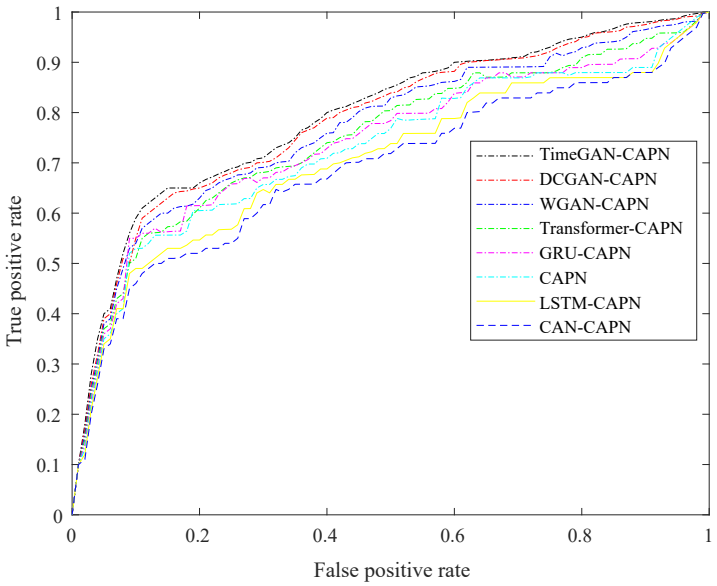


Fig. 12. ROC curve of different classification methods.

To analyze the infrasound recognition performance, eight classification models were evaluated using 6-fold cross-validation to obtain the accuracy of real labels and predicted labels from six validation runs. The confusion matrix for infrasound classification is shown in Fig. 13. From the perspective of single-class classification performance, TimeGAN-CAPN demonstrates significant advantages in classifying earthquake, tsunami, and volcanic infrasound signals. This result thoroughly confirms the robustness and generalization ability of the proposed method in handling different types of infrasound signals.

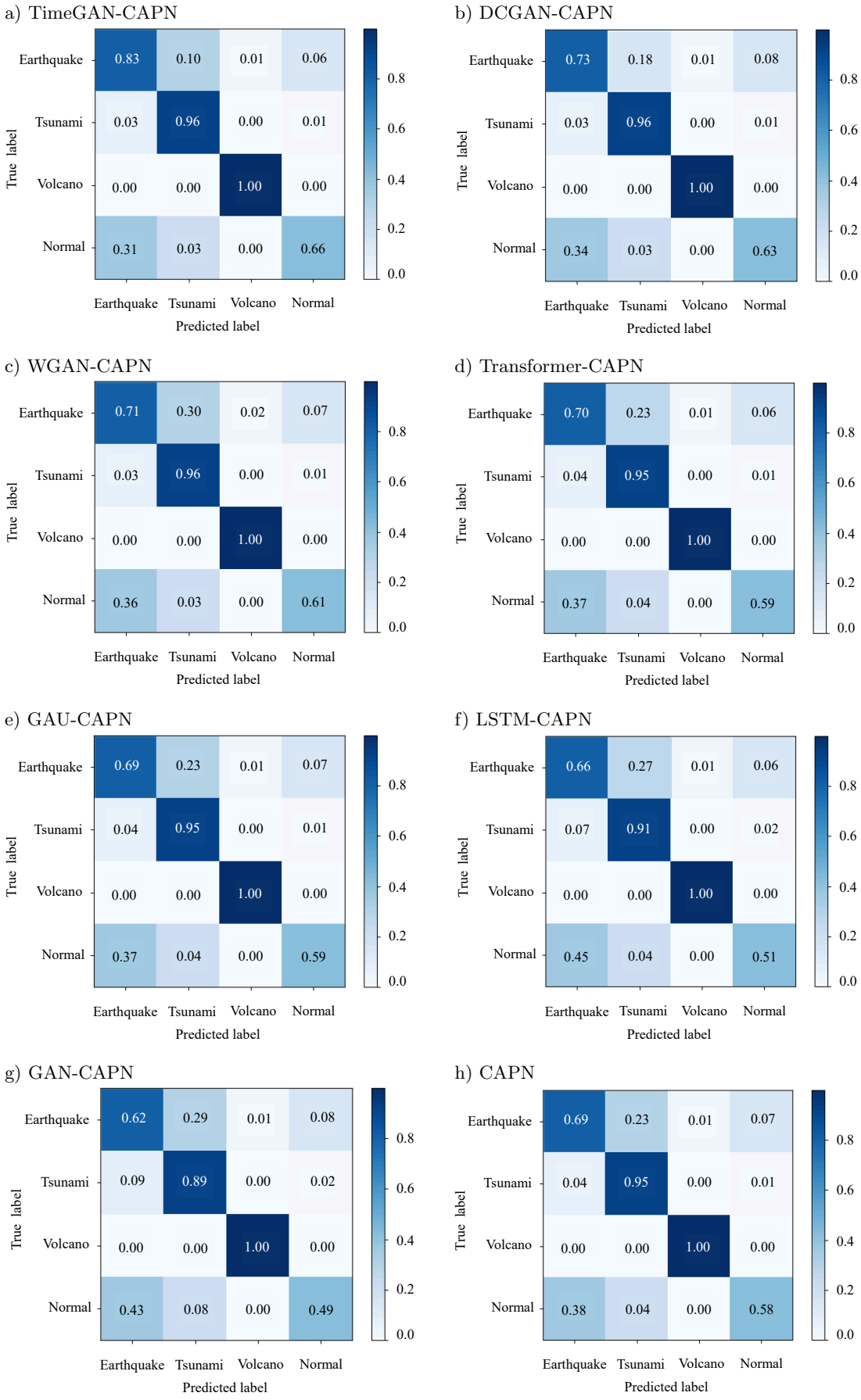


Fig. 13. Confusion matrix for infrasound classification.

4. Conclusion and future work

To further enhance the accuracy of infrasound disaster prediction, this paper proposes the TimeGAN-CAPN prediction model. The TimeGAN-CAPN model combines unsupervised and supervised learning, where the autoencoder component provides an embedding space for temporal features. The generative component operates within this embedding space to produce high-quality sequential data. By augmenting the sample data, the model increases the information content, and these new samples are then input into the CAPN to more effectively capture class prototypes, further improving prediction performance. The quality and diversity of the generated data are quantitatively and qualitatively assessed using the MMD metric and visualization methods, demonstrating that TimeGAN-CAPN generates data that closely approximates the original distribution. Comparative experiments highlight the superior predictive performance of TimeGAN-CAPN.

Although the TimeGAN-CAPN model improves infrasound prediction accuracy, it does not transform data into 2D images as seen in traditional fault diagnosis models due to the limited quantity and lack of periodicity in the data. As a result, the TimeGAN-CAPN model is specifically tailored for infrasound data. Future research could explore the use of transfer learning techniques to apply the trained model to different disaster datasets, thereby improving the model's generalizability.

Fundings

This work was supported by the Project Supported by the Chongqing Natural Science Foundation General Project (grant no. CSTB2025NSCQ-GPX0115), the Scientific and Technological Research Program of Chongqing Municipal Education Commission (grant no. KJQN202403106), and the 'Spark' Program Project for Teachers' Independent Innovation at Chongqing Polytechnic University of Electronic Technology (grant no. 25XJJSCX11).

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.

Authors' contributions

Quanbo Lu conceptualized the study, performed the data analysis and interpretation, wrote the original draft. Xiaojuan Huang completed literature review, performed the data collection and analysis, and assisted in manuscript revision. Rao Li led in-depth data analysis and result interpretation, verified research hypotheses, and drafted Sec. 3. Mei Li advised on the research methodology and tool selection, and assisted in manuscript revision. Dong Zhu completed literature review and theoretical foundation construction, and summarized research background and existing studies. All authors reviewed and approved the final manuscript.

Acknowledgments

We would like to thank the Comprehensive Nuclear-Test-Ban Treaty Beijing National Data Center for providing the data.

References

1. Baeza Moyano D., González Lezcano R.A. (2022), Effects of infrasound on health: Looking for improvements in housing conditions, *International Journal of Occupational Safety and Ergonomics*, **28**(2): 809–823, https://doi.org/10.1080/10803548.2020.1831787.

2. Dong H. *et al.* (2024), Enhanced infrasound denoising for debris flow analysis: Integrating empirical mode decomposition with an improved wavelet threshold algorithm, *Measurement*, **235**: 114961, https://doi.org/10.1016/j.measurement.2024.114961.

3. Friedrich B., Joost H., Fedtke T., Verhey J.L. (2023), Effects of infrasound on the perception of a low-frequency sound, *Acta Acustica*, **7**: 60, https://doi.org/10.1051/aacus/2023061.

4. Hou Q., Zhou D., Feng J. (2021), Coordinate attention for efficient mobile network design, [in:] *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13713–13722, https://doi.org/10.1109/CVPR46437.2021.01350.

5. Hupe P., Ceranna L., Le Pichon A., Matoza R.S., Mialle P. (2022), International monitoring system infrasound data products for atmospheric studies and civilian applications, *Earth System Science Data Discussions*, **14**(9): 4201–4230, https://doi.org/10.5194/essd-14-4201-2022.

6. Ji Z., Liu X., Pang Y., Ouyang W., Li X. (2020), Few-shot human-object interaction recognition with semantic-guided attentive prototypes network, *IEEE Transactions on Image Processing*, **30**: 1648–1661, https://doi.org/10.1109/TIP.2020.3046861.

7. Jiang Y.H. *et al.* (2025), Recursive prototypical network with coordinate attention: A model for few-shot cross-condition bearing fault diagnosis, *Applied Acoustics*, **231**: 110442, https://doi.org/10.1016/j.apacoust.2024.110442.

8. Listowski C. *et al.* (2022), Remote monitoring of Mediterranean hurricanes using infrasound, *Remote Sensing*, **14**(23): 6162, https://doi.org/10.3390/rs14236162.

9. Lu Q.B., Li M. (2023), VMD and CNN-based classification model for infrasound signal, *Archives of Acoustics*, **48**(3): 403–412, https://doi.org/10.24425/aoa.2023.145247.

10. Macpherson K.A., Fee D., Colwell J.R., Witsil A.J. (2023), Using local infrasound to estimate seismic velocity and earthquake magnitudes, *Bulletin of the Seismological Society of America*, **113**(4): 1434–1456, https://doi.org/10.1785/0120220237.

11. Mitropoulos S., Toulas V., Douligeris C. (2022), A prototype network monitoring information system: Modelling, design, implementation and evaluation, *International Journal of Information and Communication Technology*, **21**(2): 111–136, https://doi.org/10.1504/IJICT.2022.124807.

12. Ruddick K.G. *et al.* (2024), WATERHYPERNET: A prototype network of automated in situ measurements of hyperspectral water reflectance for satellite validation and water quality monitoring, *Frontiers in Remote Sensing*, **5**: 1347520, https://doi.org/10.3389/frsen.2024.1347520.

13. Sehar U., Xiong J., Xia Z. (2025), Automatic tooth labeling after segmentation using prototype-based meta-learning, *Machine Intelligence Research*, **22**: 539–552, https://doi.org/10.1007/s11633-024-1520-6.

14. Sharma S.K., Alenizi A., Kumar M., Alfarraj O., Alowaidi M. (2024), Detection of real-time deep fakes and face forgery in video conferencing employing generative adversarial networks, *Heliyon*, **10**(17): e37163, https://doi.org/10.1016/j.heliyon.2024.e37163.

15. Sovilla B. *et al.* (2025), The dominant source mechanism of infrasound generation in powder snow avalanches, *Geophysical Research Letters*, **52**(2): e2024GL112886, https://doi.org/10.1029/2024GL112886.

16. Tang T. *et al.* (2023), An improved prototypical network with L2 prototype correction for few-shot cross-domain fault diagnosis, *Measurement*, **217**: 113065, https://doi.org/10.1016/j.measurement.2023.113065.

17. Vuletić M., Prenzel F., Cucuringu M. (2024), Fin-GAN: Forecasting and classifying financial time series via generative adversarial networks, *Quantitative Finance*, **24**(2): 175–199, https://doi.org/10.1080/14697688.2023.2299466.

18. Wang W. *et al.* (2021), Rethinking maximum mean discrepancy for visual domain adaptation, *IEEE Transactions on Neural Networks and Learning Systems*, **34**(1): 264–277, https://doi.org/10.1109/TNNLS.2021.3093468.

19. Watson L.M. *et al.* (2022), Volcano infrasound: Progress and future directions, *Bulletin of Volcanology*, **84**: 44, https://doi.org/10.1007/s00445-022-01544-w.

20. Wilson T.C., Petrin C.E., Elbing B.R. (2023), Infrasound and low-audible acoustic detections from a long-term microphone array deployment in Oklahoma, *Remote Sensing*, **15**(5): 1455, https://doi.org/10.3390/rs15051455.

21. Yang S. *et al.* (2025), Correlation between and mechanisms of gas desorption and infrasound signals, *Natural Resources Research*, **34**: 515–537, https://doi.org/10.1007/s11053-024-10417-2.

22. Yoon J., Jarrett D., van der Schaar M. (2019), Time-series generative adversarial networks, [in:] *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, **32**.

23. Zajamsek B., Hansen K.L., Nguyen P.D., Lechat B., Micic G., Catcheside P. (2023), Effect of infrasound on the detectability of amplitude-modulated tonal noise, *Applied Acoustics*, **207**: 109361, https://doi.org/10.1016/j.apacoust.2023.109361.

24. Zhang B., Xu M., Zhang Y., Ye S., Chen Y. (2024), Attention-ProNet: A prototype network with hybrid attention mechanisms applied to zero calibration in rapid serial visual presentation-based brain–computer interface, *Bioengineering*, **11**(4): 347, https://doi.org/10.3390/bioengineering11040347.

Archives of Acoustics Vol. 51, No. 1, pp. 57–74 (2026), https://doi.org/10.24423/archacoust.2026.4222

Research Paper

Comparative Perceptual Assessment of Sound Quality: Cone versus Distributed Mode Loudspeakers

Piotr Kleczkowski, Paweł Małecki*, Dorota Młynarczyk

AGH University of Krakow
Kraków, Poland

*Corresponding Author: pawel.malecki@agh.edu.pl

Received March 25, 2025; revised November 26, 2025; accepted January 12, 2026; available online January 19, 2026; version of record March 13, 2026; published issue March 27, 2026.

A conventional cone loudspeaker has a limited capacity for creating the impression of spatiality, while a distributed mode loudspeaker (DML) has an inherent ability to evoke it. DMLs have their specific drawbacks, but some of these can be compensated for. A key question arises – is it a cone loudspeaker or a compensated DML that is preferred by listeners? A listening experiment with carefully controlled conditions was carried out to answer this question; 30 subjects participated. The participants evaluated three stereo systems: one based on a DML speaker (with its power response equalized) and two conventional two-way active systems. Two perceptual attributes were evaluated: overall preference and spatial impression. A graded pairwise comparison was used as an experimental paradigm; the results were analyzed according to the law of comparative judgment. The findings indicated that, even though the DMLs achieved slightly lower ratings than the conventional systems on average, the perceptual differences were very small. This was confirmed by the hypothesis testing that was performed on the raw results of the pairwise comparisons.

Keywords: distributed mode loudspeakers, loudspeaker evaluation, spatial sound, pairwise comparison, listening experiment.

1. Introduction

The relationships among the directivity of loudspeakers, the acoustics of home listening rooms, and perception have been investigated by a number of researchers (Evans *et al.*, 2009; Toole, 1986a; 1986b; 2018; Bertland, 1985; Olive, 2004a; 2004b; Zacharov, 1998). Key findings can be summarized as follows: a joint indicator that combines both the directivity of a loudspeaker and the acoustic properties of a room is the ratio of the direct to the early-reflected sounds. This indicator is closely related to the perception of reproduced sounds, with its higher values (more-direct sound) favoring the accurate localizations of sound sources and its lower values favoring the perception of space. Widening the radiation of the loudspeaker and increasing the reflectivity of the room boundaries both reduce the ratio of the direct to the early-reflected sounds. Reducing the lateral reflections in a listening room tends to have the same effect as narrowing the loudspeaker dispersion (Moulton, 1986). There are several works that support the opinion that the wide directivity of loudspeakers is preferable when compared to the directivity of conventional loudspeakers (Allison, 1995; Bertland, 1985; Ferralli, Moulton, 1995; Flindell *et al.*, 1991; Linkwitz, 2007; Moulton, 1986). Extensive work on the subject was carried out by Toole (1986a; 1986b; 2018), who noticed that wide dispersion loudspeakers were preferred by listeners – especially for recreational listening (but not exclusively). The preference for more- or less-dry listening conditions depends on the purpose of the listening; audio engineers favor drier spaces for their work, while more reverberance is preferred for recreational listening.

Loudspeaker systems that are based on electrodynamic units with cone-shaped diaphragms are incapable of wide radiation – even when they employ dome-shaped tweeter units. There is an unconventional type of loudspeaker that offers very wide radiation as its inherent property: the distributed mode loudspeaker (DML), which belongs to a wider class of flat panel loudspeakers.

A DML employs a different sound-radiating element than a conventional dynamic one does. In the DML, the radiator is a stiff flat panel with a rectangular shape and considerable mass. An electrodynamic or piezoelectric exciter is attached to the panel and induces uniformly distributed bending wave vibrations. This is entirely different than with a dynamic speaker, which was designed to vibrate like a rigid piston. In what follows, a loudspeaker system that is based on dynamic units will be referred to as a piston loudspeaker system (PLS).

DML technology began to attract the interest of researchers at the turn of this century, and numerous works on the subject have been published. An introduction to the technology can be found in (Angus, 2000; Bank, Harris, 1998; Harris, Hawksword, 2000; Newell, Holland, 2019), and a review of its history was written by Heilemann *et al.* (2021).

DMLs are normally mounted in walls or ceilings (i.e., they become architectural loudspeakers – this mounting is also referred to as flush mounting), which is one of their advantages. With this mounting, their directivity can be described as quasi-omnidirectional in the hemisphere. Comprehensive anechoic measurements were presented in (Bai, Huang, 2001; Czesak, Kleczkowski, 2023); the directivity characteristics were irregular, but the property of omnidirectionality was maintained.

DMLs have another unique property besides quasi omnidirectionality: they behave like many sound sources; thus, their radiation is incoherent (Azima, Harris, 1997; Gontcharov, Hill, 2000; Harris *et al.*, 2000). Its advantage is that the interference of the direct sound with the first reflections is largely suppressed; thus, the comb-filtering effect is reduced, and the advantage of the reflections (i.e., the perception of space) is maintained. The effect of using an incoherent sound source is similar to applying acoustic diffusors. Wendt and Höldrich (2021) analyzed the consequences of specular and diffuse reflections on the precedence effect.

As a consequence of their principle of operation (which consists of the excitation of modal frequencies), DMLs have irregular frequency responses that are far from being flat and bring coloration. This drawback could be the main reason why DMLs have never been considered to be high-end devices. Nevertheless, their development is ongoing (Bai, Huang, 2001; Heilemann *et al.*, 2021; Jeon *et al.*, 2020; Jung *et al.*, 2021; Lu, Shen, 2009; Lu *et al.*, 2012; Yu *et al.*, 2023; Zenker *et al.*, 2020).

It is likely that improvements to the DML technology, new areas of application (like the screens in OLED TVs), and the current trend in the home-entertainment market toward the use of architectural loudspeakers with multi-speaker and multi-room installations will bring another wave of interest in DMLs. Therefore, a comparison of the perceptual qualities of DMLs and PLSs is a timely topic.

There have been a few works that have been published on the perceptual properties of DMLs – especially as compared to PLSs. Harris *et al.* (1998) found that DMLs improved the stereo localization of pink noise stimuli when compared to PLSs in an untreated room. Flanagan and Moore (2000) showed that detecting a spectral ripple was easier from a DML than it was from a PLS, but the accuracy of the vowel identification was similar for the two loudspeakers (Flanagan, Moore, 2001). Flanagan and Harris (1999) proposed a hypothesis that the loudness attenuation with distance in a given space was reduced by the use of a DML.

Heilemann *et al.* (2018) performed anechoic measurements of three different types of DMLs (one- and multi-exciter) and a two-way PLS. They used a prediction model of loudspeaker preferences (Olive, 2004b) and obtained an objective evaluation that strongly favored the PLS. Roessner *et al.* (2019) performed a listening comparison among two multi-exciter DML prototypes, one one-exciter commercial DML, and two two-way passive PLSs, in monophonic reproduction. The two PLSs obtained the highest scores, while the two prototype DMLs scored about 10 % lower. The one-exciter DML scored distinctively lower than the others.

Newell and Holland (2019) reported an experiment where conventional loudspeakers were used as a stereo pair and four DMLs were used to reproduce separately recorded ambient signals (with a remarkably realistic effect).

The purpose of this work was to evaluate the perceptual properties of an example DML by a subjective perceptual comparison with two examples of professional quality active PLSs in a stereo listening format. A key

assumption for the experiment was to reduce the main disadvantage of DMLs (i.e., an irregular frequency response), as it was likely that it could override the acoustic advantages of the DML technology. The smoothing of the frequency response was achieved by equalization (which is an easily accessible technological option).

Although the DML technology has found wide range of applications, no thorough perceptual experiment on DMLs with carefully controlled conditions satisfying the requirements for an objective comparison, participated by a relatively large panel of listeners, and other than a monophonic format was known to the authors at the time of writing this article.

2. Method

2.1. Assumptions for experiment

The experimental design was based on the following assumptions:

1. It is widely agreed that the flatness of the amplitude response is the key factor in the preference ratings of loudspeakers (Gabrielsson *et al.*, 1991; Olive, 2004a). Therefore, it was decided to implement equalization in order to avoid the overriding of the results by the inherent irregularity of the frequency responses of DMLs. The possible generalization of results required a simple and robust method of equalization. Magnitude-only equalization was chosen.
2. The most widespread listening format was assumed, i.e., stereophonic.
3. It was decided to use pairwise comparison as an experimental method due to its sensitivity to small perceptual differences (International Telecommunication Union [ITU], 2015). To keep the durations of the listening sessions within recommended limits (Bech, Zacharov, 2006), only three loudspeaker systems could be used; this resulted in the decision to compare three pairs.
4. The pilot listening evaluation of an equalized DML speaker indicated that the overall sound quality was comparable to a PLS of good quality. Therefore, three systems were elected for comparison: one DML, and two near-field two-way active monitors as examples of high-end PLSs. As an example of DML, a commercial unit from the line of the most widely available full DML-type panels from Amina Technologies Ltd. was selected – a one-exciter Edge 5 model. Two PLSs were chosen: a pair of Dynaudio BM15s, and a pair of Genelec 8030s. To minimize any commercialism, the DML is referred to as DML, the Dynaudio PLS as PLS_D, and the Genelec PLS as PLS_G throughout the rest of this paper.
5. To implement the pairwise comparison, fast switching between sound sources was necessary. Therefore, all units had to be permanently installed during each experimental session.

2.2. Positions of loudspeakers

An important advantage of DMLs is the ease of their flush mounting, which provides a number of acoustic advantages (Newell, Holland, 2019). However, it was decided not to use this option since it would limit the objectivity of the comparison. The flush mounting of all of the compared loudspeakers was not feasible.

The loudspeakers of both stereo channels were positioned as close to each other as possible (as can be seen in Fig. 1), but this still introduced experimental biases – from the different spacings between the loudspeaker units in the stereo pairs and the slightly different positions of the loudspeakers in the room (thus, exciting different room modes and reflections). In order to reduce these biases, the spatial arrangements of all of the loudspeakers were changed from session to session so that the participants listened to different arrangements.

The geometric centers of the speaker triplets were positioned at a height of 120 cm (corresponding to the average ear level of the seated participants). These centers formed a stereo triangle with the listener's head. The distance from the listener to the center of the triplet was precisely measured, and the distances of the loudspeakers from the walls were also controlled. The stereo base was 2.5 m (measured from the center of the arrangement). The distance from the walls was no less than 1 m.

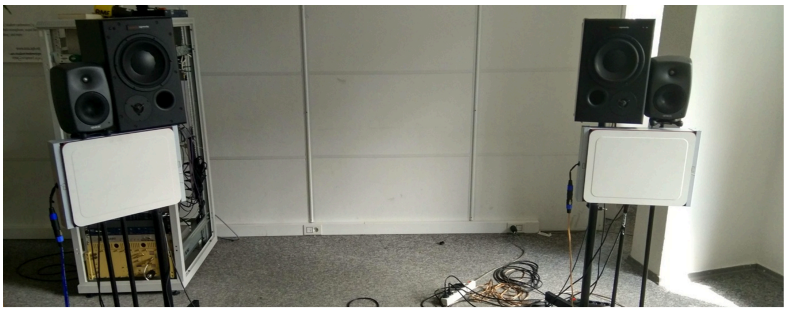


Fig. 1. Exemplary arrangement of loudspeakers. During the experiment, loudspeakers were hidden behind an acoustically transparent curtain.

There were $n = 3! = 6$ possible speaker arrangements, and the arrangements of both channels were made to be mirror images in order to maintain symmetry (cf. Fig. 5). With 12 experimental sessions, each arrangement was repeated twice in random order. A carefully designed protocol for exchanging the positions of the speakers between sessions was implemented (as shown in Fig. 2).

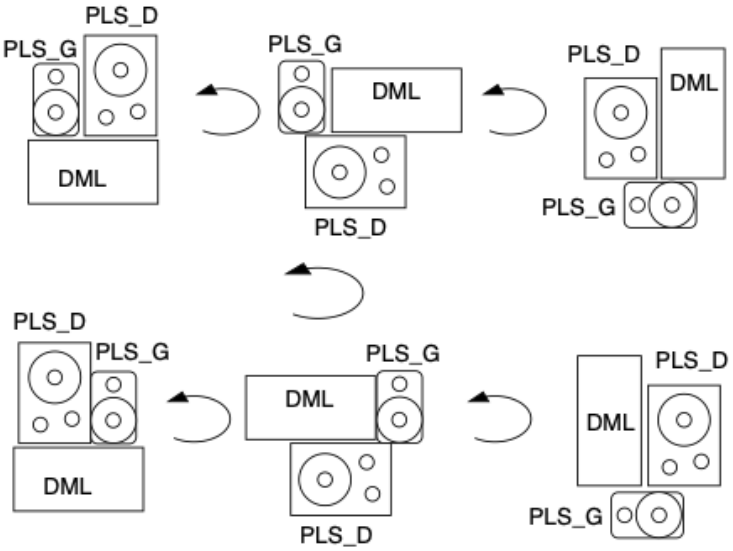


Fig. 2. Changing positions of speakers between experiment sessions.

2.3. Equalization of frequency responses

2.3.1. Equalization of DML *loudspeakers*

Efforts toward loudspeaker equalization may be grouped into two types: equalization of the loudspeaker itself (as measured under anechoic conditions), and loudspeaker + room equalization (based on the loudspeaker's in-room frequency response). In principle, loudspeaker + room equalization provides more of a flat-amplitude response at the point of listening; however, room equalization is sometimes questioned from the point of view of psychoacoustics. This also introduces an experiment-specific factor, thus limiting the external validity of an experiment. Therefore, we chose loudspeaker-only equalization.

The frequency responses of a DML measured from different directions largely differ (Czesak, Kleczkowski, 2023). Therefore, a routine measurement at the axis that is perpendicular to the loudspeaker surface is inappropriate for DMLs; multipoint measurements with averaging should be used instead.

There have been numerous works on loudspeaker equalization (e.g., Karjalainen *et al.*, 1999; Norcross *et al.*, 2004; Bank, 2013), but fewer attempts have been published on equalizing DMLs. Pueo *et al.* (2009) and Heilemann *et al.* (2017) studied the equalization of multiactuator DMLs, and their measurements were limited

to one point (as is typical for piston loudspeakers). Ho and Berkhoff (2015) investigated a new honeycomb structure of a DML panel with multiple actuators and applied velocity feedback controllers for each actuator. Hörchens and de Vries (2011) compared measurement methods for equalizing DMLs and concluded that this could not be based on measurements taken at a single position (or only a few positions) in front of the panel but rather on its average radiation spectrum.

2.3.2. Equalization above 100 Hz

In this work, it was chosen to implement the magnitude equalization of the frequency response. The procedure presented below was limited to the range above 100 Hz, since the efficiency of the DML used drops off rapidly below this frequency and the equalization would require excessive power.

A precise multipoint measurement was performed in order to obtain a reliable average magnitude of the frequency response (Czesak, Kleczkowski, 2023). Each of the DMLs used in the left and right stereo channels was analyzed. The frontal hemisphere radiation was investigated, as the units were supplied in an enclosure that reduced backward radiation. The measurements were performed on a dense grid of 325 points over the hemisphere, with an angular resolution of 10° (both in azimuth and elevation). The grid of measurement points is presented in Fig. 3. There was one measurement point at the axis perpendicular to the loudspeaker surface, and 9×36 points were distributed along nine circles (representing the parallels of the hemisphere).

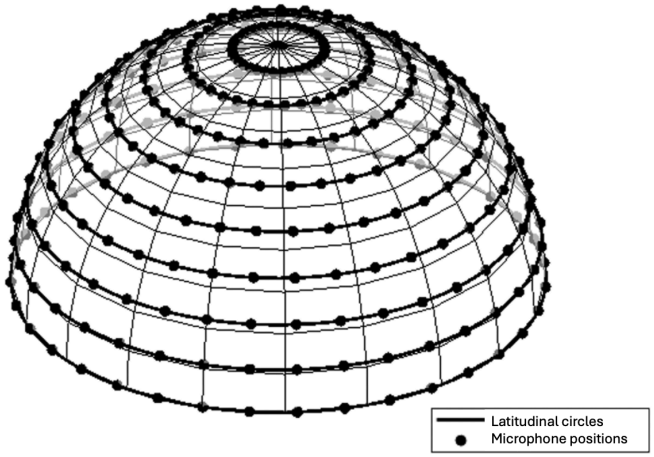


Fig. 3. Grid of measurement points of the DML on the hemisphere (from (Czesak, 2025)).

The measurements were carried out in a 1000 m^3 anechoic chamber at the Department of Mechanics and Vibroacoustics, AGH University of Krakow, using a custom-made automated system for positioning a measurement microphone. To minimize near-field effects, the maximum available measurement radius was used (2.5 m).

Due to constant angular resolution, geometric correction was required: points at higher elevation angles (e.g., 80°) are denser than those at 0° and represent smaller areas. Correction coefficients derived in (Czesak *et al.*, 2022) are presented in Table 1.

Table 1. Relative correction factors for parallel circles in measurement hemisphere.

Elevation angle [°]	0	10	20	30	40	50	60	70	80	90
Correction factor	1.000	0.985	0.940	0.866	0.766	0.643	0.500	0.342	0.174	0.786

Averaging acoustic pressure over the 325 points of the measurement hemisphere was calculated as

$$\overline{p}_h(f) = \sqrt{\sum_{i=0}^{8} k_i \sum_{j=0}^{35} |p_{i,j}(f)|^2 + k_9 |p_{9,0}(f)|^2}, \tag{1}$$

where $\overline{p}_h(f)$ is the average acoustic pressure over the hemisphere, k_i is the correction factors for elevations from Table 1, $p_{i,j}(f)$ is the acoustic pressure at point (i,j), f is the frequency.

At each measurement point, amplitude-frequency responses were obtained using narrow-band noise excitation in 1/20th-octave bands (193 bands, center frequencies: 63 Hz to 16 200 Hz). Custom software developed at the anechoic chamber (Pilch, Kamisiński, 2011) provided sound pressure values in dB for each band, offering constant-Q resolution across the frequency range.

An equalizing filter was implemented using a 220-point FFT filter. The FFT filter, with linear-phase property and numerical stability, allowed accurate equalization and easy generalization. Since offline filtering was used, causality was not a requirement, and the zero-phase property was advantageous for perceptual comparisons. Informal listening revealed no pre-ringing effects.

The equalization procedure was as follows:

- conversion of 1/20-octave frequency scale (63 Hz to 16 200 Hz) to a linear scale of 220 points,
- linear interpolation to obtain $A(f)$ – the linear-frequency amplitude response,
- computation of $A_c(f) = 1/A(f)$ – the equalizing filter,
- offline filtering of each musical excerpt by $A_c(f)$ using a 220-point zero-phase FFT filter,
- conversion back to the time domain by IFFT and storage as 24-bit/44.1 kHz WAV files. It was decided to requantize the floating-point results of filtering not to 16 bit but to 24 bit with appropriate dithering, as D/A conversion during reproduction was of 24 bit resolution.

Figure 4 shows the equalizing filters for both DMLs. The right channel was equalized from 100 Hz, while the left channel from 85 Hz due to the better low-end response. The filters are highly similar, indicating consistency between the DML units. Flat sections below 100 Hz and above 16 kHz were manually limited to avoid overload.

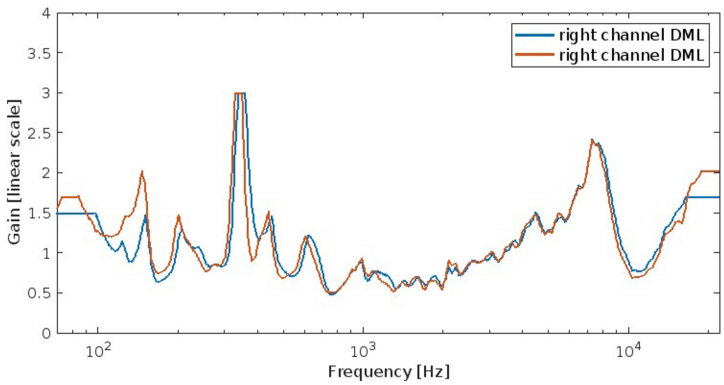


Fig. 4. Frequency responses of equalizing filters for DML loudspeakers.

To preserve objectivity, musical excerpts used for comparing DML with the two PLS systems were also passed through the same 220-point FFT/IFFT pipeline. Above 100 Hz, this procedure was transparent.

2.3.3. Correction at frequencies below 100 Hz

As the frequency responses of the DMLs could not be corrected below 100 Hz, an alternative approach to equalization in this range was adopted. The low-frequency responses were essentially flat down to 40 Hz in the case of PLS_D and to 54 Hz in the case of PLS_G (both at −2 dB). The responses of the PLSs were corrected to match those of the DMLs using high-pass filtering of the test material. This filtering was implemented within the FFT/IFFT procedure described earlier.

Analysis of the spatially averaged frequency responses of the DMLs revealed that their downward slope below 100 Hz (right speaker) and 85 Hz (left speaker) was approximately −18 dB/oct. Therefore, this slope was applied in the high-pass filtering of material reproduced by both PLS systems.

2.4. Listeners

The experiment involved a total of 30 participants, primarily fourth-year students from the Acoustical Engineering program, with an average age of 23 years. The distribution of participants across experimental sessions

was as follows: 7 attended 1 session, 8 attended 2 sessions, 7 attended 3 sessions, 6 attended 4 sessions, and 2 attended 5 sessions. We decided to divide the entire panel of listeners into two groups. The first consisted of those who participated in one session, as well as those who participated in two sessions, but in the latter case only the results from the first session were included – these participants were classified as one-time listeners (1-TL). The second group consisted of those who participated in at least three sessions. For those who took part in more than three sessions, only the first three sessions were included – these were classified as three-times listeners (3-TL). The advantage of this division was that the two groups were mutually exclusive, with the 3-TL group containing, on average, somewhat more experienced listeners. Both groups included an equal number of participants (15 each). Most of the participants had similar moderate levels of experience due to their completion of courses in ear training and sound engineering (although some individuals had additional experience working in the industry or through other relevant experiences). However, they had little experience with listening tests. About half of the members of the 1-TL group were not students of acoustical engineering and had no experience at all. Because participation was voluntary, it may be hypothesized that more experienced listeners were more interested in taking part and thus were more likely to appear in the 3-TL group. None of the participants reported any hearing problems.

2.5. Setup

During the experiments, the loudspeakers were concealed behind an acoustically transparent curtain, rendering them invisible to the listeners. The experimental setup is illustrated in Fig. 5. The experiment was conducted in a room with a floor area of 44 m^2 and a ceiling height of 2.8 m, yielding a volume of 123 m^3. The room was acoustically treated with appropriate panels and materials, resulting in a reverberation time of approximately 0.3 s. Based on its volume and reverberation characteristics, the room conformed to the standards for listening rooms as specified in ITU-R BS.1116-2 (ITU, 2015) and EBU Tech. 3276 (European Broadcating Union [EBU], 1998). However, the background noise level in the room, influenced by a neighboring street, was higher than recommended by these standards – approximately 35 dB SPL(A), comparable to that of an empty office space. Nonetheless, the test signal level was approximately 80 dB SPL(A), ensuring a sufficient signal-to-noise ratio for perceptual testing. While the room did not fully comply with the noise level specification, the elevated test level mitigated any potential perceptual masking effects.

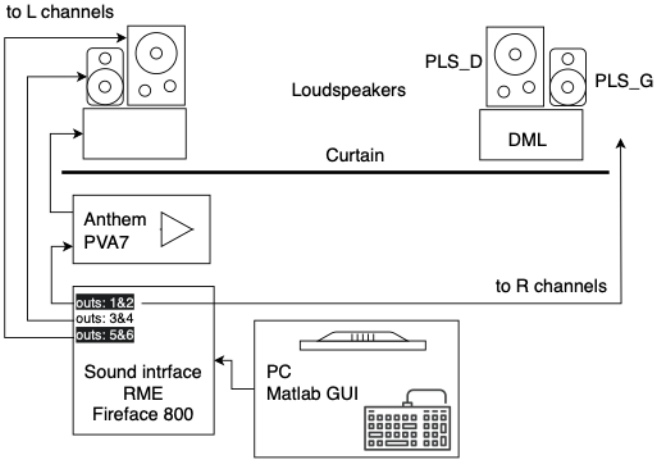


Fig. 5. Experimental setup.

2.6. Evaluation attributes

The evaluation process was facilitated through a graphical user interface (GUI) developed in MATLAB (see Fig. 6). The interface enabled participants to assess two perceptual attributes: 'overall preference' and 'spatial impression.' The test was conducted in Polish, and the respective original terms were: 'preferencja ogólna' and 'wrażenie przestrzenności,' respectively. Both attributes were evaluated in the same session. The position

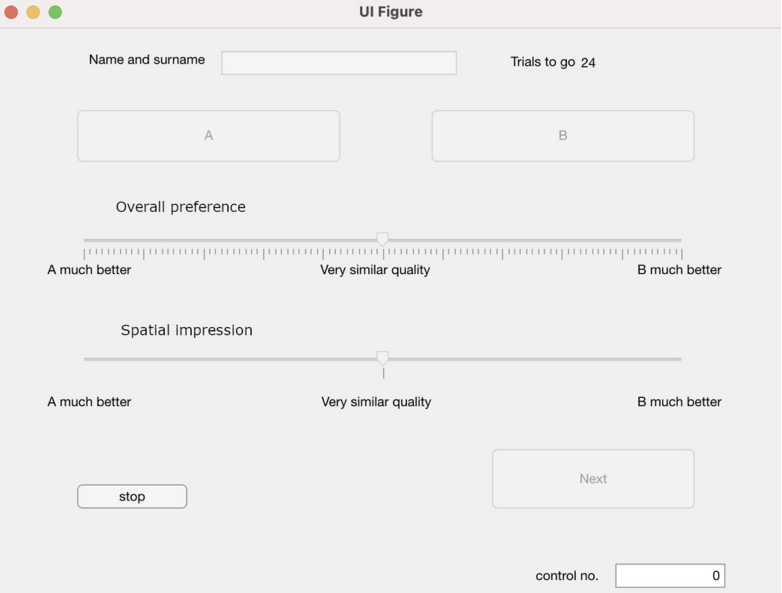


Fig. 6. GUI used in the experiment.

(upper/lower) of each attribute in the graphical interface was not randomized, but users were free to evaluate attributes in the order of their choice. These attributes were rated using a continuous slider scale ranging from 'A – much better' through 'Very similar quality' to 'B – much better.' This design allowed for nuanced judgments and captured subtle perceptual differences between loudspeaker systems.

The experiment followed a double-blind protocol in which neither participants nor administrators knew which speaker system was assigned to version A or B during any given trial. The GUI assigned control numbers to each evaluation, tracked trial progression, and maintained systematic consistency across sessions.

The overall preference attribute referred to the overall auditory experience delivered by the loudspeaker system. The spatial impression attribute addressed the spatial impression perceived by the listener.

ITU (2019) standard offers a general framework for sound quality evaluation, including aspects such as clarity, spatiality, and timbre. It extends the methodologies outlined in (ITU, 2015) which is focused on the subjective assessment of small impairments in audio systems, including multichannel configurations. Additionally, the EBU (1998) recommendation provides detailed guidance on listening conditions and the assessment of sound program material for mono and stereo reproduction, with emphasis on spatial attributes and perceptual clarity.

In our experiment, the focus was narrowed to two key evaluation dimensions: overall preference and spatial impression. This decision was informed by informal pretests, in which listeners struggled to distinguish timbral and clarity differences after equalization of the DMLs. Therefore, such aspects were assumed to be implicitly accounted for in the general quality rating. This also helped reduce the cognitive load for participants with limited critical listening experience.

The most pronounced audible differences stemmed from the contrasting directivity patterns of the loudspeakers. These were considered to be captured in the spatial impression attribute. The incoherent radiation of the DMLs, their longer impulse responses (Anderson, Bocko, 2015), and the use of non-coaxial configurations likely contributed to broader perceptual differences, all of which were assumed to be reflected within the overall preference ratings.

2.7. Method for graded pairwise sound-quality comparison

Two most often used experimental paradigms in research on perception are independent rating (referred to as multiple comparisons in audio evaluation) and pairwise comparison. Multiple comparisons are more prevalent in audio evaluation, as can be inferred from the published literature. According to Pérez-Ortiz *et al.* (2019),

pairwise comparison eliminates observer bias. We chose the latter paradigm, with an extension referred to as graded pairwise comparison (or scaled pairwise comparison) (Koczkodaj, 2016; Pérez-Ortiz *et al.*, 2019). Below we outline the basics of pairwise comparison, while information on how we used the graded version is included at the end of this section.

The raw results of the pairwise comparisons had to be transformed into a set of scalar parameters assigned to each of the three compared stimuli, allowing for direct ranking. This is typically achieved using statistical models, most notably the Thurstone law of comparative judgment (Thurstone, 1927a; 1927b; Tsukida, Gupta, 2011) or the Bradley–Terry model (later extended to the Bradley–Terry–Luce model) (Tsukida, Gupta, 2011).

In this study, the Thurstone model was employed. According to Thurstone, each comparison between two stimuli Ψ_1 and Ψ_2 evokes a discriminal process in the listener, assumed to be normally distributed. Consequently, the perceived difference between two stimuli is also normally distributed, and its mean reflects the most frequent judgments made by the listeners. If Ψ_1 is perceived as stronger or better than Ψ_2, the probability area under the $\Psi_1 - \Psi_2$ distribution is greater than that under the reverse comparison.

Thus, the perceptual magnitude of difference between two stimuli can be described by the following expression (Thurstone, 1927b):

$$\overline{\Psi}_1 - \overline{\Psi}_2 = z_{12} \cdot \sigma_{\Psi_1 - \Psi_2}, \tag{2}$$

where $\overline{\Psi}_1, \overline{\Psi}_2$ is the means of the normally distributed discriminal processes, z_{12} is the z-score corresponding to the observed probability $p_{\Psi_1 > \Psi_2}$, and $\sigma_{\Psi_1 - \Psi_2}$ is the standard deviation of the distribution of $\Psi_1 - \Psi_2$.

Thurstone's original formulation was based on dichotomous judgments (i.e., A > B or B > A), which yielded raw scores in the form of probability values (p), later transformed into z-scores. As the model does not provide values for $\sigma_{\Psi_1 - \Psi_2}$, several assumptions were required. Thurstone proposed five model cases, depending on the assumptions concerning standard deviations. The model used in this study is a hybrid of case I (a single observer with repeated judgments) and case II (multiple observers, single judgment per pair), because the design involved multiple observers each providing repeated judgments.

In graded pairwise comparison listeners express the magnitudes of their preferences toward A or B with a certain numeric scale. With a graded comparison, the distribution of the values of $\Psi_1 - \Psi_2$ is readily available, and the p-values are calculated from the sums of the respective scores (unlike in a dichotomous judgement). We assumed that the distribution of $\delta_{\Psi_1 - \Psi_2}$ calculated this way provided an estimate of $\delta_{\Psi_1 - \Psi_2}$ from the Thurstone dichotomous model. This assumption allowed direct computation of $\delta_{\Psi_1 - \Psi_2}$ without making any assumptions about the underlying distributions. Thus, direct estimation of $\overline{\Psi_1} - \overline{\Psi_2}$ was possible. Formally, the model that was used in this work was a combination of the Thurstone case I (no simplifying assumptions about the distributions of the data, a single observer, and repeated judgements) and case II (no simplifying assumptions, and many observers making single judgements).

It seems that graded pairwise comparison have seldom been used in audio evaluation. Schuck *et al.* (1993) collated practical evaluations of loudspeakers using multiple comparisons and pairwise comparison, as part of a study on the interaction between the loudspeaker and listening room. Pairwise comparison had primarily been used as the dichotomous choice, but the authors also asked subjects to give each loudspeaker a rating. They found that similar results were obtained using the multiple-comparisons paradigm, the graded paired comparison paradigm (analyzed with MANOVA), and dichotomous pairwise comparison analyzed with the Bradley–Terry method. They noticed that multiple comparisons resulted in a wider spread of ratings than pairwise comparison, but that the power to detect differences between loudspeakers was greater in the multiple-comparisons paradigm.

Francombe *et al.* (2017) used a continuous scale rating during pairwise comparisons. The authors admitted that this was more demanding for participants than a forced-choice task. No further details were provided regarding the use of the rating scale in the context of the Thurstone case V. In (Lee, Rumsey, 2004) multiple comparisons were used, but technically the rating involved a pairwise comparison of each stimulus with a reference stimulus. However, multivariate ANOVA – not the Thurstone probabilistic model – was used to quantify the effect under investigation.

All computations and model implementations were carried out in Matlab.

2.8. Design and course of experiment

Each subject participated on different days, i.e., in different sessions. Sessions differed only in the loudspeaker setup (see Fig. 2), so that each subject experienced a different spatial arrangement of loudspeakers in each session. Listeners were free to choose the dates of their participation, but sessions with repeated arrangements were excluded. Subjects were unaware of the actual loudspeaker arrangement on the days they selected. Thus, the arrangement factor was randomized. The experiment employed a randomized pairwise AB test methodology without repetitions, focused on the perceptual evaluation of audio reproduced by different loudspeaker systems. Prior to the test phase, each participant completed a training session comprising several examples, which served to familiarize them with the experimental interface and procedure. Participants were presented with eight distinct audio samples. Each sample was played in two versions, labeled 'A' and 'B,' corresponding to a pairwise comparison between two of the three systems (DML, PLS_D, and PLS_G). All possible system pairings were included: DML/PLS_D, DML/PLS_G, and PLS_D/PLS_G. To avoid bias, each pairing was tested in both assignment orders (e.g., A = DML, B = PLS_D; and A = PLS_D, B = DML). However, each participant only experienced one fixed order per pairing during their session to limit session length. The order of sample presentations was randomized individually for each participant, ensuring unbiased and diverse evaluation sequences. Participants could listen to each sample as many times as needed and switch freely between versions A and B, in accordance with ITU (2015) recommendations. After making a selection, they could not return to that trial, preserving the integrity of each evaluation. The GUI was designed to avoid audible artifacts during switching. Generally, participants operated the system with ease. Post-session remarks often indicated that the differences between versions were subtle but perceptible, validating the sensitivity and clarity of the test environment. Data from each session contributed to the global analysis across all 29 participants. Each audio sample was limited to a duration of 48 s and shaped with fade-in and fade-out envelopes. This design follows findings by Koehl and Paquier (2013), who demonstrated that excerpts longer than five seconds enhanced the discrimination sensitivity between loudspeakers. The musical material represented a variety of genres: two classical music excerpts (symphonic and chamber), choral music, instrumental and vocal jazz, stage music (Latin and guitar), blues rock, and fado. All samples were in CD-quality WAV format (16-bit, 44.1 kHz) and normalized based on their RMS levels after equalization. Differences between samples were reduced to below 1 dB, and all samples were presented at the same calibrated level.

Detailed information regarding the musical excerpts that were used is provided in Table 2.

Table 2. Music excerpts used in the experiment.

Excerpt	Title	Artist	Music genre
1	*Piano Concerto No. 1 in E Minor, Op. 11: I*	Sinfonietta Cracovia – dir. Jerzy Dybał, sol. Szymon Nehring	Classical – symphonic music
2	*Nyne otpushchayeshi*	Church Slavonic School Choir	Classical – choral music
3	*Uma Casa Portuguesa by Artur Fonseca*	Amalia Rodrigues	Fado
4	*Machine Gun*	Jimi Hendrix	Rock/funk fusion
5	*Get It While You Can by J. Ragovoy and C. Taylor*	Janis Joplin	Blues rock and soul
6	*Leaving*	Mateusz Pałka Trio	Contemporary jazz
7	*Candeeiro de Saudade by Roque Ferreira*	Thais Macedo	Samba/MPB
8	*Piano Trio No. 2 in E minor*	AMKP Piano Trio	Classical – chamber music

Most individual sessions lasted between 20 min and 30 min. While the length of samples and sessions slightly exceeded the recommendations from ITU (2019), participants were permitted to pause at any time. The procedure closely followed the paired comparisons method suggested therein.

3. Results

The structure of the data that was obtained in this experiment is presented in Fig. 7. This was organized into two $75 \times 8 \times 3$ matrices (indexed as $i \times j \times k$), with the individual participants' sessions as rows (the total number

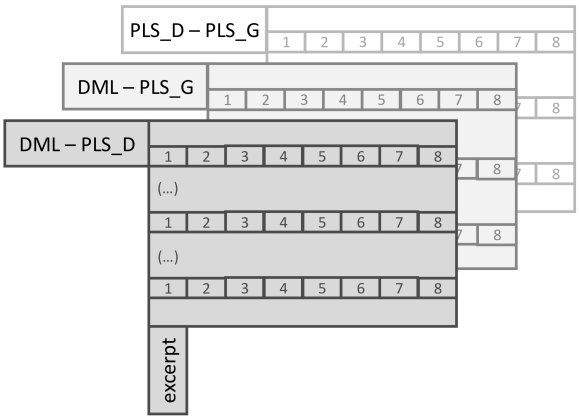


Fig. 7. Structure of data obtained in the experiment, for both matrices (GQ and FS).

of evaluations was 75 – see Subsec. 2.4), eight musical excerpts as columns, and the three pairs of compared systems as layers (3D).

One matrix represented assessments of overall preference (**OP**), the other spatial impression (**SI**). The matrices contained all participants' results from the sessions they participated in. Analyzing data in the 1-TL and 3-TL groups (see Subsec. 2.4) required dividing OP and SI matrices into appropriate submatrices. In this work, experimental variables z_{12} and $\delta_{\Psi_1-\Psi_2}$ (Eq. (2)) were determined individually from the data that was contained in each of the columns of the **OP** and **SI** matrices.

When the pairs of perceptual differences were determined according to Eq. (2), the final set of quality scores for all three systems was calculated with the least squares solution (Tsukida, Gupta, 2011). These results are presented in Fig. 8 to Fig. 11 separately for each of the excerpts. The mean values of the quality scores for each loudspeaker pair are also shown. The values in those figures are perceptual units. According to the Thurstone model, the placement of the zero point of a perceptual scale is completely arbitrary, so the scale is an inter-

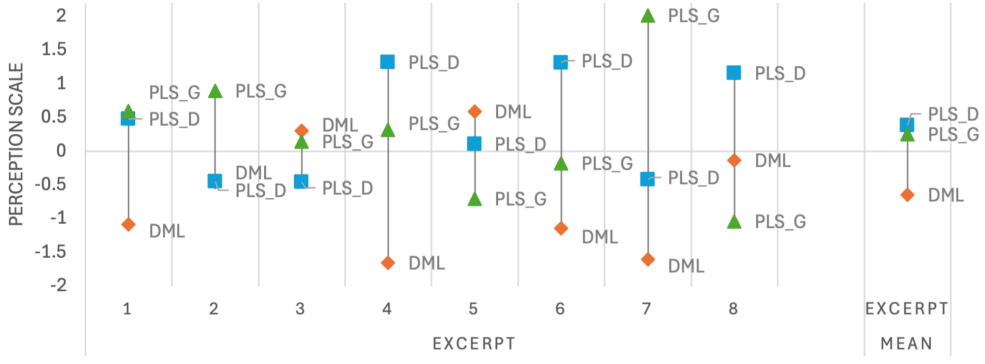


Fig. 8. Quality scores for all three systems: overall preference, 1-TL group.

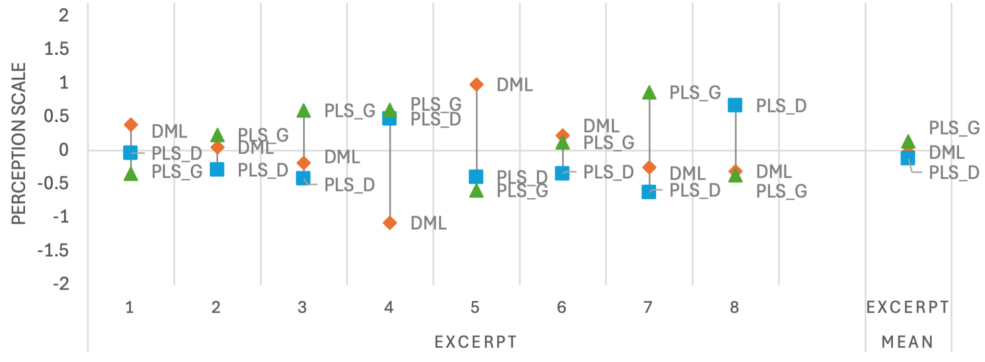


Fig. 9. Quality scores for all three systems: overall preference, 3-TL group.

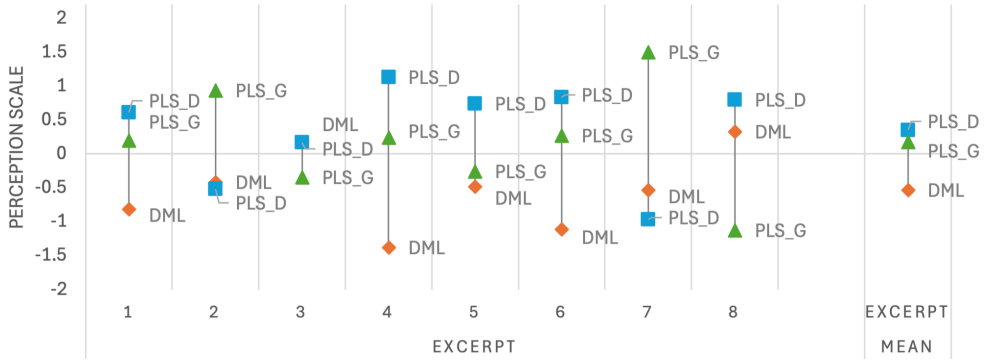


Fig. 10. Quality scores for all three systems: spatial preference, 1-TL group.

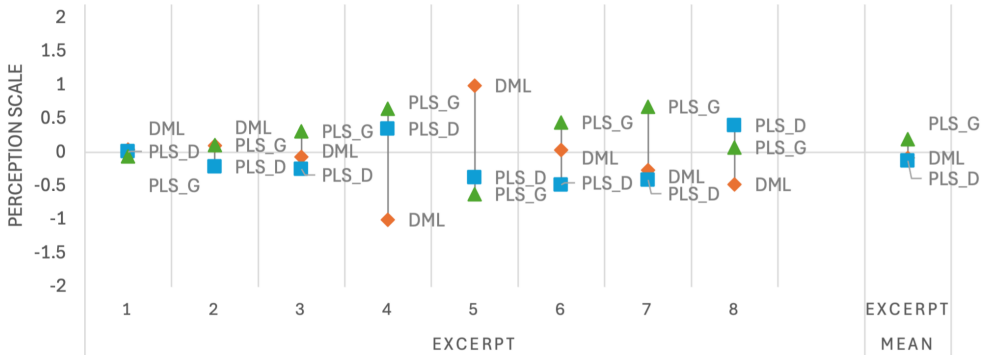


Fig. 11. Quality scores for all three systems: spatial preference, 3-TL group.

vallic one. It is convenient to perform calculations so that the mean of each of the three quality scores is zero for each excerpt. This is the way that the scores are presented.

The value of one (z-score = 1) in the Thurstone model is interpreted as one perceptual unit; therefore, most of the scores should be interpreted as being low; consequently, the perceptual differences that could be found in this experiment were low. The scores in the 1-TL group exhibited a substantially higher spread than those in the 3-TL group. This can be attributed to lower experience of participants in the 1-TL group and to intra-subject variability in assessments, which is reduced in the 3-TL group by averaging three independent evaluations. In the 3-TL group, only in five cases were the perceptual differences $\overline{\Psi_1} - \overline{\Psi_2}$ between loudspeaker pairs close to 1.5. For overall preference, these cases were DML and PLS_G (in excerpts 4 and 5), PLS_D and PLS_G (in excerpt 7); for spatial impression: DML and PLS_G (in excerpts 4 and 5). Excerpts 4, 5 and 7 appeared to be the most revealing in the listening evaluations according to both attributes. It is noteworthy that for both attributes, DML scored substantially lower than the other systems in excerpt 4, while it scored substantially higher than the others in excerpt 5. This demonstrates considerable effect of the excerpt on evaluation.

Another observation is that when scores were averaged over excerpts, for both attributes, the 1-TL group evaluated DML lower than both cone speaker systems (although the difference, expressed in perceptual units, was low), whereas the 3-TL group ranked it between the two cone systems.

With a graded pairwise comparison, it is possible to perform hypothesis testing based on the raw perceptual comparisons using parametric tools. The same procedure was performed on the data from **OP** and **SI** matrices, containing results for the 1-TL and 3-TL groups. The columns of each matrix (see Fig. 7) were assumed to be samples, and μ_{jk} sample averages were found. For each column j, the average of means $\mu_j = (\mu_{j1}+\mu_{j2}+\mu_{j3})/3$ was calculated. Then, individual divergencies from μ_j were calculated according to $\Delta_{j1} = \mu_{j1} - \mu_j$, $\Delta_{j2} = \mu_{j2} - \mu_j$, $\Delta_{j3} = \mu_{j3} - \mu_j$. The Δ_{jk} values were effects of interest; they were used in the standard procedure for evaluating the statistical significance at a $p = 0.05$ level, with the two-tailed t-test ($df = 14$ for the 1-TL group and $df = 44$ for the 3-TL group), and with the Bonferroni correction ($m = 3$). All of the Δ_{jk} values for both matrices turned out to be insignificant; this confirmed that perceptual differences that are presented in Fig. 8 to Fig. 11 were low.

Assessments of the perceptual differences among the loudspeaker systems were dispersed among the individual listeners, with fairly consistent values of standard deviation (which were calculated from the columns of the **OP** and **SI** matrices). The average values of the standard deviation were for **OP** – 1-TL group: 2.06 and 3-TL group: 2.23; for **SI** – 1-TL group: 1.93 and 3-TL group: 2.09.

Although eliminating less-reliable listeners from the results of listening tests is not recommended in general (Bech, Zacharov, 2006), such a possibility was analyzed in this work. At first, intra-listener consistency was evaluated by calculating the pairwise correlations between the rows of the **OP** and **SI** matrices that represented the sessions of the same listener. Correlations were not expected, as the sessions of an individual listener were carried out with different arrangements of loudspeakers (see Subsec. 2.2). All of the correlations were low, so care should be taken when drawing conclusions from them.

In principle, the credibility results of the triple pairwise comparisons can be verified by the transitivity test Birnbaum (2023). Preferences are said to be transitive when if $X > Y$ and $Y > Z$ for all X, Y, and Z, then $X > Z$. However, this is not a reliable measure of the consistency of the preferences, as individual preferences contain random errors. Moreover, sets of X, Y, and Z may be transitive by chance, so listeners who did not hear any differences between the stimuli might pass the test.

When applied to a particular problem, the quality of the Thurstone model can be verified by reversing the procedure. The final values in the perceptual scale $\overline{\Psi_1}$, $\overline{\Psi_2}$ can be used to determine z_{12} from Eq. (2), and the latter value can be converted to estimated value $p_e(\Psi_1 > \Psi_2)$. Finally, the two values of p that were obtained experimentally ($p_{\Psi_1 > \Psi_2}$) and estimated ($p_e(\Psi_1 > \Psi_2)$) may be used to evaluate the quality of the model. The close correspondence between the two values indicate the applicability of the model.

The reverse procedure was performed on the entire **OP** and **SI** matrices, with the 1-TL and 3-TL groups combined, and the pairs of the results were tested for statistical significance by a χ^2 test. Forty-eight pairs were tested; these results are given in Table 3. The analysis produced all insignificant values except for one instance, which indicated that the used model was applicable.

Table 3. χ^2 values from tests of the fit between experimental and predicted preference proportions. The only significant value (excerpt 7, OP, DML vs. PLS_G) is shown in bold.

	Comparison	1	2	3	4	5	6	7	8
Overall preference	DML/PLS_D	0.28	0.06	0.06	0.98	0.10	1.08	3.36	1.07
	DML/PLS_G	0.27	0.06	0.06	1.03	0.09	1.04	**4.21**	0.85
	PLS_D/PLS_G	0.30	0.06	0.06	0.98	0.10	0.94	3.23	0.66
Spatial impression	DML/PLS_D	0.04	0.00	0.02	1.53	0.24	0.12	0.63	0.85
	DML/PLS_G	0.00	0.00	0.02	0.85	0.42	0.01	0.58	0.64
	PLS_D/PLS_G	0.00	0.01	0.05	1.53	0.26	0.21	0.99	0.59

The only significant value (indicating the inapplicability of the model that was used) occurred in just one comparison – in the evaluation of overall preference regarding excerpt 7. The other two comparisons in this group also produced considerably higher χ^2 values than could be found in all of the other groups. Excerpt 7 is a samba piece with numerous percussion instruments, featuring a high amount of high-frequency content. This genre may benefit from the narrow directional characteristics of PLS monitors. It can be noticed that the lowest values of χ^2 (which indicated the very good performance of the model in both of the perceptual attributes) were obtained with excerpts 1, 2, 3, and 5.

4. Discussion and conclusions

Two key findings can be observed in the results. The primary finding is that the perceived differences between loudspeaker systems were small. This outcome is somewhat unexpected, considering the simple construction of DML speakers and the level of technological advancement in professional-grade two-way active monitors. The second finding is that the 3-TL group evaluated the DML higher than the 1-TL group for both attributes, rating it at a level comparable to that of the cone systems. Whether this difference is due to the greater experience of the 3-TL participants or to more consistent results within this group remains unknown.

After the completion of this study, a follow-up investigation designed to verify and extend our findings was conducted by our research team, with a partly different set of authors (Kleczkowski *et al.*, 2025). It was carried out under completely different experimental conditions. Important conclusions can be drawn by referencing our results to those presented in (Kleczkowski *et al.*, 2025). The overall findings of both works are similar: equalized DMLs attain sound quality close to that of high-quality loudspeaker systems based on cone drivers. This similarity was observed despite nearly all possible differences in the applied methods. In (Kleczkowski *et al.*, 2025) the loudspeaker arrangement was fixed, the DMLs (two were examined) were flush-mounted, both the DMLs and cone systems were equalized for their position in the listening room, the experimental paradigm was multiple comparisons with parametric and non-parametric statistical evaluation, six evaluation attributes were used, the musical excerpts differed, five out of six loudspeakers used in both experiments were different, the listening room was quite different, and finally all listeners were distinct.

A specific finding of (Kleczkowski *et al.*, 2025) was that for two out of three spatial attributes (envelopment and stage width) the DMLs significantly outperformed the cone system, whereas the cone systems were significantly favored in the localization attribute. For two timbral attributes and the global attribute pleasantness, the loudspeakers of both types were rated close.

The comparison of this study with (Kleczkowski *et al.*, 2025) leads to an important conclusion: loudspeaker and room correction seems to improve the perceived quality of DMLs compared to loudspeaker-only (anechoic) correction. The overall relative evaluation of DMLs was moderately higher in (Kleczkowski *et al.*, 2025) than in the current study. The type of equalization is likely to affect perceptual evaluation more than any of the following factors: flush mounting, listening room and competing cone loudspeakers. Another conclusion concerns the use of pairwise comparison: the combination of findings from both studies confirms the observation reported in (Schuck *et al.*, 1993) that the power to detect loudspeaker differences is greater for multiple comparisons than for pairwise comparison.

Several works that were cited in the Introduction indicated that a wide directivity of loudspeakers is preferable. This was not confirmed in our experiment, as the listeners did not evaluate the spatial impression of the DMLs higher than its overall preference.

Another factor that should be taken into consideration when interpreting our results is that the conditions eliminated a considerable shortcoming of DMLs, i.e., their insufficient low-end extension (see Subsubsec. 2.3.3). This shortcoming can be overcome by using a subwoofer, as implemented in (Kleczkowski *et al.*, 2025).

The χ^2 test results showed almost all comparisons to be insignificant, affirming the Thurstone model's applicability in this context (except for one instance in the overall preference evaluation).

The main conclusion from the experiment was that the widely applied technology of DML loudspeakers, after a relatively simple improvement, achieves sound quality that is evaluated as very close to that of professional quality two-way active loudspeaker systems based on the traditional cone technology. With its significantly more-versatile mounting options, the DML transducer technology offers a unique potential – particularly in home and commercial applications. This technology allows the loudspeakers to blend seamlessly into the decor of a room, serving as a piece of art or a graphic rather than a freestanding piece of furniture (like traditional speakers). This suggests that DML technology could be particularly advantageous in settings where space and aesthetic integration are critical – especially as part of a multichannel sound system.

Fundings

This work was supported by the Faculty of Mechanical Engineering and Robotics, AGH University of Krakow, Research Subsidy No. 16.16.130.942.

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.

Authors' contributions

Piotr Kleczkowski: conceptualization, methodology, formal analysis, validation, writing – original draft, review and editing, supervision. Paweł Małecki: methodology, data curation, software, resources, investigation, writing – original draft. Dorota Młynarczyk: investigation, resources, visualization, writing – review and editing. All authors reviewed and approved the final manuscript.

Data availability

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

References

1. Allison R. (1995), Imaging and loudspeaker directivity: To beam or not to beam, [in:] *Proceedings of the 99th Audio Engineering Society Convention*.
2. Anderson D.A., Bocko M.F. (2015), A model for the impulse response of distributed mode loudspeakers and multi-actuator panels, [in:] *Proceedings of the 139th Audio Engineering Society Convention*.
3. Angus J.A. (2000), Distributed mode loudspeaker radiation mechanisms, [in:] *Proceedings of the 108th Audio Engineering Society Convention*.
4. Azima H., Harris N.J. (1997), Boundary interaction of diffuse field distributed-mode radiators, [in:] *Proceedings of the 103rd Audio Engineering Society Convention*.
5. Bai M.R., Huang T. (2001), Development of panel loudspeaker system: Design, evaluation and enhancement, *The Journal of the Acoustical Society of America*, **109**(6): 2751–2761, https://doi.org/10.1121/1.1371544.
6. Bank B. (2013), Audio equalization with fixed-pole parallel filters: An efficient alternative to complex smoothing, *Journal of the Audio Engineering Society*, **61**(1/2): 28–37.
7. Bank G., Harris N. (1998), The distributed mode loudspeaker – Theory and practice, [in:] *Proceedings of the UK 13th Conference: Microphones & Loudspeakers*.
8. Bech S., Zacharov N. (2006), *Perceptual Audio Evaluation – Theory, Method and Application*, John Wiley & Sons.
9. Bertland B. (1985), Loudspeaker directionality and the perception of reality, *Journal of the Audio Engineering Society*, **33**(5): 342–350.
10. Birnbaum M.H. (2023), Testing transitivity of preference in individuals, *Decision*, **10**(2): 153–180, https://doi.org/10.1037/dec0000185.
11. Czesak K. (2025), *Measurement analysis of distributed mode laudspeakers for the purpose of improving their sound properties* [in Polish: *Analiza pomiarowa głośników modów rozproszonych w celu udoskonalenia ich własności brzmieniowych*], Unpublished Ph.D. Thesis, AGH University of Krakow, Kraków.
12. Czesak K., Kleczkowski P. (2023), Directivity characteristics of distributed mode loudspeakers. Selected aspects [in Polish: Wybrane aspekty charakterystyk kierunkowości głośników modów rozproszonych], [in:] *Advances in Research on Audio and Video Engineering. Measurement, Processing, Classification, and Quality Assessment of Audio-Video Signals* [in Polish: *Postępy Badań w Inżynierii Dźwięku i Obrazu. Pomiary, Przetwarzanie, Klasyfikacja i Ocena Jakości Sygnałów Audio-Wideo*], Opieliński K.J. [Ed.], pp. 41–56, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław, http://fbc.pionier.net.pl/id/oai:dbc.wroc.pl:125457 (access: 3.11.2025).
13. Czesak K., Kleczkowski P., Król-Nowak A. (2022), Methodology for determining the directivity characteristics of dispersed mode loudspeakers [in Polish: Metodyka wyznaczania charakterystyk kierunkowości głośników modów rozproszonych], [in:] *Postępy w Inżynierii Dźwięku i Psychoakustyce. Advances in Audio Engineering and Psychoacoustics*, Król-Nowak A. [Ed.], Wydawnictwa AGH, Kraków, https://doi.org/10.7494/978-83-67427-06-7_11.
14. European Broadcating Union [EBU] (1998), *Listening conditions for the assessment of sound programme material: Monophonic and two-channel stereophonic*, Technical report EBU Tech. 3276, https://tech.ebu.ch/docs/tech/tech3276.pdf (access: 3.11.2025).

15. Evans W., Dyreby J., Bech S., Zielinski S., Rumsey F. (2009), Effects of loudspeaker directivity on perceived sound quality – A review of existing studies, [in:] *Proceedings of the 126th Audio Engineering Society Convention.*

16. Ferralli M.W., Moulton D. (1995), 360 degree dispersion frequency invariant acoustic transducer system, [in:] *Proceedings of the 99th Audio Engineering Society Convention.*

17. Flanagan S., Harris N. (1999), Loudness: A study of the subjective difference between dml and conventional loudspeakers, [in:] *Proceedings of the 106th Audio Engineering Society Convention.*

18. Flanagan S., Moore B.C.J. (2000), The influence of loudspeaker type on timbre perception, [in:] *Proceedings of the 109th Audio Engineering Society Convention.*

19. Flanagan S., Moore B.C.J. (2001), The effect of loudspeaker type on the identification of vowel-like harmonic complexes, [in:] *Proceedings of the 111th Audio Engineering Society Convention.*

20. Flindell I.H., McKenzie A.R., Negishi H., Jewitt M., Ward P. (1991), Subjective evaluations of preferred loudspeaker directivity, [in:] *Proceedings of the 90th Audio Engineering Society Convention.*

21. Francombe J., Brookes T., Mason R., Woodcock J. (2017), Evaluation of spatial audio reproduction methods (Part 2): Analysis of listener preference, *Journal of the Audio Engineering Society*, **65**(3): 212–225, https://doi.org/10.17743/jaes.2016.0071.

22. Gabrielsson A., Lindström B., Till O. (1991), Loudspeaker frequency response and perceived sound quality, *The Journal of the Acoustical Society of America*, **90**(2): 707–719, https://doi.org/10.1121/1.401941.

23. Gontcharov V., Hill N. (2000), Diffusivity properties of distributed mode loudspeakers, [in:] *Proceedings of the 108th Audio Engineering Society Convention.*

24. Harris N., Flanagan S., Hawksford M.O.J. (1998), Stereophonic localization in rooms, comparing conventional and distributed-mode loudspeakers, [in:] *Proceedings of the 105th Audio Engineering Society Convention.*

25. Harris N., Gontcharov V., Hawksford M.O.J. (2000), Measurement and simulation results comparing the acoustics of various direct radiators in the presence of a dominant specular reflection, [in:] *Proceedings of the 109th Audio Engineering Society Convention.*

26. Harris N.J., Hawksword M.O.J. (2000), Introduction to distributed mode loud speaker (dml) with first-order behavioural modeling, *IEEE Proceedings – Circuits, Devices and Systems*, **147**(3): 153–157.

27. Heilemann M., Anderson D., Roessner S., Bocko M.F. (2018), Quantifying listener preference of flat-panel loudspeakers, [in:] *Proceedings of the 145rd Audio Engineering Society Convention.*

28. Heilemann M.C., Anderson D., Bocko M.F. (2017), Equalization of localized sources on flat-panel audio displays, [in:] *Proceedings of the 143rd Audio Engineering Society Convention.*

29. Heilemann M.C., Anderson D.A., Roessner S., Bocko M.F. (2021), The evolution and design of flat-panel loudspeakers for audio reproduction, *Journal of the Audio Engineering Society*, **69**(1–2): 27–39, https://doi.org/10.17743/JAES.2020.0057.

30. Ho J.-H., Berkhoff A.P. (2015), Flat acoustic sources with frequency response correction based on feedback and feedforward distributed control, *The Journal of the Acoustical Society of America*, **137**(4): 1948–1957, https://doi.org/10.1121/1.4914997.

31. Hörchens L., de Vries D. (2011), Comparison of measurement methods for the equalization of loudspeaker panels based on bending wave radiation, [in:] *Proceedings of the 130th Audio Engineering Society Convention.*

32. International Telecommunication Union [ITU] (2015), *Methods for the subjective assessment of small impairments in audio systems*, Recommendation BS.1116-3, Technical report.

33. International Telecommunication Union (2019), *General methods for the subjective assessment of sound quality*, Recommendation BS.1284-2, Technical report.

34. Jeon O., Ryu H., Kim H.-G., Wang S. (2020), Vibration localization prediction and optimal exciter placement for improving the sound field optimization performance of multi-channel distributed mode loudspeakers, *Journal of Sound and Vibration*, **481**: 115424, https://doi.org/10.1016/j.jsv.2020.115424

35. Jung J., Jensen J.S., Jeong C.-H., Jeon O., Wang S. (2021), Optimizing a distribution of resonators on a thin plate for the desired sound radiation, *Journal of Sound and Vibration*, **496**: 115926, https://doi.org/10.1016/j.jsv.2020.115926.

36. Karjalainen M., Piirilä E., Jäärvinen A., Huopaniemi J. (1999), Comparison of loudspeaker equalization methods based on DSP techniques, *Journal of the Audio Engineering Society*, **47**(1/2): 14–31.
37. Kleczkowski P., Makuch T., Król-Nowak A., Czesak K. (2025), Equalized distributed mode loudspeakers rival conventional ones in listening tests, *The Journal of the Acoustical Society of America*, **157**(4): 2528–2541, https://doi.org/10.1121/10.0036373.
38. Koczkodaj W.W. (2016), Pairwise comparisons rating scale paradox, [in:] *Transactions on Computational Collective Intelligence XXII. Lecture Notes in Computer Science*, Nguyen N.T., Kowalczyk R. [Eds.], Vol. 9655, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-49619-0˙1.
39. Koehl V., Paquier M. (2013), A comparative study on different assessment procedures applied to loudspeaker sound quality, *Applied Acoustics*, **74**(12): 1448–1457, https://doi.org/10.1016/j.apacoust.2013.06.008.
40. Lee H.K., Rumsey F. (2004), Elicitation and grading of subjective attributes of two-channel phantom images, [in:] *116th AES Convention.*
41. Linkwitz S. (2007), Room reflections misunderstood?, [in:] *Proceedings of the 123rd Audio Engineering Society Convention.*
42. Lu G., Shen Y. (2009), Model optimization of orthotropic distributed-mode loud speaker using attached masses, *The Journal of the Acoustical Society of America*, **126**(5): 2294–2300, https://doi.org/10.1121/1.3212943.
43. Lu G., Shen Y., Liu Z. (2012), Optimization of orthotropic distributed-mode loud speaker using attached masses and multi-exciters, *The Journal of the Acoustical Society of America*, **131**(2): EL93–EL98, https://doi.org/10.1121/1.3672642.
44. Moulton D. (1986), The significance of early high-frequency reflections from loudspeakers in listening rooms, [in:] *Proceedings of the 81st Audio Engineering Society Convention.*
45. Newell P., Holland K. (2019), *Loudspeakers for Music Recording and Reproduction*, Routledge.
46. Norcross S.G., Soulodre G.A., Lavoie M.C. (2004), Subjective investigations of inverse filtering, *Journal of the Audio Engineering Society*, **52**(10): 1020–1033.
47. Olive S.E. (2004a), A multiple regression model for predicting loudspeaker preference using objective measurements: Part I – Listening test results, [in:] *Proceedings of the 116th Audio Engineering Society Convention.*
48. Olive S.E. (2004b), A multiple regression model for predicting loudspeaker preference using objective measurements: Part II – Development of the model, [in:] *Proceedings of the 117th Audio Engineering Society Convention.*
49. Pérez-Ortiz M., Mikhailiuk A., Zerman E., Hulusic V., Valenzise G., Mantiuk R.K. (2019), From pairwise comparisons and rating to a unified quality scale, *IEEE Transactions on Image Processing*, **29**: 1139–1151, https://doi.org/10.1109/TIP.2019.2936103.
50. Pilch A., Kamisiński T. (2011), The effect of geometrical and material modification of sound diffusers on their acoustic parameters, *Archives of Acoustics*, **36**(4): 955–966, https://doi.org/10.2478/v10168-011-0065-1.
51. Pueo B., López J.J., Ramos G., Escolano J. (2009), Efficient equalization of multiexciter distributed mode loudspeakers, *Applied Acoustics*, **70**(5): 737–746, https://doi.org/10.1016/j.apacoust.2008.09.005.
52. Roessner S., Heilemann M.C., Bocko M.F. (2019), Evaluating listener preference of flat-panel loudspeakers, [in:] *Proceedings of the 147th Audio Engineering Society Convention.*
53. Schuck P.L. *et al.* (1993), Perception of perceived sound in rooms: some results of the Athena project, [in:] *Audio Engineering Society Conference: 12th International Conference: The Perception of Reproduced Sound.*
54. Thurstone L.L. (1927a), A law of comparative judgement, *Psychological Review*, **34**(4): 273–286, https://doi.org/10.1037/h0070288.
55. Thurstone L.L. (1927b), The method of paired comparisons for social values, *Journal of Abnormal and Social Psychology*, **21**(4): 384–400, https://doi.org/10.1037/h0065439.
56. Toole F.E. (1986a), Loudspeaker measurements and their relationship to listener preferences: Part 1, *Journal of the Audio Engineering Society*, **34**(4): 227–235.
57. Toole F.E. (1986b), Loudspeaker measurements and their relationship to listener preferences: Part 2, *Journal of the Audio Engineering Society*, **34**(5): 323–348.

58. Toole F.E. (2018), *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms*, 3rd ed., Routledge.

59. Tsukida K., Gupta M.R. (2011), *How to analyze paired comparison data*, Technical Report No. UWEETR-2011-0004, University of Washington.

60. Wendt F., Höldrich R. (2021), Precedence effect for specular and diffuse reflections, *Acta Acustica*, **5**: 1–9, https://doi.org/10.1051/aacus/2020027.

61. Yu Z., Zhu Q., Wu M., Yang J. (2023), Exploring the limits of virtual source localization with amplitude panning on a flat panel with actuator array: Implications for future research, *The Journal of the Acoustical Society of America*, **154**(3): 1362–1371, https://doi.org/10.1121/10.0020827.

62. Zacharov N. (1998), Subjective appraisal of loudspeaker directivity for multichannel reproduction, *Journal of the Audio Engineering Society*, **46**(4): 288–303.

63. Zenker B., Schurmann R., Merchel S., Altinsoy M.E. (2020), Improved sound radiation of flat panel loudspeakers using the local air spring effect, *Applied Sciences*, **10**(24): 1–18, https://doi.org/10.3390/app10248926.

Archives of Acoustics Vol. 51, No. 1, pp. 75–91 (2026), https://doi.org/10.24423/archacoust.2026.4173

Research Paper

Influence of Facial, Head, and Neck Dimensions on Vocal Acoustic Parameters in Polish Speakers

Łukasz Pawelec, Kamila Słowik, Anna Lipowicz*

Department of Anthropology, Wroclaw University of Environmental and Life Sciences
Wroclaw, Poland

*Corresponding Author: anna.lipowicz@upwr.edu.pl

Received January 7, 2025; revised October 4, 2025; accepted January 2, 2026; available online January 29, 2026; version of record March 17, 2026; published issue March 27, 2026.

The relationships between human voice parameters and body dimensions have been previously described, but the connections between voice and face geometry remain poorly researched. This study aims to determine the relationships between face dimensions and acoustic parameters in both sexes and examines 111 adult participants (30 males). Each participant undergoes voice recording, which includes five sustained vowels, along with anthropometric measurements of the neck, head, and face regions. Comparisons between voice parameters and the head, face, and neck regions are conducted employing Pearson's correlation coefficients (r) and a multiple linear regression model. The results reveal significant relationships between head, neck, face dimensions and acoustic parameters in both sexes. Males with higher noses, greater head circumferences, and wider faces tend to have lower formants and more stable voices. Females with larger head circumferences had lower formant values, and those with greater neck circumferences tend to have more stable voices. Also, females with increased nose height have a lower fourth formant (F4). Moreover, females with wider faces, noses, and jaws tend to have less rough voices (lower jitter) and longer maximum phonation time (MPT). These findings may be useful for scientists and law enforcement authorities in creating algorithms that build face models based on voice signals.

Keywords: biometry, formants, fundamental frequency, pitch, personal identification.

1. Introduction

There are known associations between vocal acoustics and body dimensions (Bruckert *et al.*, 2006; Evans *et al.*, 2006; González, 2004; 2007; Graddol, Swann, 1983; Pawelec *et al.*, 2022; Pisanski *et al.*, 2014; 2016; Rendall *et al.*, 2005), as well as body composition (Hamdan *et al.*, 2012; 2013). Such relationships depend on the correlation between vocal tract length (VTL), shape, and body size (Fitch, Giedd, 1999), and on the relationship between vocal tract morphology and laryngeal and vocal fold size, as well as acoustic voice parameters such as fundamental or formant frequencies (Fitch, 1997; Titze, 2011). Linear physical characteristics such as body height and weight, and the circumference of the shoulders, chest, waist, and hips, as well as proportions, e.g., the waist-to-hip ratio (WHR), are crucial for describing the appearance of an individual, but they are not as significant as the face in individual identification (Young, Bruce, 2011). Moreover, when given two stimuli for personal identification (face and body), people rely more often on the face (Burton *et al.*, 1999; O'Toole *et al.*, 2010; Robbins, Coltheart, 2012). Additionally, judges rely more on facial features, especially nose and face shape, than on body build characteristics (Rice *et al.*, 2013).

