eISSN 2300-262X
ISSN 0137-5075

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

ARCHIVES of ACOUSTICS

QUARTERLY

Vol. 50, No. 4, 2025

WARSAW

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

ARCHIVES of ACOUSTICS

QUARTERLY, Vol. 50, No. 4, 2025

In Memoriam

Research Papers

OSA 2025

Review Paper

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

Editorial Board Office

Pawińskiego 5B, 02-106 Warsaw, Poland
phone (48) 22 826 12 81 ext. 206
e-mail: akustyka@ippt.pan.pl https://acoustics.ippt.pan.pl

Indexed in BazTech, Science Citation Index-Expanded (Web of Science Core Collection),
ICI Journal Master List, Scopus, PBN – Polska Bibliografia Naukowa,
Directory of Open Access Journals (DOAJ)
Recognised by The International Institute of Acoustics and Vibration (IIAV)

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

PUBLISHED IN POLAND
Typesetting in LaTeX: Jezierska Katarzyna (Institute of Fundamental Technological Research PAS, Poland)

Archives of Acoustics Vol. 50, No. 4, pp. 419–420 (2025), https://doi.org/10.24425/aoa.2025.157220

In Memoriam

Professor Eugeniusz Kozaczka

1942 – 2025

Professor Eugeniusz Tadeusz Kozaczka, member of the Polish Academy of Sciences, passed away on May 27, 2025, in Gdynia. We write these words with profound sorrow and deep regret, mourning the loss of such an exceptional person as Professor Eugeniusz Kozaczka.

Eugeniusz Tadeusz Kozaczka was born on July 22, 1942, in Ćwików, Dąbrowa County, in the Lesser Poland Voivodeship. He completed his primary education in his hometown of Ćwików, then attended the Mining Vocational School in Katowice, after which he continued his studies at the Mining Technical Secondary School in Dąbrowa Górnicza.

While studying at the Technical School, one of his teachers encouraged him to pursue higher education, and he successfully passed the entrance examinations for both the AGH University of Science and Technology and the Jarosław Dąbrowski Military University of Technology in Warsaw (WAT). He chose the Military University of Technology, which he graduated from in 1971, earning a Master of Science in Engineering in the field of Technical Physics on the basis of a thesis entitled 'Studies of Thin BiMn Layers' (Bismuth-Manganese).

In 1976, at the Szewalski Institute of Fluid-Flow Machinery of the Polish Academy of Sciences, he obtained his PhD after defending his dissertation entitled 'Studies of Hydroacoustic Disturbances in Bounded Media'.

In 1980, at the Faculty of Mechanical Engineering of the Military University of Technology, after presenting his dissertation 'Study of Underwater Acoustic Disturbances Generated by a Ship Propeller', he earned his habilitation degree. In 1990, he was awarded the title of Professor of Technical Sciences, conferred through a procedure conducted at the Polish Naval Academy in Gdynia.

He served with the rank of Commander in the Polish Navy and worked for many years at the Polish Naval Academy named after the Heroes of Westerplatte. Later, he was employed at several other universities, including the Faculty of Ocean Engineering and Ship Technology of the Gdańsk University of Technology, the Bydgoszcz University of Science and Technology, and the Koszalin University of Technology.

Eugeniusz Kozaczka was a specialist in hydromechanics and hydroacoustics, and a passionate explorer and observer of the marine environment.

He conducted research on the generation and propagation of elastic waves in water, including: underwater noise produced by a moving vessel; the transmission of mechanical energy from the hull into the water; the transfer of vibrations into the marine environment by rotating blade systems; hydroacoustic noise generated by a cavitating propeller; and the propagation of underwater noise produced by a moving ship in shallow seas. His work also encompassed methodologies

and systems for measuring underwater noise; nonlinear hydroacoustics; the theory of high-intensity waves propagating in the specific nonlinear medium of seawater, including low-salinity seawater; the development of unique measurement and research tools for exploring the marine environment; the study of seabed structure using methods of linear and nonlinear acoustics; the theory of waveguide (modal) propagation of waves in shallow seas; and variability in the propagation conditions of elastic waves in the southern Baltic, resulting in unique studies of the acoustic climate of this region of the Baltic Sea.

His major theoretical achievements included: developing the theory of underwater noise generation by a ship's propeller; advancing the theory of nonlinear interactions of elastic waves in low-salinity seawater; and contributing to the theory of wave propagation in bounded aquatic environments.

His most important practical achievements included: a series of solutions for reducing vibrations in ship machinery and the associated underwater noise, implemented in the production of mine sweepers for the Polish Navy (7 patents); the construction of a shore-based hydroacoustic control and measurement station for naval vessels, which was introduced into service in the Navy in 1991 and is still used today to measure underwater noise from ships; the development and construction of the first measurement hydrophone in Poland; and carrying out globally unique measurements of the seabed sediments in the Gulf of Gdańsk using nonlinear acoustic methods.

He reviewed 11 applications for professorships, 22 habilitation theses (including 6 publication-based ones), and 16 doctoral dissertations, and he supervised 18 PhD graduates himself.

He participated in numerous research projects funded by the Committee for Scientific Research, the Ministry of Science and Higher Education, and the National Center of Research and Development.

He was a member of numerous Scientific Councils: the Faculty of Navigation and Naval Weapons of the Polish Naval Academy, the Institute of Oceanology of the Polish Academy of Sciences, the Faculty of Mechanical Engineering of the Koszalin University of Technology, the Faculty of Telecommunications and Electronics at ATR in Bydgoszcz, the Institute of Fluid-Flow Machinery of the Polish Academy of Sciences, and the Faculty of Ocean Engineering and Ship Technology of the Gdańsk University of Technology.

He was a member of numerous scientific societies and organizations, in which he held important positions. From 2002 to 2011, he served as Chairman of the Main Board of the Polish Acoustical Society; from 2003 to 2007, as Vice-Chairman of the Main Board of the European Acoustics Association; from 2007 to 2013, as a member of the board of the International Commission for Acoustics.

From 2011 to 2020, he was Chairman of the Acoustics Committee of the Polish Academy of Sciences, and since 2018 he has served as Chairman of the Scientific Commission of the PAS Branch in Gdańsk, 'Acoustics in Technology, Medicine, Marine Research, and Underwater Security Systems', actively contributing to the development of the national scientific community and international cooperation.

He was also a member of the Editorial Committee of Hydroacoustics and Chairman of the Editorial Board of Archives of Acoustics.

Professor Kozaczka, a corresponding member of the Polish Academy of Sciences since 2016, was awarded the Knight's Cross and the Officer's Cross of the Order of Polonia Restituta, as well as medals from the National Education Commission, the Ignacy Malecki Medal, and the Mikhail Lomonosov Medal. He also received numerous other awards, including those named after Xawery Czernicki, as well as awards from the Minister of Science and Higher Education, the Minister of National Defence, and the Rectors of the Gdańsk University of Technology and the Commandant of the Polish Naval Academy.

His scientific and educational accomplishments have left a lasting mark on Polish science. He will remain in our memory as a person of profound knowledge, devoted to his work, and service.

Professor Bogumił Linde, D.Sc.
Dr. Iwona Kochańska, D.Sc.

Archives of Acoustics Vol. 50, No. 4, pp. 421–433 (2025), https://doi.org/10.24425/aoa.2025.154831

Research Paper

Comparison of Methods for Determining the Airflow Resistivity of Porous and Covering Materials

Mykhaylo MELNYK[(1)], Jarosław RUBACHA[(2)*], Artur FLACH[(2)], Aleksandra CHOJAK[(2)], Tadeusz KAMISIŃSKI[(2)], Wojciech ZABIEROWSKI[(3)], Marek IWANIEC[(4)], Andriy KERNYTSKYY[(1)]

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

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

[(3)] *Department of Microelectronics and Computer Science, Lodz University of Technology*
Łódź, Poland

[(4)] *Department of Biocybernetics and Biomedical Engineering, Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Krakow*
Krakow, Poland

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

Received July 15, 2024; revised March 21, 2025; accepted September 21, 2025; published online November 13, 2025.

This article compares two methods for determining the airflow resistivity of porous and coating materials – a key parameter in sound absorption modelling. The analysis involves a modified static airflow measurement procedure in accordance with International Organization for Standardization [ISO] (2018), using a linear approximation algorithm (PLA), and a reverse method consisting of matching the measured absorption coefficient in an impedance tube to the Miki model. The analysis was conducted on both porous materials utilised in acoustic panel fillings and thin coverings. It is evident that both methods yield analogous outcomes for materials exhibiting low resistivity. However, for materials characterised by higher resistivity, discrepancies of up to 50 % were observed. Nevertheless, a high degree of agreement was obtained between the calculated and measured absorption coefficients. For thin coating materials, an air gap of at least 70 mm is required. For materials with a thickness of up to approximately 30 mm, differences in resistivity do not significantly affect the absorption coefficient. It is evident that both methods can be used to determine the airflow resistivity of porous materials and layered structures, supporting the effective selection of materials according to requirements.

Keywords: airflow resistivity; specific airflow resistance; sound absorption coefficient; impedance tube; porous materials.

1. Introduction

Porous media are used in various practical applications such as sound absorption and noise control (Allard, Atalla, 2009; Gibson, Ashby, 1997; Tao *et al.*, 2021). Porous structures present exceptional sound-absorbing properties in the mid-to-high-frequency ranges (Cao *et al.*, 2018; Zhao *et al.*, 2016). Porous materials are mesh-like structures with interconnected pores (Oliva, Hongisto, 2013). The pro-

cesses inside the pores, associated with the fluid's viscosity, generate heat from sound energy (Crocker, 2007; Huang *et al.*, 2023). Porous structures can be organic, inorganic, or mixed composite materials, including stone, wood, sponge, foam, rubber, non-woven fabrics, and textiles (Doutres *et al.*, 2010; Johnson *et al.*, 1987; Lafarge *et al.*, 1997).

The basic parameter describing the sound-absorbing properties of a material is its sound absorption coefficient a, defined as the part of incident energy that is not reflected (Allard, Atalla, 2009):

$$a = 1 - \frac{E_r}{E_{\text{tot}}}, \tag{1}$$

where E_r and E_{tot} are the acoustic energies of the reflected and incident waves, respectively.

The sound absorption coefficient can be determined by measurements in an impedance tube, using either the standing wave method (ISO, 1996) or the transfer function method (ISO, 2023). Measurements in the impedance tube are characterized by very good accuracy, cost-effectiveness, and testing flexibility. The sound absorption coefficient can also be established in a reverberation room (ISO, 2003; Vorländer, 2008). This procedure allows for the measurement of both flat and spatial elements, including auditorium chairs (Cuenca *et al.*, 2022; Rubacha *et al.*, 2012). However, this technique requires the use of specialized reverberation rooms. As an alternative to physical measurements, the sound absorption coefficient can also be determined using empirical models. These models are based on a large number of measurements of different materials, and the interpretation of the physical processes occurring in these materials. Delany and Bazley (1970), and later Miki (1990) proposed the simplest empirical models, which require only one parameter: airflow resistivity. Johnson *et al.* (1987), as well as Allard and Champoux (1992), suggested a more accurate physical model (the Johnson–Champoux–Allard (JCA) model) with five input parameters. Unfortunately, these parameters are usually difficult to estimate accurately. Bonfiglio and Pompoli (2013) presented an inverse method for determining the physical parameters of porous materials for use with the JCA model. Also, Vorländer (2008) discussed the difficulties in determining the parameters of porous media due to the complexity and variability of these materials, including factors such as geometrical configuration, porosity, and tortuosity for the JCA model. This means that more complex models may give worse results, with errors that are difficult to estimate.

Airflow resistivity is one of the basic parameters describing porous and nonwoven materials. It is used as input parameter for models, enabling the calculation of the sound absorption coefficient of single- and multi-layer materials using the transfer matrix method (Cox, D'Antonio, 2016; Dell *et al.*, 2021; Herrero-Durá *et al.*, 2019; Hou *et al.*, 2017). Kamisiński *et al.* (2012) showed that it can also be used to calculate the sound absorption coefficient of materials with coverings. The measurement of airflow resistivity is conducted according to two standards: ASTM C522-03 (2022) and ISO 9053-1 (2018). The primary distinctions between these two standards pertain to the measurement conditions specified for assessing airflow resistivity. ASTM C522-03 allows measurements in flow directions other than perpendicular, provided the airflow remains constant. On the other hand, ISO 9053-1 allows variable airflow measurements, but solely for flow that is perpendicular to the sample.

Both standards mandate measurements to be conducted under laminar flow conditions, ensuring the airflow resistance remains constant as flow speed varies. The ISO standard specifies that airflow resistance should be determined at a flow speed of $0.5 \cdot 10^{-3}$ m/s, either directly or by extrapolation from higher values if direct measurement is unfeasible at such a low velocities. Meanwhile, the ASTM standard mandates measurements at three different laminar flow rates, each differing by at least 25 %.

Melnyk *et al.* (2018) proposed a modification to the standardized method of measuring airflow resistivity. They suggested using the previous linear approximation method (PLA) method to improve the accuracy of the technique for measuring airflow resistance under static airflow conditions. Moreover, they analyzed an inverse method for determining airflow resistivity based on fitting sound absorption coefficients.

Similarly, Sebaa *et al.* (2005) proposed a method for determining the airflow resistivity of porous materials by analyzing the reflection of a plane wave from the porous material. The described method involves fitting the measured impulse reflected from the tested material to an impulse calculated with a time-domain model. The model, developed by Johnson *et al.* (1987), integrates porosity and airflow resistivity and was used in the computations. Sensitivity analysis showed that sound reflection is most sensitive to airflow resistivity, while the influence of porosity is minimal. Jeong (2020) also presented a parallel technique for estimating airflow resistivity. However, his method was based on fitting the measurements of the sound absorption coefficient acquired in a reverberation chamber. Currently, machine learning (ML) techniques are widely used in the inverse method to characterize porous acoustic materials (Müller-Giebeler *et al.*, 2024; Zea *et al.*, 2023).

This paper aims to compare two methods used to estimate airflow resistivity. The first one is a modified standardized method that calculates static airflow through a porous material. The second method is an inverse method, based on fitting the sound absorption coefficient calculated from the airflow resistivity to the values measured in an impedance tube. The paper also

includes an extension of the inverse method for determining airflow resistivity in thin upholstery materials. The obtained results allow for an analysis of the accuracy of the investigated methods for determining airflow resistivity across different types of materials.

2. Methods for determining airflow resistivity

2.1. Standardized method – static airflow method

The standardized method for testing the airflow resistivity of porous materials, as outlined in (ASTM, 2022; ISO, 2018), is based on static airflow. This method involves the control of the airflow through the sample under examination while simultaneously measuring the pressure before and after the sample. The measurement should be conducted with airflow velocities ranging from $5 \cdot 10^{-4}$ m/s to $4 \cdot 10^{-3}$ m/s (ASTM, 2022) or 50 mm/s (ISO, 2018). By recording the sound pressure drop Δp and the volumetric airflow rate q_v, it is possible to calculate the airflow resistance R:

$$R = \frac{\Delta p}{q_v}. \tag{2}$$

Then, the specific airflow resistance (R_s) is determined, a parameter independent of the area of the sample:

$$R_s = R \cdot A = \frac{\Delta p}{q_v} \cdot A = \frac{\Delta p}{Au} \cdot A = \frac{\Delta p}{u}, \tag{3}$$

where $u = q_v/A$ is the airflow velocity, and A is the area of the sample perpendicular to the airflow.

Finally, a parameter independent of the thickness of the sample – the airflow resistivity r_s is determined:

$$r_s = \frac{R_s}{D}, \tag{4}$$

where D is the thickness of the material.

The standard method for measuring the pressure drop involves the use of the smallest possible value of airflow velocity, $u = 5 \cdot 10^{-4}$ m/s. An alternative approach involves a gradual reduction of the airflow velocity. In this approach, the relationship between airflow resistance and airflow velocity is determined for each sample using linear regression $R_s(u)$, and the final value of R_s is taken at $5 \cdot 10^{-4}$ m/s.

2.2. Modification of the standardized method

Melnyk *et al.* (2018) conducted a study that proposed modifications to the airflow resistivity measurement method described in ISO (2018). Their research demonstrated that at low airflow velocities u, there is a significant nonlinearity in the relationship between the airflow resistance R_s and airflow velocity u (see Fig. 1). As a result, applying linear regression for extrapolation at low airflow velocities can lead to substantial errors. To circumvent this issue, a modification to the static airflow method was proposed, called the PLA iteration method. This method involves first the determination of a polynomial that describes the relationship between airflow velocity and pressure drop $q(\Delta p)$, followed by the transformation of this polynomial into a linear approximation of airflow resistance as a function of airflow velocity. The process commences at the highest airflow velocity, and in each iteration, the range is extended to include lower airflow velocity values. The error between the measured results and the linear regression model is calculated at each iteration, and iterations continue until the error does not exceed a predefined value. Finally, the determined polynomial is extrapolated to obtain the airflow resistance value at $u = 5 \cdot 10^{-4}$ m/s.

Figure 1 illustrates the relationship between airflow resistance and the linear velocity of a porous material, as determined in the experiment. The graph also

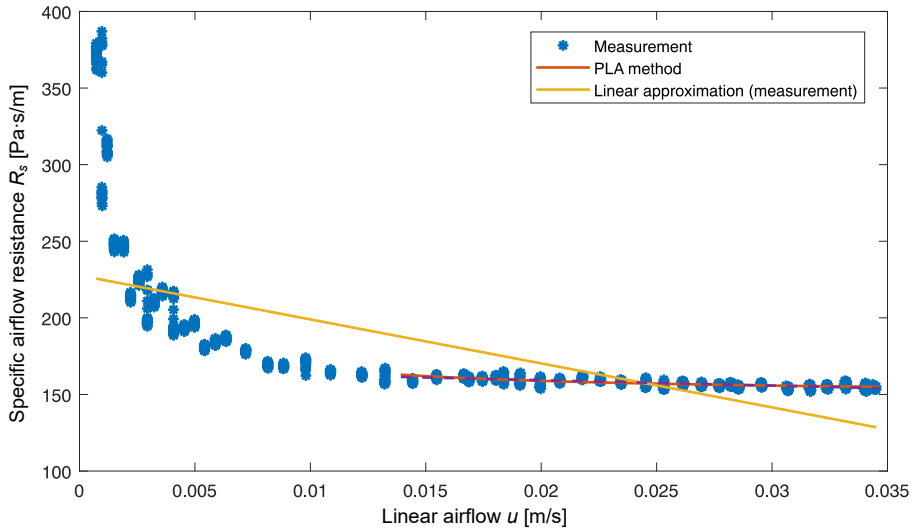

Fig. 1. Airflow resistance R_s of a porous material as determined in the measurement and a comparison of the regression line determined for the entire measurement range with the proposed PLA interpolation method.

presents regression lines calculated for the entire set of measurement data (yellow line), as well as for the data sets restricted by the proposed PLA method (red line).

The proposed PLA interpolation method, improves the alignment of the regression line with the measurement data in comparison to the linear approximation over the entire range of linear velocities u. This method effectively identifies the range of linear relationship between airflow resistance and airflow velocity by eliminating outliers. Consequently, it enhances the accuracy of airflow resistance determinations.

To automate the process of measuring airflow resistance, the laboratory setup (Fig. 2) uses sensors that are compatible with data acquisition cards. The measurement procedure involves the following steps: first, the test sample is mounted inside a cylinder. Air is then passed through the sample while simultaneously recording both the airflow rate and the pressure difference across the sample. For each test sample, a graph is created to illustrate the relationship between pressure and airflow. Finally, the airflow resistance at a linear velocity $u = 5 \cdot 10^{-4}$ m/s is determined using the PLA method.

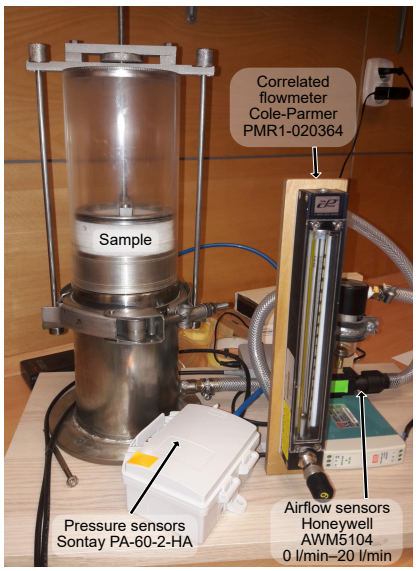

Fig. 2. Laboratory stand for determining the airflow resistivity of porous materials.

2.3. Inverse method for determining the airflow resistivity of porous materials

The proposed inverse method involves determining the flow resistivity through fitting the theoretical sound absorption coefficient to the measured value. To estimate the theoretical sound absorption coefficient, the Miki empirical model (Miki, 1990) was employed, which is a modification of the Delany–Bazley model (Delany, Bazley, 1970). The empirical model is based on experimental results and is a simplified description of the complex acoustic phenomena that occur in porous materials. It describes the relationship between key acoustic parameters of porous materials, such as characteristic impedance and propagation constant, and the physical parameter of airflow resistivity. The accuracy of the model is limited by the range of experimental data from which it was developed.

This model requires only a single parameter: the flow resistivity of the porous material. The model assumes that the solid phase is perfectly rigid and only considers the motion of the fluid. It is applicable to fibrous porous media with a porosity close to unity and provides the best fit to the experimental data in the range $0.01 < (f/r_s) < 1$. These models enable the prediction of the acoustic properties of porous materials, including specific impedance (z_c) and propagation constant (k_c).

In order to solve the inverse problem, it is necessary to find the minimum of the cost function $U(r_s)$, which is defined as follows:

$$U(r_s) = \sqrt{\frac{\sum_{i=1}^{n} \left(\alpha_t(f_i, r_s) - a_m(f_i)\right)^2}{N}}, \tag{5}$$

where $\alpha_t(f_i, r_s)$ is the predicted sound absorption coefficient for the i-th frequency band and for a given airflow resistivity r_s; $a_m(f_i)$ is the experimental sound absorption coefficient for the i-th frequency band, and N is the number of frequency bands. The bisection method is used to find the minimum of the cost function $U(r_s)$. By subdividing the initial interval into ten subintervals, calculations are accelerated, leading to faster results. The algorithm is fully detailed in the comprehensive study by Melnyk *et al.* (2018).