The number of published papers examining the connections between facial morphology and voice features (Bommarito *et al.*, 2019; Macari *et al.*, 2015; 2017; Reinheimer *et al.*, 2021) is limited, as are the facial mea-

surements presented in them. Based on results of those studies weak (0.2–0.3) to moderate (0.4–0.6) correlations between facial dimensions, such as jaw width (go-go), face width (zy-zy), maxilla width (j-j), or mandibular length (co-gn/co-me), were reported. Evidence that facial structure significantly affects voice parameters also comes from studies comparing facial measurements of voice professionals and control subjects (Brattström *et al.*, 1991; Wyganowska-Świątkowska *et al.*, 2013). Professional singers tend to have larger maxilla and mandible dimensions or greater lower face height. Furthermore, evidence indicates that faces can be correctly linked to voice with a probability greater than random using static (Kogelschatz, Barenholtz, 2013) or dynamic facial images (Kamachi *et al.*, 2003).

Some studies examined relationships between head and neck circumferences and vocal characteristics, but their findings are ambiguous – some of them showed a lack of such relationships (Evans *et al.*, 2006; Rendall *et al.*, 2005), whereas others revealed a modest inverse association between neck circumference and voice pitch ($r \sim -0.3$; Pawelec *et al.*, 2022) or stronger associations between the first three formants (F1-F3) and head circumference ($r \sim -0.6$ to 0.7; Reinheimer *et al.*, 2021). Bommarito *et al.* (2019) found that Martin's facial index (face height (n-gn)/face width (zy-zy)) correlated inversely with the second (F2) and the third formant (F3) in males ($r \sim -0.22$ and $r \sim -0.25$, respectively) and inversely with the first formant (F1) and positively with the third formant (F3) in females ($r \sim -0.31$ and $r \sim 0.27$, respectively). Macari *et al.* (2017) observed an inverse association between mandibular width and vocal pitch (F0) or habitual pitch (HP) in female participants ($r \sim -0.35$ and $r \sim -0.39$, respectively) and inverse correlations between F0 or HP and maxillary width in male participants ($r \sim -0.57$ and $r \sim -0.66$, respectively). This study also revealed negative correlations between face width and HP in males and females ($r \sim -0.54$ and $r \sim -0.35$, respectively). The reason why such relationships between head/neck anatomical structures and voice parameters may be observed is that facial morphology is associated with the height of vocal tract cavities (i.e., oral, nasal) and pharyngeal airways (Kikuchi, 2008). In addition, vocal tract structures are related to voice parameters (Fitch, 1997); therefore, one can find collinearity between head/neck dimensions and voice quality. Studies also demonstrate that machine learning methods can estimate face geometry from voice signals (Oh *et al.*, 2019; Wu *et al.*, 2022).

Accounting for the aforementioned scientific reports on the strength and directions of relationships between face and voice, the present study aims to investigate the interrelationship between head, face, and neck dimensions and various acoustic parameters. The approach used two voice stimuli: five sustained vowels and a short sentence without emotional overtones, rather than only sustained vowel phonation as used in similar studies (Bommarito *et al.*, 2019; Macari *et al.*, 2015; 2017; Reinheimer *et al.*, 2021). Moreover, anthropometric measurements were used to determine the facial dimensions of live persons using calipers, and the measurements did not rely on lateral radiographs (Macari *et al.*, 2015; 2017; Reinheimer *et al.*, 2021), other indicators (Bommarito *et al.*, 2019) or photogrammetric technology (Lucas *et al.*, 2023). According to Krauss *et al.* (2002), to identify criminals making anonymous threatening or blackmailing calls, law enforcement agencies consult voice and speech analysis experts to identify the characteristics of the speaker. Previous research has attempted to estimate facial features based solely on voice signal (Li *et al.*, 2023; Ning *et al.*, 2021; Wen *et al.*, 2021; Zheng *et al.*, 2021), but these studies used some algorithms to match voice and face without identifying the strongest determinants of acoustic parameters, and their findings are inconsistent. Therefore, the current study focuses on identifying head and facial features most strongly associated with voice parameters and determining the direction of these relationships. The relationships revealed could help future researches develop an algorithm to identify the dimensions of a speaker's face solely from a recording of their voice.

2. Material and methods

2.1. Participants

The study participants included 135 participants (40 males) from Wroclaw, Lower Silesia, Poland. The study group consisted of students from the Wroclaw University of Environmental and Life Sciences, Faculty of Biology and Animal Science, and adult inhabitants of Wroclaw invited to the research. All volunteers were examined at the

same time of day (9 AM to 12 AM) and under the same conditions (the same silent room, angle, and distance from the recorder, with the same sound recording equipment used). All participants filled in a preliminary questionnaire containing inclusion/exclusion criteria, basic questions (sex, date of birth), and questions on all possible factors that could affect their acoustic parameters, including history of trauma and surgery of the head and neck regions, speech defects, hearing deficits, and occlusion defects; any illness during the examination, use of cigarettes or e-cigarettes, significant alcohol consumption on the day prior, use of hormonal drugs such as anabolic steroids, use of growth hormones or hormonal contraception, and history of voice work as a teacher, sales representative, professional singer, or amateur singer. Females also answered questions about their current menstrual cycle, pregnancy, and menopause. No participants declared a history of head and neck trauma or surgery, hearing or speech defects, or voice work. However, excessive smoking excluded eight males and six females, while illness on the examination day excluded two males and one female. Due to the influence of sex hormones on receptors of the vocal fold epithelium (Newman *et al.*, 2000; Voelter *et al.*, 2008) and the consequent impact on human voice parameters (Abitbol *et al.*, 1999; Dabbs, Mallinger, 1999; Evans *et al.*, 2008; Raj *et al.*, 2010), the use of hormonal contraceptives (n = 4) and being in the fertile phase of the menstrual cycle (n = 3) excluded seven females from further examination. This left a total of 30 males and 81 females for further study.

Research was conducted in accordance with the Declaration of Helsinki. The personal information of all participants was anonymized by assigning each participant an individual anonymous code. This study was approved by the Bioethics Committee at the Wroclaw Medical University (consent number: KB – 25/2021).

2.2. Voice recording and analysis procedure

The voice recording of each participant was conducted under the same standardized conditions, and the recorded material consisted of five vowels (/ɑ:/, /ɛ:/, /i:/, /ɔ:/, /u:/) sustained for three seconds, with a one-second break between them. This method is one of the most commonly used in such studies (e.g., Pisanski *et al.*, 2014), although other variants are also used, for example /bVt/ context words (bat, bet, beet, bot, boot) or whole sentences. Speakers were asked to announce the vowels using a comfortable pitch and loudness. The equipment used included a Shure SM58 SE dynamic cardioid microphone (bandwidth 50 Hz to 15 kHz) connected to an IMG Stageline MPA-202 amplifier and the sound card from a Dell Latitude E6400 computer. Each participant was recorded while positioned in front of a microphone mounted on a height-adjustable tripod, with the distance between the tip of the mouth and the recording device set at 15 cm and an angle of 0°. Standardization of the recording conditions was achieved using a Mozos Mshield acoustic cabin. The sampling frequency was the same for all recordings and was set to 44.1 kHz (16-bit resolution), and all tracks were saved as uncompressed mono format (.wav). The Benetech GM1351 (Benetech Poland) sound level meter indicated that the background noise level of the recording room used was ~38 dB.

All recorded soundtracks were subsequently analyzed using Praat software version 3.9.2. (Boersma, Weenink, 2019). In the first instance, each vowel was analyzed separately. The middle (most stable) part of each vowel, of equal length (0.2 s), was extracted for analysis, and the 'voice report' function was used to compute basic acoustic parameters, including fundamental frequency F0 [Hz], defined as the lowest and strongest harmonic produced by vocal fold vibration and perceived as vocal pitch (mean F0, median F0, standard deviation (SD)-F0, min F0, and max F0), jitter [%], defined as the degree of variation in sound wave frequency from period to period (LOCAL, RAP, PPQ5, DDP), shimmer [%], defined as the degree of the amplitude variation of the acoustic wave from period to period (LOCAL, APQ3, APQ5, APG11, DDA), and mean harmonics-to-noise ratio (HNR) [dB], an indicator of the relationship between harmonic and noise components in the voice (Teixeira *et al.*, 2013; Titze, 1994). Other measurements included formant frequencies (formants: F1–F4 [Hz]), which are formed by filtering F0 in the supralaryngeal vocal tract (Fant, 1960), and their derivatives, including formant number F_n [Hz], formant position P_f [Z], formant spacing ΔF [Hz], and formant dispersion D_f [Hz]. Apparent VTL [cm] estimated from formants values, and voice intensity (loudness [dB]) was also calculated (see the appendix of (Pisanski *et al.*, 2014)). Final values of acoustic parameters were calculated as the mean across all five vowels. Additionally, mean MPT [s] was measured as the maximum duration of sustaining the vowel /ɑ:/ across three

trials, with a five-second break between trials. The pitch floor was set to 75 Hz for males and 100 Hz for females, and the pitch ceiling was set to 300 Hz and 500 Hz for male and female speakers, respectively. The formant ceiling values were set to 5000 Hz for males and 5500 Hz for females. All acoustics parameters were computed using Praat default algorithms.

2.3. Anthropometric measurements

All measurements followed standard anthropometric procedures (Martin, 1914), with each participant subjected to two series of anthropometric measurements of the body, head, and neck. The concordance of measurements derived from two series was greater than 88 % (r = 0.94, $p < 0.001$). Body measurements included body height [cm] measured with an anthropometer to an accuracy of 0.1 cm, and body weight [kg], recorded using an electronic InBody 270 body composition analyzer device (InBody Poland) to the nearest 0.1 kg. Body mass index (BMI [kg/m^2]) was then calculated based on body height and weight values. Head and neck measurements included circumference [cm] and width-length face measurements [mm]. Neck and head circumferences measurements used an anthropometric measuring tape ranging from 0 cm to 150 cm and a precision of 0.5 cm. Neck circumference was recorded at the laryngeal prominence ('Adam's apple'), and head circumference was measured using two points, glabella (g), the most forward point on the lower part of the forehead between the superciliary arches, and opisthocranion (op), the most posterior and inferior point on the occipital bone. Facial measurements were taken using a sliding caliper and a spreading caliper with a precision of 0.1 mm and included face height nasion-gnathion (n-gn), nose height nasion-subnasale (n-sn), nose width alare-alare (al-al), upper lip height subnasale-stomion (sn-sto), lower lip height stomion-supramentale (sto-sm), total lip vermillion height labiale superius-labiale inferius (ls-li), labial fissure length cheilion-cheilion (ch-ch), face width bizygomatic diameter zygion-zygion (zy-zy), and jaw width gonion-gonion (go-go) (Fig. 1). Based on measurements analyzed by Bommarito *et al.* (2019), Macari *et al.* (2015; 2017), and Reinheimer *et al.* (2021), as well as knowledge about relationship between head/face/neck dimensions and vocal tract parameters (Fitch, Giedd, 1999), the above-mentioned measurements were selected for analyses. The rationale for choosing these head/neck measurements was: a) their strong association with vocal tract dimensions (i.e., oral cavity), and b) their use by other authors addressing similar research topics.

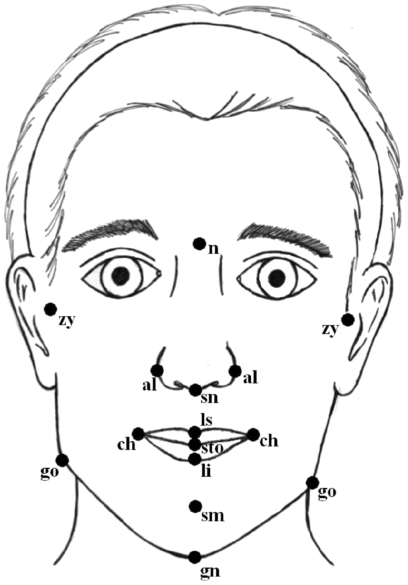


Fig. 1. Facial anthropometric landmarks used for measurements (taken from (Graja, Król, 2022) and modified by the authors).

The measurement values from the two series were averaged, and the means were used for all further analyses. For each of above-mentioned head/neck measurements, the intra-evaluator technical error of measurement (TEM)

was calculated with reference to the method presented by Perini *et al.* (2005). Both absolute and relative [%] errors were calculated based on the following equations:

$$\text{Absolute TEM} = \sqrt{\frac{\sum d_i^2}{2n}}, \tag{1}$$

where Σd_i^2 is the sum of squared differences between the first and second measurements, n is the number of measured participants, and i denotes the number of measurements;

$$\text{Relative TEM} = \frac{\text{Absolute TEM}}{\text{VAV}} * 100\,\%, \tag{2}$$

where VAV is the variable average value (the arithmetic mean calculated based on average values obtained from the two measurements).

2.4. Statistical methods

All analyses were performed using Statistica 13.5 software (1984–2017, TIBCO Software Inc, Palo Alto, California, USA). The Shapiro–Wilk W-test was applied to check the normality of distribution of body/facial features and acoustics parameters in male and female groups. To compare males and females, if the distribution of each variable was normal (Gaussian) in both groups, an independent sample t-test was applied, and the measures of central tendency and variability were reported as mean ±SD. When the distribution was not Gaussian in at least one studied group, the Mann–Whitney U-test was applied, and median and lower/upper quartiles (Q1-Q3) were reported. In both cases, the range (min-max) was also shown. The relationship between neck and head measurements and voice parameters were assessed using Pearson's partial linear correlation, due to the continuous character of the variables, with age and body height included as confounders because of their known associations with voice parameters and head/face morphology (i.e., Fitch, Giedd, 1999; Pisanski *et al.*, 2014; 2016; Jandová, Urbanová, 2016; Rojas *et al.*, 2020; Pawelec *et al.*, 2022; Cazacu *et al.*, 2025). Furthermore, multiple linear regression models verified which head, face, and neck measurements had the greatest impact on voice parameters in males and females, analyzed separately. Several models were created for both sexes, with a single voice parameter as the dependent variable in each model (one model used one voice parameter as a dependent variable). The acoustic parameters included mean F0 (pitch), jitter parameters, shimmer parameters, HNR, intensity, P_f, ΔF, D_f, apparent VTL, and MPT. For males, the explanatory variables (predictors) included head circumference and nose length (n-sn), while for females, three predictors were used: lip vermillion height (ls-li), mandible breadth (go-go), face height (n-gn). The independent variables used in the regression models were selected based on the highest correlation coefficients (r) between voice parameters and anthropometric features for each sex. Collinearity among predictors was tested by correlation coefficients, and redundant variables were excluded from the models. The highest allowed collinearity between predictors was $r \leq 0.3$. In each model, age and body height were included as a confounders. For each regression model, the adjusted coefficient of determination (R^2) was also computed. Only models with significant predictors are reported. Results at a significance level of $p < 0.05$ were considered statistically significant.

3. Results

3.1. Descriptive statistics

The mean participant age was 29.4 ±12.7 years (range from 18 to 65 years), with males being on average five years older than females. Moreover, males had greater values of body height, body mass, neck circumference, head circumference, face width, and most head measurements, except for total lip vermillion height (ls-li; Table 1). The intra-evaluator TEM-absolute [cm or mm] and relative [%] for head and face dimensions were as follows: head circumference (0.3 cm; 0.5 %), neck circumference (0.6 cm; 1.9 %), n-gn (2.1 mm; 1.8 %), n-sn (1.8 mm; 3.6 %), al-al (1.0 mm; 3.2 %), sn-sto (0.9 mm; 4.0 %), sto-sm (1.4 mm; 8.8 %), ls-li (1.0 mm; 6.3 %), ch-ch (0.9 mm; 1.9 %),

zy-zy (2.0 mm; 1.5 %), go-go (0.7 mm; 0.7 %). According to Perini *et al.* (2005), the obtained relative TEM values are below 10 % and are therefore considered acceptable.

Table 1. Physical characteristics of study participants.

Feature	Males (n = 30)*	Range	Females (n = 81)*	Range	p-value**
General data					
Age [years]	27.0 [22.0; 36.0]	20–65	22.0 [22.0; 25.5]	18–64	**0.046**
Body height [cm]	180.5 ±6.5	168.1–197	165.6 ±6.5	146–187	**<0.001**
Body mass [kg]	89.2 [68.9; 107.9]	51.5–164.7	63.6 [55.5; 75.4]	47.2–145.2	**<0.001**
BMI [kg/m^2]	26.8 [22.2; 32.4]	17–46.4	23.4 [20.5; 27.2]	17.3–51.4	**0.037**
Head/neck measurements					
Circumferences [cm]					
Head	57.5 ±1.5	54–60	55.2 ±1.5	52–58.8	**<0.001**
Neck	38.8 [36.1; 41.5]	32–48	32.0 [31.0; 34.0]	28–41	**<0.001**
Face measurements [mm]					
n-gn	114.0 [110; 122]	65–131	108.5 [104.0; 112.5]	60–124	**<0.001**
n-sn	52.0 [54.0; 55.0]	44–65	49.0 [47.0; 52.5]	38–59	**0.005**
al-al	35.5 [33.0; 38.0]	30–42	32.0 [30.5; 34.0]	27–48	**<0.001**
sn-sto	19.5 [18.0; 22.0]	12–27	20.0 [18.0; 21.0]	15–25	0.443
sto-sm	16.0 [15.0; 19.0]	11–47	16.0 [14.0; 17.0]	11–35	0.147
ls-li	13.3 ±4.1	5–23	15.1 ±2.9	8–21	0.075
ch-ch	52.0 [49.0; 56.0]	42–62	50.0 [48.0; 52.0]	31–62	**0.025**
zy-zy	140.0 [132.0; 146.0]	120–153	132.0 [128.0; 136.5]	114–184	**0.002**
go-go	112.0 [106.0; 118.0]	98–125	101.0 [94.5; 105.0]	87–130	**<0.001**

* Mean ±SD for t-Student test, or median (Q1–Q3) for Mann–Whitney test.

** Significance of male-female differences.

Table 2. Voice parameters computed from vowels for males and females.

Acoustic parameter	Males (n = 30)*	Range	Females (n = 81)*	Range	p-value**
Median pitch [Hz]	110.0 [103.9; 118.3]	87.1–174.5	207.8 [187.7; 220.8]	158.6–253.4	**<0.001**
Mean pitch (F0) [Hz]	110.0 [103.9; 120.6]	87.0–174.6	207.8 [187.7; 221.0]	159.1–256.3	**<0.001**
SD [Hz]	0.9 [0.7; 2.1]	0.4–13.7	1.2 [1.0; 1.7]	0.3–14.6	**0.045**
Minimum pitch [Hz]	107.6 [102.0; 117.0]	84.5–172.2	205.8 [185.9; 216.7]	156.3–248.6	**<0.001**
Maximum pitch [Hz]	111.6 [105.6; 122.0]	87.7–177.0	209.5 [191.0; 224.0]	163.0–285.2	**<0.001**
Jitter (LOCAL) [%]	0.43 [0.35; 0.62]	0.23–2.3	0.38 [0.29; 0.46]	0.16–0.79	**0.013**
Jitter (RAP) [%]	0.21 [0.16; 0.24]	0.12–1.32	0.21 [0.16; 0.26]	0.08–0.45	0.950
Jitter (PPQ5) [%]	0.25 [0.2; 0.32]	0.15–1.67	0.22 [0.17; 0.25]	0.09–0.43	**0.006**
Jitter (DDP) [%]	0.63 [0.48; 0.73]	0.36–3.97	0.63 [0.48; 0.77]	0.25–1.36	0.886
Shimmer (LOCAL) [%]	3.17 [1.94; 4.93]	0.83–8.33	2.41 [1.83; 3.59]	0.77–7.22	0.086
Shimmer (APQ3) [%]	1.55 [1.01; 2.2]	0.40–3.97	1.3 [0.96; 1.82]	0.40–3.42	0.148
Shimmer (APQ5) [%]	1.94 [1.23; 2.71]	0.55–6.41	1.4 [1.06; 2.08]	0.45–4.60	**0.041**
Shimmer (APQ11) [%]	2.98 [1.83; 3.95]	0.83–5.52	1.92 [1.45; 2.71]	0.57–6.74	**0.015**
Shimmer (DDA) [%]	4.66 [3.04; 6.59]	1.21–11.92	3.9 [2.88; 5.47]	1.21–10.25	0.149
Mean HNR [dB]	18.1 ±3.3	11.5–23.9	22.4 ±2.8	15.8–28.8	**<0.001**
Intensity [dB]	82.6 [75.2; 87.5]	63.9–90.8	78.9 [72.7; 84.7]	63.0–89.7	0.102
F1 [Hz]	570.3 ±74.5	463.4–801.9	590.3 ±60.8	404.2–718.7	0.152
F2 [Hz]	2771.2 [2685.2; 2891.3]	1268.8–2190.2	1559.5 [1519.1; 1648.3]	1306.4–1981.0	0.290
F3 [Hz]	2772.5 ±162.3	2393.9–3079.8	2901.0 ±144.5	2561.1–3194.6	**<0.001**
F4 [Hz]	3730.3 ±163.5	3425.6–4072.7	4012.0 ±196.2	3561.7–4461.6	**<0.001**
F_n [Hz]	2160.8 [2090.1; 2248.2]	2008.9–2422.5	2276.3 [2209.8; 2336.1]	2089.1–2462.8	**<0.001**
P_f [Z]	−0.8 [−1.5; 0.0]	−2.2–1.6	0.3 [−0.4; 0.8]	−1.5–2.0	**<0.001**
ΔF [Hz]	1101.9 ±80.2	965.8–1270.9	1137.3 ±52.6	1018.7–1273.6	**0.030**
D_f [Hz]	1053.3 ±50.7	962.0–1154.4	1140.6 ±68.4	986.5–1324.4	**<0.001**
Apparent VTL [cm]	15.3 ±1.1	13.3–17.5	14.8 ±0.7	13.2–16.6	**0.019**
MPT [s]	20.0 [12.1; 26.8]	4.0–43.8	10.6 [7.1; 14.3]	2.4–30.2	**<0.001**

The mean pitch (F0) computed for male participants was 110.0 Hz, while for female participants, it was 207.8 Hz. Also, formants, their derivative values, and HNR values were lower in males. On the other hand, instability (jitter and shimmer), intensity, and apparent VTL were higher in males. MPT values were approximately twice as long in males compared to females (Table 2). Most head/face/neck measurements, as well as acoustics parameters, significantly differed between the sexes (Table 1 and Table 2); therefore, all subsequent analyses were conducted separately for each sex group.

3.2. Head and neck dimensions vs. acoustic parameters from vowels – partial correlations

For male participants, head circumference positively correlated with voice parameters from the shimmer parameter group, indicating that a larger head circumference is associated with greater amplitude in the sound wave from period to period (i.e., a less stable or more hoarse voice), independent of age and body height. Additionally, males with larger head circumferences had lower HNR values (indicating less stable voice quality), although this relationship did not reach statistical significance. Males with greater nose height (n-sn) had lower values of the second formant (F2, $r = -0.43$, $p = 0.023$) and higher values of apparent VTL ($r = 0.41$, $p = 0.03$; Fig. 2). In contrast, males with wider faces did not show significant changes in HNR ($r = -0.36$, *ns*) or in the third formant (F3, $r = 0.28$, *ns*). Also, male participants with increased nose height (n-sn) had shorter MPT, though this correlation was not statistically significant ($r = -0.36$, $0.05 < p < 0.1$; Table 3).

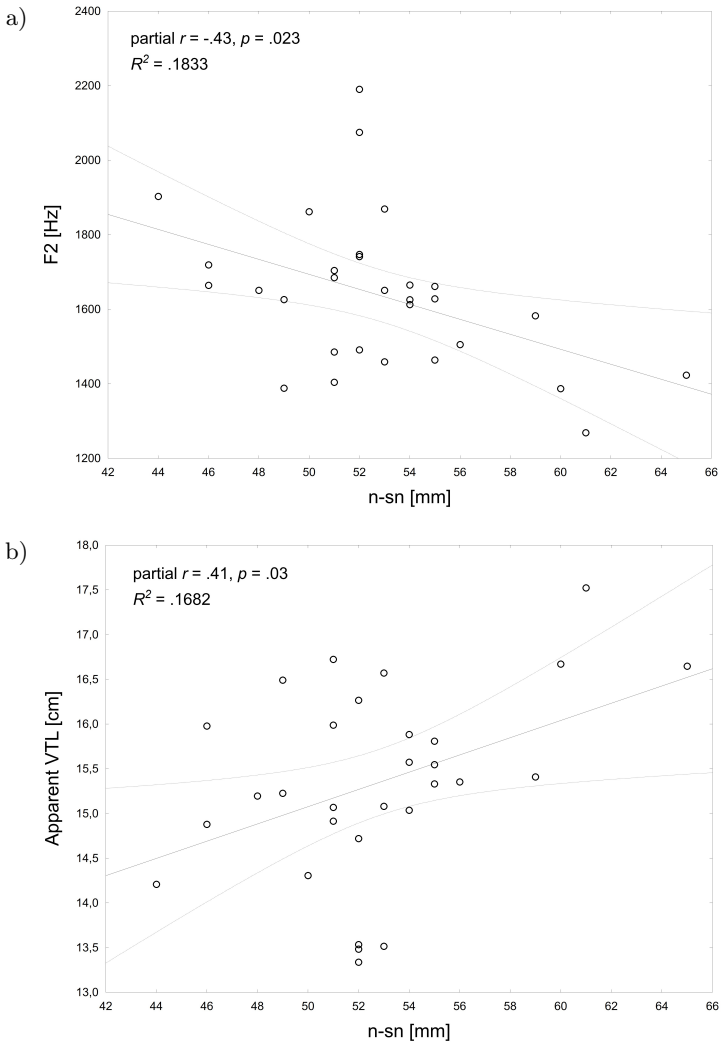


Fig. 2. Nose height (n-sn) and voice formants (from vowels) partial correlations controlling for age and body height as confounders: a) second formant (F2): $r = -0.43$, $p = 0.023$, b) apparent VTL: $r = 0.41$, $p = 0.03$ (male participants).

Table 3. Head/neck circumferences and face measurements vs. voice parameters from vowels. Pearson's correlation coefficients for males ($n = 30$) and females ($n = 81$); $p < 0.05$ results are written in **bold**, ^{b} – borderline significance ($0.05 < p < 0.06$).

Acoustic parameter	Sex (male/female)	Circumferences [cm]		Face measurements [mm]								
		Head	Neck	n-gn	n-sn	al-al	sn-sto	sto-sm	ls-li	ch-ch	zy-zy	go-go
Mean pitch (F0) [Hz]	M	−0.02	−0.03	0.12	0.15	0.27	−0.21	0.03	−0.25	0.008	0.12	0.15
	F	0.06	0.08	−0.04	−0.01	−0.10	0.03	−0.20	−0.09	0.08	0.04	0.08
Jitter [%]												
LOCAL	M	0.15	0.12	−0.05	0.11	−0.15	−0.02	0.18	−0.35	0.10	0.21	0.20
	F	−0.10	−0.05	−0.07	−0.09	0.06	−0.10	0.09	0.06	0.03	−0.12	−0.14
RAP	M	0.23	0.31	0.05	−0.01	−0.07	−0.02	0.08	−0.16	0.11	0.31	**0.45**
	F	−0.09	−0.001	−0.13	−0.13	0.03	−0.10	0.11	0.04	0.05	−0.14	−0.19
PPQ5	M	0.23	0.11	0.04	0.07	−0.10	0.06	0.10	−0.20	0.13	0.18	0.21
	F	−0.10	−0.08	−0.08	−0.04	0.03	−0.08	0.04	0.08	0.03	**−0.22**	−0.19
DDP	M	0.23	0.31	0.05	−0.006	−0.007	−0.002	0.12	−0.06	0.01	0.001	0.31
	F	−0.04	0.04	−0.12	−0.11	0.04	−0.07	0.13	0.06	0.10	**−0.23**	−0.15
Shimmer [%]												
LOCAL	M	**0.41**	0.21	−0.11	0.08	0.04	0.05	0.03	−0.18	**0.36**^{b}	0.20	0.21
	F	−0.17	**−0.27**	0.12	0.14	−0.01	0.17	0.01	**0.29**	−0.03	0.007	**−0.22**^{b}
APQ3	M	**0.43**	0.26	−0.02	0.03	0.13	0.03	0.06	−0.21	**0.38**	0.26	0.33
	F	−0.18	**−0.27**	0.10	0.07	−0.03	0.18	0.02	**0.25**	−0.04	−0.03	**−0.22**^{b}
APQ5	M	**0.42**	0.22	−0.006	0.01	0.19	0.01	0.04	−0.19	0.11	0.23	0.24
	F	−0.01	**−0.25**	0.11	0.09	0.07	0.19	0.03	**0.29**	−0.01	0.03	−0.20
APQ11	M	**0.39**	0.22	−0.19	0.01	0.01	0.09	−0.02	−0.21	**0.38**	0.24	0.14
	F	−0.14	**−0.25**	0.16	0.12	−0.11	0.15	0.01	**0.32**	0.002	0.06	−0.14
DDA	M	**0.43**	0.26	0.02	0.03	0.13	0.03	0.06	−0.21	0.21	0.26	0.33
	F	−0.18	**−0.27**	0.10	0.07	−0.03	0.18	0.02	**0.25**	−0.04	−0.03	**−0.22**^{b}
HNR [dB]	M	−0.32	**−0.40**	0.03	0.08	−0.19	−0.07	0.03	0.26	−0.29	−0.36	−0.28
	F	0.12	0.17	−0.19	−0.03	−0.07	−0.14	0.04	−0.13	0.03	0.13	0.13
Intensity [dB]	M	−0.17	−0.08	−0.005	−0.15	0.18	−0.15	0.01	−0.21	−0.29	−0.12	−0.16
	F	0.06	0.08	**−0.46**	−0.21	0.10	−0.11	0.11	−0.11	−0.07	−0.13	0.03
Formants and their derivatives												
F1 [Hz]	M	0.003	0.11	−0.24	−0.25	0.22	−0.05	0.06	−0.21	−0.06	−0.06	−0.08
	F	−0.11	−0.16	0.02	−0.05	−0.07	0.14	−0.07	−0.08	−0.06	−0.08	−0.10
F2 [Hz]	M	0.17	0.25	−0.14	**−0.43**	0.30	0.13	0.02	−0.26	0.01	0.18	0.10
	F	−0.15	−0.04	0.07	−0.21	−0.006	−0.05	−0.09	−0.15	−0.005	−0.07	0.03
F3 [Hz]	M	0.03	0.19	−0.13	−0.21	0.07	−0.16	−0.26	−0.26	0.12	0.28	0.24
	F	−0.10	−0.13	0.03	0.11	0.13	−0.06	0.11	−0.05	0.10	−0.12	0.01
F4 [Hz]	M	0.05	0.17	0.08	−0.09	0.15	0.09	−0.06	0.03	0.07	0.13	0.01
	F	−0.003	0.08	−0.16	**−0.24**	0.11	0.03	−0.20	−0.17	−0.08	−0.08	0.12
P_f [Z]	M	0.10	0.25	−0.12	−0.33	0.24	0.03	−0.09	−0.27	0.06	0.21	0.12
	F	−0.11	−0.05	−0.05	−0.15	0.10	0.002	−0.13	−0.13	−0.01	−0.13	0.08
ΔF [Hz]	M	0.09	0.22	−0.19	−0.35	0.27	0.01	−0.03	−0.29	0.01	0.13	0.06
	F	−0.15	−0.13	0.003	−0.14	0.02	0.06	−0.10	−0.10	0.05	−0.12	−0.07
D_f [Hz]	M	0.05	0.12	0.20	0.03	0.04	0.12	0.25	0.14	0.10	0.17	0.05
	F	0.03	0.12	−0.15	−0.17	0.12	−0.01	−0.07	−0.05	−0.05	−0.05	0.15
Apparent VTL [cm]	M	−0.11	−0.25	0.22	**0.41**	−0.28	−0.04	0.02	0.27	−0.02	−0.16	−0.08
	F	0.14	0.12	0.01	0.14	−0.03	−0.07	0.12	0.14	0.04	0.14	0.07
MPT [s]	M	0.08	−0.006	−0.03	−0.36	0.12	0.001	−0.35	−0.08	0.27	0.22	0.12
	F	0.08	0.19	0.17	0.08	−0.04	0.07	0.04	−0.03	−0.07	−0.07	**0.25**

For female participants, more significant associations between head and face dimensions and acoustic parameters extracted from vowels were observed (Table 3). Female participants with larger neck circumferences had lower values of shimmer parameters (APQ3, DDA: $r = -0.27$, $p = 0.018$, and APQ5, APQ11: $r = -0.25$, $p = 0.032$). Furthermore, females with higher face height (n-gn) had lower voice intensity (loudness: $r = -0.46$, $p < 0.001$).

Additionally, a higher nose height (n-sn), lower fourth formant (F4), and longer apparent VTL showed a significant positive relationship with lip vermilion height (ls-li) and voice parameters. A higher value of this characteristic

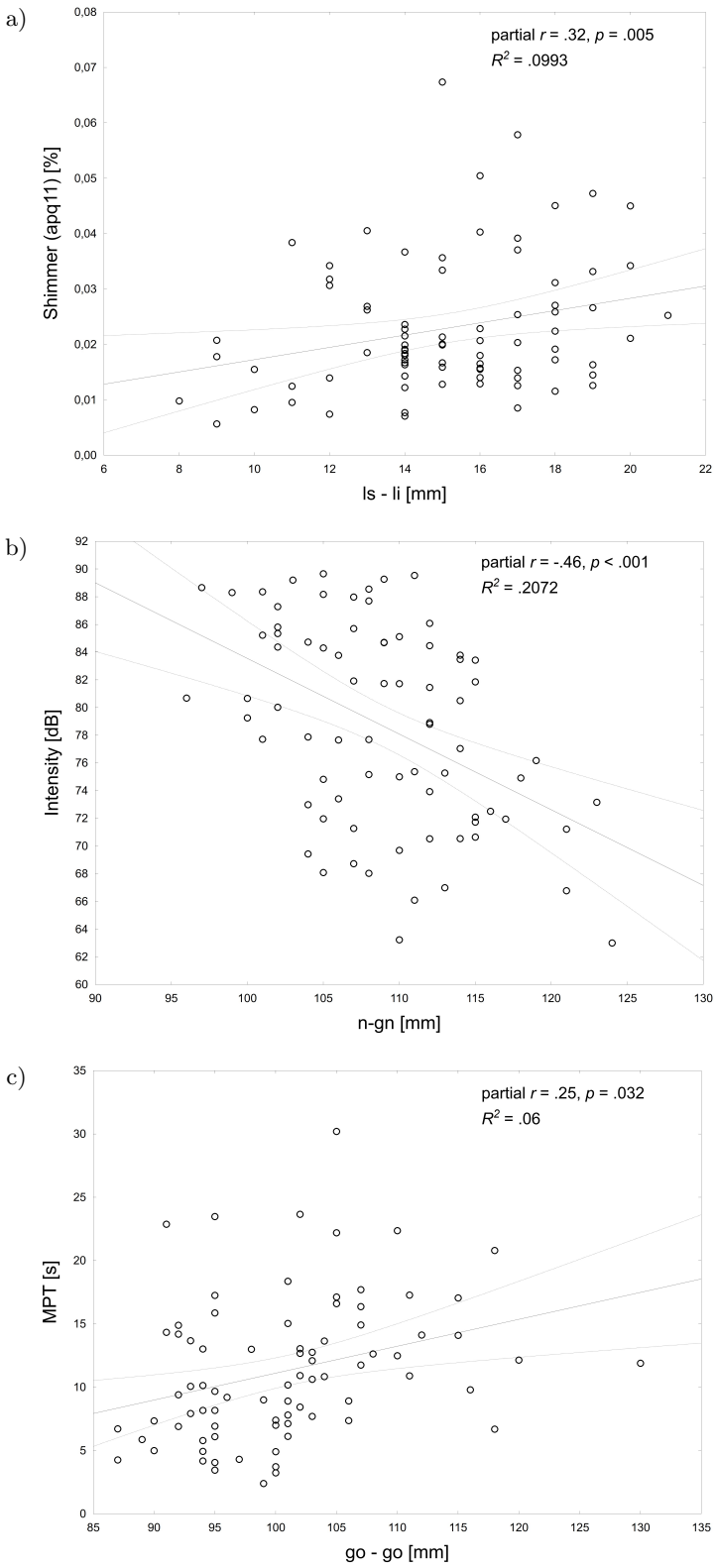


Fig. 3. Partial linear correlations controlling for age and body height as confounders in female participants: a) lip vermilion height (ls-li) and shimmer (APQ11): $r = 0.32$, $p = 0.005$, b) face height (n-gn) vs. voice intensity (loudness): $r = -0.46$, $p < 0.001$, c) jaw width (go-go) and MPT: $r = 0.25$, $p = 0.032$.

was positively correlated with shimmer parameters, indicating that greater lip vermillion was associated with an unstable (hoarse) voice. This was related to a greater face width (zy-zy) and lower jitter values (RAP, PPQ5, DDP). Also, jaw width (go-go) showed similar trends with jitter parameters, although these correlations did not reach statistical significance. Meanwhile, female participants with wider jaws (go-go) had higher values of MPT ($r = 0.25$, $p = 0.032$; Table 3, Fig. 3).

Most of the aforementioned correlations were weak (~0.2–0.3) as it was expected, but some of them (i.e., nose height (n-sn) vs. F2 showed a moderate effect size for male participants.

3.3. Head and neck dimensions vs. acoustic parameters for vowels: Multiple regression models

For males, multiple regression modeling indicated a significant positive relationships between head circumference and shimmer (LOCAL), meaning that males with larger head circumferences had more unstable (hoarse) voices ($\beta = 0.41$, $t = 2.25$, $p = 0.034$; see Table 4, model 1). The second model revealed a significant correlation between nose height (n-sn) and apparent VTL, indicating that males with higher noses had longer estimated vocal tracts ($\beta = 0.42$, $t = 2.28$, $p = 0.031$; see Table 4, model 2).

Table 4. Multiple general regression models (GRMs) of acoustic parameters (dependent variables) and head/neck measurements (predictors) for males ($n = 30$) and females ($n = 81$).

Sex	Regression model	Predictors	β	SE_β	t	p-value
Males	1. Dependent variable: shimmer (LOCAL) [%] $R^2_{\text{Adj.}} = 13.29\,\%$, $\varepsilon = 0.017$ $F = 3.22$, $p < 0.06$	Intercept	–	–	−1.73	0.097
		Head circumferences [cm]	0.41	0.18	2.25	**0.034**
		n-sn [mm]	0.07	0.19	0.38	0.704
		Age [years]	0.09	0.19	0.46	0.650
		Body height [cm]	0.02	0.18	0.10	0.918
	2. Dependent variable: apparent VTL [cm] $R^2_{\text{Adj.}} = 12.42\,\%$, $\varepsilon = 0.99$ $F = 3.06$, $p < 0.06$	Intercept	–	–	1.15	0.260
		Head circumferences [cm]	−0.12	0.18	−0.65	0.522
		n-sn [mm]	0.42	0.18	2.28	**0.031**
		Age [years]	−0.13	0.19	−0.70	0.489
		Body height [cm]	0.16	0.18	0.91	0.373
Females	3. Dependent variable: shimmer (APQ11) [%] $R^2_{\text{Adj.}} = 13.45\,\%$, $\varepsilon = 0.011$ $F = 3.43$, $p < 0.008$	Intercept	–	–	−2.52	**0.014**
		ls-li [mm]	2.99	0.004	2.99	**0.004**
		go-go [mm]	−1.52	0.132	−1.52	0.132
		n-gn [mm]	1.84	0.070	1.84	0.070
		Age [years]	0.33	0.13	2.44	**0.017**
		Body height [cm]	0.31	0.12	2.71	**0.008**
	4. Dependent variable: intensity [dB] $R^2_{\text{Adj.}} = 0.16\,\%$, $\varepsilon = 7.15$ $F = 1.02$, $p < 0.410$	Intercept	–	–	4.44	**<0.001**
		ls-li [mm]	−0.13	0.12	−1.07	0.288
		go-go [mm]	0.06	0.13	0.48	0.632
		n-gn [mm]	−0.23	0.12	−1.96	**0.054**^{b}
		Age [years]	−0.20	0.14	−1.37	0.175
		Body height [cm]	−0.09	0.13	−0.73	0.467
	5. Dependent variable: MPT [s] $R^2_{\text{Adj.}} = 8.59\,\%$, $\varepsilon = 5.52$ $F = 2.47$, $p < 0.040$	Intercept	–	–	−2.18	**0.033**
		ls-li [mm]	−0.02	0.12	−0.16	0.875
		go-go [mm]	0.24	0.12	1.98	**0.052**^{b}
		n-gn [mm]	0.14	0.11	1.23	0.221
		Age [years]	0.06	0.14	0.46	0.650
		Body height [cm]	0.20	0.12	1.67	0.099

β – standardized regression coefficient, SE_β – standard error of β, $R^2_{\text{Adj.}}$ – adjusted R^2 of the model, ε – standardized random error, t – t-statistic of β showing predictor significance, F – Fisher test value showing overall model significance. The results significant on the $p < 0.05$ level are written in **bold**, ^{b} – borderline significance ($0.05 < p < 0.06$).

For females, the first multiple regression model revealed that lip vermillion height (ls-li) significantly correlated with shimmer (APQ11), indicating that a larger ls-li ($\beta = 0.35$, $t = 2.99$, $p = 0.004$) was associated with a more hoarse voice ($\beta = 0.35$, $t = 2.99$, $p = 0.004$; see Table 4, Model 3). The second model showed a borderline

significance negative correlation between face height (n-gn) and voice intensity (loudness), which means that females with higher faces had quieter voices ($\beta = -0.23$, $t = -1.96$, $p = 0.054$, see Table 4, Model 4). The last significant model created for females showed a borderline significance correlation between jaw width (go-go) and MPT. This associations means that women with wider jaws had longer maximum time of phonation ($\beta = 0.24$, $t = 1.98$, $p = 0.052$, see Table 4, Model 5).

4. Discussion

Significant relationships between head, neck, and face dimensions and voice parameters were observed for both sexes. The strongest relationships for males occurred between shimmer parameters and head circumference, F2, nose height (n-sn) and apparent VTL, as well as jitter (RAP) and jaw width (go-go). Thus, males with larger heads exhibited more hoarse (unstable) voices, and those with higher noses were associated with lower values of F2 and longer apparent VTL. Males with wider jaws had greater voice roughness. For females, the strongest connections included: face height (n-gn) and voice intensity, neck circumference and lip vermilion height (ls-li) with shimmer parameters, nose height (n-sn) and F4, and jaw width (go-go) and MPT. These results suggest that females with smaller face heights had higher voice intensity, and those with greater neck circumferences and lower lip height had more stable (less hoarse) voices. Finally, females with longer noses had lower F4, while those with wider jaws – longer MPT. Furthermore, the relationships between lip vermilion height and shimmer remained significant after applying multiple linear regression models. There were no significant relationships between F0, formant frequencies (F1-F4), or their derivatives (P_f or D_f) and head or neck circumferences for males and females, although such relationships were found for shimmer parameters. These results partly confirm Evans *et al.* (2006), who reported the lack of such relationships for males, and support Rendall *et al.* (2005), who found no significant connections for both sexes. In contrast, Pawelec *et al.* (2022) reported significant associations between neck circumference and formants and F0 in males, even after applying multiple regression models.

For females, the strongest relationships based on multiple linear regression models were reported for jaw width and MPT (jaw width had the strongest effect on MPT). Interestingly, no significant associations were found between voice pitch (F0) and head, neck, or face dimensions. Macari *et al.* (2017) found a significant negative correlation between face width (zy-zy) and habitual pitch (F0 for the sentence) in males and inverse correlations between habitual pitch and total face height (n-me measurement corresponding to our n-gn) and jaw width for females. Moreover, one study showed a negative relationship between Martin's facial index (n-gn/zy-zy) and F3 in males and a positive correlation for females (Bommarito *et al.*, 2019). Consequently, females with higher and narrower faces and males with shorter and wider faces had higher values of F3. Wu *et al.* (2022) found that voice may indicate wider or thinner faces and stated that 'the best indicative attribute voice can hint is the head width.' This finding is consistent with the results of the current study, which indicate many significant correlations between acoustic parameters and face width measurements (go-go, zy-zy, ch-ch) in males and females.

The observed connections between facial morphology and voice parameters found in this study are logical and result from the anatomy and physiology of the speech apparatus. It is known that the length of the vocal tract is closely related to body size and shape, including body height and mass (Fitch, Giedd, 1999), and taller and heavier individuals tend to have lower voices, expressed by lower values of fundamental and formant frequencies (Evans *et al.*, 2006; González, 2004; 2007; Pawelec *et al.*, 2022; Pisanski *et al.*, 2014; Rendall *et al.*, 2005). Larger individuals also have a bigger larynx and a longer vocal tract (Fitch, Giedd, 1999), and such anatomical structures affect the voice through longer vocal folds, lower F0, a longer vocal tract, and lower formant frequencies (Fitch, 1997; Titze, 2011). Therefore, an inverse relationship between body size and voice parameters is observed (Bruckert *et al.*, 2006; Rendall *et al.*, 2005), which has also been noted in subjective assessment of body build conducted by judges (Pawelec *et al.*, 2023). A positive correlation has also been found between head length (g-op), face height (n-gn), and growth indexes of vocal tract structures during early childhood (Vorperian *et al.*, 1999). Thus, the significant associations between head and face dimensions and acoustic parameters found in this study seem justified and logical. Moreover, all presented results show pure effect of head/neck/face dimensions on

vocal characteristics, independent of body size and age. Body size (expressed as body height) and chronological age of the participants were controlled as confounders in the partial correlations and regression models; thus, their apparent impact on head/neck/face – voice relationships was reduced.

Relationships between jaw width and acoustic parameters in females may seem difficult to explain at first glance. However, females with greater face width tend to have lower jitter values (i.e., more stable voices). Moreover, females with wider mandibles have lower shimmer values (i.e., less hoarse voice) and greater MPT. In another study, it was also reported that female vocalists had significantly greater facial dimensions and mandible width than a control group of non-vocalists (Wyganowska-Świątkowska *et al.*, 2013). These results may suggest a role for facial and jaw morphology in voice production (especially articulation) and speech processes. Larger jaw size is likely associated with stronger development of the muscles attached to the mandible and greater development of the entire stomatognathic apparatus. Consequently, this may affect the biomechanics of these structures during voice production. Some support for this thesis is provided by results showing that nonsingers using a low mandible maneuver (LMM), a technique employed by top-ranked professional singers to enhance vocal output by altering oral, pharyngeal, and laryngeal configurations, demonstrated increased voice intensity (sound-pressure level (SPL)) and lower F1 and F2 values (Mercer, Lowell, 2020).

The significant negative relationships between face height, nose height, intensity, and formants, as well as the positive correlations between these dimensions and apparent VTL in both males and females, can be explained by the anatomy of the speech apparatus. Facial morphology, such as face and nose height, is associated with the height of oral and nasal cavity and the pharyngeal airway (Kikuchi, 2008). Total VTL is highly positively correlated with oral cavity length, pharyngeal cavity length, and palate height in both sexes (Roers *et al.*, 2009). Furthermore, VTL is inversely related to formant frequencies and their derivatives (Fitch, 1997). Thus, the aforementioned relationships likely result in an apparent correlation between face and nose height and the formant frequencies of the voice.

No associations between lip vermilion height and formants or their derivatives in either sex were found. This result is puzzling, as some studies have shown a relationship between lip shape and F1 and F2 (Ladefoged *et al.*, 1978), or a significant association between formants and lip rounding (Wood, 1986). A significant relationship was also reported between mouth shape and the first two formants, with F1 closely related to mouth height and F2 corresponding to mouth width (Kim *et al.*, 2002). Another study showed a high correlation between F1 and mouth height, but no correlation between F2 and mouth height (Erber, 1979). In the present study, lip vermilion height was positively correlated with shimmer parameters, but only in females.

4.1. Study limitations and future perspectives

The study was limited by the restricted number of head and neck measurements, and future studies should examine additional head and face dimensions, such as head length (glabella-opisthocranion (g-op)), physiognomic face height (trichion-gnation (tr-gn)), forehead width (frontotemporal-frontotemporal (ft-ft)), and forehead height (trichion-nasion (tr-n)). Furthermore, it is essential to examine the impact of additional vocal tract structures, such as the frontal sinuses, on voice characteristics. Another limitation is the lack of head and neck imaging, such as X-ray, computed tomography, and magnetic resonance imaging, which would have allowed the true dimensions of the participants' vocal tracts to be determined and their influence on voice acoustic parameters to be highlighted. In addition, the unequal number of male and female participants makes cross-sex comparisons difficult. Some of the relationships presented in this study may have limited practical applicability due to the methodology used – isolated sustained vowels. Correlations with acoustic parameters computed from short sentence or spontaneous speech are likely to have greater ecological validity, as these signals more closely resemble natural conditions. Another limitation of the study is the lack of control for other factors that may influence voice quality, such as hormones (Damrose, 2009; Kirgezen *et al.*, 2017; Newman *et al.*, 2000; Voelter *et al.*, 2008), social context (Sorokowski *et al.*, 2019), emotions (Klasmeyer, Sendlmeier, 2000; Raine *et al.*, 2019; Rothkrantz *et al.*, 2004; Sondhi *et al.*, 2015), use of stimulants (Byeon, Cha, 2020; Moreira *et al.*, 2015), or various aspects of physical and mental health status (Arnocky *et al.*, 2018).

Finally, the dimensions of the palate, including height, width, and length, could be included in future studies, as there is an evidence that palate shape influences voice type in opera singers (Bottalico *et al.*, 2021; Marunick, Menaldi, 2000).

5. Conclusions

Some significant correlations were found between head and face morphology and vocal acoustic parameters in males and females, which suggests a relationship between vocal tract structure and function and voice production. Indeed, both facial width and length measurements, in particular, showed significant negative relationships with voice parameters (Bommarito *et al.*, 2019; Macari *et al.*, 2015; 2017; Reinheimer *et al.*, 2021). Connections were also observed between head and neck circumferences and acoustic parameters (Pawelec *et al.*, 2022; Reinheimer *et al.*, 2021). VTL and shape were associated with voice parameters, especially formant frequencies (Fitch, 1997; Story *et al.*, 2001), as well as body size, height, and weight (Fitch, Giedd, 1999). Moreover, head and face size and proportions were related to body size, VTL, and shape, indicating that correlations may be observed between face and head dimensions and acoustic parameters. All observed associations were quite weak in most cases and some were moderate. These relationships could potentially be applied by forensic scientists to estimate the facial morphology of offenders based solely on voice recordings (Bunker, 2017; Oh *et al.*, 2019). It is necessary to expand this research by increasing the number of head and face measurements and sample size to allow for meta-analyses of the relationship between facial morphology and voice acoustic parameters. Another promising approach is to use geometric morphometrics, which considers not only the size of the head/face but also their shape through indices describing facial geometry. The results of such studies would enable investigators to identify perpetrators and victims based solely on audio recordings or to evaluate voice changes in people following surgical interventions of the head and face (medical and aesthetic plastic surgery). Recent research indicates some degree of success in estimating facial morphology from the voice signal (Li *et al.*, 2023).

Funding

This work was supported by the Wrocław University of Environmental and Life Sciences, Poland, as the Ph.D. research program 'Bon doktoranta SD UPWr' no. N020/0008/20 from the subsidy increased by the Ministry of Science and Higher Education for the period 2020–2026 in the amount of 2 % of the subsidy referred to Art. 387 (3) of the Act of 20 July 2018 – Law on Higher Education and Science, obtained in 2019.

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.

Authors' contributions

Łukasz Pawelec conceptualized the study and wrote the original draft, performed the analysis, and contributed to data interpretation, investigation, visualization, review, and editing. Kamila Słowik contributed to the investigation, conceptualization, performed the analysis. Anna Lipowicz contributed to conceptualization, wrote the original draft, project administration and supervision. All authors reviewed and approved the final manuscript.

Acknowledgments

We would like to express our gratitude to all the participants who have given their voice as well as anthropometric measurements for this research.

References

1. Abitbol J., Abitbol P., Abitbol B. (1999), Sex hormones and the female voice, *Journal of Voice*, **13**(3): 424–446, https://doi.org/10.1016/S0892-1997(99)80048-4.
2. Arnocky S., Hodges-Simeon C.R., Ouellette D., Albert G. (2018), Do men with more masculine voices have better immunocompetence?, *Evolution and Human Behavior*, **39**(6): 602–610, https://doi.org/10.1016/j.evolhumbehav.2018.06.003.
3. Boersma P., Weenink D. (2019), *Praat: Doing phonetics by computer*, Computer program, Version 6.0.56, http://www.praat.org/ (access: 20.06.2019).
4. Bommarito S. *et al.* (2019), Correlation between voice, speech, body and facial types in young adults, *Global Journal of Otolaryngology*, **20**(4): 556041, https://doi.org/10.19080/gjo.2019.20.556041.
5. Bottalico P., Marunick M.T., Nudelman C.J., Webster J., Jackson-Menaldi M.C. (2021), Singing voice quality: The effects of maxillary dental arch and singing style, *Journal of Voice*, **35**(3): 501.e11–501.e18, https://doi.org/10.1016/j.jvoice.2019.09.015.
6. Brattström V., Odenrick L., Leanderson R. (1991), Dentofacial morphology in professional opera singers, *Acta Odontologica Scandinavica*, **49**(3): 147–151, https://doi.org/10.3109/00016359109005899.
7. Bruckert L., Liénard J.-S., Lacroix A., Kreutzer M., Leboucher G. (2006), Women use voice parameters to assess men's characteristics, *Proceedings. Biological Sciences*, **273**(1582): 83–89, https://doi.org/10.1098/rspb.2005.3265.
8. Bunker D. (2017), Speech2Face: Reconstructed lip syncing with generative adversarial networks. Data reflexions: Thoughts and Projects.
9. Burton A.M., Wilson S., Cowan M., Bruce V. (1999), Face recognition in poor-quality video: Evidence from security surveillance, *Psychological Science*, **10**(3): 243–248, https://doi.org/10.1111/1467-9280.00.
10. Byeon H., Cha S. (2020), Evaluating the effects of smoking on the voice and subjective voice problems using a meta-analysis approach, *Scientific Reports*, **10**(1): 4720, https://doi.org/10.1038/s41598-020-61565-3.
11. Cazacu C.J. *et al.* (2025), Morphology of facial aging: A shape-based quantification, *Romanian Journal of Morphology and Embryology = Revue Roumaine de Morphologie et Embryologie*, **66**(2): 367–373, https://doi.org/10.47162/RJME.66.2.10.
12. Dabbs Jr. J.M., Mallinger A. (1999), High testosterone levels predict low voice pitch among men, *Personality and Individual Differences*, **27**(4): 801–804, https://doi.org/10.1016/S0191-8869(98)00272-4.
13. Damrose E.J. (2009), Quantifying the impact of androgen therapy on the female larynx, *Auris Nasus Larynx*, **36**(1): 110–112, https://doi.org/10.1016/j.anl.2008.03.002.
14. Erber N.P. (1979), Real-time synthesis of optical lip shapes from vowel sounds, *The Journal of the Acoustical Society of America*, **66**(5): 1542–1544, https://doi.org/10.1121/1.383511.
15. Evans S., Neave N., Wakelin D. (2006), Relationships between vocal characteristics and body size and shape in human males: An evolutionary explanation for a deep male voice, *Biological Psychology*, **72**(2): 160–163, https://doi.org/10.1016/j.biopsycho.2005.09.003.
16. Evans S., Neave, N. Wakelin D., Hamilton C. (2008), The relationship between testosterone and vocal frequencies in human males, *Physiology & Behavior*, **93**(4–5): 783–788, https://doi.org/10.1016/j.physbeh.2007.11.033.
17. Fant G. (1960), *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations*, 2nd ed., Walter de Gruyter.
18. Fitch W.T. (1997), Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques, *The Journal of the Acoustical Society of America*, **102**(2): 1213–1222, https://doi.org/10.1121/1.421048.
19. Fitch W.T., Giedd J. (1999), Morphology and development of the human vocal tract: A study using magnetic resonance imaging, *The Journal of the Acoustical Society of America*, **106**(3): 1511–1522, https://doi.org/10.1121/1.427148.
20. González J. (2004), Formant frequencies and body size of speaker: A weak relationship in adult humans, *Journal of Phonetics*, **32**(2): 277–287, https://doi.org/10.1016/S0095-4470(03)00049-4.

21. González J. (2007), Correlations between speakers' body size and acoustic parameters of voice, *Perceptual and Motor Skills*, **105**(1): 215–220, https://doi.org/10.2466/pms.105.1.215-220.

22. Graddol D., Swann J. (1983), Speaking fundamental frequency: Some physical and social correlates, *Language and Speech*, **26**(4): 351–366, https://doi.org/10.1177/002383098302600403.

23. Graja K., Król B. (2022), *Anthropometry and Anthroposcopy. Lecture Notes* [in Polish: *Antropometria i Antroposkopia. Skrypt do Ćwiczeń*], Unpublished teaching materials of the Division of Anthropology, Wroclaw University of Environmental and Life Sciences.

24. Hamdan A.L. *et al.* (2012), Relationship between acoustic parameters and body mass analysis in young males, *Journal of Voice*, **26**(2): 144–147, https://doi.org/10.1016/j.jvoice.2011.01.011.

25. Hamdan A.L.H. *et al.* (2013), Formant frequency in relation to body mass composition, *Journal of Voice*, **27**(5): 567–571, https://doi.org/10.1016/j.jvoice.2012.09.005.

26. Jandová M., Urbanová P. (2016), The relationship between facial morphology, body measurements and socio-economic factors, *Anthropological Review*, **79**(2): 181–200, https://doi.org/10.1515/anre-2016-0014.

27. Kamachi M., Hill H., Lander K., Vatikiotis-Bateson E. (2003), 'Putting the face to the voice': Matching identity across modality, *Current Biology*, **13**(19): 1709–1714, https://doi.org/10.1016/j.cub.2003.09.005.

28. Kikuchi Y. (2008), Three-dimensional relationship between pharyngeal airway and maxillo-facial morphology, *The Bulletin of Tokyo Dental College*, **49**(2): 65–75, https://doi.org/10.2209/tdcpublication.49.65.

29. Kim T., Kang Y., Ko H. (2002), Achieving real-time lip-synch via SVM-based phoneme classification and lip shape refinement, [in:] *Proceedings Fourth IEEE International Conference on Multimodal Interfaces*, pp. 299–304, https://doi.org/10.1109/6ICMI.2002.1167010.

30. Kirgezen T., Sunter A.V., Yigit O., Huq G.E. (2017), Sex hormone receptor expression in the human vocal fold subunits, *Journal of Voice*, **31**(4): 476–482, https://doi.org/10.1016/j.jvoice.2016.11.005.

31. Klasmeyer G., Sendlmeier W.F. (2000), Voice and emotional states, [in:] *Voice Quality Measurement*, Kent R.D., Ball M.J. [Eds.], pp. 339–357, Singular Publishing.

32. Kogelschatz L., Barenholtz E. (2013), Matching voice and face identity from static images, *Journal of Vision*, **12**(9): 1023, https://doi.org/10.1167/12.9.1023.

33. Krauss R.M., Freyberg R., Morsella E. (2002), Inferring speakers' physical attributes from their voices, *Journal of Experimental Social Psychology*, **38**(6): 618–625, https://doi.org/10.1016/S0022-1031(02)00510-3.

34. Ladefoged P., Harshman R., Goldstein L., Rice L. (1978), Generating vocal tract shapes from formant frequencies, *The Journal of the Acoustical Society of America*, **64**(4): 1027–1035, https://doi.org/10.1121/1.382086.

35. Li X., Wen Y., Yang M., Wang J., Singh R., Raj B. (2023), Rethinking voice-face correlation: A geometry view, [in:] *MM '23: Proceedings of the 31st ACM International Conference on Multimedia*, pp. 2458–2467, https://doi.org/10.1145/3581783.3611779.

36. Lucas T., Hatfield D., Henneberg M. (2023), A morphological comparison between a death mask of the American Prophet Joseph Smith and a photograph likely to depict him, *Anthropological Review*, **85**(4): 1–13, https://doi.org/10.18778/1898-6773.85.4.01.

37. Macari A.T. *et al.* (2017), Association between facial length and width and fundamental frequency, *Journal of Voice*, **31**(4): 410–415, https://doi.org/10.1016/j.jvoice.2016.12.001.

38. Macari A.T., Karam I.A., Tabri D., Sarieddine D., Hamdan A.L. (2015), Formants frequency and dispersion in relation to the length and projection of the upper and lower jaws, *Journal of Voice*, **29**(1): 83–90, https://doi.org/10.1016/j.jvoice.2014.05.011.

39. Martin R. (1914), A *Textbook of Anthropology: A Systematic Presentation* [in German: *Lehrbuch der Anthropologie in Systematischer Darstellung*], Jena, Gustav Fischer.

40. Marunick M.T., Menaldi C.J. (2000), Maxillary dental arch form related to voice classification: A pilot study, *Journal of Voice*, **14**(1): 82–91, https://doi.org/10.1016/S0892-1997(00)80097-1.

41. Mercer E., Lowell S.Y. (2020), The low mandible maneuver: Preliminary study of its effects on aerodynamic and acoustic measures, *Journal of Voice*, **34**(4): 645.e1–645.e9, https://doi.org/10.1016/j.jvoice.2018.12.005.

42. Moreira T. de C. *et al.* (2015), Substance use, voice changes and quality of life in licit and illicit drug users [in Portuguese], *Revista CEFAC*, **17**(2): 374–384, https://doi.org/10.1590/1982-021620156714.

43. Newman S.R., Butler J., Hammond E.H., Gray S.D. (2000), Preliminary report on hormone receptors in the human vocal fold, *Journal of Voice*, **14**(1): 72–81, https://doi.org/10.1016/S0892-1997(00)80096-X.

44. Ning H., Zheng X., Lu X., Yuan Y. (2021), Disentangled representation learning for cross-modal biometric matching, *IEEE Transactions on Multimedia*, **24**: 1763–1774, https://doi.org/10.1109/TMM.2021.3071243.

45. Oh T.H. *et al.* (2019), Speech2Face: Learning the face behind a voice, [in:] *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7539–7548, https://doi.org/10.1109/cvpr.2019.00772.

46. O'Toole A., Weimer S., Dunlop J., Barwick R., Ayyad J., Phillips J. (2010), Recognizing people from dynamic video: Dissecting identity information with a fusion approach, *Journal of Vision*, **10**(7): 643, https://doi.org/10.1167/10.7.643.

47. Pawelec Ł.P., Graja K., Lipowicz A. (2022), Vocal indicators of size, shape and body composition in Polish men, *Journal of Voice*, **36**(6): 878.e9–878.e22, https://doi.org/10.1016/j.jvoice.2020.09.011.

48. Pawelec Ł., Kierczak K., Lipowicz A. (2023), Assessment of the obesity based on voice perception, *Anthropological Review*, **85**(4): 43–60, https://doi.org/10.18778/1898-6773.85.4.04.

49. Perini T.A., Oliveira G.L. de, Ornellas J. dos S., Oliveira F.P. de (2005), Technical error of measurement in anthropometry [in Portuguese], *Revista Brasileira De Medicina Do Esporte*, **11**(1): 81–85, https://doi.org/10.1590/S1517-86922005000100009.

50. Pisanski K. *et al.* (2014), Vocal indicators of body size in men and women: A meta-analysis, *Animal Behaviour*, **95**: 89–99, https://doi.org/10.1016/j.anbehav.2014.06.011.

51. Pisanski K. *et al.* (2016), Voice parameters predict sex-specific body morphology in men and women, *Animal Behaviour*, **112**: 13–22, https://doi.org/10.1016/j.anbehav.2015.11.008.

52. Raine J., Pisanski K., Simner J., Reby D. (2019), Vocal communication of simulated pain, *Bioacoustics*, **28**(5): 404–426, https://doi.org/10.1080/09524622.2018.1463295.

53. Raj A., Gupta B., Chowdhury A., Chadha S. (2010), A study of voice changes in various phases of menstrual cycle and in postmenopausal women, *Journal of Voice*, **24**(3): 363–368, https://doi.org/10.1016/j.jvoice.2008.10.005.

54. Reinheimer D.M. *et al.* (2021), Formant frequencies, cephalometric measures, and pharyngeal airway width in adults with congenital, isolated, and untreated growth hormone deficiency, *Journal of Voice*, **35**(1): 61–68, https://doi.org/10.1016/j.jvoice.2019.04.014.

55. Rendall D., Kollias S., Ney C., Lloyd P. (2005), Pitch (F_0) and formant profiles of human vowels and vowel-like baboon grunts: The role of vocalizer body size and voice-acoustic allometry, *The Journal of the Acoustical Society of America*, **117**(2): 944–955, https://doi.org/10.1121/1.1848011.

56. Rice A., Phillips P.J., Natu V., An X., O'Toole A.J. (2013), Unaware person recognition from the body when face identification fails, *Psychological Science*, **24**(11): 2235–2243, https://doi.org/10.1177/0956797613492986.

57. Robbins R.A., Coltheart M. (2012), The effects of inversion and familiarity on face versus body cues to person recognition, *Journal of Experimental Psychology: Human Perception and Performance*, **38**(5): 1098–1104, https://doi.org/10.1037/a0028584.

58. Roers F., Mürbe D., Sundberg J. (2009), Voice classification and vocal tract of singers: a study of x-ray images and morphology, *The Journal of the Acoustical Society of America*, **125**(1): 503–512, https://doi.org/10.1121/1.3026326.

59. Rojas S., Kefalianos E., Vogel A. (2020), How does our voice change as we age? A systematic review and meta-analysis of acoustic and perceptual voice data from healthy adults over 50 years of age, *Journal of Speech, Language, and Hearing Research*, **63**(2): 533–551, https://doi.org/10.1044/2019_jslhr-19-00099.

60. Rothkrantz L.J.M., Wiggers P., van Wees J.W.A., van Vark R.J. (2004), Voice stress analysis, [in:] *Text, Speech and Dialogue. TSD 2004. Lecture Notes in Computer Science*, Sojka P., Kopeček I., Pala K. [Eds.], Vol. 3206, pp. 449–456, https://doi.org/10.1007/978-3-540-30120-2_57.

61. Sondhi S., Khan M., Vijay R., Salhan A.K. (2015), Vocal indicators of emotional stress, *International Journal of Computer Applications*, **122**(15): 38–43, https://doi.org/10.5120/21780-5056.

62. Sorokowski P. *et al.* (2019), Voice of authority: Professionals lower their vocal frequencies when giving expert advice, *Journal of Nonverbal Behavior*, **43**(2): 257–269, https://doi.org/10.1007/s10919-019-00307-0.

63. Story B.H., Titze I.R., Hoffman E.A. (2001), The relationship of vocal tract shape to three voice qualities, *The Journal of the Acoustical Society of America*, **109**(4): 1651–1667, https://doi.org/10.1121/1.1352085.

64. Teixeira J.P., Oliveira C., Lopes C. (2013), Vocal acoustic analysis – Jitter, shimmer and HNR parameters, *Procedia Technology*, **9**: 1112–1122, https://doi.org/10.1016/j.protcy.2013.12.124.

65. Titze I.R. (1994), Fluctuations and perturbations in vocal output, [in:] *Principles of Voice Production*, Prentice Hall, pp. 209–306.

66. Titze I.R. (2011), Vocal fold mass is not a useful quantity for describing F_0 in vocalization, *Journal of Speech Language and Hearing Research*, **54**(2): 520–522, https://doi.org/10.1044/1092-4388(2010/09-0284).

67. Voelter C. *et al.* (2008), Detection of hormone receptors in the human vocal fold, *European Archives of Oto-Rhino-Laryngology*, **265**: 1239–1244, https://doi.org/10.1007/s00405-008-0632-x.

68. Vorperian H.K., Kent R.D., Gentry L.R., Yandell B.S. (1999), Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: Preliminary results, *International Journal of Pediatric Otorhinolaryngology*, **49**(3): 197–206, https://doi.org/10.1016/S0165-5876(99)00208-6.

69. Wen P., Xu Q., Jiang Y., Yang Z., He Y., Huang Q. (2021), Seeking the shape of sound: An adaptive framework for learning voice-face association, [in:] *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16347–16356, https://doi.org/10.1109/CVPR46437.2021.01608.

70. Wood S. (1986), The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels, *The Journal of the Acoustical Society of America*, **80**(2): 391–401, https://doi.org/10.1121/1.394090.

71. Wu C.Y., Hsu C.C., Neumann U. (2022), Cross-modal perceptionist: Can face geometry be gleaned from voices?, [in:] *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10452–10461, https://doi.org/10.1109/CVPR52688.2022.01020.

72. Wyganowska-Świątkowska M., Kowalkowska I., Mehr K., Dąbrowski M. (2013), An anthropometric analysis of the head and face in vocal students, *Folia Phoniatrica et Logopaedica*, **65**(3): 136–142, https://doi.org/10.1159/000354939.

73. Young A.W., Bruce V. (2011), Understanding person perception, *British Journal of Psychology*, **102**(4): 959–974, https://doi.org/10.1111/j.2044-8295.2011.02045.x.

74. Zheng A., Hu M., Jiang B., Huang Y., Yan Y., Luo B. (2021), Adversarial-metric learning for audio-visual cross-modal matching, *IEEE Transactions on Multimedia*, **24**: 338–351, https://doi.org/10.1109/TMM.2021.3050089.

ARCHIVES OF ACOUSTICS Vol. 51, No. 1, pp. 93–105 (2026), https://doi.org/10.24423/archacoust.2026.4363

Research Paper

Ultrasonic Estimation of Pressure Dependent Non-Linearity Index in Liver

Andrzej NOWICKI(1)*, Jurij TASINKIEWICZ(1), Piotr KARWAT(1),(2), Norbert ŻOŁEK(1), Ihor TROTS(1), Ryszard TYMKIEWICZ(1)

(1) *Institute of Fundamental Technological Research, Polish Academy of Sciences*
Warsaw, Poland

(2) *us4us Ltd.*
Warsaw, Poland

*Corresponding Author: anowicki@ippt.pan.pl

Received October 25, 2025; accepted November 14, 2025;
version of record January 13, 2026; published issue March 27, 2026.

This study introduces a proof-of-concept methodology for evaluating pressure-dependent non-linear acoustic properties of liver tissue. The proposed non-linearity index (NLI) is derived from echo amplitudes obtained at two substantially different acoustic pressures. Unlike previous harmonic-based approaches, the method relies solely on the fundamental frequency band, allowing clinical implementation without additional system modifications. The image acquired for the lower pressure is then amplified to correct for the pressure difference between the beams. Next, the NLI is estimated as a ratio of local amplitudes of the amplified low-pressure image (ALPI) to the high-pressure image (HPI). In the case of nonlinear media some energy of the wave is transferred from the pulse fundamental frequency to higher harmonics, which affects mainly the HPI. With the harmonics being filtered out from the signal, the HPI amplitude becomes lower than the ALPI amplitude. As a result, the NLI becomes higher than 1 and increases with the non-linearity of the imaged tissue. The hydrophone measurements were compared to the simulation (k-Wave) of the ultrasonic field in water and vegetable oil. Next, we performed NLI imaging of healthy and fatty livers using SonixTouch (Ultrasonix) systems and two acoustic pressures of 390 kPa and 1590 kPa. Preliminary studies – imaging healthy and fatty livers using SonixTouch (Ultrasonix) systems were performed on the 4 livers of the authors of the article showed that for 'healthy' livers the NLI was below 1.1, while in one of the authors with previously diagnosed steatosis falling between score 1 and 2, the NLI locally exceeded 1.3.

These results show that the obtained NLI values increase with the degree of steatosis, which agrees with theoretical expectations based on tissue B/A coefficients. The work emphasizes methodological feasibility and physical consistency rather than clinical validation, given the limited number of volunteers and ethical restrictions on patient recruitment.

Keywords: ultrasound imaging, abdominal ultrasound, non-linear propagation.

1. Introduction

The scientific goal of the paper is to determine the relationship between the ultrasonic pressure dependent backscattering in liver tissue in vivo in subjects with diagnosed non-alcoholic fatty liver disease (NAFLD) and categorized to the specific steatosis score. An improvement of ability to non-invasively distinguish the fatty tissues will positively affect the specificity and sensitivity of the ultrasound diagnostics of liver. As a result, it will reduce a number of unnecessary biopsies. NAFLD affects 25.24 % of the world's population and 23.71 % of people in Europe, 24.13 % in North America, 30.45 % in South Africa, 31.79 % in the Middle East, and 27.37 % in Asia (YOUNOSSI *et al.*, 2016). The NAFLD prevalence of 55 % to 80 % is estimated in patients

with type 2 diabetes (Younossi *et al.*, 2002; 2016; Estes *et al.*, 2018). This results in over 1 billion individuals affected by fatty liver disease worldwide (Saadeh *et al.*, 2002). The problem of NAFLD affects people of all ages, including children, but it may be more pronounced in the elderly population.

About 10 % to 20 % of NAFLD cases are classified as non-alcoholic steatohepatitis (NASH), which must be diagnosed and treated proactively because it can lead to cirrhosis and liver cancer. A liver biopsy is currently the only method that allows for a definitive diagnosis of NASH (Saadeh *et al.*, 2002). However, routine invasive liver biopsy in all patients with such a common disease as NAFLD is unjustified. In the case of the symptoms of liver disease, a liver ultrasound – echogenicity of the liver image and recently the ultrasound attenuation of ultrasound in liver – is a first-choice method of a simple and painless imaging test of liver and its surroundings.

We believe that the implementation of a new ultrasound method for estimating tissue fat content based on the non-linear properties of tissues will provide important additional diagnostic information on the degree of steatosis and will allow for earlier detection of NAFLD, thus improving the prognosis for effective treatment of patients with NAFLD and reducing the need for liver biopsy sampling and testing. The proposed non-linear ultrasound imaging may enable the implementation in clinical practice of a new qualitative signal analysis enabling the assessment of the degree of steatosis.

Up to now, there are four main types of liver ultrasound imaging including:

- B-mode – fatty liver disease might make it brighter, while inflammation might make it denser and darker,
- vascular Doppler ultrasound – exposing the blood flow in liver,
- elastography ultrasound – allowing measuring of the stiffness or fibrosis in liver,
- contrast enhanced ultrasound (CEUS) – allowing for clearer images of liver lesions.

However, studies have shown that although B-mode images alone are useful for assessing steatosis at levels of ≥30 % (Hernaez *et al.*, 2011) they have low sensitivity at lower levels of steatosis (Dasarathy *et al.*, 2009; Bohte *et al.*, 2011). Some development work has been carried out to introduce the transfer learning with a deep convolutional neural network for determination of the hepato-renal index related to the amount of fat in the liver (Byra *et al.*, 2018).

Ultrasound attenuation imaging (ATI) is an emerging method used for detection of hepatic steatosis. It quantifies ultrasound signal attenuation and generates color-coded maps which visualize alterations in liver tissue properties, that may result from changes in hepatic composition (such as increased fat content). The ATI value is defined as dB/cm/MHz. The values increase with increasing levels of steatosis. First studies of the ultrasonic attenuation coefficients of liver have been initiated by Parker *et al.* in 1988. Their results showed that normal attenuation values for the human liver are 0.47 dB/cm/MHz with a frequency-dependent coefficient of 1.05. A trend towards higher attenuation in diffuse liver disease was observed in correlation with increasing fibrosis and fat content, based on liver biopsy grading (Parker *et al.*, 1988). Since then, a numerous reports were published on attenuation in normal and fatty liver (e.g., Tada *et al.*, 2019; Yoo *et al.*, 2020; Hsu *et al.*, 2021; Ferraioli *et al.*, 2022). As for now the guidelines for steatosis score are average (0.52, 0.63, 0.74, and ≥0.78) dB/cm/MHz for S0, S1, S2, and S3, respectively (Jesper *et al.*, 2020; Ogino *et al.*, 2021). The final results differ between research centers using different scanning machines.

Important factors influencing/modifying the measurements of liver attenuation include: respiratory phase, subject positioning, effect of hydration, liver inflammation, subcutaneous fat thickness, scanning technique, frequency dependence attenuation, and bandwidth of the transducer.

We believe that enhancing the currently used algorithm for assessing the degree of steatosis – based on attenuation estimation – by incorporating an independent parameter related to fat-dependent nonlinear propagation of ultrasound in liver tissue, may significantly improve diagnostic accuracy.

The proposed method for assessing the non-linear properties of tissues is, to our knowledge, our original idea, never published before in the world literature.

Modern ultrasound is based on linear and non-linear (tissue harmonic imaging, THI) properties of ultrasound wave propagation in tissue. The basic derivation of the B/A non-linearity coefficient is given in Appendix.

In the field of linear acoustics, particularly in lossless linear media, neglecting wave distortions and the generation of higher harmonics results in $B/A = 0$. In real materials B/A takes finite values. The documented B/A values for some fluids and biological tissue are presented in (Duck, 2002): water (20 °C and 40 °C) $B/A = 4.96$ and 5.38, 3.5 % saline solution (20 °C) $B/A = 5.25$, blood plasma (30 °C) $B/A = 5.74$, whole blood (26 °C) $B/A = 6.1$, fat-free soft tissue $B/A = 6.3$–8.0, and fatty soft tissue $B/A = 9.6$–11.3 (Varray *et al.*, 2010). Dong *et al.* (1999) reported the ultrasonic non-linearity parameter B/A for nine versions of water-based, fat-free and fat-containing, macroscopically homogeneous, ultrasonically tissue-mimicking (TM) materials. The B/A parameter is 1.5 to 2 times larger in adipose tissue than in other tissues, partially explaining the benefits of using harmonic imaging in difficult cases involving excess body fat.

The THI based on acquisitions of the second harmonic components of the ultrasonic echoes was proposed to improve the scanning resolution (van Wijk *et al.*, 2002; Varray *et al.*, 2010). The detection of the nonlinear part of the echoes include different modalities such as: amplitude modulation, pulse inversion or second harmonic inversion (Simpson *et al.*, 1999). THI suffers from lower imaging sensitivity due to the limited frequency bandwidth of available ultrasound transducers. In practice, the lower part of the transducer bandwidth is used in the transmission sequence, while the upper part of the available bandwidth is used during reception and image formation, reducing final imaging sensitivity (Nowicki *et al.*, 2007). The importance of the reduced dynamic range and penetration encountered in THI has also been pointed out (Averkiou *et al.*, 1997; Akiyama, 2000; Averkiou, 2001). THI images are reconstructed solely based on second harmonic components, which results in a reduction of the echo amplitude by more than 20 dB compared to the fundamental components.

A comprehensive review of the estimation of the B/A non-linearity parameter is given in (Panfilova *et al.*, 2021). In addition to the established medical applications of non-linearity, new ultrasound modalities for quantifying the B/A non-linearity parameter (Madigosky *et al.*, 1981) have been continuously developed for several years. Studies on aqueous solutions have shown that the B/A ratio is dependent on the chemical composition and molecular structure of the solutes (Sehgal *et al.*, 1986; Sarvazyan *et al.*, 1990; Gong *et al.*, 2004) and has been found useful for assessing the structure of silicone oil used in ocular surgery (Zhe *et al.*, 2014). It has also been shown to be useful for tissue characterization, showing distinct values for adipose (Law *et al.*, 1985), malignant, healthy, and hepatic tissues in the liver (Errabolu *et al.*, 1988; Panfilova *et al.*, 2021).

It has been shown that B/A is proportional to the change in the speed of sound that accompanies the adiabatic change in pressure. This relation is the basis of the thermodynamic method of measuring B/A (Beye, 1973; Zhu *et al.*, 1983; Zhang, Dunn, 1991; Gong *et al.*, 1993; Everbach, Apfel, 1995; Lu *et al.*, 1998). The finite amplitude method (FAM) exploits the dependence of the speed of sound on B/A and the excess density of the medium (or particle velocity, or excess pressure).

As a consequence of non-linear propagation, the effective attenuation of the wave increases due to the transfer of energy from the fundamental to higher harmonics, which experience stronger attenuation than the fundamental (Coila, Oelze, 2020). In this way, the energy of a high-amplitude non-linear wave is attenuated to a greater extent than the energy of a small-amplitude wave that can be recognized as linear. FAM exploit all these changes, quantifying the distortions by direct observation of the wave profile (Hunter *et al.*, 2016; Takahashi, 1995), by harmonic content (Adler, Hiedemann, 1962; Zhang, Gong, 1999; Fujii *et al.*, 2004; Wallace *et al.*, 2007; Varray *et al.*, 2011) or by observing the non-linearly induced attenuation (Nikoonahad, Liu, 1989; Byra *et al.*, 2017). The FAM approach is relatively simple, though less accurate than the thermodynamic method, carrying potential for clinical applications, enabling B/A tomography for transmission measurements. Recently, Coila *et al.* (2025) proposed a pulse-echo method to estimate B/A based on the theory of the fundamental band amplitude reduction of low amplitude signals.

In this paper we propose a new method for assessing the nonlinear properties of a medium. Our proposal consists in assessing the non-linearity of the medium by comparing echoes from the examined organ using waves that significantly differ in transmitted pressures. In a linear medium, the echo amplitude should be directly proportional to the amplitude of the transmitted signals. The deviation from the linear relationship between the transmitted signal amplitude and the backscatter amplitude depends on the non-linear physical properties of the tissue region being examined. We have developed a method for imaging nonlinear tissue properties that

involves comparing signals reflected/scattered in tissues for two or more transmissions that significantly differ in radiated intensity (Nowicki *et al.*, 2024). In tissues with low acoustic non-linearity, ultrasound images will vary in amplitude only proportionally to the amplitude of the transmitted wave. After normalizing echoes proportionally to the amplitude of the transmitted ultrasonic wave, the amplitude ratios of images obtained for low and high sound transmitted pressures should be close to unity. If the examined tissue contains areas with $B/A > 0$, the ratios of subsequent echoes recorded for different transmitted acoustic pressures will differ from unity, and this value should increase with the increasing non-linearity coefficient of the examined tissue. For example, the non-linearity parameter in the liver, adipose tissue, and fatty liver is approximately 7, greater than 8, and close to 11, respectively. This gives hope that this technique can support the diagnosis of fatty liver disease.

The rest of the paper is organized as follows. The following section outlines the proposed method and describes its validation process, which includes experimental measurements and numerical simulations for nonlinear propagation in water and in vegetable oil, as well as liver measurements in-vivo. The results obtained using the proposed method are presented in Sec. 3, while Sec. 4 contains the discussion.

2. Materials and methods

2.1. Method overview

Let us assume that we deal with a linear medium. In such a case, detected echoes would be linearly proportional to the pressure of the transmitted ultrasonic wave. In turn, in the case of a non-linear medium, part of the energy of the propagating wave is transferred from the carrier wave frequency f_0 to higher harmonics $2f_0$, $3f_0$, etc. The energy transfer to higher harmonics increases significantly with the increase in the initial pressure of the transmitted wave scanning the medium. The relative energy transfer to higher harmonics for low pressures of the scanning wave is small compared to this transfer for significantly higher pressures of the scanning wave. As a result, the amplitude of echoes at the fundamental frequency decreases faster with propagation distance for waves transmitted at higher pressures than for those transmitted at lower pressures. We have attempted to use this intuitively simple relationship between the transmission pressure and the rate of energy loss in the fundamental frequency.

Our approach involves acquiring pairs of ultrasound images $E1$ and $E2$, where $E1$ image is obtained using low excitation pressure $P1$, and $E2$ using high pressure $P2$. The reconstruction of these images must rely solely on the fundamental frequency band; higher harmonics should be eliminated – either due to the limited bandwidth of the transducer or through the application of additional low-pass filtering.

The further data processing includes smoothing the $E1$ and$E2$ images using moving average filtration. The final non-linearity index (NLI) image is calculated as the quotient of amplitudes in the 'low pressure' image $E1$ and in the 'high pressure' image $E2$. The echo amplitude ratio for both images is normalized by the factor equal to the transmission pressure ratio for both scans, $P2/P1$:

$$\mathrm{NLI} = \frac{P2}{P1}\frac{E1}{E2}. \tag{1}$$

2.2. Experiments overview

In research (Zhe *et al.*, 2014) the results of experiments conducted in water and in sunflower oil for two different acoustic pressures of 0.39 MPa and 1.55 MPa generated by the ultrasonic head were presented. In-vitro measurements were also performed in a fresh beef liver sample with a small area, into which 1 ml of sunflower oil was injected. Then the course of the NLI was determined in these media. These studies confirmed the correlation between the non-linearity coefficient B/A of the media and the NLI introduced.

In the current work we present results comparing hydrophone measurements in water and sunflower oil with the results of simulations performed in the k-Wave toolbox (Zhu *et al.*, 1983). Subsequently, an in vivo study of the hepatic NLI was conducted on five volunteers, including one with clinically confirmed fatty liver. The three steps leading to the validation od the proposed method are schematically presented in Fig. 1.

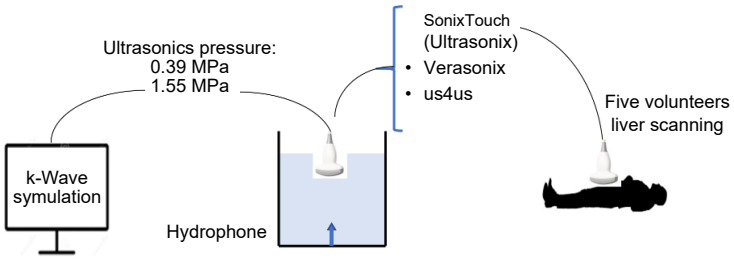


Fig. 1. Schematic diagram of the experimental studies presented in the text.

Throughout our study, a consistent set of excitation signals was employed in both experimental measurements and numerical simulations. We used the 2-cycle tone bursts with a center frequency of 3.3 MHz. The pressure amplitudes $P1$ and $P2$, measured in water at a distance of 1 cm from the transducer surface, were 0.39 MPa and 1.55 MPa, respectively.

2.3. k-Wave *simulations*

Numerical simulations were performed using k-Wave toolbox (Treeby, Cox, 2010), which is designed for modeling acoustic wave fields in complex media. k-Wave solves the full-wave acoustic propagation equations in the time domain, using FFT-based pseudospectral methods to compute spatial derivatives. This approach offers high accuracy and efficiency, especially for media with smoothly varying properties. The toolbox supports modeling of linear and non-linear wave propagation, frequency-dependent absorption, and heterogeneous media, making it well-suited for comparing propagation in different biological and non-biological environments.

In our simulation study, each medium was modeled by assigning appropriate values of sound speed, density, and absorption (Table 1). The acoustic signal was generated by a focused linear array transducer.

Table 1. Acoustic properties typical for water and vegetable oil.

Medium	Speed of sound [m/s]	Density [kg/m^3]	Absorption [dB/cm/MHz]	Alpha	B/A
Water	1500	1000	0.002	2	6
Vegetable oil	1450	920	0.5	1.8	10

The transducer consists of 32 active elements, each with a width of 480.48 µm (12 grid points), separated by a kerf of 80.08 µm (2 grid points). This gives a pitch of 560.56 µm (14 grid points), and an overall width of approximately 17.86 mm (446 grid points). Each element extends 8.008 mm in elevation (200 grid points), and the focus distance is specified as 20 m, effectively modeling a quasi-plane wave within the near-field of the computational domain, as no explicit time delay laws are applied. The transducer is modeled using k-Wave's function for 3D simulations, allowing control over its aperture, element layout, and elevation profile. Sensors are placed throughout the domain to measure the pressure field, including peak pressure, RMS pressure, and time-varying pressure signals, which are useful for evaluating wavefront shape, focal gain, and attenuation. A 2-cycle tone burst is generated and applied uniformly to all active elements. The simulation is run in 3D, with an input grid size of $1998 \times 462 \times 206$ grid points, representing a physical domain of approximately $80\,\text{mm}^3 \times 18.5\,\text{mm}^3 \times 8.25\,\text{mm}^3$. This domain is embedded within a larger computational grid of $2048 \times 512 \times 256$ points to include perfectly matched layers (PML) and ensures stability. The grid spacing is 30 µm, which provides fine spatial resolution and supports a maximum reliable frequency of 17.73 MHz – well above the 3 MHz center frequency of the source, allowing for accurate modeling of waveform distortion and higher harmonic content generated during nonlinear propagation. The time resolution is equally precise. The simulation uses a time step $\Delta t = 8.4592$ µs, with a total simulation time of 59.9924 µs, covering 7093 time steps. This duration allows sufficient time for the wave to propagate through the domain and interact with the surrounding medium or tissue.

Overall, this simulation setup enables high-fidelity modeling of ultrasound wave propagation through homogeneous or heterogeneous media. It captures essential physical phenomena such as attenuation, dispersion, non-linear distortion, and focusing effects. The fine spatial and temporal resolution, combined with a realistic

source and medium configuration, ensures that the model accurately reflects real-world acoustic wave behavior within the tested soft tissue or fluid environments.

2.4. Hydrophone measurements

The experimental measurements were performed using a needle hydrophone – sensor diameter 0.075 mm (Precision Acoustics, Great Britain). The acoustic excitations were generated using the SonixTouch (Ultrasonix-USA) scanner with a convex probe. The hydrophone measurements were carried out in water and sunflower oil. All the acoustic output parameters were exactly the same as in the k-Wave simulation.

2.5. In-vivo measurements

As a preliminary in-vivo validation of the presented approach, we performed the NLI imaging of healthy and fatty livers (of the authors of this article) using the SonixTouch (Ultrasonix) system. The corresponding mechanical indices MI, calculated for each excitation pulse based on the peak negative pressures measured 1 cm below the transducer face, are given in Table 2. For both applied transmitted pressures, the mechanical index MI was significantly lower than the FDA limit for diagnostic ultrasound, MI < 1.9.

Table 2. MI values for pressures used in the experiments.

Peak to peak pressure [MPa]	Peak negative pressure [MPa]	MI
0.39	0.196	0.11
1.55	0.68	0.37

Standard ultrasound images and backscattered raw radio frequency (RF) echoes up to 14 cm were recorded at the frame rate (RF) of 40. The recording of images for two different pressures was performed as follows. The examination was started for a lower pressure $P1$. After the liver was correctly visualized, a rapid change in the acoustic wave pressure was made by quickly changing the position of the power regulator knob on the scanner control panel.

The $E1$ image was selected from the acquired image sequence as the last frame unaffected by the change in power settings. Similarly, $E2$ was chosen as the first frame in which the amplitude stabilized after the power adjustment. It was experimentally found that the driving voltage applied to the transducer stabilized completely after a time corresponding to three image frames, i.e., after 75 ms (for FR = 40). In such a short time, it is reasonable to assume that the position of the examined liver did not change.

The $E1$ and $E2$ images were subjected to smoothing filtration using 5 mm × 5 mm window. The NLI was calculated according to Eq. (1), and the resulting NLI scan is shown in color scale in Fig. 2.

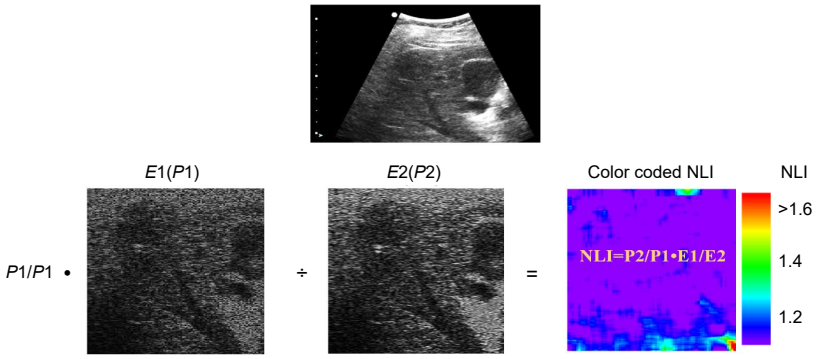


Fig. 2. B-mode image in sector/convex format (top image) is converted to rectangular grid coordinates (bottom images) so that the scan lines are regularly spaced. For each scan, 128 lines are acquired. Both rectangular scans represent the envelope of the raw RF data. The bottom left scan $E1$ is acquired at low transmission pressure $P1$, and after 75 ms the second scan $E2$ (bottom middle) is acquired at high transmission pressure $P2$. NLI color coded scan is shown at bottom right.

3. Results

As stated in the previous section, two experiments were carried out. We measured the first harmonic amplitudes in water and sunflower oil using a needle hydrophone placed at various depths from the face of the convex transducer, from 1 cm to 6 cm. Next the amplitudes of the first harmonic were simulated using k-Wave software. The measurements and simulation were done for two different pressure amplitudes of 0.39 MPa and 1.55 MPa. The results are shown in Fig. 3.

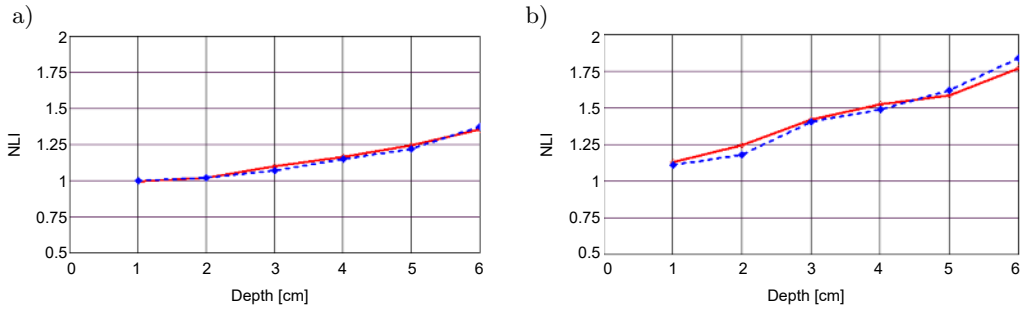


Fig. 3. Plots of NLIs in water (a) and sunflower oil (b), for a pair of acoustic output pressures of 0.39 MPa and 1.55 MPa. Experimental results (solid red lines) are compared with simulation results (dashed blue line).

The NLI scans of four 'healthy' and one fatty liver are shown in Fig. 4. In four imaging studies of 'normal' livers (in Fig. 4a – counting from the top) and one liver with steatosis (Fig. 4b), we determined the mean NLI values and standard deviation (StdDev) at three different depths: 3 cm, 5 cm, and 8 cm counting from the upper

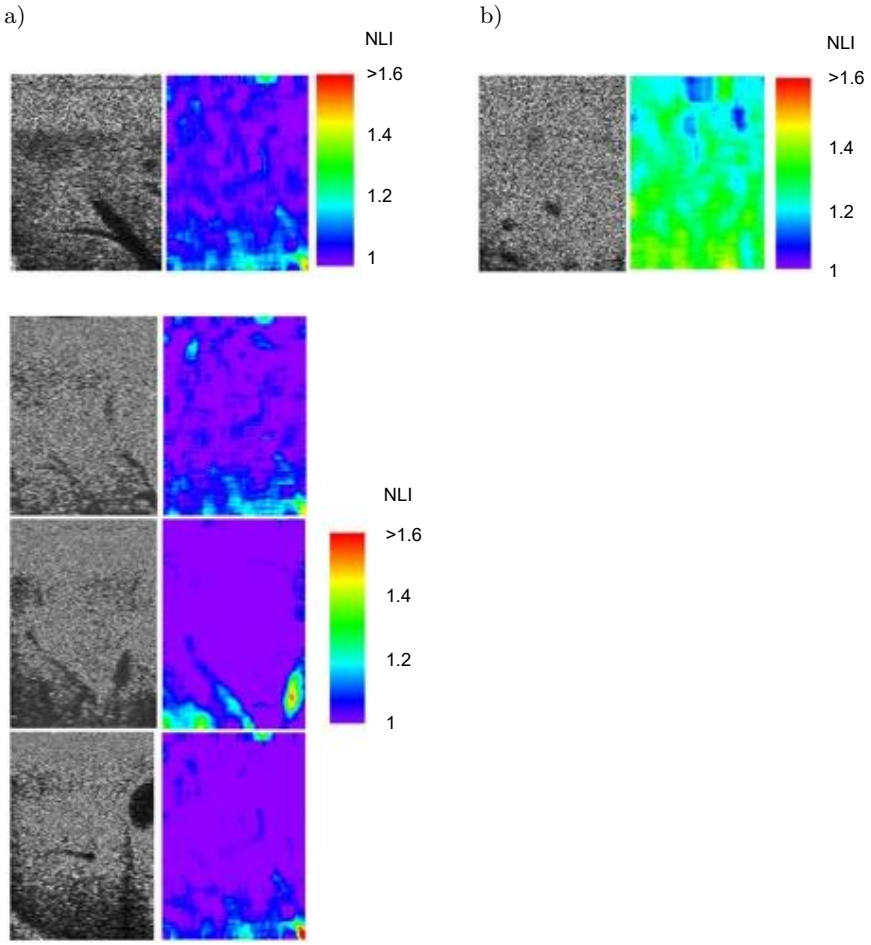


Fig. 4. Preliminary studies on the livers of five authors of the article showed that for 'healthy' livers the NLI was below 1.1 (a), while in one of the authors with previously diagnosed steatosis the NLI locally exceeded 1.3 (b).

border of the liver. The results are presented in Table 3. Both means and StdDev were calculated in 5 mm × 5 mm windows located along the center of the scan.

Table 3. Mean NLI values and standard deviation in scanned livers.

Depth [cm]	Mean StdDev	NLI Liver 1	NLI Liver 2	NLI Liver 3	NLI Liver 4	NLI Liver-steatosis
3	Mean	1.10	1.07	1.09	1.1	1.36
	StdDev	0.03	0.03	0.03	0.03	0.04
5	Mean	1.12	1.09	1.10	1.1	1.4
	StdDev	0.02	0.02	0.025	0.01	0.02
8	Mean	1.11	1.16	1.11	1.15	1.45
	StdDev	0.07	0.03	0.04	0.05	0.03

For healthy volunteers, the mean NLI had values around 1.1, while for the subject with fatty liver it reached 1.45, in agreement with theoretical predictions and simulation data.

4. Discussion

The obtained results confirm the feasibility of the proposed NLI approach for assessing the pressure-dependent nonlinear acoustic response of liver tissue. The method effectively differentiates between healthy and steatotic liver regions using only fundamental frequency information.

Our preliminary results in water, vegetable oil and fatty oil modified beef liver samples presented previously (Zhe *et al.*, 2014) showed clear increase in the NLI in 'fatty' media. The NLI increases faster in oil than in water. This result is in line with expectations, since the non-linearity coefficient B/A is more than 1.5 times higher in oil than in water (water ≈5, vegetable oil >8).

The presented study introduces a novel ultrasound-based NLI for the assessment of pressure-dependent nonlinear properties of liver tissue. We have obtained good convergence between the experimental results and the k-Wave simulations, regarding the amplitude of the first harmonic components of the ultrasound wave propagating in water and sunflower oil.

The simulation results obtained with k-Wave demonstrated the same monotonic dependence of NLI on the nonlinear coefficient (or equivalently B/A), confirming the theoretical expectation that NLI increases with stronger nonlinear behavior.

Preliminary results demonstrate that healthy livers exhibit NLI values around 1.1, while livers with diagnosed steatosis show NLI values exceeding 1.4. This clear distinction highlights the potential of the NLI method as a supportive, non-invasive diagnostic tool for detecting hepatic steatosis. The use of echo signals in the fundamental frequency domain, without reliance on harmonic imaging, further simplifies implementation and adapts well to a variety of clinical ultrasound systems.

Adoption of this approach may significantly enhance the early detection and monitoring of NAFLD, improving patient outcomes and reducing the need for invasive procedures such as liver biopsy.

We are aware that the proposed method has several limitations. Most notably, its outcomes are highly dependent on the imaging depth. Additionally, we estimate that the presence of side-lobes and grating-lobes may significantly affect the results obtained in the NLI mode. These characteristics currently make our method qualitative in nature. In future work, we aim to address these limitations to develop a truly quantitative approach for assessing medium non-linearity. Thanks to the confirmed agreement between the simulations and the measurements, we can support these future efforts with simulations – particularly when attempting acoustic measurements in media with parameters that are difficult to reproduce in laboratory conditions.

In addition to the limitations of the method itself, its validation is constrained by the small sample size – only 5 patients. Consequently, our findings require validation in a considerably larger cohort.

Ongoing and future studies in larger group of patients will be essential to validate the clinical applicability of the pressure-dependent NLI and establish standardized diagnostic thresholds for liver steatosis.

This method opens promising opportunities for quantitative, operator-independent ultrasound diagnostics of liver tissue, potentially setting new standards in non-invasive hepatology.

In summary, the present study should be regarded as a methodological proof-of-concept demonstrating that pressure-dependent analysis of fundamental ultrasound signals can provide reliable insight into the nonlinear behavior of biological tissues.

Further research involving larger patient cohorts and controlled phantom experiments is required to establish calibration between NLI and B/A values and to define diagnostic thresholds for various grades of steatosis.

Appendix

The linear wave equation was derived with three fundamental assumptions:

1. Linearized relationship between pressure and density, $p \cong (\rho - \rho_0)c^2$, where p is acoustic pressure equal to the difference between instantaneous and resting pressure, $p = P - P_0$, ρ_0 is the density of the medium in equilibrium, c is the velocity of longitudinal wave propagation in the medium.
2. Law of conservation of mass $\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\nabla \cdot (\rho v)$, where v is the acoustic particle velocity. This is a nonlinear equation because ρ and v are functions of position variables. The second term is the product of partial velocity and instantaneous density, these are acoustic variables.
3. Euler's equation of motion, $\frac{\mathrm{d}v}{\mathrm{d}t} = -\frac{\nabla p}{\rho}$.

The linearized equation approximates the actual wave propagation only when the speed of sound is much greater than the acoustic particle velocity: $c \gg v$, i.e., when the Mach number M is much smaller than unity, $\mathrm{M} = {}^{v}\!/_{c} \ll 1$. In reality, in the range of applied ultrasonic pressures, i.e., in the range from hundreds of kPascals to MPascals, the tissue volume is compressed and stretched and the speed of sound depends on the pressure. In the equation describing wave propagation, there appears:

1. A nonlinear term describing tissue elasticity (the relationship between pressure and tissue deformation).
2. A nonlinear term related to the deformation of the tissue volume element under the influence of a pressure wave with a strongly distorted / curved wave front. A wave with a finite amplitude propagates slightly faster during positive pressure, when the tissue is compressed and its density increases locally.

Let us assume that the acoustic particle velocity v_0 wave and the wave propagation velocity c have the same direction. If the phases of both waves are the same, then the part of the wave for positive pressure moves faster, with velocity $c_0 + v_0$, while for negative pressure it moves slower, $c_0 - v_0$. For the equation of state $p = f(\rho)$ without linearization, i.e., the assumptions that $p \ll p_0$ and that the changes $\rho - \rho_0$ are negligible, after expanding into a Taylor series, we obtain the expression:

$$p = (\rho - \rho_0)\left(\frac{\partial P}{\partial \rho}\right)_{\rho_0} + \frac{(\rho - \rho_0)^2}{2}\left(\frac{\partial^2 P}{\partial \rho^2}\right)_{\rho_0} + \ldots = A\left(\frac{\rho - \rho_0}{\rho_0}\right) + \frac{B}{2}\left(\frac{\rho - \rho_0}{\rho_0}\right)^2 + \ldots, \tag{2}$$

where $A = \rho_0(\partial P/\partial \rho)_{\rho=\rho_0}$ and $B = \rho_0^2(\partial^2 P/\partial^2 \rho)_{\rho=\rho_0}$.

The ratio B/A is called the non-linearity parameter of the medium. In the linear acoustic regime, where $p \ll P_0$, the changes $\rho - \rho_0$ are very small. Therefore, we can ignore the higher order terms, retaining only $(\rho - \rho_0)$. As a result, we obtain a linear equation:

$$p \approx (\rho - \rho_0)\left(\frac{\partial P}{\partial \rho}\right)_{\rho_0}. \tag{3}$$

The pressure is expressed in Pascals, $1\,\mathrm{Pa} = 1\,\mathrm{N/m^2}$, while the density is expressed in $\mathrm{kg/m^3}$. The quotient of pressure and density has the dimension $(\mathrm{m/s})^2$, i.e., it expresses the velocity in the square, $P/\rho = (\text{velocity})^2$. So,

$$\left(\frac{\partial P}{\partial \rho}\right)_{\rho_0} = c^2. \tag{4}$$

After substituting Eq. (2) describing the nonlinear relationship between pressure and density into Eq. (4) for the wave velocity c_0 and taking into account the convective partial velocity v, we obtain the resultant wave propagation velocity:

$$c = \sqrt{\frac{A}{\rho_0} + \frac{B}{\rho_0}\left(\frac{\rho-\rho_0}{\rho_0}\right) + \ldots} + v = c_0\sqrt{1 + \frac{B}{A}\left(\frac{\rho-\rho_0}{\rho_0}\right) + \ldots} + v. \tag{5}$$

Using the binomial series approximation $(1+x)^n = 1+nx+n(n-1)\frac{x^2}{2!}+\ldots$ we obtain an approximate relationship for the wave velocity in a nonlinear medium:

$$c \cong c_0\left[1 + \frac{B}{2A}\left(\frac{\rho-\rho_0}{\rho_0}\right) + \ldots\right] + v. \tag{6}$$

Taking into account that for a plane wave $\rho c = p/v$ and $p \cong (\rho-\rho_0)\,c^2$, the expression for the velocity in the medium can be written in the form:

$$c \cong c_0 + \frac{B}{2A}v + v = c_0 + \beta v, \tag{7}$$

where

$$\beta = 1 + \frac{B}{2A} \tag{8}$$

is the non-linearity coefficient.

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.

Authors' contributions

A. Nowicki conceptualized the study and wrote the original draft. A. Nowicki, J. Tasinkiewicz, and P. Karwat developed the experimental setup and performed the analysis and contributed to data interpretation. I. Trots and R. Tymkiewicz performed measurements in vitro. N. Żołek and A. Nowicki performed data processing and simulation. All authors reviewed and approved the final manuscript.

References

1. Adler L., Hiedemann E. (1962), Determination of the nonlinearity parameter B/A for water and m-Xylene, *Journal of the Acoustical Society of America*, **34**(4): 410–412, https://doi.org/10.1121/1.1918142.
2. Akiyama I. (2000), Reflection mode measurement of nonlinearity parameter B/A, *AIP Conference Proceedings*, **524**(1): 321–324, https://doi.org/10.1063/1.1309232.
3. Averkiou M.A. (2001), Tissue harmonic ultrasonic imaging, *Comptes Rendus de l'Académie des Sciences – Series IV – Physics*, **2**(8): 1139–1151, https://doi.org/10.1016/S1296-2147(01)01259-8.
4. Averkiou M.A., Roundhill D.R., Powers J.E. (1997), A new imaging technique based on the nonlinear properties of tissues, [in:] *1997 IEEE Ultrasonics Symposium Proceedings. An International Symposium (Cat. No.97CH36118)*, **2**: 1561–1566, https://doi.org/10.1109/ULTSYM.1997.663294.
5. Beye R.T. (1973), Nonlinear acoustics, *American Journal of Physics*, **41**(9): 1060–1067, https://doi.org/10.1119/1.1987473.

6. Bohte A.E., van Werven J.R., Bipat S., Stoker J. (2011), The diagnostic accuracy of US, CT, MRI and ^{1}H-MRS for the evaluation of hepatic steatosis compared with liver biopsy: A meta-analysis, *European Radiology*, **21**: 87–97, https://doi.org/10.1007/s00330-010-1905-5.
7. Byra M. *et al.* (2018), Transfer learning with deep convolutional neural network for liver steatosis assessment in ultrasound images, *International Journal of Computer Assisted Radiology and Surgery*, **13**(12): 1895–1903, https://doi.org/10.1007/s11548-018-1843-2.
8. Byra M., Wojcik J., Nowicki A. (2017), Ultrasound nonlinearity parameter assessment using plane wave imaging, [in:] *2017 IEEE International Ultrasonics Symposium (IUS)*, pp. 1511–1516, https://doi.org/10.1109/ULTSYM.2017.8092733.
9. Coila A., Oelze M.L. (2020), Effects of acoustic nonlinearity on pulse-echo attenuation coefficient estimation from tissue-mimicking phantoms, *The Journal of the Acoustical Society of America*, **148**(2): 805–814, https://doi.org/10.1121/10.0001690.
10. Coila A., Romero A., Oelze M.L., Lavarello R. (2025), Nonlinearity parameter estimation method from fundamental band signal depletion in pulse-echo using a dual-energy model, *The Journal of the Acoustical Society of America*, **157**(3): 1969–1980, https://doi.org/10.1121/10.0036215.
11. Dasarathy S., Dasarathy J., Khiyami A., Joseph R., Lopez R., McCullough A.J. (2009), Validity of real time ultrasound in the diagnosis of hepatic steatosis: A prospective study, *Journal of Hepatology*, **51**(6): 1061–1067, https://doi.org/10.1016/j.jhep.2009.09.001.
12. Dong F., Madsen E.L., MacDonald M.C., Zagzebski J.A. (1999), Nonlinearity parameter for tissue-mimicking materials, *Ultrasound in Medicine & Biology*, **25**(5): 831–838, https://doi.org/10.1016/s0301-5629(99)00016-2.
13. Duck F.A. (2002), Nonlinear acoustics in diagnostic ultrasound, *Ultrasound in Medicine & Biology*, **28**(1): 1–18, https://doi.org/10.1016/s0301-5629(01)00463-x.
14. Errabolu R.V., Sehgal C.M., Bahn R.C., Greenleaf J.F. (1988), Measurement of ultrasonic nonlinear parameter in excised fat tissues, *Ultrasound in Medicine & Biology*, **14**(2): 137–146, https://doi.org/10.1016/0301-5629(88)90181-0.
15. Estes C. *et al.* (2018), Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030, *Journal of Hepatology*, **69**(4): 896–904, https://doi.org/10.1016/j.jhep.2018.05.036.
16. Everbach E.C., Apfel R.E. (1995), An interferometric technique for B/A measurement, *The Journal of the Acoustical Society of America*, **98**(6): 3428–3438, https://doi.org/10.1121/1.413794.
17. Ferraioli G., Kumar V., Ozturk A., Nam K., de Korte Ch.L., Barr R.G. (2022), US attenuation for liver fat quantification: An AIUM-RSNA QIBA pulse-echo quantitative ultrasound initiative, *Radiology*, **302**(3): 495–506, https://doi.org/10.1148/radiol.210736.
18. Fujii Y., Taniguchi N., Akiyam I., Tsao J., Itoh K. (2004), A new system for *in vivo* assessment of the degree of nonlinear generation using the second harmonic component in echo signals, *Ultrasound in Medicine & Biology*, **30**(11): 1511–1516, https://doi.org/10.1016/j.ultrasmedbio.2004.08.016.
19. Gong X., Zhang D., Liu J., Wang H., Yan Y., Xu X. (2004), Study of acoustic nonlinearity parameter imaging methods in reflection mode for biological tissues, *The Journal of the Acoustical Society of America*, **116**(3): 1819–1825, https://doi.org/10.1121/1.1781709.
20. Gong X.F., Liu X.Z., Zhang D. (1993), Influences of tissue composition and structural features of biological media on the ultrasonic nonlinearity parameter, *Chinese Journal of Acoustics*, **12**(3): 265–270.
21. Hernaez R. *et al.* (2011), Diagnostic accuracy and reliability of ultrasonography for the detection of fatty liver: A meta-analysis, *Hepatology*, **54**(3): 1082–1090, https://doi.org/10.1002/hep.24452.
22. Hsu P.-K. *et al.* (2021), Attenuation imaging with ultrasound as a novel evaluation method for liver steatosis, *Journal of Clinical Medicine*, **10**(5): 965, https://doi.org/10.3390/jcm10050965.
23. Hunter C. *et al.* (2016), An ultrasonic caliper device for measuring acoustic nonlinearity, *Physics Procedia*, **87**: 93–98, https://doi.org/10.1016/j.phpro.2016.12.015.
24. Jesper D. *et al.* (2020), Ultrasound-based attenuation imaging for the non-invasive quantification of liver fat – A pilot study on feasibility and inter-observer variability, *IEEE Journal of Translational Engineering in Health and Medicine*, **8**: 1800409, https://doi.org/10.1109/JTEHM.2020.3001488.

25. Law W.K., Frizzell L.A., Dunn F. (1985), Determination of the nonlinearity parameter B/A of biological media, *Ultrasound in Medicine & Biology*, **11**(2): 307–318, https://doi.org/10.1016/0301-5629(85)90130-9.

26. Lu Z., Daridon J.L., Lagourette B., Ye S. (1998), A phase-comparison method for measurement of the acoustic nonlinearity parameter B/A, *Measurement Science and Technology*, **9**: 1699–1705, https://doi.org/10.1088/0957-0233/9/10/009.

27. Madigosky W.M., Rosenbaum I., Lucas R. (1981), Sound velocities and B/A in fluorocarbon fluids and in several low density solids, *The Journal of the Acoustical Society of America*, **69**(6): 1639–1643, https://doi.org/10.1121/1.385941.

28. Nikoonahad M., Liu D.C. (1989), Pulse-echo B/A measurement using variable amplitude excitation, [in:] *Proceedings of the 1989 IEEE Ultrasonics Symposium*, pp. 1047–1051.

29. Nowicki A., Tasinkiewicz J., Karwat P., Trots I., Zołek N., Tymkiewicz R. (2024), Ultrasound imaging of nonlinear media response using a pressure-dependent nonlinearity index, *Archives of Acoustics*, **49**(4): 557–563, https://doi.org/10.24425/aoa.2024.148814.

30. Nowicki A., Wójcik J., Secomski W. (2007), Harmonic imaging using multitone nonlinear coding, *Ultrasound in Medicine & Biology*, **33**(7): 1112–1122, https://doi.org/10.1016/j.ultrasmedbio.2007.02.001.

31. Ogino Y., Wakui N., Nagai H., Igarashi Y. (2021), The ultrasound-guided attenuation parameter is useful in quantification of hepatic steatosis in non-alcoholic fatty liver disease, *JGH Open*, **5**: 947–952, https://doi.org/10.1002/jgh3.12615.

32. Panfilova A., van Sloun R.J.G., Wijkstra H., Sapozhnikov O.A., Mischi M. (2021), A review on B/A measurement methods with a clinical perspective, *The Journal of the Acoustical Society of America*, **149**(4): 2200–2237, https://doi.org/10.1121/10.0003627.

33. Parker K.J., Asztely M.S., Lerner R.M., Schenkh E.A., Waag R.C. (1988), In-vivo measurements of ultrasound attenuation in normal or diseased liver, *Ultrasound in Medicine & Biology*, **14**(2): 127–136, https://doi.org/10.1016/0301-5629(88)90180-9.

34. Saadeh S. *et al.* (2002), The utility of radiological imaging in nonalcoholic fatty liver disease, *Gastroenterology*, **123**(3): 745–50, https://doi.org/10.1053/gast.2002.35354.

35. Sarvazyan A.P., Chalikian T.V., Dunn F. (1990), Acoustic nonlinearity parameter B/A of aqueous solutions of some amino acids and proteins, *The Journal of the Acoustical Society of America*, **88**(3): 1555–1561, https://doi.org/10.1121/1.400314.

36. Sehgal C.M., Brown G.M., Bahn R.C., Greenleaf J.F. (1986), Measurement and use of acoustic nonlinearity and sound speed to estimate composition of excised livers, *Ultrasound in Medicine & Biology*, **12**(11): 865–874, https://doi.org/10.1016/0301-5629(86)90004-9.

37. Simpson D.H., Chien T.C., Burns P.N. (1999), Pulse inversion Doppler: A new method for detecting nonlinear echoes from microbubble contrast agents, *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, **46**(2): 372–382, https://doi.org/10.1109/58.753026.

38. Tada T. *et al.* (2019), Usefulness of attenuation imaging with an ultrasound scanner for the evaluation of hepatic steatosis, *Ultrasound in Medicine & Biology*, **45**(10): 2679–2687, https://doi.org/10.1016/j.ultrasmedbio.2019.05.033.

39. Takahashi S. (1995), Measurement of acoustic nonlinearity parameter by observation waveforms, *Japanese Journal of Applied Physics*, **34**: 2790–2792, https://doi.org/10.1143/JJAP.34.2790.

40. Treeby B.E., Cox B.T. (2010), k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave-fields, *Journal of Biomedical Optics*, **15**(2): 021314, https://doi.org/10.1117/1.3360308.

41. van Wijk M.C., Thijssen J.M. (2002), Performance testing of medical ultrasound equipment: Fundamental vs. harmonic mode, *Ultrasonics*, **40**(1–8): 585–591, https://doi.org/10.1016/s0041-624x(02)00177-4.

42. Varray F., Cachard C., Tortoli P., Basset O. (2010), Nonlinear radio frequency image simulation for harmonic imaging: CREANUIS, [in:] *2010 IEEE International Ultrasonics Symposium*, pp. 2179–2182, https://doi.org/10.1109/ULTSYM.2010.5935538.

43. Varray F., Chenot J., Basset O., Tortoli P., Melodelima D., Cachard C. (2011), Nonlinear parameter imaging to characterize HIFU ablation: Preliminary in vitro results in porcine liver, [in:] *Proceedings of the 2011 IEEE International Ultrasonics Symposium*, pp. 1361–1363, https://doi.org/10.1109/ULTSYM.2011.0336.

44. Wallace K., Lloyd C., Holland M., Miller J.G. (2007), Finite amplitude measurements of the nonlinear parameter B/A for liquid mixtures spanning a range relevant to tissue harmonic mode, *Ultrasound in Medicine & Biology*, **33**(4): 620–629, https://doi.org/10.1016/j.ultrasmedbio.2006.10.008.

45. Yoo J. *et al.* (2020), Reproducibility of ultrasound attenuation imaging for the noninvasive evaluation of hepatic steatosis, *Ultrasonography*, **39**(2): 121–129, https://doi.org/10.14366/usg.19034.

46. Younossi Z.M., Diehl A.M., Ong J.P. (2002), Nonalcoholic fatty liver disease: An agenda for clinical research, *Hepatology*, **35**(4): 746–752, https://doi.org/10.1053/jhep.2002.32483.

47. Younossi Z.M., Koenig A.B., Abdelatif D., Fazel Y., Henry L., Wymer M. (2016), Global epidemiology of nonalcoholic fatty liver disease: Meta-analytic assessment of prevalence, incidence, and outcomes, *Hepatology*, **64**(1): 73–84, https://doi.org/10.1002/hep.28431.

48. Zhang D., Gong X. (1999), Experimental investigation of the acoustic nonlinearity parameter tomography for excised pathological biological tissues, *Ultrasound in Medicine & Biology*, **25**(4): 593–599, https://doi.org/10.1016/s0301-5629(98)00185-9.

49. Zhang J., Dunn F. (1991), A small volume thermodynamic system for B/A measurement, *The Journal of the Acoustical Society of America*, **89**(1): 73–79, https://doi.org/10.1121/1.400370.

50. Zhe Z., Gong C., Dong Z. (2014), Molecular structure dependence of acoustic nonlinearity parameter B/A for silicone oils, *Chinese Physics B*, **23**(5): 054302, https://doi.org/10.1088/1674-1056/23/5/054302.

51. Zhu Z., Roos M., Cobb W.N., Jensen K. (1983), Determination of the acoustic nonlinearity parameter B/A from phase measurements, *The Journal of the Acoustical Society of America*, **74**(5): 1518–1521, https://doi.org/10.1121/1.390154.

Archives of Acoustics Vol. 51, No. 1, pp. 107–130 (2026), https://doi.org/10.24423/archacoust.2026.4369

Research Paper

An Empirical Approach to Investigate Environmental Effects on Acoustic Signal Speed in Oceanic Layers

Khondoker Munim Salehin, Md. Khalid Mahbub Khan*, Anisur Rahman

Department of Computer Science and Engineering, East West University
Dhaka, Bangladesh

*Corresponding Author: khalidmahbub.khan@yahoo.com

Received November 2, 2025; revised January 3, 2026; accepted January 4, 2026; available online January 26, 2026; version of record March 17, 2026; published issue March 27, 2026.

This paper investigates and demonstrates the effects of three significant environmental contributors: temperature, depth and salinity impact on the acoustic signal propagation across distinctive ocean layers: mixed, thermocline, and deep layers. In the field of underwater wireless sensor networks (UWSN), exact and precise determination of coordinates for sensor localization is very crucial for data validation. Temperature dominates the upper layers; depth becomes the prime factor for the deeper domain with minimal thermal variations. Salinity while having a diminished effect, facilitates minor changes in propagation and deviation of acoustic signal speed. In our work we have analyzed these interdependencies by using different empirical models (e.g., Mckenzie, Medwin) customized to each layer, accounting to their incomparable environmental parameters. In mixed layers, changes in sound speed are mainly caused by thermal factors, where depth is of minimal importance, and the influence of salinity is insignificant, but with increasing depth, the temperature begins to decrease, and depth (pressure) begins to become important, and changes in salinity and temperature become almost equivalent. By evaluating ocean layer specified empirical formulas, we have calculated the average speed of sound and measure the corresponding contribution of all parameters. Our work has provided a substructure which helps to optimize the identification or localization of UWSN nodes. The results of this work underscored the essential to have an adaptive sound speed modeling in order to achieve enhanced and precise acoustic signal communication systems.

Keywords: acoustic signal speed, ocean layers, salinity, empirical formulas, sound speed modeling.

1. Introduction

The underwater sensor network has been a pivotal field of scientific research for decades. The vast expansion of ocean has brought both known and yet-to-be known exploration and monitoring challenges to us. At present marine monitoring is increasing promptly. As it is more difficult to establish and monitor an underwater sensor network than a terrestrial network because of its positing in the rough environments. It is important to gather marine precise information of underwater location as it helps with underwater surveillance, ocean life exploration, natural disaster study and so on (Tan *et al.*, 2011). The understanding of the accurate node location in the underwater network is very important for both tracking and validating collected data. Electromagnetic radio waves have very poor performance underwater due to sea water because it also faces a high attenuation which makes it expensive and also affects the propagation for long ranges (KukuChuku *et al.*, 2018). As radio signal has a low propagation range, we generally use an acoustic signal as a substitute (Paul *et al.*, 2020). So, in underwater communication and other necessities such as distance measurements we usually use acoustic signals. The history of underwater velocity determination was started in the early 19th century, when scientists

used a tube to take measurements underwater which was suggested by the famous artist Da Vinci and scientists recorded the speed of the submerged bell proceeding across Lake Geneva (Scowcroft *et al.*, 2015). Another invention in the field of underwater communication was the invention of 'Gertude' or we could say a marine telephone that uses analog modulation and its carrier frequency was between 2 kHz to 15 kHz (Stojanovic, 2007). So, numerous researches have scrutinized diverse perspective of UWSN, furthering in the understanding and utilization of underwater communication with an array of different ranges and fields for AUV (autonomous underwater vehicles) and regardless of the type of positioning (submerged, indoor, outdoor, underneath). Many studies have proclaimed that environmental factors can bring variation to the acoustic signal speed. According to Chen and Millero (1997), the speed of sound has been affected by temperature depth and salinity. The scope of this article is to understand and analyze environmental variable relationship with an acoustic signal as it is the core of underwater communication. This paper highlights the process of sound speed calculation and how in the ocean each layer is governed by a dominating ecological parameter. This work shows an empirical calculation using the values of different indicator available or introduced by many prominent researchers. Our work can be useful in fields of UWSN localization, ocean engineering and also for understanding underwater signal processing. In the article also the sound profiles of different empirical formulas are implemented. The major contributions of this article are arranged as follows:

1. This paper demonstrates how an average speed of sound is calculated underwater using divergent empirical formulas.
2. The analysis of the effect of temperature, depth and salinity for different oceanic layers is exhibited to observe ecological variable impact domination for different coverings.
3. Finally, the evaluation of empirical formulas' behaviors of the environmental parameter for different oceanic layers is also performed in this research.

The remaining of the paper has been organized in different sections structurally. Section 2 which is divided into two parts: the problem field and review of the previous research work. In the problem field we have evaluated the issues that arose during the measurement of the acoustic speed velocity for UWSN and included a network architecture for a visualization purpose and also the way to examine the effects of environmental parameters on the speed of sound for marine communication using empirical formulas. Section 3 has shown a method of the work, and also including mathematical equation along with additional acronym and indicators. In Sec. 4 all the simulation works are shown with detail information. Section 5 offers a rivaling discussion and finally, in Sec. 6, the article concludes.

2. Background

2.1. Problem domain

In our proposed UWSN structure in Fig. 1, we consider a single beacon on the top of the water surface column and four submerged sensors are deployed underwater that require distance measurement. The sensor nodes can be considered in any layer, provided that the components meet the constrains from empirical formulas. For accurate measurement of acoustic speed or speed of sound we required the values of temperature, depth and salinity from both beacon and sensor region. Determining these variables for beacon is uncomplicated as temperature and salinity of the surface water traditionally determined at nil depth. However, if we do not integrate specific sensors (for example, temperature, depth, and brininess sensors) for deployed sensors it becomes very much difficult to grasp those variables accurately. In this paper, our focus is to investigate the effect of these three environmental variables on the acoustic velocity for three different ocean layers such as mixed layer, thermocline layer and deep layer using different empirical equations. For ease of comprehension, we think that all the deployed sensor are stationary at the moment of implementation time calculation. We have neglected the motility of the sensor nodes because our main concern here is to measure the average acoustic velocity and analyze the effect of environmental variables using various empirical equations. Our work uses the value of three underwater layer (mixed, thermocline, and deep layer) ranges to make our desired simulation.

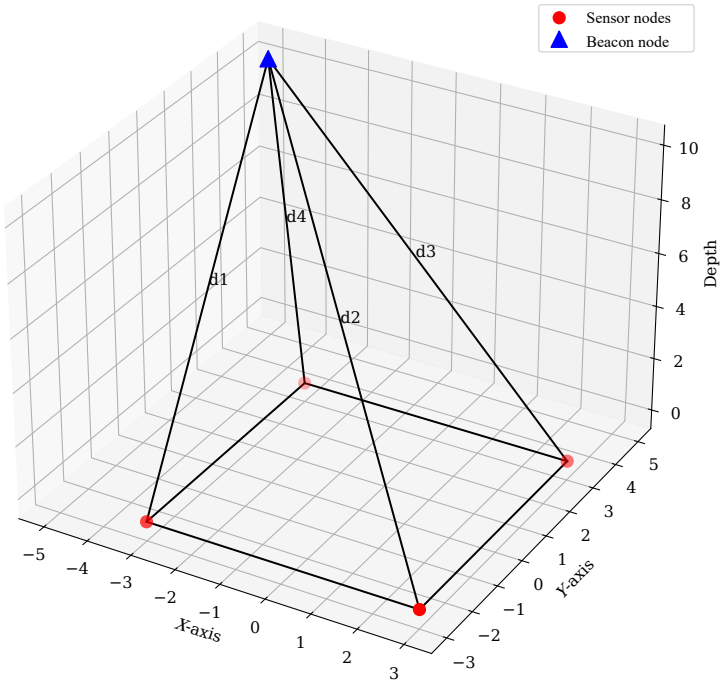


Fig. 1. Underwater network deployment considering one beacon and multiple submerged sensors.

2.2. Literature review

Numerous researches have scrutinized diverse perspective of UWSN, furthering in the understanding and utilization of empirical equations throughout an array of different ranges and fields. Multitudinous empirical formula exists that are established throughout the years to determine the acoustic velocity on the base of oceanographic variable like temperature, depth and salinity. Rahman *et al.* (2013a) has analyzed the ecological variables to see which parameters have high influence on an acoustic signal velocity for a vertical water column for a specific problem field, they considered one beacon on the top of the water surface and few submerged sensors which were deployed and needed to measure their distance. They used the Mackenzie equation and the triple integral method to measure the average speed of sound to analyze which environmental parameter among temperature depth and salinity has the greatest impact on the speed of sound for the considered configuration of UWSN with a single beacon and few submerged sensors. In this work, only one formula with a specific range was visited to investigate the influence of environmental elements. Another pivot research of Rahman (2014) has proposed an innovative manner to resolve sensor localization using the measurement of in-situ acoustic velocity with the Mackenzie formula and utilizing as minimal, as a single beacon. The coordinates of the mathematical system were solved using Cayley–Menger determinant followed by linearization and solving nonlinear equations, as it is considered that no node has information of the other node position after that for the distance measurement, he considered both radio and acoustic signals but as we all know radio signal has a minimal propagation range underwater and he used this to only synchronize the clock between beacon and deployed sensors. He has considered the surface plane as parallel but also shown a non-parallel state of subsets for both configurations, he has considered a minimal single beacon which is on the top of the water surface and submerged sensors that are installed underwater. Correspondingly, Talib *et al.* (2011) has determined the value of speed of sound using a different empirical equation such as Del Grosso, Mackenzie, and Medwin as the efficacy of speed of sound is highly sensitive to temperature, density and salinity. They talk about the on-site monitoring of speed of sound for various types of water categories such as sea water, fresh water and inlet. Their work could be very helpful for hydrographer which will save time while electing the appropriate speed of sound in adjusting echo sound equipment and this observation will also be valid in countries with moderate atmospheric variations. Huang *et al.* (2024) investigated the field of an underwater sound speed profile or in short SSP as synchronous and precise

establishment of zonal SSP plays a significant role in marine positioning, navigation and timing (PNT) systems as it appreciably influenced the signal propagation manner for an instance trajectory. There are generally two methods for construction of SSP, one is the direct SSP measurement another one is SSP inversion. For the direct SSP measurement they have used some efficiently functional empirical formulas, such as Wilson, Leroy, Medwin, Del Grosso, etc. These methods have helped promoting the rapid advancement in the field of underwater sensing apparatus.

3. Methodology

3.1. System overview and workflow

To deliver consistent insights of the research procedure, Fig. 2 demonstrated the general workflow applied in this study. Firstly, we have considered three oceanic layers named mixed layer, thermocline layer, and deep layer. The first layer of ocean is mixed layer or we can call it surface layer, it has a direct connection with atmosphere and it also has almost homogeneous vertical qualities in terms of temperature, depth, and salinity (Gill, 2016). The second layer thermocline layer is the middle layer between warmer seawater and deeper cold water in other words we can state that stratified, unstable water forms a thermocline layer, which develops frequent temperature drops (Lana *et al.*, 2017). This applies to oceans worldwide that are deeper than 200 m and also it is noted that the large area of the marine habitat can have their own general understanding, either negatively or positively depending on the circumstances (Jamieson *et al.*, 2025). For each layer we have considered a range for the main three environmental parameters: temperature, depth, and salinity. Now, for the calculation of velocity of sound underwater we need to use formulas which are incorporated with these parameters and calculate the velocity of a vertical water column, and for that we have considered empirical formulas, such as Mackenzie, Medwin, Wilson, Leroy, etc. After calculating the speed value of vertical column we use the triple integral method to find the value of the average sound of the speed. Now we have analyzed the effect of temperature, depth, and salinity and for that we have considered some fixed values of these indicators and examined the change of the average speed for each layer for different empirical formulas. For the temperature change we have considered the fixed range surface temperature, for depth we considered a specific range of depth across three oceanic layers and for salinity for three layers we considered a range where we calculate the average speed for a slight variation of salinity.

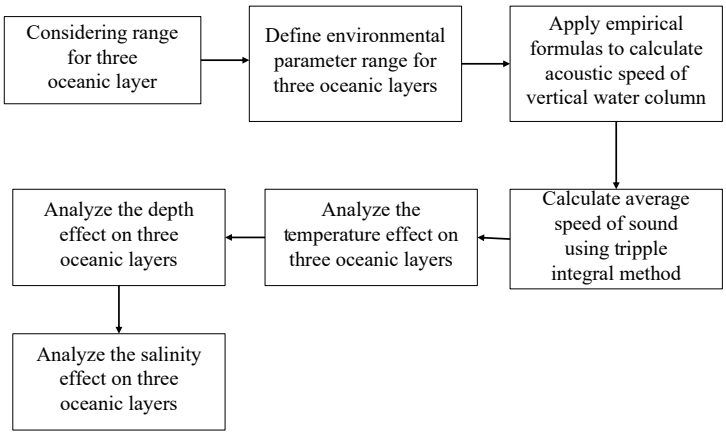


Fig. 2. Overview of the research workflow.

3.2. Analytical framework

In order to determine the distance from, the beacon to sensor position traditionally we consider the speed of sound and travel time and the equation becomes (Talib *et al.*, 2011):

$$\text{Distance}(d) = \frac{1}{2} \times \text{Speed of sound}(v) \times \text{Travel time}(t). \tag{1}$$

In general, the acoustic signal is used instead of radio signal as it has the lowest propagation range than the acoustic signal (Rahman *et al.*, 2013a) the speed of sound or in this case the acoustic wave near the sea plane is considered about 1500 m/s which is four times faster than in air (Rahman, 2014). Nevertheless, the effect of environmental variable on the acoustic speed cannot be ignored. To calculate the sound speed velocity underwater we need to inspect the accurate value of temperature, depth, and salinity and for that we cannot use traditional formulas. So, to calculate the acoustic velocity we can use empirical formulas which take these three environmental values into account in the calculations. They are as follows:

1. Mackenzie empirical formula (Mackenzie, 1981):

$$\begin{aligned} v_m = {} & 1448.96 + 4.591T - 5.304 \times 10^{-2}T^2 + 2.374 \times 10^{-4}T^3 \\ & + 1.34(S-35) + 1.63 \times 10^{-2}D + 1.675 \times 10^{-7}D^2 \\ & - 1.025 \times 10^{-2}T(S-35) - 7.139 \times 10^{-13}TD^3, \end{aligned} \tag{2}$$

where v_m is the speed of the Mackenzie formula under consideration, T stands for temperature, and D stands for depth.

2. Medwin empirical formula (Medwin, 1975):

$$v_m d = 1449.2 + 4.6T - 0.055T^2 + 0.00029T^3 + (1.34 - 0.01T)(S-35) + 0.016D, \tag{3}$$

where $v_m d$ is the speed of the Medwin formula under consideration.

3. Chen–Millero empirical formula (National Physical Laboratory [NPL], n.d.):

$$v_C = C_1(T,P) + C_2(T,P) \cdot S + C_3(T,P) \cdot S^{3/2} + C_4(T,P) \cdot S^2, \tag{4}$$

where

$$\begin{aligned} C_1(T,P) = {} & 1402.388 + 5.0383T - 0.058109T^2 + 0.00033432T^3 - 1.47797 \times 10^{-6}T^4 + 3.1419 \times 10^{-9}T^5 \\ & + (0.153563 + 0.00068999T - 8.1829 \times 10^{-6}T^2 + 1.3632 \times 10^{-7}T^3 - 6.126 \times 10^{-10}T^4)P \\ & + (3.126 \times 10^{-5} - 1.7111 \times 10^{-6}T + 2.5986 \times 10^{-8}T^2 - 2.5353 \times 10^{-10}T^3 + 1.0415 \times 10^{-12}T^4)P^2 \\ & + (-9.7729 \times 10^{-9} + 3.8513 \times 10^{-10}T - 2.3654 \times 10^{-12}T^2)P^3, \end{aligned}$$

$$\begin{aligned} C_2(T,P) = {} & 1.389 - 0.01262T + 7.166 \times 10^{-5}T^2 + 2.008 \times 10^{-6}T^3 - 3.21 \times 10^{-8}T^4 \\ & + (9.4742 \times 10^{-5} - 1.2583 \times 10^{-5}T - 6.4928 \times 10^{-8}T^2 + 1.0515 \times 10^{-8}T^3 - 2.0142 \times 10^{-10}T^4)P \\ & + (-3.9064 \times 10^{-7} + 9.1061 \times 10^{-9}T - 1.6009 \times 10^{-10}T^2 + 7.994 \times 10^{-12}T^3)P^2 \\ & + (1.1 \times 10^{-10} + 6.651 \times 10^{-12}T - 3.391 \times 10^{-13}T^2)P^3, \end{aligned}$$

$$C_3(T,P) = -0.01922 - 4.42 \times 10^{-5}T + (7.3637 \times 10^{-5} + 1.795 \times 10^{-7}T)P,$$

$$C_4(T,P) = 0.001727 - 7.9836 \times 10^{-6}P,$$

where v_c is the speed of thee Chen–Millero formula, P stands for pressure, and S denotes for salinity. In the formula under consideration C_1 to C_4 are its coefficients.

4. Wilson empirical formula (Wilson, 1977):

$$v_W = 1449.14 + v_1 + v_2 + v_3 + v_{tpS}, \tag{5}$$

where

$$v_1 = 4.5721T - 4.4532 \times 10^{-2}T^2 - 2.604 \times 10^{-4}T^3 + 7.9851 \times 10^{-6}T^4,$$

$$v_2 = 1.60272 \times 10^{-1} P + 1.0268 \times 10^{-5} P^2 + 3.5216 \times 10^{-9} P^3 - 3.3603 \times 10^{-12} P^4,$$

$$v_3 = 1.39799(S-35) + 1.69202 \times 10^{-3}(S-35)^2,$$

$$\begin{aligned} v_{tpS} = {} & (S-35)\big(-1.1244 \times 10^{-2} T + 7.7711 \times 10^{-7} T^2 + 7.7016 \times 10^{-5} P - 1.2943 \times 10^{-7} P^2 \\ & + 3.158 \times 10^{-8} PT + 1.579 \times 10^{-9} PT^2\big) \\ & + P\big(-1.8607 \times 10^{-4} T + 7.4812 \times 10^{-6} T^2 + 4.5283 \times 10^{-8} T^3\big) \\ & + P^2\big(-2.5294 \times 10^{-7} T + 1.8563 \times 10^{-9} T^2\big) + P^3(-1.9646 \times 10^{-10} T), \end{aligned}$$

where v_w is the speed of the Wilson formula. In the formula under consideration v_1, v_2, and v_{tpS} are its coefficients.

5. Leroy empirical formula (Leroy, 1969):

$$v_L = 1492.9 + 3(T-10) - 0.006(T-10)^2 - 0.04(T-18)^2 + 1.2(S-10) - 0.01(S-35)(T-18) + D/61, \tag{6}$$

where v_L is the speed of the Leroy formula.

6. Coppens empirical formula (Coppens, 1981):

$$v_{Co} = v_0 + (16.23 + 0.0253T) \cdot 0.001Z + (0.213 - 0.01T) \cdot 0.000001Z^2, \tag{7}$$

$$v_0 = 1449.05 + 4.57T - 0.0521T^2 + 0.00023T^3 + (1.333 - 0.0126T + 0.00009T^2)(S-35).$$

7. Del Grosso empirical formula (NPL, n.d.):

$$v_D = C_{000} + \Delta C_T + \Delta C_S + \Delta C_P + \Delta C_{STP}, \tag{8}$$

where

$$C_{000} = 1402.392,$$

$$\Delta C_T = 5.012285T - 0.0551184T^2 + 0.000221649T^3,$$

$$\Delta C_S = 13.2953S + 0.0001288598S^2,$$

$$\Delta C_P = 0.1560592P + 0.0002449993P^2 - 8.833959 \times 10^{-8} P^3,$$

$$\begin{aligned} \Delta C_{STP} = {} & 0.006353509TP - 4.383615 \times 10^{-7} T^3 P - 0.00001593895TP^2 + 2.656174 \times 10^{-8} T^2 P^2, \\ & + 5.222483 \times 10^{-10} TP^3 - 0.01275936ST + 9.688441 \times 10^{-5} ST^2 - 0.0003406824STP, \\ & + 4.857614 \times 10^{-6} S^2 TP - 1.616745 \times 10^{-8} S^2 P^2, \end{aligned}$$

where v_D is the speed of the Del Grosso formula, and C_{000} to C_T, C_P, C_{STP} are its coefficients.

Each equation has different types of range for the three environmental parameters. In this work our desired environmental parameter unit for temperature is Celsius, for salinity it is p.s.u, and for depth it is meter but as discussed in (Huang *et al.*, 2024) for Eq. (5) and Eq. (7) the depth unit is 1000 kg/cm^3 and Eq. (3) the unit is 1000 bar. To determine the velocity of vertical water column and also to measure the average speed of sound it is needed to convert them to our desired parameter. To convert the depth value into meters we use the formula from (NPL, n.d.), to convert pressure to depth in meters:

$$Z_s(P,\Phi) = \frac{9.72659 \times 10^2 P - 2.512 \times 10^{-1} P^2 + 2.279 \times 10^{-4} P^3 - 1.82 \times 10^{-7} P^4}{g(\Phi) + 1.092 \times 10^{-4} P}, \tag{9}$$

where $g(\Phi)$ (variation of gravity with latitude) is considered 1000 m/s^2 and P is around 9.81×10^6 n/m^2, and as the equation with these values provided a yield value which do not align with expected physical depths,

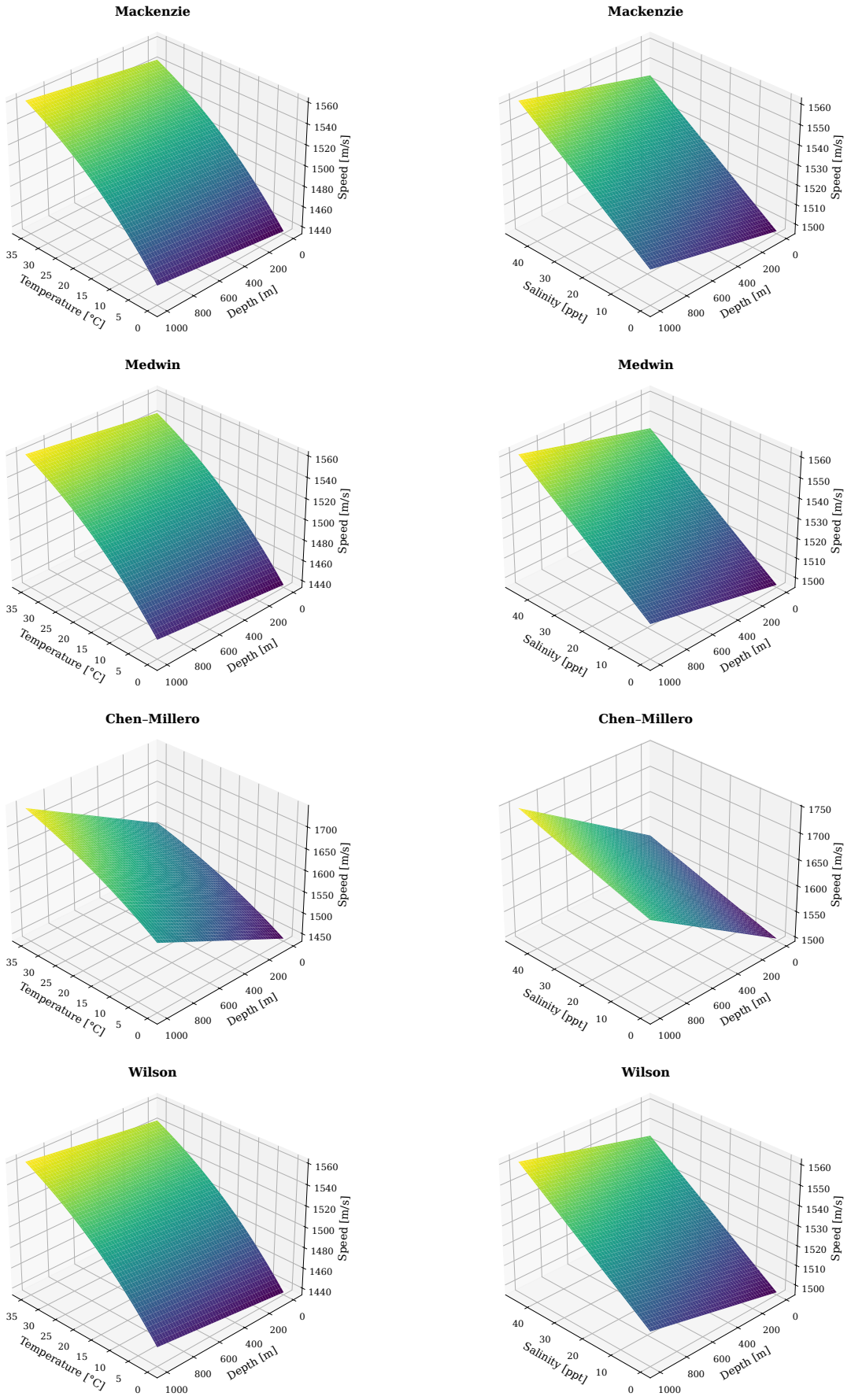


Fig. 3. Acoustic profile for Mackenzie, Medwin, Chen–Millero, and Wilson (temperature/salinity vs. depth).

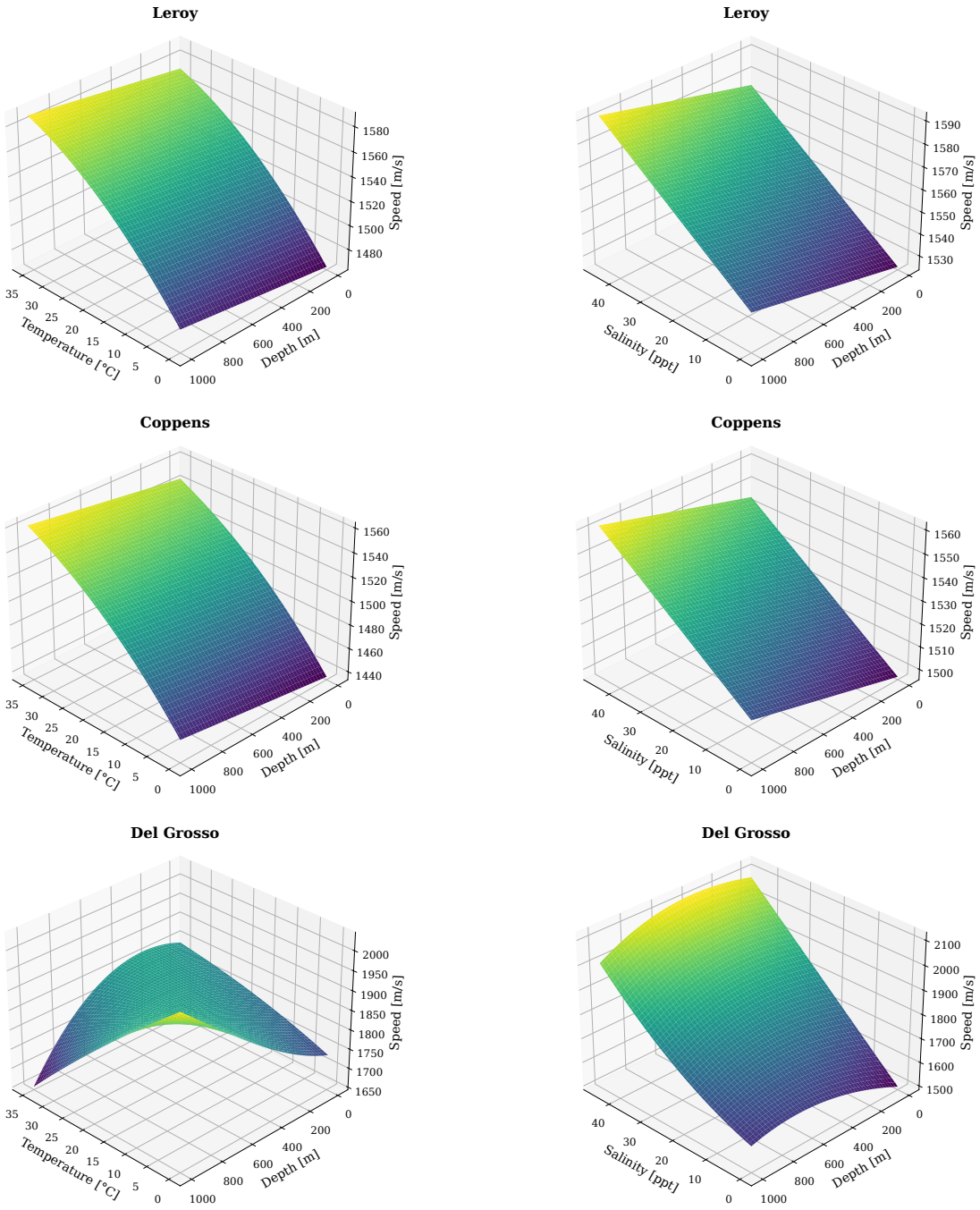


Fig. 4. Acoustic profile for Leroy, Coppens, and Del Grosso (temperature/salinity vs. depth).

we have used the scaling factor of 0.08593 which gives the depth value as 8000 m for 1000 kg/cm^3 and for 1000 bar the scaling factor is considered 0.008618 which give us a depth of same 8000 m. In Table 1, we can see the comparison of different empirical formula range for temperature, depth and salinity.

After determining and transferring the numeric values of temperature, depth and salinity with the top beacon which has all the information from top to the bottom region, we can calculate the average speed of an acoustic signal by Rahaman *et al.* (2013b) using:

$$V\text{avg} = f_{\text{avg}}(T,D,S) = \frac{1}{A}\int\int\int_{R} f(T,D,S) = \frac{1}{A}\int_{S_i}^{S_f}\int_{D_i}^{D_f}\int_{T_i}^{T_f} f_{\text{avg}}(T,D,S)\,\mathrm{d}T\,\mathrm{d}D\,\mathrm{d}S. \tag{10}$$

Table 1. Comparison of different empirical formulas for speed of sound.

Empirical formula	Proposed year	Applicable range		
		Temperature [°C]	Depth [m]	Salinity (p.s.u)
Mackenzie (Mackenzie, 1981)	1981	2 to 30	0 to 8000	25 to 40
Medwin (Medwin, 1975)	1975	0 to 30	0 to 1000	0 to 40
Chen–Millero (NPL, n.d.)	1980	0 to 40	0 to 8000	5 to 40
Wilson (Wilson, 1977)	1960	0 to 30	0 to 1000	0 to 37
Leroy (Leroy, 1969)	1969	−2 to 40	0 to 1000	0 to 42
Coppens (Coppens, 1981)	1981	0 to 35	0 to 4000	0 to 45
Del Grosso (NPL, n.d.)	1974	0 to 30	0 to 8000	30 to 40

Considering the ranges of all empirical formulas from Table 1, we can initialize the upper and lower ranges for all empirical formulas and plot there acoustic profile using the Python simulation environment and Fig. 3 and Fig. 4 have demonstrated that.

In this work, A considers the area produced by the limits of T, D, S; $f(T, D, S)$ which is for all empirical equation the same as in Eq. (2) to Eq. (8). The right sides of Fig. 3 and Fig. 4 represent the acoustic profile for salinity and depth where temperature is constant and in the left side acoustic profile is for temperature and depth where salinity is constant. So, by the help of the acoustic profile we can visualize the three environment parameters.

4. Simulation results

4.1. Environmental constrains

In this work, the entire underwater environment has been taken into account in three factors, which are temperature, depth, and salinity, and these variables fluctuated depending on different layers such as mixed layer, thermocline layer, and deep layer. According to Kundu (2016), each ocean layer has its own range for environmental parameters and it highly impacts the ecological paramets which are listed in Table 2.

Table 2. Environmental variable range for different layers.

Ocean layers	Depth [m]	Temperature [°C]	Salinity (p.s.u)
Mixed	0 to 450	30 to 25.5	25 to 25.045
Thermocline	450 to 1000	10 to 4	34 to 34.5
Deep	1000 to8000	4 to 2	34.8 to 35.1

To investigate the effect of environmental parameters on an acoustic speed, we have simulated different empirical models which align with our ranges of sea layers given in Table 2 using the Python enviroment. The simulation was conducted by applying Python 3.12 utilizing a cloud-based environment facilitated by Google Collab. Throughout the computation, a suite of Python libraries was implemented; each of the library was selected based on specific computational conditions. The NumPy library was applied to perform efficient computational operations, specifically for optimizing numerical arrays for the environmental variable range. The SciPy library was implemented to perform mathematical integration for calculating average sound speed by utilizing the quad function. The Mathplotlib library was also applied for visualization purposes like, originating computational plots to differentiate environmental variables to observe their characteristics. Each empirical equations have been developed as a custom function in the Phyton that takes temperature, depth, and salinity as input and return a comparable sound speed. The resulting simulation calculates the average sound speed for different empirical formulas, then visualizes the outcomes into the labeled subplot to help with comparative analysis. In this work Table 2 shows the simulation environment parameter which later was simulated using the Python environment mentioned above. An average speed of sound was calculated for six empirical models for mixed layers and thermocline layers (Mackenzie, Medwin, Leroy, Wilson, Coppens, and Chen–Millero) and for deep layers there

are only five (Mackenzie, Leroy, Wilson, Coppens, Chen–Millero) that are consistent with the simulation ranges and the Del Grosso empirical formula was not included in the simulation or analysis and its range does not satisfy our desired range of simulation. To calculate the average speed for mixed, thermocline, deep layers the temperature range is considered from 25.5 °C, 4 °C, and 2 °C for the bottom temperature and 30 °C, 10 °C, and 2 °C for the surface temperature. Respectively, in Fig. 5, Fig. 6, and Fig. 7 for mixed, thermocline, and deep layers the variation of salinity are 0.045 p.s.u, 0.5 p.s.u, and 0.3 p.s.u with the added Gaussian noise, estimated at 1 % of the mean acoustic speed of each empirical model in order to reproduce measurement uncertainties, it has been directly added in the 'average speed' of sound. After 100 iterations, the mean average speed has been

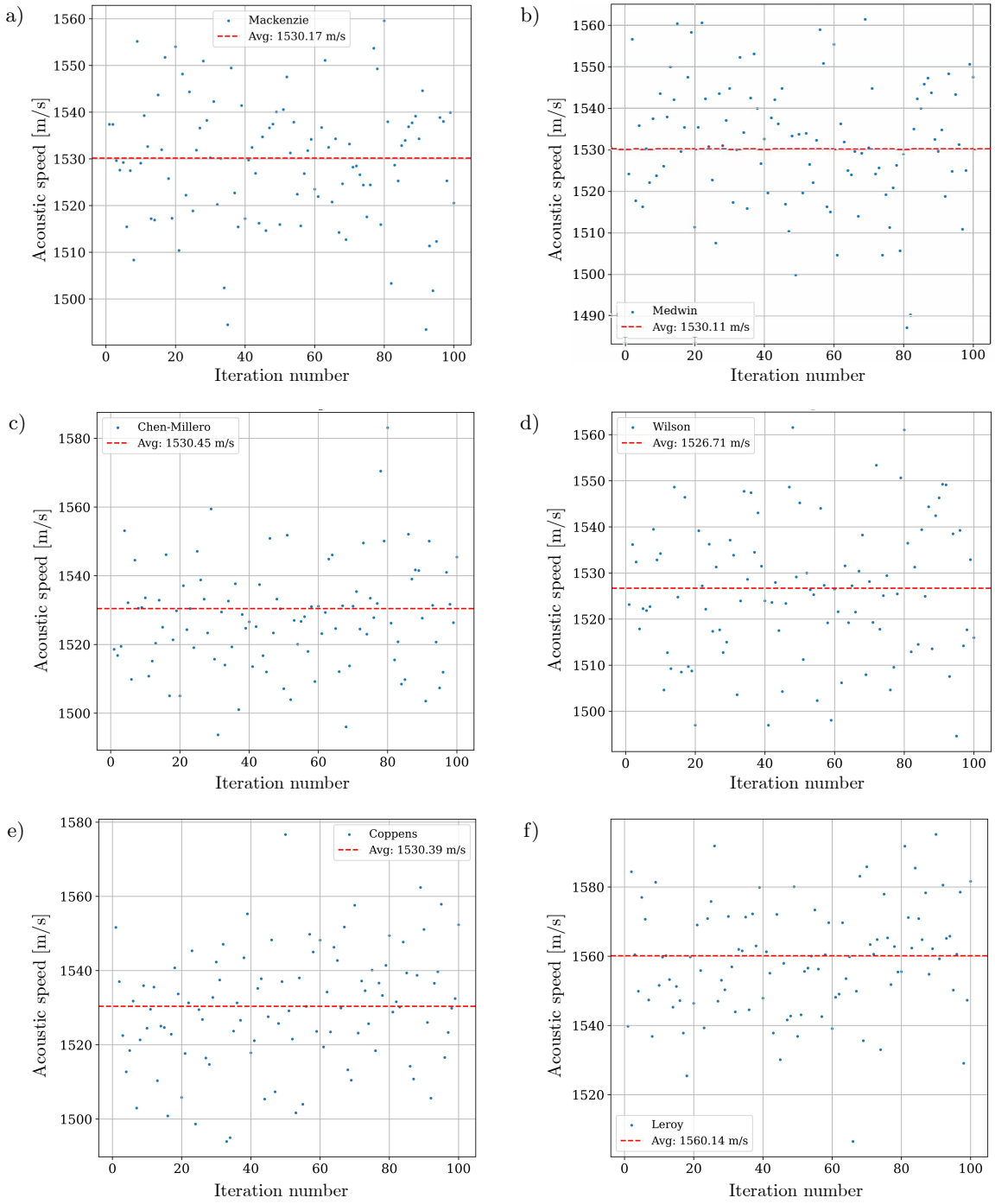


Fig. 5. Average sound speed variation for mixed layers:
a) Mackenzie, b) Medwin, c) Chen–Millereo, d) Wilson, e) Coppens, f) Leroy.

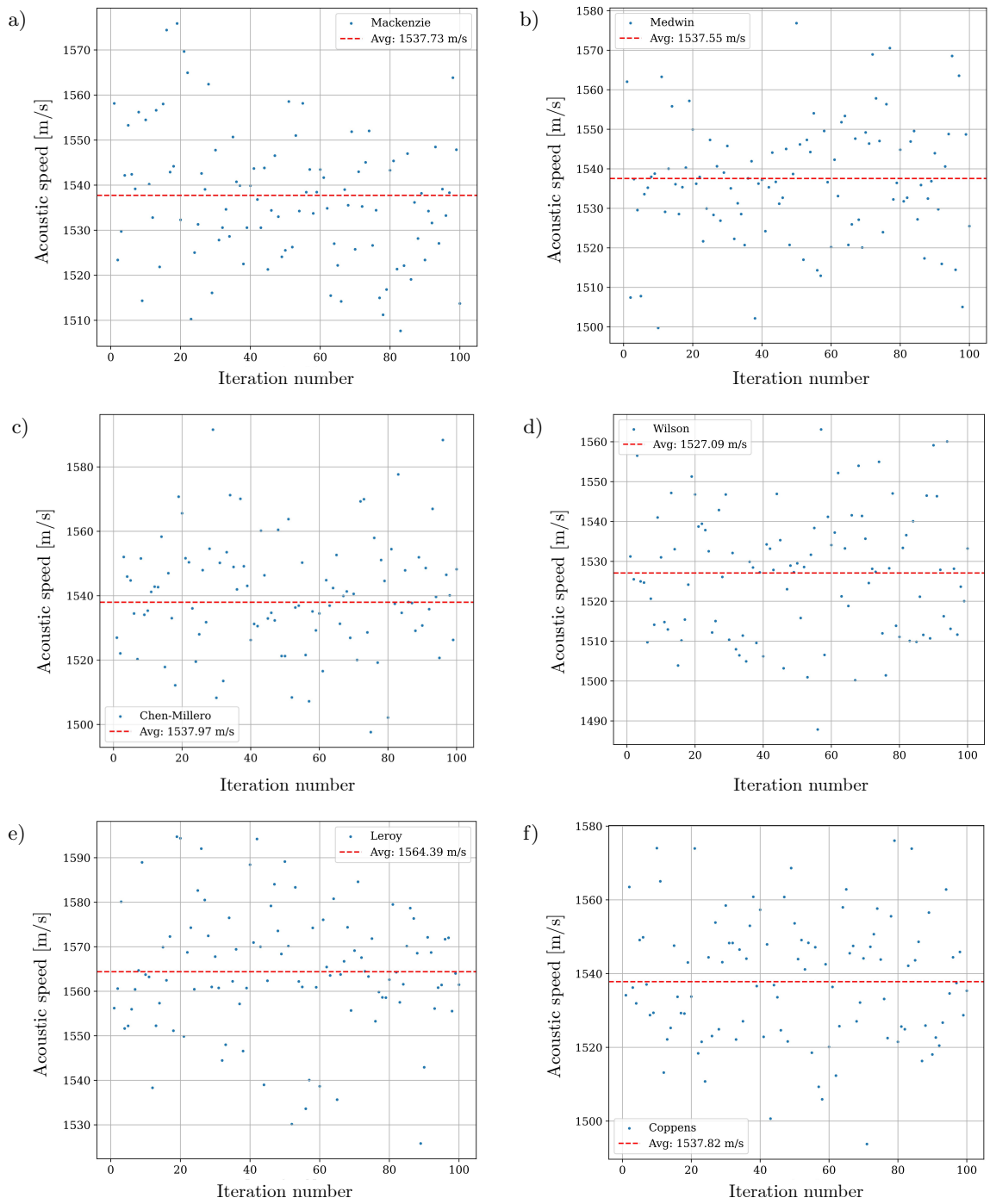


Fig. 6. Average sound speed variation for thermocline layers:
a) Mackenzie, b) Medwin, c) Chen–Millereo, d) Wilson, e) Leroy, f) Coppens.

in the range of 1526.71 m/s to 1560.14 m/s for mixed layers, 1527.09 m/s to 1564.39 m/s for thermocline layers, and 1524.19 m/s to 1597.12 m/s for deep layers.

Figure 8 has portrayed the effect of temperature on the average speed of sound for mixed layers. The bottom temperature is considered from 10 °C to 30 °C under the surface temperature fixed to 25 °C and 28 °C. Additionally, in Fig. 9, for thermocline layers the bottom temperature −8 °C to 2 °C and the surface temperature is fixed to 4 °C as well as 6 °C whereas, 10 is for deep layer where the surface temperature is fixed at 2 °C and 4 °C where a bottom is varying for −20 °C to 0 °C. In mixed layers the average speed increases by 2 m/s to 2.25 m/s per 1 °C increment in surface temperature for Leroy and Coppens formulas and on the other-hand, Mackenzie,

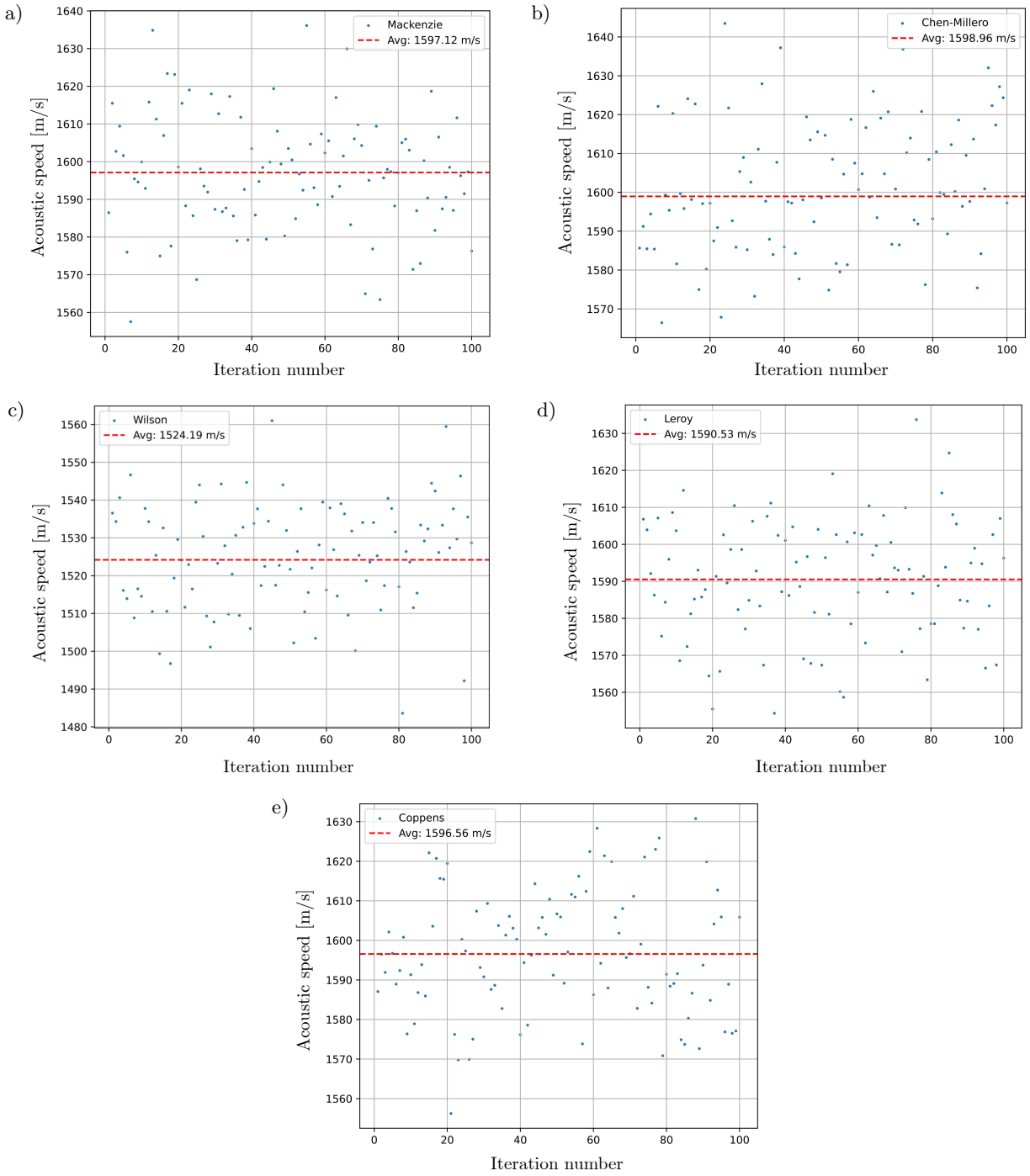


Fig. 7. Average sound speed variation for deep layers:
a) Mackenzie, b) Chen–Millereo, c) Wilson, d) Leroy, e) Coppens.