According to Miki's model, the propagation of sound in an isotropic, homogenous material is determined by two complex quantities:

– the characteristic impedance:

$$z_c(f) = \rho_0 c_0 \left[R(f) + jX(f)\right], \tag{6}$$

– the propagation constant (wavenumber) $k_c(f)$:

$$k_c(f) = \alpha + j\beta. \tag{7}$$

Based on the airflow resistivity r_s for a given material, the characteristic quantities can be determined using:

$$R(f) = \rho_0 c_0 \left[1 + 0.070 \left(\frac{f}{r_s}\right)^{-0.632}\right], \tag{8}$$

$$X(f) = \rho_0 c_0 \left[-0.107 \left(\frac{f}{r_s}\right)^{-0.632}\right], \tag{9}$$

$$\alpha = \frac{\omega}{c_0} \left[1 + 0.109 \left(\frac{f}{r_s}\right)^{-0.618}\right], \tag{10}$$

$$\beta = \frac{\omega}{c_0} \left[0.160 \left(\frac{f}{r_s}\right)^{-0.618}\right], \tag{11}$$

where $\omega = 2\pi f$, f is the frequency, ρ_0 is the air density, c_0 is the sound speed in air, and r_s is the airflow resistivity. Equations (8)–(11) were derived from regression models fitted to the relationship between the real parts $R(f)$ and $a(f)$, and the imaginary parts $X(f)$ and $b(f)$, of the acoustic impedance and propagation constant, respectively, with respect to the normalised flow resistance – expressed as (f/r_s) (Miki, 1990). The values for flow resistivity, acoustic impedance, and propagation constant were determined from measurements conducted on different materials (Delany, Bazley, 1970).

The surface impedance is then determined by

$$z_w = -i \cdot z_c \cdot \cot(kc \cdot D), \tag{12}$$

where D is the thickness of the material.

The formula to calculate the sound reflection coefficient R is

$$R = \frac{\frac{z_w}{\rho_0 c_0}\cos(\theta) - 1}{\frac{z_w}{\rho_0 c_0}\cos(\theta) + 1}, \tag{13}$$

where θ is the incidence angle, and for normal incidence $\theta = 0$.

The formula for the sound absorption coefficient $a_{t,i}$ is

$$\alpha_{t,i} = 1 - \left|R^2\right|. \tag{14}$$

2.4. Inverse method for determining the airflow resistivity of covering materials

The inverse method has also been used to calculate the airflow resistivity of covering materials. Similar to porous materials, a numerical model is required to find the sound absorption coefficient. For porous materials, it was assumed that the sound absorption coefficient can be determined for a model of the covering material placed on an air gap with a thickness of D. Therefore, the model describing the acoustic impedance of the surface of the material z_w, placed at a distance, was used to determine the sound absorption coefficient:

$$z_w = \frac{-j z_{s0} z_c \cot(k_c L) + z_c^2}{z_{s0} - j z_c \cot(k_c L)}, \tag{15}$$

where $z_{s0} = -j z_0 \cot(k_0 L)$ is the surface impedance at the top of the air layer of thickness L behind the material, $z_0 = \rho_0 c_0$ is the acoustic impedance of air, k_0 is the wave number in air, and z_c, k_c are the characteristic impedance and wave number of the covering material, respectively. Then, using Eqs. (13) and (14), the reflection coefficient R and sound absorption coefficient $a_{t,i}$ were calculated. The values of airflow resistivity were determined by the inverse method by finding the minimum of the cost function.

2.5. Sensitivity analysis of the porous material models

A sensitivity analysis was conducted for both the porous material model and the covering model where the covering material is mounted with an air gap behind, to investigate their applicability for the inverse method. The sensitivity of the models to changes in airflow resistivity was investigated. To evaluate the sensitivity of the models, the sensitivity index S_i was determined using the following relationship (Saltelli *et al.*, 2004):

$$S_i = \frac{\partial \alpha_i}{\partial r_s}\frac{r_s}{\alpha_i}, \tag{16}$$

where the differential $\frac{\partial \alpha_i}{\partial r_s}$ was calculated numerically for the sound absorption coefficient a_i at the i-th frequency for a given airflow resistivity r_s. The index provides a non-dimensional measure of sensitivity, showing how much the sound absorption coefficient is affected by a unit change in airflow resistivity. A higher S_i value indicates greater sensitivity, meaning that small variations in airflow resistivity lead to significant changes in the sound absorption properties of the material.

The analysis of porous materials was performed for materials with:

- low airflow resistivity ($r_s = 5000\,\mathrm{Pa\cdot s/m^2}$),
- medium airflow resistivity ($r_s = 15\,000\,\mathrm{Pa\cdot s/m^2}$),
- high airflow resistivity ($r_s = 50\,000\,\mathrm{Pa\cdot s/m^2}$)

for two material thicknesses: 15 mm and 30 mm (Figs. 3a and 3b, respectively). For the covering material (thickness $D = 2.5$ mm), the analyses were performed with no distance behind the material and with a distance $L = 100$ mm behind the material (Figs. 3c and 3d, respectively).

The obtained results show the frequency ranges where the sound absorption coefficient is the most sensitive to changes in airflow resistivity. These results were used to formulate recommendations for the inverse method for determining airflow resistivity.

The analysis shows that the sensitivity index S_i of the sound absorption coefficient for porous materials changes with frequency and depends on the airflow resistivity r_s and the material thickness D (Figs. 3a, 3b). Sensitivity analysis has revealed that for materials with low airflow resistivity, the greatest sensitivity to changes in airflow resistivity occurs in the frequency range where the sound absorption coefficient increases from 0 to its maximum value. In contrast, for materials with high airflow resistivity, the range of greatest sensitivity shifts to the range where the absorption coefficient reaches its maximum. The value of airflow resistivity also determines the maximum value of the sound absorption coefficient and the frequency at which the maximum occurs. As a result, the value and frequency of the maximum absorption coefficient serve as key indicator for adjusting the sound absorption characteristics.

On the other hand, high negative values of the sensitivity index S_i can be observed in the low-frequency

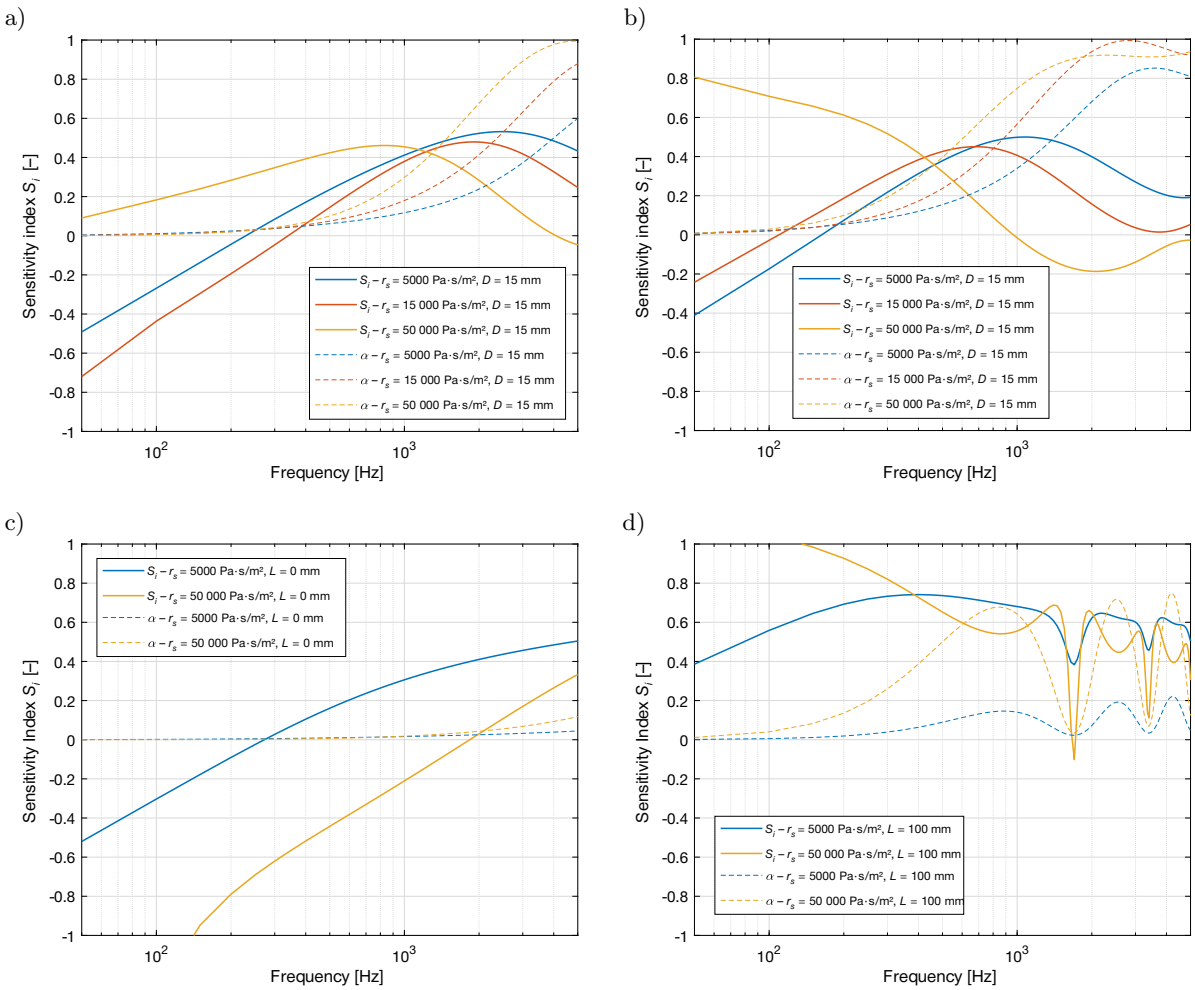

Fig. 3. Sensitivity indexes S_i of the sound absorption coefficient model to changes in airflow resistivity calculated for porous materials and covering materials across different airflow resistivity ranges. The results of calculations for porous material thicknesses: a) D = 15 mm, b) D = 30 mm; and for the covering material with thickness D = 2.5 mm: c) with no gap behind the material (L = 0 mm), d) with a gap (L = 100 mm).

range, indicating relatively large changes in the absorption coefficient. However, the absolute change in the absorption coefficient is small. Therefore, in this frequency range, the effect of changes in airflow resistivity is minimal, making it less significant when fitting the absorption characteristics to the measured values. To increase the accuracy of the inverse airflow resistivity determination, it is advantageous to fit the results over a frequency range that includes both the range in which the absorption coefficient increases and the region of the local maximum of the absorption coefficient.

In the case of thin materials and coverings that are directly backed by a rigid surface, the frequency at which the maximum occurs may be outside the measurement range (Figs. 3a, 3c). It is therefore advantageous to fit the sound absorption coefficient of a material sample placed over an air gap. The presence of an air gap behind the material has been shown to result in a shift of the sound absorption coefficient's maximums to lower frequencies (Fig. 3d). This effect serves to enhance the frequency range in which sensitivity is high, thereby facilitating improved matching across a broader frequency range.

3. Comparison of the methods

This section is concerned with a comparison of the two methods for determining airflow resistivity. The research was performed on seven different kinds of thick materials, which are primarily used as acoustic panel fillings, and two thin materials – polyester felts, which are mostly used as acoustic panel coverings or furniture upholstering (Fig. 4).

3.1. Covering materials

The airflow resistivity of thin materials was determined by employing both of the previously described methods. Measurements were taken for two different polyester felts, each with fibers of a different diameter and, presumably, different airflow resistivities. The

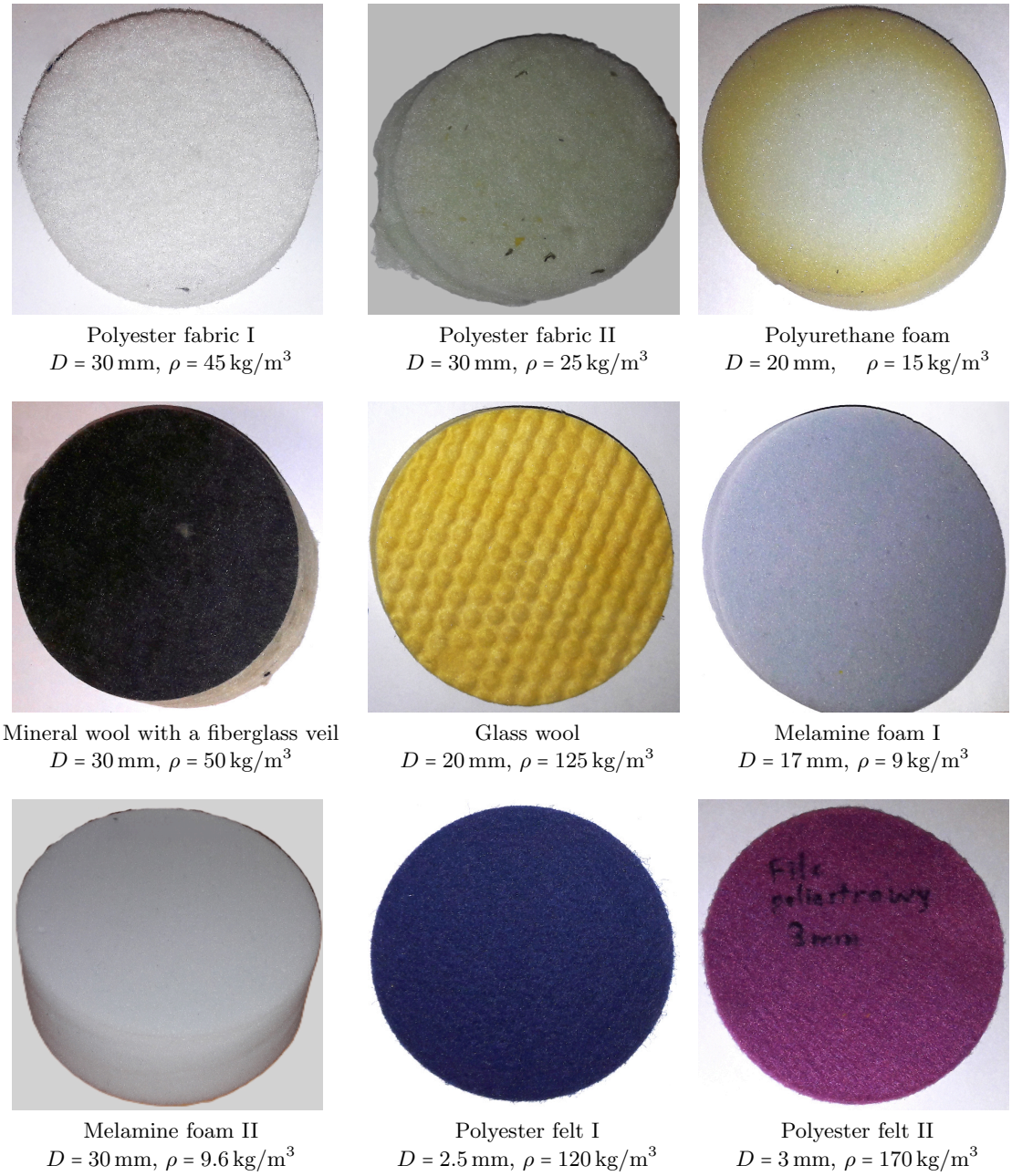

Fig. 4. Description of the measurement samples.

sound absorption coefficient was measured by placing the felts at the back of the impedance tube, with a distance ring to maintain the desired air gap L. The measurements of the sound absorption coefficient were performed for 10 different distances within the range L = 10 mm–200 mm (Fig. 5).

The matching was performed separately for three frequency ranges: f = 50 Hz–1600 Hz, which corresponds to measurements in the impedance tube of a large diameter (ø = 100 mm), f = 1600 Hz–6400 Hz (small impedance tube, ø = 29 mm), and for the wide frequency range f = 100 Hz–6400 Hz.

As the results show, increasing the air gap behind the sample L produces more local maxima associated with quarter-wavelength resonances (Fig. 5). In the inverse method, these maxima represent important points for obtaining more accurate matching. The choice of the frequency range for which the matching was performed is also important. The results from the large tube (Fig. 5a) include both the range in which the sound absorption coefficient increases and, depending on the distance L, also the local maxima. In contrast, the results from the small tube contain mainly local maxima, not always including the range in which the sound absorption coefficient increase starts from a minimum (Fig. 5b). Choosing a wide frequency range ensures that both the information about the sound absorption coefficient increase and the local maxima are

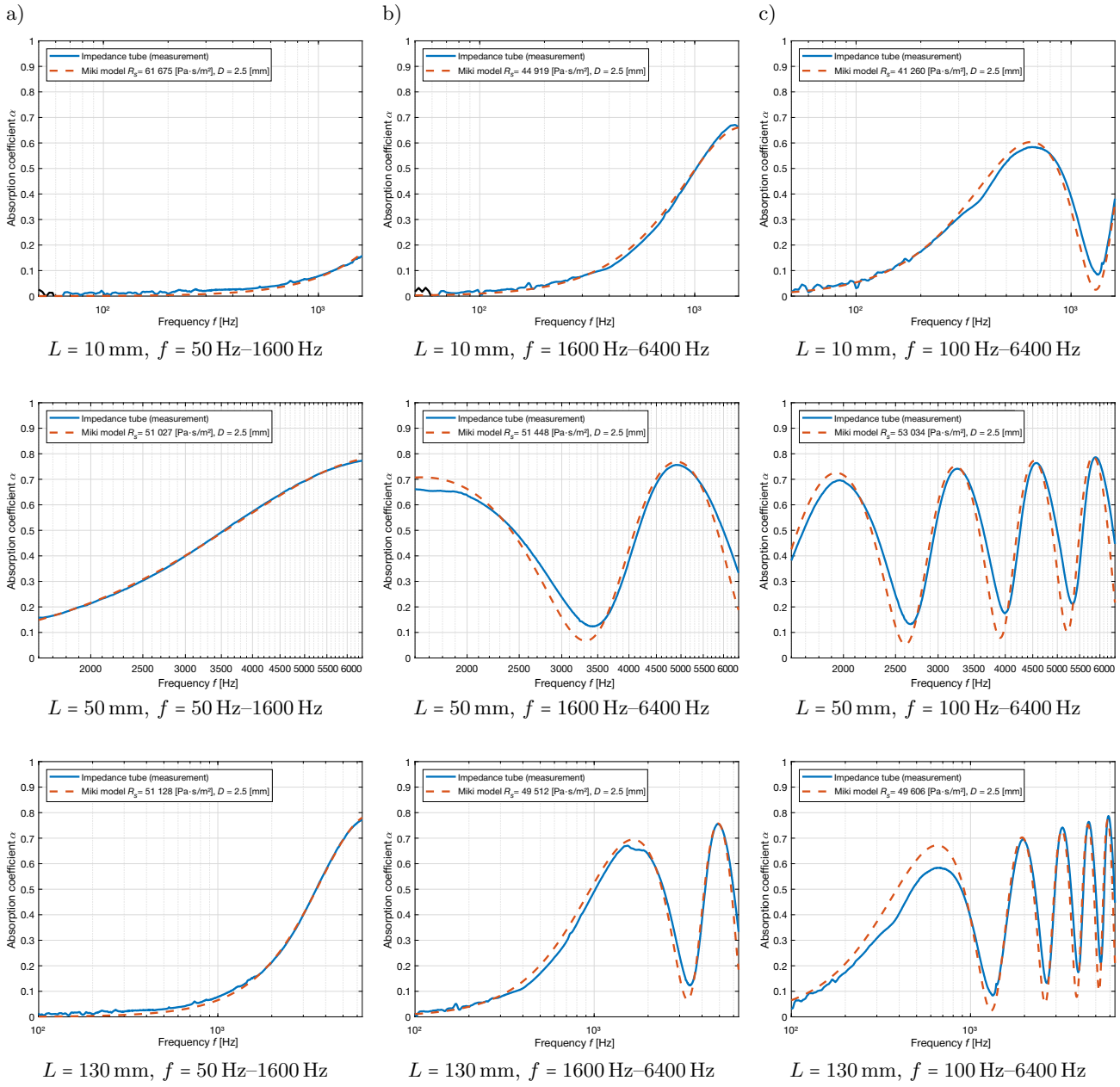

Fig. 5. Sound absorption coefficient of 2.5 mm-thick felt, where L is the air gap behind the specimen and f is frequency range: a) for large (ø = 100 mm) diameter tube, b) small (ø = 29 mm) diameter tube, c) for wide frequency range – combined results from both tubes.

acquired (Fig. 5c). As a result, the choice of the air gap behind the sample L and the frequency range translate into the value of airflow resistivity determined by matching the sound absorption coefficient.

The influence of the air layer L behind the material on the calculation of airflow resistivity using the inverse method is shown in Fig. 6.

The analysis showed that the 3 mm-thick felt with thinner fibers has significantly higher airflow resistivity. For the large tube, the results are mainly lower than those from the small tube. The results of the calculations performed for the wide frequency range are the average of the airflow resistivity obtained from both large and small tubes (Fig. 6). The obtained results of airflow resistivity differ most for air gaps of 10 mm and 50 mm for both felts. For the remaining air gaps, the results did not change significantly with the change of L. This confirms that the air gap improves the repeatability of the inverse method for determining airflow resistivity.

Table 1 compares the airflow resistivity values obtained with both methods. The results for the inverse method are the average values calculated for all the air gaps. A comparison of the determined airflow resistance values for both coverings reveals that the 2.5 mm-thick felt obtained values that are approx-

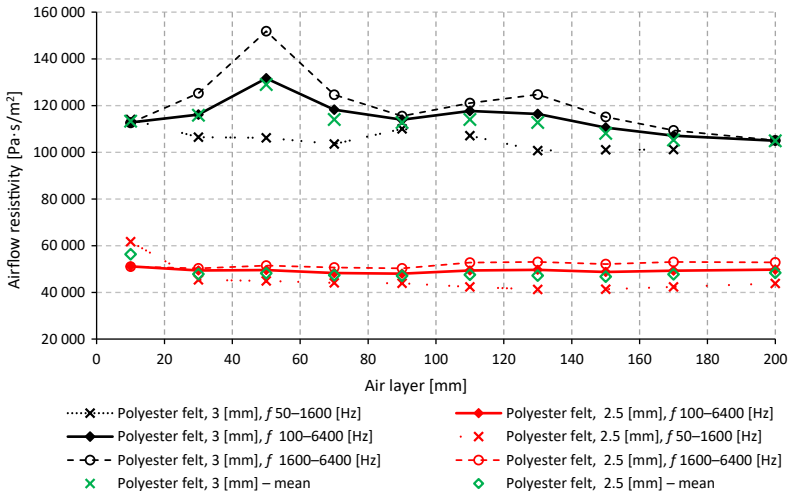

Fig. 6. Airflow resistivity of the porous materials, depending on the air gap behind the measurement sample and the frequency range (impedance tube measurement).