Medwin, Wilson, and Chen–Millero showcase a moderate level sensitivity which is approximately 1.37 m/s for the set bottom temperature range. A 5 °C rise in bottom temperature will increase the average sound speed by 0.46 % to 0.66 % depending on the formula while formula like Leroy and Coppens reacts most with am increase around 0.62 % and 0.66 %. The rest of the formulas has an overall increment around 0.46 %. Meanwhile for thermocline layers the increase in average sound speed has been increased, as 1 °C ranges from 1.5 m/s to 2 m/s. Again Mackenzie, Medwin, Wilson, and Chen–Millero recorded a slight increase of 1.5 m/s but Leroy and Coppens recorded the high increase of 2 m/s so, this is observed with the highest sensitivity. Again for a 5 °C bottom temperature increases the speed around 0.29 % to 0.41 % for all the formula. It is 0.29 % to 0.31 % for Medwin,

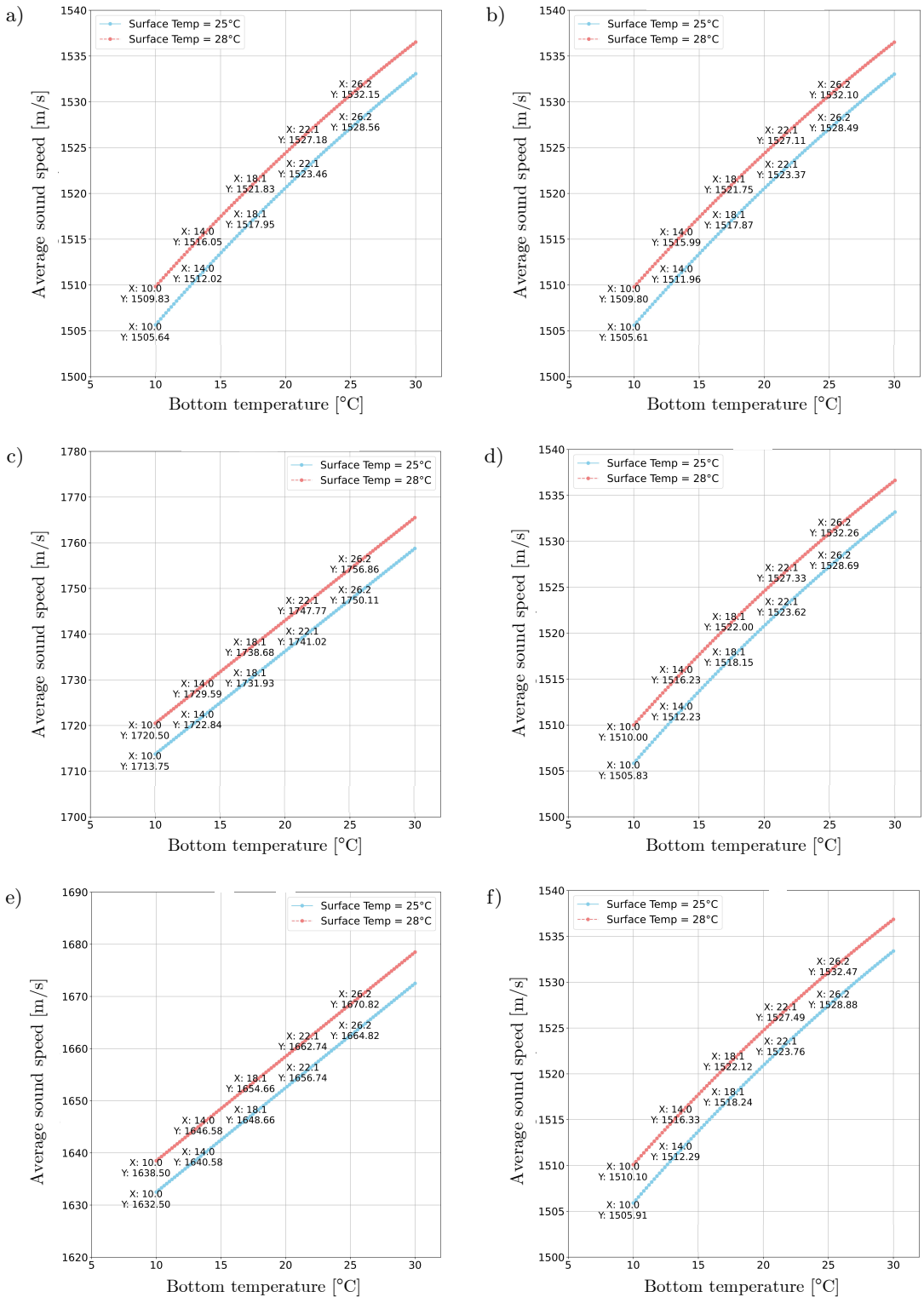


Fig. 8. Average sound speed variation due to temperature changes for mixed layers: a) Mackenzie, b) Medwin, c) Leroy, d) Wilson, e) Coppens, f) Chen–Millero.

Wilson, Chen–Millero, and for Leroy and Coppens it is around 0.37 % and 0.41 %. Lastly, for deep layers the effect of temperature on average speed becomes weaker. The average sound speed increased only 1 m/s to 1.2 m/s, while with a 5 °C increase in bottom temperature, the speed of sound increased again 0.21 % to 0.31 %, with Leroy and Coppens showing the highest sensitivity.

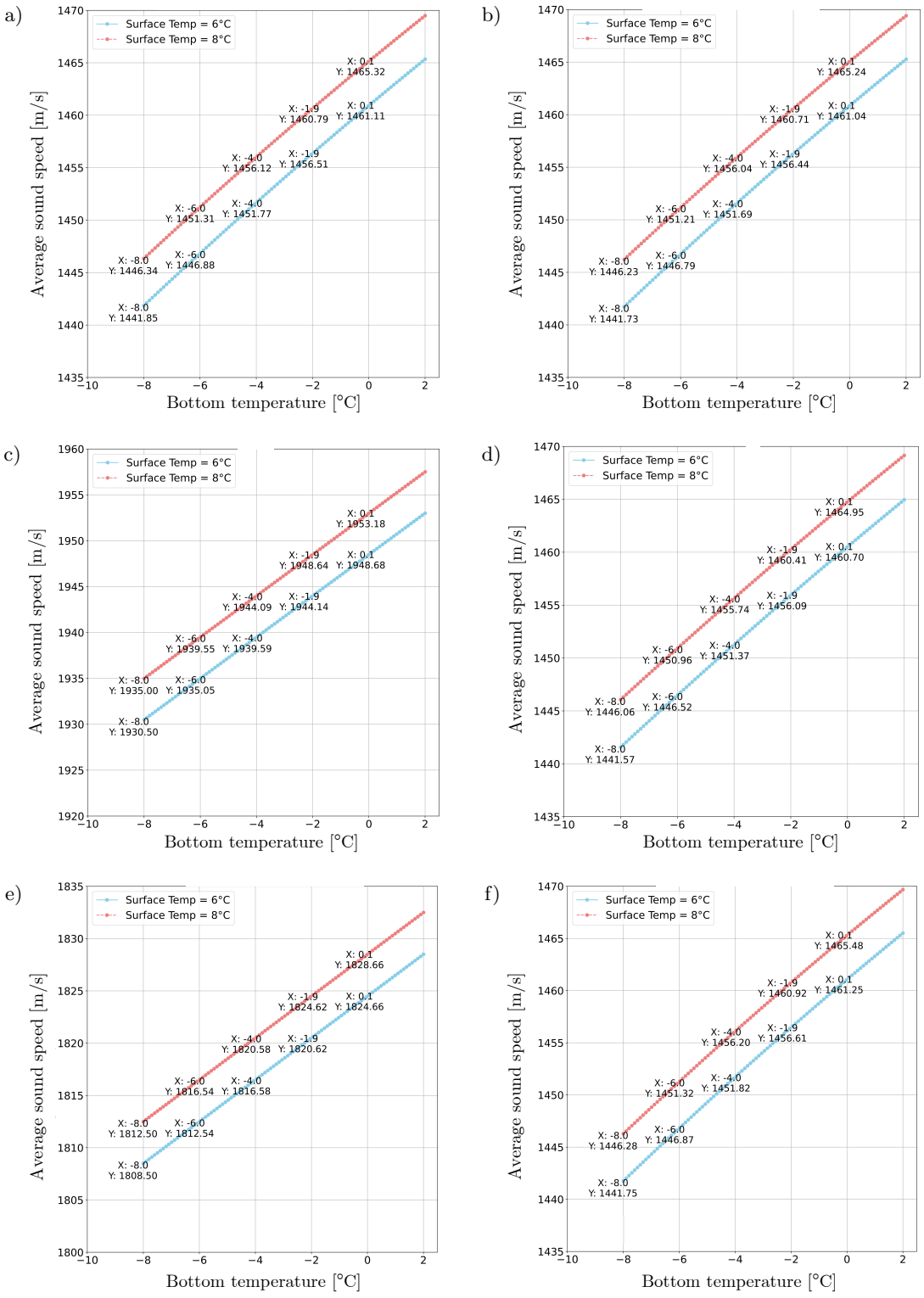


Fig. 9. Average sound speed variation due to temperature changes for thermocline layers: a) Mackenzie, b) Medwin, c) Leroy, d) Wilson, e) Coppens, f) Chen–Millero.

In Fig. 11, Fig. 12, and Fig. 13 the depth measurement accuracy on the computed speed of sound was systematically for 450 m water column for mixed layers, 1000 m for thermocline layers and 8000 m for deep layers. In mixed layer the speed of sound increment is 0.08 m/s to 0.085 m/s for each 10 m increase in the depth and for thermocline and deep layers the increase of sound per 10 m increment is 0.16 m/s to 0.166 m/s and 0.12 m/s

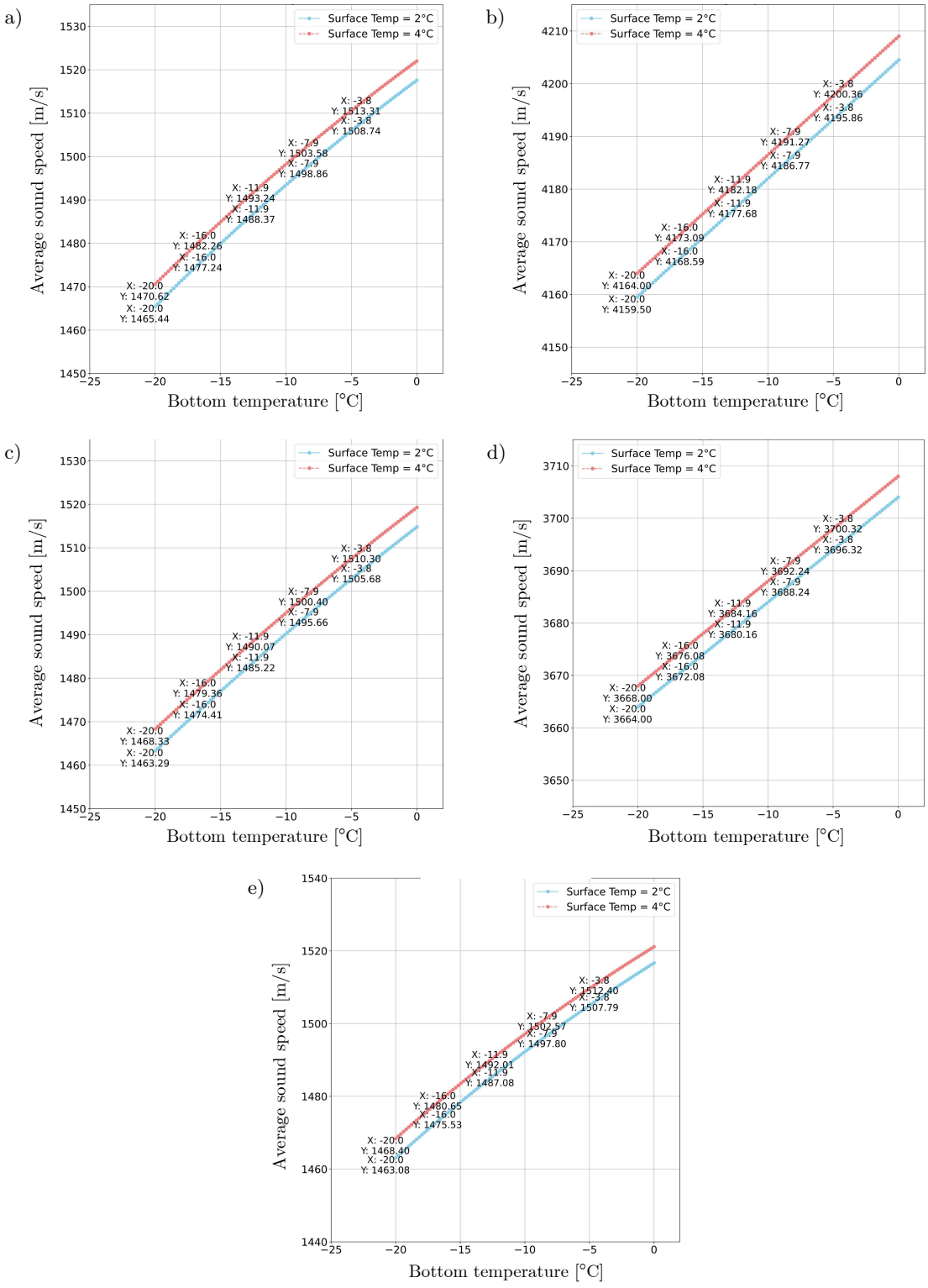


Fig. 10. Average sound speed variation due to temperature changes for deep layers: a) Mackenzie, b) Medwin, c) Leroy, d) Wilson, e) Coppens, f) Chen–Millero.

to 0.13 m/s. In all layers Coppens has the high response and for thermocline and deep layers along with Coppens Mackenzie also shows a great impact around 0.022 % to 0.4 % and 0.019 % to 0.4 %. Now if we consider 5 % error for mixed (22.2 m), thermocline (50 m), and deep layers the effect on the average speed increment is 0.18 m/s to 0.19 m/s (0.012 %), 0.8 m/s to 0.83 m/s (0.055 %), and 4.8 m/s to 5.3 m/s (0.3 % to 0.35 %). The negligible

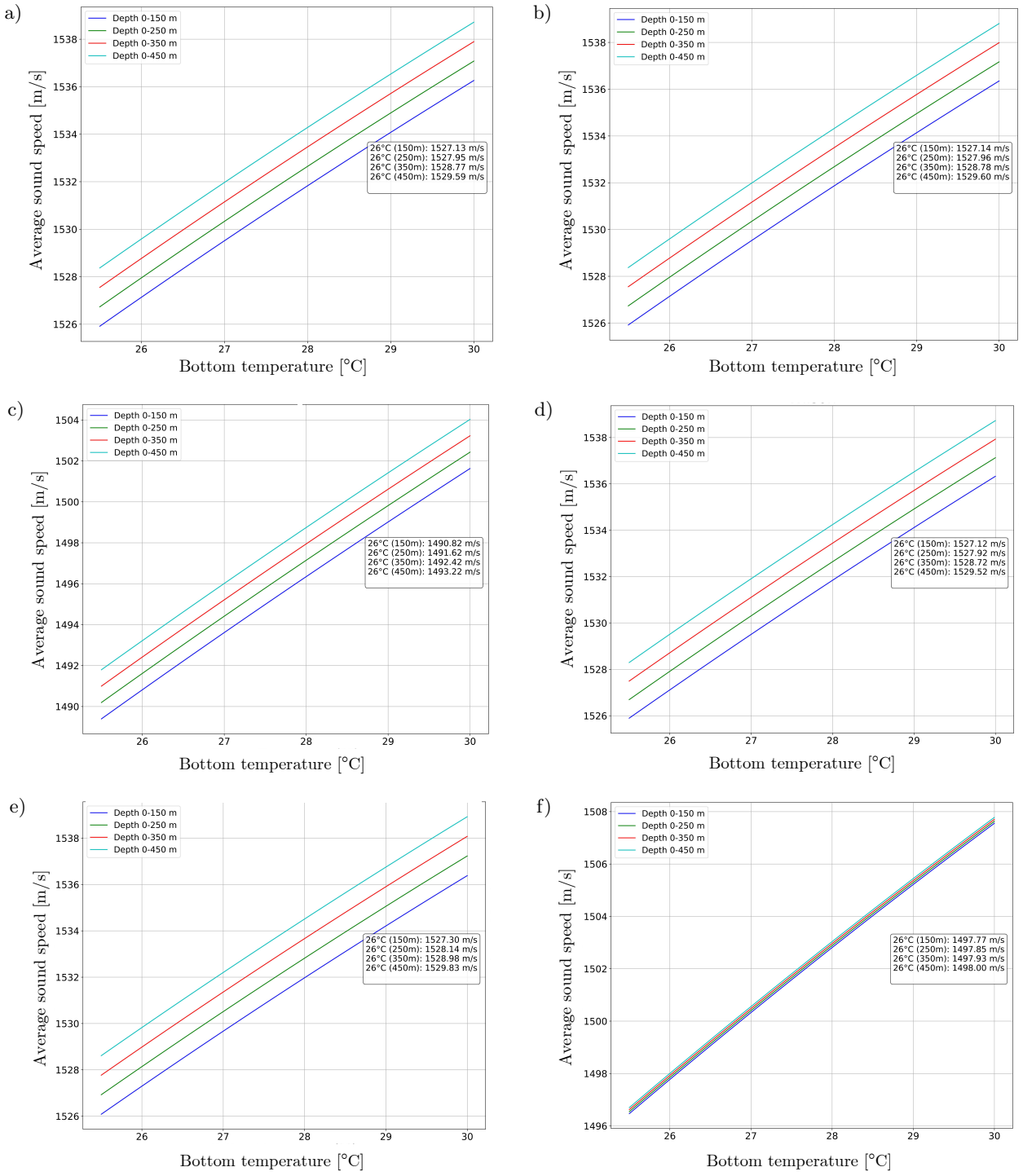


Fig. 11. Average sound speed variation due to depth changes for mixed layers: a) Mackenzie, b) Medwin, c) Leroy, d) Wilson, e) Coppens, f) Chen–Millero.

depth can be for each layer such as: [≤50 m], [≤100 m], and finally [≤200 m]. So, finally we can say that the rate of change of sound for mixed, thermocline, and deep layers is accordingly to 0.0083 $\mathrm{m\,s^{-1}\,m^{-1}}$, 0.0165 $\mathrm{m\,s^{-1}\,m^{-1}}$, and 0.013 $\mathrm{m\,s^{-1}\,m^{-1}}$ which means the sound speed rises to 0.083 m/s, 0.0165 m/s, and 0.013 m/s for each 1 m increment on the depth.

The effect of salinity has been portrayed in Fig. 14, Fig. 15, and Fig. 16, respectively. To observe the salinity impact on the three oceanic layers we consider a fixed temperature for all the three layers which are 27 °C, 5 °C, and 2 °C, respectively. Considering the salinity increases to 0.045 ppt for mixed layer shown in Fig. 14,

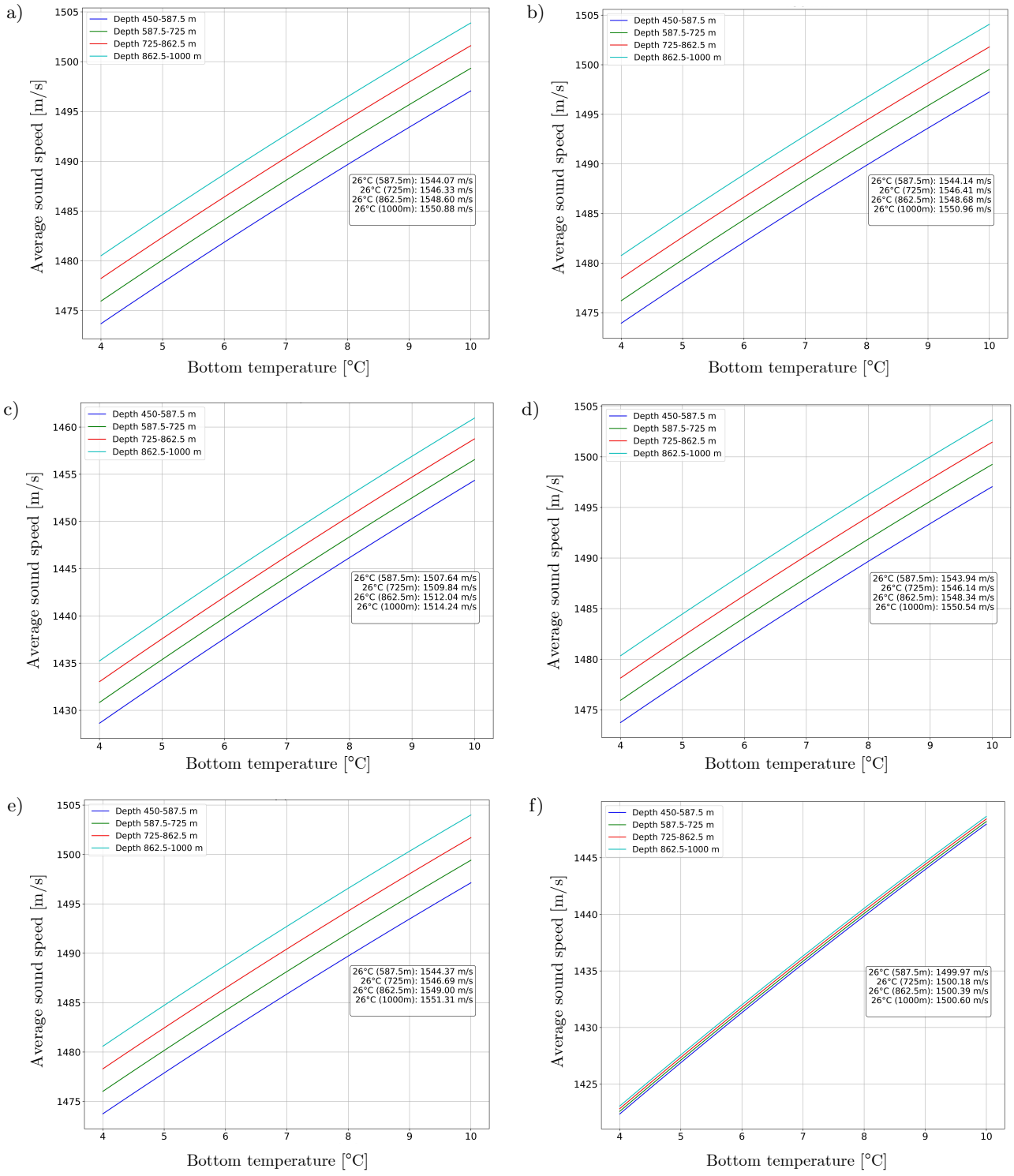


Fig. 12. Average sound speed variation due to depth changes for thermocline layers: a) Mackenzie, b) Medwin, c) Leroy, d) Wilson, e) Coppens , f) Chen–Millero.

for thermocline in Fig. 15, and deep layer from the Fig. 16 it is shown to be 0.5 ppt and 0.3 ppt. All formulas show a positive rise in the average speed of sound. In mixed layer the rise in average speed is +0.049 m/s to +0.068 m/s when Mackenzie and Medwin show a positive rise of +0.049 m/s (0.0033 %) Leroy and Wilson show a very high effect around +0.059 m/s (0.00398 %) and +0.068 m/s (0.00441 %) although this time Coppens has the lowest impact which is +0.049 m/s (0.00185 %). The effect of salinity is still negligible. In thermocline layer with

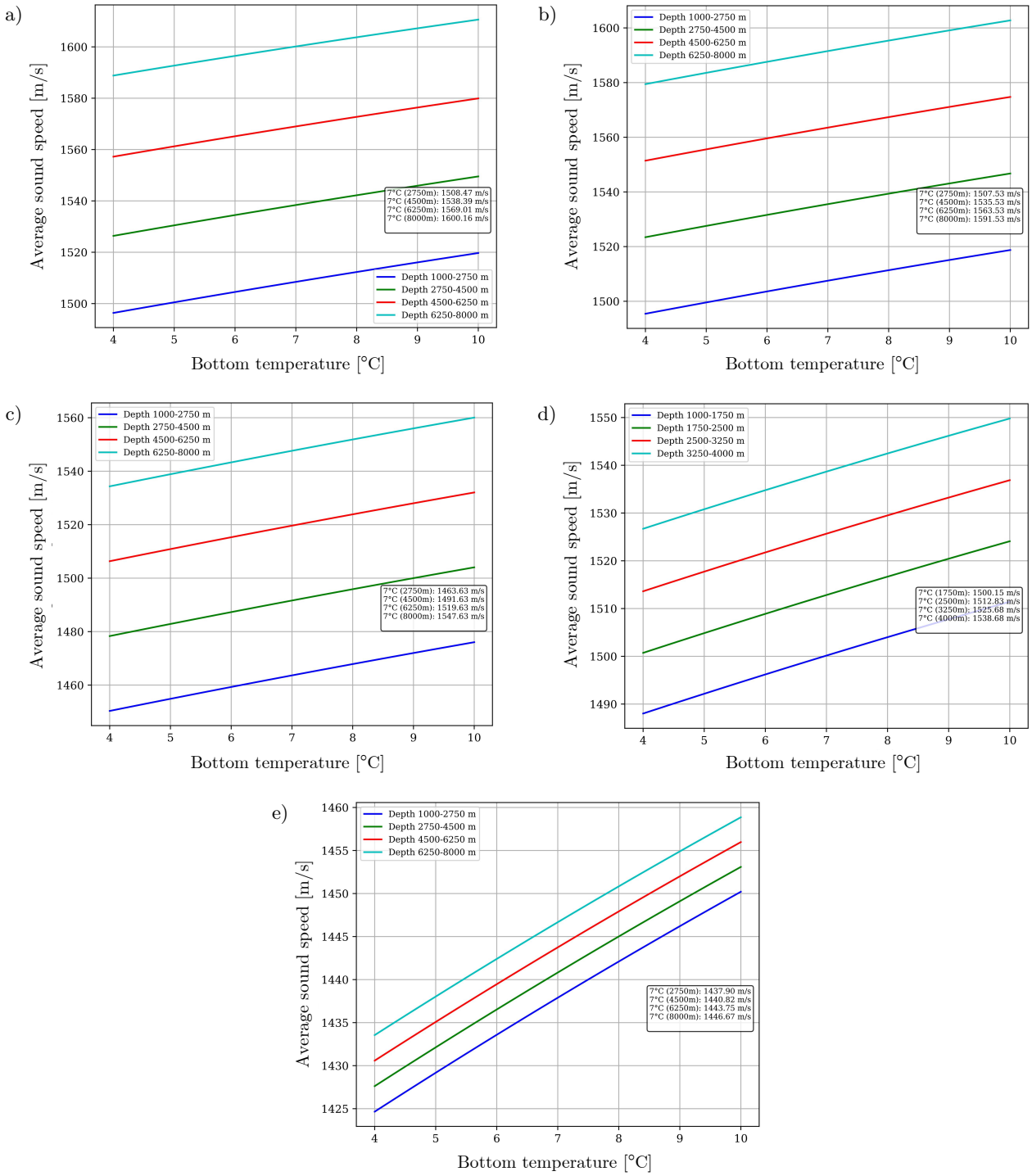


Fig. 13. Average sound speed variation due to depth changes for deep layers: a) Mackenzie, b) Medwin, c) Leroy, d) Wilson, e) Coppens, f) Chen–Millero.

an increase in salinity variation the domination increases. The average speed rise for all formula is still positive and it is around 0.73 m/s to 0.9 m/s. Mackenzie shows +0.75 m/s (0.05 %), Medwin it is +0.73 m/s (0.048 %) and for Leroy, Wilson, Coppens, and Chen–Millero it is correspondingly +0.82 m/s (0.055 %), +0.9 m/s (0.058 %), +0.7 m/s (0.026 %), and +0.78 m/s (0.053 %). The influence of salinity is still less than that of temperature, but it cannot be ignored. Lastly, for deep layer the variation is just 0.3 ppt but its influence is noticeable for an increase of the speed of the sound. All formulas still show a positive increase from +1.2 m/s to +1.35 m/s, where Mackenzie, Leroy, Wilson, Coppens, and Chen–Millero are concerned, respectively, +1.2 m/s (0.08 %), +1.28 m/s

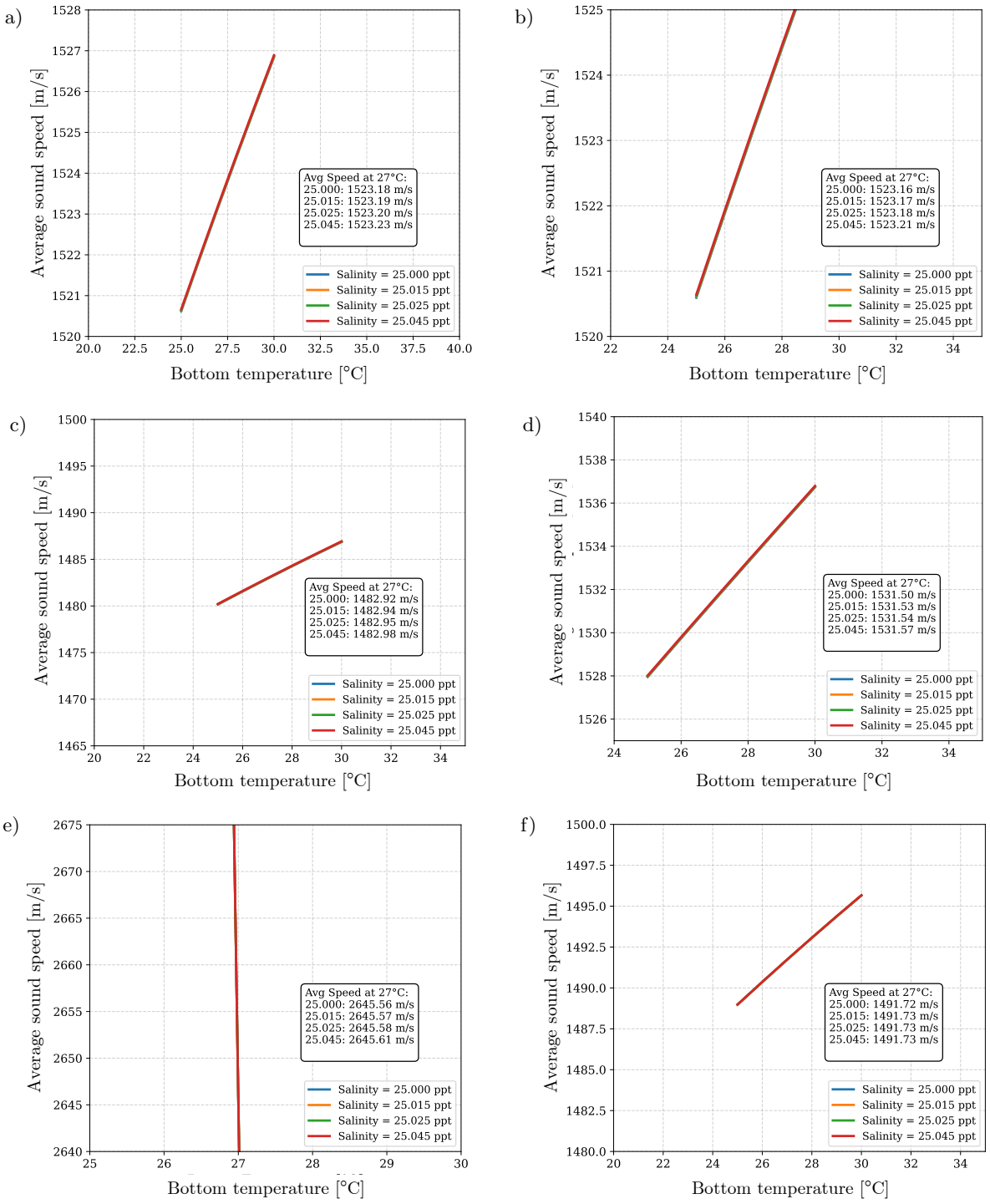


Fig. 14. Average sound speed variation due to salinity changes for mixed layers: a) Mackenzie, b) Medwin, c) Leroy, d) Wilson, e) Coppens, f) Chen–Millero.

(0.087 %), +1.35 m/s (0.089 %), +1.15 m/s (0.043 %), +1.3 m/s (0.085 %), so we can say Mackenzie has the least effect and Wilson has the highest effect. Comprehensively, the salinity effect increases layer by layer.

5. Discussion

In this study, we have systematically analyzed the effect of three environmental parameters which are temperature, depth, and salinity. The effects of these parameters on the acoustic signal for three ocean layers are

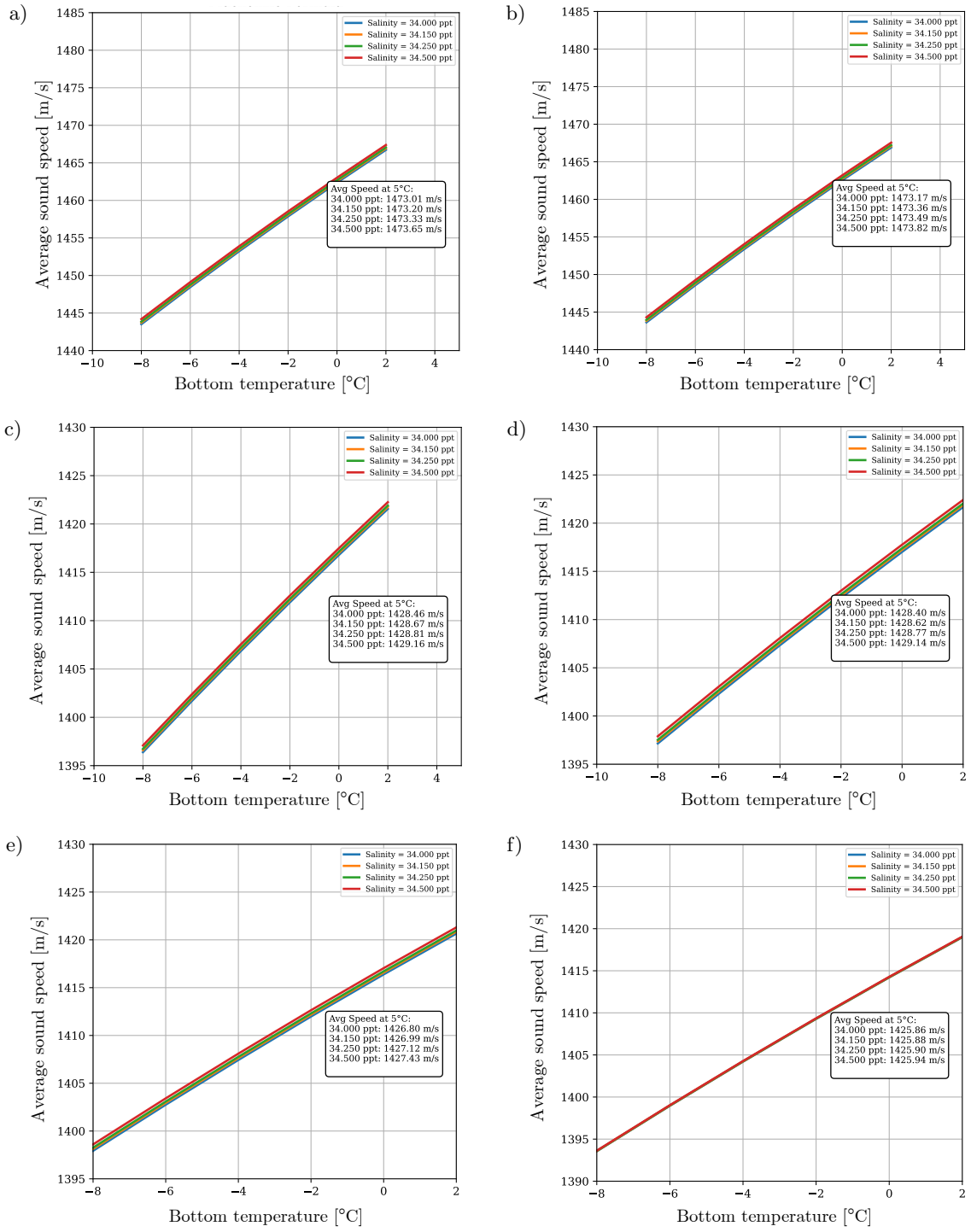


Fig. 15. Average sound speed variation due to salinity changes for thermocline layers: a) Mackenzie, b) Medwin, c) Leroy, d) Wilson, e) Coppens, f) Chen–Millero.

demonstrated using different empirical formulas. Each formulas compute the mean speed of sound considering the ranges of deep-ocean conditions. After considering the different empirical formulas and all parameters were calculated using the average speed for empirical formulas which satisfied the simulation range and got the mean average sound speed around 1526.71 m/s to 1560.14 m/s for all three oceanic layers and for practicality we introduced the Gaussian noise with a mean 0 and standard deviation for 100 iterations. For the computation of average sound speed the triple integral method was considered. In Table 3 a detailed interpretation of this research work has been presented. The three prime oceanic layers (mixed, thermocline, deep) have been observed to estimate

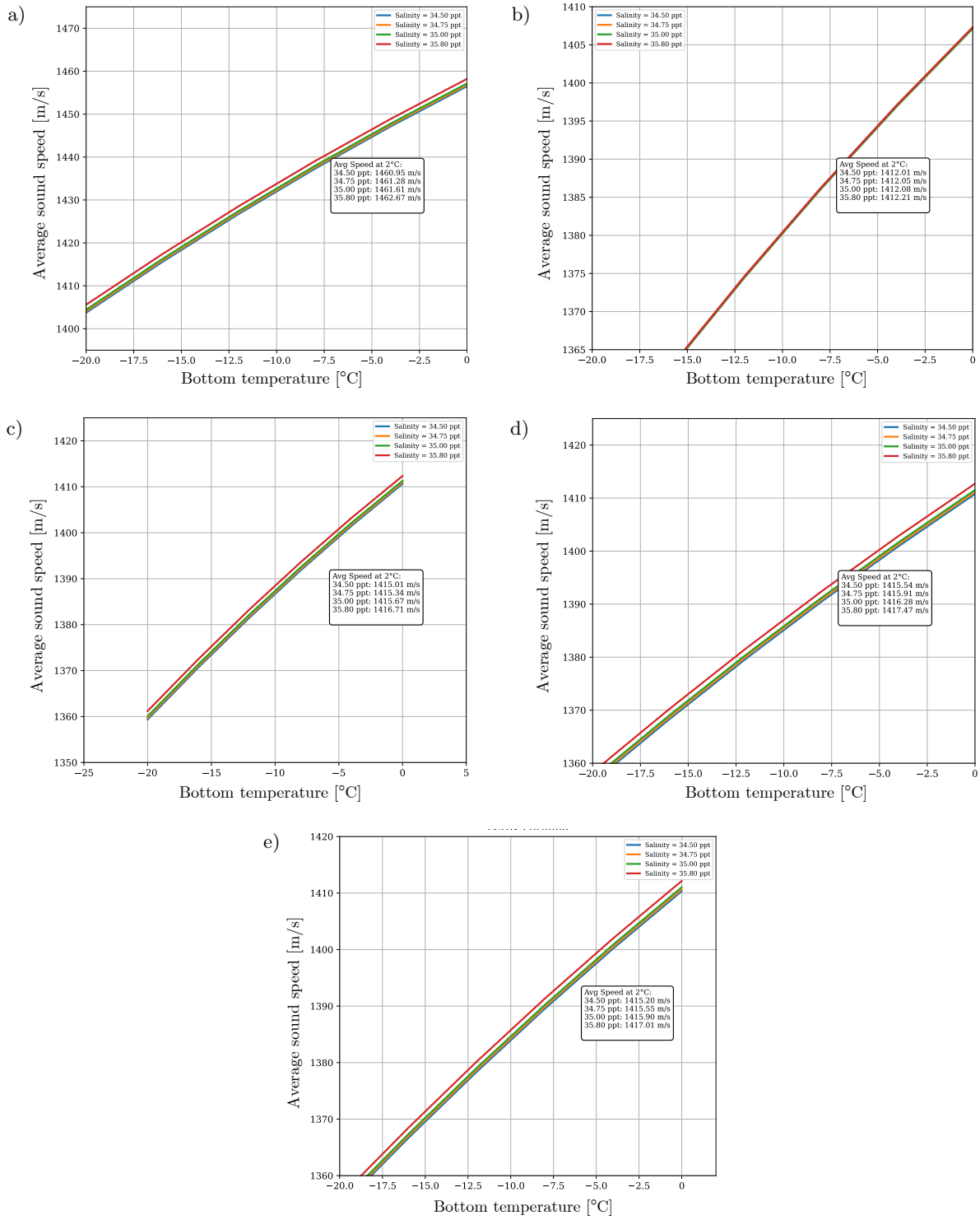


Fig. 16. Average sound speed variation due to salinity changes for deep layers: a) Mackenzie, b) Medwin, c) Leroy, d) Wilson, e) Coppens, f) Chen–Millero.

the variation in the mean sound speed and each layer has its own range of the environmental parameter which has been demonstrated in Table 2. Each layer displays a unique response to the major environmental variables: temperature, depth, and salinity. The effects of these parameters depend on difference between each layer range and also upon different empirical formulas including Mackenzie, Medwin, Leroy, Wilson, Coppens, and Chen–Millero. Table 3 highlights the amplitude of changes in the average speed of sound due to gradual variation in the values of temperature, depth, and salinity.

Table 3. Analysis of environmental parameter influence on oceanic layers.

Oceanic layer	Temperature effect	Depth effect	Salinity effect	Constrain effect on formulas	Parameter role
Mixed layer	1 °C rise in the temperature can cause a significant impact on the average speed around 2 m/s to 2.5 m/s	10 m rise can cause the arise of average sound speed is only 0.08 m/s to 0.085 m/s	Salinity has a minimal effect of only 0.049 m/s to 0.068 m/s for variation of 0.045 ppt	The Mackenzie and Medwin formulas show a high sensitivity in the mixed layer where temperature effect is around 0.32 %	Mixed layers temperature is the most domination parameter among them and depth and salinity effect can be ignored as they influence the speed just 0.04 % and 0.003 %
Thermocline layer	1 °C rise in the temperature cause fluctuation on average speed around 1.5 m/s to 2 m/s	10 m rise can cause the arise of average sound speed is only 0.16 m/s to 0.166 m/s	Salinity effect is only 0.73 m/s to 0.9 m/s for a variation of 0.5 ppt	In thermocline layer, formulas like Wilson, Leroy, and Coppens show sensitivity to all parameters like for temperature it is 0.3 % and for depth is 0.05 % (specially for Wilson)	In thermocline layer temperature effects starts to fall and slowly around 0.5 m/s, depth and salinity started to influence the speed more the increment can be shown almost 0.8 m/s and 0.3 m/s
Deep layer	1 °C temperature hikes the average speed is just 1 m/s to 1.2 m/s	10 m rise in depth causes a change in average speed is 4.8 m/s to 5.3 m/s	Salinity is 1.2 m/s to 1.35 m/s for a small variation of 0.3 ppt	Mackenize, Coppens, and Chen–Millero show a sensitivity and this time is for depth around 0.6 %	Deep layer temperature is considered as the weakest impact and depth has the highest effect around 0.6 % and salinity effect also increase to 0.01 % which is higher than the temperature effect of 0.001 %

6. Conclusion

This work has drawn attention to the fact that the three main environmental parameters have different effect on an acoustic sound speed based on the oceanic layer. When the signal passes from the layer close to a surface, it has the maximum effect due to temperature and the salinity effect can be ignored but as we start thinking of going deeper the influence of temperature gradually decreases and other two parameters which are depth and salinity effect began to increase in the deepest part of the ocean. The sound seed is very sensitive to depth and temperature and the depth effect becomes the same and also minimal. This understanding can be crucial for deploying and developing underwater communication networks much more effectively as this ecological variation can introduce unpredictable noise which will also bring an effect on the speed, direction, supervision and the timing of the marine communication and can also helps in the development of more optimized sonar and navigation algorithms.

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.

Authors' contributions

Khondoker Munim Salehin contributed in experimental data analysis and manuscript drafting. Md. Khalid Mahbub Khan contributed in manuscript writing, modeling, and also in conceptualization. Anisur Rahman supervised as well as conceptualized the entire research. All authors reviewed and approved the final manuscript.

References

1. Chen C.T., Millero F.J. (1977), Speed of sound in seawater at high pressures, *The Journal of the Acoustical Society of America*, **62**(5): 1129–1135, https://doi.org/10.1121/1.381646.
2. Coppens A.B. (1981), Simple equations for the speed of sound in Neptunian waters, *The Journal of the Acoustical Society of America*, **69**(3): 862–863, https://doi.org/10.1121/1.385486.
3. Gill A.E. (2016), *Atmosphere – Ocean Dynamics*, Elsevier.
4. Huang W. *et al.* (2024), Underwater SSP measurement and estimation: A survey, *Journal of Marine Science and Engineering*, **12**(12): 2356, https://doi.org/10.3390/jmse12122356.
5. Jamieson A.J., Swanborn D.J.B., Bond T., Arasu P., Partridge J.C. (2025), Reconsidering the term 'deep sea', *ICES Journal of Marine Science*, **82**(6): fsaf080, https://doi.org/10.1093/icesjms/fsaf080.
6. KukuChuku S., Idoniboyeobu D.C., Sunny O., Lewis O.T. (2018), Improved underwater wireless communication system using OFDM technique, *American Journal of Engineering Research (AJER)*, **7**(9): 82–95, https://www.ajer.org/papers/Vol-7-issue-9/K07098295.pdf (access: 29.10.2025).
7. Kundu S. (2016), *In-situ determination of acoustic average speed in underwater environment*, East West University, http://103.133.167.5:8080/bitstream/handle/123456789/1946/Shrabonti_Kundu.pdf?sequence=1&isAllowed=y (access: 29.10.2025).
8. Lana A.B., Kurniawati N., Purba N.P., Syamsuddin M.L. (2017), Thermocline layers depth and thickness in Indonesian waters during the Southeast Monsoon, *Omni-Akuatika*, **13**(2): 65–72, https://doi.org/10.20884/1.oa.2017.13.2.70.
9. Leroy C.C. (1969), Development of simple equations for accurate and more realistic calculation of the speed of sound in seawater, *The Journal of the Acoustical Society of America*, **46**(1B): 216–226, https://doi.org/10.1121/1.1911673.
10. Mackenzie K.V. (1981), Nine-term equation for sound speed in the oceans, *The Journal of the Acoustical Society of America*, **70**(3): 807–812, https://doi.org/10.1121/1.386920.
11. Medwin H. (1975), Speed of sound in water: A simple equation for realistic parameters, *The Journal of the Acoustical Society of Americ*a, **58**(6): 1318–1319, https://doi.org/10.1121/1.380790.
12. National Physical Laboratory [NPL] (n.d.), *Technical guides – Speed of sound in sea water – Underlying physics*, http://resource.npl.co.uk/acoustics/techguides/soundseawater/underlying-phys.html (access: 31.10.2025).
13. Paul A., Islam M., Rahman Md.F., Rahman A. (2020), Localization of mobile submerged sensors using Lambert-W function and Cayley-Menger determinant, *International Journal of Advanced Computer Scienceand Applications*, **11**(2), https://doi.org/10.14569/IJACSA.2020.0110266.
14. Rahman A. (2014), The effect of in-situ underwater acoustic speed on localization of submerged sensors, [in:] *Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, https://doi.org/10.1109/WoWMoM.2014.6918936.
15. Rahman A., Muthukkumarasamy V., Sithirasenan E. (2013a), Coordinates determination of submerged sensors using Cayley-Menger determinant, [in:] *2013 IEEE International Conference on Distributed Computing in Sensor Systems*, pp. 466–471, https://doi.org/10.1109/dcoss.2013.62.
16. Rahman A., Muthukkumarasamy V., Sithirasenan E. (2013b), The analysis of temperature, depth, salinity effect on acoustic speed for a vertical water column, [in:] *2013 IEEE International Conference on Distributed Computing in Sensor Systems*, pp. 310–312, https://doi.org/10.1109/DCOSS.2013.63.
17. Scowcroft G., Vigness-Raposa K., Knowlton C., Morin H. (2015), *Discovery of sound in the sea booklet*, University of Rhode Island, https://dosits.org/wp-content/uploads/2017/07/DOSITS-Booklet-2015-web.pdf (access: 29.10.2025).
18. Stojanovic M. (2007), Frequency reuse underwater: capacity of an acoustic cellular network, [in:] *Proceedings of the 2nd Workshop on Underwater Networks*, pp. 19–24, https://doi.org/10.1145/1287812.1287818.

19. Talib K.H., Othman M.Y., Sulaiman S.A.H., Wazir M.A.M., Azizan A. (2011), Determination of speed of sound using empirical equations and SVP, [in:] *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, pp. 252–256, https://doi.org/10.1109/CSPA.2011.5759882.

20. Tan H.-P., Diamant R., Seah W., Waldmeyer M. (2011), A survey of techniques and challenges in underwater localization, *Ocean Engineering*, **38**(14): 1663–1676, https://ink.library.smu.edu.sg/sis_research/2956.

21. Wilson D.W. (1977), Equation for the speed of sound in sea water, *The Journal of the Acoustical Society of America*, **32**(10): 1357, https://doi.org/10.1121/1.1907913.

Archives of Acoustics Vol. 51, No. 1, pp. 131–143 (2026), https://doi.org/10.24423/archacoust.2026.4295

Research Paper

An Improved Dynamic Time Warping Algorithm for Active Sonar Signal Matching

Tongjing Sun*, Hunyuan Wang, Lei Chen, Haoran Xu

Department of Automation, Hangzhou Dianzi University
Hangzhou, China

*Corresponding Author: stj@hdu.edu.cn

Received August 7, 2025; revised November 27, 2025; accepted December 10, 2025; available online January 16, 2026; version of record February 10, 2026; published issue March 27, 2026.

Active sonar signal matching is a critical technique for measuring inter-signal similarity and enhancing target detection and classification performance. However, in complex underwater environments, noise, reverberation, and prolonged signal durations often degrade matching accuracy and computational efficiency. To address these challenges, this paper proposes an adaptive extremum-aligned boundary-constrained dynamic time warping (AEB-DTW) algorithm, based on the classical dynamic time warping (DTW) framework. The algorithm extracts significant extrema from signal envelopes to suppress noise and reverberation while capturing salient features. By integrating the position and amplitude of extrema, an adaptive weighted matching strategy is introduced to enhance feature discrimination. In addition, spline fitting is applied to the residuals of the extremum matching path to dynamically generate upper and lower boundary constraints, thus restricting DTW computation to a meaningful region and achieving a balance between accuracy and efficiency. Experiments using lake-trial active sonar data under signal-to-reverberation ratios (SRRs) from 0 dB to 30 dB show that AEB-DTW outperforms Euclidean distance (ED), DTW, and its variants in matching accuracy, robustness, and angular resolution, while significantly improving computational efficiency, particularly for long-duration signals.

Keywords: active sonar signal matching, dynamic time warping (DTW), time series similarity, adaptive boundary constraints.

1. Introduction

Active sonar signals, as a typical type of time-series data, play a crucial role in underwater navigation, target detection, and marine resource exploration (Bian *et al.*, 2024; Zhang *et al.*, 2023; Zhu *et al.*, 2025). In complex marine environments, efficiently and accurately measuring the similarity between signals is a fundamental basis for supporting applications such as target detection, localization, feature extraction, and classification using active sonar (Sun *et al.*, 2024; Dau *et al.*, 2018). It also has a profound impact on improving the processing efficiency and recognition performance of subsequent tasks (Silva *et al.*, 2018).

Dynamic time warping (DTW) is a widely used algorithm for time-series alignment that addresses limitations of traditional Euclidean distance (ED) metrics by allowing nonlinear temporal warping (Wang *et al.*, 2022; Chen *et al.*, 2017). DTW effectively handles sequences of differing lengths, temporal shifts, and amplitude variations, enabling robust and precise matching (Li, Guo, 2013). Consequently, it has attracted extensive research attention worldwide.

To improve computational efficiency, methods such as LB_Yi, LB_Kim, and LB_Keogh have been proposed by Yi *et al.* (1998), Kim *et al.* (2001), and Keogh, Ratanamahatana (2005), respectively. These methods define lower-bound constraints on sequence distances to exclude dissimilar sequences and to reduce the search space for matching. Salvador and Chan (2007) introduced the fast-DTW algorithm, which reduces computational cost by

projecting sequences into low-dimensional spaces to approximate the warping path. Lahreche and Boucheham (2021) developed LEDTW, which significantly reduces complexity by performing DTW only on local extrema. Tang and Gao (2023) proposed EWDTW, incorporating adaptive weighting strategies based on local extrema to enhance both efficiency and classification accuracy. Sakoe and Chiba (1978) introduced the Sakoe–Chiba band, which constrains the warping path's search area to reduce matrix computations.

For improving alignment accuracy, Jeong *et al.* (2011) and Li (2021) proposed WDTW and TWDTW, which adjust feature weights to enhance alignment precision. Keogh and Pazzani (2001) introduced DDTW, which focuses on the trend of sequence variation by incorporating first-order derivatives to avoid misalignments. Li *et al.* (2020) proposed ACDTW, which dynamically adjusts the warping window to better match local features. Zhao and Itti (2018) proposed shapeDTW, which uses local extremum structures to enhance the ability to identify complex shapes. Hong *et al.* (2020) proposed SSDTW, which combines sparse representation with spatial structural constraints for efficient high-dimensional sequence alignment.

Despite these developments, existing DTW variants still struggle to balance alignment accuracy and computational efficiency for active sonar echoes. In complex underwater environments, strong reverberation and noise distort signal envelopes, causing unstable similarity estimation, while fixed or manually tuned warping constraints fail to adapt to variable signal durations. Therefore, the core technical problem addressed in this work is achieving robust and efficient similarity matching for active sonar signals under low signal-to-reverberation ratio (SRR) conditions while preserving key structural features.

To solve this problem, this study proposes the adaptive extremum-aligned boundary-constrained dynamic time warping (AEB-DTW) algorithm. The method extracts significant extrema to suppress noise and highlight salient envelope structures, introduces adaptive weighting to enhance feature discrimination, and constructs dynamic boundary constraints to reduce computation while maintaining alignment precision. The main contributions are developing a structure-aware matching framework, achieving a robustness–efficiency balance, and validating its effectiveness through comprehensive lake-trial experiments against ED, DTW, and its variants.

2. DTW algorithm

The core objective of the DTW algorithm is to find an optimal alignment path that minimizes the overall distance between two sequences. Let the features of the envelope of two 1D active sonar signals be $\mathbf{X} = \{x_1, ..., x_i, ..., x_m\}$ and $\mathbf{Y} = \{y_1, ..., y_i, ..., y_n\}$, with lengths m and n, respectively, as shown in Fig. 1a.

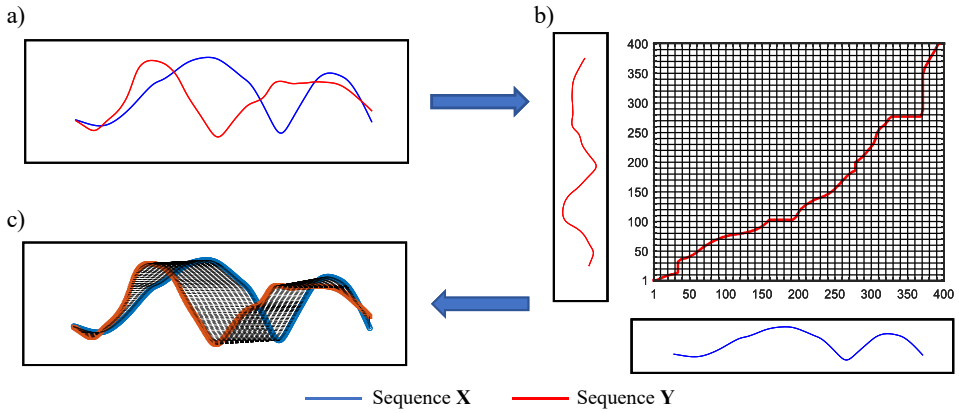


Fig. 1. DTW alignment effect and warping path.

To implement this optimal alignment, we turn to the DTW matching process, which is illustrated in Fig. 1b; sequences $\mathbf{X}$ and $\mathbf{Y}$ are mapped to the horizontal and vertical axes of a 2D grid to construct a distance matrix. Using the principle of minimum cumulative distance, the shortest path from the bottom-left corner to the top-right corner is found, representing the optimal nonlinear alignment between the two sequences. Figure 1c shows the pointwise correspondence along the optimal warping path.

To compute the DTW distance between sequences $\mathbf{X}$ and $\mathbf{Y}$, we first define the point-wise distance function between elements x_i and y_j:

$$d(i,j) = \|x_i - y_j\|_\beta , \tag{1}$$

where $\beta \in \mathbb{R}^+$, and $\beta = 2$ corresponds to the Euclidean distance. To compute the DTW distance between sequences $\mathbf{X}$ and $\mathbf{Y}$, an optimal warping path must be identified. Let this optimal path be denoted as

$$\mathbf{W} = \{(i_1, j_1), (i_2, j_2), ..., (i_k, j_k), (i_K, j_K)\}, \tag{2}$$

where K is the length of the warping path, and each (i_k, j_k) represents a matched pair of indices from sequences $\mathbf{X}$ and $\mathbf{Y}$, respectively. The warping path $\mathbf{W}$ is an ordered set of index pairs that aligns the two sequences in a nonlinear manner. Its formal constraints and cost computation are described as follows:

- boundary condition: $(i_1, j_1) = (1,1)$, and $(i_K, j_K) = (m,n)$,
- monotonicity: $i_{k+1} \geq i_k$, $j_{k+1} \geq j_k$,
- continuity: $(i_{k+1} - i_k, j_{k+1} - j_k) \in \{(1,0),(0,1),(1,1)\}$.

Under these constraints, the DTW distance is calculated as

$$\mathrm{DTW}(\mathbf{X}, \mathbf{Y}) = \min_{\mathbf{W}} \left\{ \sum_{k=1}^{K} d(i,j) \right\}. \tag{3}$$

To solve this, a cumulative cost matrix $\mathbf{D}$ is constructed recursively as

$$D(i,j) = d(i,j) + \min\{D(i-1,j), D(i,j-1), D(i-1,j-1)\}, \tag{4}$$

where $i = 1,2,...,m$, $j = 1,2,...,n$. The initial conditions are set as $D(0,0) = 0$, and $D(i,0) = D(0,j) = \infty$. Equation (4) indicates that the cumulative distance at a given point equals the sum of the current pointwise distance and the minimum cumulative distance from its neighboring positions. $D(m,n)$ denotes the DTW distance between sequences $\mathbf{X}$ and $\mathbf{Y}$, and the computational complexity of the algorithm is $O(m \cdot n)$.

3. AEB-DTW algorithm

The core idea of the AEB-DTW algorithm is to adaptively extract significant extrema from active sonar signal envelopes in order to capture key structural features. A weighted matching model is constructed using amplitude and phase differences to enhance the discrimination ability of extrema alignment. Based on the weighted matching results, a dynamic upper and lower boundary constraint is constructed to restrict the DTW search region, thereby enhancing both efficiency and robustness.

Figure 2 illustrates the processing flow of AEB-DTW. The signal envelope is first extracted to highlight major structures and suppress noise (Fig. 2a). Then, significant extrema are adaptively identified to form a reduced

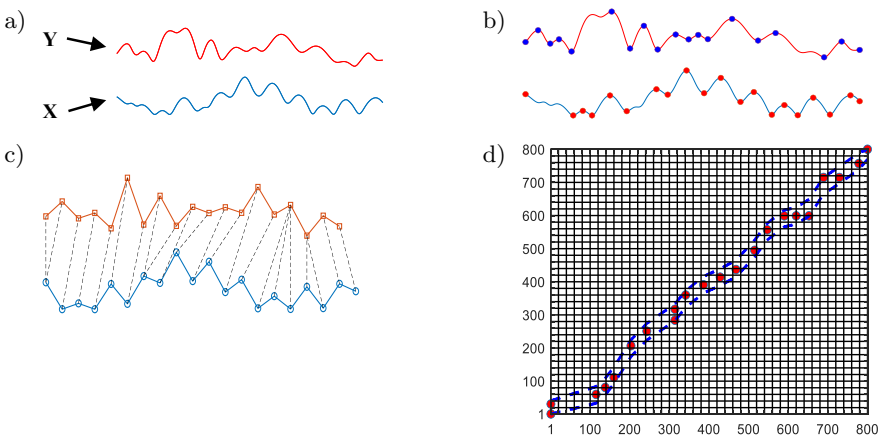


Fig. 2. Execution flow of the AEB-DTW algorithm: a) input two time series, b) extraction of significant extreme values, c) adaptive weighted alignment of extreme values, d) constraint boundary generation and distance measurement.

feature sequence (Fig. 2b). These extrema are assigned adaptive weights according to their position and amplitude to guide the warping path (Fig. 2c). Finally, boundary constraints are constructed from the residuals between the matched extrema path and the smoothed trend line (Fig. 2d).

3.1. Extraction of significant extrema

To extract structurally meaningful local extrema from active sonar signal envelopes, we adopt a prominence-based adaptive detection method. For a given sequence $\mathbf{X} = \{x_1, ..., x_k, ..., x_m\}$, each point x_k $(k = 2, 3, ..., m-1)$ is identified as a local maximum if $x_{k-1} < x_k > x_{k+1}$, or a local minimum if $x_{k-1} > x_k < x_{k+1}$.

To determine whether a local extremum is significant, a dynamic threshold is introduced based on the statistical characteristics of the signal. Let μ, $x_{\min}$, and $x_{\max}$ denote the mean, minimum, and maximum values of $\mathbf{X}$, respectively. A scaling factor γ is defined as

$$\gamma = \frac{\mu - x_{\min}}{x_{\max} - x_{\min}}. \tag{5}$$

Based on this, the prominence threshold δ is computed as

$$\delta = \gamma \cdot \mathrm{MAD}(\mathbf{X}), \tag{6}$$

where $\mathrm{MAD}(\mathbf{X})$ is the mean absolute deviation, which measures average signal fluctuation and is robust to outliers. Local extrema are retained only if their prominence exceeds δ, effectively suppressing noise-induced oscillations and weak fluctuations.

The prominence of a local extremum quantifies how much it stands out from its surroundings. For a maximum, it is defined as the vertical distance between the peak and the lowest point on either side before reaching a higher peak. For a minimum, it is the distance between the trough and the highest point between it and a deeper trough.

Let $\mathbf{X}_E = \{(p_i, a_i)\}_{i=1}^{m'}$ and $\mathbf{Y}_E = \{(q_j, b_j)\}_{j=1}^{n'}$ denote the sequences of significant extrema extracted from $\mathbf{X}$ and $\mathbf{Y}$, respectively, where p_i and q_j are the positions (indices) of the i-th and j-th significant extreme points, and a_i and b_j are the corresponding amplitudes. For two sequences of lengths m and n, the numbers of significant extrema extracted are m' and n', respectively, where $m' \ll m$ and $n' \ll n$. As illustrated in Fig. 3, peaks and troughs correspond to local maxima and minima of the signal. Only extrema exceeding the adaptive saliency threshold are retained, while non-significant extrema caused by noise are excluded. The start and end points are added to the extrema sets $\mathbf{X}_E$ and $\mathbf{Y}_E$ to preserve boundary characteristics. Although some local extrema are not detected (indicated by arrows in Fig. 3), these omissions do not affect the extraction of essential structural features, ensuring algorithmic correctness and stability while enabling dimensionality reduction.

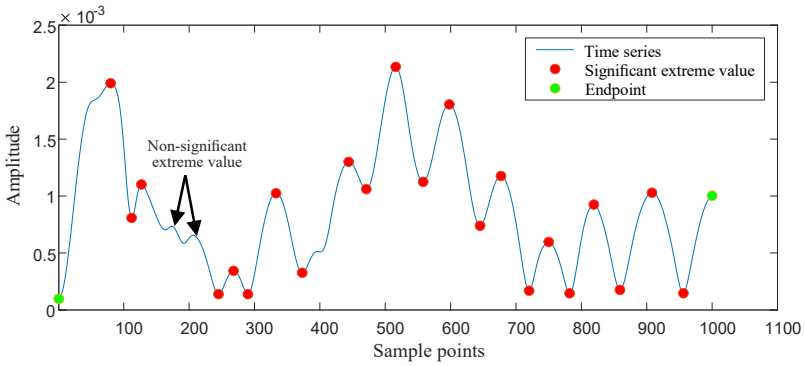


Fig. 3. Extraction process of significant local extrema.

3.2. Adaptive weighted alignment of significant extrema

After extracting significant extrema from both sequences, the next step is to establish reliable correspondences. As key structural features, the accurate alignment of these extrema provides a foundation for reducing the

search space and enhancing the robustness of similarity matching. To this end, a weighted alignment strategy is proposed, which jointly models the positional and amplitude attributes of each extremum to characterize structural correspondences more precisely. Significant extrema are regarded as fundamental matching units, and their fused features are used to construct weighted correspondences. The resulting position mappings serve as the structural basis for generating boundary constraints, as follows:

$$D_E(i,j) = \omega_b \cdot |a_i - b_j| + \omega_a \cdot |p_i - q_j| + \min \begin{cases} D_E(i-1,j), \\ D_E(i,j-1), \\ D_E(i-1,j-1), \end{cases} \tag{7}$$

where $\omega_a = \lambda/(1+r)$ and $\omega_b = 1-\omega_a$ represent the weighting coefficients for positional and amplitude differences, respectively. Here, $\lambda = 0.6$ is a weighting constant used to establish a stable numerical relationship between the two components. The ratio r reflects the relative difference between amplitude and position, and is computed as follows:

$$r = \frac{\Delta p_x + \Delta p_y}{\Delta a_x + \Delta a_y}, \tag{8}$$

where Δp_x, Δa_x, Δp_y, and Δa_y represent, respectively, the position average difference change rate and the amplitude average difference change rate of the extreme value sequences $\mathbf{X}_E$ and $\mathbf{Y}_E$. The average change rates of the position and amplitude of the significant extreme value points are defined as follows:

$$\Delta p = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| \frac{\text{pos}\,(i+1) - \text{pos}\,(i)}{L} \right|, \tag{9}$$

$$\Delta a = \frac{1}{N-1} \sum_{i=1}^{N-1} |\text{value}\,(i+1) - \text{value}\,(i)|, \tag{10}$$

where $L = \lfloor (m+n)/2 \rfloor$ is the normalized length of the time axis, N represents the number of significant extreme value points, $\text{pos}\,(i)$ and $\text{value}\,(i)$ represent the position and amplitude of the i-th significant extreme value point. For $\mathbf{X}_E$, $N = m'$, $\text{pos}\,(i) = p_i$, $\text{value}\,(i) = a_i$. Similarly, for $\mathbf{Y}_E$, $N = n'$, $\text{pos}\,(i) = q_i$, $\text{value}\,(i) = b_i$.

Through the above-mentioned distance measurement, an optimal warping path will be obtained. Let the extremum alignment path be denoted by $\mathbf{W}_E = \{(p_k, q_k)\}_{k=1}^{T}$, where T is the number of aligned pairs in the path; $p_k \in P_X$ (with $P_X = \{p_1, p_2, ..., p_{m'}\}$ being the position set of significant extrema in $\mathbf{X}_E$) and $q_k \in P_Y$ (with $P_Y = \{q_1, q_2, ..., q_{n'}\}$ being the position set of significant extrema in $\mathbf{Y}_E$) are the temporal positions of significant extrema in $\mathbf{X}_E$ and $\mathbf{Y}_E$, respectively. The path $\mathbf{W}_E$ satisfies the standard DTW constraints and is derived by minimizing the weighted distance metric, thus establishing the structural correspondence between the two extremum sequences.

3.3. DTW *with boundary constraints*

To improve efficiency and robustness, AEB-DTW incorporates boundary constraints using matched significant extrema as anchors. Weighted extrema alignments are treated as 2D points, to which a smoothing spline is fitted; residuals define discrete upper and lower boundaries, which are linearly interpolated over the reference sequence. This confines the DTW path to a meaningful region, reducing computational load and misalignments. The extrema path captures global trends, while the residual-based envelope reflects local variations, enabling semantically meaningful alignment under noise. As shown in Fig. 4, AEB-DTW produces fewer spurious matches than the standard DTW.

The construction process of the upper and lower boundaries is as follows: first, the significant extremum matching path $\mathbf{W}_E$ is treated as discrete points in the 2D coordinate space. These points, shown as the red dots in Fig. 5a, have coordinates (p_k, q_k) that denote the positions of significant extrema in sequences $\mathbf{X}$ and $\mathbf{Y}$, respectively. Then, a smoothing spline is applied to these points to obtain the fitted curve $f(k)$. This curve

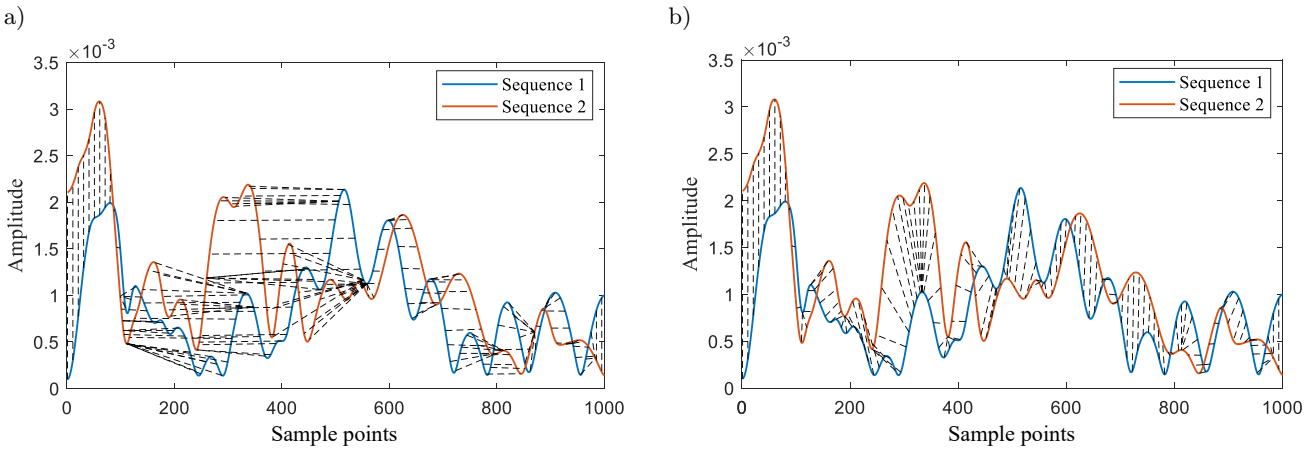


Fig. 4. Comparison of alignment effects between DTW (a) and AEB-DTW (b).

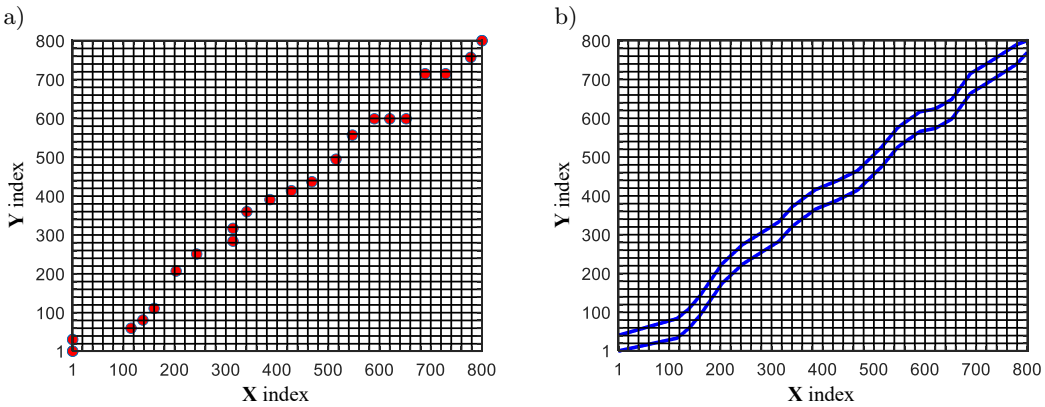


Fig. 5. Formation process of the boundary constraint: a) extreme position alignment path, b) upper and lower boundary constraints.

captures the positional alignment trend between the two sequences and predicts the corresponding position q_k in $\mathbf{Y}$ for each p_k in $\mathbf{X}$.

Subsequently, the absolute residuals between the actual matched points and the fitted curve are calculated, and the 95th percentile of these residuals, denoted as res_{95}, is used as the width of the constraint band. Using this width, discrete upper and lower boundary points are constructed, which are then extended to the full index range of the reference sequence via linear interpolation. This results in the complete boundary functions LB and UB, as illustrated in Fig. 5b.

The construction of the discrete upper and lower boundaries is as follows:

$$\text{UB}(p_k) = \min(m, \max(q_k, f(p_k) + \text{res}_{95})), \tag{11}$$

$$\text{LB}(p_k) = \max(1, \min(q_k, f(p_k) - \text{res}_{95})). \tag{12}$$

At this stage, the upper and lower boundary functions are only defined at the discrete points p_k and have not yet covered the entire range of the reference sequence. To address this, linear interpolation is applied to the sets of boundary points $(p_k, \text{LB}(p_k))$ and $(p_k, \text{UB}(p_k))$, respectively, resulting in continuous functions that are fully defined over the index range of the reference sequence. The interpolated functions ensure that at each time index i, a valid upper and lower constraint boundary is provided. Accordingly, the DTW path search region is finally constrained within the following area:

$$B = \{(i,j) \,|\text{LB}(i) \le j \le \text{UB}(i),\ i \in [1,m]\}. \tag{13}$$

Within this region, the DTW alignment path P must satisfy the standard DTW conditions of boundary, monotonicity, and continuity. The AEB-DTW distance is finally computed as

$$\text{AEB} - \text{DTW}(\mathbf{X}, \mathbf{Y}) = \min_{P \subseteq B} \sum_{(i,j) \in P} d(x_i, y_j), \tag{14}$$

where $d(x_i, y_j)$ is the local distance function between points x_i and y_j, as defined in the standard DTW. This boundary-constrained DTW framework not only significantly reduces computational overhead, but also improves alignment reliability by preventing implausible matches.

4. Performance validation based on measured data

4.1. Overview of measured data

The lake trial was conducted on October 25, 2024, at the Moganshan Reservoir in Deqing County, Huzhou City, Zhejiang Province. The test target was a submarine-like model. The transmitted signal was a linear frequency modulated (LFM) pulse with a frequency range of 10 kHz to 20 kHz, a pulse width of 4 ms, and a sampling rate of 500 kHz.

As shown in Fig. 6a, the experiment employed a bistatic sonar configuration, with the transmitter and hydrophone separated by 7 m, and the target located 14 m from the hydrophone, all positioned in the same horizontal plane at a water depth of approximately 5 m. Figure 6b shows a photograph of the experimental data acquisition system. The dry-end electronic equipment on the transmitting side includes a power amplifier, a signal source, and an oscilloscope, while that on the receiving side includes a data acquisition unit and a computer. The transmitting device is a transmitting transducer, and the receiving device is a receiving hydrophone.

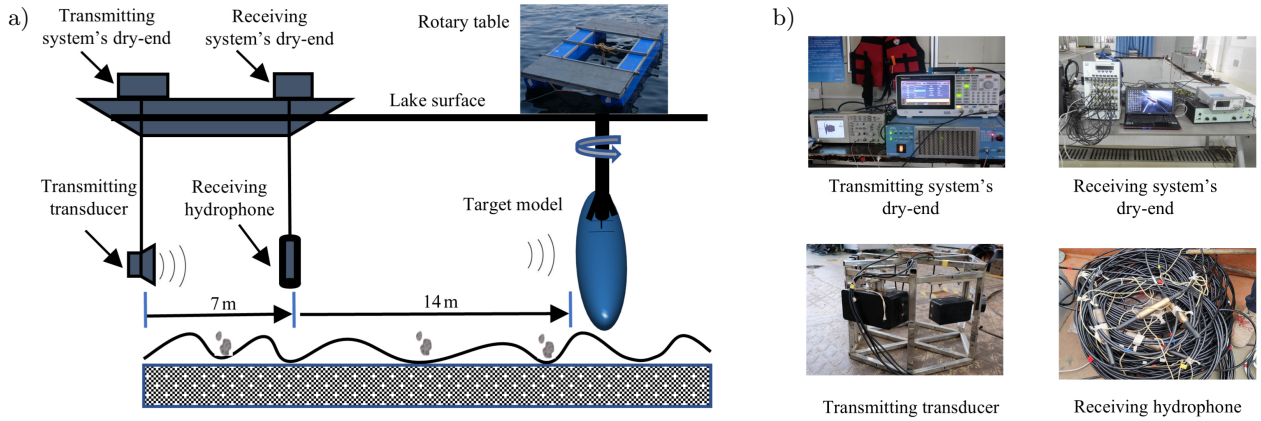


Fig. 6. Experimental setup (a) and equipment configuration (b).

The target was mounted on a servo-driven platform rotating uniformly counterclockwise. The platform completed a 180° turn in about 3 min; with a pulse repetition interval of 0.6 s, this corresponds to a rotation speed of approximately 1 °/s and an azimuth step of about 0.6° per transmission. Consequently, the 300 recorded echoes form a full 180° azimuth scan of the target, with each echo corresponding to a fixed angular increment relative to the transmitter–receiver baseline.

The signal processing procedure is illustrated in Fig. 7a. The original signal refers to the segment extracted from the raw measurements that contains the target echo. Sparse reconstruction is first applied to this segment to enhance signal quality. The envelope of the reconstructed signal is then obtained to characterize the overall amplitude variation and improve stability, providing a reliable input for subsequent matching. Figure 7b shows the matched-filtering results of reconstructed signals from different bearing angles. The matched-filter responses reveal the target's attitude variations with bearing, thereby supporting the rationality of the subsequent matching results.

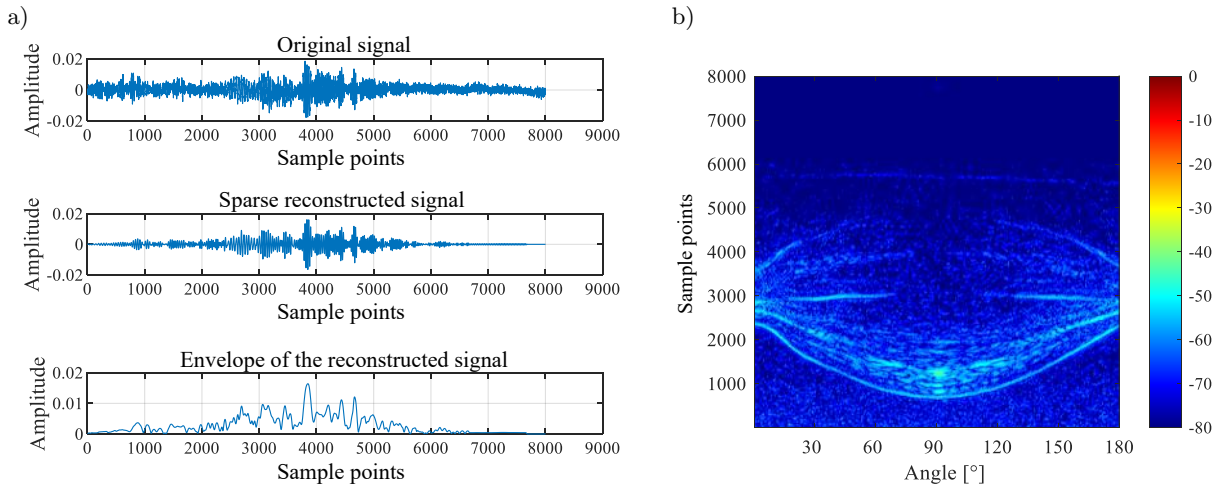


Fig. 7. Signal processing flow.

4.2. Validation of matching accuracy and anti-interference capability

To evaluate the robustness of the algorithm under different interference conditions, echo signals with specified SRRs were synthesized from sparsely reconstructed lake-trial echoes. Each measured echo was first processed by the sparse reconstruction procedure described in Subsec. 4.1. The main reflection of a high signal-to-noise ratio (SNR) reconstructed echo was then used as the clean target signal, and a late-time segment of the same reconstructed record containing diffuse reverberation served as the reverberation component. SRR values from 0 dB to 30 dB in 2 dB steps were obtained by linearly superimposing these two signals and scaling the reverberation amplitude to the desired level. This preserves the measured reverberation characteristics while enabling precise SRR control. Figure 8 illustrates one example of the construction process, showing the reconstructed target echo,

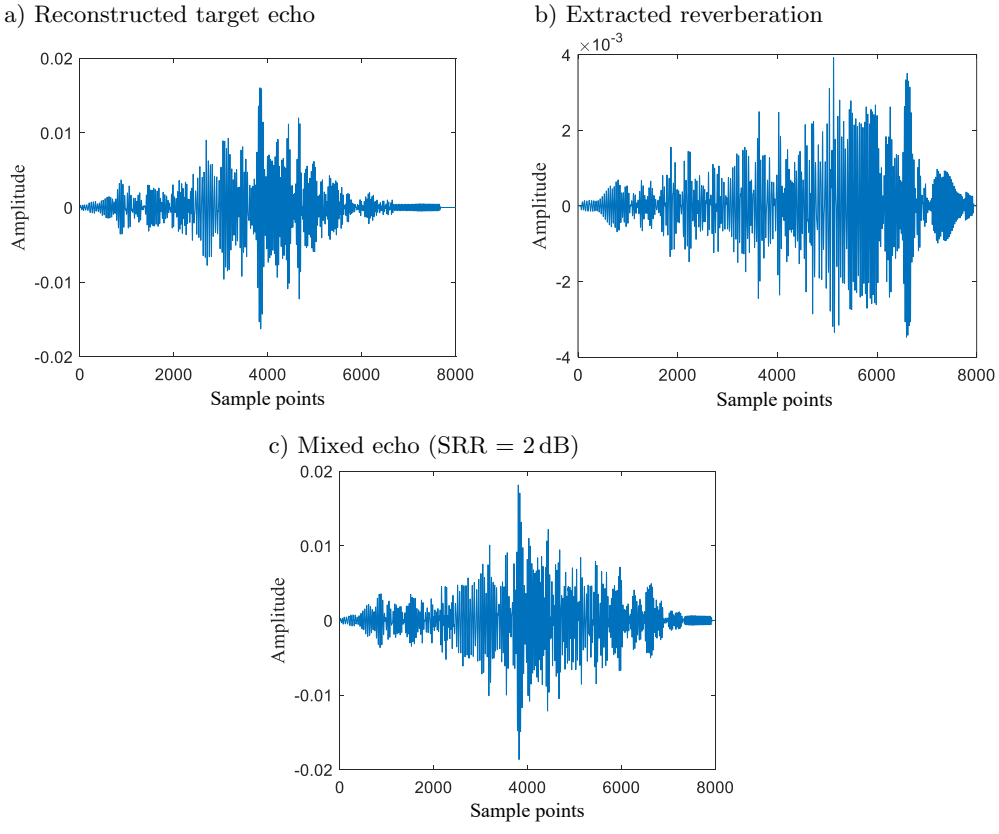


Fig. 8. Waveform example of a sonar mixed echo (SRR = 2 dB).

the extracted reverberation segment, and the synthesized mixed echo. Datasets at other SRR levels are generated in the same way.