Table 1. Airflow resistivity of the covering under study.

No.	Material	Density ρ [kg/m^3]	Thickness D [mm]	r_s [Pa·s/m^2] (Inverse method)	r_s [Pa·s/m^2] (PLA – static airflow)
1	Polyester felt I	120	2.5	49 309	36 864
2	Polyester felt II	170	3	114 913	142 189

imately 25 % lower when measured using the PLA algorithm, while the 3 mm felt obtained values are 23 % higher for this method.

The comparison of sound absorption coefficients calculated using the airflow resistivities determined for both presented methods are shown in Fig. 7. The results of the sound absorption coefficients calculated and measured for the 3 mm-thick felt with a 70 mm air gap are in a good agreement for medium and high frequencies. Larger differences (though still not exceeding 0.1) can be observed at frequencies below 500 Hz. This means that the model is not very sensitive to changes in airflow resistivity and the value of sound absorption coefficient can be determined with reasonable accuracy.

3.2. Porous materials for filling acoustics panels

As demonstrated in Table 2, the values of airflow resistivities obtained by the inverse method and the

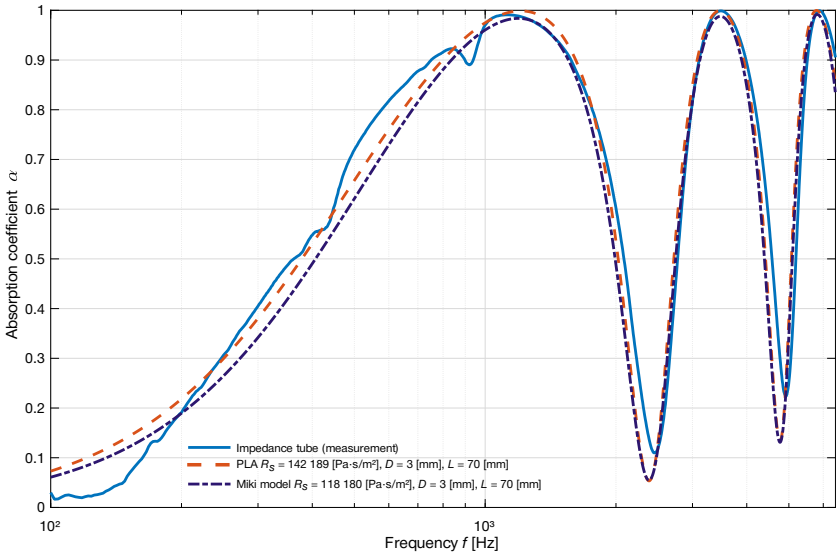

Fig. 7. Sound absorption coefficients measured and calculated based on the airflow resistivity determined with the inverse method and airflow resistivity measurement (PLA algorithm) for a 3 mm-thick polyester felt mounted at a 70 mm air gap.

Table 2. Airflow resistivity of materials used for filling the acoustic panels.

No.	Material	Density ρ [kg/m^3]	Thickness D [mm]	R^2	r_s [Pa·s/m^2] (Inverse method)	r_s [Pa·s/m^2] (PLA – static airflow)	Relative error d [%]
1	Polyester fabric I	45	30	0.993	2436	2435	0.04
2	Polyester fabric II	25	30	0.973	5024	4840	3.80
3	Mineral wool with fiberglass veil	50	30	0.983	16 474	18 603	11.44
4	Glass wool	125	20	0.960	102 893	126 154	18.44
5	Polyurethane foam (CME = CV)	15	20	0.965	5099	5040	1.17
6	Melamine foam I	9	17	0.995	6941	7922	12.38
7	Melamine foam II	9.6	30	0.988	8701	8652	0.57

static airflow measurement method (PLA algorithm) for a wide frequency range are shown. In order to evaluate the results, the relative error between the values obtained from the two methods was determined. This was done by using the following equation:

$$\delta = \frac{|r_{s1} - r_{s2}|}{r_{s1}}, \tag{17}$$

where r_{s1} is the airflow resistivity from the inverse method and r_{s2} is the value obtained from modified static airflow measurements. It is acknowledged that the true value of airflow resistivity is unknown, so the relative error serves a comparison between the two different measurement techniques. The analysis of airflow resistivity shows that for materials with low airflow resistivity (up to around 10 000 Pa·s/m^2), the dif-

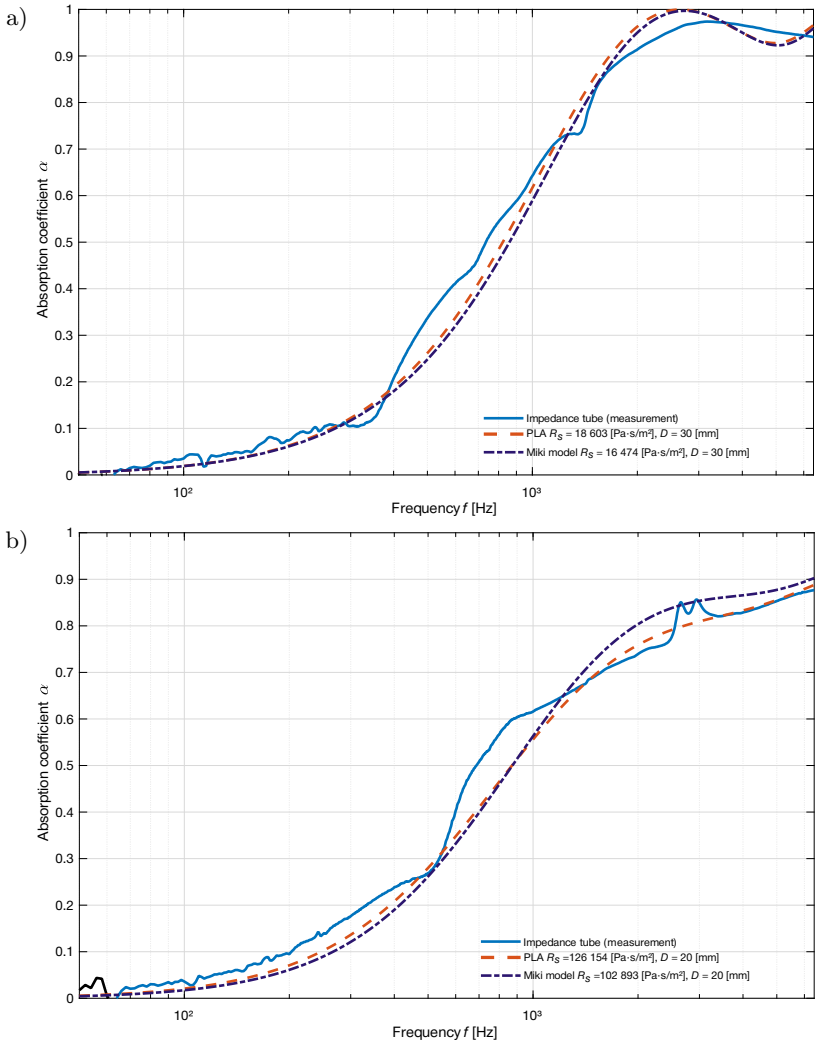

Fig. 8. Sound absorption coefficient: measured and calculated based on the airflow resistivity determined with the inverse method and with the airflow resistivity measurement method (PLA algorithm) for: a) mineral wool, b) glass wool.

ferences in the determined values of airflow resistivity do not exceed 12 %. However, as the airflow resistivity values increase, the discrepancy between the values obtained by the two methods also increases, reaching up to 18 % for glass wool.

However, the comparison of the sound absorption coefficient values determined and measured for mineral wool and glass wool shows good agreement (Fig. 8).

The largest differences between the determined values do not exceed 0.1 and can be observed at high frequencies. Mineral wool demonstrates a better match compared to glass wool, particularly in the range where the absorption coefficient is increasing (Fig. 8a). To improve the fitting of the curves, a different model specifically developed for glass wools could be used (Fig. 8b). Similarly to the covering materials, the calculation model is not very sensitive to changes in airflow resistivity, and even for large changes in airflow resistivity, the values of the sound absorption coefficient can be determined with good accuracy.

4. Discussion

The findings of the research indicate that both methodologies employed for the estimation of the airflow resistivity of porous materials, with densities ranging from 9.6 kg/m^3 to 45 kg/m^3, facilitate the precise calculation of the sound absorption coefficient. However, a higher discrepancy is observed in the estimation of airflow resistivity for materials characterized by high density and high airflow resistivity (Table 2).

The discrepancy between the methods may be attributed to the selection of the computational model for the inverse method. Empirical models are fitted to specific data sets, thus constraining their ability to predict the behavior of materials with significantly different properties or under conditions significantly different from those under which measurements were made (Komatsu, 2008). Consequently, these models may be less accurate in predicting acoustic properties across a broader range of material parameters, such as density or flow resistance. The accuracy of the inverse method is contingent on the execution of the fitting procedure within the applicable range of the relevant approximations (Bonfiglio, Pompoli, 2013). Conversely, the outcomes derived from the inverse method are also directly influenced by the quality of the experimental data, such as the sound absorption coefficient. Errors in the input data can propagate to the inversion results (Pelegrinis *et al.*, 2016).

The complexity and imprecision inherent in the determination of the physical parameters of porous materials constitute a substantial challenge, especially for more complicated models (Bonfiglio, Pompoli, 2013). Consequently, the present study opted for a more straightforward model that necessitates only flow resistivity for the inversion method.

The flow resistivity measurement method under discussion is subject to factors that can influence measurement error, with one potential source of error being the leakage of air through the side of the material sample mounted in the holder during the airflow method. Other issues arise from non-linearities in the relationship between airflow resistance R_s and airflow velocity u (Melnyk *et al.*, 2018).

However, it is worthwhile to analyze the significance of the observed discrepancies and their impact on the prediction of sound absorption coefficients. The findings for both upholstery and thicker porous materials demonstrate that the discrepancies in the determined airflow resistivity values for low-density materials (up to $\rho < 50\,\mathrm{kg/m^3}$) and low-resistivity airflows (up to $r_s < 30\,000\,\mathrm{Pa\cdot s/m^2}$) do not exceed 15 % between the methods. For materials with higher densities and flow resistivities, the differences can reach up to 18 %. Nevertheless, a comparison of sound absorption coefficients calculated from airflow resistivity values obtained by both methods with values measured in the impedance tube revealed that these differences do not significantly affect the sound absorption coefficient values. According to these results, the process of determining airflow resistivity can be simplified by using the inverse method for measurement samples with a diameter of 100 mm only. The study also investigated the impact of the method used to mount thin and covering materials on the accuracy of determining airflow resistivity using the inverse method. It was found that mounting the material with an air gap is necessary for obtaining accurate results, and that for achieving repeatable results, a minimum distance of 70 mm is required.

5. Conclusions

In the present study, two methods for determining the airflow resistivity of porous materials were compared and validated. The first method was a modified version of the standardized method based on static airflow, as proposed by Melnyk *et al.* (2018). In this method, a linear approximation was used to improve the fit between the measured airflow resistivity and sound absorption coefficient results. The second method, known as the inverse method, involved matching the theoretical sound absorption coefficient with the impedance tube measurement results. The primary objective of the research was to evaluate the accuracy of airflow resistivity measurement for certain materials and to assess the effect of this parameter on the agreement between the predicted sound absorption coefficient and the impedance tube measurement results.

The study was carried out on different types of porous materials with thicknesses ranging from 2.5 mm to 30 mm and densities from 9 kg/m^3 to 170 kg/m^3.

Thin covering materials, used for upholstery, were investigated as well as thicker porous materials typically used in acoustic panels or as furniture infill. The selection of materials for testing was based on their different airflow resistivities.

The results of the research suggest that for porous materials up to approximately 30-mm thick, variations in measured airflow resistivity values do not have a significant effect on the sound absorption coefficient. Consequently, both methods can be used to determine the airflow resistivity required to calculate the sound absorption coefficient of porous materials and layered structures, including upholstery, without the need for repeated measurements of specific configurations. This will greatly speed up the process of selecting materials and upholstery for specific acoustic purposes.

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

Mykhaylo Melnyk – conceptualization, development of a PLA method, measurement and data analysis, writing the manuscript; Jarosław Rubacha – conceptualization, development of a invers method, data analysis, writing the manuscript; Artur Flach – measurement and data analysis, verification of measurement, manuscript review; Aleksandra Chojak – measurement and data analysis, writing the manuscript; Tadeusz Kamisiński – supervision, results analysis, manuscript review; Wojciech Zabierowski – supervision, manuscript review; Marek Iwaniec – supervision, manuscript review; Andriy Kernytskyy – manuscript review.

References

1. Allard J.F., Atalla N. (2009), *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials*, Wiley, https://doi.org/10.1002/9780470747339.
2. Allard J.-F., Champoux Y. (1992), New empirical equations for sound propagation in rigid frame fibrous materials, *The Journal of the Acoustical Society of America*, **91**(6): 3346–3353, https://doi.org/10.1121/1.402824.
3. ASTM C522-03 (2022), *Standard test method for airflow resistance of acoustical materials*, ATM International.
4. Bonfiglio P., Pompoli F. (2013), Inversion problems for determining physical parameters of porous materials: Overview and comparison between different methods, *Acta Acustica United with Acustica*, **99**(3): 341–351, https://doi.org/10.3813/AAA.918616.
5. Cao L., Fu Q., Si Y., Ding B., Yu J. (2018), Porous materials for sound absorption, *Composites Communications*, **10**: 25–35, https://doi.org/10.1016/j.coco.2018.05.001.
6. Cox T., D'Antonio P. (2016), *Acoustic Absorbers and Diffusers. Theory, Design and Application*, 3rd ed., CRC Press, https://doi.org/10.1201/9781315369211.
7. Crocker M.J. [Ed.] (2007), *Handbook of Noise and Vibration Control*, Wiley, https://doi.org/10.1002/9780470209707.
8. Cuenca J., Göransson P., De Ryck L., Lähivaara T. (2022), Deterministic and statistical methods for the characterisation of poroelastic media from multi-observation sound absorption measurements, *Mechanical Systems and Signal Processing*, **163**: 108186, https://doi.org/10.1016/j.ymssp.2021.108186.
9. Delany M.E., Bazley E.N. (1970), Acoustical properties of fibrous absorbent materials, *Applied Acoustics*, **3**(2): 105–116, https://doi.org/10.1016/0003-682X(70)90031-9.
10. Dell A., Krynkin A., Horoshenkov K.V. (2021), The use of the transfer matrix method to predict the effective fluid properties of acoustical systems, *Applied Acoustics*, **182**: 108259, https://doi.org/10.1016/j.apacoust.2021.108259.
11. Doutres O., Salissou Y., Atalla N., Panneton R. (2010), Evaluation of the acoustic and non-acoustic properties of sound absorbing materials using a three-microphone impedance tube, *Applied Acoustics*, **71**(6): 506–509, https://doi.org/10.1016/j.apacoust.2010.01.007.
12. Gibson L.J., Ashby M.F. (1997), *Cellular Solids: Structure and Properties*, 2nd ed., Cambridge University Press, https://doi.org/10.1017/CBO9781139878326.
13. Herrero-Durá I., Cebrecos Ruíz A., García-Raffi L.M., Romero-García V. (2019), Matrix formulation in acoustics: The transfer matrix method [in Spanish: Formulación matricial en Acústica: El método de la matriz de transferencia], *Modelling in Science Education and Learning*, **12**(2): 153, https://doi.org/10.4995/msel.2019.12148.
14. Hou X., Du S., Liu Z., Guo J., Li Z. (2017), A transfer matrix approach for structural – Acoustic correspondence analysis of diesel particulate filter, *Advances in Mechanical Engineering*, **9**(9), https://doi.org/10.1177/1687814017722495.
15. Huang S., Li Y., Zhu J., Tsai D.P. (2023), Sound-absorbing materials, *Physical Review Applied*, **20**(1): 010501, https://doi.org/10.1103/PhysRevApplied.20.010501.

16. International Organization for Standardization (1996), *Acoustics – Determination of sound absorption coefficient and impedance in impedance tubes. Part 1: Method using standing wave ratio* (ISO Standard No. 10534-1:1996), https://www.iso.org/standard/18603.html.

17. International Organization for Standardization (2003), *Acoustics – Measurement of sound absorption in a reverberation room* (ISO Standard No. 354:2003), https://www.iso.org/standard/34545.html.

18. International Organization for Standardization (2018), *Acoustics – Determination of airflow resistance. Part 1: Static airflow method* (ISO Standard No. 9053-1:2018), https://www.iso.org/standard/69869.html.

19. International Organization for Standardization (2023), *Acoustics – Determination of sound absorption coefficient and impedance in impedance tubes. Part 2: Two-microphone technique for normal sound absorption coefficient and normal surface impedance* (ISO Standard No. 10534-2:2023, https://www.iso.org/standard/81294.html.

20. Jeong C.-H. (2020), Flow resistivity estimation from practical absorption coefficients of fibrous absorbers, *Applied Acoustics*, **158**: 107014, https://doi.org/10.1016/j.apacoust.2019.107014.

21. Johnson D.L., Koplik J., Dashen R. (1987), Theory of dynamic permeability and tortuosity in fluid-saturated porous media, *Journal of Fluid Mechanics,* **176**(1): 379–402, https://doi.org/10.1017/S0022112087000727.

22. Kamisiński T., Brawata K., Pilch A., Rubacha J., Zastawnik M. (2012), Sound diffusers with fabric covering, *Archives of Acoustics*, **37**(3): 317–322, https://doi.org/10.2478/v10168-012-0040-5.

23. Komatsu T. (2008), Improvement of the Delany–Bazley and Miki models for fibrous sound-absorbing materials, *Acoustical Science and Technology*, **29**(2): 121–129, https://doi.org/10.1250/ast.29.121.

24. Lafarge D., Lemarinier P., Allard J.F., Tarnow V. (1997), Dynamic compressibility of air in porous structures at audible frequencies, *The Journal of the Acoustical Society of America*, **102**(4): 1995–2006, https://doi.org/10.1121/1.419690.

25. Melnyk M., Rubacha J., Kamisiński T., Majchrzak A. (2018), Application of MEMS sensors for the automation of a laboratory stand for the measurement of the flow resistance of porous materials, [in:] *2018 XIV-th International Conference on Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, pp. 28–34, https://doi.org/10.1109/MEMSTECH.2018.8365695.

26. Miki Y. (1990), Acoustical properties of porous materials. Modifications of Delany–Bazley models, *Journal of the Acoustical Society of Japan (E)*, **11**(1): 19–24, https://doi.org/10.1250/ast.11.19.

27. Müller-Giebeler M., Berzborn M., Vorländer M. (2024), Free-field method for inverse characterization of finite porous acoustic materials using feed forward neural networks, *The Journal of the Acoustical Society of America*, **155**(6): 3900–3914, https://doi.org/10.1121/10.0026239.

28. Oliva D., Hongisto V. (2013), Sound absorption of porous materials – Accuracy of prediction methods, *Applied Acoustics*, **74**(12): 1473–1479, https://doi.org/10.1016/j.apacoust.2013.06.004.

29. Pelegrinis M.T., Horoshenkov K.V., Burnett A. (2016), An application of Kozeny–Carman flow resistivity model to predict the acoustical properties of polyester fibre, *Applied Acoustics*, **101**: 1–4, https://doi.org/ 10.1016/j.apacoust.2015.07.019.

30. Rubacha J., Pilch A., Zastawnik M. (2012), Measurements of the sound absorption coefficient of auditorium seats for various geometries of the samples, *Archives of Acoustics*, **37**(4): 483–488, https://doi.org/10.2478/v10168-012-0060-1.

31. Saltelli A., Tarantola S., Campolongo F., Ratto M. (2004), *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, Wiley.

32. Sebaa N., Fellah Z.E.A., Fellah M., Lauriks W., Depollier C. (2005), Measuring flow resistivity of porous material via acoustic reflected waves, *Journal of Applied Physics*, **98**(8): 084901, https://doi.org/10.1063/1.2099510.

33. Tao Y., Ren M., Zhang H., Peijs T. (2021), Recent progress in acoustic materials and noise control strategies – A review, *Applied Materials Today*, **24**: 101141, https://doi.org/10.1016/j.apmt.2021.101141.

34. Vorländer M. (2008), *Auralization. Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-48830-9.

35. Zea E., Brandão E., Nolan M., Cuenca J., Andén J., Svensson U.P. (2023), Sound absorption estimation of finite porous samples with deep residual learning, *The Journal of the Acoustical Society of America*, **154**(4): 2321–2332, https://doi.org/10.1121/10.0021333.

36. Zhao X.-D., Yu Y.-J., Wu Y.-J. (2016), Improving low-frequency sound absorption of micro-perforated panel absorbers by using mechanical impedance plate combined with Helmholtz resonators, *Applied Acoustics*, **114**: 92–98, https://doi.org/10.1016/j.apacoust.2016.07.013.

Archives of Acoustics Vol. 50, No. 4, pp. 435–443 (2025), https://doi.org/10.24425/aoa.2025.156932

Research Paper

Acoustic Silencer for a Dedicated Frequency

Adam BRAŃSKI, Edyta PRĘDKA-MASŁYK*

Department Faculty of Electrical and Computer Engineering
Technical of Electrical and Computer Engineering Fundamentals, University of Rzeszow
Rzeszów, Poland

*Corresponding Author e-mail: edytap@prz.edu.pl

Received April 8, 2025; revised October 16, 2025; accepted October 20, 2025; published online November 17, 2025.

Acoustic resonators are useful for damping low frequencies. In cylindrical silencers (mufflers), the implementation of the resonance concept consists in selecting such a length of the expansion chamber (EC) that a wave of opposite phase is created in it, and with this opposite phase the incident wave is damped. Based on the plane wave theory (1D) and simple analytical calculations, it is possible to approximately determine the shortest length of the EC for a selected frequency; such a chamber represents the simplest silencer. Its efficiency is measured by the transmission loss (TL) value; increasing the TL value indicates that the silencer efficiency increases as well. The efficiency was improved in two ways: first, in single EC, by adding inlet, outlet, or both horizontal extensions, and second, by adding another EC. In the first case, the influence of the length of the horizontal extensions on TL was analyzed. In the second study, another dedicated EC was added, and the influence of the width and orifice diameter of the transverse partition on TL was analyzed. All analytical results were confirmed experimentally. The results indicate that, first of all, a simple silencer (single EC) is found to damp a dedicated frequency. In addition, simple changes in the structure of such a silencer significantly increase its efficiency.