Using this method, 16 datasets were constructed with SRRs ranging from 0 dB to 30 dB in 2 dB increments. For each dataset, 50 test samples with known azimuth angles were uniformly selected from the target's 0° to 180° rotation range, with an angular interval of approximately 3.6°. During the similarity matching phase, each test sample was compared with all other samples within the same SRR dataset using a distance-based metric. The sample yielding the minimum cumulative distance was defined as the 'optimal matching sample,' and the matching error was calculated as the absolute azimuth difference between the two samples. A match was regarded as correct if the azimuth error was less than 10°. This threshold is motivated by the observation that, for the considered target and measurement geometry, the echo characteristics vary only weakly over azimuth intervals of about ±5°, so angular errors within 10° do not lead to perceptible differences in the envelope features used for matching and are acceptable in practical active sonar applications.

The matching accuracy under each SRR condition was defined as the ratio of correctly matched samples to the total number of test samples, serving as a quantitative indicator of the algorithm's robustness to reverberation. Table 1 summarizes the comparative results for AEB-DTW against ED, DTW, and its variants. If two or more methods achieve the same maximum value under a specific SRR condition, each is credited with one win in the overall count. AEB-DTW achieves over 50 % accuracy at SRR = (0–2) dB and exceeds 80 % when SRR ≥ 24 dB, consistently outperforming all baselines in most conditions.

Table 1. Comparison of matching accuracy for different algorithms.

SRRs	AEB-DTW	DTW	DDTW	WDTW	Sakoe–Chiba	LEDTW	EWDTW	ED	shapeDTW
0	0.54	0.44	0.2	0.54	0.42	0.32	0.36	0.58	0.42
2	0.68	0.4	0.34	0.44	0.42	0.28	0.4	0.52	0.32
4	0.62	0.46	0.4	0.46	0.48	0.34	0.54	0.68	0.46
6	0.7	0.48	0.44	0.54	0.52	0.36	0.44	0.68	0.42
8	0.72	0.46	0.38	0.64	0.52	0.42	0.52	0.7	0.32
10	0.76	0.44	0.36	0.6	0.44	0.36	0.74	0.72	0.46
12	0.74	0.68	0.42	0.72	0.66	0.48	0.7	0.76	0.46
14	0.72	0.74	0.34	0.78	0.80	0.5	0.78	0.8	0.68
16	0.76	0.66	0.42	0.68	0.7	0.4	0.8	0.78	0.66
18	0.8	0.78	0.48	0.78	0.74	0.44	0.86	0.7	0.7
20	0.82	0.7	0.44	0.72	0.66	0.44	0.76	0.78	0.74
22	0.78	0.64	0.4	0.72	0.66	0.5	0.78	0.78	0.7
24	0.84	0.7	0.46	0.82	0.64	0.5	0.78	0.78	0.74
26	0.86	0.64	0.46	0.82	0.6	0.46	0.76	0.8	0.72
28	0.82	0.74	0.48	0.78	0.7	0.46	0.76	0.8	0.72
30	0.84	0.7	0.48	0.8	0.74	0.54	0.76	0.76	0.68
Mean	0.750	0.603	0.406	0.677	0.605	0.425	0.671	0.726	0.575
Wins	11	0	0	0	1	0	3	3	0

To provide a more intuitive comparison of matching accuracy under different SRR conditions, Fig. 9 presents pairwise comparisons between AEB-DTW and ED, DTW, and its variants. Each subplot contains 16 circular markers, representing the matching accuracy under 16 distinct SRR levels. The horizontal and vertical coordinates of each marker correspond to the accuracies of AEB-DTW and the compared algorithm, respectively. Markers on the diagonal line indicate equal accuracy under a given SRR; those below the diagonal indicate that AEB-DTW outperforms the baseline method, while those above indicate that the baseline method achieves higher accuracy than AEB-DTW.

The overall distribution of these points indicates that AEB-DTW tends to deliver superior performance across the tested reverberation conditions. As shown in Fig. 9, it demonstrates higher robustness to reverberation than the other evaluated algorithms.

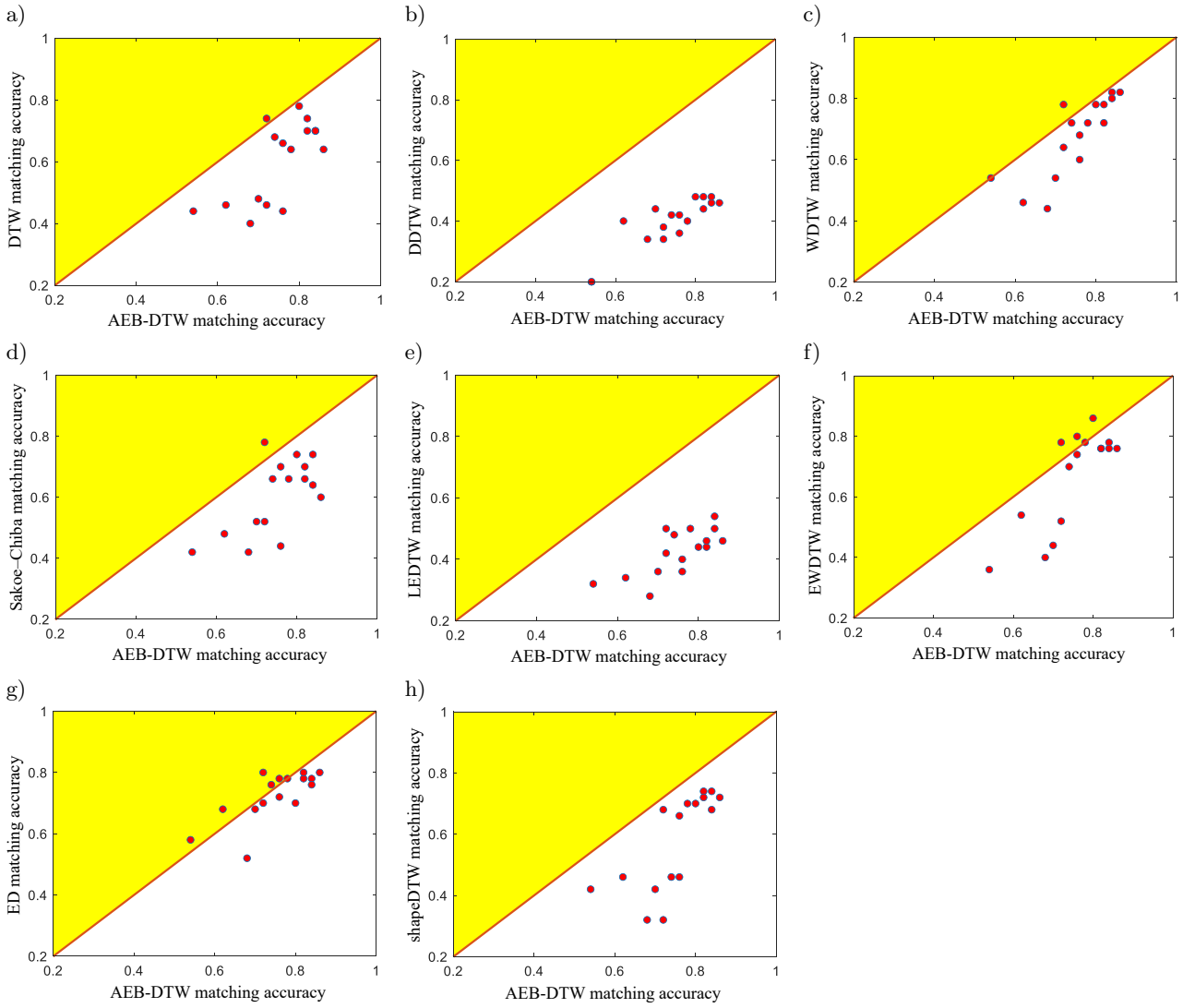


Fig. 9. Matching accuracy of AEB-DTW vs. ED, DTW, and its variants.

4.3. Evaluation of similarity measurement based on angular resolution

In active sonar signal recognition, angular resolution measures an algorithm's ability to distinguish echoes from different azimuths. To evaluate this, a dataset with an SRR of 24 dB was used. The 90° echo, representing the median azimuth and exhibiting stable structural features, was selected as the reference. For each algorithm, distances between the reference and other azimuth samples were computed and normalized to the range [0, 1]. The normalized distances were then divided into nine 20° intervals to balance angular granularity and statistical robustness. Within each interval, the mean normalized distance was calculated, and its absolute deviation from 1 was taken as the interval similarity, providing an intuitive measure of signal discriminability-larger deviations indicate stronger angular separability. Interval-wise similarity profiles for all algorithms are shown in Fig. 10.

DTW shows poor angular resolution due to minimal similarity variation, while DDTW, being insensitive to amplitude, produces abnormally high similarity in low-angle regions. WDTW and Sakoe–Chiba DTW are constrained by fixed weights or rigid boundaries, limiting their performance in high-angle scenarios. ED and EWDTW offer only basic angular discrimination, as ED is restricted to equal-length sequences, and EWDTW relies heavily on hyperparameter tuning. LEDTW and shapeDTW also exhibit limited sensitivity to angular differences. In contrast, AEB-DTW effectively distinguishes signals with large angular separations while maintaining robustness to local variations. Its adaptive extrema extraction and boundary constraints enable smoother similarity transitions across angles, thereby enhancing angular resolution.

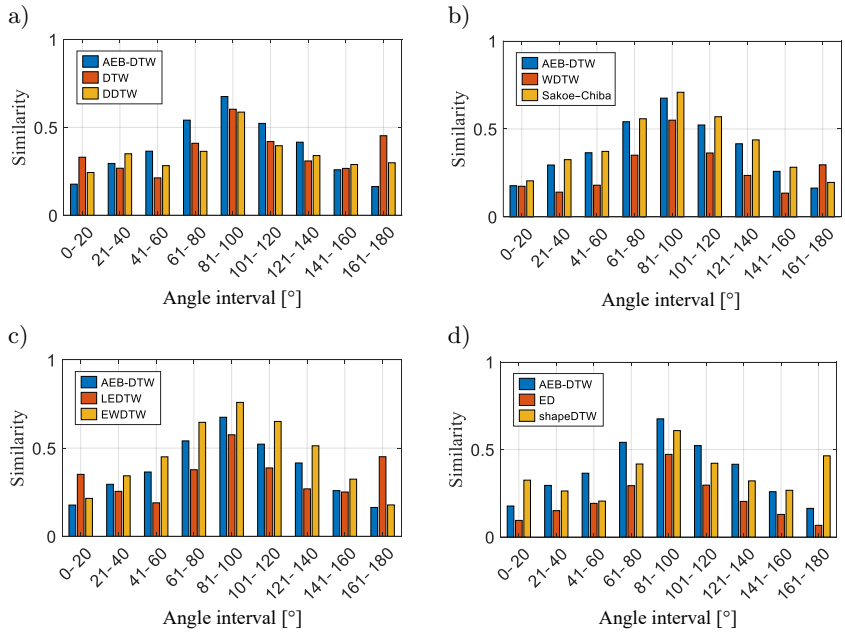


Fig. 10. Angular resolution comparison of AEB-DTW and baseline methods.

4.4. Comparison of algorithm runtime

Figure 11 shows the comparison of average runtime between AEB-DTW and ED, DTW, and its variants across 16 SRR datasets. Although ED has the shortest runtime, it is only applicable to sequences of equal length and cannot handle asynchronous time axes. LEDTW and EWDTW only account for extrema features without considering structural relationships between extrema, thus lacking global alignment capacity. Among algorithms that support global asynchronous alignment, AEB-DTW demonstrates the best runtime efficiency, verifying its computational advantages.

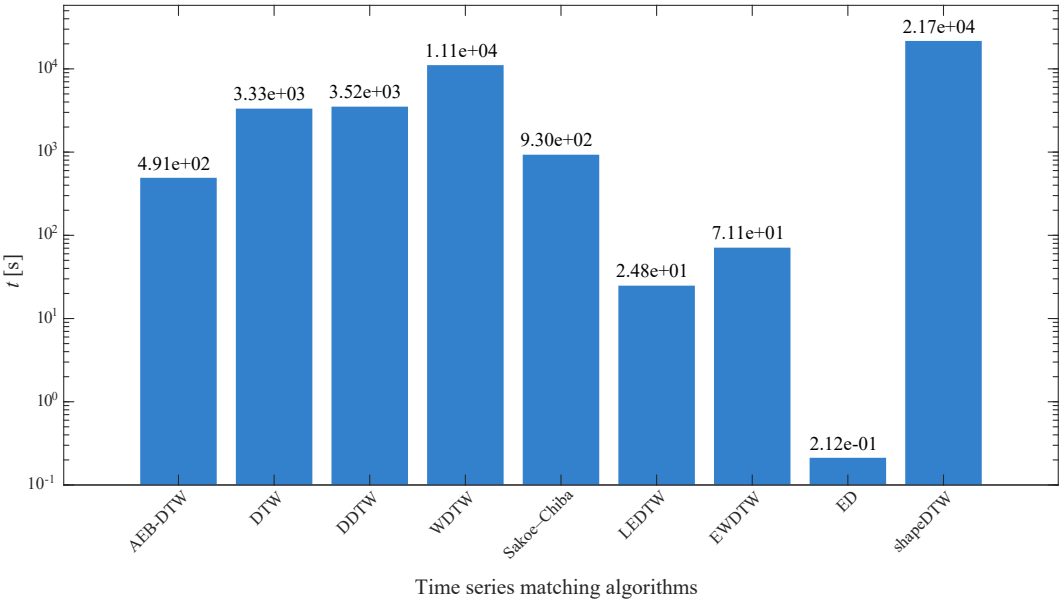


Fig. 11. Average running time comparison.

AEB-DTW shows a clear runtime advantage over other global alignment algorithms, as indicated by its lower average computation time in Fig. 11. Although its runtime is slightly higher than LEDTW and EWDTW, it remains about one order of magnitude lower than DTW and achieves higher matching accuracy (see Table 1), demonstrating a favorable balance between efficiency and accuracy.

5. Conclusion

This paper proposes the AEB-DTW algorithm to mitigate the sensitivity of sonar signal matching to reverberation and the high computational cost of long-duration echoes. The method extracts salient local extrema using an adaptive threshold and aligns them via a joint time–amplitude weighting strategy. Dynamic boundary constraints derived from the extrema alignment path restrict the DTW search region, improving matching efficiency and robustness to structural distortions.

The main findings of the experimental evaluation can be summarized as follows:

1. Compared with classical distance-based methods, AEB-DTW achieves higher matching accuracy and stronger robustness to interference; under strong interference conditions with an SRR value of 2 dB, the algorithm still achieves around 70 % matching accuracy, demonstrating strong resilience to reverberation.
2. Angular-resolution-based similarity evaluation shows that AEB-DTW is more sensitive to azimuth changes and can effectively distinguish echoes originating from different bearing angles, thereby improving the angular discrimination capability of active sonar echo matching.
3. Among algorithms that support global asynchronous alignment, AEB-DTW exhibits the shortest runtime, confirming its significant computational efficiency and indicating that it can reduce computational cost while maintaining high matching performance.

Overall, the algorithm improves both matching accuracy and angular discrimination ability while reducing computational cost, providing an efficient and reliable solution for active sonar echo matching and 1D time-series alignment. Future work will explore multi-target scenarios, nonlinear target motion, variable acoustic environments, and the integration with deep learning methods to further enhance adaptability and recognition performance.

Fundings

This work was supported by the Joint National Natural Science Foundation of China (no. U22A2044) and the Key Laboratory Fund from Underwater Test and Control Technology (no. 2023-JCJQ-LB-030).

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.

Authors' contributions

Tongjing Sun: data curation, formal analysis, investigation, resources, writing – review and editing, funding acquisition, supervision; Hunyuan Wang: conceptualization, data curation, investigation, methodology, writing – original draft; Lei Chen: investigation, methodology, validation; Haoran Xu: data curation, formal analysis, writing – review and editing. All authors reviewed and approved the final manuscript.

Acknowledgments

This article uses experimental data collected at Shanghai Jiaotong University (China), and we gratefully acknowledge our colleagues for their experimental skills.

References

1. Bian J.C., Xue S.Q., Zhao S., Xiao Z., Gao J.L. (2024), Accuracy analysis of active sonar navigation for different underwater beacon configurations, *Navigation and Control*, **23**(Z1): 78–83, https://doi.org/10.3969/j.issn.1674-5558.2024.h5.008.

2. Chen H.Y., Liu C.H., Sun B. (2017), Survey on similarity measurement of time series data mining [in Chinese], *Control and Decision*, **32**(1): 1–11, https://doi.org/10.13195/j.kzyjc.2016.0462.
3. Dau H.A. *et al.* (2018), Optimizing dynamic time warping's window width for time series data mining applications, *Data Mining and Knowledge Discovery*, **32**: 1074–1120, https://doi.org/10.1007/s10618-018-0565-y.
4. Hong J.Y., Park S.H., Baek J.-G. (2020), SSDTW: Shape segment dynamic time warping, *Expert Systems with Applications*, **150**: 113291, https://doi.org/10.1016/j.eswa.2020.113291.
5. Jeong Y.-S., Jeong M.K., Omitaomu O.A. (2011), Weighted dynamic time warping for time series classification, *Pattern Recognition*, **44**(9): 2231–2240, https://doi.org/10.1016/j.patcog.2010.09.022.
6. Keogh E., Ratanamahatana C.A. (2005), Exact indexing of dynamic time warping, *Knowledge and Information Systems*, **7**: 358–386, https://doi.org/10.1007/s10115-004-0154-9.
7. Keogh E.J., Pazzani M.J. (2001), Derivative dynamic time warping, [in:] *Proceedings of the 2001 SIAM International Conference on Data Mining*, pp. 1–11, https://doi.org/10.1137/1.9781611972719.1.
8. Kim S.-W., Park S., Chu W.W. (2001), An index-based approach for similarity search supporting time warping in large sequence databases, [in:] *Proceedings 17th International Conference on Data Engineering*, pp. 607–614, https://doi.org/10.1109/ICDE.2001.914875.
9. Lahreche A., Boucheham B. (2021), A fast and accurate similarity measure for long time series classification based on local extrema and dynamic time warping, *Expert Systems with Applications*, **168**: 114374, https://doi.org/10.1016/j.eswa.2020.114374.
10. Li H. (2021), Time works well: Dynamic time warping based on time weighting for time series data mining, *Information Sciences*, **547**: 592–608, https://doi.org/10.1016/j.ins.2020.08.089.
11. Li H.H., Liu J.X., Yang Z.L., Liu R.W., Wu K.F., Wang Y. (2020), Adaptively constrained dynamic time warping for time series classification and clustering, *Information Sciences*, **534**: 97–116, https://doi.org/10.1016/j.ins.2020.04.009.
12. Li H.L., Guo C.H. (2013), A review of research on feature representation and similarity measurement in time series data mining [in Chinese], *Computer Application Research*, **30**(5): 1285–1291, https://doi.org/10.3969/j.issn.1001-3695.2013.05.002.
13. Sakoe H., Chiba S. (1978), Dynamic programming algorithm optimization for spoken word recognition, *IEEE Transactions on Acoustics, Speech, and Signal Processing,* **26**(1): 43–49, https://doi.org/10.1109/TASSP.1978.1163055.
14. Salvador S., Chan P. (2007), Toward accurate dynamic time warping in linear time and space, *Intelligent Data Analysis: An International Journal*, **11**(5): 561–580, https://doi.org/10.3233/IDA-2007-11508.
15. Silva D.F., Giusti R., Keogh E., Batista G.E.A.P.A. (2018), Speeding up similarity search under dynamic time warping by pruning unpromising alignments, *Data Mining and Knowledge Discovery*, **32**: 988–1016, https://doi.org/10.1007/s10618-018-0557-y.
16. Sun S., Qiu J.X., Liu Z., Wang J.Y., Wang J.H. (2024), A method for sonar pulse matching recognition based on time-frequency matrix, *Journal of Electroacoustics*, **48**(11): 64–69, https://doi.org/10.16311/j.audioe.2024.11.020.
17. Tang W.F., Gao C.F. (2023), Dynamic time warping algorithm with adaptive weighting of extreme points, *Computer Engineering*, **49**(7): 150–160, https://doi.org/10.19678/j.issn.1000-3428.0065500.
18. Wang C., Long Y.W., Yin W.H., Huang B. (2022), Research on distance measurement method based on improved DTW lower bound function, *Computer Engineering and Applications*, **58**(23): 316–326, https://doi.org/10.3778/j.issn.1002-8331.2106-0036.
19. Yi B.-K., Jagadish H.V., Faloutsos C. (1998), Efficient retrieval of similar time sequences under time warping, [in:] *Proceedings 14th International Conference on Data Engineering*, pp. 201–208, https://doi.org/10.1109/ICDE.1998.655778.
20. Zhang M.Y., Cai W.Y., Wang Y.H., Zhu J.F. (2023), A level set method with heterogeneity filter for side-scan sonar image segmentation, *IEEE Sensors Journal*, **24**(1): 584–595, https://doi.org/10.1109/JSEN.2023.3334765.
21. Zhao J.P., Itti L. (2018), shapeDTW: Shape dynamic time warping, *Pattern Recognition*, **74**: 171–184, https://doi.org/10.1016/j.patcog.2017.09.020.
22. Zhu Z., Tao C., Wu T., Schneider von Dämeling J., Zhang J., Zhang G. (2025), Seafloor classification by fusing AUV acoustic and magnetic data: Towards complex deep-sea environments, *IEEE Transactions on Geoscience and Remote Sensing*, **63**: 1–15, https://doi.org/10.1109/TGRS.2025.3545128.

Archives of Acoustics Vol. 51, No. 1, pp. 145–159 (2026), https://doi.org/10.24423/archacoust.2026.4275

Research Paper

Study on Dual-Stage Amplification Cascaded Piezoelectric Transducer for High-Power Applications

Guo Li[(1)]*, Peiyu Tan[(1)], Ruihui Ma[(1)], Feilong Li[(1)], Xiaoli Zhang[(2)], Hua Tian[(3)]

[(1)] *School of Automation, Xi'an Key Laboratory of Advanced Control and Intelligent Processing, Xi'an University of Posts & Telecommunications*
Xi'an, China

[(2)] *Department of Electronic & Information Engineering, Ankang University*
Ankang, Shaanxi, China

[(3)] *Key Laboratory of Ultrasound of Shaanxi Province, School of Physics and Information Technology, Shaanxi Normal University*
Xi'an, China

*Corresponding Author: liguo@xupt.edu.cn

Received July 10, 2025; revised December 25, 2025; accepted January 1, 2026; available online January 15, 2026; version of record March 17, 2026; published issue March 27, 2026.

To reduce the size and enhance the efficiency of cascaded sandwich transducers with conical horns, a novel structural configuration of such transducers is investigated. This transducer incorporates two sets of piezoelectric stacks, enabling two-stage amplification to improve efficiency. An equivalent circuit model of the cascaded sandwich transducer with a conical horn is established, and analytical expressions for key performance parameters, including input impedance, velocity amplification ratio, and resonant characteristics, are systematically derived. Through theoretical and simulation analyses, the dynamic influence of key structural parameters on electromechanical energy conversion efficiency is determined, specifically including the output radius of the second stage, the relative position of the variable cross-sections of two sets of piezoelectric ceramic sandwich structures, and the spacing between the two sets of piezoelectric stacks. Furthermore, a performance optimization strategy based on piezoelectric single-crystal materials is proposed. Numerical simulation results, validated against the theoretical models, reveal the governing principles of piezoelectric material properties on transducer performance. Experimental results demonstrate excellent agreement between the operational characteristics of the optimized transducer and predictions obtained from both theoretical models and finite element simulations. This work provides significant guidance for the optimization of multi-mode transducers and demonstrates promising application potential in high-power ultrasonic fields.

Keywords: cascaded transducer, multi-mode, velocity amplification ratio, finite element analysis.

1. Introduction

In recent years, power ultrasonic technology has demonstrated extensive application prospects across diverse fields, including ultrasonic machining, precision cleaning, underwater acoustics, sonochemistry, food processing, and biodiesel production (Shah, Liu, 2019; Tan *et al.*, 2022; Li *et al.*, 2018b; Pokhrel *et al.*, 2016; Subhedar, Gogate, 2015; Chemat *et al.*, 2011). Particularly, within ultrasonic motor research, piezoelectric transducer-based ultrasonic systems have attracted significant scientific attention (Li *et al.*, 2023b; Roland, 2024; Jiang *et al.*, 2024). As critical components of ultrasonic systems, sandwich piezoelectric ultrasonic transducers exhibit remarkable advantages in electromechanical energy conversion through the piezoelectric effect. These advantages are a simplified structural design, high conversion efficiency, and large vibration amplitude, making them predom-

inant in high-power ultrasonic applications (Lin, 2009; Kuang *et al.*, 2014; Du *et al.*, 2020; Davari *et al.*, 2012). Their applications further extends to piezoelectric nanopositioning systems and robotic manipulators (Su *et al.*, 2018; Jiang *et al.*, 2018; Li *et al.*, 2023a).

To meet escalating demands in high-power applications, research focus has shifted toward optimizing conventional sandwich transducers and developing novel vibration systems. Enhanced radiation coverage and ultrasonic intensity have been achieved through geometric modifications and vibration mode innovations, leading to the proposal of coupled-vibration transducers and mode-conversion composite systems, including longitudinal-torsional, radial-torsional, flexural-torsional, longitudinal-flexural, radial-flexural, and interdigital configurations (Li *et al.*, 2018a; 2023a; Meshkinzar, Al-Jumaily, 2021; Watanabe *et al.*, 1993; Lin *et al.*, 2011; 2013; Tressler *et al.*, 2006; Lin, 2007; Tang *et al.*, 2024; Xu *et al.*, 2019; Hunter *et al.*, 2008). Lin (2017) pioneered the concept of cascaded piezoelectric transducers, where multiple half-wavelength sandwich transducers synergistically amplify input electrical power and ultrasonic intensity. Subsequent work by Lin *et al.* (2018) introduced tunable longitudinal transducers with comprehensive performance characterization. Li *et al.* (2022) conducted a multifrequency analysis of stepped conical transducers, deriving equivalent impedance models and resonance frequency equations through theoretical modeling and experimental validation. Cascaded transducers can operate at different resonant frequencies to meet the requirements of multifrequency modes. This cascaded vibration structure can also double the power and intensity of ultrasonic waves, while the intermediate mass block can assist in heat dissipation. However, systematic investigation of the multi-mode vibration characteristics – crucial for optimizing energy transfer and operational stability – of conical cascaded sandwich transducers remains limited. Specifically, the correlations between the multifrequency performance and key geometric parameters are not well understood, hindering their optimized design for high-power, multi-mode applications.

To address this gap, this study investigates the relationship between performance parameters, geometric dimensions, and piezoelectric materials of a conical cascaded sandwich piezoelectric transducer operating in multifrequency mode. A theoretical analysis model is established by deriving the frequency equations and vibration velocity amplification ratio expression for cascaded transducers using Kirchhoff's law. Keeping the conical amplitude rod structure unchanged, the geometric dimensions of the piezoelectric stack radius near the output end of the amplitude rod and the spatial configuration of the sandwich structure formed by the two piezoelectric stacks are varied, and the piezoelectric material PZT-8 is replaced with a relaxor ferroelectric single crystal (RFSC). The impact of combining PZT-8 with an RFSC on vibration performance are analyzed, and variations in key performance parameters are analyzed using theoretical models and numerical simulations. Based on the theoretical and simulation results, an optimized transducer prototype is manufactured and experimentally tested. The results are in good agreement with the predictions and thus validate the proposed design method. This work provides a practical optimization framework for cascaded transducers with conical horn and offers important guidance for the development of efficient, high-power ultrasonic systems for precision machining, sonochemistry, and other applications.

2. Theoretical analysis

The ultrasonic composite transducer is formed by the longitudinal cascade of two sandwich-type structures, a conical horn, and a prestressed bolt. These two sandwich structures serve as two-stage excitation mechanisms. In the design, the influence of bolt prestress is neglected; consequently, the two sandwich structures and the conical horn can be considered as a solid rod bonded together. Figure 1 illustrates the structural diagram of the ultrasonic transducer, comprising seven components: back metal cylinder I, piezoelectric stack II, middle metal cylinder III, middle metal cylinder IV, piezoelectric stack V, front metal cylinder VI, and conical horn VII. The parameter L_i (i = 1, 2, 3, 4, 5, 6, 7) corresponds to the length of each component, while the parameter R_i (i = 1, 2, 3, 4, 5, 6, 7) corresponds to the cross-sectional radius of each component. N_i (i = 1, 2) represents the electromechanical conversion coefficient of piezoelectric ceramic stack.

The equivalent circuit diagram of the transducer is presented in Fig. 2. Here, V denotes the excitation voltage, I_1 and I_2 represent the currents entering into the two piezoelectric stacks, respectively; v_i (i = 1, 2,

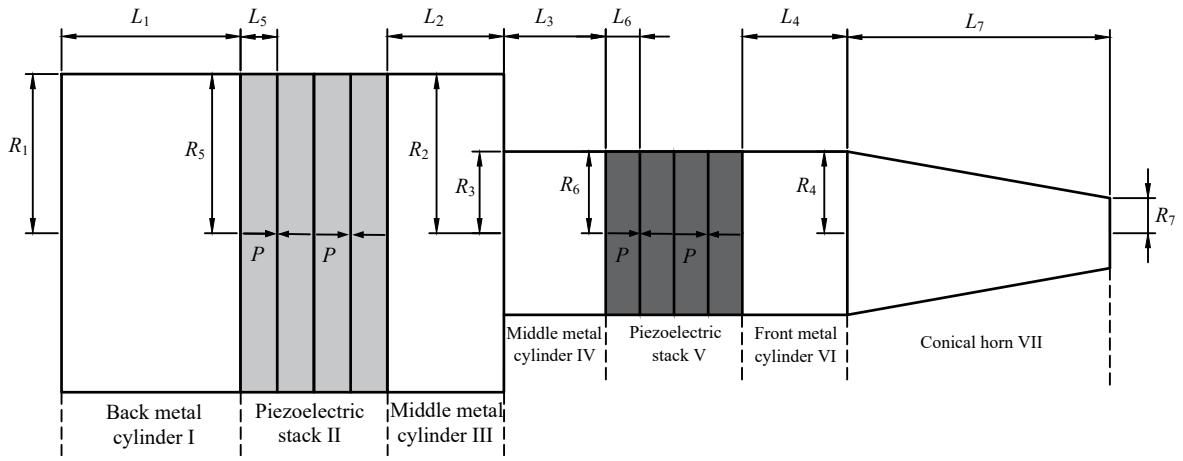


Fig. 1. Structural diagram of the cascaded-excitation longitudinal output system.

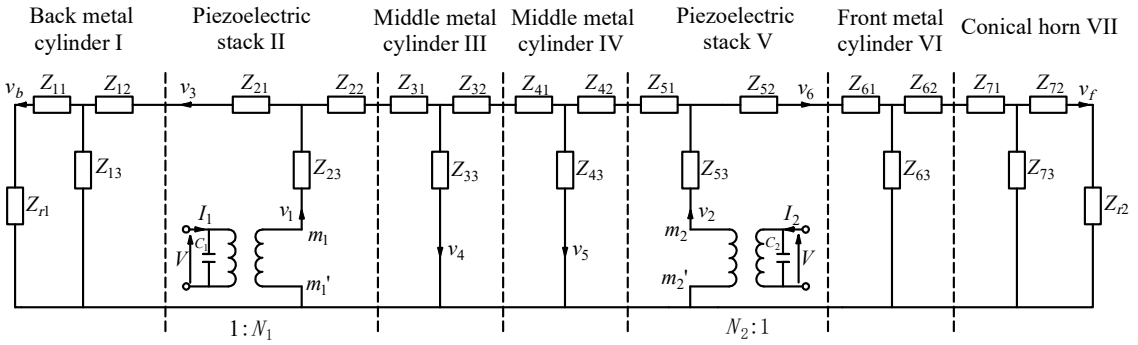


Fig. 2. Equivalent circuit of the cascaded-excitation longitudinal output system.

3, 4, 5, 6) denotes the vibration velocity at each designated section, while v_b and v_f denote the longitudinal vibration velocities at the outer surfaces of the rear metal cylinder and the conical horn, respectively; C_i represents the clamped capacitance of the piezoelectric stacks, N_i is the electromechanical transformation ratio, p is the number of piezoelectric rings, ρ is the material density, E is Young's modulus, k is the wavenumber, and c is the longitudinal wave sound velocity; Z_{i1}, Z_{i2}, Z_{i3} (i = 1, 2, 3, 4, 5, 6, 7) denote the impedances of the metal cylinders and piezoelectric ceramic stacks, respectively, where $Z_{i1} = Z_{i2} = jZ_{0i}\tan\left(\frac{k_i L_i}{2}\right)$, $Z_{i3} = \frac{Z_{0i}}{j\sin(k_i L_i)}$.

Li *et al.* (2024) conducted research on the equivalent circuit of the transducer. From this model, the mechanical impedances Z_{m1} and Z_{m2} between the mechanical terminals m_1, m_1' and m_2, m_2' can be obtained, leading to the derivation of the total equivalent mechanical impedance Z_m:

$$Z_{m1} = Z_{23} + \frac{v_3}{v_1}\left[\frac{(Z_{r1}+Z_{11})Z_{13}}{(Z_{r1}+Z_{11})+Z_{13}} + Z_{12} + Z_{21}\right], \tag{1}$$

$$Z_{m2} = Z_{53} + \frac{v_6}{v_2}\left[\frac{(Z_{71}+Z_{62})(Z_{72}+Z_{73})Z_{32}+Z_{72}Z_{73}}{(Z_{71}+Z_{62}+Z_{63})(Z_{72}+Z_{73})+Z_{72}Z_{73}} + Z_{61} + Z_{52}\right], \tag{2}$$

$$Z_m = \frac{Z_{m1}Z_{m2}}{Z_{m1}(jwC_2Z_{m2}+N_2^2)+Z_{m2}(jwC_1Z_{m1}+N_1^2)}, \tag{3}$$

where $w = 2\pi f$ and f is the frequency.

The resonant frequency f_r is obtained when Z_m approaches zero, and the anti-resonant frequency f_a is calculated when Z_m approaches infinity. The effective electromechanical coupling coefficient is $k_{\text{eff}} = \sqrt{\left(1-\left(\frac{f_r}{f_a}\right)^2\right)}$, and the velocity amplitude magnification ratio is $M = \left|\frac{v_b}{v_f}\right|$.

3. Relationships between performance parameters and geometric dimensions of the cascaded piezoelectric transducer with conical horn

To investigate the influence of the transducers geometric dimensions on its performance, the metal block material is 40Cr steel, while other metal blocks and the conical metal component are made of hard aluminum. The piezoelectric materials employed are PZT-8 and an RFSC. PZT-8 is widely utilized in ultrasonic transducer research due to its advantageous properties, including a high electromechanical coupling coefficient, high mechanical strength, and fast response speed. Compared to PZT-8, the RFSC exhibits higher piezoelectric and dielectric constants along with lower dielectric loss, making it the preferred material for high-precision ultrasonic applications. Their typical characteristic parameters are listed in Table 1. Through theoretical analysis, the function of impedance with respect to frequency can be obtained. By assigning a value to the frequency, the corresponding impedance can be obtained, producing an impedance-frequency curve. From this curve, the resonant and anti-resonant frequencies can be identified, and the electromechanical coupling coefficient can be calculated. By substituting the resonant frequency into the amplification ratio formula, the amplification ratio can be calculated. This theoretical calculation process is implemented using MATLAB. These performance parameters were also obtained through numerical simulation using ANSYS. The model was established in an axisymmetric configuration. The metal components were discretized with PLANE42 elements, a 4-node axisymmetric structural solid element suitable for linear elastic analysis. The piezoelectric ceramic stacks were modeled using PLANE13 elements, a 4-node axisymmetric coupled-field solid element that can directly solve the interaction between electrical and mechanical degrees of freedom. Free meshing with a node size of 0.001 m was applied to each part of the transducer. The effects of pre-tightening bolt stress and structural constraints were neglected during the simulation.

Table 1. Characteristic parameters of the transducer.

Material	ρ [kg/m^3]	E [N/m^2]	v	s_{33}^E [m^2/N]	d_{33} [C/N]	ε_{33}^T [C^2/(N$\cdot$m^2)]	k_{33}
40Cr steel	7840	20.9×10^{10}	0.28	–	–	–	–
Duralumin	2790	7.15×10^{10}	0.34	–	–	–	–
PZT-8	7600	–	–	13.9×10^{-12}	225×10^{-12}	8.84×10^{-9}	0.62
Single crystal	8122	–	–	49×10^{-12}	1285×10^{-12}	42.1×10^{-9}	0.87

3.1. Influence of output radius of the second stage on performance parameters

Using PZT-8 as the piezoelectric material, the initial dimensional parameters of the transducer are: L_1 = 0.02 m, L_2 = L_3 = 0.0075 m, L_4 = 0.021 m, L_5 = L_6 = 0.005 m, L_7 = 0.054 m, R_1 = R_2 = R_3 = R_4 = 0.0255 m, R_5 = R_6 = 0.025 m, R_7 = 0.009 m, p_1 = p_2 = 4, τ_1 = R_2 m. The relationships between the performance parameters and the output radius of the second stage are shown in Fig. 3 to Fig. 5, where the subscripts denote: t for theoretical values, s for simulated values, r for resonance, a for anti-resonance, and the numerical suffix i (i = 1, 2, 3) for the first three longitudinal vibration modes. For example, f_{tr1} refers to the theoretical fundamental resonant frequency, f_{sa2} refers to the simulated second anti-resonant frequency, k_{t1} refers to the theoretical fundamental effective electromechanical coupling coefficient, k_{s2} refers to the simulated second-order effective electromechanical coupling coefficient, M_{t1} refers to the theoretical fundamental velocity amplitude, and M_{s2} refers to the simulated second-order velocity amplitude magnification.

Figure 3 shows that as the τ_1 increases, the fundamental longitudinal vibration resonant and anti-resonant frequencies both increase, while the second-order longitudinal vibration resonant and anti-resonant frequencies decrease. For the third-order longitudinal vibration, the theoretical resonance frequency increases, and the simulated resonance frequency first increases and then decreases, but the changes are very small, whereas the anti-resonant frequency decreases. In Fig. 4, the fundamental longitudinal vibration electromechanical coupling coefficient increases, whereas the effective electromechanical coupling coefficients of the second- and third-order longitudinal vibrations exhibit decreasing trends. Within smaller radius ranges, both fundamental and second-order longitudinal vibrations maintain relatively large effective electromechanical coupling coefficients, with the simulated third-order coupling coefficient showing a distinct minimum. Figure 5 indicates that the fundamental longitudinal

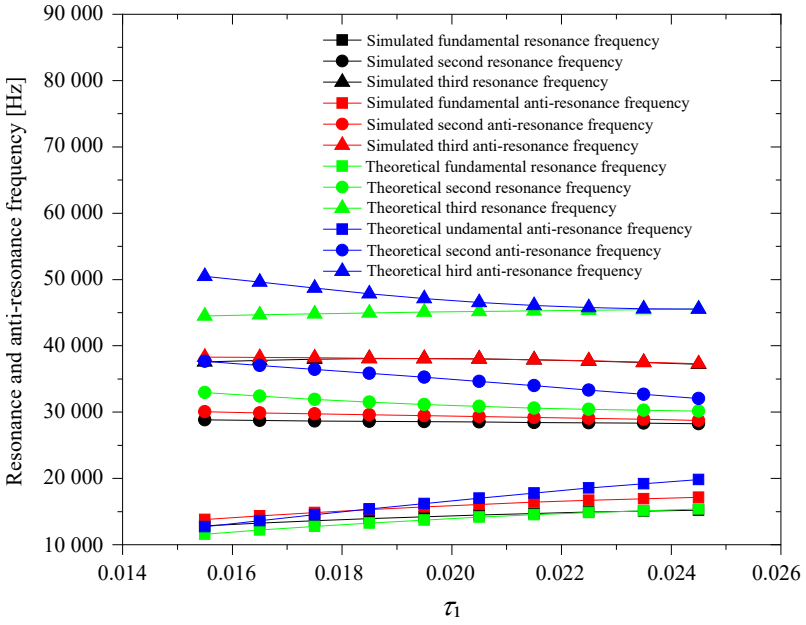


Fig. 3. Relationship between the resonance and anti-resonance frequency and τ_1.

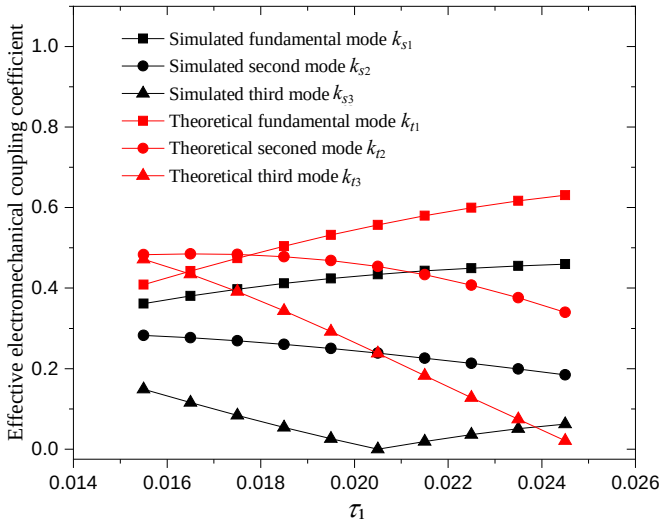


Fig. 4. Relationship between the effective electromechanical coupling coefficient and τ_1.

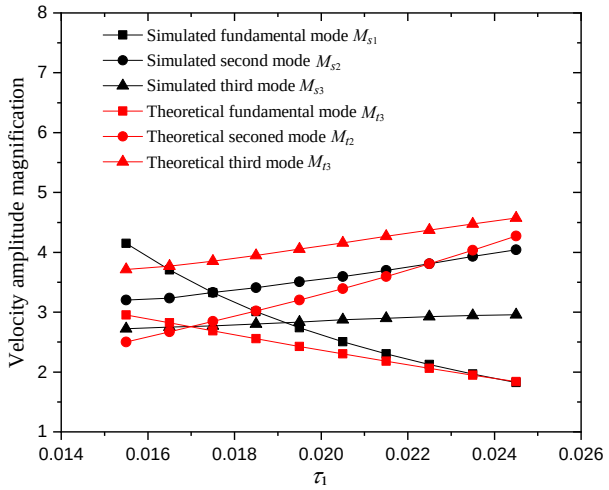


Fig. 5. Relationship between the velocity amplitude magnification and τ_1.

vibration velocity amplification ratio decreases with increasing τ_1, while the second- and third-order velocity amplification ratios increase proportionally. When τ_1 reaches approximately 20.5 mm, the transducer simultaneously achieves large effective electromechanical coupling coefficients and velocity amplification ratios for the first two longitudinal vibration modes, establishing this value as the optimized radius for the right-side sandwich-type piezoelectric stack. Therefore, $R_3 = R_4 = 0.0205$ m and $R_6 = 0.02$ m were selected as the optimized dimensional parameters.

3.2. Influence of the relative position of the variable cross-section of two sets of piezoelectric ceramic sandwich structures on performance parameters

As described in the previous section, the radii of the metal blocks and piezoelectric ceramics are kept unchanged. The lengths L_2 and L_3 are varied, but the total length L_2 and L_3 remains constant. The initial dimensional parameters of the transducer are: $L_1 = 0.02$ m, $L_2 = L_3 = 0.0075$ m, $L_4 = 0.021$ m, $L_5 = L_6 = 0.005$ m, $L_7 = 0.054$ m, $R_1 = R_2 = 0.0255$ m, $R_3 = R_4 = 0.0205$ m, $R_5 = 0.025$ m, $R_6 = 0.02$ m, $R_7 = 0.009$ m, $p_1 = p_2 = 4$, and $\tau_2 = L_2/L_3$. The relationships between the resonance frequency, anti-resonance frequency, effective electromechanical coupling coefficient, and velocity amplification ratio with respect to τ_2 are shown in Fig. 6 to Fig. 8.

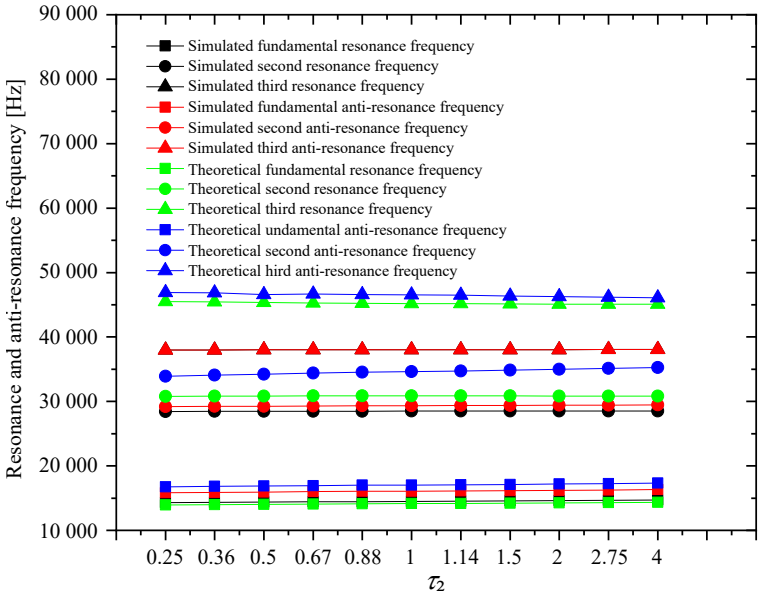


Fig. 6. Relationship between the resonance and anti-resonance frequency and τ_2.

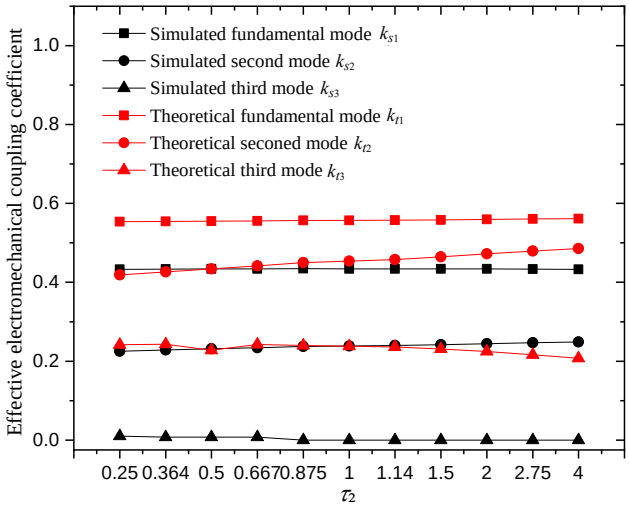


Fig. 7. Relationship between the effective electromechanical coupling coefficient and τ_2.

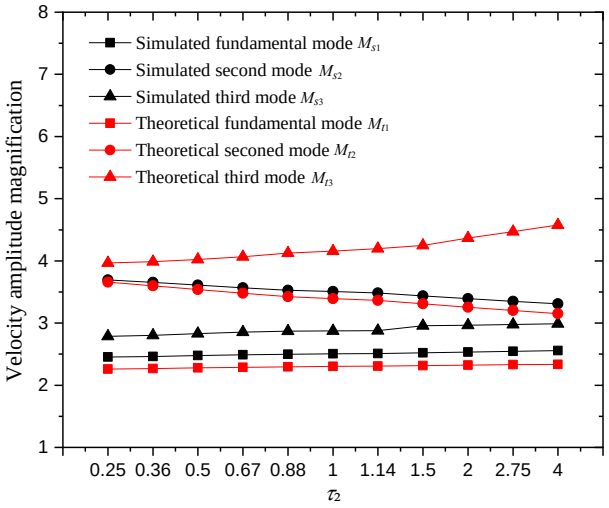


Fig. 8. Relationship between the velocity amplitude magnification and τ_2.

Figure 6 demonstrates that as τ_2 increases, the resonant and anti-resonant frequencies of the first three longitudinal vibration modes all exhibit gradual increases. In Fig. 7, the electromechanical coupling coefficients of the first- and second-order longitudinal vibrations show slow increases with τ_2, while the third-order longitudinal vibration electromechanical coupling coefficient gradually decreases. The velocity amplification ratios of the first- and third-order longitudinal vibrations rise slowly with τ_2. Figure 8 reveals a gradual decrease in the second-order longitudinal vibration velocity amplification ratio as τ_2 increases. Overall, variations in the relative position of the variable cross-sections in the two sandwich-type piezoelectric stacks exert minimal influence on the resonant/anti-resonant frequencies, electromechanical coupling coefficients, and velocity amplification ratios. The optimal relative position is determined as $\tau_2 = 1$ for the variable cross-sections in the two sandwich-type piezoelectric stack structures.

3.3. Influence of the spacing between two groups of piezoelectric ceramics on performance parameters

As discussed in the preceding section, while maintaining the other structural dimensions of the vibration system constant, the spacing between the two sets of piezoelectric ceramics was varied. This spacing corresponds to the length of the middle metal cylinders III and IV. The length changes of these two metal cylinder were identical, while the total length of the transducer remained unchanged. Consequently, the lengths of the back metal cylinder I and the front metal cylinder VI were adjusted correspondingly to accommodate the changes in the middle metal cylinder III and IV. The initial dimensional parameters of the transducer are: $L_1 = 0.02$ m, $L_2 = L_3 = 0.0075$ m, $L_4 = 0.021$ m, $L_5 = L_6 = 0.005$ m, $L_7 = 0.054$ m, $R_1 = R_2 = 0.0255$ m, $R_3 = R_4 = 0.0205$ m, $R_5 = 0.025$ m, $R_6 = 0.02$ m, $R_7 = 0.009$ m, $p_1 = p_2 = 4$, and $\tau_3 = L_2 + L_3$ m. The relationships between the resonance frequency, anti-resonance frequency, effective electromechanical coupling coefficient, velocity amplification ratio with respect to τ_3 are shown in Fig. 9 to Fig. 11.

Figure 9 indicates that, as τ_3 increases, the resonant and anti-resonant frequencies of the first three longitudinal vibration modes exhibit gradual decreases. In Fig. 10, the effective electromechanical coupling coefficients of the first- and second-order longitudinal vibrations decrease with increasing τ_3, while the third-order coupling coefficient initially decreases and then increases. Figure 11 demonstrates that the velocity amplification ratios for all three longitudinal vibration modes decreases as τ_3 increases, with the first- and second-order modes exhibiting smaller reductions compared to the third-order mode. Both theoretical and simulation results indicate that a smaller τ_3 between the two piezoelectric stacks enhances the effective electromechanical coupling coefficients and the velocity amplification ratios for the first two vibration modes. Considering manufacturing and assembly constraints, as well as the vibrational coupling intensity between structural components, the piezoelectric ceramic spacing is optimized to 10 mm.

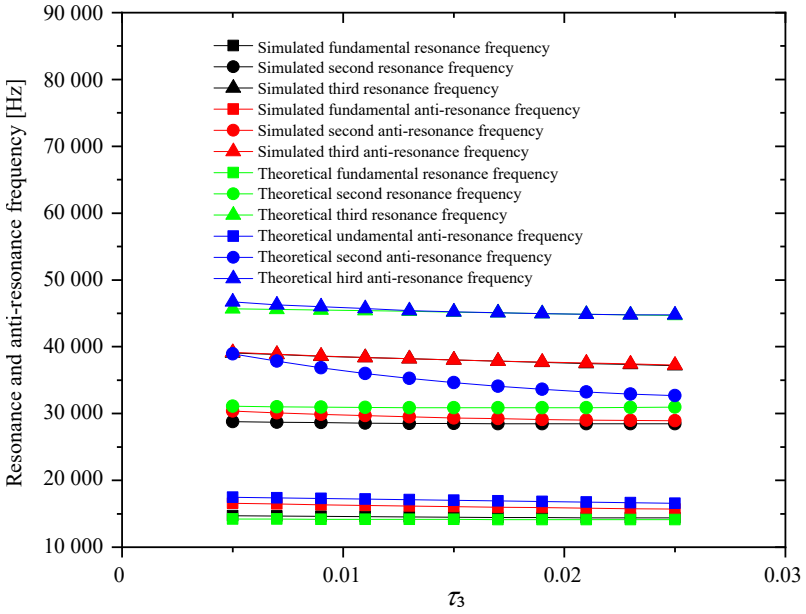


Fig. 9. Relationship between the resonance and anti-resonance frequency and τ_3.

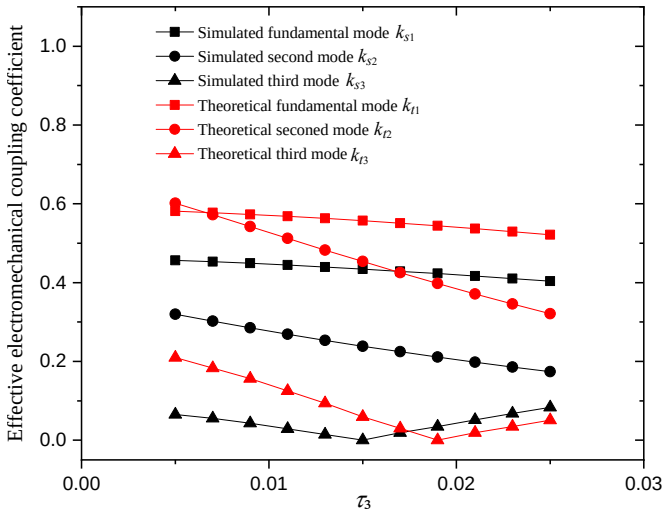


Fig. 10. Relationship between the effective electromechanical coupling coefficient and τ_3.

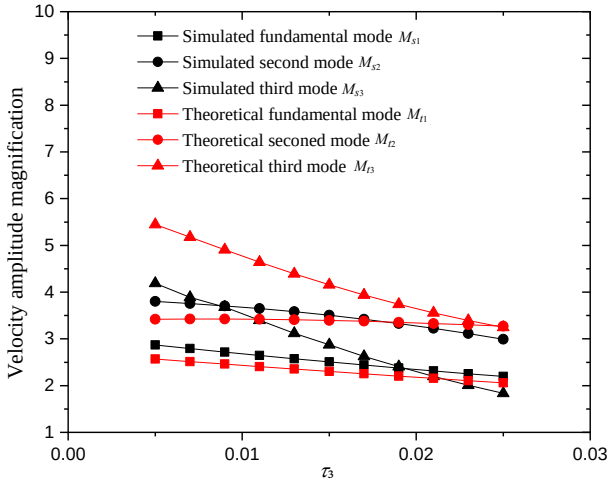


Fig. 11. Relationship between the velocity amplitude magnification and τ_3.

3.4. Influence of piezoelectric materials on performance parameters

Replacing piezoelectric ceramics with RFSCs alters the transducer performance, as quantified by theoretical and simulated parameters listed in Table 2. The parameters f_{tri}, f_{tai}, f_{sri}, f_{sai}, k_{ti}, k_{si}, M_{ti}, and M_{si} represent the theoretical/simulated resonance/anti-resonance frequencies, effective electromechanical coupling coefficients, and velocity amplitude amplification ratios, respectively. Here, $i = 1$, $i = 2$, and $i = 3$ denote configurations with bilateral piezoelectric ceramics, a left-side piezoelectric ceramic/right-side RFSC, and bilateral RFSCs, respectively. Table 2 demonstrates that replacing the right-side piezoelectric ceramic with an RFSC increases the first- and third-order electromechanical coupling coefficients but reduces the second-order coupling coefficients, while the amplification ratios exhibit negligible variation. In contrast, substituting both sides with RFSCs induces a pronounced upward trend in all electromechanical coupling coefficients, accompanied by reduced amplification ratios for the fundamental and second-order modes and an increased third-order amplification ratio.

Theoretical and simulated results for bilateral piezoelectric ceramics.

Table 2. Theoretical and simulated results for different piezoelectric material configurations.

Theoretical and simulated results for bilateral piezoelectric ceramics.								
Mode	f_{tr1} [Hz]	f_{ta1} [Hz]	f_{sr1} [Hz]	f_{sa1} [Hz]	k_{t1}	k_{s1}	M_{t1}	M_{s1}
Fundamental	14 177	17 265	14 585	16 305	0.571	0.447	2.43	2.68
Second	30 939	36 411	28 606	29 769	0.527	0.277	3.34	3.68
Third	45 453	45 883	38 481	38 506	0.137	0.036	4.77	3.54
Theoretical and simulated results for the left-side piezoelectric ceramic/right-side RFSC.								
Mode	f_{tr2} [Hz]	f_{ta2} [Hz]	f_{sr2} [Hz]	f_{sa2} [Hz]	k_{t2}	k_{s2}	M_{t2}	M_{s2}
Fundamental	12 509	18 578	11 113	15 271	0.739	0.686	2.68	2.69
Second	33 799	37 198	27 042	27 298	0.418	0.137	3.18	3.37
Third	43 554	44 996	30 577	31 366	0.251	0.223	4.78	3.56
Theoretical and simulated results for the bilateral RFSCs.								
Mode	f_{tr3} [Hz]	f_{ta3} [Hz]	f_{sr3} [Hz]	f_{sa3} [Hz]	k_{t3}	k_{s3}	M_{t3}	M_{s3}
Fundamental	8953	14 259	10 046	14 704	0.778	0.730	1.88	1.89
Second	20 824	31 883	20 615	21 767	0.757	0.321	1.66	1.77
Third	32 740	33 469	29 647	30 690	0.208	0.258	5.72	4.63

Figure 12 to Fig. 14 display the harmonic displacement responses from the front to the rear radiation surfaces of transducers with three piezoelectric ceramic-RFSC configurations across the first three vibration modes. When both sides employ piezoelectric ceramics, symmetrical displacement distributions emerge, with lower amplitudes in the fundamental mode, while higher-order modes exhibit progressively amplified vibrations and potential multi-peak displacement patterns. For the asymmetric configuration (left-side piezoelectric ceramic/right-side RFSC), the RFSC demonstrates localized amplitude enhancement in the second- and third- modes due to its superior mechanical compliance. However, the ceramic's rigidity restricts vibration energy transfer to the rear surface, resulting in insignificant improvement in the amplification ratio compared with the bilateral ceramic configurations. The bilateral RFSC achieves the maximum overall vibration amplitudes, as the high material compliance promotes efficient energy propagation to the rear surface, generating substantial posterior displacement. Concurrently, reduced wave reflection and increased energy dissipation at the front surface diminish its vibration amplitude, ultimately causing significant degradation of the amplification ratio.

In summary, replacing the right-side piezoelectric ceramic with RFSCs predominantly enhances the electromechanical coupling coefficients while minimally affecting the amplification ratios, enabling transducers to achieve both high coupling coefficients and velocity amplification at low-order modes. Complete substitution with bilateral RFSCs further optimizes electromechanical coupling performance, demonstrating the material's potential for advanced transducer applications. In practical manufacturing processes, the size of RFSCs is usually designed to be smaller. The size optimization presented in this study makes it possible to apply RFSCs in transducers. Compared with conventional piezoelectric ceramics, the performance can be comparable or even superior. In reality, the material combination can be chosen according to the specific application needs.

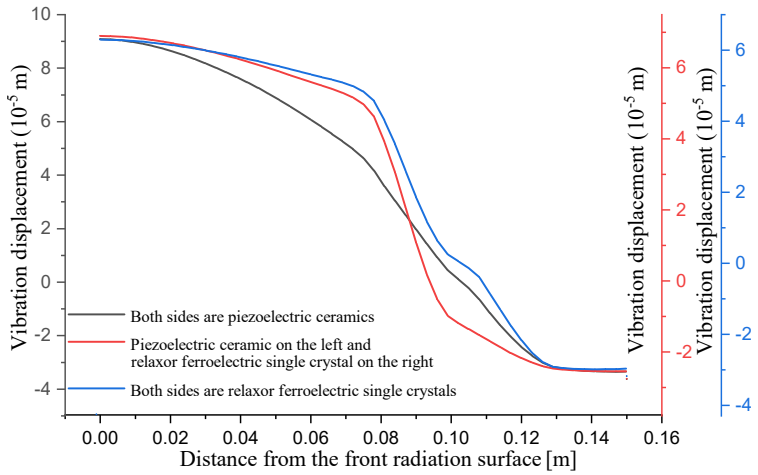


Fig. 12. Fundamental harmonic displacement response curve.

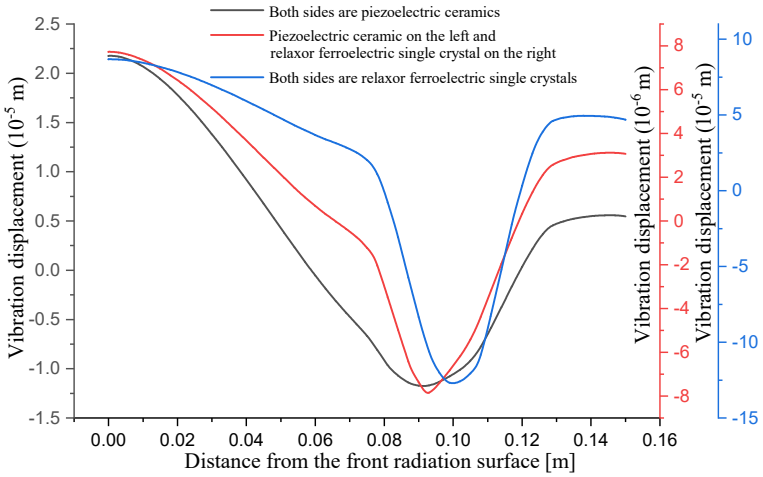


Fig. 13. Second-order harmonic displacement response curve.

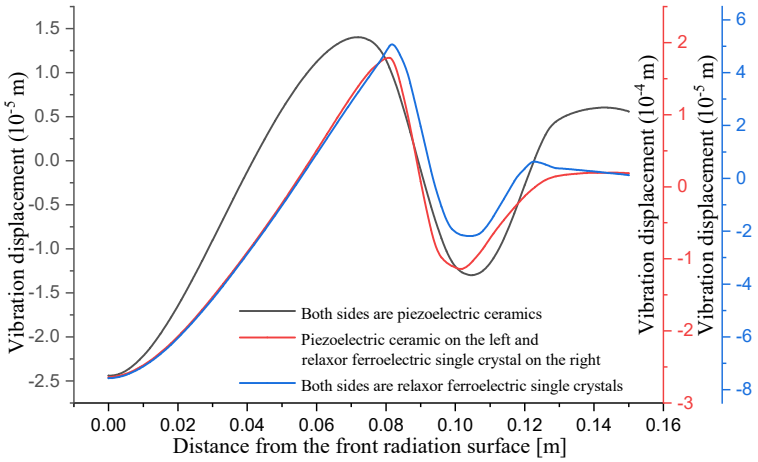


Fig. 14. Third-order harmonic displacement response curve.

4. Experiment

Based on the transducer dimensions designed in Sec. 3, a physical prototype was fabricated and tested. The corresponding structural dimensions are listed in Table 3, while Fig. 15. First three mode shapes illustrates the mode shapes of the transducer under three distinct vibration modes. In this transducer, the capacitances of the left

Table 3. Structural dimensions of the cascaded piezoelectric transducer [mm].

L_1	L_2	L_3	L_4	L_{01}	L_{02}	L_7	R_1	R_2	R_{01}	R_{02}	R_7
22.5	5	5	23.5	5	5	54	25.5	20.5	25	20	9

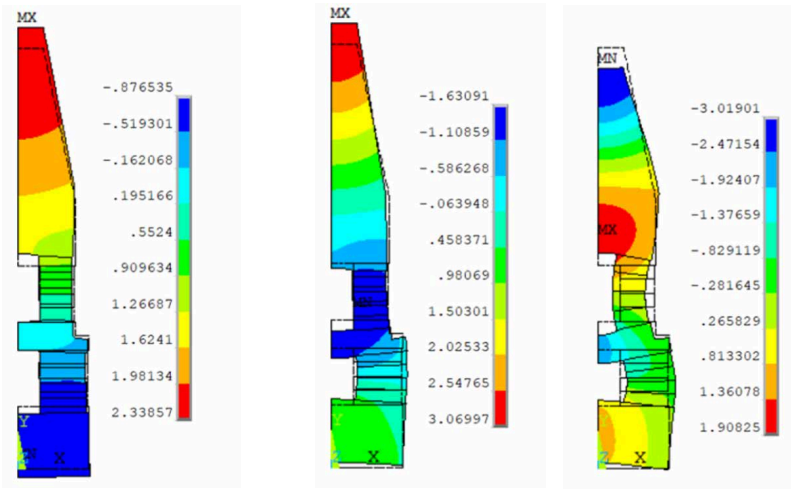


Fig. 15. First three mode shapes of the transducer.

piezoelectric ceramic are 3.414 nF, 3.409 nF, 3.418 nF, and 3.406 nF, respectively, and the capacitances of the right piezoelectric ceramic are 1.606 nF, 1.625 nF, 1.608 nF, and 1.635 nF, respectively. The total capacitance of the transducer is 20.01 nF, indicating good electrical connectivity.

Reactance-frequency responses were acquired using a precision impedance analyzer (E4990A) under 1 V excitation over the range of 10 kHz to 45 kHz. The experimental setup and measured reactance-frequency characteristics are presented in Fig. 16 and Fig. 17. Notably, spectral peaks correspond to anti-resonance states while troughs indicate resonance conditions, with the first three resonance points matching the aforementioned vibration modes. Table 4 provides comparative resonance/anti-resonance frequency data obtained using different analytical approaches.

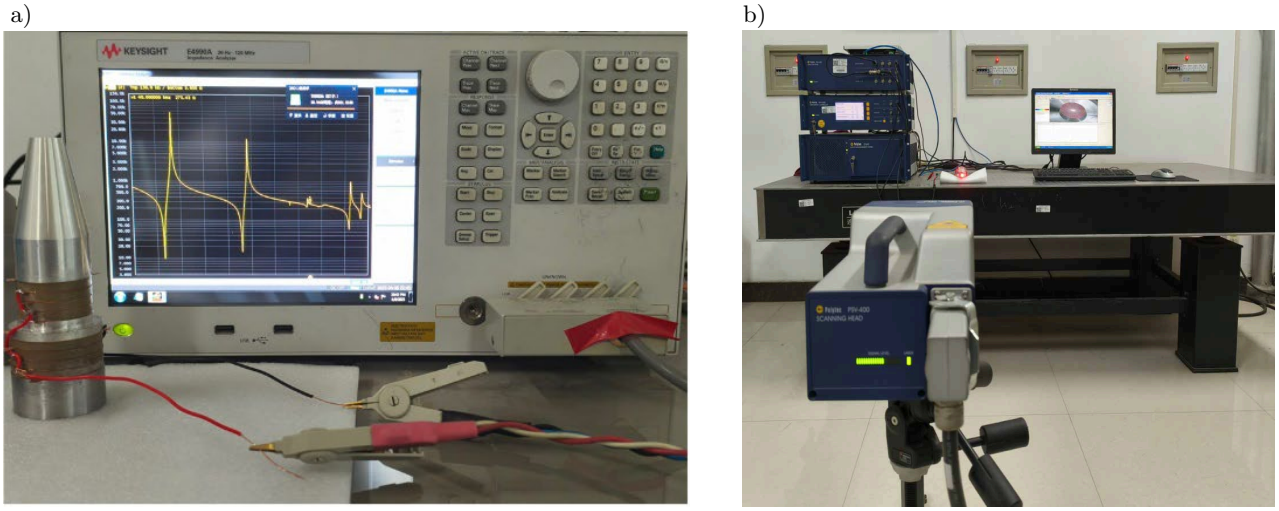


Fig. 16. a) Measuring electrical impedance-frequency response of the transducer by using the E4990A precision impedance analyzer, b) measuring vibration mode of the transducer by using the Polytec scanning vibrometer.

The vibration velocities at the metal block end face and conical output terminal of the transducer were measured using a Polytec laser vibrometer. Figure 16 shows the experimental setup, Fig. 17 illustrates the experimental electrical impedance-frequency response curve and experimental longitudinal vibration displacement

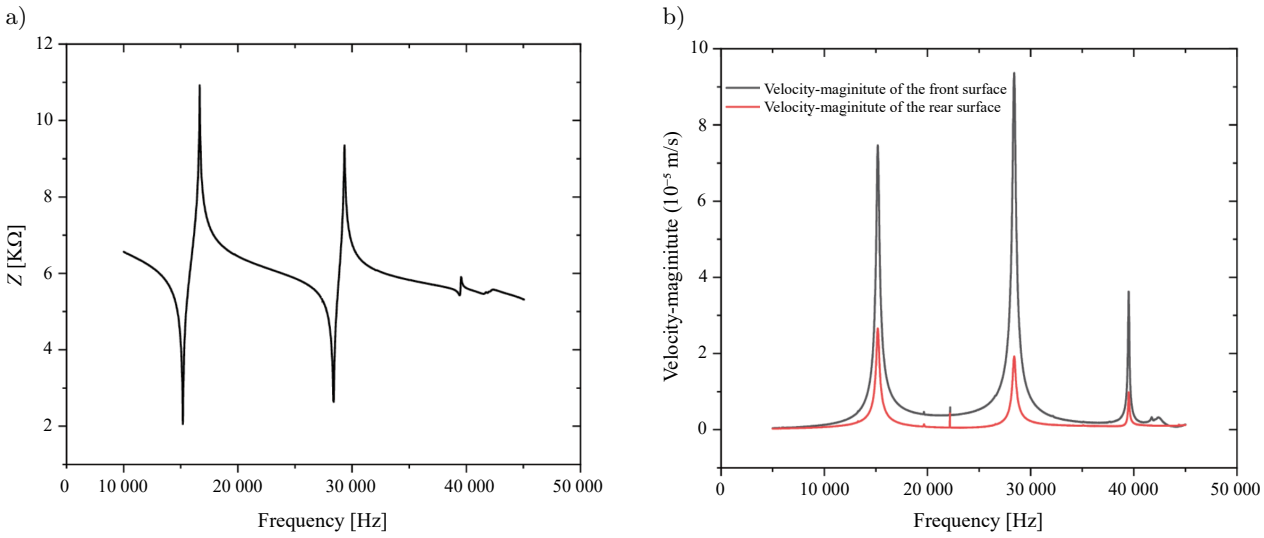


Fig. 17. a) Experimental electrical impedance-frequency response curve, b) experimental longitudinal vibration displacement frequency-response results.

Table 4. Theoretical, simulated, and measured frequencies of the multi-frequency mode transducer.

Mode	f_{tr} [Hz]	f_{ta} [Hz]	f_{sr} [Hz]	f_{sa} [Hz]	f_{mr} [Hz]	f_{ma} [Hz]
Fundamental	14 177	17 265	14 585	16 305	15 187	16 660
Second	30 939	36 411	28 606	29 769	28 367	29 343
Third	45 453	45 883	38 481	38 506	39 418	39 549

frequency-response results. The longitudinal vibration displacement distributions are presented in Fig. 18. The first two longitudinal vibration modes exhibit relatively pure characteristics, while the third-order demonstrates coupling vibration effects caused by the larger transverse dimension at the rear end. Nevertheless, the front end maintains satisfactory longitudinal vibration output. The effective electromechanical coupling coefficients and velocity amplification ratios were calculated using $k_{\text{eff}} = \sqrt{\left(1 - \left(\frac{f_r}{f_a}\right)^2\right)}$ and $M = \left|\frac{v_b}{v_f}\right|$, with comparative results from theoretical, simulated, and experimental approaches listed in Table 5, k_t, k_s, k_m, and M_t, M_s, M_m denote the theoretical, simulated, and measured values, respectively.

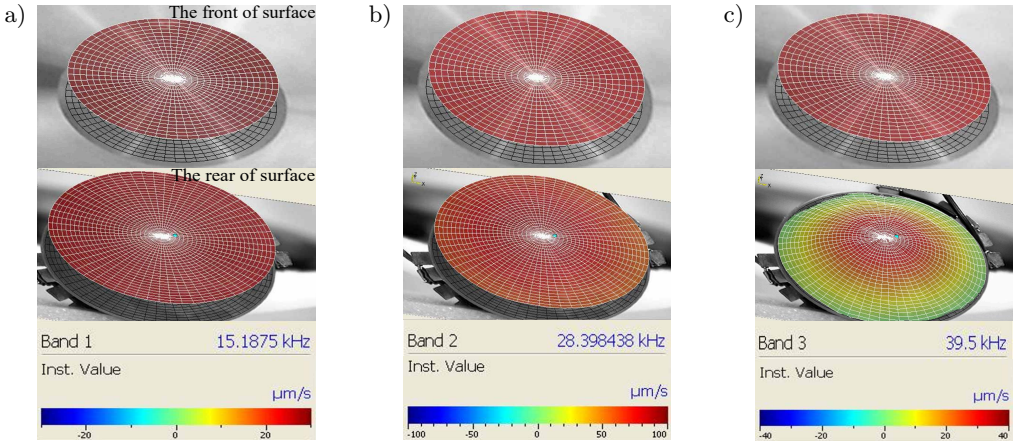


Fig. 18. Longitudinal vibration displacement distributions on the front and rear radiation surfaces of the transducer: a) fundamental mode, b) second-order, c) third-order.

As shown in Table 4, the theoretical, simulated, and experimental values for the first two resonant/anti-resonant frequencies show excellent agreement. However, a larger discrepancy is observed for the third-order

Table 5. Theoretical, simulated, and measured effective electromechanical coupling coefficient and velocity amplitude magnifications of the multi-frequency mode transducer.

Mode	k_t	k_s	k_m	M_t	M_s	M_m
Fundamental	0.571	0.447	0.412	2.403	2.68	2.87
Second	0.527	0.277	0.256	3.34	3.68	4.86
Third	0.137	0.036	0.081	4.77	3.54	3.68

frequency, where the theoretical value deviates significantly from both the simulated and experimental results. This discrepancy arises because the theoretical model is based on 1D analysis, which neglects the influence of transverse vibrations. As the mode order increases, the impact of transverse coupling effects becomes progressively more significant, leading to increased error. Furthermore, discrepancies between the simulated and experimental values exist due to inevitable deviations of the actual material parameters from their nominal values, combined with the neglect of bolt stress effects in the simulation.

As shown in Table 5, the simulated and experimental values of the effective electromechanical coupling coefficient exhibit good agreement, while the theoretical values are consistently higher. This discrepancy arises because the theoretical model neglects dielectric losses and assumes idealized fixed boundary conditions. The experimental velocity amplification ratio for the second-order is significantly higher than both theoretical and simulated values. This occurs because the measured resonant frequency is lower than theoretically predicted, indicating that the actual system stiffness is lower than anticipated. This measured frequency likely approaches the transducer's true optimal resonant point, where vibration phase matching between the piezoelectric ceramic stack and metal blocks is improved, resulting in enhanced amplification efficiency. Conversely, the theoretical amplification ratio for the third-order substantially exceeds both simulated and experimental values. This overestimation stems from the theoretical assumption of perfectly rigid metal blocks, whereas at higher frequencies, material compliance and micro-slip at connections become significant, leading to an overestimation of amplification performance. The remaining amplification ratios show good consistency across all three methods.

5. Conclusions

This paper investigated the multi-modal vibration characteristics of a longitudinal cascaded sandwich transducer incorporating a conical horn. Through theoretical and simulation analyses, the influence of key structural parameters on performance was determined, specifically the output radius of the second-stage, the relative position of the variable cross-section of two sets of piezoelectric ceramic sandwich structures, and the spacing between the two sets of piezoelectric ceramics. Optimal transducer dimensions were derived and subsequently validated experimentally. Furthermore, a hybrid configuration combining PZT-8 ceramics with RFSCs was proposed, revealing the impact of piezoelectric material selection on performance metrics. The established optimization methodology provides guidance for multi-mode transducer design, while the optimized transducer demonstrates promising application potential in high-power ultrasonic systems. The principal conclusions are as follows:

1. Reducing the output radius of the second-stage facilitates the concurrent maintenance of substantial effective electromechanical coupling coefficients for the fundamental and second-order longitudinal vibrations, while the third-order effective electromechanical coupling coefficient simulations reach minimal values within smaller radius ranges. This configuration simultaneously enhances fundamental vibration velocity amplification ratios while diminishing higher-order responses.
2. The relative positions of the two variable cross-section piezoelectric stacks demonstrated negligible influence on operational parameters.
3. Minimizing the spacing between the two piezoelectric stacks improves both effective electromechanical coupling coefficients and velocity amplification ratios in primary vibration modes, representing a critical design consideration for practical implementations.
4. Replacing conventional ceramics with RFSCs in the right-side configurations yields enhanced effective electromechanical coupling coefficients without compromising velocity amplification. Bilateral substitution

further increases effective electromechanical coupling coefficients at the expense of reduced amplification ratios, offering flexible material selection strategies for application-specific requirements. These findings establish a methodological framework for performance-adaptive transducer design in high-intensity ultrasonic applications.

Fundings

This work was supported by the National Natural Science Foundation of China (grant nos. 12104369 and 12174004), the Key Research and Development Project of Shaanxi Province in China (no. 2025SF-YBXM-470).

Conflict of interests

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.

Authors' contributions

Guo Li conceptualized the study, supervised the research and acquired funding. Peiyu Tan performed the theoretical analysis, numerical simulations, and experimental measurements, and wrote the original draft. Ruihui Ma and Feilong Li assisted in simulations and data validation. Xiaoli Zhang contributed to resources and manuscript review. Hua Tian provided laboratory support and administered the project. All authors reviewed and approved the final manuscript.

References

1. Chemat F., Zill-e-Huma, Khan M.K. (2011), Applications of ultrasound in food technology: Processing, preservation and extraction, *Ultrasonics Sonochemistry*, **18**(4): 813–835, https://doi.org/10.1016/j.ultsonch.2010.11.023.
2. Davari P., Ghasemi N., Zare F., O'Shea P., Ghosh A. (2012), Improving the efficiency of high power piezoelectric transducers for industrial applications, *IET Science, Measurement & Technology*, **6**(4): 213–221, https://doi.org/10.1049/iet-smt.2011.0209.
3. Du Y.D., Xu L., Zhou G.P. (2020), Vibration characteristics of two-dimensional orthogonal composite sandwich piezoelectric circular transducer [in Chinese], *SCIENTIA SINICA Physica, Mechanica & Astronomica*, **51**(11): 115311, https://doi.org/10.1360/SSPMA-2020-0417.
4. Hunter G., Lucas M., Watson I., Parton R. (2008), A radial mode ultrasonic horn for the inactivation of *Escherichia coli* K12, *Ultrasonics Sonochemistry*, **15**(2): 101–109, https://doi.org/10.1016/j.ultsonch.2006.12.017.
5. Jiang C.R., Cheng S.N., Shen Z., Jin L. (2024), Design and analysis of impedance matching for a radial standing wave piezoelectric ultrasonic motor, *Review of Scientific Instruments*, **95**(1): 015003, https://doi.org/10.1063/5.0180662.
6. Jiang Z., Wang L., Jin J.M. (2018), A novel robotic arm driven by sandwich piezoelectric transducers, *Ultrasonics*, **84**: 373–381, https://doi.org/10.1016/j.ultras.2017.11.019.
7. Kuang Y., Jin Y., Cochran S., Huang Z. (2014), Resonance tracking and vibration stablilization for high power ultrasonic transducers, *Ultrasonics*, **54**(1): 187–194, https://doi.org/10.1016/j.ultras.2013.07.001.
8. Li F.-M., Liu S.-Q., Xu L., Zhang H.-D., Zeng X.-M., Chen Z.-J. (2023a), Open-slit circular tube piezoelectric ultrasonic transducer with longitudinal bending mode conversion [in Chinese], *SCIENTIA SINICA Physica, Mechanica & Astronomica*, **53**(11): 114311, https://doi.org/10.1360/SSPMA-2023-0233.
9. Li G. *et al.* (2024), Investigation of a cascaded piezoelectric transducer with cone horn for multi-mode power ultrasound application, *Physica Scripta*, **99**: 105997, https://doi.org/10.1088/1402-4896/ad6d05.
10. Li G., Gong J.H., Wang T., Qiu C.R., Xu Z. (2018a), Study on the broadband piezoelectric ceramic transducer based on radial enhanced composite structure, *Ceramics International*, **44**(Supplement 1): S250–S253, https://doi.org/10.1016/j.ceramint.2018.08.101.
11. Li G., Qu J.S., Xu L., Zhang X.L., Gao X.Y. (2022), Study on multi-frequency characteristics of a longitudinal ultrasonic transducer with stepped horn, *Ultrasonics*, **121**: 106683, https://doi.org/10.1016/j.ultras.2022.106683.

12. Li G., Xu Z., Gong J.H., Li F., Liu Y.B. (2018b), Modeling and experiment of a small size dual mode transducer for underwater acoustic communication and detection, *Acta Acustica*, **104**(6): 947–955, https://doi.org/10.3813/AAA.919260.

13. Li J.Q., Zheng J.D., Pan H., Tong J., Feng K., Ni Q. (2023b), Two-dimensional composite multi-scale time–frequency reverse dispersion entropy-based fault diagnosis for rolling bearing, *Nonlinear Dynamics*, **111**: 7525–7546, https://doi.org/10.1007/s11071-023-08250-y.

14. Lin S.Y. (2007), Radial vibration of the composite ultrasonic transducer of piezoelectric and metal rings, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **54**(6): 1276–1280, https://doi.org/10.1109/TUFFC.2007.381.

15. Lin S.Y. (2009), Analysis of multifrequency Langevin composite ultrasonic transducers, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **56**(9): 1990–1998, https://doi.org/10.1109/TUFFC.2009.1275.

16. Lin S.Y. (2017), A new type of cascaded high intensity piezoelectric ultrasonic transducers [in Chinese], *Journal of Shaanxi Normal University (Natural Science Edition)*, **45**(06): 22–28, https://doi.org/10.15983/j.cnki.jsnu.2017.06.261.

17. Lin S.Y., Fu Z.Q., Zhang X.L., Wang Y., Hu J. (2013), Radially sandwiched cylindrical piezoelectric transducer, *Smart Materials and Structures*, **22**: 015005, https://doi.org/10.1088/0964-1726/22/1/015005.

18. Lin S.Y., Guo H., Xu J. (2018), Actively adjustable step-type ultrasonic horns in longitudinal vibration, *Journal of Sound and Vibration*, **419**: 367–379, https://doi.org/10.1016/j.jsv.2018.01.033.

19. Lin S.Y., Xu L., Hu W.X. (2011), A new type of high power composite ultrasonic transducer, *Journal of Sound and Vibration*, **330**(7): 1419–1431, https://doi.org/10.1016/j.jsv.2010.10.009.

20. Liu S.Q., Yao Y., Lin S.Y. (2009), Radial-torsional vibration mode of disc-type piezoelectric transducer, *Journal of Mechanical Engineering*, **45**(06): 176–180, https://doi.org/10.3901/JME.2009.06.176.