Keywords: acoustic silencer; transmission loss coefficient (TL); expansion chamber (EC); transverse partition; horizontal inlet/outlet extensions to a single D-EC.

1. Introduction

Acoustic silencers are used in many areas of live, e.g., in the automotive industry, HVAC ducts, and firearms (Munjal, 1987; Nilsson *et al.*, 2021; Karami *et al.*, 2024). They are mainly dissipative silencers, which work on the phenomenon of successive reflection of sound waves and the conversion of their energy into heat.

General requirements for the design of silencers are described in many studies (e.g., Potente, 2005; Rahman *et al.*, 2005; Munjal, 2013; 2014; Jokandan *et al.*, 2023). The desirable properties of a silencer are, above all, simple construction, small size and sound attenuation over a wide frequency range. To meet the first two requirements the main challenge is to reduce the volume of the silencer's expansion chamber (EC), in practice its length.

Generally, the effectiveness of a silencer is measured, by, e.g., the transmission loss (TL) coefficient (Lee *et al.*, 2020). There are many analytical and numerical methods to calculate TL (at the silencer design stage), as well as experimental TL measurements on a real silencer. Among analytical methods, 1D (in simple structures), 2D (cylindrical wave), and 3D (three-dimensional wave) theories are used. Also, numerical methods such as FEM/BEM (Selamet, Radavich, 1997; Strek, 2010; Cui, Huang, 2012; Wei, Guo, 2016) and computational programs, e.g., SYSNOISE, COMSOL, and ANSYS (Swamy *et al.*, 2014), are widely used. In the aforementioned methods, only the problem of reflection is taken into account, while other aspects of sound propagation in silencers are omitted (Rahman *et al.*, 2005). Three experimental methods are also used, i.e., the 'traditional' laboratory method, the four-pole transfer ma-

trix method and the three-point method; they are compared in (Bilawchuk, Fyfe, 2003; Tao, Seybert, 2003; Zalte, Sature, n.d.).

The TL of a single circular EC can be increased through a variety of simple internal configuration. For example, the TL value was analyzed depending of the following parameters: EC length (Selamet, Radavich, 1997), surface absorption coefficient (Chiu, Chang, 2014), locations of horizontal partitions (Selamet *et al.*, 1998; Yu, Cheng, 2015), horizontal inlet/outlet extensions (Chaitanya, Munjal, 2011; Munjal, 2013; Rafique *et al.*, 2022), and also multi-chamber silencers with transverse partitions (Selamet *et al.*, 2003; Yu, Cheng, 2015; Yu *et al.*, 2015; Xiang *et al.*, 2016). In the mentioned studies, the influence of silencer structure on TL in a certain frequency range was considered.

The aim of this article is to demonstrate that it is possible to build a structurally simple silencer for a dedicated frequency, using of course conclusions from previous studies. This is important because, apart from starting and breaking, mechanical devices typically generate noise at an approximately constant frequency. Such a silencer should be therefore most effective at this dedicated frequency compared to other similar designs. Assuming that an objective function is TL, maximizing TL will indicate the optimal silencer for the dedicated frequency.

2. TL of the cylindrical EC

Due to the purpose of silencers, it is advisable to predict the maximum TL at the design stage. It turns out that the most important parameter is the geometry of the EC. For a given diameter of a cylindrical EC, the remaining task is to determine its length (Bilawchuk, Fyfe, 2003).

To define TL, we first define the sound power transmission coefficient (TC), $a_{\mathrm{tr}} = W_{\mathrm{out}}/W_{\mathrm{in}}$, where $W_{\mathrm{tr}} = W_{\mathrm{out}}$ is the outgoing (transmitted) acoustic power, and W_{in} is the incident (incoming) acoustic power. The TL is then expressed in terms of the TC (in dB) (Barron, 2003; Swamy *et al.*, 2014):

$$\mathrm{TL} = 10\log_{10}\left(W_{\mathrm{in}}/W_{\mathrm{out}}\right) = 10\log(1/a_{\mathrm{tr}}). \tag{1}$$

For a plane wave, at the inlet and outlet one has:

$$W_{\mathrm{in}} = \frac{p_{\mathrm{in}}^2}{2z_0}S_{\mathrm{in}}, \qquad W_{\mathrm{out}} = \frac{p_{\mathrm{out}}^2}{2z_0}S_{\mathrm{out}}, \tag{2}$$

where $z_0 = \rho c$ is the characteristic impedance, S is the surface area, p_{in} and p_{out} are the average (root mean square (RMS)) pressures at the inlet and outlet, respectively.

Hence,

$$\frac{1}{a_{\mathrm{tr}}} = \frac{W_{\mathrm{in}}}{W_{\mathrm{out}}} = \frac{p_{\mathrm{in}}^2}{p_{\mathrm{out}}^2}\frac{S_{\mathrm{in}}}{S_{\mathrm{out}}}. \tag{3}$$

Primary approach to sound transmission through the EC is the 1D theory (Selamet, Radavich, 1997; Barron, 2003; Tao, Seybert, 2003; Zhang *et al.*, 2020; Rafique *et al.*, 2022). After some calculations, the following useful equation is obtained:

$$\frac{1}{a_{\mathrm{tr}}} = \frac{1}{4}\frac{S_1}{S_3}\left\{\left(1+\frac{S_3}{S_1}\right)^2 + \left[\left(\frac{S_2}{S_1}+\frac{S_3}{S_2}\right)^2 - \left(1+\frac{S_3}{S_1}\right)^2\right]\sin^2(k_2\ell_2)\right\}, \tag{4}$$

where, see Fig. 1, $S_\nu = \pi r_\nu^2$, $\nu = 1, 2, 3$ are the cross-sectional areas of the inlet, EC, and outlet, and $u_{\nu,i}$ and $u_{\nu,e}$ denote the incident and reflected plane waves, respectively.

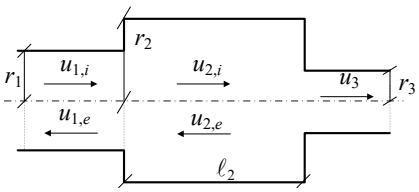

Fig. 1. Plane wave transmission through the EC.

Note that the TL, Eq. (1), reaches a maximum if $1/a_{\mathrm{tr}}$ is also a maximum. For this to happen, $\sin^2(k\ell_2)$ ought to be one. So:

$$k\ell_2 = \frac{\pi}{2} + n\pi, \qquad n = 0, 1, 2, \ldots \tag{5}$$

Hence,

$$\ell_2 = (1+2n)\frac{\lambda}{4}, \qquad n = 0, 1, 2, \ldots \tag{6}$$

The minimum chamber length ℓ_{min} is for $n = 0$:

$$\ell_{\mathrm{min}} = \lambda/4 = c/(4f). \tag{7}$$

In this way, the minimal length of the EC is obtained, for which the TL reaches its maximum values. However, note that the 1D theory is valid only up to the 'cut off' frequency (Potente, 2005).

In fact, sound transmission through a single EC is somewhat different from what the 1D theory suggests. As indicated in (Kang, Ji, 2008; Chaitanya, Munjal, 2011), the difference between 1D analysis and experimental, 3D, or numerical analyses is due to the presence of three-dimensional waves. Therefore, as pointed out in (Yu, Cheng, 2015), the 1D model can be used to approximately calculate the TL maxima, but only if the cross-section of the EC is sufficiently small.

3. Numerical calculations and experiments

The construction of a structurally simple silencer for a dedicated frequency was realized in the following steps:

1) Based on the 1D theory, the minimum length of the EC was found. Due to the inaccuracies of

this theory, this length was then experimentally corrected, so the length of the dedicated expansion chamber (D-EC) was obtained.

2) The TL was increased by adding horizontal inlet/outlet extensions to a single D-EC.
3) The TL was further increased by adding another D-EC, which was achieved by adding a transverse partition to the corresponding EC length:
   – the influence of the transverse partition widths was determined at a fixed orifice diameter,
   – the influence of the transverse partition orifice diameters was determined at a fixed width.

   All measurements below were performed using the Brüel & Kjær set, based on the four-pole matrix. They were conducted for frequencies $f = \{1, 2, 3, 4, 5, 6\} \times 10^3$ Hz, while results were presented at selected frequencies, i.e., $f = \{1, 3, 5\} \times 10^3$ Hz.

3.1. Attached length of the EC → D-EC

The influence of the single EC length $l_{\min}$, Eq. (7), on the TL was analyzed, where

$$\ell_{\min} = \{8.5, 2.83, 1.7\} \cdot 10^{-2}\,\text{m}.$$

Furthermore, the TL was calculated according to Eq. (1), using the following parameters: $r_1 = 0.003$ m, $r_2 = 0.018$ m, $r_1 = r_3$, hence $S_1 = S_3 = 2.827 \cdot 10^{-5}$ m^2, $S_2 = 1.0179 \cdot 10^{-3}$ m^2, $k_2 = k = (2\pi f)/c$, $\ell_{\min} = \lambda/4$. The results are presented in Fig. 2.

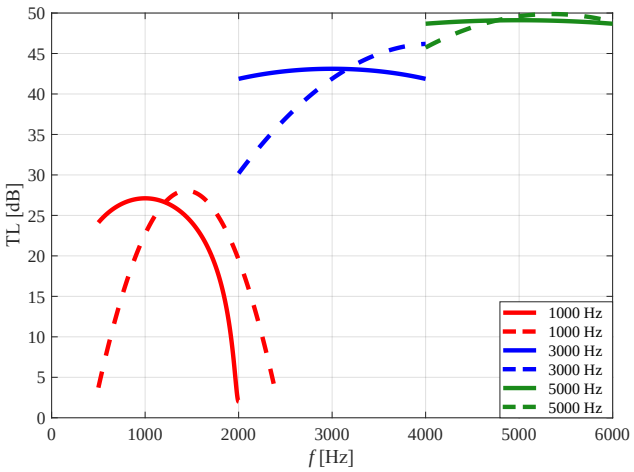

Fig. 2. TL for a single EC, solid line – calculated values Eq. (1); dashed line – measured values.

As can be seen in Fig. 2, the experimental results do not agree with the 1D theory, which predict the TL maximum occurs at the dedicated frequency. So, in order to account for the influence of three-dimensional wave effects, the length $\ell_{\min}$ ought to be increased by some length ℓ_a, so that the chamber length $\ell = \ell_{\min} + \ell_a$ corresponds exactly to a quarter-wave length; this adjusted length leads to the D-EC.

The attached length ℓ_a can be estimated based on numerical calculations, such as the finite element method (FEM) (Komkin *et al.*, 2012), or through theoretical considerations (Selamet, Radavich, 1997; Kang, Ji, 2008). In this study, ℓ_a was determined experimentally. For this purpose, the TL was measured as a function of frequency for different values of ℓ_a, Fig. 3.

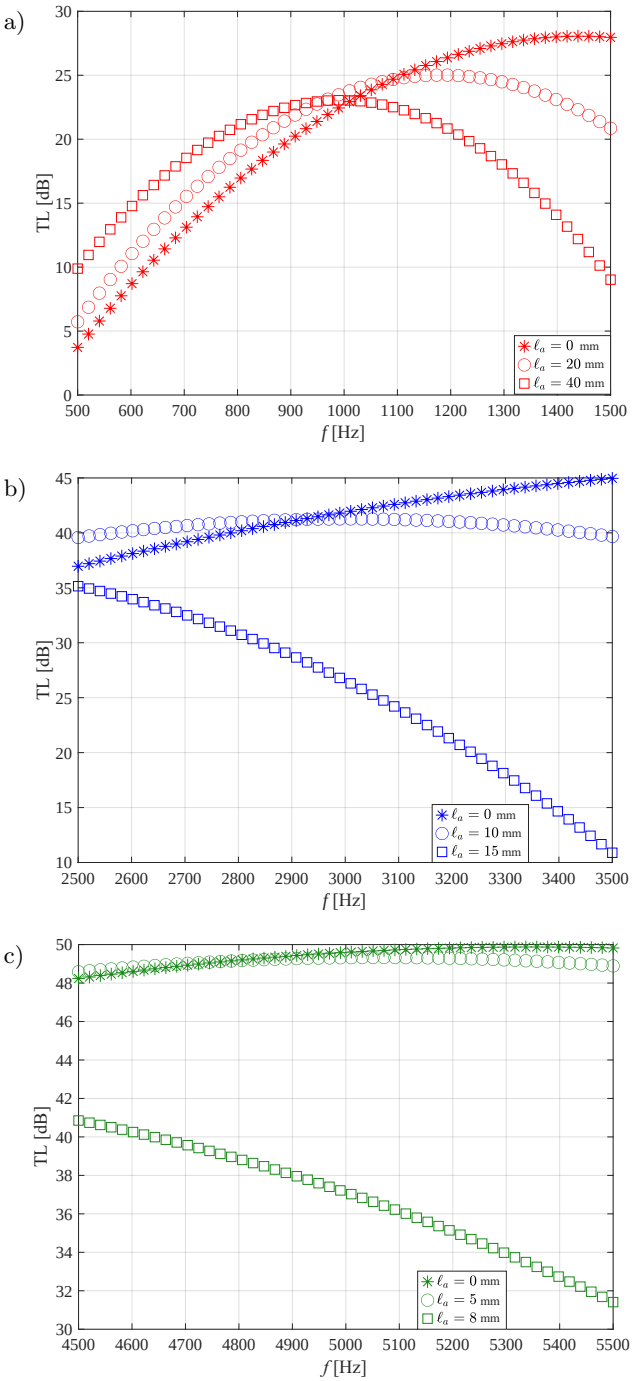

Fig. 3. Influence of different ℓ_a values on the maximum TL at selected frequencies: a) 1000 Hz, b) 3000 Hz, c) 5000 Hz.

For each frequency, the value of ℓ_a was chosen, which produced a TL value closest to its maximum. So, these results were $\ell_a = \{40, 23, 10, 7, 5, 0\}$ mm for $f = \{1, 2, 3, 4, 5, 6\} \times 10^3$ Hz, respectively. From discrete ℓ_a values, based on an approximation theory, an empirical formula was derived, as a function of frequency f, i.e., $\ell_a = \ell_a(f)$. This relationship is given based on an approximation theory and depicted in Fig. 4:

$$\ell_a = -4.6895 + \frac{46398.7}{f}.$$

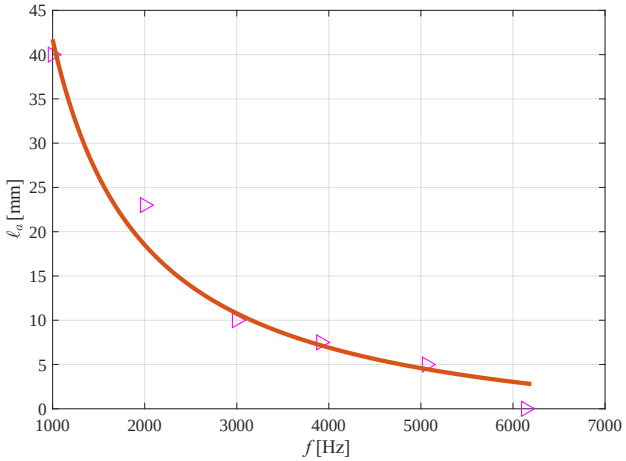

Fig. 4. Approximate value of ℓ_a as a function of frequency f.

3.2. *Influence of the horizontal inlet/outlet extensions on a single* D-EC

First, the influence of the length $\ell_{p,i}$ or $\ell_{p,o}$ or both of the horizontal extensions of the D-EC on the TL was analyzed. These considerations are similar to those published in (Selamet *et al.*, 2003; Łapka, 2007; Chaitanya, Munjal, 2011; Munjal, 2013; Xiang *et al.*, 2016; Chang *et al.*, 2019; Zhao, Li, 2022) but here they refer to the dedicated frequency.

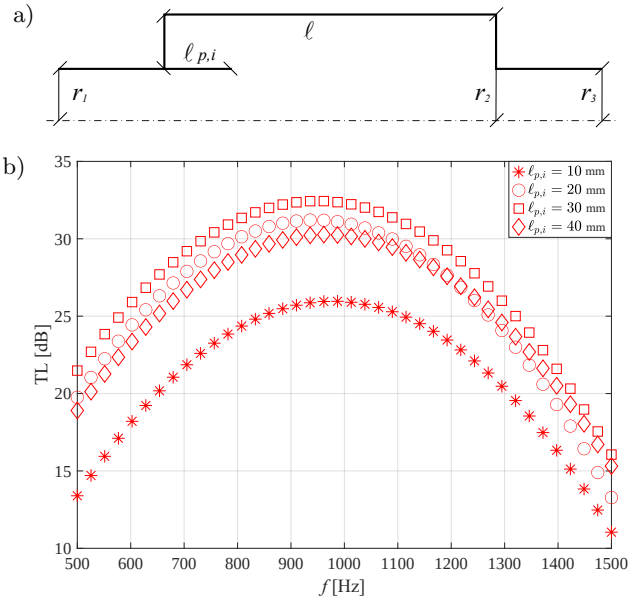

Fig. 5. Cross-section of the silencer with the horizontal inlet extension $\ell_{p,i}$ (a); effect of $\ell_{p,i} = \{10, 20, 30, 40\}$ mm on the TL, $f = 1000$ Hz (b).

At a frequency of 1000 Hz, the same horizontal extensions length $\ell_{p,i} = 30$ mm (first case) or $\ell_{p,o} = 30$ mm (second case) resulted in the same TL increase of about 9 dB; further increase in these lengths did not yield additional TL increase (Figs. 5 and 6). Whereas, using both horizontal extensions of the inlet and outlet lengths $\ell_{p,i} = \ell_{p,i} = 40$ mm (third case) produced a TL increase of about 21 dB (Fig. 7). However, if in the

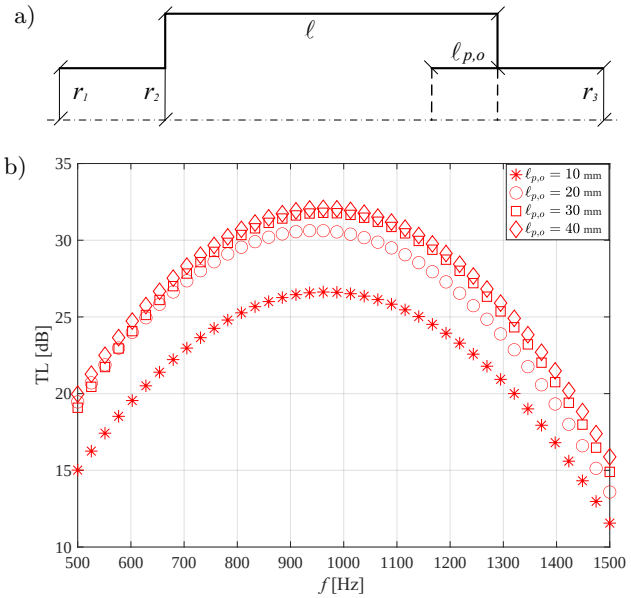

Fig. 6. Cross-section of the silencer with the horizontal outlet extension $\ell_{p,o}$ (a); effect of $\ell_{p,o} = \{10, 20, 30, 40\}$ mm on the TL, $f = 1000$ Hz (b).

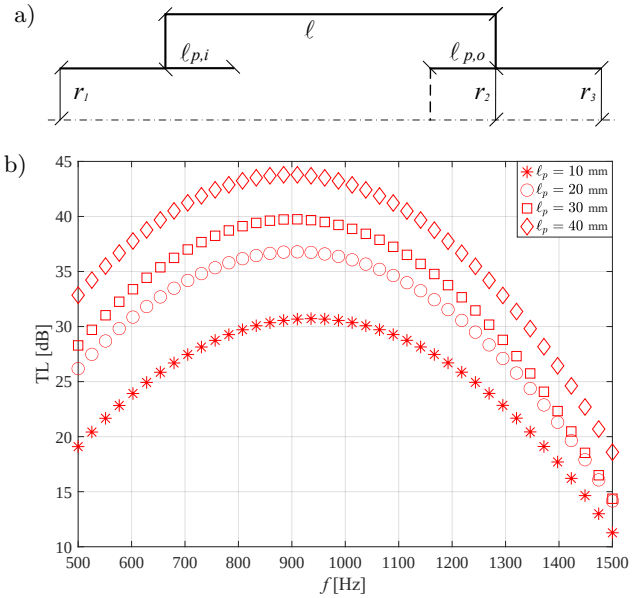

Fig. 7. Cross-section of the silencer with both horizontal inlet and outlet extensions $\ell_{p,i}$ and $\ell_{p,o}$ (a); effect of the $\ell_{p,i} = \ell_{p,o} = \ell_p = \{10, 20, 30, 40\}$ mm on the TL, $f = 1000$ Hz (b).

third case the sum of these lengths, i.e., $\ell_{p,i} + \ell_{p,o}$, is approximately equal to the length of $\ell_{p,i}$ (first case) or $\ell_{p,o}$ (second case), i.e., about 30 mm, then the TL increase is about 14 dB.

For frequency 3000 Hz the same horizontal extensions length $\ell_{p,i}$ = 30 mm (Fig. 8) or $\ell_{p,o}$ = 30 mm (Fig. 9) and for frequency 5000 Hz the same horizontal extensions length $\ell_{p,i}$ = 15 mm (Fig. 11) or $\ell_{p,o}$ = 15 mm (Fig. 12) gave the same maximum TL increase of about 9 dB. However extensions of the inlet and outlet by the same length $\ell_{p,i} = \ell_{p,o}$ = 15 mm for 3000 Hz (Fig. 10) and $\ell_{p,i} = \ell_{p,o}$ = 5 mm–10 mm for 5000 Hz (Fig. 13) gave the TL increase also of about 9 dB (cf. Chaitanya, Munjal, 2011).

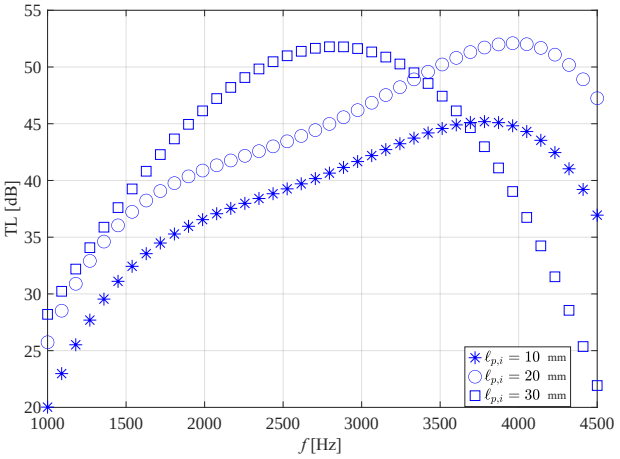

Fig. 8. Effect of $\ell_{p,i} = \{10, 20, 30\}$ mm on the TL, f = 3000 Hz.

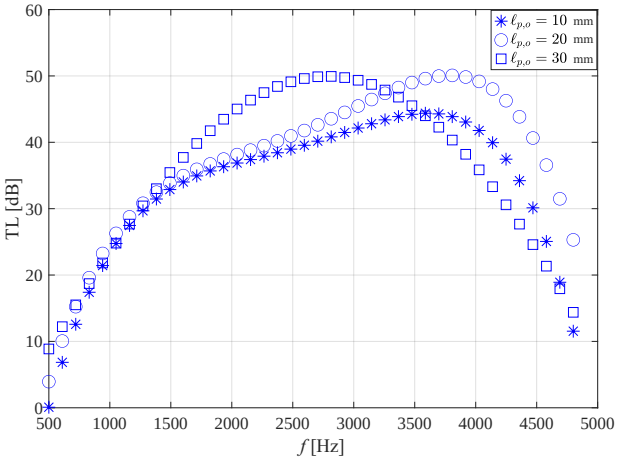

Fig. 9. Effect of $\ell_{p,o} = \{10, 20, 30\}$ mm on the TL, f = 3000 Hz.

3.3. *Influence of the second* D-EC

The simplest way to increase the TL of a silencer at the dedicated frequency is to connect two D-ECs in series. This is possible by inserting a transverse partition into the EC of the appropriate length, so that two D-ECs are formed. However, the geometric parameters of this partition also affect the TL value.

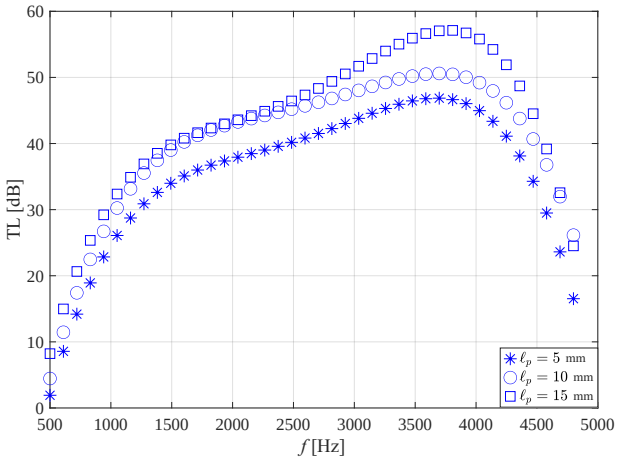

Fig. 10. Effect of $\ell_{p,i} = \ell_{p,o} = \ell_p = \{5, 10, 15\}$ mm on the TL, f = 3000 Hz.

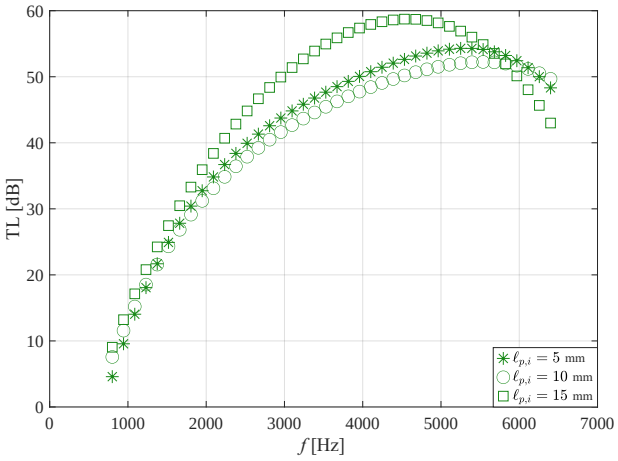

Fig. 11. Effect of $\ell_{p,i} = \{5, 10, 15\}$ mm on the TL, f = 5000 Hz.

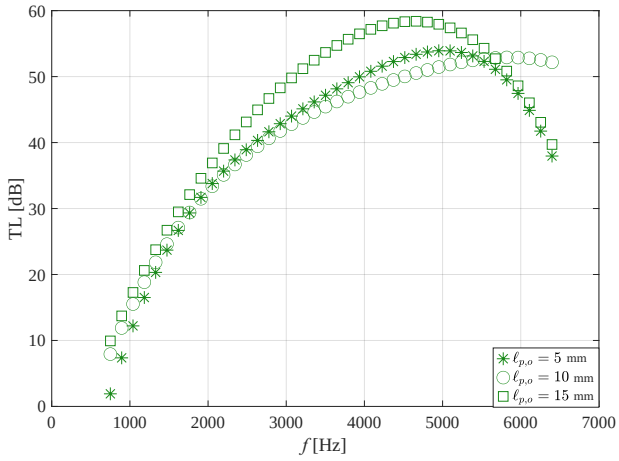

Fig. 12. Effect of $\ell_{p,o} = \{5, 10, 15\}$ mm on the TL, f = 5000 Hz.

First, for a selected baffle width of h = 5 mm and with the orifice diameter d_0 equal to the inlet and outlet diameters, i.e., $d_0 = 2r_1 = 2r_3$ = 6 mm, the TLs of a single D-EC and of two D-ECs were compared, Fig. 14.

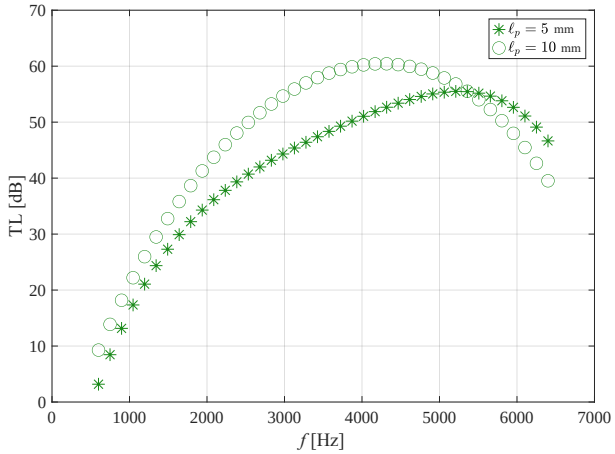

Fig. 13. Effect of $\ell_{p,i} = \ell_{p,o} = \ell_p = \{5, 10\}$ mm on the TL, $f = 5000$ Hz.

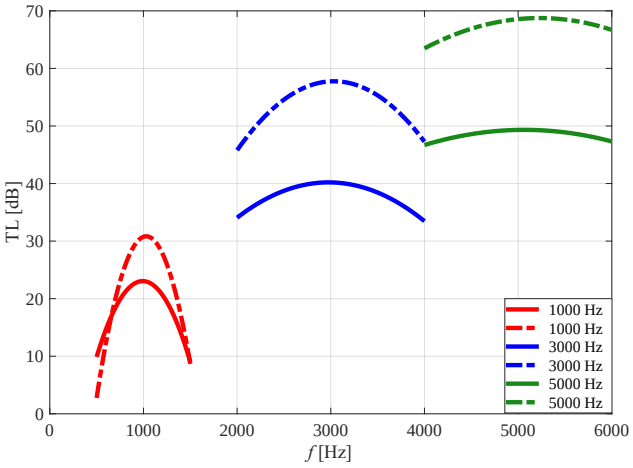

Fig. 14. Influence of the number of D-ECs on the TL at selected frequencies; solid lines – single D-ED, dashed lines – two D-EC.

It can be seen that an increase in the number of D-ECs from one to two causes an increase in the TL; this conclusion is qualitatively obvious. Furthermore, the double D-EC does not significantly shift the maximum TL, and it still functions as a dedicated silencer. Moreover, with an increase of dedicated frequency, the difference in maximum TL between one D-EC and double D-EC also increases, i.e., at 1000 Hz – the difference is about 7 dB, at 3000 Hz – about 17 dB, and at 5000 Hz – about 19 dB.

Next, the effect of the transverse partition width h between the D-ECs on the TL is analyzed. It is assumed that the partition orifice, as well as the inlet and outlet diameters, are the same as aforementioned; the results are depicted in Fig. 15.

As can be seen from Fig. 15, assuming a fixed transverse partition orifice diameter d_0, the transverse partition width h between the D-ECs influences the TL value at the dedicated frequency. In the analyzed frequencies, the optimal width h is about $h = 10$ mm, while a TL increase is about 7 dB–8 dB.

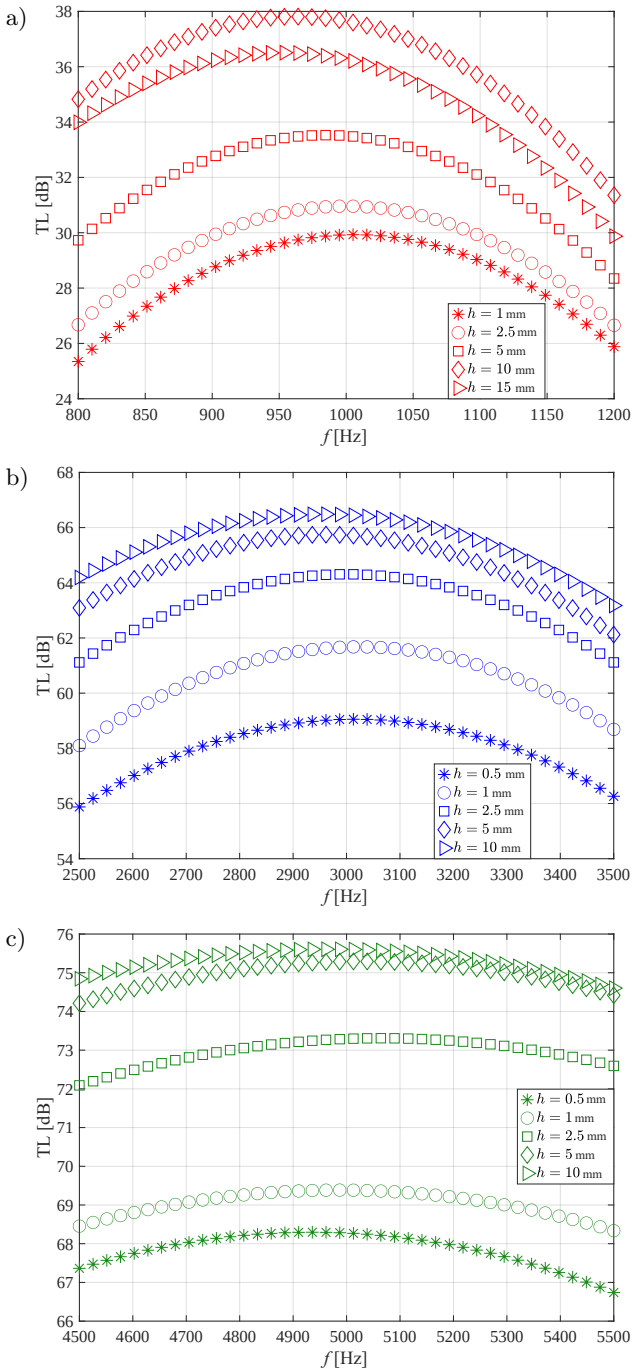

Fig. 15. Effect of the transverse partition width h [mm], $d_0 = 6$ mm, between D-ECs on the TL for selected frequencies: a) 1000 Hz, b) 3000 Hz, c) 5000 Hz.

Finally, the influence of the transverse partition orifice diameter d_0 between the D-ECs on the TL is analyzed. It is assumed that the partition orifice width is $h = 5$ mm, with the inlet and outlet diameters as aforementioned; the results are presented in the Fig. 16.

From Fig. 16, it follows that assuming a fixed transverse partition width h, the smallest orifice diameter d_0 of the transverse partition between the D-ECs provides the largest TL value at the dedicated frequency; here

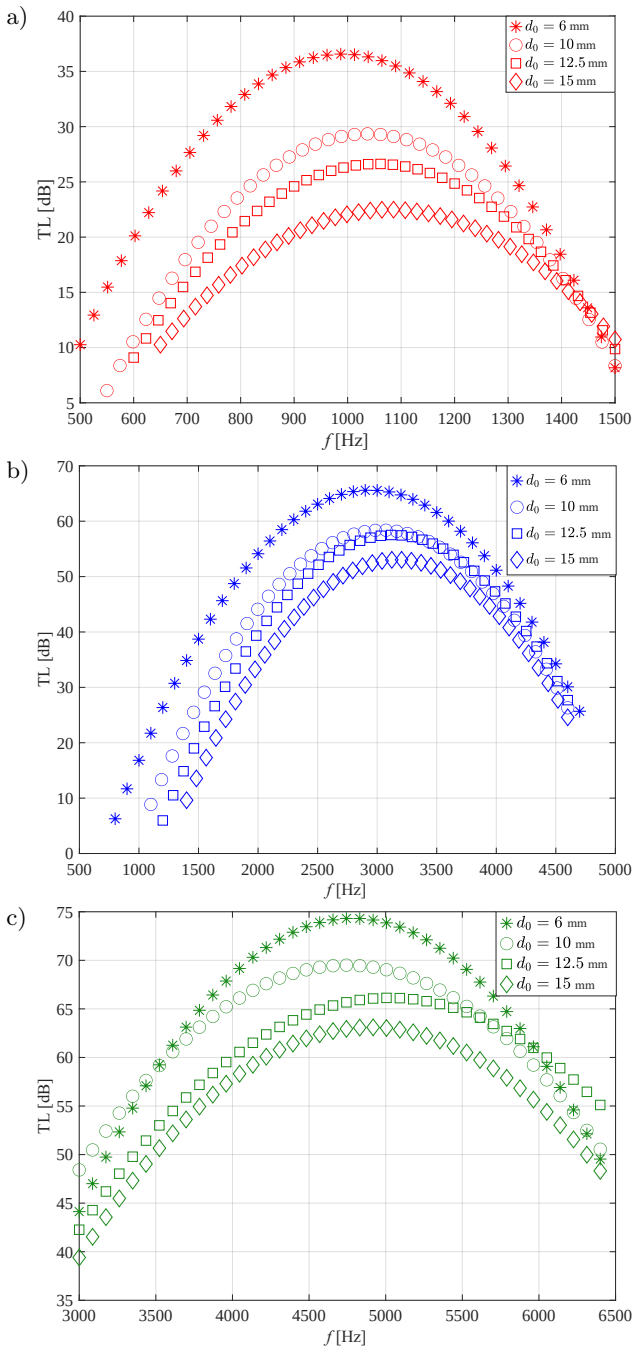

Fig. 16. Effect of the diameter of the transverse partition $d_0 = \{6, 10, 12.5, 15\}$ mm, $h = 5$ mm, between D-ECs on the TL for selected frequencies: a) 1000 Hz, b) 3000 Hz, c) 5000 Hz.

it is $d_0 = 6$ mm. However, the smallest diameter is dictated by technical operating conditions. By doubling the diameter d_0, e.g., from 6 mm to 12.5 mm, the TL value decreases by 10 dB–8 dB and the TL maximum slightly shifts towards higher frequencies.

4. Summary and general conclusions

It was shown that it is possible to build a simple silencer to damp noise at a dedicated frequency; it may even consist of a single EC. The effectiveness of such a silencer can also be easily increased, for example, by adding horizontal extensions to the inlet, the outlet, or both. Another simple method to improve noise reduction efficiency is to connect identical silencers in series. The most important conclusions from this study are as follows:

1) The plane wave theory gives a basis for determining the EC length for the dedicated frequency, and by adding an additional length, the D-EC is obtained. The D-EC is the simplest silencer for a dedicated frequency. The attached length was obtained from an empirical formula based on approximation theory for discrete experimentally obtained data.
2) For all analyzed frequencies, horizontal extension lengths, either $\ell_{p,i}$ or $\ell_{p,o}$, different for different frequencies, gave a TL increase of about 9 dB. A similar increase in TL was obtained for horizontal inlet and outlet extensions, provided that their combined length is the same as in the first and second case. Only at 1000 Hz, this increase is slightly greater.
3) Increasing the number of D-ECs obviously increases the TL. Moreover, as the dedicated frequency increases, the TL also increases.
4) For a fixed orifice diameter d_0 of the transverse partition between the D-ECs, there is an optimal width h that maximizes the TL value at the dedicated frequency.
5) For a fixed width h of the transverse partition between the D-ECs, the smallest orifice diameter d_0 provides the largest maximum TL value at the dedicated frequency. However, the smallest diameter d_0 is most often imposed due to technical reasons.

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

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

References

1. Barron R.F. (2003), *Industrial Noise Control and Acoustics*, Marcel Dekker, Inc., New York.

2. Bilawchuk S., Fyfe K.R. (2003), Comparison and implementation of the various numerical methods used for calculating transmission loss in silencer system, *Applied Acoustic*, **64**: 903–916, https://doi.org/10.1016/S0003-682X(03)00046-X.
3. Chaitanya P., Munjal M.L. (2011), Effect of wall thickness on the end corrections of the extended inlet and outlet of a double-tuned expansion chamber, *Applied Acoustic*, **72**(1): 65–70, https://doi.org/10.1016/j.apacoust.2010.09.001.
4. Chang Y.-C., Chiu M.-C., Huang S.-E. (2019), Numerical analysis of circular straight mufflers equipped with three chambers at high-order-modes, *Applied Acoustics*, **155**: 167–179, https://doi.org/10.1016/j.apacoust.2019.05.021.
5. Chiu M.-C., Chang Y.-C., (2014), An assessment of high-order-mode analysis and shape optimization of expansion chamber mufflers, *Archives of Acoustics*, **39**(4): 489–499, https://doi.org/10.2478/aoa-2014-0053.
6. Cui Z., Huang Y. (2012), Boundary element analysis of muffler transmission losses with LS-DYNA, [in:] *12-th International LS-DYNA Users Conference*.
7. Jokandan M.R., Variani A.S., Ahmadi S. (2023), Study of acoustic and aerodynamic performance of reactive silencer with different configurations: Theoretical, modeling and experimental, *Heliyon*, **9**(9): e20058, https://doi.org/10.1016/j.heliyon.2023.e20058.
8. Kang Z., Ji Z. (2008), Acoustic length correction of duct extension into a cylindrical chamber, *Journal of Sound and Vibration*, **310**(4–5): 782–791, https://doi.org/10.1016/j.jsv.2007.11.005.
9. Karami F., Rad M.S., Karimipour I. (2024), Review on attenuation of low – Frequency noise in passive silencer, *Journal of Low Frequency Noise, Vibration and Active Control*, **43**(4): 1679–1695, https://doi.org/10.1177/14613484241228373.
10. Komkin A.I., Mironov M.A., Yudin S.I. (2012), On the attached length of orifices, *Acoustical Physics*, **58**(6): 628–632, https://doi.org/10.1134/S1063771012050090.
11. Lee C.H., Han M.J., Park T.W. (2020), A comparative study on the transmission loss of Helmholtz resonator and quarter, half, conical half-wave resonator using acoustic analysis model, *International Journal of Mechanical Engineering and Robotics Research*, **9**(1): 153–157, https://doi.org/10.18178/ijmerr.9.1.153-157.
12. Łapka W. (2007), Acoustic attenuation performance of a round silencer with the spiral duct at the inlet, *Archives of Acoustics,* **34**(4(S)): 247–252.
13. Munjal M.L. (1987), *Acoustics of Ducts and Mufflers with Application to Exhaust and Ventilation System Design*, John Wiley & Sons, Inc., New York.
14. Munjal M.L. (2013), Recent advances in muffler acoustics, *International Journal of Acoustics and Vibration*, **18**(2): 71–85, https://doi.org/10.20855/ijav.2013.18.2321.
15. Munjal M.L. (2014), *Acoustics of Ducts and Mufflers*, John Wiley & Sons, New York.
16. Nilsson E., Vardaxis N.G., Ménard S., Hagberg D.B. (2021), Sound reduction of ventilation ducts through walls: Experimental results and updated models, *Acoustics*, **3**(4): 695–716, https://doi.org/10.3390/acoustics3040044.
17. Potente D. (2005), General design principles for an automotive muffler, [in:] *Proceedings of Acoustics 2005*, pp. 153–158, https://www.acoustics.asn.au/conference_proceedings/AAS2005/papers/34.pdf.
18. Rafique F., Wu J.H., Liu C.R., Ma F. (2022), Transmission loss analysis of a simple expansion chamber muffler with extended inlet and outlet combined with inhomogeneous micro-perforated panel (iMPP), *Applied Acoustic*, **194**: 108808, https://doi.org/10.1016/j.apacoust.2022.108808.
19. Rahman M., Sharmin T., Hassan A.F.M.E., Nur M.Al. (2005), Design and construction of a muffler for engine exhaust noise reduction, [in:] *Proceedings of the International Conference on Mechanical Engineering 2005 (ICME2005)*, https://me.buet.ac.bd/public/old/icme/icme2005/Proceedings/PDF/ICME05-TH-47.pdf.
20. Selamet A., Denia F.D., Besa A.J. (2003), Acoustic behaviour of circular dual-chamber mufflers, *Journal of Sound and Vibration*, **265**(5): 967–985, https://doi.org/10.1016/S0022-460X(02)01258-0.
21. Selamet A., Ji Z.L, Radavich P.M. (1998), Acoustic attenuation performance of circular expansion chambers with offset inlet/outlet: II. Comparison with experimental and computational studies, *Journal of Sound and Vibration*, **213**(4): 619–641, https://doi.org/10.1006/jsvi.1998.1515.
22. Selamet A., Radavich P.M. (1997), The effect of length on the acoustic attenuation performance of concentric chambers: an analytical, numerical and experimental investigation, *Journal of Sound and Vibration*, **201**(4): 407–426, https://doi.org/10.1006/jsvi.1996.0720.
23. Strek T. (2010), *Finite element modelling of sound transmission loss in reflective pipe*, [in:] *Finite Element Analysis*, Moratal D. [Ed.], https://doi.org/10.5772/10236.
24. Swamy M., van Lier L.J, Smeulers J. (2014), Optimisation of acoustic silencer for the screw compressor system, [in:] *Excerpt from the Proceedings of the 2014 COMSOL Conference in Cambridge*.
25. Tao Z., Seybert A.F. (2003), A review of current techniques for measuring muffler transmission loss, *SAE Transactions*, **12**: 2096–2100, https://doi.org/10.4271/2003-01-1653.
26. Wei F., Guo L.-X., (2016), An investigation of acoustic attenuation performance of silencers with mean flow based on three-dimensional numerical simulation, *Shock and Vibration*, **2016**: 6797593, https://doi.org/10.1155/2016/6797593.