21. Meshkinzar A., Al-Jumaily A.M. (2021), Vibration and acoustic radiation characteristics of cylindrical piezoelectric transducers with circumferential steps, *Journal of Sound and Vibration*, **511**: 116346, https://doi.org/10.1016/j.jsv.2021.116346.

22. Pokhrel N., Vabbina P.K., Pala N. (2016), Sonochemistry: Science and engineering, *Ultrasonics Sonochemistry*, **29**: 104–128, https://doi.org/10.1016/j.ultsonch.2015.07.023.

23. Roland R. (2024), Improving output performance of the ultrasonic multicell piezoelectric motor by development the multi-rotor structure, *IEEE Access*, **12**: 32080–32088, https://doi.org/10.1109/ACCESS.2024.3370560.

24. Shah U.H., Liu X. (2019), Ultrasonic resistance welding of TRIP-780 steel, *Journal of Materials Processing Technology*, **274**: 116287, https://doi.org/10.1016/j.jmatprotec.2019.116287.

25. Su Q., Liu Y.X., Wang L., Deng J., Li H. (2018), A three-dimensional piezoelectric nanopositioner using a sandwich transducer, *Ceramics International*, **44**(Supplement 1): S108–S111, https://doi.org/10.1016/j.ceramint.2018.08.226.

26. Subhedar P.B., Gogate P.R. (2015), Ultrasound assisted intensification of biodiesel production using enzymatic interesterification, *Ultrasonics Sonochemistry*, **29**: 67–75, https://doi.org/10.1016/j.ultsonch.2015.09.006.

27. Tan W.X., Tan K.W., Tan K.L. (2022), Developing high intensity ultrasonic cleaning (HIUC) for post-processing additively manufactured metal components, *Ultrasonics*, **126**: 106829, https://doi.org/10.1016/j.ultras.2022.106829.

28. Tang Y.F., Chen C., Tian H., Lin S. (2024), Radial vibration analysis for piezoceramic shell-stacked spherical transducer with thick walls, *Smart Materials and Structures*, **33**: 035002, https://doi.org/10.1088/1361-665X/ad1e8f.

29. Tressler J.F., Howarth T.R., Huang D. (2006), A comparison of the underwater acoustic performance of single crystal versus piezoelectric ceramic based "cymbal" projectors, *The Journal of the Acoustical Society of America*, **119**(2): 879–889, https://doi.org/10.1121/1.2150153.

30. Watanabe Y., Tsuda Y., Mori E. (1993), A longitudinal-flexural complex-mode ultrasonic high-power transducer system with one-dimensional construction, *Japanese Journal of Applied Physics*, **32**: 2430, https://doi.org/10.1143/JJAP.32.2430.

31. Xu L., Qiu X.J., Zhou J.C., Li F.M., Zhang H.D., Wang Y.B. (2019), A 2D dual-mode composite ultrasonic transducer excited by a single piezoceramic stack, *Smart Materials and Structures*, **28**: 025017, https://doi.org/10.1088/1361-665X/aaf275.

Archives of Acoustics Vol. 51, No. 1, pp. 161–181 (2026), https://doi.org/10.24423/archacoust.2026.4254

Technical Note

Laboratory Evaluation of Underwater Acoustic Attenuation Provided by Air Bubble Curtains

Alfio Yori*, José Luis Barros, Rodrigo Torres, Felipe Figueroa

Instituto de Acústica, Universidad Austral de Chile
Valdivia, Chile

*Corresponding Author: ayori@uach.cl

Received June 3, 2025; revised December 26, 2025; accepted January 1, 2026; available online January 16, 2026; version of record March 9, 2026; published issue March 27, 2026.

Air bubble curtains are widely used in Chilean salmon farming to protect fish from water pollutants. Due to new requirements from the Environmental Assessment Service, their use for underwater noise mitigation has become very important. This study aimed to evaluate the acoustic attenuation provided by a type of curtain widely used in salmon farming and seawater desalination under laboratory conditions, obtaining results independent of the conditions encountered in field work. This allowed for comparative measurements of different bubble curtain configurations. This study presents the results of an acoustic evaluation of underwater sound attenuation provided by different bubble curtain configurations, conducted in the test channel of the Engineering Department at the Universidad Austral de Chile. The parameter evaluated was the insertion loss (IL). It was found that the evaluated bubble curtains provided a broadband insertion loss between 11 dB and 12 dB in the simplest configuration and IL per third-octave band equal to or greater than 10 dB. The 10 dB attenuation in the source level represents a significant reduction in the area of acoustic impact, reducing the physiological damage distances for the marine mammal species considered by at least 78.4 %.

Keywords: air-bubble curtain, underwater sound attenuation, reduction of physiological risk, underwater noise control.

1. Introduction

Currently, in Chile, bubble curtains are widely used in salmon farming centers to protect salmon from algae blooms or waterborne solid and liquid contaminants. Another use of these bubble curtains, which has been steadily increasing in our country, is in the water intake pipes of desalination plants to prevent the suction of organic matter. Due to new requirements of the Environmental Assessment Service [SEA] (2022), regarding the impact of underwater noise, a new need arises for the use of bubble curtains to mitigate anthropogenic underwater noise.

The efficacy of bubble curtains in regulating the dispersal of both solid and oil contaminants can be demonstrated visually. However, validating their effectiveness in sound attenuation poses a greater challenge. Despite the ambiguity surrounding the quantification of the effectiveness of bubble curtains in reducing underwater noise, they are generally recommended as a possible strategy for mitigating the acoustic impact on marine organisms. One of the reasons why information about the acoustic effectiveness of these curtains presents high variability is the lack of attenuation measurements under controlled conditions (Rustemeier *et al.*, 2012; Beelen *et al.*, 2025), contrasted with the high number of field measurements (Molnar *et al.*, 2020; Würsig *et al.*, 2000; Dähne *et al.*, 2017; Lucke *et al.*, 2011) and computerized models studies (Novarini *et al.*, 1998; Hall, 1989; Göttsche *et al.*, 2013; Qi *et al.*, 2021; Tsouvalas, 2020).

Air bubble curtains can be used as underwater acoustic barriers to control underwater noise sources and reduce acoustic impacts in the marine environment. Air bubble curtains are one of the few mitigation measures available for controlling anthropogenic underwater noise (Strietman *et al.*, 2018; Koschinski, Lüdemann, 2013; Merck *et al.*, 2014; Bellmann, 2014; Joint Nature Conservation Committee [JNCC], 2010; Werner, 2010; Lucke, Siemensma, 2013). Currently, the Environmental Assessment Agency (SEA, 2022) requires that projects with potential underwater acoustic impacts include mitigation measures, with air bubble curtains being a possible solution.

Due to this lack of information or clarity about sound attenuation provided by the air-bubble curtains, the main objective of this work is to assess, under controlled laboratory conditions, the acoustic attenuation provided by a type of curtain widely used by both the salmon industry and the industries of sea-water desalinization as well; seeking to achieve results independent of those variable conditions so commonly present in field measurements. Thus, attenuation results obtained from different curtain configurations may be compared.

This work shows the results of the acoustic assessment of underwater sound attenuation given by different configurations of air-bubble curtains, which was carried out in the test-channel belonging to the Engineering Faculty from Universidad Austral de Chile. The assessed parameter was insertion loss (IL), estimated from the sound pressure level generated by a sound source over a hydrophone, with and without the presence of the evaluated attenuating element (Fahy, 2001).

2. Objectives

To measure, under controlled conditions, underwater sound attenuation provided by air-bubble curtains presently used in the control of solid and liquid particles in salmon farming and water desalination plants.

To evaluate the reduction in the acoustic impact area of a project involving underwater noise and marine mammals when air bubble curtains are used as a mitigation measure.

3. Theoretical framework

3.1. Underwater sound generalities

In air acoustics, the standard reference pressure $p_0 = 20\,\mu\text{Pa}$ corresponds to the threshold of human hearing. In contrast, underwater acoustics uses a different reference pressure (Robinson *et al.*, 2014; Möser, Barros, 2004) equal to $p_0 = 1\,\mu\text{Pa}$. It is important to note that, due to significant differences in the acoustic impedance of air and water, sound pressure levels between these two media cannot be directly compared. For example, even in extremely quiet underwater environments, the broadband background noise level is generally around 90 dB re 1 µPa. Under high wind conditions, this value may reach approximately 120 dB re 1 µPa (Richardson *et al.*, 1995; Urick, 2010).

3.2. Sound and marine mammals

Anthropogenic underwater noise in the oceans can negatively affect mammals and invertebrates. In the oceans, sound propagates efficiently, just as light does in air (Richardson *et al.*, 1995). Marine mammals have evolved, taking advantage of the phenomenon of low sound attenuation using their hearing for most tasks, whereas land mammals use their eyesight. Thus, marine mammals use hearing for critical tasks such as navigation, communication, prey hunting, predator detection, and spatial localization in the dark ocean (Richardson *et al.*, 1995; Urick, 2010; Department of Arts, Heritage and the Gaeltacht [DAHG], 2014). Marine mammals can be mainly divided into (Department of Planning, Transportation and Infrastructure [DPTI], 2012):

– mysticetes or baleen whales, where this group includes large whales such as right whales, sei whales, humpback whales and blue whales. These species produce sounds mainly under 1000 Hz; with notable exceptions such as the humpback whale, which emits frequencies exceeding 1000 Hz, and the blue whale, which emits frequencies as low as 10 Hz to 15 Hz;

– odontocetes or toothed whales, in which dolphins, killer whales, sperm whales and porpoises are included. This type of cetacean communicates using frequencies under 20 kHz and use high-frequency echolocation to perceive their surroundings, using frequencies over 20 kHz;
– pinnipeds, which include sea lions and sea elephants, communicate using frequencies ranging from 1 kHz to 4 kHz.

3.3. Physiological and behavioral impact due to underwater noise

The effects of underwater noise of anthropogenic origin over marine mammals may be divided into five categories, which depend on the distance between the receiver and the sound source (DAHG, 2014; DPTI, 2012; McKenzie, 2015):

– physiological stress response,
– behavioral response,
– masking,
– temporary threshold shift (TTS),
– permanent threshold shift (PTS) or physical damage.

The frequency ranges through which these five groups of animals communicate are not well defined, and the levels of sound vary significantly between species. The range over which sounds can be detected depends largely on background noise levels and the animal's hearing threshold. Figure 1 shows a comparative graph of the average hearing thresholds of different mammal groups.

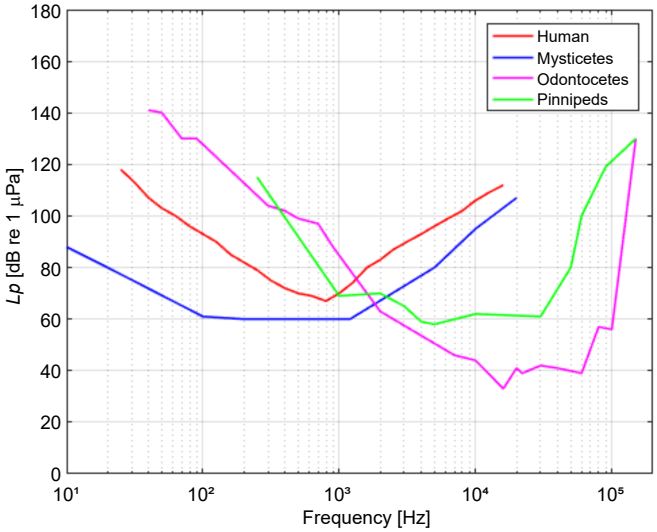


Fig. 1. Average hearing thresholds of different groups of marine mammals and humans when underwater (Parvin, Nedwell, 1995; Nedwell *et al.*, 2004).

Masking is an effect whereby noise masks essential auditory signals for animals, such as communication, obstacles or predator detection.

Changes in behaviour are observed across a wide range of sound levels but are difficult to evaluate objectively, since an animal's reaction to a noise depends on factors such as the type of sound, exposure time, and the season.

A TTS indicates temporary deafness, whereas a PTS implies a irreversible hearing loss (DAHG, 2014; DPTI, 2012; McKenzie, 2015).

It is important to note that not all high-level sounds will be harmful for all species. The auditory perception of a sound is considered harmful only if its frequencies fall within the audible range of the species concerned. Secondly, the level must exceed the initial level at which the species begins to perceive sound, around 50 dB (Richardson *et al.*, 1995; Nedwell *et al.*, 2007).

Figure 2 shows a comparison of the frequency spectrum of pile driving noise at a distance of 25 m (Yori, 2018) with the hearing thresholds of different species (Nedwell *et al.*, 2004). Frequency spectrum measured during the driving of piles 0.61 m in diameter, with a D62 diesel hammer and 40 blows per minute.

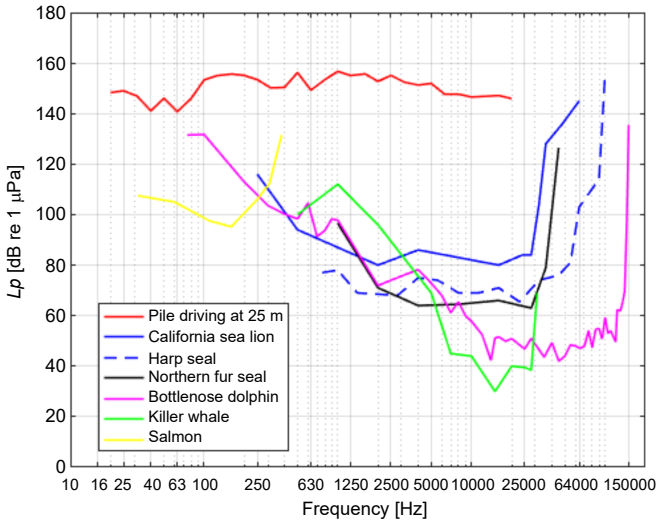


Fig. 2. Spectrum of a pile driving noise in 1/3-octave bands, in contrast with the audiograms of different species (Yori, 2018; Nedwell *et al.*, 2004).

3.3.1. Criteria for the assessment of underwater noise impact NOAA NMFS 2024

The National Marine Fishing Service (NMFS) of United States, belonging to the National Office of Ocean and Atmospheric Administration (NOAA), organized, interpreted and synthesized available scientific information regarding the impact of underwater noise over marine mammals, in order to specify the maximum thresholds to avoid the occurrence of a temporal hearing shifts TTS or an auditory injury AUD INJ, where AUD INJ includes, but is not limited to PTS. Marine mammals are grouped into seven groups as shown in Table 1.

Table 1. Hearing groups and ranges according to criterion by National Marine Fisheries Service [NMFS] (2024).

Hearing group	Hearing range
Low-frequency (LF) cetaceans (baleen whales)	7 Hz to 36 kHz
High-frequency (HF) cetaceans (dolphins, toothed whales, beaked whales, bottlenose whales)	150 Hz to 160 kHz
Very high-frequency (VHF) cetaceans (true porpoises, Kogia, river dolphins, cephalorhynchid, Lagenorhynchus cruciger, L. australis)	200 Hz to 165 kHz
Phocid pinnipeds (PW) (underwater) (true seals)	40 Hz to 90 kHz
Otariid pinnipeds (OW) (underwater) (sea lions and fur seals)	60 Hz to 39 kHz
Phocid pinnipeds (PA) (air) (true seals)	42 Hz to 52 kHz
Otariid pinnipeds (OA) (air) (sea lions and fur seals)	90 Hz to 40 kHz

The noise levels used by NMFS (2024) criteria to assess the impact of underwater noise on marine mammals are summarized in Table 2. Levels presented in Table 2 are cumulative and weighted levels, therefore the frequency

Table 2. Thresholds for assessing the acoustic impact of underwater noise.
Auditory injury INJ and TTS according to NMFS (2024) criterion [dB re 1 $\mu Pa^2 \cdot s$].

	Non-impulsive noise [dB]		Impulsive noise [dB]	
	Threshold TTS	Threshold AUD INJ	Threshold TTS	Threshold AUD INJ
Group	SEL_{cum24h}	SEL_{cum24h}	SEL_{cum24h}	SEL_{cum24h}
LF	177	197	168	183
HF	181	201	178	193
VHF	161	181	144	159
OW	179	199	170	185
PW	175	195	168	183

spectrum amplitudes of the assessed noise must be weighted by the auditory weighting functions corresponding to each group in Table 1. These auditory weighting function curves represent the hearing characteristics of the species classified in each one of these groups.

4. Materials and methods

4.1. Measurement system

The system employed to measure of underwater acoustic attenuation, provided by the evaluated air bubble curtain, was as follows:

- loudspeaker Lubell Labs, model LL916H (frequency response: 200 Hz to 23 kHz (±15 dB) and 500 Hz to 21 kHz (±10 dB)),
- Amplifier Peavey, model 2600, 75 watt rms,
- sound level meter NTI, type 1, model XL2,
- hydrophone Cetacean Research technology, model C55 (frequency response: 15 Hz to 44 kHz (±3 dB) and 8 Hz to 100 kHz (±12 dB)),
- eight-track recorder Tascam, model DR680,
- power source Cetacean Research technology, model 736,
- white noise source.

It is important to note that the speaker used is capable of emitting sound below 200 Hz. According to the sound pressure level plot provided by the manufacturer, the speaker is capable of producing a sound pressure level of 130 dB re 1 µPa · m below 40 Hz.

4.2. Hydrodynamic testing channel and bubble curtain installation

The measurement was carried out in the hydrodynamic testing channel of the Engineering Faculty at Universidad Austral de Chile (see Fig. 3). The channel is 50 m in length, 3 m in width, and 1.7 m in depth, and is constructed from steel. In principle, the water depth is capable of providing a work range above 200 Hz, given that lower frequencies are not propagated along the channel (Richardson *et al.*, 1995; Urick, 2010). However, near the source, there is propagation below this cut-off frequency, but its amplitude decays after a few wavelengths (Möser, Barros, 2004).

Fig. 3. Hydrodynamic testing channel.

As demonstrated in Fig. 4 and Fig. 5, the distance between the sound source and the hydrophone is 20 m. The evaluation involved four configurations of the curtain, with each line comprising two diffusers and hoses.

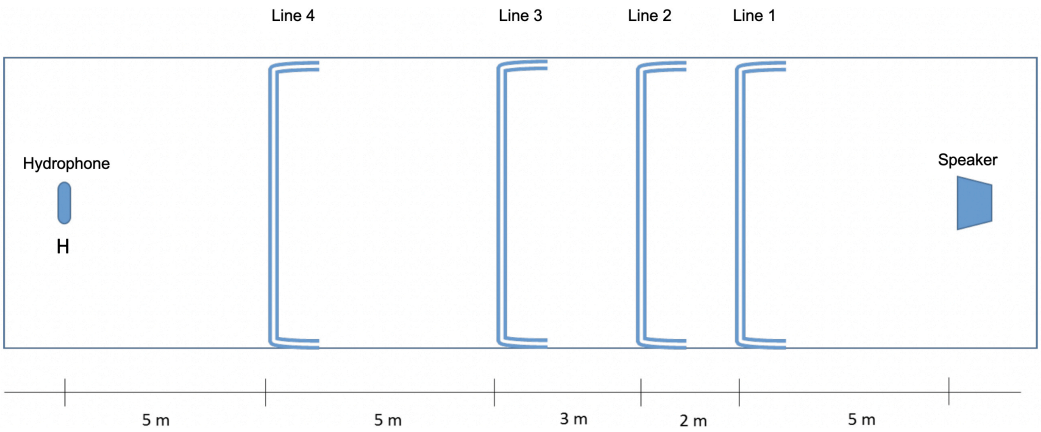


Fig. 4. Spatial arrangement of the four lines assessed in the study.

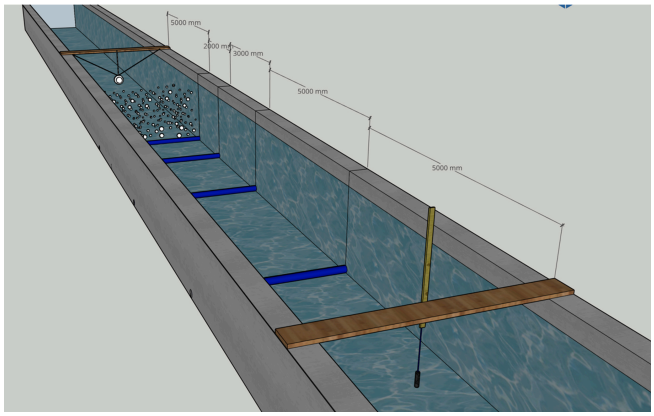


Fig. 5. Operational sequence of line 1 producing its bubble curtain.

The lines are designated line 1, line 2, line 3, and line 4, with their respective locations being 5 m, 7 m, 10 m, and 15 m from the sound source. Figure 6 shows one of these lines, where each hose has a 3/4-inch diameter and is supplied with an air flow rate of 27 l/min. The hoses have membranes producing microbubbles with diameters between 1 mm and 5 mm.

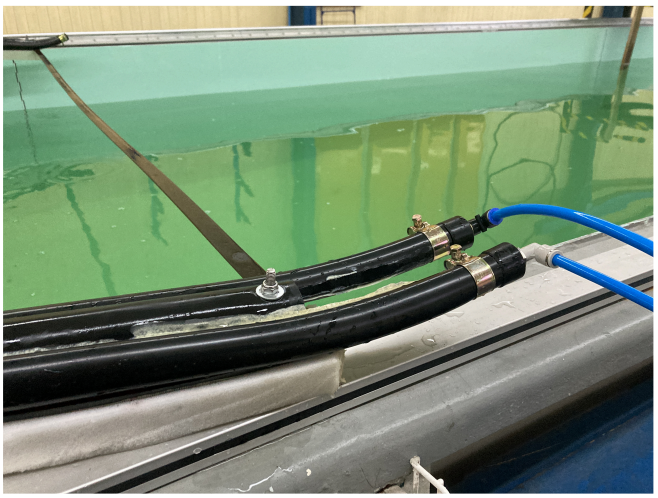


Fig. 6. Configuration of the double-diffusor line.

Each measurement was repeated four times, with the hydrophone positioned in four different locations, as illustrated in Fig. 7. The objective of these four positions was to minimize the effect of finding the hydrophones

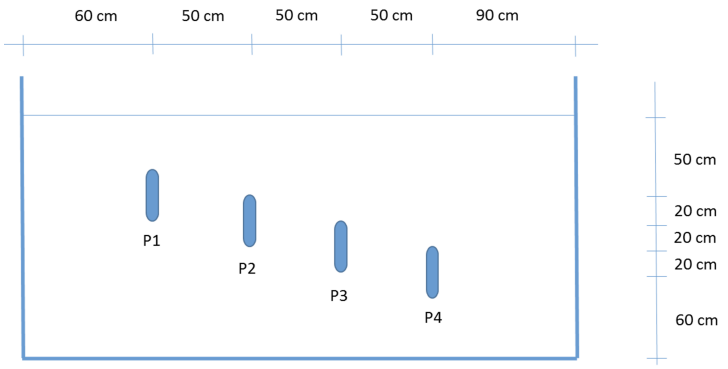


Fig. 7. Hydrophone positions.

at a position coinciding with a sound field diminishing at the point due to interference. This phenomenon was also reported in previous studies (Beelen *et al.*, 2025), so the goal here was to improve the ratio of sound signal to background noise. To get an average value reflecting both time and space, the results from these four different positions were combined by calculating their mean energy.

4.3. Curtains configuration and measurement methodology

The formula used to measure IL is the one set by Eq. (1), which is given by:

$$\mathrm{IL} = Lp_s - Lp_{x.x} \quad [\mathrm{dB}], \tag{1}$$

where Lp_s is the sound pressure level generated by the loudspeaker at the position of the hydrophone and in the absence of the evaluated curtain [dB re 1 µPa], and $Lp_{x.x}$ is the sound pressure level generated by the loudspeaker at the position of the hydrophone and in the presence of the evaluated curtain [dB re 1 µPa].

The IL is defined as the logarithmic ratio of the sound power transmitted by a system before the insertion of a noise control device to the sound power transmitted after insertion. This measure takes into account not only the performance of the noise control device, but also the effects of insertion, such as the alteration of the source's sound power or the generation of sound by the attenuator itself (Fahy, 2001).

In the subindex $x.x$ in Eq. (1), the first x indicates the number of the line considered, which may be 1, 2, 3 or 4. The second x indicates the number of active diffusors per line, which may be 1 or 2. For example, $Lp_{1.1}$ corresponds to the noise level measured at the hydrophone position, when line 1 presents only one of its two bubble curtains working. $Lp_{1.2}$ refers to line 1 with its two bubble curtains operating. Figure 8a shows an example of line 1 with only one of its diffusers or hoses active, and Fig. 8b, with both of its diffusers or hoses working.

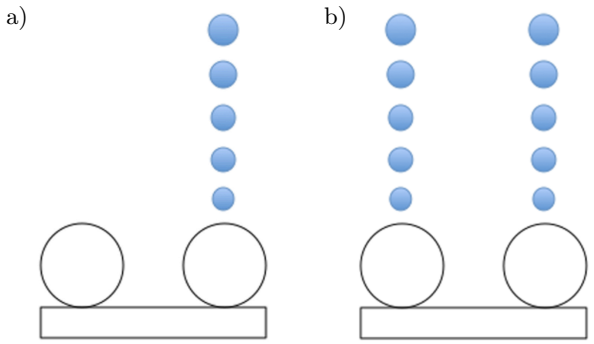


Fig. 8. a) Line 1 ($L_{1.1}$) with only one of its diffusors generating a bubble curtain, b) line 1 ($L_{1.2}$) with its two diffusors generating bubble curtains.

The measurements were carried out in a laboratory environment, under controlled conditions, allowing comparative measurements of different air bubble curtain configurations.

Table 3 shows the different configurations of bubble curtains that were assessed during this study.

Table 3. Curtains configuration and measurement methodology.

Lp_{RF}	Background noise level present in the laboratory. Loudspeaker off.
Lp_{RFB}	Background noise level with line 1, line 2 and line 3 working; each line with only one of their bubble curtains active. Loudspeaker off.
Lp_s	Level generated by the loudspeaker at the hydrophone position, without the presence of active curtains.
$Lp_{1.1}$	Level generated by the speaker at the hydrophone position, with line 1 operating with only one of its bubble curtains active.
$Lp_{1.2}$	Level generated by the speaker at the hydrophone position, with line 1 operating with both bubble curtains active.
$Lp_{3.1}$	Level generated by the speaker at the hydrophone position, with line 3 operating with one of its bubble curtains active.
$Lp_{4.1}$	Level generated by the speaker at the hydrophone position, with line 4 operating with one of its bubble curtains active.
$Lp_{1.1} + Lp_{2.1}$	Level generated by the speaker at the hydrophone position with line 1 and line 2 operating, with one of its bubble curtains active.
$Lp_{1.1} + Lp_{3.1}$	Level generated by the speaker at the hydrophone position with line 1 and line 3 operating, with one of its bubble curtains active.
$Lp_{2.1} + Lp_{3.1}$	Level generated by the speaker at the hydrophone position with line 2 and line 3 operating, with one of its bubble curtains active.
$Lp_{1.1} + Lp_{2.1} + Lp_{3.1}$	Level generated by the speaker at the hydrophone position, with line 1, line 2, and line 3 operating, with one of its curtains active.
$Lp_{1.1} + Lp_{4.1}$	Level generated by the speaker at the hydrophone position with line 1 and line 4 operating, with one of its bubble curtains active.
$Lp_{1.1} + Lp_{3.1} + Lp_{4.1}$	Level generated by the speaker at the hydrophone position, with line 1, line 3, and line 4 operating, with one of its curtains active.
$Lp_{1.1}$, 25 %	Level generated by the speaker at the hydrophone position, with line 1 operating with the diffuser 1 at 25 % of maximum airflow and the diffuser 2 at 0 % of its maximum airflow.
$Lp_{1.1}$, 50 %	Level generated by the speaker at the hydrophone position, with line 1 operating with the diffuser 1 at 50 % of maximum airflow and the diffuser 2 at 0 % of its maximum airflow.
$Lp_{1.1}$, 75 %	Level generated by the speaker at the hydrophone position, with line 1 operating with the diffuser 1 at 75 % of maximum airflow and the diffuser 2 at 0 % of its maximum airflow.
$Lp_{1.1}$, 100 %	Level generated by the speaker at the hydrophone position, with line 1 operating with the diffuser 1 at 100 % of maximum airflow and the diffuser 2 at 0 % of its maximum airflow.

5. Results

5.1. Measured sound pressure levels

5.1.1. Background noise levels Lp_{BN} and Lp_{BNB}

Table 4 shows the sound pressure level Lp_{BN} recorded at the hydrophone location when the sound source is off and all bubble curtains are disconnected. It also provides the sound pressure level Lp_{BNB}, recorded under the same conditions except for all bubble curtains active. This allows for an evaluation of the noise produced by the bubble curtains themselves.

Table 4. Background noise levels measured at the four positions of the hydrophones with the loudspeaker off.

	Hydrophone position				Measured level [dB re 1 µPa]
	P1	P2	P3	P4	
Lp_{BN}	117.9	117.6	118.3	118.0	118.0
Lp_{BNB}	119.8	119.9	120.4	120.2	120.1

5.1.2. Loudspeaker level Lp_s without active curtains

Table 5 shows the sound pressure level measured at the hydrophone position with the sound source operating and all bubble curtains deactivated.

Table 5. Sound pressure level measured at the hydrophone position with the loudspeaker driven by white noise and all bubble curtains disconnected.

	Hydrophone position				Measured level [dB re 1 µPa]
	P1	P2	P3	P4	
Lp_s	160.8	161.1	163.3	161.2	161.7
Lp_s	160.9	161.6	163.5	161.4	161.9
Average value					161.8

Figure 9 shows the frequency spectrum of the laboratory background noise, measured at the hydrophone position with no noise source present (Lp_{BN}). Similarly, it displays the frequency spectrum of the background noise with all bubble curtains active (Lp_{BNB}) and the frequency spectrum of the sound measured at the hydrophone position with the sound source driven by white noise, with no active curtain (Lp_s).

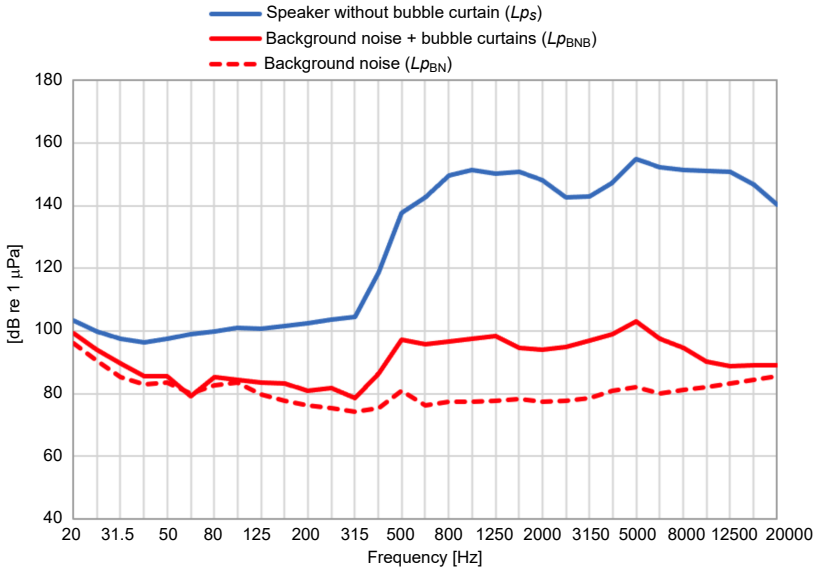


Fig. 9. Third-octave band frequency spectra of the sound source and background noise of the laboratory and air generation system.

Although the speaker specifications state that its frequency response ranges from 200 Hz to 23 kHz, Fig. 9 shows that below 200 Hz the speaker does emit sound and is capable of producing a sound pressure level at the hydrophone position higher than the background noise level. From 40 Hz upwards, the sound pressure level produced by the sound source at the hydrophone position is always higher than the background noise level, by 13 dB or more.

5.1.3. Levels measured for different curtain configurations

Table 6 shows the sound pressure levels at the hydrophone position during the different settings of the evaluated curtains, as shown in Table 3. We see that the sound pressure levels obtained by using the two bubble curtains or diffusers of line 1 ($Lp_{1.2}$) are very similar to the levels obtained by using only one of the bubble curtains of line 1 ($Lp_{1.1}$). This means that the line provides almost the same broadband attenuation in both cases. The data shows that reducing the air flow, from initially at a rate of 27 l/min, by half has virtually no effect on the achieved sound attenuation. Therefore, in all subsequent measurements, it was determined that only one of the two bubble curtains per line should be used.

Table 7 shows the levels obtained by modifying the percentage of the air flow of 27 l/min, applied to hose 1 of line 1. Figure 10 to Fig. 15 present frequency band attenuation for each bubble curtain configuration listed in Table 6.

Table 6. Sound pressure level measured at the hydrophone position with the loudspeaker driven by white noise and with different curtain configurations.

	Hydrophone position				Measured level [dB re 1 µPa]
	P1	P2	P3	P4	
$Lp_{1.1}$	150.6	150.3	150.4	150.2	150.4
$Lp_{1.2}$	150.2	150.0	150.4	149.9	150.1
$Lp_{3.1}$	151.2	151.0	150.9	150.8	150.9
$Lp_{4.1}$	151.2	150.9	151.0	150.2	150.8
$Lp_{1.1} + Lp_{2.1}$	149.4	149.4	149.6	149.7	149.5
$Lp_{1.1} + Lp_{3.1}$	147.8	147.8	147.7	148.1	147.9
$Lp_{2.1} + Lp_{3.1}$	149.7	149.2	1479.2	149.5	149.4
$Lp_{1.1} + Lp_{2.1} + Lp_{3.1}$	147.8	147.7	147.7	147.6	147.7
$Lp_{1.1} + Lp_{4.1}$	146.2	146.3	146.42	146.4	146.3
$Lp_{1.1} + Lp_{3.1} + Lp_{4.1}$	145.6	145.3	145.8	145.5	145.6

Table 7. Sound pressure level measured at the position P4 of the hydrophone, with the loudspeaker driven by white noise and modifying the air flow of diffuser 1 of line 1.

Hydrophone position P4			
$Lp_{1.1}$ 25 % [dB re 1 µPa]	$Lp_{1.1}$ 50 % [dB re 1 µPa]	$Lp_{1.1}$ 75 % [dB re 1 µPa]	$Lp_{1.1}$ 100 % [dB re 1 µPa]
155.4	152.6	152.0	150.8

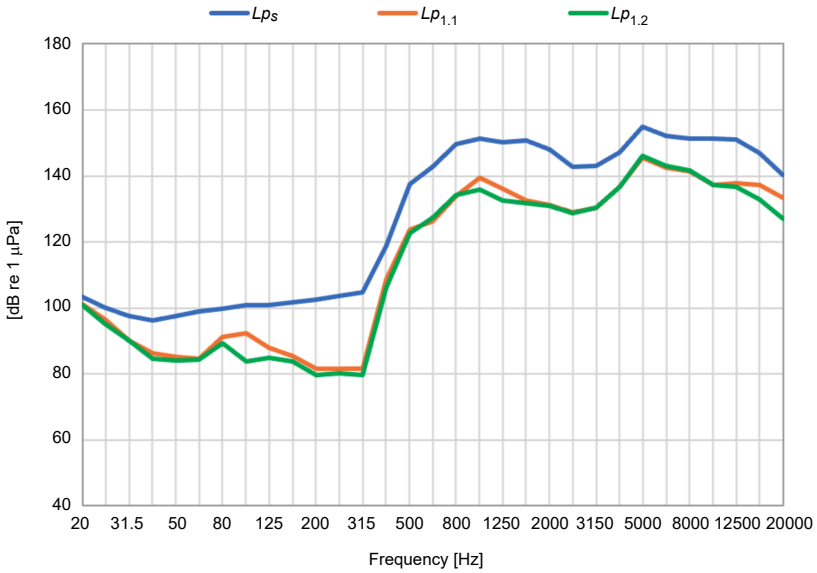


Fig. 10. Third-octave band frequency spectra obtained for curtain configurations Lp_s, $Lp_{1.1}$, and $Lp_{1.2}$ (see Table 3).

5.2. Insertion loss

The IL is estimated as the difference between the sound pressure level (Lp_s) measured without a curtain and the sound pressure level ($Lp_{x.x}$) measured when the curtain is active (Eq. (1)). Table 8 shows the broadband IL values obtained for the different combinations of bubble curtains. The results demonstrate that using a single line does not affect the broadband IL, regardless of the number of curtains per line ($Lp_{1.1}$ or $Lp_{1.2}$) or the distance of the line from the source or receiver. The various configurations of a single curtain result in an IL ranging from 11 dB to 12 dB. These attenuation values are consistent with some of the results obtained in field measurements (Nehls *et al.*, 2015). When using two lines, the greater separation between the lines results in the greater broadband IL, with IL values increasing by up to 15.5 dB. The use of three lines causes only a slight further increase in ILs compared to two lines, reaching a level of 16.2 dB.

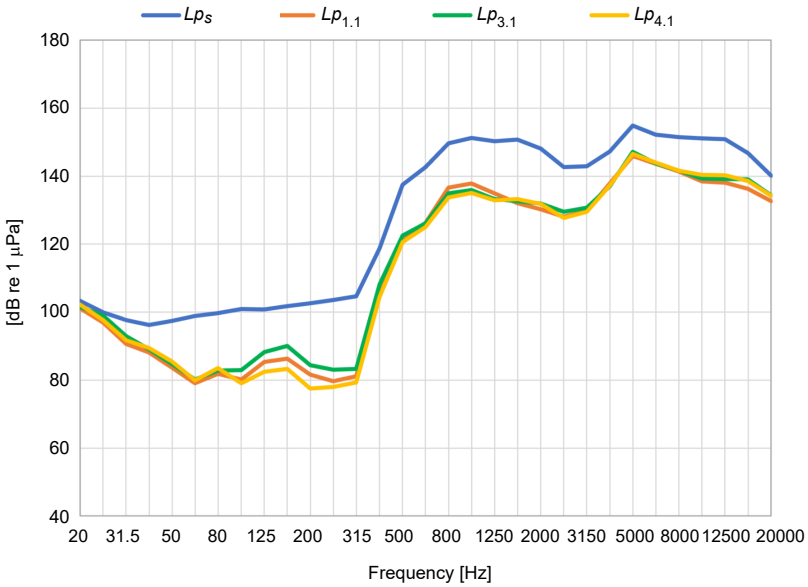


Fig. 11. Third-octave band frequency spectra obtained for curtain configurations Lp_s, $Lp_{1.1}$, $Lp_{3.1}$, and $Lp_{4.1}$ (see Table 3).

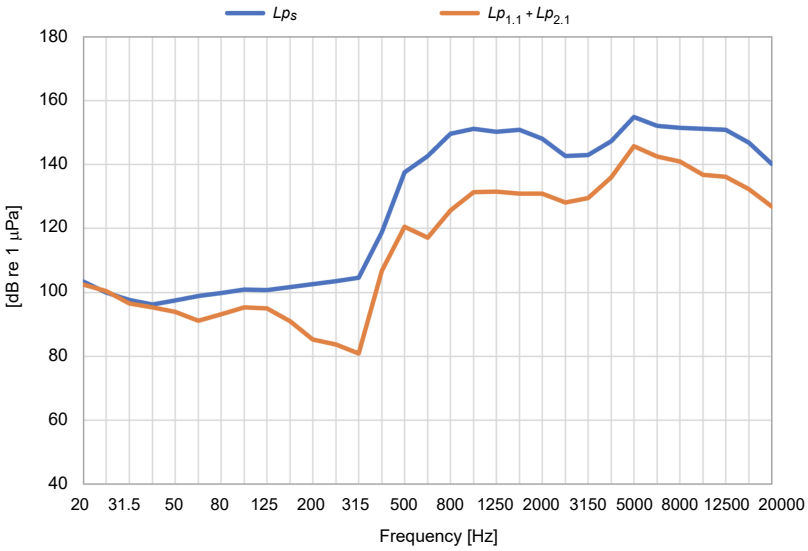


Fig. 12. Third-octave band frequency spectra obtained for curtain configurations Lp_s and $(Lp_{1.1} + Lp_{2.1})$.

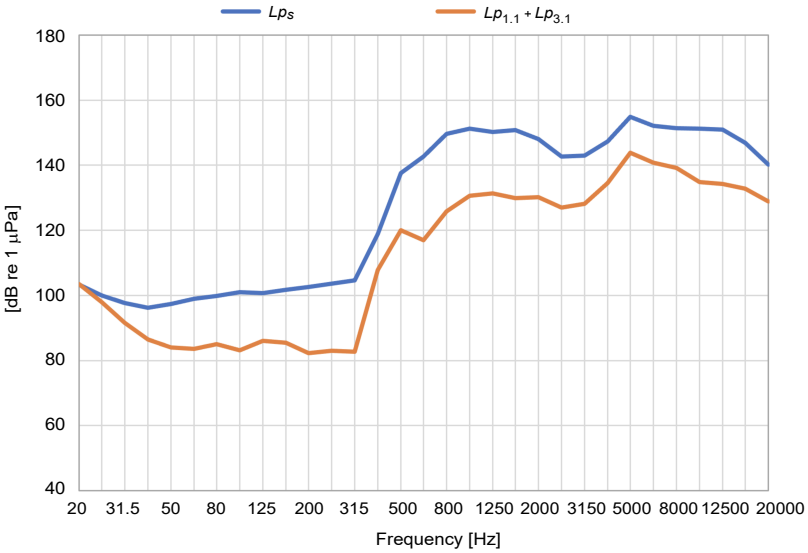


Fig. 13. Third-octave band frequency spectra obtained for curtain configurations Lp_s and $(Lp_{1.1} + Lp_{3.1})$.

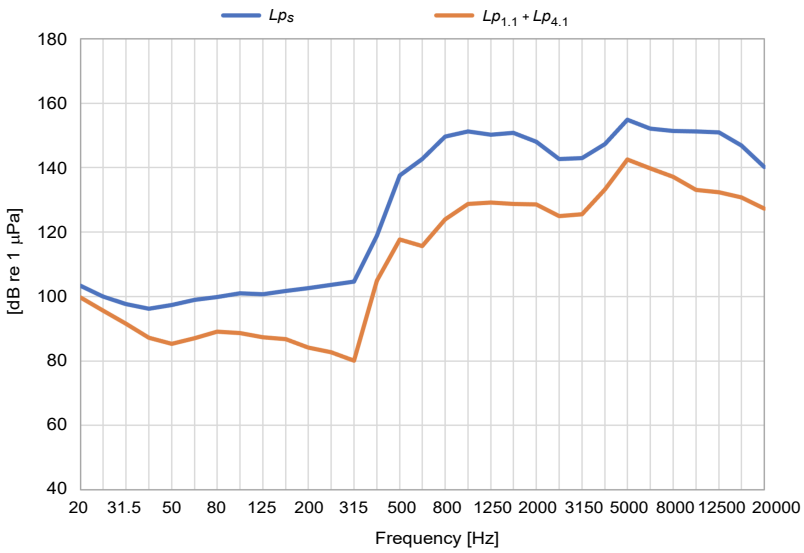


Fig. 14. Third-octave band frequency spectra obtained for curtain configurations Lp_s and $(Lp_{1.1} + Lp_{4.1})$.

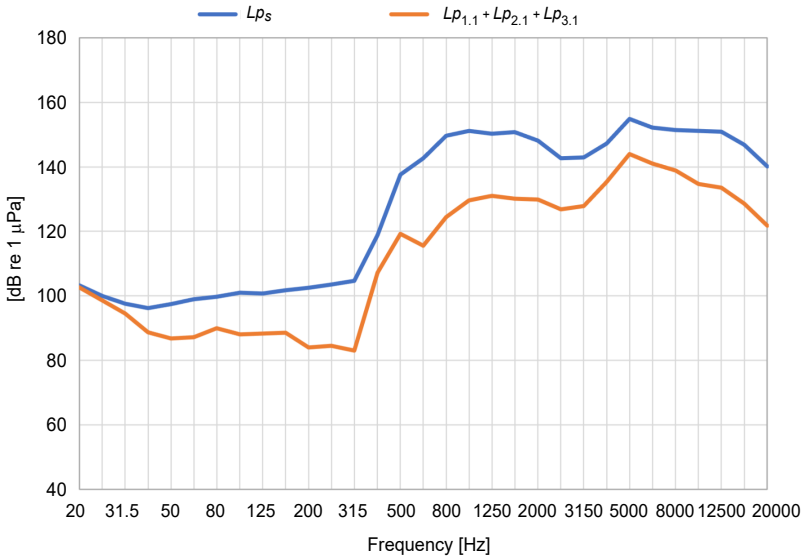


Fig. 15. Third-octave band frequency spectra obtained for curtain configurations Lp_s and $(Lp_{1.1} + Lp_{2.1} + Lp_{3.1})$.

It is imperative to establish the IL for each frequency band. As illustrated in Fig. 16, the IL per the third-octave band for line 1 is presented, considering both configurations: one active curtain $Lp_{1.1}$ and two active curtains $Lp_{1.2}$. As previously indicated in Table 6 and Table 8, the IL per frequency band provided by line 1 in both configurations

Table 8. Broadband IL for the different configurations of bubble curtains evaluated [dB re 1 µPa].

Curtain configuration evaluated	Lp_s [dB]	$Lp_{x.x}$ [dB]	Broadband IL [dB]
$Lp_{1.1}$	161.8	150.4	11.4
$Lp_{1.2}$	161.8	150.1	11.7
$Lp_{3.1}$	161.8	150.9	10.9
$Lp_{4.1}$	161.8	150.8	11.0
$Lp_{1.1} + Lp_{2.1}$	161.8	149.5	12.3
$Lp_{1.1} + Lp_{3.1}$	161.8	147.9	13.9
$Lp_{1.1} + Lp_{4.1}$	161.8	146.3	15.5
$Lp_{1.1} + Lp_{2.1} + Lp_{3.1}$	161.8	147.7	14.1
$Lp_{1.1} + Lp_{3.1} + Lp_{4.1}$	161.8	145.6	16.2

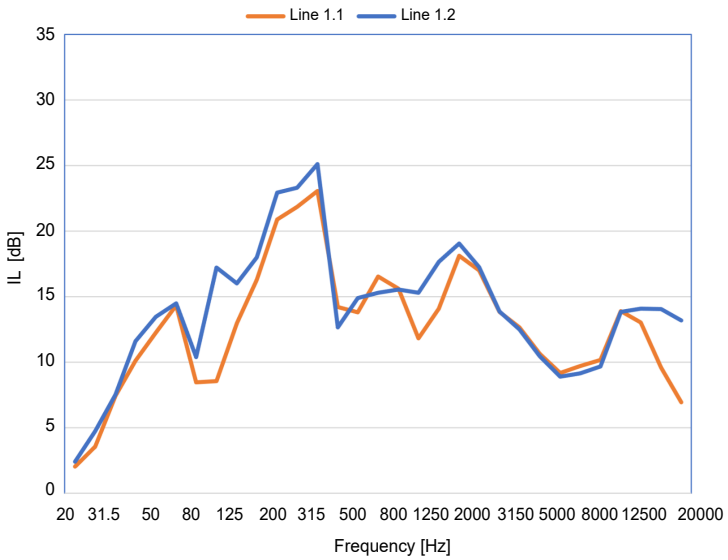


Fig. 16. IL by frequency bands given by line 1, configurations $Lp_{1.1}$ and $Lp_{1.2}$.

is practically the same. The $Lp_{1.2}$ configuration is the most prevalent in salmon farming, where its attenuation is equal to or greater than 10 dB in virtually all frequency bands from 40 Hz upwards. This minimum attenuation is consistent with that reported in another study for similar curtains (Rustemeier *et al.*, 2012).

Figure 17 shows that, for the case of a single line, the position of the line between the source and the receiver will produce a change in the IL per frequency band. The closer the curtain is to the receiver, the greater the attenuation at low frequencies. This can be explained as an increase in the attenuation of low-frequency components mechanically transmitted through the bottom of the test channel.

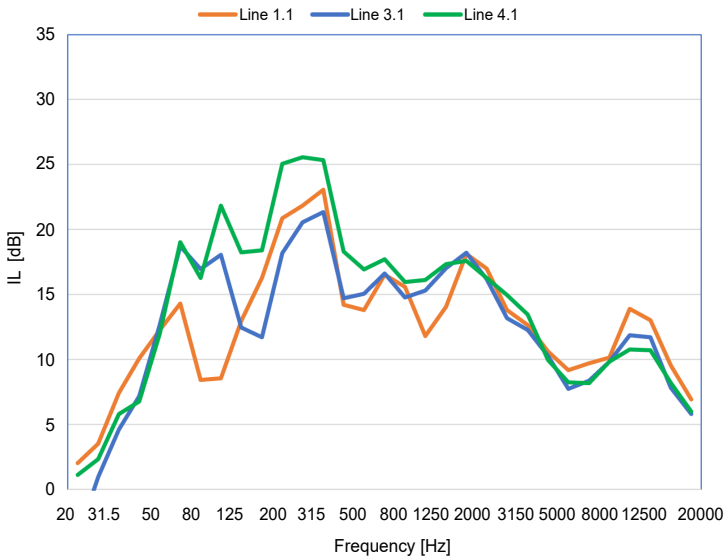


Fig. 17. IL by frequency bands given by line 1, configurations $Lp_{1.1}$, $Lp_{3.1}$, and $Lp_{4.1}$.

The employment of dual lines as opposed to a single line results in an increase in the IL per frequency band, with an increase from 400 Hz. The attenuation is directly proportional to the separation of the lines. As illustrated in Fig. 18, utilising two lines instead of a single line results in a substantial enhancement of the IL, with the maximum occurring within the frequency range from 400 Hz to 2 kHz. This leads to IL values ranging from 15 dB to 27 dB. This frequency range corresponds with another study of bubble curtains (Beelen *et al.*, 2025). However, there, the IL values are considerably lower due to the limitations imposed by the low signal-to-noise ratio (SNR) employed.

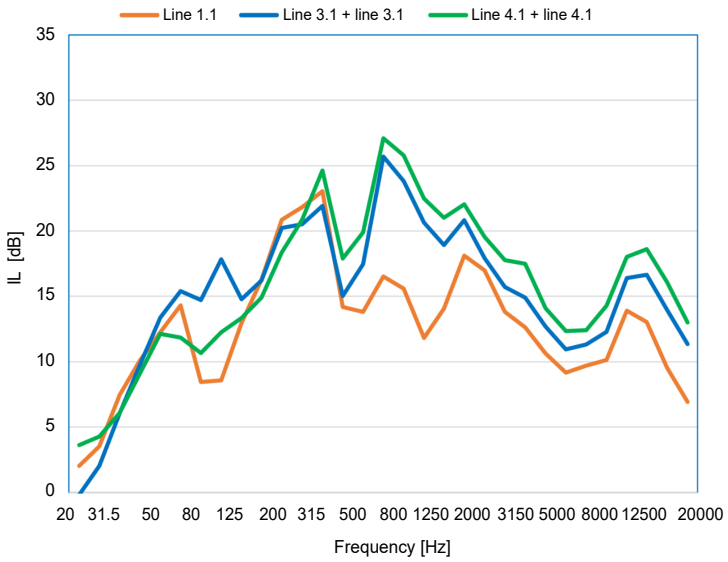


Fig. 18. IL by frequency bands given by line 1, configurations ($Lp_{1.1} + Lp_{3.1}$) and ($Lp_{1.1} + Lp_{4.1}$).

5.3. Air-bubble curtains and acoustic impact reduction

These days, bubble curtains are extensively employed as an underwater noise mitigation system; however, the extent to which the noise level of the sound source will be mitigated remains uncertain. The results of this study demonstrate that the evaluated bubble curtain will provide the IL attenuation greater than or equal to 10 dB in all frequency bands and all configurations. To comprehend the implications of a 10 dB reduction of a noise level, one must consider the acoustic impact area that would result from, for instance, pile driving operations, with and without the use of bubble curtains, as evaluated in this study. In order to assess the potential impact of the project, e.g., on marine species, it is first necessary to estimate the sound pressure level to which the species under study will be exposed. Equation (2) provides the noise level for each frequency band that a noise source generates at a given distance, considering the sound propagation model characteristic of the site:

$$Lr = \mathrm{SL} - N\log(r) - \alpha \cdot r \quad [\mathrm{dB}], \tag{2}$$

where Lr is the sound level per frequency band at a distance r [dB re 1 µPa], SL is the source level per frequency band (L_{eq}, Lp_{rms}, SEL) [dB re 1 µPa·m or dB re 1 µPa^2·s·m], N is the local attenuation ratio, r is the distance from the source to the receiver [m], α is the sound absorption per frequency band [dB/m].

In deep waters and in absence of sound channels, a spherical propagation is considered with $N = 20$. In presence of sound channels, a cylindrical propagation is considered with $N = 10$.

In shallow waters, sound propagation involves a large number of reflections of the acoustic signal at the surface and seabed, making it difficult to determine the N value that correctly describes the propagation (Jastrzębski, 2007; Khalilabadi, 2022; Lippert *et al.*, 2018). In the case of shallow waters, field measurements of attenuation at various distances from a noise source are typically conducted to obtain an empirical approximation of N, where values ranging from 13 to 35 are typically observed (Richardson *et al.*, 1995). As demonstrated in the relevant literature, $N = 15$ has been shown to provide a high degree of fit with a significant number of measurements in shallow waters (MMO, 2015; NMFS, 2021; Molnar *et al.*, 2020). Because of this, to describe sound propagation in our example a value $N = 15$ will be used.

Once the value of N has been selected or determined, the level Lr at different distances r from a source of known level SL can be predicted. Furthermore, the maximum distance at which the emitted sound can be perceived without generating negative effects in the receiver can be determined. To evaluate the physiological impact on the considered marine species, the sound levels per frequency band Lr obtained from Eq. (2), or the source levels SL of the spectrum, must be weighted by the auditory weighting functions of each animal.

Subsequently, a comparison of the results with the highest permitted thresholds according to the applied criterion is required. These thresholds are indicated in Table 2, corresponding to specific species of marine mammals.

Pile driving is a significant source of underwater noise, generating high levels of sound pressure. In the context of construction, this noise source is typically the primary focus of attention in the initial phase of a project, as it is often the most distinct. Pile driving can easily reach broadband levels close to 200 dB re 1 µPa at a distance of 1 m. Table 9 presents the levels emitted at a distance of 10 m during the driving of a steel pile with a diameter of 1.52 m. It should be noted that each strike lasted approximately 0.1 s on average.

Table 9. Sound pressure levels generated during the driving of 1.52 m diameter piles, measured at a distance of 10 m and with a strike duration of 0.1 s (URS, 2011; Rodkin, Pommerenck, 2014).

Source	Lp_{peak} [dB re 1 µPa]	Lp_{rms} [dB re 1 µPa]	SEL [dB re 1 µPa^2 · s]
Pile ∅1.52 m	210	195	185

By normalizing these values to a distance of 1 m from the source, the SL can be obtained for the various noise descriptors, as demonstrated in Table 10.

Table 10. Estimated sound pressure levels at a distance of 1 m when driving a pile with a diameter of 1.52 m. Calculated considering a sound propagation between spherical and cylindrical 15 log(r) (Molnar *et al.*, 2020).

Source	Lp_{peak} [dB re 1 µPa]	Lp_{rms} [dB re 1 µPa · m]	SEL [dB re 1 µPa^2 · s · m]
Pile ∅1.52 m	225	210	200

For instance, consider a pile driving rate of three piles per day, with an estimated 270 blows required for each pile. This results in a total of 810 blows, and the unweighted cumulative noise exposure level over a 24-hour period is SEL_{cum24h} = 229.1 dB.

Figure 19 presents the third-band used for our noise source, based on data from 1.6 m diameter pile driving (Nehls *et al.*, 2007). The amplitude was adjusted so its broadband level matches the cumulative sound exposure level in our example.

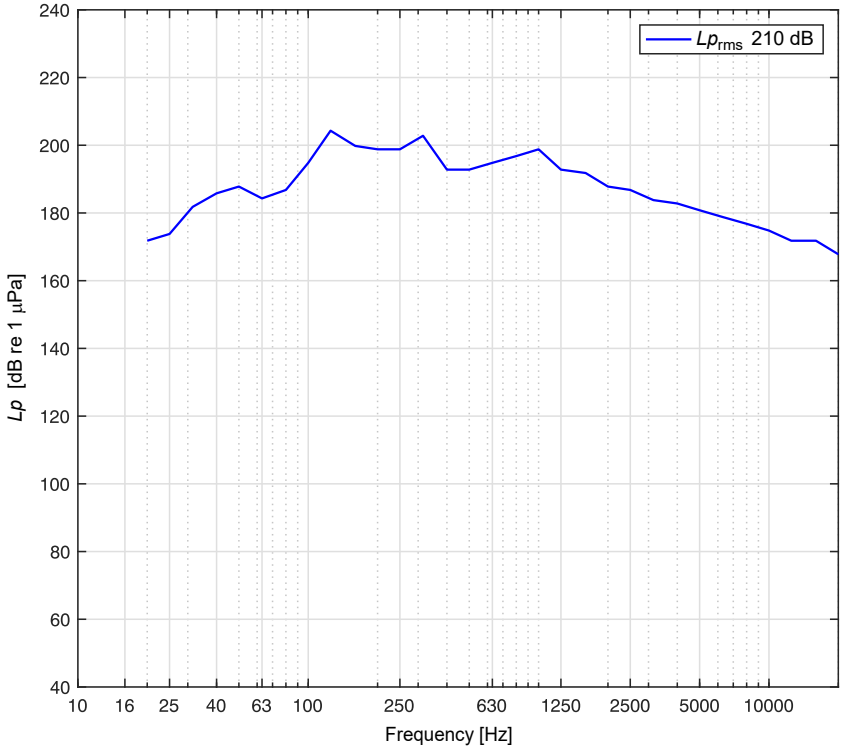


Fig. 19. Third-octave band frequency spectrum obtained for driving 1.6 m diameter piles (Nehls *et al.*, 2007).

Table 11 summarizes the distances at which the TTS and AUD INJ thresholds are reached for five groups of marine mammals due to pile driving of 1.52 m, according to the NMFS (2024) criteria with auditory weighting for the five animal groups. Figure 20 and Fig. 21 show the distance dependence of the weighted SEL for LF and HF mammals, respectively, and the threshold distances assuming an attenuation ratio of $15\log(r)$.

Table 11. Evaluation of physiological effects as a result of driving 1.52 m diameter piles without the use of air bubble curtains. Auditory injury INJ and temporal threshold shift TTS according to NMFS (2024) criterion. Estimated considering sound propagation as $15\log(r)$ (Molnar *et al.*, 2020), [dB re 1 μPa$^2\cdot$s].

Group	SEL_{cum24h} weighted [dB]	Threshold TTS [dB]	Distance TTS [m]	Threshold AUD INJ [dB]	Distance AUD INJ [m]
LF	227.2	168	5490	183	821
HF	214.7	178	273	193	28
VHF	201.3	144	3497	159	588
OW	212.4	170	630	185	67
PW	219.6	168	2231	183	268

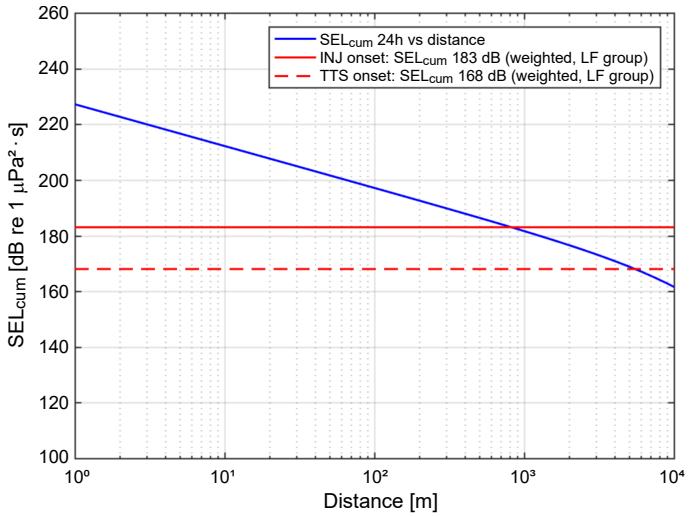


Fig. 20. Physiological impact ranges of sound emissions during pile driving on low-frequency (LF) marine mammals. Attenuation factor $15\log(r)$.

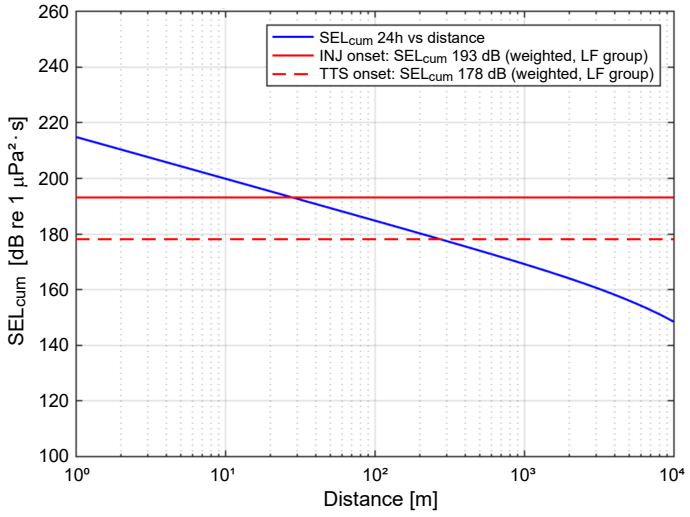


Fig. 21. Physiological impact ranges of sound emissions during pile driving on high-frequency (HF) marine mammals. Attenuation factor $15\log(r)$.

Using a bubble curtain as tested here can reduce pile driving noise by at least 10 dB per frequency band, even in its basic form $Lp_{1.2}$. Using a conservative criterion, the spectrum levels in Fig. 19 can be reduced by 10 dB,

thus reducing the SEL_{cum24h} level from 229.1 dB to 219.1 dB. Table 12 and Fig. 22 illustrate how such attenuation decreases distances of occurrence of exceeding TTS and AUD INJ thresholds for high-frequency cetaceans.

Table 12. Assessment of physiological effects on the species considered as a result of driving 1.52 m diameter piles and using air bubble curtains as a mitigation measure. Estimated considering a sound propagation between spherical and cylindrical $15\log(r)$ (Molnar *et al.*, 2020), [dB re 1 $\mu Pa^2 \cdot s$].

Group	SEL_{cum24h} weighted [dB]	Threshold TTS [dB]	Distance TTS [m]	Threshold AUD INJ [dB]	Distance AUD INJ [m]
LF	217.2	168	1183	183	177
HF	204.7	178	59	193	6
VHF	191.3	144	753	159	126
OW	202.4	170	136	185	14
PW	209.6	168	481	183	58

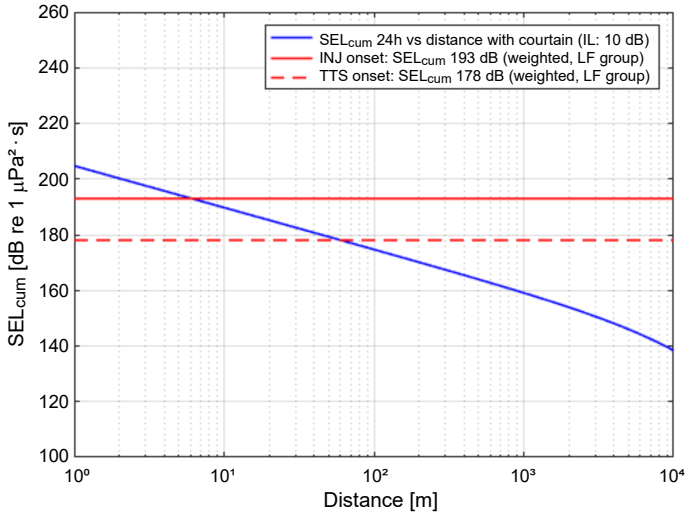


Fig. 22. Evaluation of the physiological effects on high frequency cetaceans HF due to pile driving and the use of an air bubble curtain (IL: 10 dB).

By comparing Table 11 and Table 12, it is clear that with a conservative approach, employing an air bubble curtain that provides at least 10 dB of attenuation in each frequency band leads to at least a 78.4 % reduction in the distances at which the TTS and AUD INJ thresholds are reached. This equates to a substantial decrease in the effects of pile driving on the region's native marine fauna.

6. Discussion

The measurements were conducted under favorable background noise conditions, yielding a SNR of at least 13 dB for frequencies above 40 Hz. The background noise level reached 120.1 dB with the bubble generation system active and 118 dB with the system inactive, both referenced to 1 µPa. Sound levels produced by the loudspeaker were measured at four distinct hydrophone positions to reduce potential wave cancellation within the test channel, thereby enhancing the signal-to-background noise ratio. On average, the loudspeaker generated a broadband level of 161.8 dB at the measurement points, resulting in a SNR of 41.7 dB.

With respect to attenuation achieved by a single line, in the case of line 1 – whether employing a single active bubble curtain $Lp_{1.1}$ or two active bubble curtains $Lp_{1.2}$ – delivers a broadband IL between 11 dB and 12 dB (Table 8). This suggests that the line can maintain its sound attenuation performance using only one hose. The dual curtain setup $Lp_{1.2}$ yields higher IL at some frequencies but does not notably improve broadband IL overall (Fig. 16).

Considering individual lines attenuation, line 1, line 3, and line 4 exhibit similar broadband IL, each presenting values in the region of 11 dB (Table 8). This observation indicates that the specific distance of a line from either the noise source or the receiver does not significantly influence the broadband IL. Despite the broadband similarities, notable differences in the IL are observed in third-octave frequency bands (Fig. 17). Specifically, the attenuation at low frequencies increases in the case of lines positioned closer to the receiver. This phenomenon can be attributed to the mechanical transmission of noise through the bottom of the test channel, followed by its re-radiation into the water. It is important to note that if this transmission pathway through the channel floor were eliminated, the frequency-dependent IL data would likely converge and become more uniform across the different line positions.

When evaluating the use of two lines, the broadband IL varies between 12.3 dB for configuration $Lp_{1.1} + Lp_{2.1}$ and 15.5 dB for configuration $Lp_{1.1} + Lp_{4.1}$. This suggests that a greater separation between the lines results in a higher IL. Addition of a third line between two active lines leads to a modest increase in total IL, with an enhancement of less than 1 dB. In these cases, the maximum IL reaches 16.2 dB (Table 8 for configurations $Lp_{1.1} + Lp_{2.1} + Lp_{3.1}$ and $Lp_{1.1} + Lp_{3.1} + Lp_{4.1}$).

For the third-octave band IL, single lines offer an IL between 10 dB and 25 dB, ensuring that single lines always achieve at least 10 dB of broadband attenuation. Enhancing the SNR at frequencies below 40 Hz could further raise this minimum attenuation. In contrast, double lines provide the IL from 11 dB to 27 dB and cover a broader frequency range than single lines. Their highest attenuation is between 400 Hz and 2 kHz.

Comparing Table 11 and Table 12 reveals that using a bubble curtain achieving 10 dB IL in each frequency band reduces the distance at which physiological damage thresholds are reached by 78.4 %.

For the LF cetacean group – which includes species like blue, humpback, and fin whales – Table 11 indicates that any individual within 821 m of pile driving activity risks auditory injury (AUD INJ) during a 24-hour exposure. With air bubble curtains providing 10 dB noise attenuation, this risk zone is reduced from 821 m to 177 m (refer to Table 12). For threshold shift (TTS) in the LF group, the affected distance decreases from 5490 m to 1183 m.

In the HF group individuals within distance below 3497 m of pile driving risk TTS, and those within 588 m risk AUD INJ; employing bubble curtains reduces these distances to 753 m and 126 m, respectively (see Table 11 and Table 12).

7. Conclusion

The results of this study demonstrate that the bubble curtains evaluated function effectively as acoustic barriers and can serve as a viable underwater noise mitigation measure, aligning with the new requirements of the Environmental Assessment Service for controlling high energy anthropogenic underwater noise.

Laboratory measurements of IL enabled systematic comparison of different bubble curtain configurations under stable and controlled conditions, thereby avoiding the variability inherent in field measurements.

Across all tested configurations, the air bubble curtains achieved the IL of at least 10 dB in every frequency band. Applying a conservative criterion, the evaluated bubble curtain – when deployed as a mitigation measure for pile driving – would provide a minimum attenuation of 10 dB per frequency band, even in its simplest configuration. This reduction translates into a decrease of at least 78.4 % in the physiological risk area for marine mammals.

It should be noted, however, that attenuation performance in the field may vary depending on factors such as SNR, seabed sound transmission, and the variation of the structural integrity of the bubble curtain with depth and ocean currents.

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.

Authors' contributions

Alfio Yori directed the study, defined the measurement methodology, and wrote the original draft. José Luis Barros performed the analysis and contributed to the interpretation of the data. Rodrigo Torres performed the analysis and the measurements. Felipe Figueroa performed the measurements. All authors reviewed and approved the final manuscript.

References

1. Beelen S., Nijhof M., de Jong C., van Wijngaarden L., Krug D. (2025), Bubble curtains for noise mitigation: One vs two, *The Journal of the Acoustical Society of America*, **157**(2): 1336–1355, https://doi.org/10.1121/10.0035817.
2. Bellmann M.A. (2014), Overview of existing noise mitigation systems for reducing pile-driving noise, [in:] *Inter-Noise and Noise-Con Congress and Conference Proceedings*, **249**: 2544–2554.
3. Department of Arts, Heritage and the Gaeltacht [DAHG] (2014), *Guidance to manage the risk to marine mammals from man-made sound sources in Irish waters*.
4. Department of Planning, Transportation and Infrastructure [DPTI] (2012), *Underwater pilling noise guide*, Government of South Australia.
5. Dähne M., Tougaard J., Carstensen J., Rose A., Nabe-Nielsen J. (2017), Bubble curtains attenuate noise from offshore wind farm construction and reduce temporary habitat loss for harbour porpoises, *Marine Ecology Progress Series*, **580**: 221–237, https://doi.org/10.3354/meps12257.
6. Environmental Assessment Service [SEA] (2022), *Evaluation criteria in the Environmental Impact Assessment System (SEIA): Prediction and evaluation of underwater noise impacts* [in Spanish: *Criterio de evaluación en el sistema de evaluación de impacto ambiental SEIA: Predicción y evaluación de impactos por ruido submarino*].
7. Fahy F.J. (2001), *Foundations of Engineering Acoustics*, Elsevier Press.
8. Göttsche K.M., Juhl P.M., Steinhagen U. (2013), Numerical prediction of underwater noise reduction during offshore pile driving by a small bubble curtain, [in:] *Proceedings Internoise: Noise Control for Quality of Life*.
9. Hall M.V. (1989), A comprehensive model of wind generated bubbles in the ocean and predictions of the effects on sound propagation at frequencies up to 40 kHz, *The Journal of the Acoustical Society of America*, **86**(3): 1103–1117, https://doi.org/10.1121/1.398102.
10. Jastrzębski S. (2007), Application of time reversal technique in shallow water environment, *Archives of Acoustics*, **32**(4(S)): 319–324.
11. Joint Nature Conservation Committee [JNCC] (2010), *Statutory nature conservation agency protocol for minimising the risk injury to marine mammals from piling noise*, Aberdeen, UK.
12. Koschinski S., Lüdemann K. (2013), *Development of Noise Mitigation Measures in Offshore Wind Farm Construction*, Federal Agency for Nature Conservation.
13. Khalilabadi M.R. (2022), 2D modeling of wave propagation in shallow water by the method of characteristics, *Archives of Acoustics*, **47**(3): 407–412, https://doi.org/10.24425/aoa.2022.142014.
14. Lippert T., Ainslie M., von Estorff O. (2018), Pile driving acoustics made simple: Damped cylindrical spreading model, *The Journal of the Acoustical Society of America*, **143**(10): 310–317, https://doi.org/10.1121/1.5011158.
15. Lucke K., Lepper P.A., Blanchet M.A., Siebert U. (2011), The use of an air bubble curtain to reduce the received sound levels for harbor porpoises (Phocoena phocoena), *The Journal of the Acoustical Society of America*, **130**(5): 3406–3412, https://doi.org/10.1121/1.3626123.

16. Lucke K., Siemensma M. (2013), *International regulations on the impact of pile driving noise on marine mammals – A literature review*, Institute for Marine Resources and Ecosystem Studies.

17. McKenzie C. (2015), *Smålandsfarvandet offshore wind farm underwater noise*, Technical report, Energinet.dk.

18. Merck T., Werner S., Koschinski S., Lüdemann K. (2014), *Inventory of measures to mitigate the emission and environmental impact of underwater noise*, Technical report, OSPAR Commission.

19. MMO (2015), *Modelled mapping of continuous underwater noise generated by activities. A report produced for the Marine Management Organization*, MMO Project No: 1097.

20. Molnar M., Buehler D., Oestman R., Reyff J., Pommerenck K., Mitchell B. (2020), *Technical guidance for the assessment of hydroacoustic effects of pile driving on fish*, Division of Environmental Analysis California Department of Transportation.

21. Möser M., Barros J.L. (2004), *Acoustic Engineering, Theory and Applications* [in Spanish: *Ingeniería Acústica, Teoría y Aplicaciones*], Springer.

22. National Marine Fisheries Service (2021), *Guidelines for non-lethally deterring marine mammals.*

23. National Marine Fisheries Service (2024), *Update to: Technical guidance for assessing the effects of anthropogenic sound on marine mammal hearing (Version 3.0): Underwater and in-air criteria for onset of auditory injury and temporary threshold shifts*, U.S. Department of Commerce, NOAA. NOAA Technical Memorandum NMFS-OPR-71.

24. Nedwell J.R., Edwards B., Turnpenny A.W.H., Gordon J. (2004), *Fish and marine mammal audiograms: A summary of available information*, Subacoustech Report ref: 534R0214.

25. Nedwell J.R. *et al.* (2007), *A validation of the dB_{ht} as a measure of the behavioural and auditory effects of underwater noise*, Subacoustech Report No. 534R1231.

26. Nehls G., Betke K., Eckelmann S., Ros M. (2007), *Assessment and costs of potential engineering solutions for the mitigation of the impacts of underwater noise arising from the construction of offshore windfarms* (Report No. ENG-01-2007), Report by BioConsult SH. Report for Collaborative Offshore Wind Research into the Environment (COWRIE).

27. Nehls G., Rose A., Diederichs A., Bellmann M., Pehlke H. (2015), *Noise mitigation during pile driving efficiently reduces disturbance of marine mammals*, *Advances in Experimental Medicine and Biology*, **875**: 755–762, https://doi.org/10.1007/978-1-4939-2981-8_92.

28. Novarini J.C., Keiffer R.S., Norton G.V. (1998), A model for variations in the range and depth dependence of the sound speed and attenuation induced by bubble clouds under wind-driven sea surfaces, *IEEE Journal of Oceanic Engineering*, **23**(4): 423–438, https://doi.org/10.1109/48.725236.

29. Parvin S.J., Nedwell J.R. (1995), Underwater sound perception and the development of an underwater noise weighting scale, *Underwater Technology*, **21**(1): 12–19, https://doi.org/10.3723/175605495783328836.

30. Qi C., Hu Z., Wang S., Jiang Z., Wu H., Yin Z. (2021), Study on the propagation law of water hammer wave in underwater blasting and the reducing effect of air curtain on water hammer wave, *Shock and Vibration*, https://doi.org/10.1155/2021/9994071.

31. Richardson W.J., Greene C.R., Malme C.I., Thomson D.H. (1995), *Marine Mammals and Noise*, Academic Press, New York.

32. Robinson S.P., Lepper P.A., Hazelwood R.A. (2014), *Good practice guide for underwater noise measurement*, National Measurement Office, Marine Scotland, The Crown Estate, NPL Good Practice Guide No. 133.

33. Rodkin R., Pommerenck K. (2014), Caltrans compendium of underwater sound data from pile driving, [in:] *Inter-Noise 2014*.

34. Rustemeier J., Griessmann T., Rolfes R. (2012), Underwater sound mitigation of bubble curtains with different bubble size distributions, *Proceedings of Meetings on Acoustics*, **17**(1): 070055, https://doi.org/10.1121/1.4772936.

35. Strietman W.J., Michels R., Leemans E. (2018), *Measures to reduce underwater noise and beach litter: An assessment of potential additional measures for the Netherlands*, Wageningen Economic Research, Report 2018-087, https://doi.org/10.18174/457443.

36. Tsouvalas A. (2020), Underwater noise emission due to offshore pile installation: A review, *Energies*, **13**(12): 3037, https://doi.org/10.3390/en13123037.

37. Urick R.J. (2010), *Principle of Underwater Sound*, 3rd ed., Peninsula Publishing.

38. URS (2011), *Report marine noise assessment*, Northern Territory Department of Lands and Planning, Australia.

39. Werner S. (2010), *Towards a precautionary approach for regulation of noise introduction in the marine environment from pile driving* [in German], Federal Environmental Agency, Stralsund, Germany.

40. Würsig B., Greene C.R., Jefferson T.A. (2000), Development of an air bubble curtain to reduce underwater noise of percussive piling, *Marine Environmental Research*, **49**(1): 79–93, https://doi.org/10.1016/S0141-1136(99)00050-1.

41. Yori A.F. (2018), Underwater assessment of anthropogenic noise sources using a field recording method, *Acta Acustica United with Acustica*, **104**(1): 13–24, https://doi.org/10.3813/AAA.919141.

Archives of Acoustics Vol. 51, No. 1, pp. 183–196 (2026), https://doi.org/10.24423/archacoust.2026.4286

Technical Note

Similarity Analysis and Target Strength Correction of Non-Proportional Scaled Models Based on Highlight Model Theory

Jin Zhang[(1),(2)], Zhenyu Li[(3)], Yinhao Li[(3)], Zilong Peng[(3)*]

[(1)] *Beijing Institute of Technology*
Beijing, China

[(2)] *Jiangsu Automation Research Institute*
Lianyungang, China

[(3)] *School of Energy and Power, Jiangsu University of Science and Technology*
Zhenjiang, China

*Corresponding Author: zlp_just@sina.com

Received July 23, 2025; revised December 30, 2025; accepted December 30, 2025; version of record March 18, 2026; published issue March 27, 2026.

Acoustic scattering scale models often fail to meet the acoustic similarity design requirements due to limitations in fabrication technology, testing facilities, and safe transportation, which restrict the accurate extrapolation of acoustic scattering characteristics between scaled models and full-scale ships. To overcome this challenge, the present study applies highlight model theory to perform acoustic similarity analysis and to correct the local target strength of simple objects in the model based on overall acoustic scattering correction. A novel method for correcting the target strength of non-proportional scaled models is proposed. The method is validated using various model geometries, including ellipsoids, finite-length cylinders, truncated elliptical cones, and complex structures. Additionally, the plate element method is employed for target strength correction and scaling conversion analysis for non-proportional scaled models. The study highlights the variation in target strength due to changes in geometric dimensions and demonstrates the effectiveness of the proposed correction method. The results indicate that the proposed correction approach allows for more accurate extrapolation of target strength from non-proportional scaled models to full-scale prototypes, thereby better satisfying the requirements of practical engineering applications.

Keywords: acoustic scattering, similarity theory, highlight theory, scaled model.

1. Introduction

Underwater target acoustic scattering experiments are categorized into full-scale ship tests and scaled model tests. Scaled model tests involve conducting experiments using a scaled-down model under controlled conditions (Tang *et al.*, 2018). These tests offer enhanced controllability and repeatability, making them ideal for in-depth analysis of scattering mechanism. Gao and Lin (2019) performed pool tests on a novel underwater vehicle, designing and validating a system capable of five degrees of freedom movement, thus confirming the feasibility of using fan-shaped wings underwater. Ke *et al.* (2024) investigated the acoustic scattering characteristics of a propeller through lake tests, validating a static acoustic scattering prediction method for ship propellers with complex linear geometries. Liu *et al.* (2023) conducted acoustic scattering tests on a 1/16 scaled model of an underwater vehicle, validating the accuracy of a patch-correction highlight model prediction approach. Xue *et al.* (2023) performed acoustic target strength tests on a benchmark scaled model, verifying the reliability of

the curved element method through comparison with the plate element method. Zhou *et al.* (2021) designed a ship shafting propulsion system using aerodynamic scaling relations, confirming the design's feasibility through simulations and experiments. Yu *et al.* (1999; 2002a) carried out similarity tests on structural vibrations in air and sound radiation in water for a ribbed cylindrical shell model, establishing corresponding similarity scaling relationships.

Scaled model multi-base elastic wave tests enable the analysis of elastic wave generation mechanisms and propagation paths (Agounad *et al.*, 2020; 2023; Ting *et al.*, 2019). Stulenkov *et al.* (2024) examined the acoustic field of a propeller, focusing on the resonance frequencies of its blades, using a combination of numerical simulations and experimental physical modeling. They also demonstrated the influence of the Q-factor on the radiation spectrum through propeller models made from different materials. Yan and Wang (2025) developed a composite structure model using thin-film metamaterials for low-frequency underwater sound absorption. The model was fabricated and validated through numerical simulations and experiments, and the effects of material properties, thickness, and mass block distribution on acoustic performance were analyzed, providing new insights for underwater stealth technology. Tao *et al.* (2025) proposed a deep learning model based on low-frequency analysis and recording (LOFAR) spectral features to mitigate the impact of ship underwater radiated noise (VURN) on underwater acoustic communication. Simulations and experiments demonstrated that the proposed method offers greater robustness and lower bit error rates under varying signal-to-noise ratios. Feng *et al.* (2025) introduced a time-difference-of-arrival (TDOA) sequence estimation method based on a hidden Markov model (HMM) for passive positioning systems with wideband underwater moving sound sources and distributed hydrophones. The effectiveness of the method was validated through simulations and lake experiments. Although these studies differ in their research objectives and methodologies, they all employ scale model experiments to empirically validate theoretical analyses, numerical simulations, or design concepts under controlled conditions. However, they focus exclusively on geometrically scaled models rather than non-geometrically scaled models.

Non-proportional scaled models can lead to inaccurate conclusions or data deviations, undermining the credibility and scientific value of research. To mitigate the impact of non-proportional scaling on experiments, large-scale models are often employed to simplify model fabrication. For example, the United States developed the large-scale 'Kamloops' model for sonar fairing design, while the large scale vehicle (LSV) series models were also used for testing to enhance design (Liu *et al.*, 2017). Wang *et al.* (2024) proposed a novel similarity method for box girders subjected to combined effects of bending moments and lateral pressure, which disregards the similarity criteria of cross-sectional properties and instead focuses solely on strength-based similarity criteria. Using this approach, similarity relations for the ultimate moment and applied lateral pressure were derived, and a step-by-step procedure for scale model design was proposed. Ge *et al.* (2024) introduced a discretized macro-element (DME) method that integrates experimental, numerical, and analytical techniques to rapidly assess the crushing behavior and protective performance of crashworthy devices in ship collisions. Based on similarity laws, the crush curve for large crashworthy devices was derived, and the advantages of the DME method in accurately capturing local deformation and improving computational efficiency were demonstrated. Liu *et al.* (2024) conducted wave pool experiments to investigate the hydrodynamic response characteristics of a scaled model of a submerged floating tunnel (SFT) and the effects of local cable breakage. The results indicated that under lower wave excitation, self-vibration of the SFT becomes more pronounced, and the response is more sensitive to variations in cable force, which is also influenced by immersion depth. Li *et al.* (2020) developed a method to assess the structural performance of beams on ultra-large container ships. A scaled model testing system based on Buckingham's theorem and Froude scaling law was implemented, and systematic experimental and numerical analyses were conducted on the stay cables, providing substantial support for the design and optimization of ultra-large cable-stayed bridges. To address potential biases introduced by non-proportional scaled models, current research is shifting from traditional geometric similarity toward a deeper focus on physical mechanism similarity, with an emphasis on developing tailored similitude criteria for specific structural responses.