27. Xiang L., Zuo S., Wu X., Zhang J., Liu J. (2016), Acoustic behaviour analysis and optimal design of a multi-chamber reactive muffler, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **230**(13): 1862–1870, https://doi.org/10.1177/0954407016630112.

28. Yu X., Cheng L. (2015), Duct noise attenuation using reactive silencer with various, *Journal of Sound and Vibration*, **335**: 229–244, https://doi.org/10.1016/j.jsv.2014.08.035.

29. Yu X., Tong Y., Pan J., Cheng L. (2015), Subchamber optimization for silencer design, *Journal of Sound and Vibration*, **351**: 57–67, https://doi.org/10.1016/j.jsv.2015.04.022.

30. Zalte Y.B., Sature M.J. (n.d.), Transmission losses in simple expansion chamber of reactive muffler analysis by numerical & experimental method, *International Engineering Research Journal*, pp. 1933–1939.

31. Zhang L., Shi H.-M., Zeng X.-H, Zhuang Z. (2020), Theoretical and experimental study on the transmission loss of a side outlet muffler, *Shock and Vibration*, **2020**: 6927574, https://doi.org/10.1155/2020/6927574.

32. Zhao B., Li H. (2022), Analysis of the influencing factors of the acoustic performance of the muffler considering acoustic-structural coupling, *Archives of Acoustics*, **47**(4): 497–490, https://doi.org/10.24425/aoa.2022.142900.

Archives of Acoustics Vol. 50, No. 4, pp. 445–454 (2025), https://doi.org/10.24425/aoa.2025.156933

Research Paper

Analysis of Decision Fusion in Speech Detection

Tomasz MAKA, Lukasz SMIETANKA*

Faculty of Computer Science and Information Technology, West Pomeranian University of Technology in Szczecin
Szczecin, Poland

*Corresponding Author e-mail: lsmietanka@zut.edu.pl

Received December 31, 2024; revised June 7, 2025; accepted November 3, 2025; published online November 19, 2025.

This article addresses the issue of detecting speech signal segments in an acoustic signal and analyzes potential decision fusion for a group of voice activity detectors (VADs). We designed ten new VADs using three different types of neural network architectures and three time-frequency signal representations. One of the proposed models has higher classification efficiency than competitive solutions. We used our VAD models to analyse data fusion and improve the final classification decision. For this purpose, we used gradient-free and gradient-based optimizers with different objective functions. The analysis revealed the impact of individual classifiers on the final decisions and the potential gains or losses resulting from VAD fusion. Compared with existing models, the models we proposed achieved higher classification accuracy at the cost of increased memory requirements. The final choice of a specific model depends on the platform constraints on which the VAD system will be deployed.

Keywords: voice activity detection (VAD); deep neural networks; data fusion.

1. Introduction

The problem of detecting speech in an acoustic signal involves identifying segments that contain speech. The detection mechanism for these segments is commonly used in various tasks where the signal serves as an input data source. This includes speech recognition, speaker identification, keyword spotting, and speech coding in telecommunications systems, all of which directly impact the effectiveness of classification. Although many speech detection systems, such as voice activity detector (VAD), have been developed so far, numerous new solutions have emerged recently. This is because VAD systems must operate under real-world conditions and incorporate adaptation mechanisms to handle varying acoustic environments. Additionally, their use in communication systems requires designers to develop models that account for hardware and time constraints. In speech detection, the main challenge arises from the non-stationary nature of speech signals and the diverse acoustic environments in which the signals are captured. Acoustic events and the momentary appearance and disappearance of sound sources influence the variability of the acoustic environment over time. Additionally, different acquisition conditions can introduce various types of noise into the speech signal at varying signal-to-noise ratios. These conditions make it difficult for machine learning models to accurately detect speech within a highly non-stationary signal.

Currently, existing and developed VAD systems are built based on different deep neural network architectures which very often use attention mechanisms (Song *et al.*, 2022; Wang *et al.*, 2022; Zhang *et al.*, 2023; Zhao, Champagne, 2022). Basic issues covered by such systems are connected with noise robustness and low use of energy and hardware resources. For example, Yang *et al.* (2024) introduced the sVAD model, which is based on an attention mechanism and achieves noteworthy robustness to noise. Moreover, as the authors state, it is characterized by low power consumption. Similarly, Zhao and Champagne (2022) described a VAD system built on top of the transformer architecture with an attention mechanism, which supports noise immunity and has moderate computational complexity. Kim *et al.* (2022) presented

ADA-VAD, which uses an adversarial domain adaptation mechanism to determine the properties of noisy signals. As a result, the proposed VAD is highly robust to various types of noise. SG-VAD model was proposed in (Svirsky, Lindenbaum, 2023). It was designed to work in a low-resource environment and comprises two neural networks. The model contains only 7800 parameters, which makes it suitable for running the system on edge devices. Despite a focus on noise robustness and low resource requirements, solutions for more complex scenarios, such a speech detection in multi-talker environments, have also been proposed (Aloradi *et al.*, 2023). Various issues related to the hardware implementation of 21 VADs, including performance criteria, limitations, and effectiveness, are discussed in (Yadav *et al.*, 2023).

In this work, we analyze the potential for decision fusion across ten VAD models by using an optimization process with three objective functions as examples. The paper is organized as follows: Sec. 2 discusses our VAD models, their architectures, and the dataset used in the experiments; in Sec. 3, we describe the fusion models, briefly discuss the optimisation process, and present the results; Sec. 4 concludes the paper.

2. Voice activity detection

This study aims to develop a VAD system capable of identifying speech segments containing speech signals in long audio recordings. Since our acoustic scene analysis system operates with a frame length of one second, the same frame length was used in the developed VAD modules and for comparisons with other VAD systems. To support this application, we created a custom dataset, generated from a variety of publicly available sources[1].

2.1. Dataset

In our dataset, we included three types of source signals: speech, music recordings without singing, and background noise. We randomly selected one-second frames from each group for the training set, converted them in to the appropriate representation, and used them for model training.

Table 1 presents the characteristics of the training and validation sets. The test set was created by randomly selecting fragments from the source data, each lasting between 5 s and 15 s, with segments ranging from 3 to 20 in number. These selected segments were then joined consecutively to form a single test signal. In total, 1000 such test signals were generated in this manner, 49 of which contained no speech segments.

Table 1. Characteristics of the one-second frame sets used in the training and validation process.

Process	Speech	Music	Background noise	Total
Train	2100	1050	1050	4200
Validation	900	450	450	1800
Total	3000	1500	1500	6000

2.2. VAD architectures

Our approach is based on signal frame classification. The input signal is divided into frames, from which one of four representations ($\widetilde{r}$) is derived. A decision module is then utilized, which outputs the probability (p) that the analyzed frame contains a speech signal. In the final stage, thresholding is applied, resulting in a binary value. If the probability exceeds 50 %, a value of 1 is generated at the output; otherwise, the output is 0. The entire process is illustrated in Fig. 1. To determine speech segments in an audio signal, we decided to use popular neural network architectures in conjunction with three time-frequency audio representations. Nine VAD models were designed in total.

Audio samples were converted in to three two-dimensional representations, which include spectrogram (spect), CQT-spectrogram (cqt), and mel-spectrogram (mel). The spectrogram uses a linear frequency scale, whereas the CQT-spectrogram uses a constant-Q transform (Schörkhuber, Klapuri, 2010), and in the mel-spectrogram, the frequency scale is mapped into mel scale (Rabiner, Schafer, 2010).

[1]The list of data sources is available at: https://github.com/staticvoice/ovad/blob/main/FusionDataSources.md

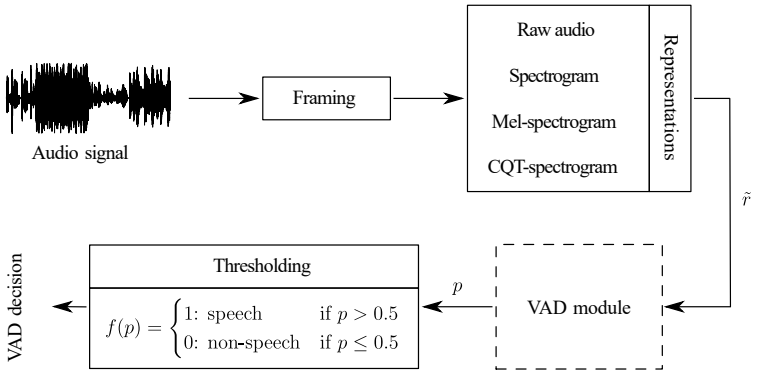

Fig. 1. General architecture for detecting speech-containing segments in acoustic signals.

All data were calculated using the Librosa Library (McFee *et al.*, 2015), and we used the following configuration of these representations:

- spectrogram (spect): n_fft: 1024, win_length: 512, $n\_features$: 513, hop_length: 512;
- CQT-spectrogram (cqt): n_bins: 90, bins_per_octave: 12, $n\_features$: 90, hop_length: 512;
- mel-spectrogram (mel): n_mels: 128, n_fft: 1024, $n\_features$: 128, hop_length: 512.

The following three neural networks architectures were used in the design of our VADmodels[2]:

1) BiLSTM (Ma *et al.*, 2022)
A simple model built with three recurrent layers, a single linear layer and a dropout layer. The first utilized architecture is a BiLSTM (Fig. 2). It is a simple model consisting of three recurrent layers: two unidirectional layers (LSTM layers) separated by a bidirectional layer (BiLSTM layer). The hidden size of the first unidirectional layer and the subsecuent BiLSTM layer is determined by the number of features ($n\_features$) in the input representation. In turn, the hidden size of the second LSTM layer is equal to $2 \cdot n\_features$. Additionally, to mitigate the phenomenon of model overfitting during the training process, the bidirectional recurrent layer is preceded by a dropout layer with a rate of 0.2. The entire model concludes with a fully connected (FC) layer with a sigmoid activation function.

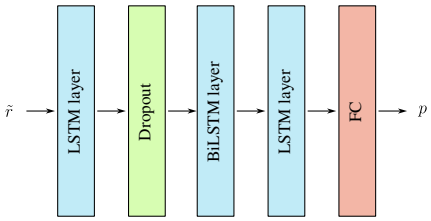

Fig. 2. VAD decision module based on the BiLSTM architecture.

2) ResNet50 (He *et al.*, 2016)
The next model, ResNet50 (Fig. 3), is a slight modification of the original architecture with the same name, differing only in changes to the first convolutional layer and the final FC layer. The first difference arises due to the type of data provided to the network's input. In the original architecture, the input consists of RGB images with three channels. In contrast, the variant used in this study takes spectrograms as input, which are single-channel images. This necessitates the use of a single input channel in the first convolutional layer instead of three. The second modification involves adapting the final FC layer of the model for binary classification. The primary component, which is the sequence of residual blocks (Fig. 6a), remains unchanged. Similarly, the layers responsible for dimensionality reduction (MaxPool, AvgPool) and the layer that converts the data into a one-dimensional vector (flattening) also remain unaltered.

[2]All proposed models can be found at: https://github.com/staticvoice/ovad/models/

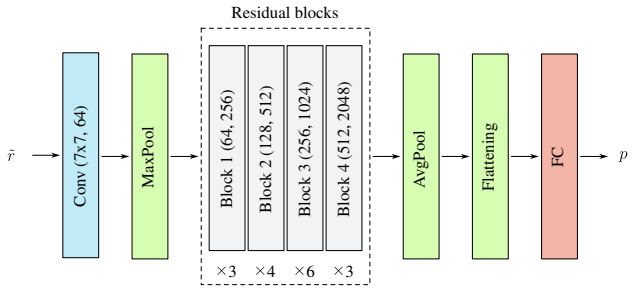

Fig. 3. VAD decision module based on the ResNet architecture.

3) ViT (Dosovitskiy *et al.*, 2021)
The third utilized architecture is the classic vision transformer (ViT), see Fig. 4. In this model, the input data is first divided into patches with a size of 16×16. Each patch is then mapped (via linear projection) to a 128-dimensional vector, which is supplemented with positional information within the sequence. Subsequently, the entire input is processed through a sequence of 12 transformer blocks (Fig. 6b). Each block consists of eight attention heads (MHA), normalization layers (norm), and linear layers (MLP). The architecture concludes with an MLP head composed of fully connected linear layers with a sigmoid activation function. Additionally, due to the need to divide the input data into patches, each input spectrogram was scaled to the following dimensions: cqt and mel to 128×128, and spect to 128×512.

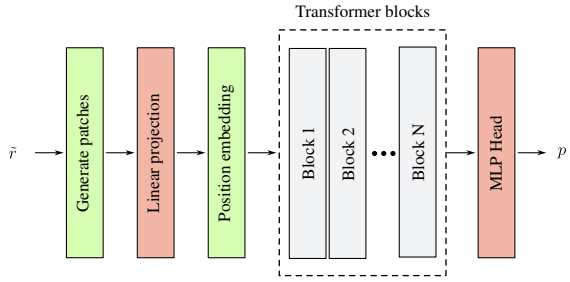

Fig. 4. VAD decision module based on the ViT architecture.

4) AugViT (Smietanka, Maka, 2023)
The final architecture used is AugViT (Fig. 5). This model is based on the standard sequence of transformer blocks, but it is preceded by a block that incorporates additional augmentation. Unlike the three previous models, the input to this

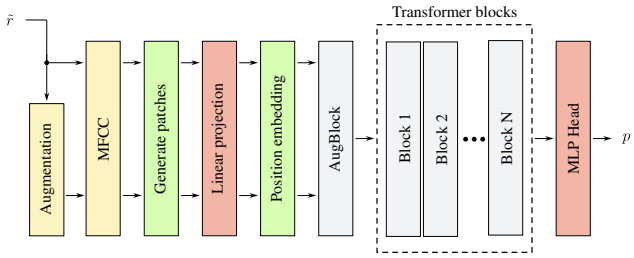

Fig. 5. VAD decision module based on the AugViT architecture.

architecture is raw audio. A random augmentation is applied to a copy of this raw signal. In the next stage, MFCC coefficients are computed separately for both the original and augmented signals. Subsequently, both MFCC representations are divided independently into patches (each patch corresponds to a single MFCC column). These patches are linearly projected into 8-dimensional vectors and supplemented with positional information within the sequence. Next, these sequences are passed to the AugBlock (Fig. 6c). Compared to the original transformer block (Fig. 6b), this block consists of two attention heads: one processes data from the original audio signal, while the other processes data from the augmented signal. The subsequent stages follow the standard ViT structure: a sequence of eight transformer blocks (each with two attention heads) followed by an MLP head.

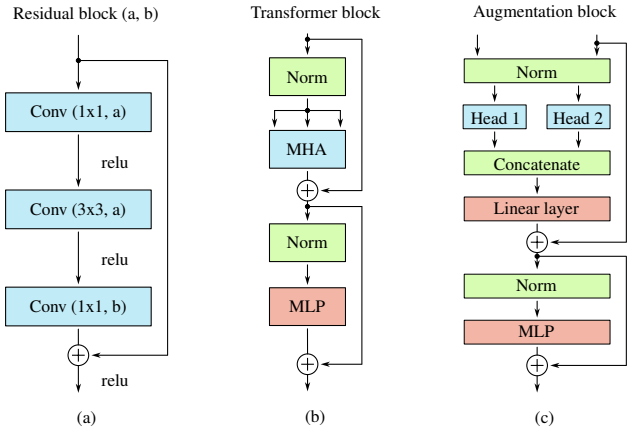

Fig. 6. Auxiliary blocks used in the VAD modules: residual (a), transformer (b), and augmentation (c) blocks.

The following parameters characterized the training procedure of each of these models:

- number of epochs in the training stage: 100 or less if, after 20 epochs, there is no improvement in classification (F_1-score not increase on the validation set);
- for further stages, the checkpoint model that obtained the highest F_1-score on the validation set was selected;
- the batch_size is equal to 16;
- Adam optimizer was selected with a learning_rate equal to 0.001;
- selected loss function: binary cross entropy (BCELoss).

2.3. Evaluation

Each of our speech detectors was tested on the entire test set. Additionally, the same data was used to carry out tests with two popular VADs: Silero (Team, 2024) and Brouhaha (Lavechin *et al.*, 2023). The results of the tests are presented as F_1-score distribution, as shown in Fig. 7. All the proposed VADs exhibit comparable classification efficacy, with the ResNet50-cqt model achieving the highest accuracy on the test set. For a detailed comparison of our best model with the Silero and Brouhaha VADs, we computed the confusion matrices, which are presented in Fig. 8.

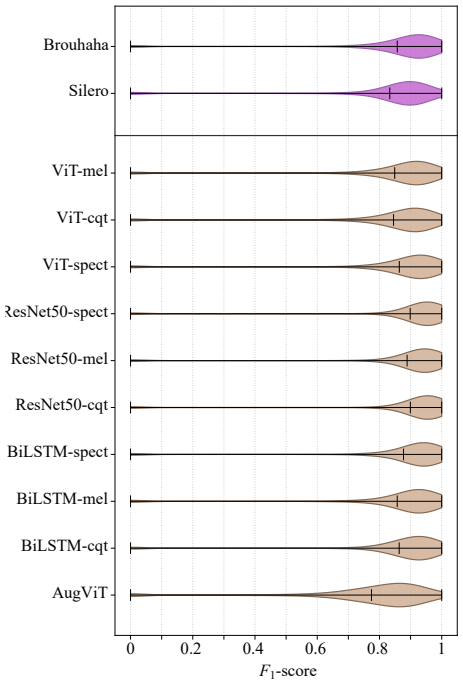

Fig. 7. Comparison of F_1-score distributions for prediction of speech segments in our proposed models and two competitive models obtained using test signals.

To compare the prediction speed and accuracy of the selected models, we predicted an audio signal of 138 seconds in length, containing four speech segments (30.07 % of the audio file) among nine other segments. The predictions were performed on a machine equipped with an i5-13600K CPU, an RTX 4070Ti GPU, and 32 GB of RAM. The results, including the models' memory requirements, are presented in Table 2. In the case of the ResNet50 architecture, there is no difference in the number of parameters or model size due to the first layers' independence from the complexity of the input data. The first layer in

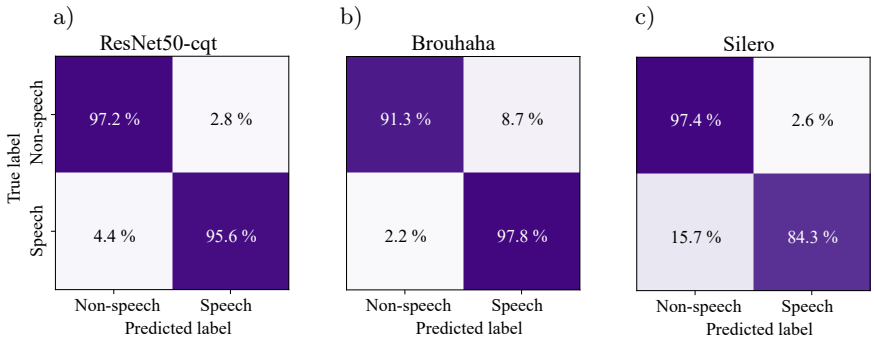

Fig. 8. Confusion matrices of our best model ResNet50-cqt (a), Brouhaha (b), and Silero (c) VADs.

Table 2. Comparison of models in predicting speech segments for an example test signal. The prediction times measured using both the CPU and GPU are presented, along with the model sizes, the number of parameters for each model, and the achieved prediction performance expressed as the F_1-score.

Model	CPU [s]	GPU [s]	F_1-score	Parameters	Size [MB]
Silero	0.6118 (±0.1102)	0.7076 (±0.1248)	0.889	462 594	2.1
Brouhaha	2.1283 (±0.0484)	0.3879 (±0.0076)	0.941	3 930 599	45
ViT-spect	2.0145 (±0.2207)	0.4116 (±0.1235)	0.895	2 446 721	9.4
ViT-mel	11.3619 (±0.4708)	0.6790 (±0.0048)	0.838	2 422 145	9.3
ViT-cqt	2.8021 (±0.2333)	2.0526 (±0.0031)	0.886	2 422 145	9.3
ResNet50-cqt	4.4269 (±0.0410)	2.0888 (±0.0321)	0.950	23 503 809	90
ResNet50-mel	4.2817 (±0.0345)	0.7287 (±0.0023)	0.925	23 503 809	90
ResNet50-spect	9.2796 (±0.0017)	0.4187 (±0.0089)	0.937	23 503 809	90
BiLSTM-cqt	2.1067 (±0.0024)	1.9555 (±0.0027)	0.817	457 381	1.8
BiLSTM-mel	0.8850 (±0.0071)	0.5373 (±0.0072)	0.865	922 881	3.5
BiLSTM-spect	7.5258 (±0.1073)	0.8784 (±0.0562)	0.897	14 759 011	56.3
AugViT	0.3921 (±0.0005)	0.5100 (±0.0720)	0.886	10 681	0.4

this architecture is a convolutional layer (Conv2d) with a fixed number of filters across all audio representations. For the ViT architecture, there is a slight difference in model size when using the spectrogram compared to other audio representations. This variation is due to differences in the number of patches into which the input can be divided. However, this number has minimal influence on the overall number of parameters. For instance, both cqt and mel spectrograms have the same number of parameters because both were interpolated to a size of 128×128, whereas the standard spectrogram was interpolated to 512×128. In contrast, for the BiLSTM models, the size of the initial LSTM layer depends on the number of rows (i.e., frequency bins) in the input representation. This, in turn, affects the total number of intermediate states in subsequent layers.

3. Data fusion

Fusing classifier outputs can be implemented in various ways (Kittler *et al.*, 1998). The basic classifier fusion techniques include so-called voting techniques: hard voting and soft voting. In the case of the first voting technique, a given class is determined as the one selected by the majority of classifiers. The second method involves averaging the probabilities and comparing them against a predefined threshold (Rokach, 2005). All of the 10 classifiers described in Subsec. 2.2 were used to fuse their individual decisions to improve speech signal detection on the test set. Because we obtained vectors with probabilities from the classifiers' outputs, we decided to use them to determine the final decision. For this purpose, for each vector, the probability of each classifier, we assigned $\alpha_n \in (0,1)$ coefficients to scale the entire vector, and thus the degree of its impact in merging the decisions of all classifiers. To determine what values α_n coefficients should be assigned to the individual vectors; we used optimization procedures and proposed the following three models for fusion. The first model is a linear combination of the probabilities from individual VAD modules:

$$\widehat{f}_1(k) = \sum_{n=1}^{N} p_n(k) \cdot \alpha_n. \tag{1}$$

The second model is also a linear combination of decisions, but only from those modules where the probability of speech presence in a given frame exceeds 60 %:

$$\widehat{f}_2(k) = \sum_{n=1}^{N} p_n(k) \cdot \widetilde{\alpha}_n; \qquad \widetilde{\alpha}_n = \begin{cases} \alpha_n & \text{if } \alpha_n > r, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $r = 0.6$, $N = 10$, $\epsilon = 10^{-8}$, $p_n(k)$ is the probability of the k-th frame of the n-th classifier, and α_n is the model coefficient for the n-th classifier.