Yu and Lu (2001) and Yu *et al.* (2002b) employed the energy method to analyze the loss factor, surface sound reflection, and scale effects caused by small mass blocks in large-scale acoustic models of underwater vehicles, controlling these effects through model design. However, the fabrication of large-scale models is costly, requires

specialized experimental equipment, and imposes high demands on the width and depth of the test water area. During the construction of scaled models, material losses inevitably affect acoustic similarity. To address the limitations of traditional proportional scaling schemes in meeting rigid-body overload similarity requirements, Liu *et al.* (2020) investigated non-proportional scaling penetration similarity laws and proposed a corresponding non-proportional scaling test design method. Wang *et al.* (2021) focused on material properties and introduced dimensionless numbers to characterize the similarity of strain hardening, strain-rate sensitivity, and temperature softening effects, based on similarity analysis of the thermal-viscosity constitutive equation. Wang *et al.* (2023) examined the scaling characteristics of geometrically distorted models using similarity criteria and design methods, analyzing various combinations of scaled models and providing guidance on the design of scaled models. Zhang *et al.* (2001) applied similarity theory to analyze key dimensionless quantities in the cylindrical shell sound field formula and employed material compensation to maintain similarity between the compensated sound field and the prototype sound field. By effectively correcting the data, the accuracy of scaled-model results can be enhanced. Huijgens *et al.* (2021) accurately simulated the interaction between the propeller and the engine by correcting viscous scale effects and torsional axis dynamics through testing.

From the early focus on performance control of large-scale acoustic models, to the establishment of scaling laws with non-geometric similarity for mechanical processes, and further to in-depth similarity analysis and compensation of material constitutive properties and geometric characteristics, research has progressively advanced in scaled-model studies. The correction methods for scaled models outlined above primarily focus on the models themselves. These methods involve utilizing large-scale models, adjusting the physical properties of materials, and correcting influencing parameters, with the aim of minimizing the potential impact of geometric scaling on scaled-model experiments. However, the effects of non-proportional scaling on target strength, as well as methods for correcting target strength in non-proportional scaled models, remain areas requiring further investigation. The concept of non-proportional scaling is illustrated in Fig. 1. To address these challenges, this study proposes a correction approach based on highlight model theory (Tang *et al.*, 1994), integrated with sound-scattering similarity, to effectively correct the results of scaled-model experiments and enhance their accuracy and reliability. The findings demonstrate that this correction method is not only theoretically sound but also holds significant practical application value in real-world experiments.

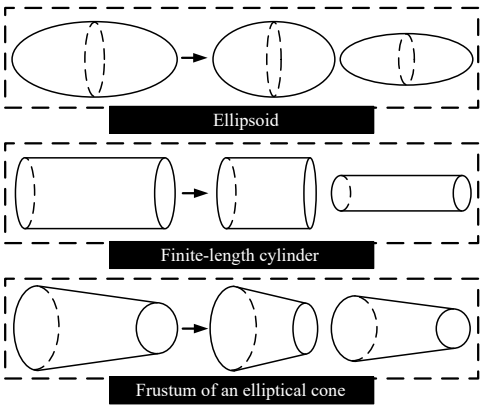


Fig. 1. Diagram of the non-proportional scaling of the base model.

The work presented in this paper is structured as follows. First, correction methods for the target strength of scaled models under various operating conditions (ellipsoid, finite-length cylinder, and frustum) are derived using highlight model theory. Second, the plate element method (Fan *et al.*, 2012) is employed to correct the target strength of the scaled models under different operating conditions, with the corrected results compared to the target strength of the prototype to assess the feasibility of the correction method. Finally, a complex computational model is developed to further correct the target strength of the scaled model. By comparing the corrected results with those of the prototype, the applicability and effectiveness of the proposed correction method are validated.

2. Geometric highlight model of simple targets

The primary theoretical modeling methods for underwater target echo characteristics can be categorized into two approaches: the 'decomposition-synthesis' method (also known as the component method) and the numerical integration method. In the component method, the echo of each individual component or sub-target is modeled using a highlight model derived from physical acoustics. This model, referred to as the 'highlight model,' is particularly suitable for engineering applications. This paper provides a comprehensive overview of the shape features of underwater vehicles and derives the geometric highlight parameters for several typical sub-targets, as illustrated in Fig. 2.

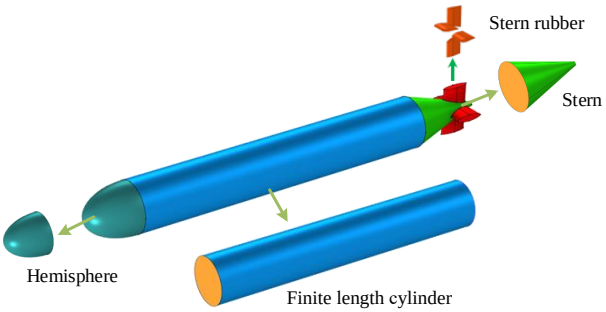


Fig. 2. 3D model of an underwater vehicle and its component decomposition.

The underwater vehicle is composed of simple geometric components, including an ellipsoid at the bow, a cylindrical section at the mid-body, and an elliptical-cone frustum at the stern, along with the sail housing, elevators, and stern rudders (Schneider *et al.*, 2003). In Fig. 2, when assembling the vehicle from components such as the hemisphere, finite-length cylinder, and sail housing, the spatial position relationships between the components are not taken into account. Instead, an incoherent superposition method is employed.

The method proposed in this paper is based on the 'highlight model' theory. The highlight model is an approximate method applicable under high-frequency conditions, and this study primarily focuses on the geometric components of underwater vehicles. These vehicles are mostly rigid structures, and the study mainly concerns structures made of structural steel.

During underwater scaled-model testing, factors such as manufacturing tolerances and transportation constraints may prevent the target strength of the scaled model from being directly converted to that of the prototype using scaling ratios. The method proposed in this paper can be applied to correct the target strength of the scaled model. The geometric highlight model for simple targets is primarily used to calculate unscaled structures. In practical applications, it can be utilized to correct the acoustic target strength of imperfectly scaled models, thereby establishing a link between the acoustic target strength of the scaled model and that of the prototype.

2.1. Ellipsoid

The ellipsoid is a fundamental geometric shape commonly used as an approximation in early echo prediction models for underwater vehicles. It is also one of the first types of sub-targets to be studied. When the major and minor axes are equal, the ellipsoid degenerates into a sphere. This shape can effectively approximate the head of an underwater vehicle.

The coordinate system is shown in Fig. 3, where the ellipsoid's semi-major axis is denoted as a and the semi-minor axis as b. When a sound wave is incident at an angle of φ, the highlight transfer function of the rotating ellipsoid can be derived using physical acoustics methods and the stationary-phase integration:

$$H(\mathbf{r}\omega) = \frac{1}{2}\frac{a^2\sin^2 v_0 + b^2\cos^2 v_0}{a} e^{2i\omega\sqrt{(b\sin v_0)^2+(a\cos v_0)^2}\cos(\varphi-a)/c}, \tag{1}$$

where $\tan v_0 = (b/a)\tan\varphi$, $\tan\alpha = (b/a)^2\tan\varphi$, and c is the sound speed in water, ω is the angular frequency. When the semi-major and semi-minor axes are equal, (i.e., $a = b$), the transfer function for a spherical target

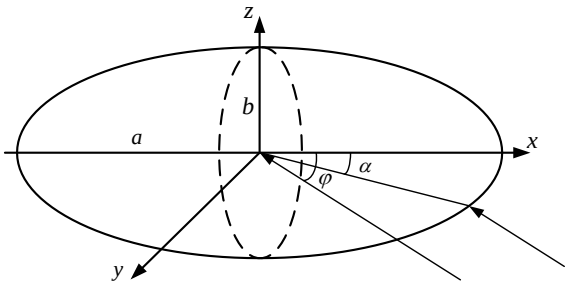


Fig. 3. Diagram for calculating the echo of an ellipsoid.

is obtained. In the transfer function, the phase factor represents the transmission path of the echo. Ignoring the phase factor, the target strength formula for the ellipsoid is expressed as

$$\mathrm{TS} = 20\log(H(\mathbf{r}\omega)) = 20\log\left(\frac{1}{2}\frac{a^2\sin^2 v_0 + b^2\cos^2 v_0}{a}\right). \tag{2}$$

2.2. Finite-length elliptical cylinder

The finite-length elliptical cylinder is a significant type of sub-target. The shapes of early submarine sail enclosures, tail fins, and diving planes closely resemble an elliptical cylinder. When the major and minor axes are equal, the elliptical cylinder degenerates into a cylindrical shape. This model can be used to approximate both the pressure hull and non-pressure hull sections of a submarine. A finite-length cylindrical shape is illustrated in Fig. 4. In this configuration, the long semi-axis (a) of the ellipse aligns with the x-axis, the short semi-axis (b) is along the y-axis, and the side length (h) is oriented along the z-axis. This discussion focuses on the case of vertical incidence with respect to the z-axis, where the sound wave propagates in the xy-plane. Under these conditions, only the echo from the side needs to be considered. The target strength of the finite-length elliptical cylinder is given by

$$\mathrm{TS} = 10\log\left(\frac{1}{\lambda^2}|I|^2\right). \tag{3}$$

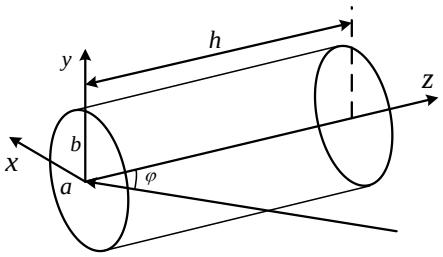


Fig. 4. Diagram for calculating the echo of a finite-length elliptical cylinder.

The phase factor represents the propagation path of the echo. After neglecting the phase factor, the expression for I becomes:

$$I = \sqrt{\frac{\pi}{k}}\frac{hab}{(a^2\cos^2\varphi + b^2\sin^2\varphi)^{3/4}}, \tag{4}$$

where k denotes the wave number ($k = \pi/\lambda$), h represents the height of the finite-length elliptical cylinder, a and b correspond to the long and short semi-axes of the elliptical cylinder, respectively, and λ is the wavelength.

2.3. Frustum of an elliptical cone

A frustum of an elliptical cone is a significant type of sub-target, from which a frustum of a circular cone can be derived. Both the front and rear edges of the frustum have a specific slope, and it can be utilized to approximate control surfaces such as the fuselage rudder, horizontal tailplane, and vertical tailplane, either individually or in combination with other basic geometric shapes. The frustum of an elliptical cone is illustrated in Fig. 5.

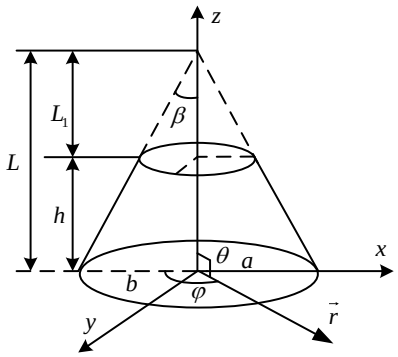


Fig. 5. Diagram for calculating the echo of a frustum of an elliptical cone.

The key parameters of the elliptical cone frustum include the semi-major axis (a) and semi-minor axis (b) of the upper base, the height (L) of the upper section, the height of the lower base (L_1) and the half-apex angle (β). The highlight on the frustum of the elliptical cone can also be equivalently regarded as the edge points of the upper and lower edges. By applying methods from physical acoustics and stationary-phase analysis, the following expression for the highlight transfer function can be derived:

$$H(\mathbf{r}\omega) = H_1(\mathbf{r}\omega) + H_2(\mathbf{r}\omega), \tag{5}$$

$$H_1(\mathbf{r}\omega) = A_1 e^{i\omega\tau_1} e^{i\varphi_1}, \tag{6}$$

$$H_2(\mathbf{r}\omega) = A_2 e^{i\omega\tau_2} e^{i\varphi_2}, \tag{7}$$

$$A_1 = \frac{1}{4}\sqrt{\frac{L}{\pi k}} \frac{\eta^{3/2}\tan^{1/2}\beta}{\sin^{1/2}\theta(\sin^2\varphi+\eta^2\cos^2\varphi)^{1/4}} \frac{\sin\theta(\sin^2\varphi+\eta^2\cos^2\varphi)^{-1/2} - \frac{1}{\eta}\tan\beta\cos\theta}{\sin\theta\tan\beta(\sin^2\varphi+\eta^2\cos^2\varphi)^{1/2}+\eta\cos\theta}, \tag{8}$$

$$A_2 = \frac{1}{4}\sqrt{\frac{L_1}{\pi k}} \frac{\eta^{3/2}\tan^{1/2}\beta}{\sin^{1/2}\theta(\sin^2\varphi+\eta^2\cos^2\varphi)^{1/4}} \frac{\sin\theta(\sin^2\varphi+\eta^2\cos^2\varphi)^{-1/2} - \frac{1}{\eta}\tan\beta\cos\theta}{\sin\theta\tan\beta(\sin^2\varphi+\eta^2\cos^2\varphi)^{1/2}+\eta\cos\theta}, \tag{9}$$

$$\tau_1 = 2L\left(\sin\theta\tan\beta\left(\cos^2\varphi+\frac{1}{\eta^2}\sin^2\varphi\right)^{1/2}+\cos\theta\right)\Big/c, \tag{10}$$

$$\tau_2 = 2L_1\left(\sin\theta\tan\beta\left(\cos^2\varphi+\frac{1}{\eta^2}\sin^2\varphi\right)^{1/2}+\cos\theta\right)\Big/c, \tag{11}$$

where $\varphi_1 = -\frac{5}{4}\pi$, $\varphi_2 = -\frac{1}{4}\pi$, $\eta = \frac{a}{b}$.

The phase factor represents the propagation path of the echo. By neglecting the phase factor, the target strength of the frustum of an elliptical cone is given by

$$\mathrm{TS} = 20\log(H(\vec{r}\omega)) = 20\log(A_1 + A_2). \tag{12}$$

3. Target strength correction for scale models

This section focuses on ellipsoids, finite-length elliptical cylinders, and frustums of elliptical cones as the primary research objects. By combining acoustic similarity theory with the geometric scattering model theory, a correction method for the target strength of non-geometrically scaled models is developed. The proposed correction method is subsequently validated using the plate element method.

3.1. Correction of target strength for the ellipsoid

The target strength formula for the ellipsoid, derived in Subsec. 2.1, demonstrates that the target strength depends on the ellipsoid's geometric dimensions and the incident angle. This section examines the variations in ellipsoid target strength under different incident angles for non-proportional scaling:

1) incidence angle 0°:
when the incident angle $\varphi = 0°$, substituting into the equation $\tan v_0 = (b/a)\tan\varphi$, we obtain $v_0 = k\pi$. At this point, the target strength of the ellipsoid is

$$\mathrm{TS} = 10\log(H(\mathbf{r}\omega)) = 20\log\left(\frac{b^2}{2a}\right). \tag{13}$$

The major semi-axes of the ellipsoid, a and b, are scaled by the scaling factors N_x and N_y, respectively, such that $a = N_x a'$ and $b = N_y b'$, where the variables in the scaled model are denoted with a prime ($'$). The difference in target strength between the scaled model and the prototype is given by

$$\Delta\mathrm{TS} = \mathrm{TS} - \mathrm{TS}' = 20\log(N_y^2/N_x); \tag{14}$$

2) incidence angle 90°:
when the incident angle $\varphi = 90°$, substituting into the equation $\tan v_0 = (b/a)\tan\varphi$, we obtain $v_0 = \pi/2 + k\pi$. At this point, the target strength of the ellipsoid is

$$\mathrm{TS} = 10\log(H(\mathbf{r}\omega)) = 20\log\left(\frac{a}{2}\right). \tag{15}$$

The major semi-axes of the ellipsoid, a and b, are scaled by the scaling factors N_x and N_y, respectively, i.e., $a = N_x a'$ and $b = N_y b'$, where the variables in the scaled model are denoted with a prime ($'$). The difference in target strength between the scaled model and the prototype is

$$\Delta\mathrm{TS} = \mathrm{TS} - \mathrm{TS}' = 20\log(N_x). \tag{16}$$

For an ellipsoid with a major axis of 38 m and a minor axis of 2 m, the major axis is scaled down by a factor of 2, while the minor axis remains unchanged. The target strength is calculated for sound wave incident angles of 0° and 90°, considering both reception and transmission alignment. The target strength is then corrected using the derived method, and the corrected results are presented in Fig. 6.

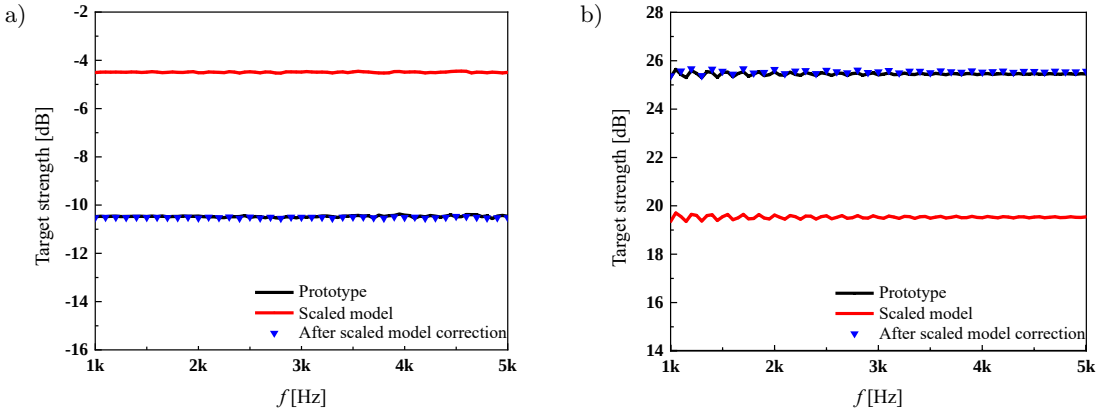


Fig. 6. Target strength of the ellipsoid with a non-isotropic scaling model after correction: a) incident angle of 0°, b) incident angle of 90°.

As shown in Fig. 6, with the reduction of the ellipsoid's major axis, the target strength correction for the scaled model at an incident angle of 0° is −6.02 dB. The corrected result is presented in Fig. 6a, and the average

absolute error between the corrected scaled model's target strength and the prototype's target strength is 0.05 dB. At an incident angle of 90°, the target strength correction for the scaled model is 6.02 dB. The correction result is shown in Fig. 6b, and the average absolute error between the corrected scaled model's target strength and the prototype's target strength is 0.10 dB.

3.2. Correction of target strength for a finite-length elliptical cylinder

According to the target strength equation for the finite-length elliptical cylinder derived in Subsec. 2.2, the target strength of the finite-length cylinder depends on the its geometric dimensions, incident angle, and incident frequency. This section examines the variation in target strength after non-isotropic scaling under different incident angles.

The variables a, b, and h are scaled by scale factors N_x, N_y, and N_z, respectively, i.e., $a = N_x a'$, $b = N_y b'$, $h = N_z h'$, where the relevant variables of the scaled model are denoted with primes ($'$). The difference in target strength between the scaled model and the prototype is expressed as

$$\Delta \mathrm{TS} = \mathrm{TS} - \mathrm{TS}' = 20\log(I\lambda'/I'\lambda). \tag{17}$$

From the above equation, it can be seen that the variation in target strength of the non-isotropically scaled model is related to the value of $I\lambda'/I'\lambda$. The he value of $I\lambda'/I'\lambda$ can be expressed as follows:

$$\frac{I\lambda'}{I'\lambda} = \left(\frac{f}{f'}\right)^{1/2} N_x N_y N_z \left(\frac{a'^2\cos^2\varphi + b'^2\sin^2\varphi}{(N_x a')^2\cos^2\varphi + (N_y b')^2\sin^2\varphi}\right)^{3/4}; \tag{18}$$

1) incidence angle is 0°:
when the incident angle is 0°, i.e., $\varphi = 0°$, the target strength formula can be simplified to:

$$\frac{I\lambda'}{I'\lambda} = \left(\frac{f}{f'}\right)^{1/2} \frac{N_y N_z}{\sqrt{N_x}}; \tag{19}$$

2) incidence angle is 90°:
when the incident angle is 90°, i.e., $\varphi = 90°$, the target strength formula can be simplified to:

$$\frac{I\lambda'}{I'\lambda} = \left(\frac{f}{f'}\right)^{1/2} \frac{N_x N_z}{\sqrt{N_y}}. \tag{20}$$

The target strength of an elliptical cylinder with a long semi-axis $a = 4$ m, a short semi-axis $b = 2$ m, and a height $h = 10$ m is calculated for frequencies ranging from 1 kHz to 10 kHz with a step size of 100 Hz. A scaled model 1 is constructed with a long semi-axis $a = 4$ m, a short semi-axis $b = 2$ m, and a height $h = 6$ m, and the target strength is calculated for the same frequency range and step size. A scaled model 2 is constructed with a long semi-axis $a = 2$ m, a short semi-axis $b = 1$ m, and a height $h = 10$ m, and the target strength is calculated again for the same frequency range and step size.

As shown in Fig. 7, as the height of the elliptical cylinder decreases, the corrected scaled-model target strength becomes 4.43 dB. The correction result is presented in Fig. 7a, where the corrected scaled model target strength closely matches the original target strength. As the semi-major and semi-minor axes of the elliptical base are scaled, the corrected scaled model target strength is 3.01 dB. The correction result is shown in Fig. 7b, where the corrected scaled model target strength closely matches the prototype target strength.

3.3. Correction of target strength for an elliptical frustum

According to the target strength formula for the truncated elliptical frustum derived in Subsec. 2.3, it is known that the target strength of the truncated elliptical frustum depends on the geometric dimensions, incident

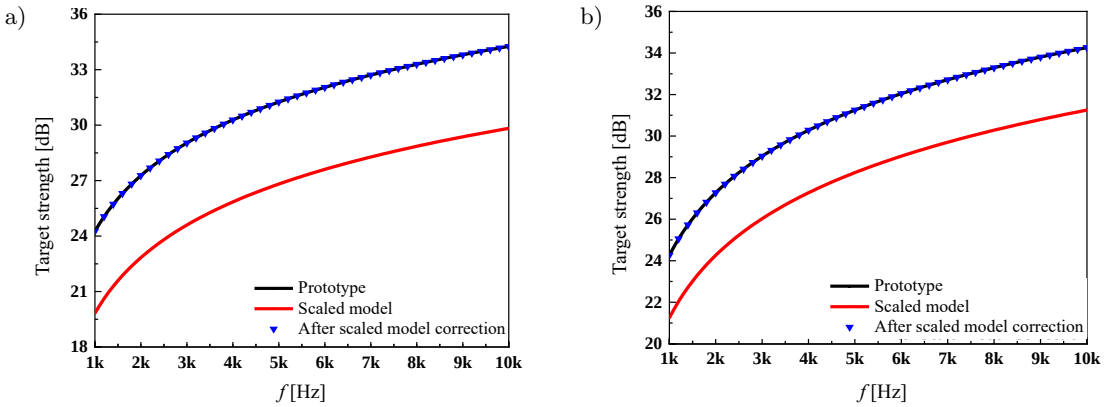


Fig. 7. Comparison of the original target strength and the corrected scaled-model target strength: a) model 1, b) model 2.

angle, and incident frequency. This section discusses the variation in the target strength of the truncated elliptical frustum after non-isotropic scaling, specifically for an incident angle of 90°.

When the incident angle is 90°, i.e., $\theta = 90°$, the target strength formula can be simplified as follows:

$$A_1 = \frac{1}{4}\sqrt{\frac{L}{\pi k}}\frac{\eta^{3/2}\tan^{1/2}\beta}{(\sin^2\varphi+\eta^2\cos^2\varphi)^{1/4}}\frac{(\sin^2\varphi+\eta^2\cos^2\varphi)^{-1/2}}{\tan\beta(\sin^2\varphi+\eta^2\cos^2\varphi)^{1/2}}, \tag{21}$$

$$A_2 = \frac{1}{4}\sqrt{\frac{L_1}{\pi k}}\frac{\eta^{3/2}\tan^{1/2}\beta}{(\sin^2\varphi+\eta^2\cos^2\varphi)^{1/4}}\frac{(\sin^2\varphi+\eta^2\cos^2\varphi)^{-1/2}}{\tan\beta(\sin^2\varphi+\eta^2\cos^2\varphi)^{1/2}}. \tag{22}$$

For the case of a circular base frustum ($\eta = 1$) and $\varphi = 0°$:

$$A_1 = \frac{1}{4}\sqrt{\frac{L}{\pi k}}\frac{1}{\tan^{1/2}\beta}, \qquad A_2 = \frac{1}{4}\sqrt{\frac{L_1}{\pi k}}\frac{1}{\tan^{1/2}\beta}.$$

From the aforementioned equations, it can be seen that the change in target strength of the non-geometrically scaled model is related to the half-apex angle β, and the value of β depends on the geometric dimensions of the truncated cone. Next, we discuss the relationship between the length scale ratio of the truncated cone and the half-apex angle β. The radius of the truncated cone base a and the height L are scaled by the scale factors N_x and N_y, respectively, i.e., $a = N_x a'$ and $L = N_y L'$, where the variables for the scaled model are denoted with a prime. The difference in target strength between the scaled model and the prototype is

$$\Delta\mathrm{TS} = \mathrm{TS} - \mathrm{TS}' = 20\log\left(\frac{A_1 + A_2}{A_1' + A_2'}\right), \tag{23}$$

$$\Delta\mathrm{TS} = 20\log\left(\frac{\tan^{1/2}\beta'}{\tan^{1/2}\beta}\frac{\sqrt{L}+\sqrt{L_1}}{\sqrt{L'}+\sqrt{L_1'}}\right). \tag{24}$$

As shown in Fig. 5, $\tan\beta = a/L$ and $\tan\beta' = a'/L'$. The relationship between the target strength difference and the scale ratio can thus be expressed as

$$\Delta\mathrm{TS} = 20\log\left(\frac{N_y}{\sqrt{N_x}}\right). \tag{25}$$

The target strength of a truncated cone with a lower base radius of 4 m, an upper base radius of 1 m, and a height of 6 m is calculated for frequencies ranging from 1 kHz to 10 kHz and a step size of 100 Hz. Scaled model 1 is constructed with a lower base radius of 2.4 m, an upper base radius of 0.6 m, and a height of 6 m, and its target strength is calculated over the same frequency range and step size. Scaled model 2 is constructed with a lower base radius of 4 m, an upper base radius of 1 m, and a height of 4 m, and its target strength is also calculated for the same frequency range and step size.

As shown in Fig. 8, with the scaling of the semi-major and semi-minor axes of the base ellipse, the correction value for the target strength of the scaled model is 2.21 dB, as illustrated in Fig. 8a. The corrected target strength of the scaled model closely matches that of the prototype. As the height of the elliptical cone is reduced, the correction value for the target strength of the scaled model increases to 3.01 dB, with the corrected result shown in Fig. 8b. The average absolute error between the corrected target strength of the scaled model and the prototype target strength is 0.24 dB.

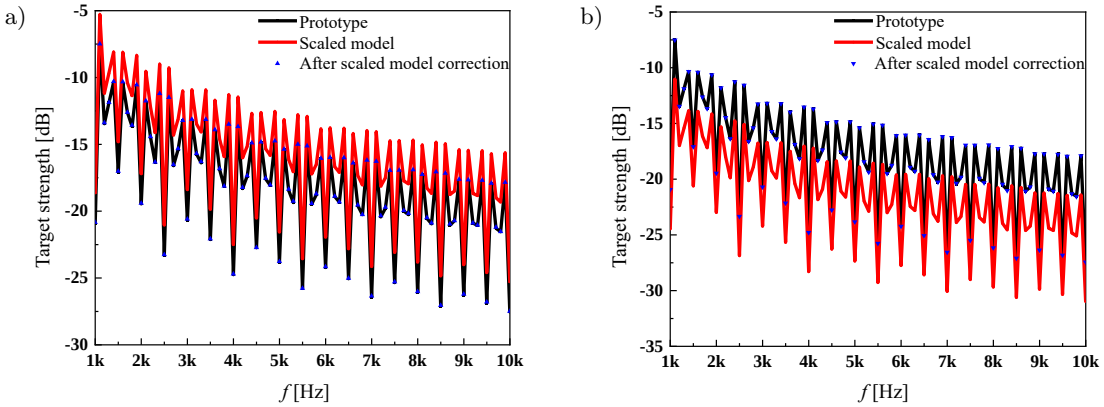


Fig. 8. Comparison of the original target strength and the corrected scaled-model target strength: a) model 1, b) model 2.

4. Target strength correction of the non-geometrically scaled model for the combined structure

A typical underwater vehicle is composed of simple geometric components such as ellipsoids, cylinders, and truncated elliptical frustum. Using the method of energy superposition, the target strength of the combined structure is given by

$$\mathrm{TS} = 10\log\left(10^{\left(\frac{\mathrm{TS}_1}{10}\right)} + 10^{\left(\frac{\mathrm{TS}_2}{10}\right)} + 10^{\left(\frac{\mathrm{TS}_3}{10}\right)}\right), \tag{26}$$

where TS represents the target strength of the combined structure, and TS_1 to TS_3 represent the target strengths of the individual geometric components. After the combined structure undergoes non-geometrical scaling, the target strengths of each component are denoted as TS', TS'_1 to TS'_3. At this point, each component has a corresponding correction factor $\Delta\mathrm{TS}_1$ to $\Delta\mathrm{TS}_3$, and the following equation holds:

$$\mathrm{TS} = 10\log\left(10^{\left(\frac{\mathrm{TS}'_1+\Delta\mathrm{TS}_1}{10}\right)} + 10^{\left(\frac{\mathrm{TS}'_2+\Delta\mathrm{TS}_2}{10}\right)} + 10^{\left(\frac{\mathrm{TS}'_3+\Delta\mathrm{TS}_3}{10}\right)}\right). \tag{27}$$

If the correction values for each component are the same ($\Delta\mathrm{TS}_1 = \Delta\mathrm{TS}_2 = \Delta\mathrm{TS}_3 = \Delta\mathrm{TS}$), the formula can be simplified to:

$$\mathrm{TS} = \mathrm{TS}' + \Delta\mathrm{TS}. \tag{28}$$

This study designs a rotational body model composed of an ellipsoid, a cylinder, and a conical frustum. In practical applications, the correction methods for different regions of a complex model are not identical, and the actual correction coefficients for each region can vary. In such cases, the scaling model correction coefficients must be solved separately for each region to complete the target strength correction. The method proposed in this paper uses the geometric highlight model formula, applying non-proportional scaling to the geometric- dimension-related variables in the formula to determine the difference between the target strength of the non-geometrically scaled model and that of the prototype. This enables rapid correction of the scaled model's target strength for practical engineering applications.

The model consists of a head ellipsoid, a central cylinder, and a tail concentric frustum. The interior of the complex model is filled with air, and the specific parameters of the model are shown in Fig. 9, where L_1 is the long radius of the model's shell ellipsoid (5 m), L_2 is the length of the model's shell cylinder (10 m), L_3 is the

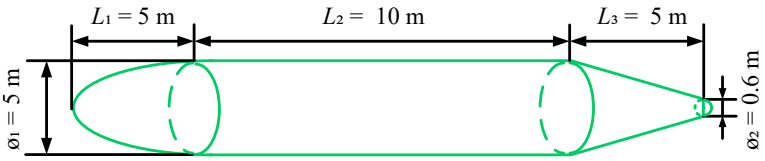


Fig. 9. Schematic diagram of the complex model dimensions.

height of the model's shell frustum (5 m), $\varnothing_1$ is the diameter of the model's shell cylinder (5 m), and $\varnothing_2$ is the diameter of the small end face of the model's shell frustum (0.6 m).

To calculate the target strength of the complex model, the calculation frequency ranges from 1 kHz to 10 kHz with a step size of 100 Hz. For length-scaled model 1 (Fig. 10a): the length of the complex model is scaled down by a factor of 0.75, while the radius of the rotational body remains unchanged. The calculation frequency is 1 kHz to 10 kHz with a step size of 100 Hz. For radius-scaled model 2 (Fig. 10b): the radius of the complex model is scaled down by a factor of 0.6, while the length of the rotational body remains unchanged. The calculation frequency is 1 kHz to 10 kHz with a step size of 100 Hz.

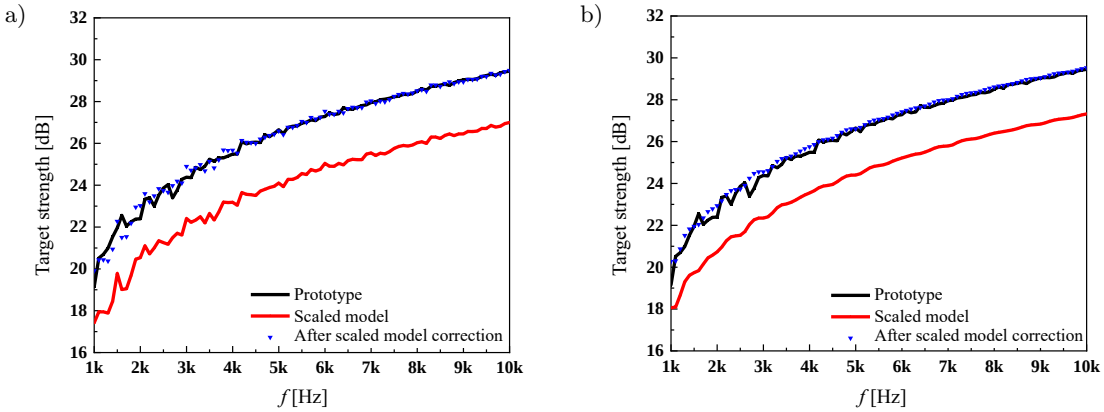


Fig. 10. Comparison of the original target strength and the corrected scaled model target strength: a) length-scaled model 1, b) radius-scaled model 2.

As shown in Fig. 10, as the height of the ellipsoid cylinder decreases, the target strength correction of the scaled model is 2.49 dB. The correction results are shown in Fig. 10a, where the average absolute error between the corrected scaled model's target strength and the prototype's target strength is 0.16 dB. As the semi-major and semi-minor axes of the bottom ellipse are scaled, the target strength correction of the scaled model is 2.21 dB. The correction results are shown in Fig. 10b, where the average absolute error between the corrected scaled model's target strength and the prototype's target strength is 0.18 dB.

5. Conclusions

This paper investigated simple targets (ellipsoids, finite-length cylinders, and frustums of elliptical cones) as well as rotating bodies. Based on the highlight model theory, the paper analyzed the impact of geometric size changes in scaled models on target strength and established a correction formula for non-proportionally scaled models. The feasibility of this target strength correction method was verified using plate element methods. The main conclusions are as follows:

1. This paper proposes a correction method suitable for the mid-to-high frequency range. The method, based on the highlight model theory, addresses geometric scales in the target strength solution formula by considering proportional changes in the geometric scales. It summarizes the target strength variation law and develops a corresponding correction method. However, the omission of low-frequency elastic resonance reduces the accuracy of the target strength correction. Additionally, geometric target strength is closely related to

the angle of incidence of the sound waves. This paper also discusses the target strength correction method for non-proportional scaled models under specific incident-angle conditions.

2. Non-proportional scaled models of simple geometric targets (ellipsoids, cylinders, and frustums of cones) can be corrected using the highlight model theory. When correcting the target strength of an ellipsoid non-proportional scaled model, the scale ratio of the semi-major and semi-minor axes must be considered. For a finite-length cylinder, the scale ratios of the cylinder's base radius and height are important. For a frustum of a cone, the scale ratios of the cone's height and base radius need to be accounted for. After correction, the target strength of these simple geometric models is generally consistent with that of the prototype.
3. The rotating body model designed in this paper consists of an ellipsoid, a cylinder, and a frustum of a cone. For the non-proportional scaled model of this combined body, the paper, based on the highlight model theory and using the energy superposition method, derives the target strength correction method. The complex model exhibits significant curvature changes at the junction of the cylinder and the frustum of the cone. When sound waves are incident at this junction, corner waves are generated. In this case, the correction method based on the energy superposition method introduces some errors.
4. The correction method proposed in this paper is designed for rapid target strength correction of scaled models under engineering-scale characteristic-angle conditions, utilizing the highlight model theory. In practical applications, the scale ratios of different parts of the same model can vary. As the incident angle changes, the model's target strength solution formula also changes. Therefore, different incident angles must be considered separately, as the target strength value varies with the incident angle, and the correction coefficient cannot be represented by a single numerical value. When correcting the target strength of complex models, it is essential to analyze the incident-reflection conditions of different components and account for the impact of various incident angles. Different components require different correction coefficients.

Currently, correction methods for scaled models are mainly applied to shell vibration and acoustic radiation similarity, while research on sound-scattering similarity correction methods is relatively limited. This paper proposes a rapid target strength correction method for scaled models in engineering applications, achieving certain correction effects. However, the accuracy and applicability of this method still require further improvement.

Fundings

This work was supported by the National Natural Science Foundation of China (grant no. 52201397) and the Open Fund of the National Key Laboratory of Ship Vibration and Noise (grant no. 6142204230709 and no. 6142204240202).

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.

Authors' contributions

Jin Zhang conceptualized the study and wrote the original draft. Zhenyu Li performed the analysis and contributed to data interpretation. Yinhao Li and Zilong Peng performed the analysis and wrote the original draft. All authors reviewed and approved the final manuscript.

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.ymssp.2022.109892.

2. Agounad S., Décultot D., Chati F., Leon F., Khandouch Y., Elhanaoui A. (2020), Acoustic scattering from cylindrical shell: Monostatic versus bistatic configurations [in French], *Forum Acusticum*, pp. 1863–1866, https://doi.org/10.48465/fa.2020.0610.
3. Fan J., Tang W.L., Zhou L.K. (2012), Planar elements method for forecasting the echo characteristics from sonar targets [in Chinese], *Journal of Ship Mechanics*, **16**(Z1): 171–180, https://doi.org/10.3969/j.issn.1007-7294.2012.01.020.
4. Feng M. *et al.* (2025), A TDOA sequence estimation method of underwater sound source based on hidden Markov model, *Applied Acoustics*, **227**: 110238, https://doi.org/10.1016/j.apacoust.2024.110238.
5. Gao T., Lin Y. (2019), Model analysis, design and experiment of a fan-wing underwater vehicle, *Ocean Engineering*, **187**: 106101, https://doi.org/10.1016/j.oceaneng.2019.06.006.
6. Ge J., Luo T., Qiu J. (2024), Discretized macro-element method to evaluate the crushing behavior and protective performance of crashworthy devices under ship collisions, *Ocean Engineering*, **306**: 118125, https://doi.org/10.1016/j.oceaneng.2024.118125.
7. Huijgens L., Vrijdag A., Hopman H. (2021), Hardware in the loop experiments with ship propulsion systems in the towing tank: Scale effects, corrections and demonstration, *Ocean Engineering*, **226**: 108789, https://doi.org/10.1016/j.oceaneng.2021.108789.
8. Ke H., Peng Z., Zhou F., Chen T. (2024), Prediction method and characteristics of static acoustic scattering from marine propellers [in Chinese], *Acta Armamentarii*, **45**(5): 1523–1533, https://doi.org/10.12382/bgxb.2023.0017.
9. Li C., Wang D., Liu J. (2020), Numerical analysis and experimental study on the scaled model of a container ship lashing bridge, *Ocean Engineering*, **201**: 107095, https://doi.org/10.1016/j.oceaneng.2020.107095.
10. Liu J., Peng Z., Fan J., Liu Y., Kong Z. (2023), Research on the acoustic scattering prediction methodology of underwater vehicle based on sliced-parameterized multi-highlights model [in Chinese], *Acta Armamentarii*, **44**(2): 517–525, https://doi.org/10.12382/bgxb.2021.0764.
11. Liu Y., Pi A., Yang H., Feng J., Huang F. (2020), Study on similarity law of non-proportionally scaled penetration/perforation test [in Chinese], *Explosion and Shock Waves*, **40**(03): 033302, https://doi.org/10.11883/bzycj-2019-0086.
12. Liu Z., Shi X., Guo T., Ren H., Zhang M. (2024), Investigation on hydrodynamic response and characteristics of a submerged floating tunnel based on hydrodynamic tests, *Ocean Engineering*, **300**: 117187, https://doi.org/10.1016/j.oceaneng.2024.117187.
13. Liu Z.X., Song Y., Yang Q.X. (2017), Research on foreign free maneuvering submarine models experiment [in Chinese], *Ship Science and Technology*, **39**(17): 194–199.
14. Schneider H.G. *et al.* (2003), Acoustic scattering by a submarine: Results from a benchmark target strength simulation workshop, [in:] *Tenth International Congress on Sound and Vibration.*
15. Stulenkov V.A. *et al.* (2024), Physical modeling of the hydroacoustic field of a marine propeller, *Acoustical Physics*, **70**(5): 872–879, https://doi.org/10.1134/s1063771024601912.
16. Tang W.L. (1994), Highlight model of echoes from sonar targets, *ACTA Acustica*, **19**(2): 92–100, https://doi.org/10.15949/j.cnki.0217-9776.1994.02.004.
17. Tang W.L., Fan J., Ma Z.C., Ji S. (2018), Acoustic scattering of underwater targets [in Chinese], *Beijing*: *Science Press*, https://doi.org/10.15949/j.cnki.0371-0025.2018.05.020.
18. Tao W., Liu Y., Long R., Cheng L. (2025), A time series CGAN denoising model targeting vessel's radiated noise for underwater acoustic communications, *Ocean Engineering*, **318**: 120076, https://doi.org/10.1016/j.oceaneng.2024.120076.
19. Ting Z., Yang T.C., Wen X. (2019), Bistatic localization of objects in very shallow water, *IEEE Access*, **7**: 180640–180651, https://doi.org/10.1109/ACCESS.2019.2947851.
20. Wang S., Xu F., Zhang X., Yang L., Liu X. (2021), Material similarity of scaled models, *International Journal of Impact Engineering*, **156**: 103951, https://doi.org/10.1016/j.ijimpeng.2021.103951.
21. Wang Y., Wang Q., Wang D. (2023), Scaling characteristics of the unified similarity criterion for geometrically distorted scale models, [in:] *ASME 2023 42nd International Conference on Ocean, Offshore and Arctic Engineering*, https://doi.org/10.1115/OMAE2023-101312.

22. Wang Y., Wei P., Wang Q., Dai Z., Wang D. (2024), A new similarity method for the ultimate strength of box girders subjected to the combined load of bending and lateral pressure, *Ocean Engineering*, **313**(Part 1): 119270, https://doi.org/10.1016/j.oceaneng.2024.119270.

23. Xue Y., Peng Z., 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, https://doi.org/10.12382/bgxb.2022.0374.

24. Yan T., Wang K. (2025), Application and control mechanism of thin film metamaterials in underwater sound-absorbing materials at low frequency under high hydrostatic pressure, *Applied Acoustics*, **231**: 110535, https://doi.org/10.1016/j.apacoust.2025.110535.

25. Yu M., Lu S. (2001), Investigation of acoustic scale effects for large size model of submarine compartment [in Chinese], *Journal of Ship Mechanics*, **5**(5): 50–55, https://doi.org/10.3969/j.issn.1007-7294.2001.05.007.

26. Yu M., Shi X., Chen K. (1999), Study on acoustic similitude of elastic stiffened cylindrical shells by FEM and BEM [in Chinese], *Shipbuilding of China*, **1999**(3): 65–71.

27. Yu M., Wang Z., Lu S. (2002a), Analysis of acoustic similarity of mechanical system for underwater vehicle by energy method [in Chinese], *Journal of Ship Mechanics*, **6**(4): 96–102, https://doi.org/10.3969/j.issn.1007-7294.2002.04.011.

28. Yu M., Wu Y., Lu S. (2002b), Experiment investigation on acoustic similarity of stiffened cylindrical shells [in Chinese], *Shipbuilding of China*, **43**(2): 52–59, https://doi.org/10.3969/j.issn.1000-4882.2002.02.008.

29. Zhang L., Yang D. (2001), Similarity of sound field of shell radiation with dissimilar item [in Chinese], *Journal of Harbin Engineering University*, **22**(3): 30–32, https://doi.org/10.3969/j.issn.1006-7043.2001.03.008.

30. Zhou L., Xiong C., Duan Y., Wu J., Sun Y., Wei Q. (2021), Design of a dynamic scaled model for marine propulsion shafting [in Chinese], *Journal of Ship Mechanics*, **25**(7): 973–980, https://doi.org/10.3969/j.issn.1007-7294.2021.07.015.

Archives of Acoustics Vol. 51, No. 1, pp. 197–209 (2026), https://doi.org/10.24423/archacoust.2026.4316

Technical Note

Cube-Shaped Sound-Insulating Enclosures: Experimental Tests and Calculation Models

Krzysztof Kosała

Department of Mechanics and Vibroacoustics, Faculty of Mechanical Engineering and Robotics, AGH University of Krakow
Kraków, Poland

e-mail: kosala@agh.edu.pl

Received September 5, 2025; revised November 10, 2025; accepted December 10, 2025; available online January 15, 2026; version of record February 19, 2026; published issue March 27, 2026.

The research described in this article concerns sound-insulating enclosures used for sound sources imitating a noisy machine or device. It is a continuation of experimental tests and modelling studies previously conducted on a prototype test stand, in which the enclosure walls measured 0.7 m × 0.7 m. The main aim of the current research was to estimate the acoustic efficiency of the enclosures through experimental testing on a new stand with walls measuring 0.55 m × 0.55 m, conducted under conditions similar to those found in an industrial facility. Tests conducted on five wall types of varying thicknesses, made of materials such as steel, aluminium, and plexiglass, enabled the development of a calculation model for insertion loss (IL), which can be used based on the material data for the enclosure walls. The model was validated through further experimental tests covering four additional material variants, and a high correlation of the results was obtained. The influence of the calculation model used for the enclosure wall transmission loss on the IL result was also investigated. The results of the experimental tests and modelling studies were also compared with those obtained for a larger enclosure made of the same wall materials. The research described in the current article may have practical applications in the selection of walls of cube-shaped enclosures and in estimating their effectiveness in a cost-free manner, assuming that appropriate material data are used in the calculations.

Keywords: acoustical enclosures, insertion loss (IL), noise protection.

1. Introduction

Homogeneous sound-insulating baffles have a number of applications in noise protection solutions. They can be used individually as shields or wall elements of acoustic barriers, or as walls of sound-insulating enclosures, usually designed to completely or partially isolate excessively noisy machines or devices (Barron, 2003; Vér, Beranek, 2006; Pawelczyk, Wrona, 2022; Chiu, 2012). The basic parameter determining the effectiveness of an acoustic enclosure is the insertion loss (IL). This parameter can be determined experimentally, based on the difference in sound power levels between the unenclosed and enclosed sound sources, or by using computational models, such as statistical energy analysis (SEA) (Nieradka, Dobrucki, 2018; Kim *et al.*, 2014; Lei *et al.*, 2012; Ming, Pan, 2004), or numerical methods (Zhou *et al.*, 2011; Agahi *et al.*, 1999).

Selecting the proper material for an enclosure wall has a significant impact on its effectiveness; however, this is often dependent on economic and functional factors. The sound-insulating properties of a single baffle can be determined in a laboratory, but this approach cannot take into account all the factors that influence its final resistance to sound transmission, such as the mounting method being identical to the intended one in a specific enclosure, external dimensions of the panel other than those resulting from the dimensions of the laboratory

measurement window, sealing method, etc. In order to estimate the sound insulation of a baffle with specific dimensions, calculation models (Kosała *et al.*, 2020a) can be used, with the accuracy dependent primarily on the material data used, which in reality may differ slightly from the commonly known and typical values given in the literature.

To test the effectiveness of material and construction solutions for baffles and enclosures, a prototype test stand was proposed within the framework of previous research, the main element being a heavy steel frame with the possibility of attaching baffles constituting the enclosure walls (Kosała *et al.*, 2020b). In addition to experimental tests of the effectiveness of an enclosed omnidirectional sound source imitating a noisy machine or device, a computational model has been developed to estimate the spectral characteristics of the IL of a sound-insulating enclosure without any cost (Kosała *et al.*, 2020c; Kosała, 2022; 2024). The experimental tests and modelling studies conducted so far concerned specific dimensions of the enclosure walls (0.7 m × 0.7 m), resulting from the developed test stand. As a continuation of this research, a smaller frame was developed, which enables the attachment of enclosure wall panels measuring 0.55 m × 0.55 m. The effectiveness of such cube-shaped sound-insulating enclosures, both experimental and model-based, is the subject of this article. A calculation model for the IL of enclosures made of five material sets of different thicknesses, such as steel, aluminium and plexiglass, was proposed and validated by further experimental studies on another four material sets. This model is a function of parameters related, among others, to the sound-insulating properties of the enclosure walls in question, such as the transmission loss (TL) and the single-number weighted sound insulation index of the wall R_w determined on its basis. These parameters were determined from the model proposed in previous studies, which is a combination of the models developed by Davy and Sharp, in specific frequency bands, respectively lower and higher, as described in (Kosała, 2019). The influence of TL, determined using the SoundFlow (Ahnert Feistel Media Group, 2011) and INSUL (Marshall Day Acoustics, 2017) programs, on the IL calculation results was also verified with respect to the experimental tests. The IL spectral characteristics obtained from the experimental tests of the enclosures were compared for four variants of wall materials, identical for the enclosure under study and the previously tested larger one with wall dimensions of 0.7 m × 0.7 m.

2. Measurement setup

The enclosure is based on a steel frame in the shape of a cube, allowing the installation of five walls measuring 0.55 m × 0.55 m with a thickness of 0.001 m to 0.09 m (Fig. 1). The frame is placed on a rubber plate, which ensures good tightness of the enclosure at the bottom. Each of the tested walls is tightened to the enclosure frame, through a rubber seal, using 0.01 m × 0.01 m square steel frames and a set of 11 holdfast mechanisms, except for the upper wall of the enclosure, which is attached using 12 mechanisms. The lower sides of the four walls mounted vertically to the frame are pressed against the frame using two holdfast mechanisms and a flat steel bar. This clamping solution allows the cubic shape of the enclosure cavity, measuring 0.55 m × 0.55 m × 0.55 m, to be maintained. Inside the enclosure, a loudspeaker is placed centrally on the floor (Fig. 1).

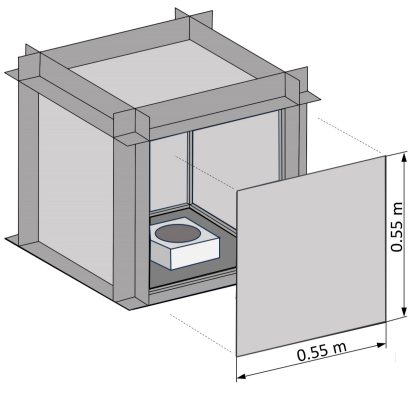


Fig. 1. Conceptual sketch of the enclosure in the form of a frame, loudspeaker, and five walls.

Figure 2 shows the enclosure frame with a loudspeaker (a) and the enclosure with steel walls (b). Acoustic tests were carried out in a room with a volume of 79 m^3. The signal, in the form of pink noise, generated using the Audacity program via an audio interface (EDIROL UA-5) and an amplifier (Bruel & Kjaer 2716-C), was fed to the sound source. A SVAN 945A sound level meter was used to measure sound pressure levels on a hemispherical measurement surface to determine the sound power levels of the source with and without the enclosure.

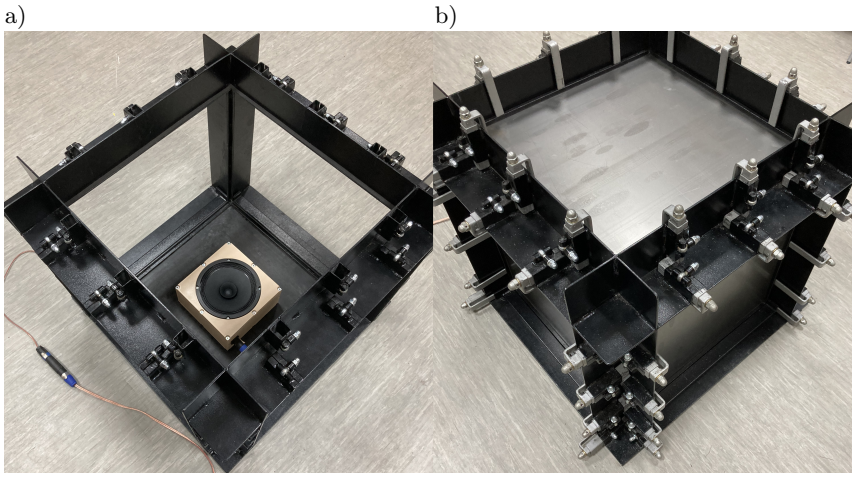


Fig. 2. View of the enclosure frame: a) without the walls, b) with steel walls installed.

For this purpose, the survey method was used in accordance with the ISO standard 3746 (International Organization for Standarization [ISO], 2010), taking into account the background noise correction factor K1 and the environmental correction factor K2. The surface sound pressure level was averaged from the results obtained from measurements carried out at four measurement points, located in accordance with the standard on a hemispherical measurement surface with a radius of 1.15 m.

The room acoustic absorption required for calculating the K2 correction factor was determined based on the reverberation time obtained from impulse responses recorded in the room. For this purpose, an omnidirectional sound source was used, consisting of six loudspeakers (BG 20/70W VISATON) assembled in a cubic casting, an amplifier (Bruel & Kjaer 2716-C), a BEHRINGHER ECM 8000 measuring microphone, and an EDIROL UA-5 audio interface, along with a DIRAC 6.0 Bruel & Kjaer system. The maximum length sequence (MLS) signal generated by the DIRAC program was used for the room excitation.

The performance of the acoustical enclosure in terms of IL was determined as the difference between the sound power levels radiated by the unenclosed and the enclosed source.

The IL of an acoustic enclosure can be determined more accurately under laboratory conditions in an anechoic room by determining sound power levels using the precision method (ISO, 2012), which was the subject of previous research (Kosała *et al.*, 2020c). In the current research, the use of the survey method was intended to reflect measurement conditions in which the nature of the sound field is similar to that occurring under industrial conditions, and it is considered a sufficient approach to validate the results obtained from the calculation model for the IL of sound-insulating enclosures proposed in this article.

An enclosure with a set of five identical walls was tested for different baffle types, with the material data shown in Table 1.

Table 1. Material data of the tested enclosure walls.

Parameter	Aluminium	Plexiglass	Steel
Density (p) [kg/m^3]	2800	1190	7850
Thickness (h) [m]	0.002 0.003	0.005 0.015	0.001
Young's modulus (E) [GPa]	70	3.5	207
Poisson's Ratio (ν) [–]	0.35	0.35	0.3
Loss factor (η) [–]	0.01	0.02	0.01

Assuming that the enclosure has perfectly rigid walls, the mode frequencies (eigenfrequencies) $f_{m,n,p}$ of the enclosure cavity can be determined according to the well-known formula (Hopkins, 2007):

$$f_{m,n,p} = \frac{c_0}{2}\sqrt{\left(\frac{m}{L_x}\right)^2 + \left(\frac{n}{L_y}\right)^2 + \left(\frac{p}{L_z}\right)^2}, \tag{1}$$

where c_0 is the speed of sound in the air [m/s], m, n, and p are the particular mode numbers, and L_x, L_y, and L_z are dimensions of the enclosure cavity [m].

The eigenfrequencies calculated in this way for a simple, rectangular-shaped cavity are an approximation. In reality, the air cavity inside the enclosure is reduced by the presence of a loudspeaker, which has a case with dimensions of 0.24 m × 0.24 m × 0.115 m. Calculations of the eigenfrequencies for such a complex geometry of the air cavity inside the enclosure were possible using a numerical model and simulations performed in COMSOL Multiphysics 6.3.

Table 2 shows the values of the eigenfrequencies calculated for the enclosure cavity with and without a loudspeaker.

Table 2. Eigenfrequencies of the enclosure cavity with and without a loudspeaker.

Mode	Eigenfrequency [Hz]		
	Enclosure cavity without a loudspeaker		Enclosure cavity with a loudspeaker
	Equation (1)	COMSOL Multiphysics	COMSOL Multiphysics
1	312.1	312.0	297.3
2	312.1	312.0	297.3
3	312.1	312.0	319.5
4	441.4	441.3	427.9
5	441.4	441.3	427.9
6	441.4	441.3	436.1
7	540.6	540.4	534.3
8	624.2	624.1	594.6
9	624.2	624.1	601.2
10	624.2	624.1	643.8

The identical dimensions of the enclosure cavity without a loudspeaker ($L_x = L_y = L_z$) show that each of the axial modes (#1–3 and #8–10) as well as each of the tangential modes (#4–6) have the same values. For an enclosure cavity reduced by the presence of a loudspeaker, the values of the eigenfrequencies are more varied.

3. IL calculation model

Previous studies have shown that, when determining the effectiveness of a sound-insulating enclosure, good results, i.e., small discrepancies between the calculation and experimental results. can be obtained using the model described in (Kosała, 2022). This model determines the effectiveness of a sound source enclosed by sound-insulating walls with dimensions of 0.7 m × 0.7 m. For cube-shaped enclosures, built of sound-insulating walls with dimensions 0.55 m × 0.55 m, a similar approach to that presented in (Kosała, 2022) is proposed, however, it is defined by the relationship:

$$\mathrm{IL} = \begin{cases} 10\log\left[\cos(kd) - 0.37\rho h\omega\sin(kd)/\rho_0 c_0\right]^2 & \text{for} \quad f \le 1.9 f_{0,0,1}, \\ 10\log\dfrac{\bar{\alpha}_{\mathrm{rand}}}{10^{-0.1\mathrm{TL}} + e^{-0.31 R_w}} & \text{for} \quad f > 1.9 f_{0,0,1}, \end{cases} \tag{2}$$

where k is the sound wavenumber, d is the distance of the top panel of the tested enclosure from the floor, ρ is the material density, h is the material thickness, ω is the angular frequency, ρ_0 is the air density, c_0 is the speed of sound in the air, $f_{0,0,1}$ is the first axial mode frequency of the enclosure cavity, $\bar{\alpha}_{\mathrm{rand}}$ is the random-incidence sound absorption coefficient of bare enclosure walls, TL is the sound transmission loss of the bare enclosure walls, R_w is the weighted sound reduction index of the enclosure wall, and f_c is the coincidence frequency of the enclosure wall (Kosała, 2022; Vér, Beranek, 2006).

A new calculation model is proposed for IL for the relevant frequency ranges, depending on the eigenfrequency $f_{0,0,1}$ enclosure cavity. For the low-frequency range, in which $f \leq 1.9 f_{0,0,1}$ of the enclosure cavity, a modified Oldham model (Oldham, Hillarby, 1991; Kosała, 2022) for a clamped boundary condition is proposed. For the high-frequency range, where $f > 1.9 f_{0,0,1}$ of the enclosure cavity, the proposed model is a function of $\bar{\alpha}_{\mathrm{rand}}$, TL, and R_w.

The quantity 0.32 in Eq. (16) from (Kosała, 2022) was replaced by 0.37 in Eq. (2) for the low-frequency range. This value was obtained based on the results of new experimental tests of the IL spectral characteristics, carried out for five variants of the wall materials given in Table 1. For the frequency range of 100 Hz to 500 Hz (centre frequencies of the ⅓-octave bands), using the assumed value of 0.37, a relatively low root mean square error (RMSE) value of 3.9 dB was obtained, averaged across five variants, with a simultaneously high value of the Pearson correlation coefficient $r = 0.88$ between the results of the model calculations and the experimental tests. Similarly, new experimental tests were used for the higher frequency range, from 630 Hz to 5 kHz. Using the assumption of a value of 0.31 for R_w resulted in a relatively low five-variant average of RMSE = 2.73 dB, with a high Pearson correlation coefficient of $r = 0.85$.

4. Results

4.1. Calculation of sound-insulating properties of the enclosure walls

To determine the TL of a homogeneous baffle with external dimensions of 0.55 m × 0.55 m, we needed to determine the IL of a sound-insulating enclosure using the proposed Eq. (1), the material data of the baffle, shown in Table 1, and the Davy–Sharp model proposed in (Kosała, 2019) were used. Previous studies have shown that such a combination of the two models, Davy (Davy, 2009) and Sharp (Sharp, 1973), for the relevant frequency ranges, resulted in smaller discrepancies in TL calculations for single homogeneous baffles with dimensions of 1 m × 2 m, compared to laboratory tests, than when using commercial software and when the models were used separately (Kosała, 2019). The calculation results, in the form of TL spectral characteristics in ⅓-octave frequency bands for five materials with thicknesses ranging from 0.001 m to 0.015 m, are shown in Fig. 3.

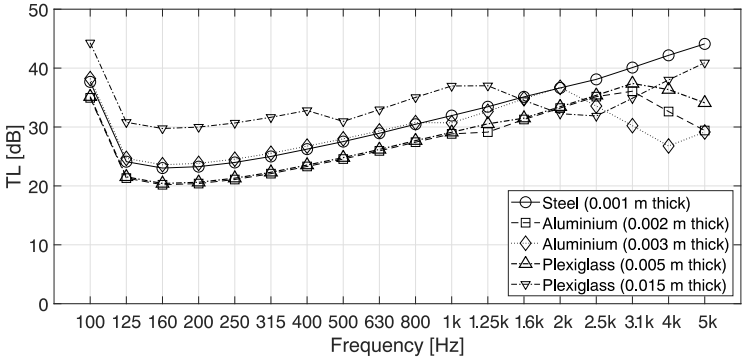


Fig. 3. TL of the tested enclosure walls obtained using the Davy–Sharp calculation model.

The curves for the aluminium baffle (0.003 m thick) and the plexiglass baffle (0.015 m thick) show a characteristic reduction in sound insulation in the higher frequency range, related to the occurrence of coincidence phenomena. For the remaining baffles, the coincidence frequency f_c is above the highest centre frequency of the ⅓-octave bands considered in this study, which is 5 kHz. The calculated f_c values for the tested enclosure walls are shown in Table 3.

In the low-frequency bands (Fig. 3), where the radiation efficiency of a finite-sized baffle is reduced, a specific increase in sound insulation can be observed. Based on the TL spectral characteristics, the weighted sound reduction index R_w was calculated for each type of enclosure wall tested. The values of this index, as well as the spectral adaptation indices C and C_{tr}, for the tested enclosure walls, are shown in Table 3. By comparing the R_w values (Table 3) for enclosure walls of dimensions 0.55 m × 0.55 m, made of steel (0.001 m thick) and

Table 3. Weighted sound reduction indices $R_w(C;C_{tr})$ and the coincidence frequencies f_c of the tested enclosure walls.

Enclosure wall material	$R_w(C;C_{tr})$ [dB]	f_c [Hz]
Steel (0.001 m thick)	32(−1;−3)	12 042
Aluminium (0.002 m thick)	29(−1;−3)	6072
Aluminium (0.003 m thick)	32(−1;−2)	4048
Plexiglass (0.005 m thick)	29(0;−2)	6843
Plexiglass (0.015 m thick)	34(0;0)	2281

aluminium (0.002 m thick) with the corresponding values of this parameter for baffles of dimensions 0.7 m × 0.7 m (Kosała, 2022), it can be seen that the smaller baffles have R_w values 1 dB higher than the larger baffles. This is influenced by the TL spectral characteristics, which in the low-frequency bands show slightly higher values for smaller baffles compared to larger ones. This also applies to baffles of larger dimensions (1 m × 2 m), for which the steel baffle (0.001 m thick) has R_w = 30 dB, and the aluminium baffle (0.002 m thick) has R_w = 27 dB (Kosała, 2019).

4.2. Calculation of IL *for the enclosure using the proposed model*

The calculation results obtained using the proposed IL Eq. (2) for the five tested variants of sound-insulating enclosures are shown in Fig. 4.

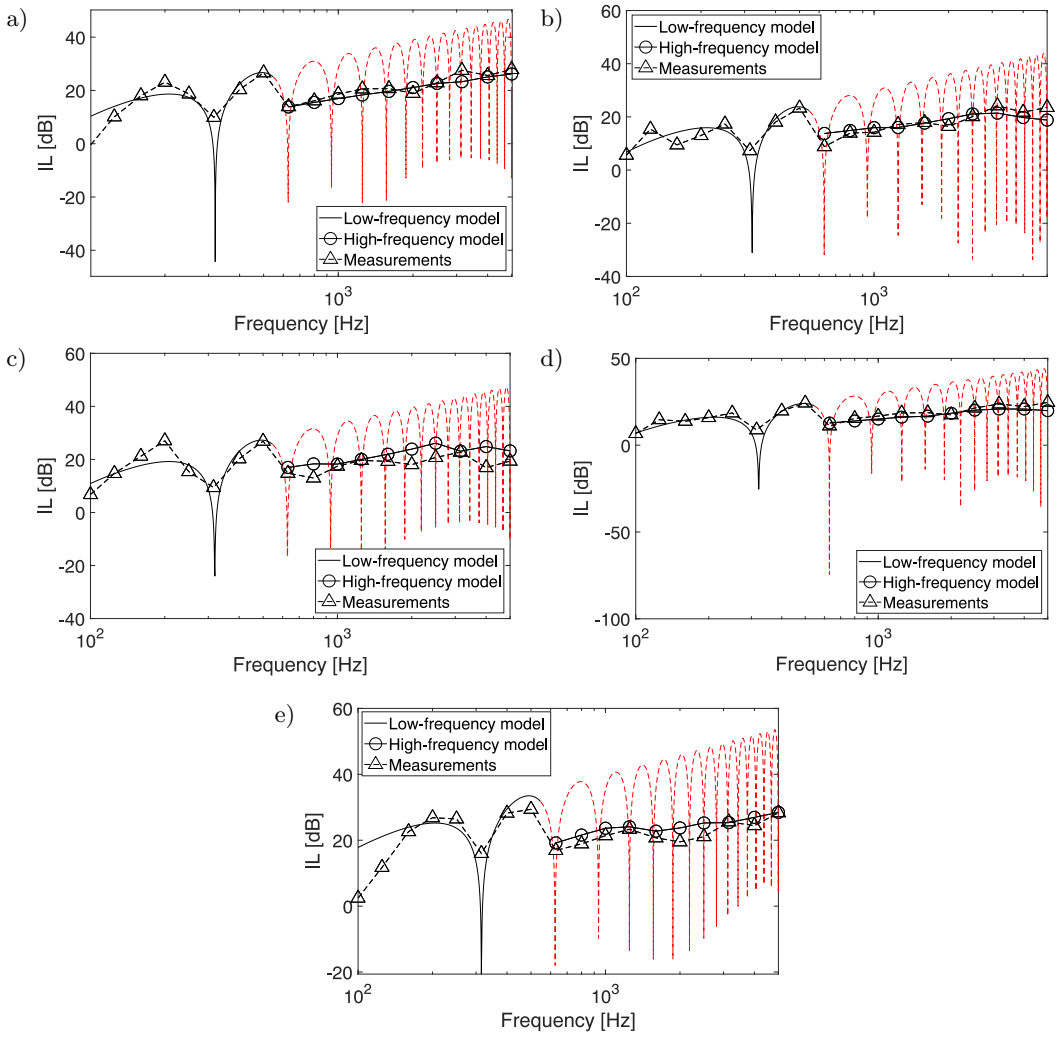


Fig. 4. IL of enclosures with walls made of different materials: a) steel (0.001 m thick), b) aluminium (0.002 m thick), c) aluminium (0.003 m thick), d) plexiglass (0.005 m thick), e) plexiglass (0.015 m thick), obtained using the low-frequency and high-frequency models, compared with experimental measurements.

After recalculating the IL values for the ⅓-octave bands in the low-frequency range, it was possible to compare the results of calculations and experimental tests using the Pearson correlation coefficient r and the RMSE. Close agreement between the calculated and measured results was obtained in both the lower and higher frequency ranges, as shown in Table 4. For the lower frequency range (centre frequencies of the ⅓-octave bands below 600 Hz), the Pearson correlation coefficients for all types of enclosure walls are high ($r \geq 0.8$). For the frequency range higher than 600 Hz, the values of $r \geq 0.86$, except for the 0.003 m thick aluminium baffle, for which $r = 0.61$. Considering the full frequency range, the Pearson coefficient remains high ($r \geq 0.73$).

Table 4. Pearson correlation coefficients r and RMSEs for predicted IL compared with experimental tests for the five enclosure wall materials.

Enclosure wall material			Steel	Aluminium	Aluminium	Plexiglass	Plexiglass
Thickness (h) [m]			0.001	0.002	0.003	0.005	0.015
r	Frequency range	low	0.88	0.87	0.85	0.95	0.80
		high	0.94	0.86	0.61	0.92	0.91
		full	0.9	0.87	0.73	0.94	0.81
RMSE [dB]	Frequency range	low	4.5	2.8	3.7	2.0	6.6
		high	1.9	2.7	4.2	2.3	2.5
		full	3.3	2.7	4.0	2.2	4.8

The RMSE values for the low-frequency range are from 2 dB (0.005 m thick plexiglass baffle) to 6.6 dB (0.015 m thick plexiglass wall). For the higher frequency range, the RMSE ranges from 1.9 dB (0.001 m thick steel baffle) to 4.2 dB (0.003 m thick aluminium baffle). Over the entire frequency range, the RMSE is between 2.2 dB and 4.8 dB.

In Fig. 4, the red dashed line indicates the further course of IL, after applying the modified Oldham formula (line 1 of Eq. (2)), but for $f > 1.9 f_{0,0,1}$. The nulls on the graphs correspond to the axial resonant frequencies of the enclosure cavity $f_{0,0,1}$, $f_{0,0,2}$, etc. Table 5 shows the eigenfrequencies of the enclosure cavity determined by this method and those calculated using Eq. (1).

Table 5. Eigenfrequencies of the enclosure cavity calculated from Eq. (1) and Eq. (2).

Axial mode number	m	n	p	Eigenfrequency [Hz] $f_{m,n,p}$	
				Equation (1)	Modified Oldham model (Eq. (2))
1	0	0	1	312	319
2	0	0	2	624	628
3	0	0	3	936	939
4	0	0	4	1248	1250
5	0	0	5	1560	1562
6	0	0	6	1873	1874
7	0	0	7	2185	2186
8	0	0	8	2497	2498
9	0	0	9	2809	2810
10	0	0	10	3121	3122
11	0	0	11	3433	3434
12	0	0	12	3745	3746
13	0	0	13	4057	4058
14	0	0	14	4369	4370
15	0	0	15	4681	4682

The eigenfrequency values of the enclosure cavity obtained by the two methods differ slightly, as shown in Fig. 5. The greatest differences are for the (0,0,1) mode; however, they decrease with increasing frequency.

It should be taken into account that the calculations using Eq. (1) are approximate, since the walls of the tested enclosure cannot be perfectly rigid when excited by a sound source.

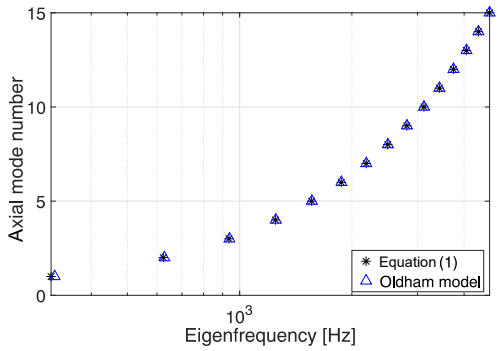


Fig. 5. Comparison of eigenfrequencies of the enclosure cavity obtained from Eq. (1) and Eq. (2).

5. Model validation and the influence of the wall's TL on the enclosure's IL

The proposed model for the sound insulation effectiveness of enclosures was tested in preliminary studies for five variants, each comprising five identical walls made of different materials. Based on the material data and the calculation of the enclosure wall's TL using the Davy–Sharp model, the IL of these enclosures was determined. In the next stage of the study, the model was validated using four additional sets of five walls made of the following materials: 1.5 mm and 2 mm thick steel, 10 mm thick plexiglass, and 1 mm thick aluminium. At this stage, the influence of other models, such as SoundFlow and INSUL, on the final results of the enclosure's IL calculations was examined. First, the TL spectral characteristics were determined and the corresponding R_w indices were calculated, as shown in Fig. 6.

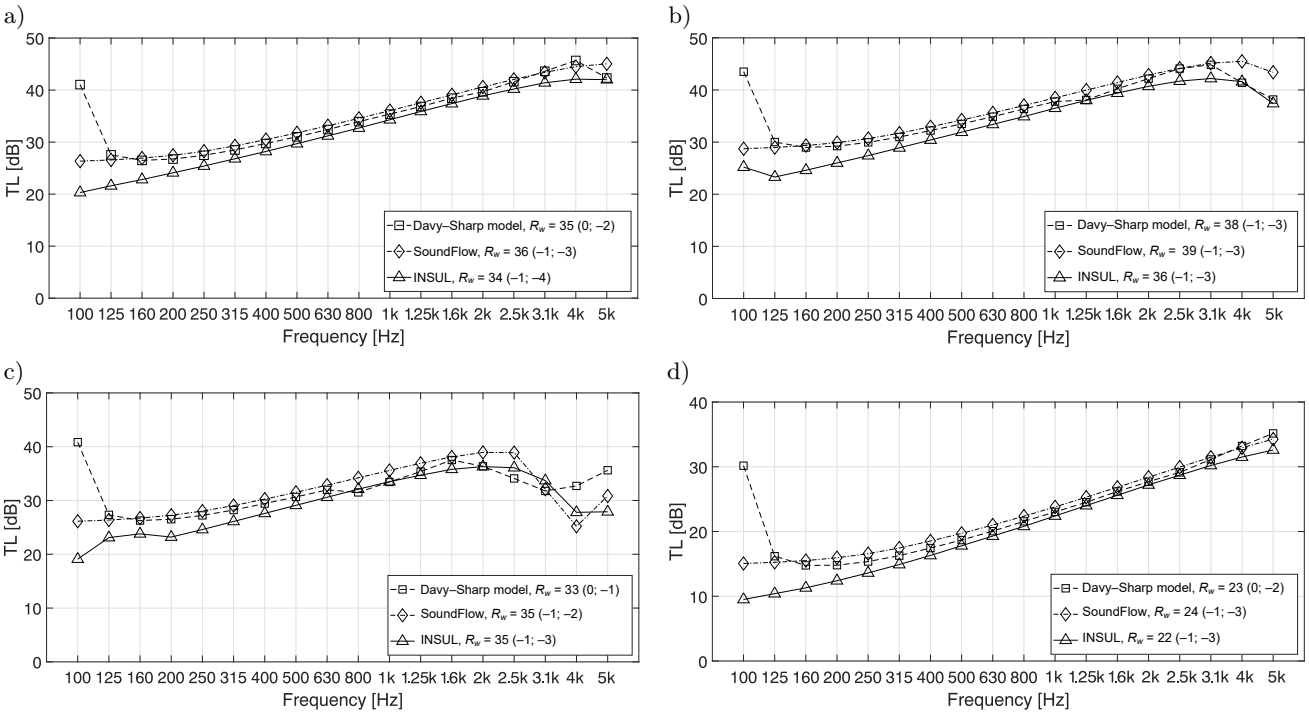


Fig. 6. TL calculated using the Davy–Sharp model, AFMG SoundFlow, and INSUL for enclosure walls made of: a) steel 0.0015 m thick, b) steel 0.002 m thick, c) plexiglass 0.01 m thick, and d) aluminium 0.001 m thick.

For the calculations in the AFMG SoundFlow and INSUL programs, the baffle dimensions were assumed to be 0.6 m × 0.6 m, because it was not possible to take into account the actual wall dimensions of 0.55 m × 0.55 m. Figure 6 shows discrepancies in the TL calculations, especially in the lower frequency bands. It should be noted that, although the IL calculations take into account the TL of the wall starting from the centre frequency of the ⅓-octave band at 630 Hz, the discrepancies are relatively small in the higher frequency range. The weighted

sound reduction index R_w, on which IL also depends, is calculated for the ⅓-octave bands from 100 Hz to 3.1 kHz. The discrepancies between the calculated R_w values range from 2 dB to 3 dB. In the higher frequency bands, a noticeable reduction in sound insulation due to coincidence is observed for a 0.01 m thick plexiglass baffle. Since the results are presented in ⅓-octave frequency bands, the coincidence frequency value f_c = 3421 Hz, cannot be precisely identified in the graphs. For the AFMG SoundFlow and INSUL models, f_c = 4 kHz, and for the Davy–Sharp model, f_c = 3.1 kHz.

The results of the enclosure IL calculations for the four tested new wall material sets, including the higher frequency ranges of the three models for which TL was calculated, are shown in Fig. 7.

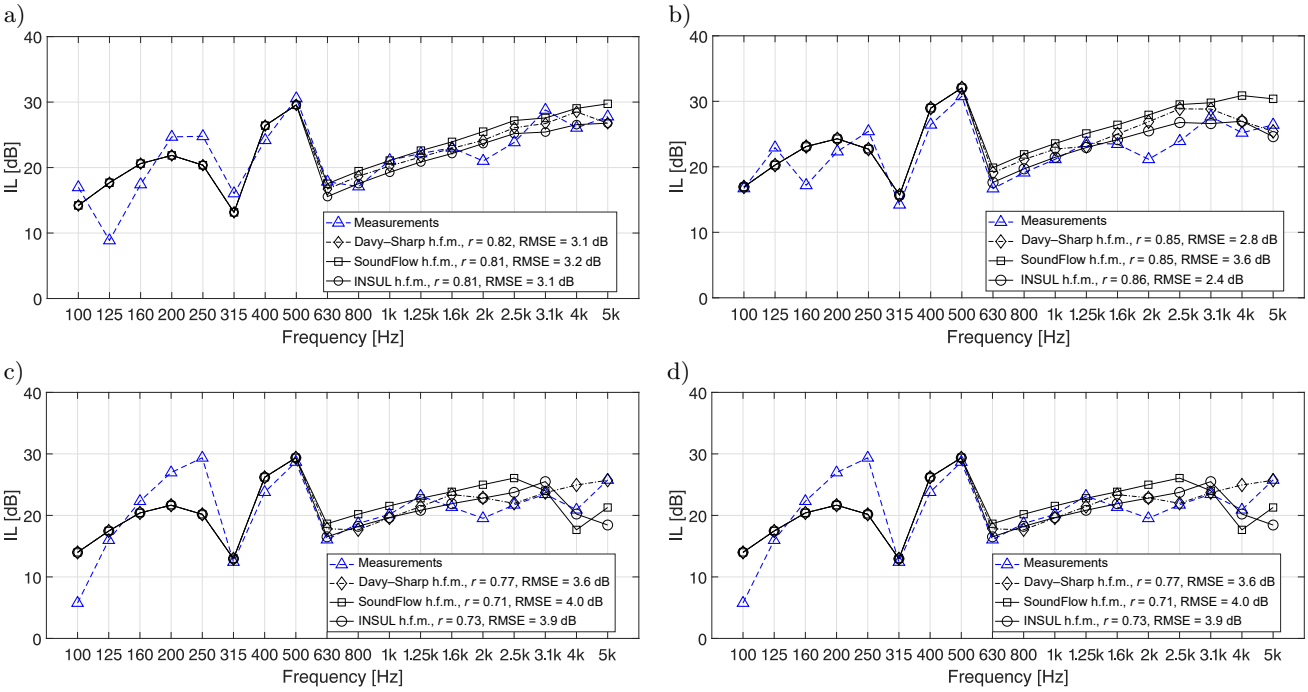


Fig. 7. IL calculated using the proposed calculation model: modified Oldham for the low-frequency range and Davy–Sharp, AFMG SoundFlow, and INSUL for the high-frequency range, for enclosure walls made of: a) steel 0.0015 m thick, b) steel 0.002 m thick, c) plexiglass 0.01 m thick, and d) aluminium 0.001 m thick.

The smallest RMSE discrepancies occurred for the 0.002 m thick steel baffle ($RMSE_{avr}$ = 2.9 dB), and the largest for the 0.001 m thick aluminium baffle ($RMSE_{avr}$ = 4.2 dB). The average Pearson correlation coefficients r for all four baffles calculated using the individual models are: 0.84 for Davy–Sharp, 0.8 for AFMG SoundFlow, and 0.8 for Insul. The average RMSEs are: 3.40 dB for Davy–Sharp, 3.73 dB for AFMG SoundFlow, and 3.43 dB for INSUL. These results indicate that the Davy–Sharp model showed both the highest correlation of IL characteristics with the experimental results among the three tested models and the lowest average RMSE.

6. Effectiveness of sound-insulating enclosures of different sizes made from the same wall materials

Experimental tests and modelling studies of the effectiveness of small enclosures (wall dimensions of 0.55 m × 0.55 m) described in this article, and large enclosures (wall dimensions of 0.7 m × 0.7 m) described in (Kosała, 2022), carried out on sets of identical wall materials, such as steel 0.001 m thick, aluminium 0.002 m thick, and plexiglas 0.005 m and 0.015 m thick, enabled a comparison of their sound-insulting properties. Figure 8 shows the results of the experimental tests, and Fig. 9 presents the results of the modelling studies.

The IL spectral characteristics (Fig. 8) show that IL is higher for smaller enclosures, especially in the mid- and high-frequency bands. For smaller enclosures in these ranges, greater variations in IL (5 dB to 10 dB) are observed for the tested materials, compared to larger enclosures (approximately 3 dB to 5 dB). In larger enclosures, the influence of the enclosure cavity eigenfrequencies on the IL spectral characteristic is observed up to 315 Hz,

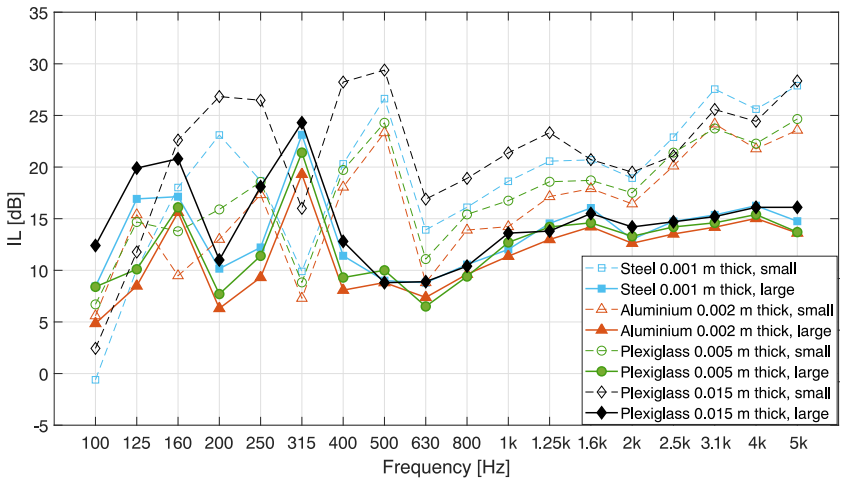


Fig. 8. IL obtained from experimental tests of small and large enclosures made of wall materials: steel (0.001 m thick), aluminium (0.002 m thick), and plexiglass (0.005 m and 0.015 m thick).