In the case of the third model, the result of the first model is used, with its decision trajectory dynamics altered by applying a logarithmic function:

$$\widehat{f}_3(k) = \log\left[\widehat{f}_1(k) + \epsilon\right]. \tag{3}$$

The process of combining decisions from the set of proposed VAD systems and decision fusion models described by Eqs. (1)–(3) is implemented as the optimization of parameters α_n to maximize the F_1-score. The mechanism for tuning these coefficients is schematically illustrated in Fig. 9. The process of determining the objective function for a single fusion model is carried out in the following steps:

1) for α_n coefficients, determine the resulting function signal, according to the specified model ($\widehat{f}_1$, $\widehat{f}_2$, $\widehat{f}_3$);
2) normalize the obtained signal to the $(0, 1)$ range;
3) apply threshold-based detection with $h = 0.5$;
4) compute the F_1-score value between the detected and target signals which is the final value of the objective function.

The F_1-score is computed as the harmonic mean of precision and recall (Rijsbergen, 1979). Using the true positives (TP), the false positives (FP), and the false negatives (FN) values, the score can be described as follows:

$$F_1\text{-score} = \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}}. \tag{4}$$

We used both gradient-free (Opt I) and gradient-based (Opt II) optimization processes to determine the coefficients of the three proposed models. The entire process of optimizing the model coefficients is depicted in Fig. 9. Since every signal in the dataset was automatically generated and labeled, the optimization process was provided with an audio signal and its corresponding valid VAD trajectory. To determine the gain or loss during optimization, we used the following rule, where the value G is expressed as a percentage $G \in (-100, 100)$:

$$G = 100 \cdot \left(1 - \frac{\widehat{F}_1\text{-score}}{\widetilde{F}_1\text{-score}}\right), \tag{5}$$

where $\widehat{F}_1$-score is the best score obtained for whole set of VAD modules, and $\widetilde{F}_1$-score is the best score for the fused architecture.

3.1. Gradient-free optimization

For this type of optimization we used the random annealing algorithm (Blanke, 2020), which uses a hill-climbing technique with a variable step in time, similarly as in the simulated annealing method. We decided to use this algorithm after conducting a series of experiments with signals generated in the same way as those from our test set. This algorithm achieved the best results for each of the proposed models. The optimization procedure was performed separately for each objective function and for all signals in the test set. The procedure was carried out for each signal by maximizing the F_1-score over 1000 iterations. The coefficients α_n were searched within the range of 0 to 1, and the step size was equal to 0.2.

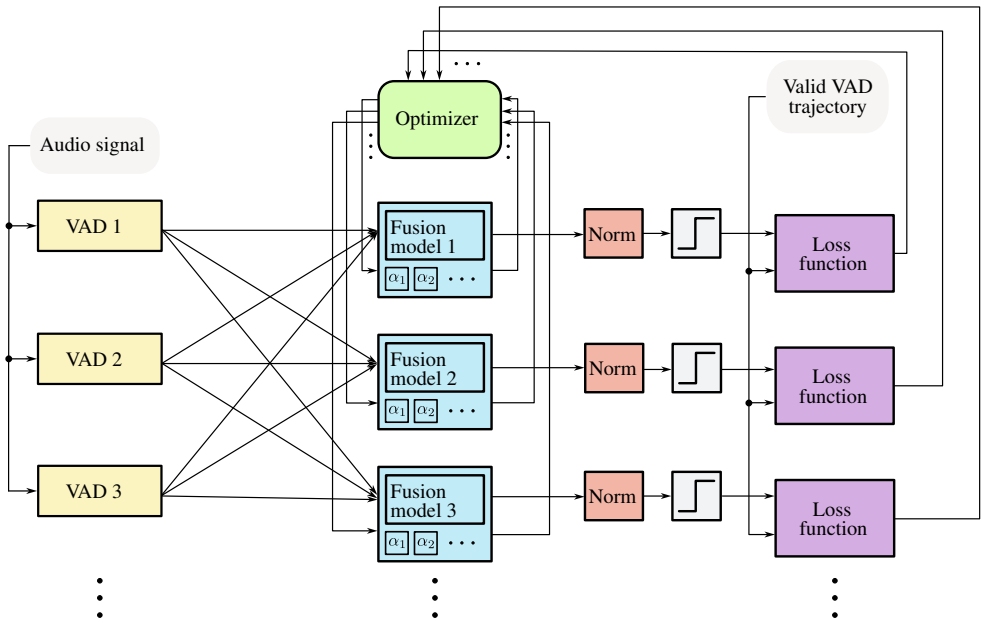

Fig. 9. General framework for optimizing decision fusion models obtained from a set of VAD modules.

3.2. Gradient-based optimization

As a gradient optimiser, we utilized the Adam (adaptive moment estimation) algorithm, an extended version of stochastic gradient descent algorithm. The Adam algorithm is known for its efficiency and robustness, and therefore we use it in the optimization process (Kingma, Ba, 2015). The optimisation procedure was performed as follows. First, we initialized the weight vector W_α with a uniform distribution with values in the $(0, 1)$ range. The variable B_f was initialized with 0; the role of this variable is to hold the highest value of the objective function. The variable B_{W_α} contains the weights for the best F_1-score. We used the binary cross entropy (BCELoss) loss function, a learning rate LR = 0.001, and the number of epochs was equal to 1 000 000. In each epoch, the given fusion model was calculated from the probabilities of ten classifiers and the weight vector W_α. From the resulting signal, the objective function was calculated. If the value of objective function (g) was higher than B_f, then $B_f = g$ and $B_{W_\alpha} = W_\alpha$. Then, an optimization of weight vector W_α using the obtained loss was performed. When, after 1000 epochs, B_f did not increase, the learning rate was reduced: LR = LR $\cdot$ 0.01. Early stopping was applied if, after 10 000 epochs, B_f did not increase. In the end, resulting B_{W_α} weights were the final coefficients of thea given model. In this case, we resigned from limiting the coefficients α_n to the range $(0, 1)$ for comparison purposes with the previous algorithm, as this limitation could have a negative impact on the optimization quality. This caused negative values in the signal after the fusion process, and in this case, it eliminated the $\widehat{f_3}$ model from use.

3.3. Results

To determine the effectiveness of the proposed fusion models, we conducted a series of experiments involving the individual fusion of each signal from the test set. The obtained coefficients were used to determine the new detection trajectory and the corresponding F_1-score, which was then compared to the F_1-score of the best of our VAD model for a given signal. Based on this comparison, gain or loss was determined. Table 3 shows its smallest, largest, and average values. The average gain in the best cases resulting from classifier fusion was less than one percent. Figure 10 depicts the gains and losses obtained on the

Table 3. Fusion results for the test set.

Fusion type	Gain(+) / Loss(−) [%]		
	Minimum	Maximum	Average
Hard voting	−48.03	7.41	−1.53
Soft voting	−41.18	7.41	−1.28
Opt I (model 1)	−10	9.71	+0.67
Opt I (model 2)	−10	10.91	+0.67
Opt I (model 3)	**−10**	**11.76**	**+0.74**
Opt II (model 1)	−15.79	12.59	+0.69
Opt II (model 2)	**−5.26**	**16.08**	**+0.89**

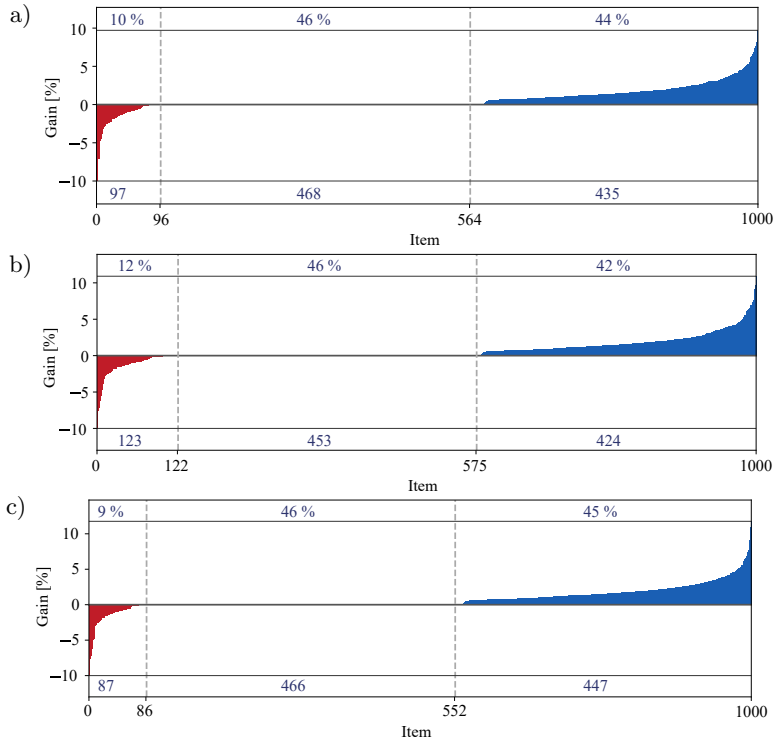

Fig. 10. Gains and losses in the test set from the optimisation process for model 1 (a), model 2 (b), and model 3 (c) using non-gradient optimization (Opt I).

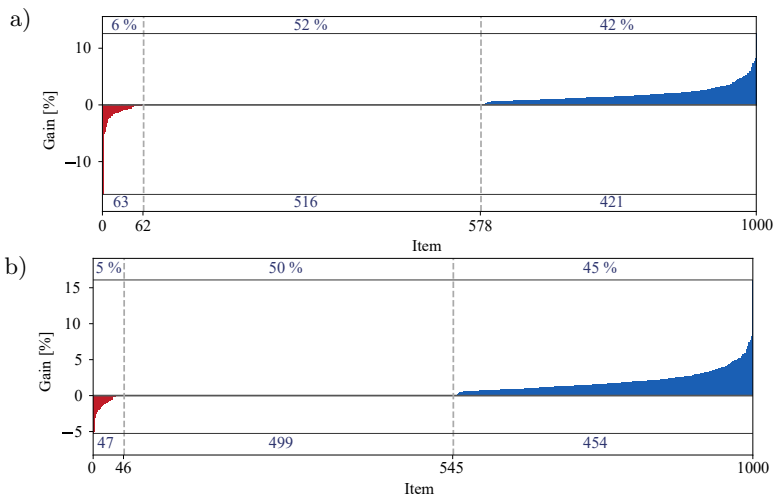

Fig. 11. Gains and losses in the test set from the optimization process for model 1 (a), and model 2 (b) using gradient optimization (Opt II).

test set for non-gradient optimization, whereas Fig. 11 for gradient optimization. In both figures, deterioration in F_1-score is marked in red, and improvement in blue, compared to the best VAD model for individual signals. As shown in Fig. 10, the highest number of cases with improved classification accuracy was achieved with model 3, where only 9 % of test signals experienced a decline in classification performance. For each of the fusion models used, the number of cases with neither improvement nor deterioration in classification was similar, amounting to approximately 46 %.

In the case of gradient optimization, the results include only two models. As mentioned in Subsec. 3.2, the possibility of weight coefficients dropping below zero and the use of a logarithm in model 3 made its inclusion in the experiments impossible. Based on the obtained results in this case, it can be observed that the number of instances where classification performance deteriorated due to fusion is almost halved compared to non-gradient optimization. Additionally, the highest gain achieved in this case exceeded 16 %.

Because, in the case of non-gradient optimization, the coefficients α_n directly influenced the significance of the probabilities in each of the VAD models, Fig. 12 shows their distribution for the entire test set in the best case where model 3 was used. Based on the obtained results, it can be concluded that the greatest contribution to the final decision comes from the ResNet50-spect model (α_7), BiLSTM-mel (α_3), and AugViT (α_1).

Interestingly, for the same architectures but different representations, there are significant differences in the distribution of weight coefficients (e.g., α_5, α_6, and α_7) determining the fusion of individual VAD modules. This indicates that the representation of the acoustic signal also plays a significant role in the effectiveness of the VAD module.

Table 4 presents the percentage contribution of individual VAD models to the correct classification of frame groups. Each group represents frames from the test set that were correctly classified by at least one and at most all classifiers. Additionally, the last row of the table shows what percentage of the entire test set each group represents.

A total of 74.1 % of frames were correctly classified by all classifiers. In 15.2 % of cases, frames were

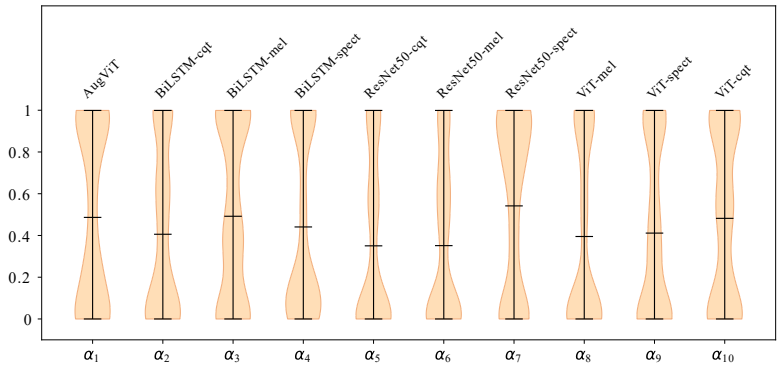

Fig. 12. Variability of the α coefficients in model 3 of the fusion (Opt I), showing the highest average value across 1000 test signals.

Table 4. Percentage contribution of VAD models to the classification of individual frame groups.

VAD	Group									
	1	2	3	4	5	6	7	8	9	10
AugViT	24.3	35.9	44.5	50.9	50.4	56.3	58.2	58.5	53.3	100
BiLSTM-cqt	16.1	23.5	31.1	41.2	45.5	57.4	68	80.5	91.7	100
BiLSTM-mel	6.1	15	24.9	33.2	41.4	49.8	61.6	74.3	95.1	100
BiLSTM-spect	8.5	20.9	35.7	48.1	54.6	67.2	75.7	82.5	95	100
ResNet50-cqt	2.7	7	24.4	36	57.3	72.8	89	95.4	99.2	100
ResNet50-mel	4.4	11.7	19.6	33.7	52.8	66.3	81.8	91.1	97.5	100
ResNet50-spect	11.8	32.6	45.6	54	67	74.9	85.1	92.5	97.7	100
ViT-cqt	10.6	21.1	30.6	36.9	45.9	58.3	64.5	73.9	84.8	100
ViT-mel	6.8	12.2	14.1	28	34.4	41.4	53.7	75	92.3	100
ViT-spect	8.7	20	29.5	38.1	50.7	55.5	62.4	76.3	93.3	100
Number of frames	0.4	0.4	0.4	0.5	0.7	1.1	1.9	4.5	15.2	74.1

correctly classified by any nine models. A smaller part of the set, 4.5 %, was correctly classified by any eight models, with ResNet VAD being the most accurate. On the other hand, 0.4 % of frames were correctly classified by only a single model, with AugViT VAD performing the best. A similar situation is observed for frames correctly classified by 2 to 7 classifiers. In these cases, each model correctly classified only a portion of the frames, but the fusion of their decisions positively affected the final result. The number of frames not correctly classified by any model was 964 (0.8 %).

4. Conclusion

In the case of analyzing the fusion mechanisms, the individual VADs learned on the same data and therefore the fusion influence in such a case was small. All the VADs we proposed were quite efficient, with an average F_1-score above 0.8, which directly impacts the fusion of decisions. This may lead to the conclusion that the chosen network architecture and signal input representation have less impact on the efficiency of VAD performance compared to the quality of the data used to train these models. When examining the resulting trajectory after detection, one can see that there many single frames that are wrongly classified. Thus, applying well-known post-processing techniques (Peinado, Segura, 2006) may improve the accuracy of frame classification. In this work, we attempted to analyze the decision fusion process in ten VAD modules. As the results demonstrate, the decision to implement the fusion process in a practical solution must be based on factors such as computational and memory resource constraints, the characteristics of the source data, and the conditions of signal acquisition. These factors directly impact the effectiveness of VAD models and, consequently, the potential contribution of fusion process in improving the overall classification performance.

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

Tomasz Maka – implementation of data fusion mechanisms, preparation of the audio dataset; Lukasz Smietanka – model implementation and training, development of the framework for the VAD components. Both authors contributed to the conceptualization of the study as well as to the development and analysis of the results. All authors reviewed and approved the final manuscript.

References

1. Aloradi A., Elminshawi M., Chetupalli S.R., Habets E.A.P (2023), Target-speaker voice activity detection in multi-talker scenarios: An empirical study, [in:] *Speech Communication – 15th ITG Conference*, pp. 250–254, https://doi.org/10.30420/456164049.
2. Blanke S. (2020), Gradient-Free-Optimizers: Simple and reliable optimization with local, global, population-based and sequential techniques in numerical search spaces, https://github.com/SimonBlanke/Gradient-Free-Optimizers (access: 16.06.2024).
3. Dosovitskiy A. *et al.* (2021), An image is worth 16x16 words: Transformers for image recognition at scale, https://arxiv.org/abs/2010.11929.
4. He K., Zhang X., Ren S., Sun J. (2016), Deep residual learning for image recognition, [in:] *2016 IEEE*

Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.

5. Kim T., Chang J., Ko J.H. (2022), ADA-VAD: Unpaired adversarial domain adaptation for noise-robust voice activity detection, [in:] *ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7327–7331, https://doi.org/10.1109/icassp43922.2022.9746755.
6. Kingma D.P., Ba J. (2015), Adam: A method for stochastic optimization, [in:] *ICLR 2015*.
7. Kittler J., Hatef M., Duin R.P.W., Matas J. (1998), On combining classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(3): 226–239, https://doi.org/10.1109/34.667881.
8. Lavechin M. *et al.* (2023), Brouhaha: Multi-task training for voice activity detection, speech-to-noise ratio, and C50 room acoustics estimation, https://doi.org/10.48550/arXiv.2210.13248.
9. Ma C., Dai G., Zhou J. (2022), Short-term traffic flow prediction for urban road sections based on time series analysis and LSTM_BILSTM method, *IEEE Transactions on Intelligent Transportation Systems*, **23**(6): 5615–5624, https://doi.org/10.1109/tits.2021.3055258.
10. McFee B. *et al.* (2015), librosa: Audio and music signal analysis in python, [in:] *Proceedings of the 14th Python in Science Conference*, https://doi.org/10.25080/Majora-7b98e3ed-003.
11. Peinado A.M., Segura J.C. (2006), *Speech Recognition Over Digital Channels: Robustness and Standards*, Wiley, https://doi.org/10.1002/0470024720.
12. Rabiner L.R., Schafer R.W. (2010), *Theory and Applications of Digital Speech Processing*, Pearson.
13. Rijsbergen C.J.V. (1979), *Information Retrieval*, 2nd ed., Butterworth-Heinemann.
14. Rokach L. (2005), Ensemble methods for classifiers, [in:] *Data Mining and Knowledge Discovery Handbook*, Maimon O., Rokach L. [Eds], pp. 957–980, Springer, https://doi.org/10.1007/0-387-25465-x_45.
15. Schörkhuber C., Klapuri A. (2010), Constant-Q transform toolbox for music processing, [in:] *7th Sound and Music Computing Conference (SMC2010)*, https://doi.org/10.5281/zenodo.849741.
16. Smietanka L., Maka T. (2023), Augmented transformer for speech detection in adverse acoustical conditions, [in:] *2023 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, pp. 14–18, https://doi.org/10.23919/spa59660.2023.10274438.
17. Song S., Desplanques B., Demuynck K., Madhu N. (2022), SoftVAD in iVector-based acoustic scene classification for robustness to foreground speech, [in:] *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 404–408, https://doi.org/10.23919/eusipco55093.2022.9909938.
18. Svirsky J., Lindenbaum O. (2023), SG-VAD: Stochastic gates based speech activity detection, [in:] *ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, https://doi.org/10.1109/icassp49357.2023.10096938.
19. Team S. (2024), Silero VAD: Pre-trained enterprise-grade voice activity detector (VAD), number detector and language classifier, https://github.com/snakers4/silero-vad.
20. Wang R., Moazzen I., Zhu W.-P. (2022), A computation-efficient neural network for VAD using multi-channel feature, [in:] *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 170–174, https://doi.org/10.23919/eusipco55093.2022.9909914.
21. Yadav S., Legaspi P.A.D., Alink M.S.O., Kokkeler A.B.J., Nauta B. (2023), Hardware implementations for voice activity detection: Trends, challenges and outlook, *IEEE Transactions on Circuits and Systems I: Regular Papers*, **70**(3): 1083–1096, https://doi.org/10.1109/tcsi.2022.3225717.
22. Yang Q., Liu Q., Li N., Ge M., Song Z., Li H. (2024), SVAD: A robust, low-power, and light-weight voice activity detection with spiking neural networks, [in:] *ICASSP 2024 – 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 221–225, https://doi.org/10.1109/icassp48485.2024.10446945.
23. Zhang Y., Zou H., Zhu J. (2023), Vsanet: Real-time speech enhancement based on voice activity detection and causal spatial attention, [in:] *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 1–8, https://doi.org/10.1109/asru57964.2023.10389633.
24. Zhao Y., Champagne B. (2022), An efficient transformer-based model for voice activity detection, [in:] *2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 1–6, https://doi.org/10.1109/mlsp55214.2022.9943501.

Archives of Acoustics Vol. 50, No. 4, pp. 455–464 (2025), https://doi.org/10.24425/aoa.2025.154828

Research Paper

Effect of Listener Head Position on Speech Intelligibility in an Automotive Cabin

Linda LIANG[(1)]*, Linghui LIAO[(1)], Jiahui SUN[(1)], Lingling LIU[(1)], Liuying OU[(2)], Xiaoyue HUANG[(1)]

[(1)] *College of Civil Engineering and Architecture, Guangxi University*
Nanning, China

[(2)] *Guangxi Vocational University of Agriculture*
Nanning, China

*Corresponding Author e-mail: ldliang@gxu.edu.cn

Received November 21, 2024; revised June 9, 2025; accepted September 16, 2025; published online October 17, 2025.