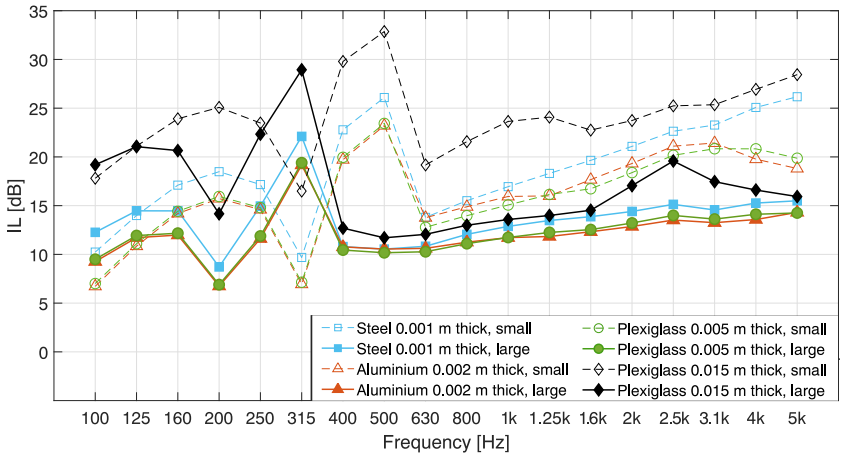


Fig. 9. IL obtained using calculation models for small and large enclosures made of wall materials: steel (0.001 m thick), aluminium (0.002 m thick), and plexiglass (0.005 mm and 0.015 m thick).

corresponding to 1.5 f_c of the enclosure wall, while for smaller enclosures this influence is more dominant, up to 500 Hz, corresponding to 1.9 f_c of the enclosure wall. Among the tested baffles, the most effective is the 0.015 m thick plexiglass wall; however, for smaller enclosures, the influence of coincidence, which is evident in the TL spectral characteristic of these walls, is more noticeable in the IL curves, whereas for larger enclosures, the IL characteristic in the 2 kHz region is flatter (f_c = 2281 Hz). The IL curves obtained from the calculation models (Fig. 9) closely match the corresponding curves obtained from the experimental tests (Fig. 8).

Table 6 summarizes the Pearson correlation coefficients r and the RMSEs for the IL calculation models compared to the experimental tests for enclosures with wall dimensions of 0.55 m × 0.55 m and 0.7 m × 0.7 m.

Table 6. Pearson correlation coefficients and RMSEs for predicted IL in comparison to experimental tests for the same enclosure wall materials, and enclosure wall dimensions of 0.7 m × 0.7 m and 0.55 m × 0.55 m.

Enclosure wall material	Thickness h [m]	Enclosure wall dimensions [m]			
		0.7 × 0.7		0.55 × 0.55	
		r	RMSE [dB]	r	RMSE [dB]
Steel	0.001	0.87	1.80	0.90	3.33
Aluminium	0.002	0.86	2.12	0.87	2.74
Plexiglass	0.005	0.89	1.75	0.94	2.17
Plexiglass	0.015	0.87	2.98	0.81	4.80
Average		0.87	2.16	0.88	3.26

The two models for smaller and larger enclosures, proposed in (Kosała, 2022), showed a high correlation coefficient between the IL spectral characteristics calculated using the models and those obtained from experimental tests. The average value of these coefficients is 0.9. The IL estimation using the calculation models showed approximately 1 dB lower average RMSE values for larger enclosures compared to smaller ones. In both cases, the average RMSE values are not high, and range from about 2 dB to 3 dB, except for the small enclosure made of 15 mm plexiglass, for which the prediction error is close to 5 dB.

7. Conclusions

As a result of experimental tests aimed at determining the effectiveness of an enclosed sound source imitating a noisy machine or device, IL characteristics were determined, first for five different wall material variants, and in the second stage of the study, for four additional variants.

In total, nine wall material variants were tested: steel with thicknesses of 0.001 m, 0.0015 m, and 0.002 m, aluminium with thicknesses of 0.001 m, 0.002 m, and 0.003 m, and plexiglass with thicknesses of 0.005 m, 0.01 m, and 0.015 m. The determined IL curves are very similar, and it can be roughly concluded that, in the medium- and high-frequency ranges, they are shifted towards higher IL values, corresponding to the single-number weighted sound insulation index R_w of the given wall. The poorest performance was observed in enclosures made of 0.001 m thick aluminium walls (R_w = 22 dB), while the best performance was achieved by enclosures made of 0.002 m thick steel walls (R_w = 38 dB). The IL curves, up to a certain centre frequency of the ⅓-octave bands, equal to 500 Hz, i.e., for $f \leq 1.9 f_{0,0,1}$, exhibit characteristic shapes, corresponding to the enclosure air cavity resonances. Above this frequency, the IL values are less differentiated across subsequent centre frequencies.

The proposed calculation model for the IL of sound-insulating enclosures was developed based on experimental tests conducted in the first stage, covering five variants of material walls. In a subsequent stage, the model was validated using four additional variants, taking into account the influence of the calculation model for the wall's TL on the final IL results. The studies showed that the model proposed in earlier studies, a combination of the Davy and Sharp models, yielded better agreement with experimental results compared to the models implemented in SoundFlow and INSUL software. The Davy–Sharp model showed both the highest correlation of IL characteristics with the experimental results of the three models tested and the lowest average RMSE. However, it should be noted that only the Davy–Sharp model could take into account the actual dimensions of the enclosure wall of 0.55 m × 0.55 m. Limitations in input data in the other software required the use of wall dimensions of 0.6 m × 0.6 m, i.e., only close to the actual ones, which could also have a certain impact on the accuracy of calculations using these models. A comparison of the effectiveness of enclosures made of the same materials, with walls measuring 0.55 m × 0.55 m, was made against larger enclosures with walls measuring 0.7 m × 0.7 m, analysed in previous studies. Based on the comparison of the IL spectral characteristics, it was found that IL is higher for smaller enclosures, especially in the mid- and high-frequency bands. Based on IL calculations using models, the average RMSE values were approximately 1 dB lower for larger enclosures compared to smaller ones. In both cases, the average RMSE values were not high, ranging from about 2 dB to 3 dB, except for the small enclosure made of 15 mm plexiglass, where the prediction error was close to 5 dB. Both the models for smaller and larger enclosures, proposed in (Kosała, 2022), showed high correlation coefficient between calculated and experimental IL spectral characteristics, with an average r = 0.9. The proposed calculation models can be helpful in cost-free estimation of the acoustic efficiency of cube-shaped sound-insulating enclosures with the studied wall dimensions, provided that appropriate material data needed for these calculations are available.

Testing the acoustic efficiency of sound-insulating enclosures is a significant challenge, as their effectiveness must be evaluated differently compared to that of sound-absorbing or insulating enclosures. In the latter, the impact of resonances within the enclosure, which sometimes counteracts their effectiveness, is mitigated by sound-absorbing material placed near the sound source. In contrast, sound-insulating enclosures rely on reflective walls, and their performance is strongly influenced by cavity resonances. Despite these challenges, sound-insulating enclosures have many practical applications. These are particularly relevant in cases where smooth, sound-reflecting transparent walls, such as plastic or glass, are required to monitor the operation of the enclosed device. Another

area of application involves situations where, due to the risk of bacterial growth in porous and fibrous materials, sound-absorbing and insulating enclosures are inappropriate.

The obtained research results may have practical applications for selecting wall materials for cube-shaped sound-insulating enclosures.

Fundings

This research was funded by a research subvention supported by the Polish Ministry of Education and Science, grant number 16.16.130.942.

Conflict of interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author's contribution

All author reviewed and approved the final manuscript.

Data availability

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

References

1. Agahi P., Singh U.P., Hetherington J.O. (1999), Numerical prediction of the insertion loss for small rectangular enclosures, *Noise Control Engineering Journal*, **47**(6): 201–208, https://doi.org/10.3397/1.599316.
2. Ahnert Feistel Media Group (2011), AFMG SoundFlow User's Guide.
3. Barron R.F. (2003), *Industrial Noise Control and Acoustics*, Marcel Dekker, New York.
4. Chiu M.-C. (2012), Shape optimization of two-layer acoustical hoods using an artificial immune method, *Archives of Acoustics*, **37**(2): 181–188, https://doi.org/10.2478/v10168-012-0024-5.
5. Davy J.L. (2009), Predicting the sound insulation of single leaf walls: Extension of Cremer's model, *The Journal of the Acoustical Society of America*, **126**(4): 1871–1877, https://doi.org/10.1121/1.3206582.
6. Hopkins C. (2007), *Sound Insulation*, Elsevier, Oxford.
7. International Organization for Standardization (2010), *Acoustics – Determination of sound power levels and sound energy levels of noise sources using sound pressure – Survey method using an enveloping measurement surface over a reflecting plane* (ISO Standard No. 3746:2010), https://www.iso.org/standard/52056.html.
8. International Organization for Standardization (2012), *Acoustics – Determination of sound power levels and sound energy levels of noise sources using sound pressure – Precision methods for anechoic rooms and hemi-anechoic rooms* (ISO Standard No. 3745:2012), https://www.iso.org/standard/45362.html.
9. Kim H.-S., Kim J.-S., Lee S.-H., Seo Y.-H. (2014), A simple formula for insertion loss prediction of large acoustical enclosures using statistical energy analysis method, *International Journal of Naval Architecture and Ocean Engineering*, **6**(4): 894–903, https://doi.org/10.2478/IJNAOE-2013-0220.
10. Kosała K. (2019), Calculation models for analysing the sound insulating properties of homogeneous single baffles used in vibroacoustic protection, *Applied Acoustics*, **146**: 108–117, https://doi.org/10.1016/j.apacoust.2018.11.012.
11. Kosała K. (2022), Experimental tests and prediction of insertion loss for cubical sound insulating enclosures with single homogeneous walls, *Applied Acoustics*, **197**: 108956, https://doi.org/10.1016/j.apacoust.2022.108956.
12. Kosała K. (2024), Sound insulation performance of cube-shaped enclosures, *Vibrations in Physical Systems*, **35**(2): 2024212, https://doi.org/10.21008/j.0860-6897.2024.2.12.

13. Kosała K., Majkut L., Mleczko D., Olszewski R., Wszołek T. (2020a), Accuracy of prediction methods for sound insulation of homogeneous single baffles, *Vibration in Physical Systems*, **31**(2): 2020210, https://doi.org/10.21008/j.0860-6897.2020.2.10.

14. Kosała K., Majkut L., Olszewski R. (2020b), Experimental study and prediction of insertion loss of acoustical enclosures, *Vibrations in Physical Systems*, **31**(2): 2020209, https://doi.org/10.21008/j.0860-6897.2020.2.09.

15. Kosała K., Majkut L., Olszewski R., Flach A. (2020c), Laboratory tests of the prototype stand to determine the acoustic properties of materials used in noise protection, [in:] *21st Century Technologies – Current Problems and New Challenges* [in Polish: *Technologie XXI Wieku – Aktualne Problemy i Nowe Wyzwania*], Vol. 1, pp. 7–20, Wydawnictwo Naukowe TYGIEL, Lublin.

16. Lei Y., Pan J., Sheng M.P. (2012), Investigation of structural response and noise reduction of an acoustical enclosure using SEA method, *Applied Acoustics*, **73**(4): 348–355, https://doi.org/10.1016/j.apacoust.2011.10.008.

17. Marshall Day Acoustics (2017), Sound Insulation Prediction Program, INSUL, User's Manual.

18. Ming R., Pan J. (2004), Insertion loss of an acoustic enclosure, *The Journal of the Acoustical Society of America*, **116**(6): 3453–34599, https://doi.org/10.1121/1.1819377.

19. Nieradka P., Dobrucki A. (2018), Insertion loss of enclosures with lined slits, [in:] *Proceedings of the 11th European Congress and Exposition on Noise Control Engineering Euronoise Conference*, pp. 893–898.

20. Oldham D.J., Hillarby S.N. (1991), The acoustical performance of small close fitting enclosure, part 1: Theoretical models, *The Journal of Sound and Vibrations*, **150**(2): 261–281, https://doi.org/10.1016/0022-460X(91)90620-Y.

21. Pawelczyk M., Wrona S. (2022), *Noise-Controlling Casings*, CRC Press, Boca Raton.

22. Sharp B.H. (1973), *A study of techniques to increase the sound insulation of building elements*, Technical Report, US Department of Commerce, National Technical Information Service, Washington.

23. Vér I.L., Beranek L.L. (2006), *Noise and Vibration Control Engineering – Principles and Applications*, 2nd ed., John Wiley & Sons, Hoboken, New Jersey.

24. Zhou L., Carter A.E., Herrin D.W., Shi J., Copley D.C. (2011), Airborne path attenuation of partial enclosures: Simulation and sensitivity study, *Applied Acoustics*, **72**(6): 380–386, https://doi.org/10.1016/j.apacoust.2010.12.012.

Archives of Acoustics Vol. 51, No. 1, pp. 211–222 (2026), https://doi.org/10.24423/archacoust.2026.4329

OSA 2025

Active Vibration Suppression of a Thin Circular Pipe

Marcin Pater*, Lucyna Leniowska, Marcin Grochowina

Department of Mechatronics and Automation, University of Rzeszów
Rzeszów, Poland

*Corresponding Author: mpater@ur.edu.pl

Received September 9, 2025; revised November 29, 2025; accepted December 16, 2025; available online January 19, 2026; version of record March 2, 2026; published issue March 27, 2026.

The article presents an active vibration damping system for a thin-walled cylindrical tube, which is a simplified model of a lightweight robot (LWR) arm. The proposed solution integrates control algorithms, piezoelectric materials and a hardware and software environment enabling real-time control. Macro fiber composite (MFC) elements were used for active vibration reduction, acting simultaneously as sensors and actuators. The object on which the research was conducted was a tube with an external diameter of 40 mm, this element was rigidly mounted at a distance of 1 meter from the free end, simulating cantilever conditions. The stimulation of the object to vibration was carried out using the MFC actuator, while the system response was recorded in the xPC Target environment. Based on the measurement data, the mathematical model of the object was identified in the discrete domain using the ARX method. The obtained model was used to design a controller based on the pole location method, which was implemented on a real test stand. The experimental results showed the effectiveness of the designed control system in reducing the amplitude of natural vibrations of the structure. The use of MFC elements as sensor elements and actuators enabled effective vibration damping in real time, confirming the usefulness of the proposed solution in the context of improving the precision of robotic systems.

Keywords: active vibration control, lightweight robot arm, macro fiber composite, PID controller, system identification, piezoelectric actuators.

Notations

ARX – AutoRegressive with eXogenous inputs,
ARMAX – AutoRegressive Moving Average with eXogenous inputs,
LWR – lightweight robot (arm),
MFC – macro fiber composite,
PID – proportional–integral–derivative (controller),
xPC – real-time MATLAB/Simulink execution environment (xPC Target),
RMS – root mean square,
FEM – finite element method,
PZT – lead zirconate titanate (classical piezoelectric ceramic),
DoF – degrees of freedom.

1. Introduction

Modern robotics, particularly in the context of lightweight robotic arm structures, demands increasingly higher precision in positioning, faster response times, and adaptability to dynamic environments. One of the

central challenges in the design of robotic systems is the effective suppression of structural vibrations. These disturbances necessitate slower operational speeds and longer cycle times, ultimately leading to increased production and maintenance costs (Iskandar *et al.*, 2020). Despite ongoing advancements in articulated robotic arms, which are widely used in manufacturing and material handling, these systems remain vulnerable to undesirable dynamic behaviour, particularly vibrations. The prevailing design trend of reducing structural mass to lower energy consumption further exacerbates the susceptibility of LWRs to vibration-induced instability (Zhong *et al.*, 2019). Vibrations reduce operational precision and cause unintentional deviations in the end-effector trajectory. In dynamic and unpredictable environments, external disturbances can compromise both accuracy and repeatability – two critical parameters in automated and precision-driven applications. This issue is particularly pronounced in lightweight robot arms (LWRs), where mass reduction is typically achieved at the expense of structural stiffness, increasing the system's susceptibility to adverse dynamic effects. The study on the control of a flexible arm manipulators started as a part of the space robots research because due to LWR light weight, ease of maneuverability and less power consumption these kind of manipulators are very useful for space applications (Sobieszczanski-Sobieski, Haftka, 1997; Cannon, Schmitz, 1984). Flexible manipulators are used also for surgical and micro-surgical operation, describes research in active instruments for enhanced accuracy in micro-surgery (Leniowska, Leniowski, 2012; Riviere *et al.*, 2003). Nowadays, in science and engineering, the pursuit of lightweight design represents both a well-established and continually evolving area of research. In contemporary applications, lightweight solutions are expected to meet not only technical and economic criteria but also align with principles of sustainability.

Designing a control law for a flexible robotic system involves reconciling conflicting requirements. On one hand, the system is expected to exhibit rapid dynamic responses. When the structure and parameters of the robot's model are accurately known, control strategies based on the control theory can be applied to achieve high-speed performance. In practice, however, the model and its parameters are often only approximately known, which complicates the synthesis of such control laws. Precise dynamic modelling plays a fundamental role in model-based control, which remains one of the most widely used approaches for industrial robotic systems. The mathematical models of flexible robots have been considered by many researchers. Typically, obtaining an accurate dynamic model involves two main stages:

1. Formulating dynamic equations using established methods such as the principle of virtual work, the Newton–Euler formalism, Kane's method (Book, 1984; De Luca, Siciliano, 1989; 1991), or the finite element method (Bayo, 1987; Chedmail, Khalil, 1989).
2. Estimation of dynamic parameters through identification techniques that rely on the linear parameterisation of the dynamic model (Ljung, 1999; Söoderström, Stoica, 1983). This approach refers to the methodology of determining a mathematical representation of a system by fitting model parameters to observed input–output data. The central idea is to capture the intrinsic dynamics of the system within a chosen model structure, such as transfer functions or state–space formulations.

Over the past several decades, identification techniques have received considerable attention, as it provides a systematic way to translate experimental measurements into relative reliable models. Such models not only offer insight into the underlying physical or engineering processes but also serve as essential tools for simulation, control design, and optimization. For these reasons there has been a strong trend towards extending or even replacing classical model-based approaches with modern data-driven and hybrid methods. These techniques enable the design of controllers directly from experimental input–output data, thereby reducing the dependency on precise analytical models, which are often difficult to obtain for lightweight flexible structures with distributed parameters and nonlinearities. A comprehensive review by Yang *et al.* (2024) summarizes the rapid development of data-driven vibration control (DDVC), highlighting subfields such as iterative learning control, reinforcement learning, and model-free adaptive control, all of which demonstrate strong potential for vibration suppression in robotic and aerospace applications. The authors emphasize that data-driven methods provide advantages in adaptability to uncertainties, reduced modelling effort, and robustness against disturbances, which are particularly valuable for thin-walled, lightweight manipulators. Similar directions are also explored in the context of pipelines and fluid-conveying structures, where system identification combined with adaptive feedback

improves damping performance under variable operating conditions (Ding, Ji, 2023). This experimental-based technique has been also used by the authors in active vibration suppression (Leniowska, Kos, 2009; Leniowska, Mazan, 2015; Leniowska *et al.*, 2022; Pater *et al.*, 2024) and it is applied herein.

An important stream of recent research focuses on advanced robust control strategies. For example, Qin and Xin (2023) proposed a data-driven H_∞ control method that ensures stability and performance guarantees without requiring an exact plant model, achieving effective vibration attenuation in flexible beam-like structures. These approaches stand in contrast to classical ARX/ARMAX-based identification, positioning themselves as scalable solutions for real-time industrial implementations. Complementary work by Zhang *et al.* (2025) has shown that combining data-driven design with hybrid piezoelectric actuators leads to significant improvements in damping efficiency, which suggests new pathways for integrating simple controllers with advanced materials in practice.

Parallel to algorithmic progress, actuator and sensor technologies have undergone significant advances. While traditional lead zirconate titanate (PZT) ceramics remain widely studied, modern macro fiber composites (MFC) and adaptive hybrid actuators offer improved flexibility, reduced fragility, and higher strain energy density, making them well suited for lightweight robotics applications (Masa'id *et al.*, 2023). Recent studies have also highlighted the importance of distributed sensor–actuator networks, where piezoelectric patches are optimally placed to target dominant modes and increase control authority. In addition, semi-active strategies such as piezoelectric shunt damping circuits have been extensively reviewed by Marakakis *et al.* (2019), who underline that shunted PZT elements – when combined with active feedback – enable hybrid approaches balancing energy efficiency with wideband damping performance. This trend towards multi-modal, hybrid suppression strategies positions classical active control within a broader landscape of contemporary solutions.

Within this broader context of active vibration and noise suppression, adaptive feedforward control methods based on least-mean-square (LMS) algorithms have been extensively investigated (Pawelczyk, 2004; Bismor, 2014a; 2014b). The limitations of the hybrid active feedforward noise control (ANC) configuration were analyzed, demonstrating that its practical performance depends heavily on careful consideration of secondary path dynamics and adaptive mechanisms. These results are particularly relevant for lightweight, distributed-parameter structures, where computational resources are constrained, and thus motivate the search for control strategies that remain effective under similar practical limitations in vibro-acoustic systems.

Taken together, these developments indicate that the state of the art in vibration suppression is moving towards integration: data-driven algorithms coupled with advanced actuators and hybrid control architectures. For lightweight robotic arms, this means that even relatively simple controllers such as PID, when embedded in modern mechatronic setups and supported by efficient identification procedures, can remain competitive. Our study, therefore, complements these advances by demonstrating how a classical control law can be effectively combined with MFC-based actuation in a real-time environment, providing a practical and experimentally validated benchmark against which more complex, data-driven strategies may be assessed.

This paper proposes the usefulness solution in the context of improving the precision of robotic systems based on the use of MFC piezoelectric elements, which serve both as sensors and actuators. The key innovation lies in the integration of experimentally derived dynamic models with a real-time digital PID controller, enabling effective damping of vibrations under realistic operating conditions. Despite the development of more advanced control methodologies in recent years, PID controllers continue to be widely applied across various industrial domains owing to their effectiveness and robustness as well as in feedback control of flexible robot manipulators (Ho, Tu, 2005; Ang *et al.*, 2005; Akyuz *et al.*, 2011). The developed control system was validated through both simulation and physical experiments. Potential applications include lightweight robotic systems and high-precision devices where vibration minimization is critical to performance and operational safety. This paper is structured as follows: In Sec. 2 the considered object, the aluminium tube that serves as a prototype for the LWR robotic arm is described and technical specifications of the MFC elements are given. In Sec. 3 a method of dynamic parameters estimation through identification techniques to develop mathematical model is provided. Section 4 provides brief notes on the PID controller used. Section 5 gives an overview on the implemented program code. View on the complete system which represents a digital PID controller with supporting hardware is given in Sec. 6. Finally, Sec. 7 provides the testing results and conclusions.

2. Materials and methods

In the context of research on active vibration damping in lightweight robotic structures of the LWR type, a thin-walled tube made of aluminum alloy Al99.5 was selected as a representative model of a robotic arm. This component, characterized by appropriately selected geometric dimensions (detailed in Table 1), was rigidly clamped at one end, simulating cantilever boundary conditions. On its surface, MFC piezoelectric elements were mounted, serving a dual role as sensors detecting vibrations and actuators for excitation and compensation of vibrations (Fig. 1). The locations of the MFC elements distributed along the object are presented in Table 2. One of the transducers was configured as an exciter, generating a chirp signal ranging from 1 Hz to 1 kHz over a period of 30 s.

Table 1. Geometric parameters of the structural element.

Parameter	Value [mm]
Outer diameter	40
Inner diameter	38
Wall thickness	2
Total length	2000
Clamping length	1000

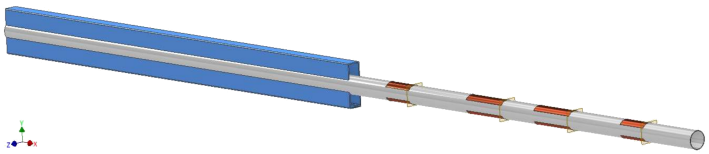


Fig. 1. Schematic representation of the tested structure: a thin-walled aluminum tube equipped with MFC piezoelectric transducers for active vibration control.

Table 2. Arrangement of MFC elements relative to the mounting point ($x = 0$ at the center of the object).

Element	x_{left} [mm]	x_{middle} [mm]	x_{right} [mm]	Width [mm]
MFC 1	109.0	142.5	176.0	67.0
MFC 2	343.0	394.5	446.0	103.0
MFC 3	540.0	591.5	643.0	103.0
MFC 4	792.5	826.0	859.5	67.0

The experimental placement of the MFC elements was based on modal analysis with a Polytec laser vibrometer. Sensors were positioned away from the modal nodes to maximize strain sensitivity, and the locations of the elements are listed in Table 2. Four transducer pairs were mounted on the upper and lower surfaces of the tube to capture the structural response and provide actuation authority for vibration suppression. Real-time experiments were conducted using the xPC Target platform. A linear parametric model of the tube was identified directly from the input–output data and subsequently used for controller design. This approach enabled a realistic representation of the system's dynamic behavior and facilitated the development of an effective vibration control algorithm.

MFC transducers consist of parallel piezoelectric fibres embedded in a polymer matrix, which provides flexibility and mechanical robustness. Constructed from parallel-aligned piezoelectric fibers embedded in a polymer composite matrix, MFC transducers exhibit suitability for operation under dynamic conditions and can be mounted on surfaces with irregular geometries.They can be bonded to curved or irregular surfaces.

In this study, MFC transducers were applied to actively suppress vibrations in a thin-walled aluminum tube which serves as a lightweight robot arm. Locations of sensors and actuators were selected from the modal analysis to avoid nodes. The transducers were mounted using a two-component adhesive, ensuring a durable bond with minimal influence on the mass and stiffness of the structure.

Compared to traditional PZT piezoceramics, MFCs exhibit several distinct advantages, primarily due to their flexibility, resistance to mechanical deformation, and high effectiveness in vibration damping. Moreover, each

element is capable of both sensing strain and generating actuation forces. The best performance was obtained with elements of 35 mm × 103 mm; smaller transducers were less effective. The parameters of the elements used are provided in Table 3.

Table 3. Technical specifications of the MFC element.

Parameter	Value	Unit
Model	06L18-044D	–
Width	34	mm
Length	103	mm
Thickness	0.3	mm
Polarization voltage	+1500 / −500	V
Piezoelectric coefficient d_{31}	−2.1E+02	pm/V
Piezoelectric coefficient d_{33}	4.6E+02	pm/V

Described above the thin-walled aluminium tube was used as a test model, selected due to its geometric and material properties, which make it highly susceptible to vibrations. The aim of this study was to develop an effective active vibration reduction algorithm for the aluminium tube which serves as a prototype of the LWR robotic arm. The foundation of the research involved the use of piezoelectric materials in the form of MFC elements, which act both as sensors and actuators responsible for active vibration suppression. To achieve the research objectives, a dedicated laboratory test stand had to be designed and constructed (Pater *et al.*, 2024). It was developed to enable vibration measurements, identification of the dynamic properties of the structure, and their active suppression using appropriately selected controllers.

3. Mathematical model

Effective vibration suppression requires a thorough understanding of the robot's dynamic characteristics. Dynamic identification of mechatronic systems is a process aimed at developing a mathematical model of the investigated object based on experimental measurement data. This method is particularly useful in cases where constructing an accurate theoretical model is challenging due to complex system geometry, numerous physical parameters, or the presence of nonlinearities. Estimation of dynamic parameters through identification techniques involves analyzing the system's response to known excitation inputs and constructing mathematical models that accurately reflect its behavior. One such approach is the use of MFC elements for sensing, combined with ARMAX-based system identification (Isermann, 2013) to model dynamic properties. These experimentally developed models are crucial for the design of active feedback-based vibration control systems and also preserve the essential features of the object dynamics within the considered vibration band. The key components of the experimental approach include the appropriate selection of the excitation signal and precise measurement of the system's response. This allows to extract essential information about the system's dynamic characteristics, such as resonance frequencies, damping properties, and transfer functions.

The integrated excitation and data-acquisition chain is shown in Fig. 2. A precomputed `.wav` signal is played by the laptop's built-in sound card; the line-out feeds a power amplifier that drives the MFC actuator. For

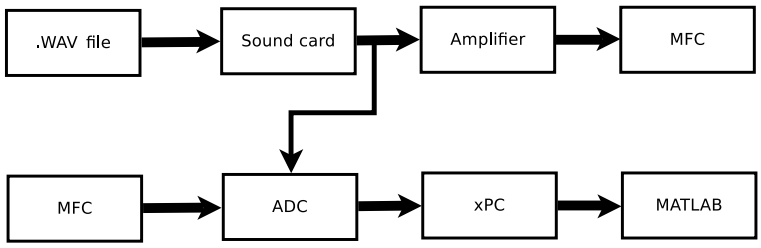


Fig. 2. Integrated system for excitation generation and data acquisition. The sound-card line-out is split: one branch drives the amplifier–MFC actuator path, the other is sampled by the ADC (measurement card of the `xPC` target) for synchronization.

synchronization, the same line-out is routed in parallel to the ADC (the measurement card integrated with the `xPC` target), which simultaneously samples it together with the MFC sensor output. The `xPC` target performs sampling and buffering and streams the data to MATLAB for recording and analysis. During identification experiments the sampling frequency was 10 kS/s for all channels, and the pre-processing consisted of removing the DC component (detrending by subtracting the sample mean) from each time series.

In the experiment, a chirp signal was used, which enables excitation of the tested system over a wide frequency range within a single measurement run. This approach allows obtaining the frequency response and identifying the system's resonant frequencies without the need to repeat the experiment for individual frequencies. Such a method is widely used in the dynamic system analysis, as it allows for efficient and rapid acquisition of the system's frequency response.

The time-domain data collected during the experiment was imported into the MATLAB environment using the System Identification Toolbox. This module enables data analysis and the identification of mathematical models, both in the form of transfer functions and state-space models. With advanced algorithms, it is possible to accurately fit the model structure to the actual behaviour of the studied system. This requires proper selection of the sampling frequency to ensure sufficient time resolution, particularly when analysing higher-frequency components. The identification process involves matching the model structure including the number of poles and zeros to the recorded data. Through an iterative parameter-fitting procedure, a model was developed that accurately reflects the dynamic properties of the robot arm. Such a model can subsequently be used for control system design, simulation, and analysis of how different parameters affect the system's dynamics. An experimental approach based on real measurement data offers several advantages over purely theoretical methods. Models developed from experimental data can accurately capture nonlinearities, complex geometries, and unpredictable dynamic effects. Measurement technologies based on piezoelectric sensors and laser vibrometry enable detailed analysis even of complex mechatronic structures. The resulting mathematical models provide a solid foundation for the further design and optimization of dynamic systems. As part of the identification process, commonly used ARX (AutoRegressive with eXogenous inputs) and ARMAX (AutoRegressive Moving Average with eXogenous inputs) models were employed. These are fundamental tools for modelling linear systems. Both models use input and output data to create a difference equation that describes the system's dynamics. The key distinction between ARX and ARMAX lies in their treatment of noise – the ARMAX model includes an additional term that accounts for the dynamics of noise components (Ljung, 1999):

$$y(t) + a_1 y(t-1) + \ldots + a_n y(t-n) = b_1 u(t-1) + \ldots + b_m u(t-m) + e(t), \tag{1}$$

where $y(t)$ denotes the output signal, $u(t)$ represents the input signal, a_i and b_i are the model coefficients, and $e(t)$ denotes the noise term, which is considered white noise. The model assumes that the disturbances are random in nature and affect only the output.

4. PID control

In the developed active vibration control system, the PID controller was implemented in a discrete-time form, consistent with the digital real-time architecture of the xPC Target platform. Instead of the continuous PID equation, the controller was realised using the following standard discrete transfer function:

$$G_{\mathrm{PID}}(z) = K_p\left(1 + \frac{T_s}{T_i}\frac{1}{1-z^{-1}} + \frac{T_d}{T_s}(1-z^{-1})\right), \tag{2}$$

where K_p is the proportional gain, T_i the integral time constant, T_d the derivative time constant, and T_s the sampling period. The integral part is implemented using the discrete accumulator $1/(1-z^{-1})$, while the derivative part uses the backward difference $(1-z^{-1})$, ensuring numerical stability and robustness to measurement noise.

This discrete transfer-function formulation is well suited for real-time vibration suppression because it provides predictable phase characteristics and stable operation near high-Q resonances. In this study, the PID controller was tuned to increase the effective damping of the dominant vibration modes (214 Hz and 574 Hz) identified

in Sec. 3. Combined with MFC sensor–actuator pairs, the discrete PID controller enabled effective attenuation of resonant oscillations under both harmonic and broadband excitations. The final values of the tuned parameters used during experiments are listed in Table 4.

Table 4. PID controller settings used in the experiment.

Parameter	Value
Proportional (P)	5.5
Integral (I)	4.0
Derivative (D)	5.0

5. Model identification

The objective of this study was to develop an effective algorithm for active vibration reduction in the LWR robotic arm. The research was based on an experimental mathematical model of the system, obtained using the ARX (AutoRegressive with eXogenous inputs) method. Data was recorded and subsequently used in the model identification process. The transfer function of the considered object takes the form:

$$G(z^{-1}) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + b_4 z^{-4} + b_5 z^{-5}}{a_0 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + a_4 z^{-4} + a_5 z^{-5} + a_6 z^{-6}}. \tag{3}$$

The obtained data for the discrete model are presented in Table 5.

Table 5. The values of the A and B coefficients adopted in the object identification process.

i	$A(z^{-i})$	$B(z^{-i})$
5	−4.473	0.0870
4	8.095	−0.3335
3	−7.392	0.4914
2	3.393	−0.3298
1	−0.6226	0.0851

6. Vibration suppression simulation and tests

In the next step, the PID controller was tuned in a simulation environment based on the identified models. The PID controller parameters used during the experiments are summarized in Table 4. These settings were selected based on the tuning procedure described in Sec. 4 and were applied to the physical object during testing. Providing these values ensures that the presented control performance can be independently verified.

The response of both the system and the controller was recorded for a first vibration resonanse frequency of 214 Hz, Fig. 3.

Similar effectiveness was observed at the second resonance frequency of 574 Hz, where the vibration amplitude was considerably reduced when the controller was turn on. In subsequent stages of the research, the object was subjected to excitation composed of two harmonics (214 Hz and 574 Hz). Despite the increased signal complexity, the PID controller effectively reduced the amplitudes of both vibration components, demonstrating its capability to operate under more complex dynamic conditions (Fig. 4).

Additionally, when applying a chirp signal with a frequency sweep ranging from 1 Hz to 1 kHz, an increase in vibration amplitude was observed near the resonance frequencies. After activating the controller, vibrations were successfully suppressed, especially at critical points within the frequency band, confirming the versatility of the control system and its ability to perform under variable excitation conditions (Fig. 5).

The results of the experimental tests are presented further. Data were recorded in the same manner as in the previous cases.

During the conducted experiments with the real mechanical system subjected to active vibration reduction, the effectiveness of the PID controller was evaluated. Tests were performed for various types of excitations,

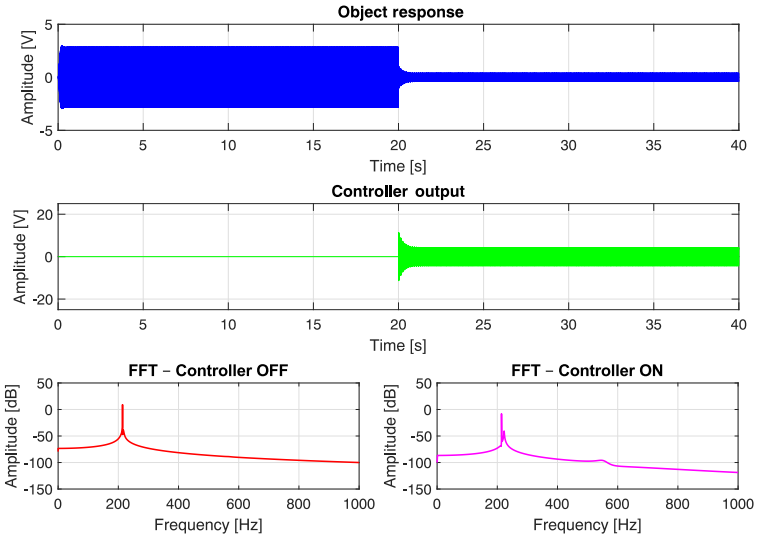


Fig. 3. Simulation result – time course and corresponding frequency spectrum for 214 Hz; 0 s to 20 s, the open loop control system, 20 s to 40 s the close loop control system.

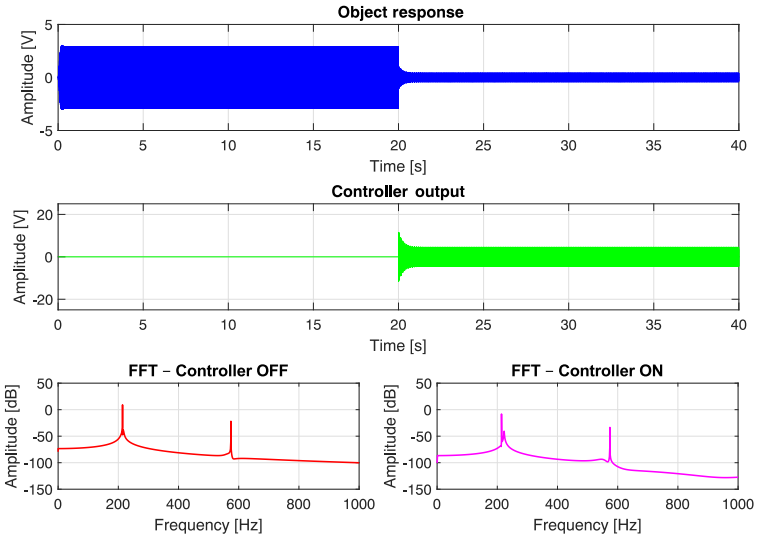


Fig. 4. Simulation result – time course and corresponding frequency spectrum for the sum of frequencies 214 Hz and 574 Hz, without a controller from 0 s to 20 s, and with a controller from 20 s to 40 s.

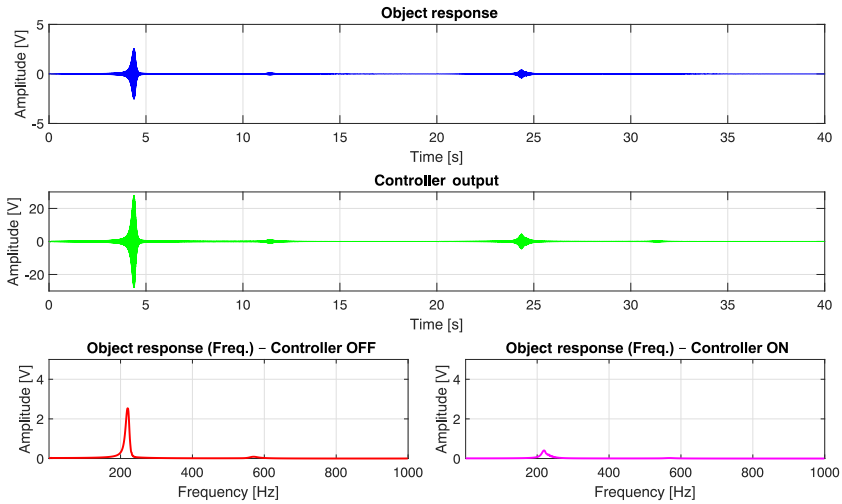


Fig. 5. Simulation result – time course and corresponding frequency spectrum of chirp signal with frequency 1 Hz to 1000 Hz, without controller from 0 s to 20 s, and with controller from 20 s to 40 s.

including harmonic signals at resonance frequencies of 214 Hz and 574 Hz, their combination, and a chirp signal. In each case, the controller was activated after a predefined time interval, allowing comparison of the system's behavior before and after control activation. For the harmonic excitation at 214 Hz, the system without control exhibited large amplitude vibrations characteristic of resonance. After enabling the PID controller, a significant reduction in the oscillation amplitude and gradual damping of vibrations were observed (Fig. 6). The damping process was smooth and free of overshoot, indicating properly tuned controller settings and its effectiveness in improving system stability. Similar phenomena were observed for the resonance frequency of 574 Hz. Without the controller, the system showed sustained oscillations with considerable amplitude. Upon activation of the PID controller, a rapid and effective vibration reduction occurred, enhancing the quality of the dynamic response. This process was also stable and free from adverse effects, confirming the efficiency of the applied control system. When the object was subjected simultaneously to two harmonic excitations at 214 Hz and 574 Hz, the system exhibited particularly unfavorable dynamic behavior with large oscillation amplitudes in the uncontrolled state. Activation of the PID controller resulted in a significant reduction of vibration amplitude and gradual suppression of oscillations caused by the superposition of the two resonant components. The vibration reduction levels obtained in the simulations were highly favorable for the considered resonant frequencies (51.11 % and 48.51 %). However, the reductions achieved in the experimental tests were significantly lower; for the 214 Hz resonant frequency, a reduction of 21.53 % was observed (Table 6). When the object is forced with a chirp signal, the results are slightly better. Detailed results are summarized in Table 7. Nevertheless, the results confirm that the proposed solution enables effective vibration reduction despite its simple structure.

Table 6. Vibration attenuation based on the object response.

Excitation	Average amplitude controller OFF [V]	Average amplitude controller ON [V]	Attenuation [%]
214 Hz (simulation)	2.25	1.10	51.11
214+574 Hz (simulation)	2.35	1.21	48.51
214 Hz (experiment)	3.53	2.77	21.53

Table 7. Vibration attenuation based on the object response.

Excitation	Maximum amplitude controller OFF [V]	Maximum amplitude controller ON [V]	Attenuation [%]
Chirp (simulation)	2.55	0.85	66.67
Chirp (experiment)	3.85	2.37	38.44

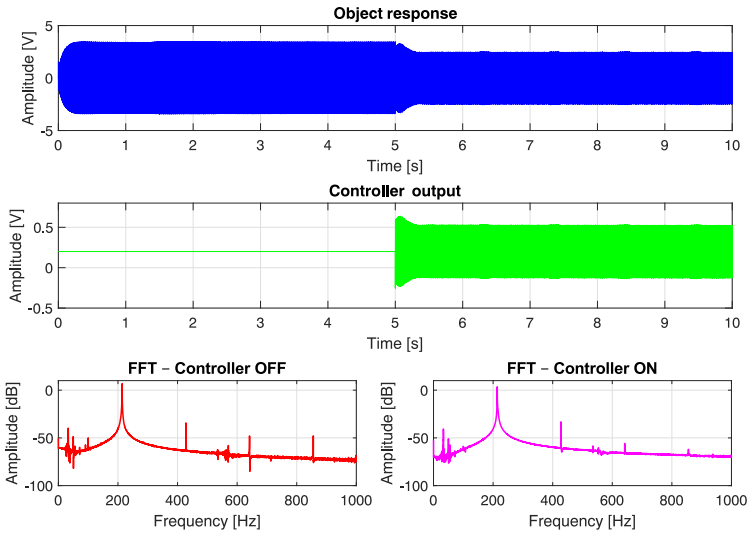


Fig. 6. Experimental results – the time response for the frequency of 214 Hz excitation, 0 s to 5 s the open loop control system, 5 s to 10 s the close loop control system.

7. Conclusion

The conducted research confirmed the effectiveness of the developed active vibration reduction system utilizing a PID controller. The control system, designed based on an ARX-type mathematical model of the object, demonstrated high efficiency both in simulation environments and real-world conditions. The obtained results validate the correctness of the adopted assumptions as well as the suitability of the chosen modeling and control methods which is confirmed by the data in the Table 6 and Table 7. The use of MFC piezoelectric materials as sensors and actuators proved to be an effective and practical solution (Fig. 7). The high sensitivity of these elements, combined with their capability for active vibration damping, enabled precise monitoring and elimination of vibrations in lightweight structures such as thin-walled tubes modeling robot arms. The developed control system demonstrated its versatility by effectively suppressing vibrations induced by both simple harmonic excitations and more complex input signals. This indicates its potential applicability across a wide range of engineering applications, particularly in robotics and mechatronics. A key factor in the overall process was the accurate modeling of the object, facilitated by tools such as the xPC Target system within the MATLAB environment. Precise identification of the dynamic properties of the structure directly contributed to the effectiveness of the developed control algorithm. The obtained results provide a solid foundation for further research and development. The presented solution can be adapted to more complex structures and integrated with advanced control strategies, such as adaptive or predictive control. This system opens new perspectives for active stabilization of lightweight robotic structures, contributing to improvements in their precision and operational reliability.

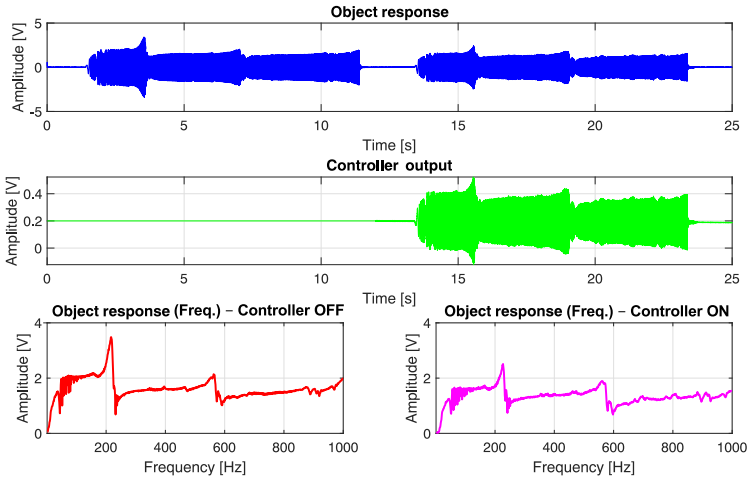


Fig. 7. Experimental results – the time response for chirp signal (from 1 Hz to 1000 Hz) excitation, 0 s to 5 s the open loop control system, 5 s to 10 s the close loop control system.

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.

Authors' contributions

Marcin Pater and Lucyna Leniowska conceptualized the study and wrote the original draft. Marcin Pater and Marcin Grochowina performed the measurements, analysis and contributed to data interpretation. All authors reviewed and approved the final manuscript.

Data availability statement

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

References

1. Akyuz I.H., Yolacan E., Ertunc H.M., Bingul Z. (2011), PID and state feedback control of a single-link flexible joint robot manipulator, [in:] *2011 IEEE International Conference on Mechatronics*, pp. 409–414, https://doi.org/10.1109/ICMECH.2011.5971320.
2. Ang K.H., Chong G., Li Y. (2005), PID control system analysis, design, and technology, *IEEE Transactions on Control Systems Technology*, **13**(4): 559–576, https://doi.org/10.1109/TCST.2005.847331.
3. Bayo E. (1987), A finite-element approach to control the end-point motion of a single-link flexible robot, *Journal of Robotic Systems*, **4**(1): 63–75, https://doi.org/10.1002/rob.4620040106.
4. Bismor D. (2014a), Comments on "A new feedforward hybrid ANC system", *IEEE Signal Processing Letters*, **21**(5): 635–637, https://doi.org/10.1109/LSP.2014.2310783.
5. Bismor D. (2014b), Partial update LMS algorithms in active noise control, [in:] *Proceedings of Forum Acusticum*.
6. Book W.J. (1984), Recursive Lagrangian dynamics of flexible manipulator arms, *The International Journal of Robotics Research*, **3**(3): 87–101, https://doi.org/10.1177/027836498400300305.
7. Cannon R.H., Schmitz E. (1984), Initial experiments on the end-point control of a flexible one-link robot, *The International Journal of Robotics Research*, **3**(3): 62–75, https://doi.org/10.1177/027836498400300303.
8. Chedmail P., Khalil W. (1989), Nonlinear decoupling control of flexible robots, [in:] *Advanced Robotics: 1989*, Waldron K.J. [Ed.], pp. 138–145, Springer, Berlin, Heidelberg.
9. De Luca A., Siciliano B. (1989), Trajectory control of a non-linear one-link flexible arm, *International Journal of Control*, **50**(5): 1699–1715, https://doi.org/10.1080/00207178908953460.
10. De Luca A., Siciliano B. (1991), Closed-form dynamic model of planar multi-link lightweight robots, *IEEE Transactions on Systems, Man, and Cybernetics*, **21**(4): 826–839, https://doi.org/10.1109/21.108300.
11. Ding H., Ji J. (2023), Vibration control of fluid-conveying pipes: A state-of-the-art review, *Applied Mathematics and Mechanics (English Edition)*, **44**(9): 1203–1236, https://doi.org/10.1007/s10483-023-3023-9.
12. He J., Gao P., Jin J., Yin H., Yu T. (2025), Experimental study on vibration control of pipeline system with enhanced active constrained layer treatment, *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, **239**(8): 757–773, https://doi.org/10.1177/09544100241307998.
13. Ho M.-T., Tu Y.-W. (2005), PID controller design for a flexible-link manipulator, [in:] *Proceedings of the 44th IEEE Conference on Decision and Control*, pp. 6841–6846, https://doi.org/10.1109/CDC.2005.1583262.
14. Isermann R. (2013), *Digital Control Systems*, 2 ed., Springer Science & Business Media.
15. Iskandar M., Ott C., Eiberger O., Keppler M., Albu-Schäffer A., Dietrich A. (2020), Joint-level control of the DLR lightweight robot SARA, [in:] *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8903–8910, https://doi.org/10.1109/IROS45743.2020.9340700.
16. Leniowska L., Grochowina M., Sierżęga M., Hołota B. (2022), An adaptive method for reducing vibrations of circular plates with recursive identification, *Applied Sciences*, **12**(5): 2723, https://doi.org/10.3390/app12052723.
17. Leniowska L., Leniowski R. (2012), The joint vibration analysis of a multi-link surgical manipulator, *Archives of Acoustics*, **37**(4): 475–482, https://doi.org/10.2478/v10168-012-0059-7.
18. Leniowska L., Kos P. (2009), Self-tuning control with regularized RLS algorithm for vibration cancellation of a circular plate, *Archives of Acoustics*, **34**(4): 613–624.
19. Leniowska L., Mazan D. (2015), MFC sensors and actuators in active vibration control of the circular plate, *Archives of Acoustics*, **40**(2): 257–265, https://doi.org/10.1515/aoa-2015-0028.
20. Leniowski R., Leniowska L. (2020), The multi-segment controller of a flexible arm, [in:] *ISR 2020; 52th International Symposium on Robotics*, pp. 1–8.

21. Ljung L. (1999), *System Identification: Theory for the User*, 2nd ed., Prentice Hall.
22. Marakakis K., Tairidis G.K., Koutsianitis P., Stavroulakis G.E. (2019), Shunt piezoelectric systems for noise and vibration control: A review, *Frontiers in Built Environment*, **5**: 64, https://doi.org/10.3389/fbuil.2019.00064.
23. Masa'id A., Lenggana B.W., Ubaidillah U., Susilo D.D., Choi S.-B. (2023), A review on vibration control strategies using magnetorheological materials actuators: Application perspective, *Actuators*, **12**(3): 113, https://doi.org/10.3390/act12030113.
24. Pater M., Leniowska L., Grochowina M. (2024), Active vibration control of pipe-shaped robot arm, *Engineering Transactions*, **72**(4): 389–402, https://doi.org/10.24423/EngTrans.3385.2024
25. Pawełczyk M. (2004), Adaptive noise control algorithms for active headrest system, *Control Engineering Practice*, **12**(9): 1101–1112, https://doi.org/10.1016/j.conengprac.2003.11.006.
26. Qin Z.-C., Xin Y. (2023), Data-driven H_∞ vibration control design and verification for an active suspension system with unknown pseudo-drift dynamics, *Communications in Nonlinear Science and Numerical Simulation*, **125**: 107397, https://doi.org/10.1016/j.cnsns.2023.107397.
27. Riviere C.N., Ang W.T., Khosla P.K. (2003), Towards active tremor canceling in handheld microsurgical instruments, *IEEE Transactions on Robotics and Automation*, **19**(5): 793–800, https://doi.org/10.1109/TRA.2003.817506.
28. Robinett R.D. III, Feddema J., Eisler G.R., Dohrmann C., Parker G.G., Wilson D.G., Stokes D. (2002), *Flexible Robot Dynamics and Controls*, Springer, New York, NY, https://doi.org/10.1007/978-1-4615-0539-6.
29. Sobieszczanski-Sobieski J., Haftka R. (1997), Multidisciplinary aerospace design optimization – Survey of recent developments, [in:] *AIAA 1996-711. 34th Aerospace Sciences Meeting and Exhibit*, https://doi.org/10.2514/6.1996-711.
30. Söoderström T., Stoica P. [Eds.] (1983), *Instrumental Variable Methods for System Identification*, Springer, Berlin, Heidelberg, https://doi.org/10.1007/BFb0009019.
31. Smith S.W. (2007), *Digital Signal Processing – A Practical Guide for Engineers and Scientists* [in Polish], Wydawnictwo BTC, Warszawa.
32. Yang W., Li S., Luo X. (2024), Data driven vibration control: A review, *IEEE/CAA Journal of Automatica Sinica*, **11**(9): 1898–1917, https://doi.org/10.1109/JAS.2024.124431.
33. Zhang R., Hyman L., Ding J., Keogh P.S., Bailey N.Y. (2025), Vibration suppression using piezoelectric actuator-based active flexure joints for high precision operations, *Mechanical Systems and Signal Processing*, **235**: 112816, https://doi.org/10.1016/j.ymssp.2025.112816.
34. Zhong Y., Wang R., Feng H., Chen Y. (2019), Analysis and research of quadruped robot's legs: A comprehensive review, *International Journal of Advanced Robotic Systems*, **16**(3), https://doi.org/10.1177/1729881419844148.

Archives of Acoustics Vol. 51, No. 1, pp. 223–232 (2026), https://doi.org/10.24423/archacoust.2026.4330

OSA 2025

Flanking Sound Transmission in Massive-Lightweight Connections

Agnieszka Wójtowicz*, Tadeusz Kamisiński, Jarosław Rubacha

AGH University of Krakow
Kraków, Poland

*Corresponding Author: awojtowicz@agh.edu.pl

Received September 11, 2025; revised January 8, 2026; accepted January 12, 2026; available online January 19, 2026; version of record March 17, 2026; published issue March 27, 2026.

Assessing the impact of flanking sound transmission is one of the most significant challenges in the process of designing building partitions. Acoustic parameters declared by manufacturers of lightweight systems are subject to errors of up to several decibels – and in the case of inaccurate construction on site, these differences can reach even higher values. One factor contributing to this is the phenomenon known as flanking sound transmission, which involves the transmission of acoustic energy through partitions connected to the partition directly dividing two adjacent rooms. For this reason, estimating the resultant acoustic insulation of a partition, taking flanking paths into account, is crucial at an early stage of the design process to ensure compliance with the requirements outlined in standard recommendations, and the literature. Currently, there are regulations and studies that provide guidance on calculating the estimated reduction in acoustic insulation due to flanking transmission. However, in practice, situations arise that have not yet been addressed in standards or the literature. Examples include partitions made of plasterboard, which are among the most common types of partition walls in Poland, yet are not covered by current normative procedures, as well as glass systems. This study aims to further explore this topic by analysing the impact of combining a massive partition with flanking lightweight partitions for selected structures (glass, plasterboard with single or double panelling, with full or partial sound-absorbing material infill, and without infill) and connection types.

Keywords: building acoustics, flanking sound transmission, sound insulation, statistical energy analysis.

1. Introduction

When estimating the sound transmission index of a partition, it is essential to take into consideration not only the direct partition between rooms, but also all alternative paths through which sound is transferred, known as flanking paths. These paths can significantly decrease the resultant sound insulation of a direct partition, particularly if the flanking partitions provide weaker insulation or have a lower density than the direct partition. A methodology for estimating the impact of flanking transmission for several types of partitions is described in the standard EN 12354-1 (International Organization for Standardization, 2017); however it does not cover all types of partitions encountered in practice. Although there have been numerous articles written on the matter, most of them discuss flanking transmission through timber or cross-laminated timber (CLT) partitions (Neusser, Bednar, 2022), which are not among the popular building materials in Poland, where gypsum board partitions prevail. Less frequently, other constructions are discussed, such as gypsum board partitions or double walls (Dijckmans *et al.*, 2019; Crispin *et al.*, 2017; Gerretsen, 2015; Schoenwald, 2008). This paper, however, examines a previously unstudied case: the use of lightweight gypsum board and glass partitions and their impact on the resultant sound insulation of a massive partition, using statistical energy analysis (SEA). Several cases are described – case 1 shows a basic situation of a single concrete partition between two rooms

without flanking paths, case 2 includes single-cladding gypsum-board-based flanking partitions. Cases 3 to 5 focus on double-cladding partitions without filling in the cavity between plates, with a cavity partially filled with sound-absorbing material, and with a cavity fully filled with a sound-absorbing material. Case 6 focuses on glass flanking partitions. All cases show results for C-, T-, and H-shaped connections.

2. Statistical energy analysis

SEA, widely described by Craik (1996) as well as Crocker and Price (1969), is one of the methods used to calculate sound and vibration transmission through a given acoustic system. The general principle of SEA is to create a model of a system consisting of smaller subsystems and then determine the equilibrium equations describing the energy flow between them. A subsystem is a set of modes with the same properties and similar modal energy; it represents a physical object, such as a partition, a room, or a void. In SEA, the measure of sound in a room or vibration of a partition is energy. The value of power transferred from subsystem i to subsystem j (W_{ij}) depends on the sound energy in the transmitting subsystem E_i, the angular frequency ω, and the energy loss factor η_{ij}:

$$W_{ij} = E_1 \omega \eta_{ij}. \tag{1}$$

Some of the energy leaving a subsystem is lost as heat or transferred to a different partition that is not a part of a system (W_{id}, W_{jd}), and some is radiated and transmitted to other subsystems (W_{ij}, W_{ji}). The energy entering the subsystem includes the external excitation (W_1) and the transmission from other subsystems (W_{ji}, W_{ij}).

In the SEA model, it is assumed that the energy is distributed evenly in all frequency bands, so that each band must contain an appropriate number of modes. The number of modes per unit frequency is defined by the modal density $n(f)$, and the number of modes in a band is expressed as ΔN for a given subsystem. The modal density depends on the type of wave and on the geometry, material, and boundary conditions of the subsystem. The number of modes depends on the width of the frequency band.

All modes in a given subsystem and frequency band are excited equally, and their responses are independent of one another. The modal density, according to Craik (1996), Kleiner, Tichy (2014), and Schoenwald (2008), is described as follows:

a) for plates:

$$n(f) = \frac{2\pi f S}{c^2}, \tag{2}$$

where S refers to the surface area of the plate and c is the wave speed. For bending waves on thin plates, modal density is described as

$$n(f) = \frac{\pi S f_c}{c_0^2}; \tag{3}$$

b) for rooms:

$$n(f) = \frac{4\pi f^2 V}{c_0^3} + \frac{\pi f S'}{2c_0^2} + \frac{L'}{8c_0}, \tag{4}$$

where V is the volume of the room, S' is the total surface of the room, and L' is the total length of all edges in the room;

c) for cavities: for thin voids and for low frequencies below the eigenfrequency of the first cross mode, calculated using the following formula:

$$f_{m,n,o} = \frac{c_0}{2}\sqrt{\frac{m^2}{l_x^2} + \frac{n^2}{l_y^2} + \frac{o^2}{l_z^2}}, \tag{5}$$

where m, n, o are positive integers (Craik, 1996; Schoenwald, 2008), and l_x, l_y, l_z are the dimensions of the void, the system is treated as 2D and Eq. (2) should be used, whereas above this frequency, the void is treated like a room, which implies the use of Eq. (4).

The number of modes in a given frequency band is calculated using:

$$N = n(f)\Delta f, \tag{6}$$

where

$$\Delta f = 0.23 f_m \tag{7}$$

for 1/3 octave bands, and

$$\Delta f = 0.707 f_m \tag{8}$$

for 1-octave bands, where f_m is the centre frequency of a given frequency band.

2.1. Damping

A key element in calculating the energy in individual subsystems is determining the energy loss coefficients, which define the energy flow between subsystems. The loss coefficient is the fraction of energy lost by a subsystem in one cycle. Damping is described by several types of loss factors: internal loss factors $\eta_{i,d}$, coupling loss factors (CLFs) η_{ij} and total loss factors η_i, where

$$\eta_i = \sum_{j=1}^{n} \eta_{ij} + \eta_{i,d}. \tag{9}$$

2.1.1. Coupling loss factors

The CLF describes the attenuation due to coupling between subsystems. It is the fraction of energy transferred from one subsystem to another in one cycle, and is generally expressed as follows:

$$\eta_{ij} = \frac{W_{ij}}{E_i \omega}. \tag{10}$$

The formulas for connections between different types of subsystems are as follows:

a) energy transfer from a room to a partition:

$$\eta_{ij} = \frac{\rho_0 c_0^2 S_j f_{c,j} \sigma_j}{8\pi f^3 V_i \rho_{s,j}}, \tag{11}$$

b) energy transfer from a partition to a room:

$$\eta_{ij} = \frac{\rho_0 c_0 \sigma_i}{\omega \rho_{s,i}}, \tag{12}$$

c) energy transfer between partitions:

$$\eta_{ij} = \frac{l_{ij} c_{g1} \tau_{ij}}{\omega \pi S_i}, \tag{13}$$

d) energy transfer from a cavity to a partition:

$$\eta_{ij} = \frac{\rho_0 c_0 f_{c,j} \sigma_j}{4\pi f^2 \rho_{s,j}}, \tag{14}$$

e) energy transfer from a cavity to a room:

$$\eta_{ij} = \frac{\tau_{ij}}{4\pi}, \tag{15}$$

f) energy transfer between rooms:

$$\eta_{ij} = \frac{c_0 S \tau_{ij}}{8\pi f V_i}, \tag{16}$$

where $f_{c,j}$ is the critical frequency for element j, S_j is the surface area of element j, V_i is the volume of the source room, $\sigma_{i/j}$ is the resonant radiation efficiency of element i/j, $\rho_{s,i/j}$ is the surface mass of element i/j, τ_{ij} is the transmission coefficient from element i to j, $c_{g,i}$ is the corrected group velocity, and l_{ij} is the length of the connection between elements i and j.

If the CLF from subsystem i to j is known, the energy flow from subsystem j to i can also be calculated using:

$$\eta_{ji} = \frac{\eta_i \eta_{ij}}{\eta_j}. \tag{17}$$

The transmission coefficient can be calculated using Eq. (18), as stated by Craik (1996):

$$\tau_{ij} = \left(\frac{\rho_0 c_0}{\pi f \rho_j (1-\mu^{-4})}\right)^2 \left\{\ln\left(\frac{2\pi f\sqrt{S}}{c_0}\right) + 0.16 + U(l_x, l_y) + \frac{1}{4\mu^6}\left[\left(2\mu^2-1\right)\left(\mu^2+1\right)^2 \ln\left(\mu^2-1\right) + \left(2\mu^2+1\right)\left(\mu^2-1\right)^2 \ln\left(\mu^2+1\right) - 4\mu^2 - 8\mu^6 \ln(\mu)\right]\right\}, \tag{18}$$

where $U(l_x, l_y)$ is a shape function that can be omitted if $0.1 < l_x, l_y < 10$, and μ is a square root of f_c/f, where f_c is the critical frequency.

2.1.2. Internal loss factors

The internal loss factor (ILF) represents the amount of energy lost by a subsystem and converted into heat, or transferred to another structure not included in the model, in one cycle. It is defined as follows:

$$\eta_{id} = \frac{W_{id}}{E_i \omega}. \tag{19}$$

ILFs for common material types can be found in Table 1.

Table 1. Internal loss factors of common materials.

Material	Internal loss factor $\eta_{id} \cdot 10^{-3}$
Steel	~0.1
Aluminium	~0.1
Glass	0.6–2
Concrete	4–8
Lightweight concrete	10–20
Autoclaved aerated concrete	10–20
Gypsum plate	10–15
Chipboard	10–30

2.1.3. Total loss factors

The total energy loss factor (TLF), denoted as η_i, is the sum of the energy loss due to coupling of subsystems (CLF) and the internal losses of a given subsystem. Equations for each type of subsystem are specified as follows:

a) TLF of a room:

$$\eta_i = \frac{2.2}{T_{60,i} f}, \tag{20}$$

b) TLF of a cavity:

$$\eta_i = \frac{c_0 \sum l\alpha'}{2\pi^2 f S_i}, \tag{21}$$

c) TLF of a partition:

$$\eta_i = \frac{c_g L \alpha}{2\pi^2 f S}, \tag{22}$$

whereas for massive partitions, it can be assumed that:

$$\eta_i \approx \frac{1}{\sqrt{f}} + 0.015, \tag{23}$$

and for a lightweight partitions:

$$\eta_i = \frac{c_0}{f\pi^2\sqrt{\frac{f_{\text{ref}}}{f}}}, \tag{24}$$

where $T_{60,i}$ is the reverberation time of element i, f is the frequency, c_0 is the speed of sound in air, l is the length of the cavity, α' is the average sound absorption coefficient in the void, S_i is the surface area of element i, c_g is the group velocity, and $f_{\text{ref}} = 1000\,\text{Hz}$. The group velocity, which describes the velocity at which energy is transported, is given by

$$c_g = \frac{\mathrm{d}\omega}{\mathrm{d}k}. \tag{25}$$

2.2. Transfer matrix

In order to calculate the energy in each subsystems, the first step is to form equilibrium equations, based on the assumption that the energy entering a subsystem is equal to the energy leaving it. Using these equations, a transfer matrix is formed:

$$\begin{bmatrix} -\eta_1 & \eta_{21} & \eta_{31} & \eta_{41} & \dots \\ \eta_{12} & -\eta_2 & \eta_{32} & \eta_{42} & \dots \\ \eta_{13} & \eta_{23} & -\eta_3 & \eta_{43} & \dots \\ \eta_{14} & \eta_{24} & \eta_{34} & -\eta_4 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \\ \dots \end{bmatrix} = \begin{bmatrix} -W_1/\omega \\ 0 \\ 0 \\ 0 \\ \dots \end{bmatrix}. \tag{26}$$

2.3. Analysed cases

For each analysed case, the same initial conditions were assumed, as presented in Table 2.

Table 2. Initial conditions assumed for the calculations.

Parameter	Symbol	Value
Initial power [W]	W_1	0.005
Speed of sound in air [m/s]	c_0	343
Density of air [kg/m^3]	ρ_0	1.2
Air temperature [°C]	T	20

It was assumed that both the source and receiving rooms were identical in terms of their dimensions and reverberation time. The assumed parameters of the rooms are listed in Table 3.

Table 3. Parameters of the source and receiving rooms assumed for the calculations.

Parameter	Symbol	Value
Width [m]	W_R	3
Length [m]	L_R	5
Height [m]	H_R	4
Volume [m^3]	V_R	60
Reverberation time [s]	$T_{60,R}$	0.6

The main partition between the source room and receiving rooms is a massive concrete wall. Flanking partitions were assumed to be gypsum board or glass-based. Their parameters are specified in Table 4.

The description of cases analysed in the study is presented in Table 5.

The connection shapes are presented in Fig. 1.

Table 4. Parameters of the walls.

Parameter	Symbol	Value		
		Massive wall	Gypsum board panel	Glass panel
Width [m]	W	3	3	3
Thickness [m]	T	0.15	0.012	0.0066
Height [m]	H	4	4	4
Surface [m^2]	S	12	12	12
Perimeter [m]	P	14	14	14
Density [kg/m^3]	ρ	2400	720	2500
Surface mass [kg/m^2]	m'	360	8.64	16.5
Critical frequency [Hz]	f_c	110	2846	1808
Young's modulus [N/m^2]	E	3.6e+10	2.4e+9	7.2e+10
Poisson's ratio	μ	0.2	0.3	0.2
Bending stiffness [Nm]	B	3.1e+6	380	1.8e+4

The cavity between the gypsum board or glass panels is assumed to be 0.05 m deep.

Table 5. Description of analysed cases.

Case number	Description
1	Basic model with no flanking partitions – source and receiving rooms divided by a 15 cm thick concrete wall.
2	One, two or four flanking partitions, consisting of a single gypsum board cladding with no absorptive filling; three connection types ('T', 'C', 'H').
3	One, two or four flanking partitions, composing of a double gypsum board cladding with no absorptive filling; three connection types ('T', 'C', 'H').
4	One, two or four flanking partitions, consisting of a double gypsum board cladding; cavity filled with an absorptive material in 50 %; three connection types ('T', 'C', 'H').
5	One, two or four flanking partitions, consisting of a double gypsum board cladding; cavity fully filled with absorptive material; three connection types ('T', 'C', 'H').

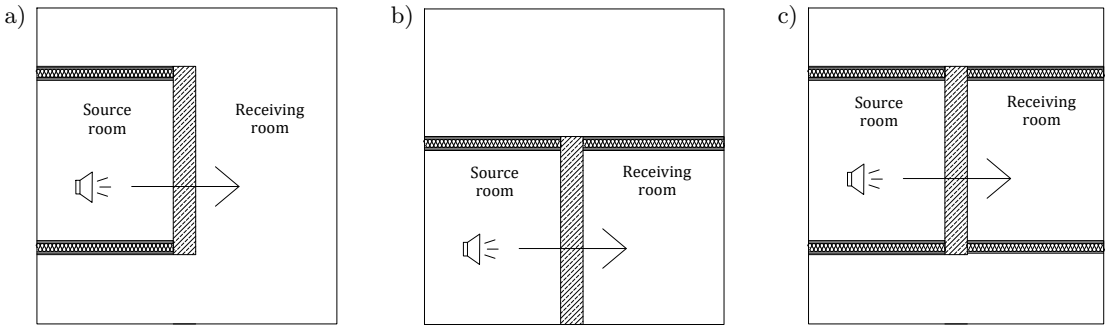


Fig. 1. Schematic layout of the analysed connections in case 2:
a) C-shaped connection, b) T-shaped connection, c) H-shaped connection.

2.3.1. Subsystems

The connections between subsystems were assumed according to Fig. 2 to Fig. 5. For cases 2 to 6, the connections are identical, as the double-cladding of gypsum board walls is considered as a single subsystem.

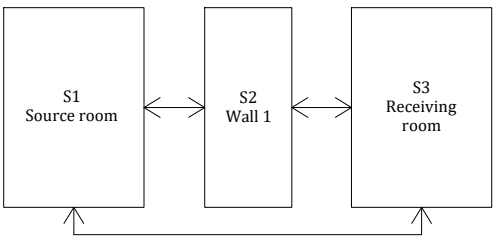


Fig. 2. Energy flow scheme between subsystems for case 1.

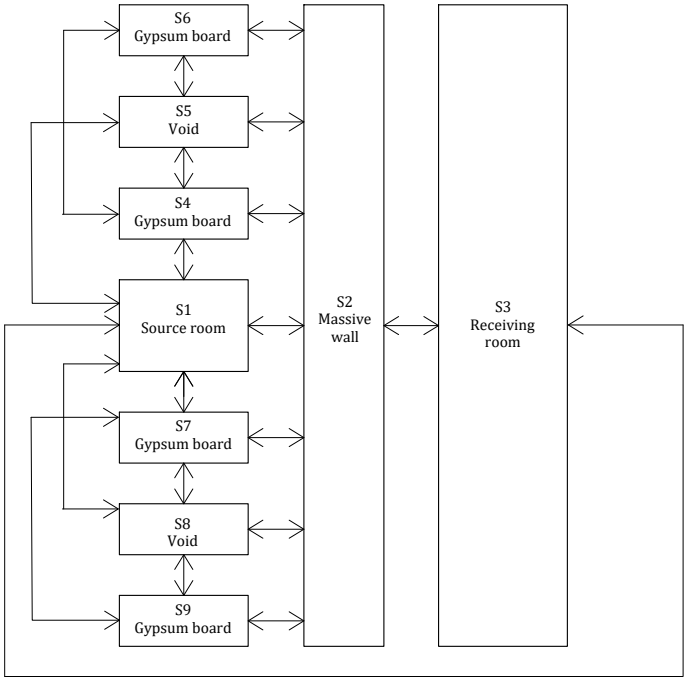


Fig. 3. Energy flow scheme between subsystems for C-shaped connections.

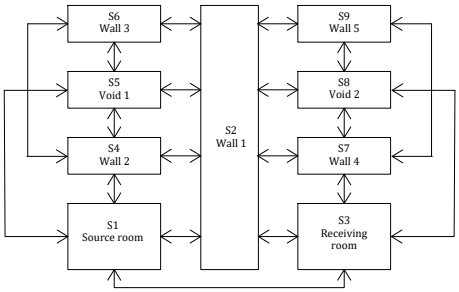


Fig. 4. Energy flow scheme between subsystems for T-shaped connections.

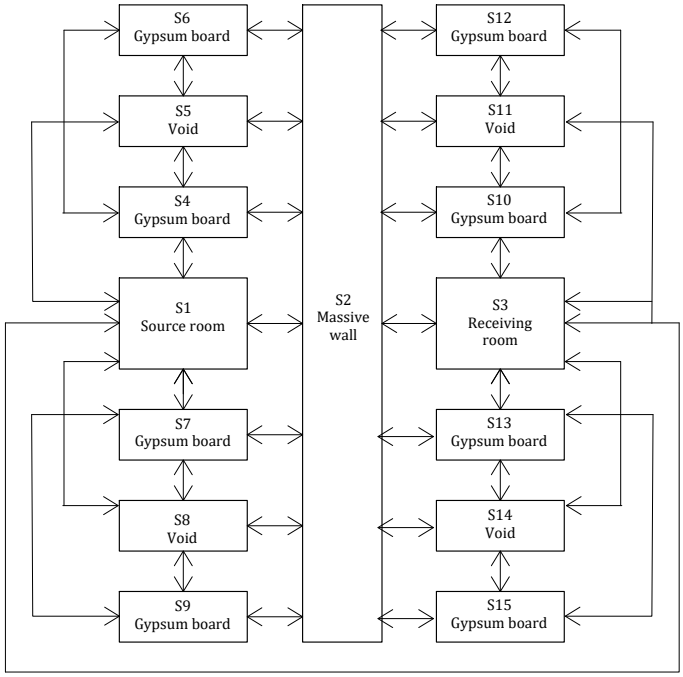


Fig. 5. Energy flow scheme between subsystems for H-shaped connections.

3. Results

For each case and each connection type, the energy in every subsystem was calculated. The results obtained for cases 2 to 6 were then compared to the basic situation (case 1) in order to show the difference between the resultant sound insulation of the massive wall with no flanking paths and the insulation obtained after adding lightweight partitions. The results are presented in Fig. 6 and Fig. 7.

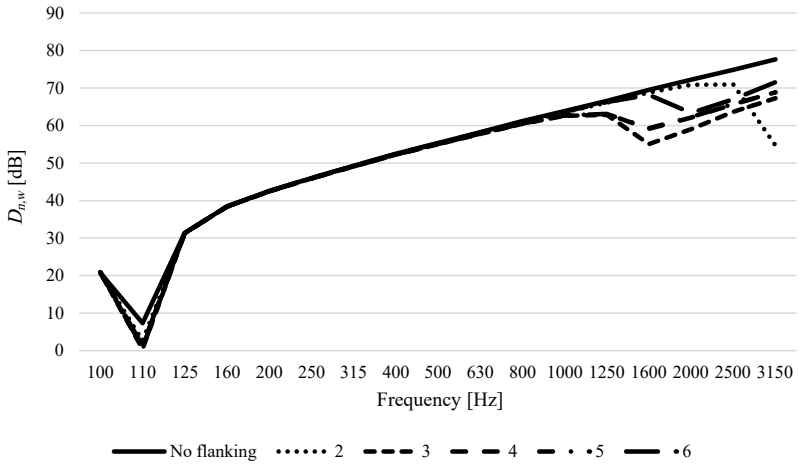


Fig. 6. Weighted sound reduction indexes for H-shaped connections (cases 1–6).

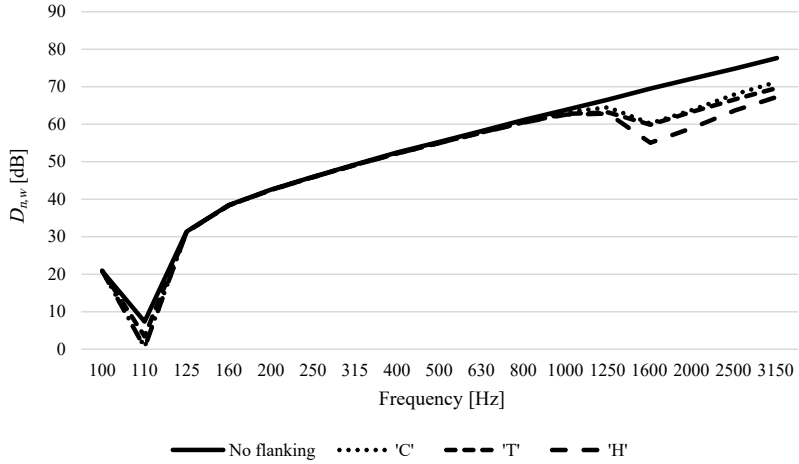


Fig. 7. Weighted sound reduction indexes for case 3 with three connection shapes.

The differences in sound insulation of H-shaped connections between the massive wall and gypsum walls with no filling, partial filling, and fully filled cavities are presented in Table 6.

The results show that the sound reduction index, compared to case 1, decreases above the critical frequency of the gypsum/glass board. In the case of single cladding, the reduction occurs above 2500 Hz, while in the case of double cladding, the reduction begins above 1250 Hz. In the case of glass flanking partitions, the decrease occurs above approximately 1850 Hz. The highest reduction of sound insulation can be observed for H-shaped connections, as these include the largest number of flanking partitions.

As for the absorptive material in the cavities, the resultant sound reduction difference between cases 3, 4, and 5 appear above the critical frequency of the gypsum board, which for the double-cladded wall is around 1400 Hz. Differences at higher frequencies are clearly noticeable and impact the overall resultant single-number sound insulation of the wall. The difference between cases 4 and 5, representing walls with partially and fully filled cavities, is also observed; however, the resultant single-number sound reduction index is $R_{A1} = 48$ dB. For case 3, the sound reduction index is $R_{A1} = 45$ dB.

Table 6. Sound insulation difference between three of the analysed cases.

Frequency [Hz]	$D_{n,e,w}$ [dB] – case 3	$D_{n,e,w}$ [dB] – case 4	$D_{n,e,w}$ [dB] – case 5
100	20.9	20.9	20.9
125	31.3	31.3	31.3
160	38.4	38.4	38.4
200	42.4	42.4	42.4
250	45.9	45.9	45.9
315	49.1	49.1	49.1
400	52.2	52.2	52.2
500	55.0	55.0	55.0
630	57.8	57.8	57.8
800	60.4	60.5	60.5
1000	62.6	62.7	62.7
1250	62.8	63.1	63.1
1600	55.1	59.1	59.3
2000	59.1	62.1	62.3
2500	63.7	65.7	65.7
3150	67.3	68.8	68.9

4. Conclusions

SEA, when applied correctly, can be an effective method for calculating the sound transmission of a partition, including contributions from flanking paths. While most recent publications focus on timber or massive constructions when discussing flanking transmission, this research analysed connections between gypsum-board-based or glass partitions and massive partitions, which are commonly used in Poland. The results for six different cases were presented, starting with a standard situation without flanking transmission. Cases 2 to 5 represent connections between a massive wall and a gypsum board wall, considering three different connection shapes (C-shaped, T-shaped, and H-shaped) and both single- and double-cladded walls. Filling the cavity with absorptive material was also taken into consideration. Case 6 represents the situation when the flanking partitions are glass systems for the same types of connections between partitions as in the case of gypsum board partitions.

The results indicate a noticeable impact of flanking paths on the resultant sound insulation of a massive wall. The reduction in sound insulation occurs above the critical frequency of the gypsum or glass board. For this reason, the damping in case 2, where the flanking walls have single cladding and therefore a higher critical frequency, is smaller than in cases 3 to 6. The shape of the connection is also significant, with H-shaped connections resulting in the greatest reduction among all analysed cases, as they include the largest number of flanking partitions.

Another factor that has an impact on sound insulation is the filling of the cavity between gypsum boards with absorptive material. The calculations imply that higher attenuation in the cavity increases the sound insulation of the massive partition, thereby the impact of flanking paths.

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.

Authors' contributions

Agnieszka Wójtowicz performed the analysis and contributed to data interpretation. Tadeusz Kamisiński and Jarosław Rubacha conceptualized the study and contributed to data interpretation. All authors reviewed and approved the final manuscript.

References

1. Craik R.J.M. (1996), *Sound Transmission Through Buildings: Using Statistical Energy Analysis*, Gower.
2. Crispin C., Mertens C., Dijckmans A. (2017), Detailed analysis of measurement results of flanking transmission across a junction composed of double walls carried out on a half scaled test bench, [in:] *24th International Congress on Sound and Vibration 2017*.
3. Crocker M.J., Price A.J. (1969), Sound transmission using statistical energy analysis, *Journal of Sound and Vibration*, **9**(3): 469–486, https://doi.org/10.1016/0022-460X(69)90185-0.
4. Dijckmans A., De Geetere L., Crispin C. (2019), Simplified prediction of the vibration reduction indices of double wall junctions, [in:] *Proceedings of 23rd International Congress on Acoustics*, https://doi.org/10.18154/RWTH-CONV-239666.
5. Gerretsen E. (2015), Extending EN 12354 sound insulation modelling to composed, light weight building systems, [in:] *44th International Congress and Exposition on Noise Control Engineering*.
6. International Organization for Standardization (2017), *Building acoustics – Estimation of acoustic performance of buildings from the performance of elements. Part 1: Airborne sound insulation between rooms* (ISO Standard No. ISO 12354-1:2017), https://www.iso.org/standard/70242.html.
7. Kleiner M., Tichy J. (2014), *Acoustics of Small Rooms*, Taylor & Francis, Boca Raton.
8. Neusser M., Bednar T. (2022), Measurement and estimation of the flanking impact sound transmission in timber frame building constructions, [in:] *Proceedings of the 24th International Congress on Acoustics*.
9. Schoenwald S. (2008), *Flanking sound transmission through lightweight framed double leaf walls: Prediction using statistical energy analysis*, Ph.D. Thesis, Technische Universiteit Eindhoven, https://doi.org/10.6100/IR637821.