When evaluating speech intelligibility (SI) in automotive cabins, binaural measurements typically employ a fixed dummy head. However, the impact of listener head positions on SI in nonuniform cabin sound fields remains unclear. This study analyzed SI under various listener head positions in an automotive cabin. An artificial mouth was regarded as the speaker, which was placed in three passenger positions. Binaural room impulse responses were measured using a dummy head in the driver's seat with various head positions. The results show that head position significantly affects SI, with variations of up to 7 dB in octave band magnitudes, more than one just-noticeable difference in the speech transmission index, and shifts of up to 2.5 dB in the speech-reception threshold. SI variability depends on the speaker's location. Directivity patterns play a crucial role in the front-passenger position, while seat occlusion affects SI at the back-right position, causing substantial decreases below a certain height threshold. At the back-left position, head positions close to the headrest enhance SI due to distance and reflections. Minor head displacements (4 cm apart) generally have insignificant effects on SI, except near seat obstructions or reach critical thresholds.

Keywords: automotive cabin; speech intelligibility; head position; speech reception threshold; speech transmission index.

1. Introduction

In recent years, the automobile has evolved from a simple means of transportation into an essential part of everyday life, often referred to as thr third space. Consequently, acoustic comfort has emerged as a notable area of concern due to increasing consumer demands (Miqueau *et al.*, 2024). Speech intelligibility (SI) is strongly associated with the level of acoustic comfort perceived by passengers within automotive cabins (Biswas *et al.*, 2022). Thus, it plays a vital role in enhancing safety and the overall travel experience.

However, the acoustic environments within automotive cabins possess unique characteristics that distinguish them from traditional rooms, thereby rendering SI in automobiles a specific concern (Parizet, 1993). The confined dimensions and intricate boundary conditions within automotive cabins result in a notable low-frequency resonance and rapid attenuation of high-frequency sounds (Granier *et al.*, 1996; Rumsey, 2016; Meissner, 2017). Many of the reflections are early reflections (Granier *et al.*, 1996; Kleiner, Tichy, 2014; Rumsey, 2016), which are considered advantageous for SI (Bradley *et al.*, 2003; Arweiler, Buchholz, 2011; Warzybok *et al.*, 2013). Consequently, the adverse effects of reverberation on intelligibility can be disregarded (Samardzic, Novak, 2011a; 2011b; Gerrera *et al.*, 2016; Ebbitt, Remtema, 2015). Furthermore, seatbacks play a pivotal role in sound absorption within automobiles

(Parizet, 1993; Visintainer, VanBuskirk, 1997; Cao *et al.*, 2022). Seat occlusions diminish speech energy transmission from rear speakers to listeners in the front (or vice versa), significantly impairing SI (Liang *et al.*, 2021). Moreover, background noise in automotive cabins has a substantial impact on SI, as it exhibits unique and fluctuating characteristics based on speed, operating conditions, and road conditions, which are absent in traditional indoor environments (Samardzic, Novak, 2011a; 2011b; Parizet, 1992; Wang *et al.*, 2012; Samardzic, 2014). The interior environment of automotive cabins demonstrates the considerable signal-to-noise ratio (SNR) variations (Dal Degan, Prati, 1988; Ferrari *et al.*, 2023). In contrast to the quieter and more constant background noise prevalent in traditional indoor settings, the SI within automotive cabins is influenced by background noise (or SNR) rather than reverberation (Ebbitt, Remtema, 2015; Samardzic, Novak, 2011a; 2011b; Gerrera *et al.*, 2016; Liang *et al.*, 2021).

Furthermore, the extremely confined dimensions of automotive cabins place the speaker and listener within the near-field zone, which further complicates the SI variations within the cabin (Liang, Yu, 2020). Specifically, SI measurements within automotive cabins are more sensitive to factors such as speaker directivity, orientation, and position compared to typical indoor environments (Bilzi *et al.*, 2005; Liang, Yu, 2023b). Moreover, binaural listening phenomena, including binaural interactions and the head shadow effect (van Wijngaarden, Drullman, 2008), introduce an effective SNR that differs between the ears of the listener (Liang, Yu, 2020). These phenomena have a direct impact on SI in automobiles. The SI in automotive cabins is strongly influenced by the direction and distance of the speaker relative to the listener's ears. The combination of the near-field head shadow effect and the unique sound field characteristics within automotive cabins (such as the nonuniform distribution of early reflections and seatback occlusions) renders SI under binaural listening conditions in automobiles more complex than in traditional indoor environments (Liang *et al.*, 2021; Liang, Yu, 2023b). Consequently, for an accurate SI evaluation within automotive cabins, it is imperative to use binaural measurements (Samardzic, Moore, 2021) and consider the orientation of the listener's head (Liang *et al.*, 2021; Liang, Yu, 2023b). Neglecting these factors can result in substantial deviations in the SI assessment.

In previous studies evaluating SI within automotive cabins, binaural signals were typically captured using a dummy head in a static position (Ebbitt, Remtema, 2015; Samardzic, Novak, 2011a; 2011b; Liang *et al.*, 2021; Liang, Yu, 2023b; Samardzic, Moore, 2021). However, the nonuniform sound pressure distribution within automotive cabins is influenced by acoustic resonances and the irregular distribution of absorptive and reflective surfaces (Granier *et al.*, 1996; Rumsey, 2016). Given the interplay between the binaural effect and the unique sound field characteristics within the confined acoustic space of an automobile, it is anticipated that the variations in listener head position would result in significant differences in the sound pressure level (SPL) experienced by the ears (Ghanati, Azadi, 2020; Granier *et al.*, 1996; Rumsey, 2016). Consequently, the SI may undergo considerable fluctuations due to the uncertainty introduced by the passenger head displacement. To the best of our knowledge, this issue has not yet been thoroughly examined.

This work aims to investigate the impact of the listener head position on SI evaluations within an automotive cabin. Specifically, the primary objective is to quantify the extent to which SI discrepancies arise due to changes in listener head positions and to establish a benchmark for SI measurements in such environments. Initially, binaural room impulse responses (BRIRs) were measured with a speaker at three passenger locations: the front passenger (FP), back left (BL), and back right (BR) seats. During these measurements, an artificial mouth was used to emulate the speaker. A dummy head was placed in the driver's seat and at various spatial locations, encompassing four different heights multiplied by five horizontal positions, resulting in a total of 20 head positions. Subsequently, the magnitude spectra of the BRIRs, speech transmission indices (STIs), and speech reception thresholds (SRTs) in Mandarin Chinese were evaluated.

2. Methods and materials

2.1. BRIR measurements

The measurements for this study were conducted within a Volkswagen Tiguan L, with dimensions of 4733 mm by 1839 mm by 1673 mm in length, width, and height, respectively. A simplified top-down view of the automobile is depicted in Fig. 1. To streamline the analysis and focus on prevalent scenarios, the listener was in the driver's seat for this study, represented by a dummy head. The dummy head used in this study is a statistical shape model-based average head model (SSMAH) (Wang, Yu, 2025) created from 100 Chinese adults (74 men and 26 women). The dummy head's primary components, including the torso, head, and shoulders, were fabricated using ABS plastic, while the pinnae were crafted from silicone rubber. To rigorously investigate the impact of head position on SI evaluation results, this study excluded the consideration of head displacement resulting from seat adjustments, which could potentially alter the sound field within the automotive cabin. To ensure stability during measurements, the seat was securely fixed in place.

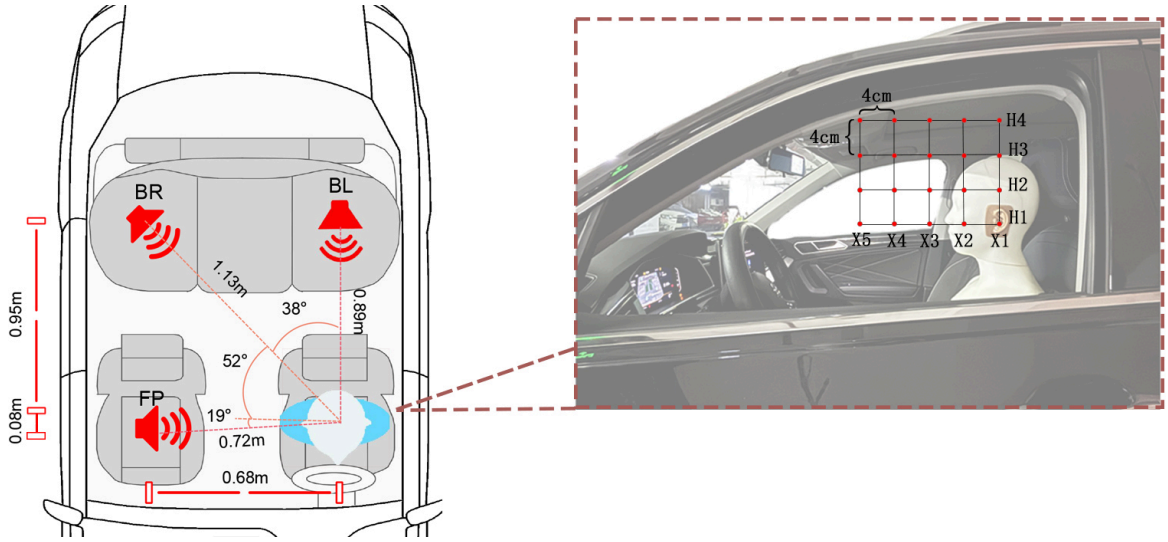

Fig. 1. Schematic of the experimental setup in the automotive cabin.

To streamline the problem and align with typical scenarios, the analysis focuses solely on the front-back and up-down dimensions of the listener's head position. A thick rectangular plastic plate with markings was laid horizontally on the driver's seat to maintain the dummy head's uniform movement in the horizontal plane. Following the measurement of one height, a 4 cm thick plastic plate was added to facilitate the dummy head's movement in the vertical direction. Consequently, the dummy head was positioned centrally in the left-right dimension of the driver's seat, facing forward. The ear canal entrance of the dummy head was systematically placed in 20 distinct locations, comprising 4 vertical levels (designated as H1 to H4, representing various heights above the seat cushion) and 5 horizontal points (designated as X1 to X5, representing different distances from the headrest). Each position was spaced 4 cm apart, as illustrated in Fig. 1. The precise location of the ear canal entrance was calibrated using a 3D laser calibrator (LSG686SPD), positioned outside the side window of the automobile. The positioning of the dummy head's head at the H1 height signifies that its ear canal entrance was 1.22 m above the ground plane.

The experiment used an artificial mouth (GRAS 44AB) as the speaker, which exhibited comparable directivity and frequency response characteristics to a human mouth. It is important to highlight that the GRAS 44AB, as initially outlined in its product documentation, was designed primarily for testing telephone mouthpieces and comparable microphones within communication systems, intended specifically for the close-proximity use. The directivity pattern of this artificial mouth might not perfectly match that of a human speaker at slightly longer distances. Nonetheless, considering that the automotive cockpit environment, which this study examines, inherently represents a unique near-field range, the influence of minor variations in directivity is expected to be relatively minor. The speaker was sequentially placed in the FP, BR, and BL locations, with its front consistently oriented towards the listener (refer to Fig. 1). The speaker was placed at a height of 1.28 m above the ground; a value determined through measurements of the mouth height of a representative sample of Chinese males with an average stature of 1.70 m. When at the FP, BR, and BL locations, the speaker was arranged at distances of 0.68 m, 1.13 m, and 0.89 m, respectively, from the listener occupying the (H1, X1) coordinate. Furthermore, the speaker was oriented at approximate angles of −19°, 52°, and 90° to the right of the listener's position.

During the measurements, all windows, doors, and the automotive air conditioning system were meticulously closed to eliminate extraneous noise. A maximum-length sequence, characterized by a 48 kHz sampling frequency, a duration of 6 s, and 24-bit quantization, served as the excitation signal. This signal was converted from digital to analog format using the Roland Studio Capture 1610 sound card and subsequently fed to the speaker. To capture the binaural signals, a pair of DPA 4060 miniature microphones were precisely positioned within the occluded ear canal entrances of the dummy head. Following this, the noise-free BRIRs were derived through deconvolution using cross-correlation between the original excitation signal and the recorded binaural signals.

2.2. STI calculation

Previous studies have comprehensively established that STI can effectively predict SI within automotive cabins (Samardzic, Novak, 2011a; Liang *et al.*, 2021), despite the negligible temporal characteristics of

the transfer function within these environments. The modulation transfer function represents the intelligibility interference arising from the temporal modulation reduction by the transmission system, as outlined in previous research (International Electrotechnical Commission [IEC], 2011; Houtgast *et al.*, 1980). Using the indirect method theory (Schroeder, 1981; Rife, 1992), the STI computation necessitates only the acquisition of the impulse response (refer to BRIRs in Subsec. 2.1) and the SNR.

Varying listener head positions and speaker positions can introduce significant variations in the speech levels received by each ear, potentially leading to discrepancies in the SNR perceived by the listener. To ensure a consistent transfer function, the BRIRs were employed to indirectly obtain the binaural speech signals. To create a monaural speech sample, pink noise was first generated and then filtered using the Chinese standard spectrum specified in (GB/T 7347-1987, 1987). Subsequently, the corresponding BRIRs obtained in Subsec. 2.1 were convolved with the monaural speech sample to produce the binaural speech signals. Additionally, background noise was sourced from actual measurements of binaural noise at 100 km/h within a fuel-powered vehicle. In reality, the SPLs of the binaural noise signals were very close between the left and right ears, with a difference of less than 0.3 dB(A). Using stationary noise and the binaural speech signals, the SNRs were indirectly derived for different listener head positions and speaker positions. Typically, it is necessary to measure the SPLs of both the noise and the binaural speech signal independently for determining the SNR needed to calculate the STI. In this study, both the speech signal and noise were considered virtual signals, making the SNR a relative value. This approach is primarily used to emphasize the changes in STI resulting from variations in the transfer function due to different head and speaker positions. Consequently, we select an appropriate value to observe the trend of STI changes. Then, the STI was calculated using the SNRs and BRIRs through the indirect STI approach, as specified in the IEC (2011) standard.

Actually, STI is essentially a monaural model. According to (IEC, 2011), when performing binaural STI measurements using an artificial head, the recommended approach is to use the results of STI for the better ear, i.e., selecting the better (larger) value from the pair of STIs. In practice, the better-ear STI cannot fully show the benefit of listening with two ears. To date, no standard for combining different STI measurements from the two ears has been developed, and so the advantages of binaural hearing in SI are always disregarded (van Wijngaarden, Drullman, 2008; Liang *et al.*, 2022). Nonetheless, the better-ear STI is still the most effective indicator compared with the existing binaural STI model (van Wijngaarden, Drullman, 2008), thus it is adopted in the present study.

2.2.1. Subjective experiment

In practice, the STI falls short of accurately capturing the impact of binaural hearing and low-frequency components on SI within the confined space of an automotive cabin, offering merely a partial estimation of the true SI level, as noted in (van Wijngaarden, Drullman, 2008; Huang *et al.*, 2023). To address this limitation, a supplementary subjective experiment was conducted to obtain the SRTs, defined as the SNR required for 50 % intelligibility. This subjective experiment encompassed 36 test conditions, factoring in 3 distinct speaker positions (namely, FP, BR, and BL) and 12 head positions, which were determined by 4 different heights (namely, H1–H4) and 3 horizontal coordinates (namely, X1, X3, and X5).

2.2.2. Participants

For this study, 12 volunteers were recruited, comprising an equal gender distribution with 6 males and 6 females. The study participants ranged in age from 20 to 25 years, with a mean age of 21.83 years. They were drawn from a pool of undergraduate and graduate students at Guangxi University. Each participant reported normal hearing and was a native Mandarin-speaking Chinese individual hailing from diverse geographical regions. As a token of appreciation, participants were compensated for their involvement in the study.

2.2.3. Stimuli and procedure

The Mandarin Chinese matrix sentence test served as the source of sentences comprising the target speech, as reported in (Hu *et al.*, 2018). Each sentence within the corpus adhered to a pre-established syntactic structure containing five words: *name*, *verb*, *number*, *adjective*, and *object*. This framework was both grammatically correct and semantically unpredictable. A total of 40 lists, each containing 20 sentences, was used. Employing auralization technology, the target speech was emitted from various passenger locations to the driver's position by convolving it with the corresponding BRIR. The interfering signal was the binaural noise captured in authentic automotive environments, which had previously been used for STI calculations. The interferer's level was set to approximately 60 dB(A) for each ear to ensure comfort. Prior to convolution with the BRIR, the target speech's level was adjusted to achieve different SNRs. The binaural interferer was then combined with the convolved binaural target speech to produce the final binaural signals.

The experiment was conducted in a room with ambient noise levels below 30 dB(A). An adaptive up-

down method, initiated with an SNR of 10 dB, was used to measure the SRTs, following the procedure outlined in (Brand, Kollmeier, 2002). Notably, the SNR referred to the difference between the interferer and target speech levels prior to convolution, rather than the actual SNR at the listener's ears. Stimuli were presented using Sennheiser HD650 headphones powered by a Roland Studio Capture 1610 sound card. Participants were instructed to independently mark the terms they heard and understood on a MATLAB graphical user interface (GUI) during the close-set response format assessments. Across the 36 test conditions considered, each participant completed a total of 72 runs, with each test condition repeated two times. The final SRT was obtained by averaging the SRTs from the two repetitions. Given that the total number of runs (72) exceeded the number of lists (40), some lists were used twice. However, this had no effect on the outcomes, as the corpus was designed to be semantically surprising and suitable for two uses by the listener, as noted in (Hu *et al.*, 2018). Participants were presented with random lists and test conditions in different sequences. To minimize listener fatigue, the 72 runs, each lasting 4 to 4.5 minutes, were divided into two sessions spaced at least 12 hours apart, with a 20-minute break after every eight runs. Each session began with a training period.

3. Results and discussion

3.1. Effect of listener head position on BRIR magnitude spectra

The magnitude spectra of the BRIRs measured in Subsec. 2.1 were computed for diverse head positions, with the speaker at the FP, BR, and BL locations. Figure 2 illustrates the variations in the magnitude spectra of the measured BRIRs relative to the reference position (H3, X3), encompassing results across the 125 Hz–8000 Hz octave bands as well as the overall results. Table 1 provides the range of magnitude variations among the 20 different head positions.

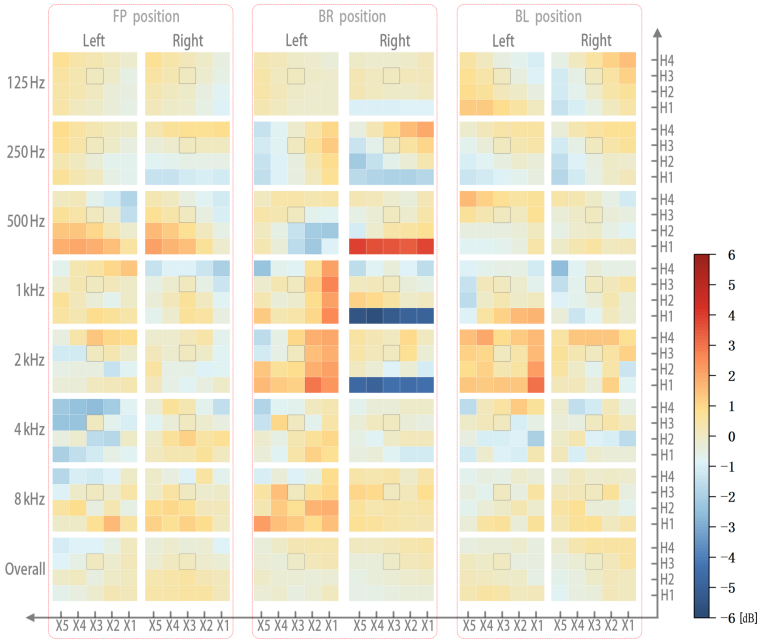

Fig. 2. Variations in the magnitude spectra (125 Hz–8000 Hz octave bands and the overall magnitude) of measured binaural room impulse responses (BRIRs) under different head positions compared to the reference position (X3, H3), when the speaker was located in the FP, BR, and BL positions.

Table 1. Magnitude variation among 20 head positions, including the 125 Hz–8000 Hz octave bands and the overall magnitude.

Speaker	Ear	Magnitude variation [dB]							
		125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	Overall
FP position	Left	1.38	1.55	3.82	2.08	2.76	2.61	3.30	1.03
	Right	1.63	2.31	3.25	2.63	1.49	2.64	1.96	1.26
BR position	Left	0.69	2.66	2.70	4.99	4.48	2.98	3.18	0.76
	Right	1.23	4.15	4.99	6.85	5.92	1.28	1.07	1.01
BL position	Left	2.28	1.89	2.41	3.46	3.12	3.24	1.43	1.36
	Right	2.95	2.58	1.65	2.91	2.64	2.05	1.50	1.40

As depicted in Fig. 2 and Table 1, irrespective of the speaker's location, the overall magnitude variation induced by the displacement of the listener's head falls within a range of 1 dB to 1.5 dB. Specifically, when the speaker is at the BL location, the magnitude variation is slightly more pronounced compared to other speaker positions, whereas it is minimal when the speaker is situated at the BR location. Furthermore, there are disparities in the magnitude fluctuations between the two ears across various listener head positions. Notably, the magnitude difference for the ipsilateral ear (i.e., the right ear in the case of the FP and BR speaker positions) is significantly greater than that for the contralateral ear, with a difference approaching 0.2 dB.

Regarding the outcomes observed for each octave band, the magnitude difference resulting from the displacement of the listener's head is more pronounced, occasionally approaching a value of 7 dB (as shown in Table 1). Notably, the octave bands ranging from 500 Hz to 4000 Hz exhibit larger magnitude variations compared to other frequency bands. When the speaker is at the BR location, the magnitude recorded in the right ear at the H1 height is significantly reduced in most frequency bands below 4 kHz (excluding the 500 Hz band) when compared to higher heights (H2 to H4). This reduction can be attributed to the direct obstruction of sound waves emitted by the speaker at the BR position by the driver's seatback when the listener's ear canal is at the H1 height. The opposite trend observed in the 500 Hz octave band may be due to standing wave phenomena within the automotive cabin, as the resonance zone typically falls within the 1 kHz range (Rumsey, 2016). Indeed, the standing wave phenomenon within the cabin often results in inconsistent trends in magnitude variations within the 125 Hz to 500 Hz octave bands compared to higher frequency bands (as depicted in Fig. 2).

3.2. Effect of listener head position on STI results

Figure 3 illustrates the better-ear STIs recorded using a dummy head at various head positions, encompassing the results obtained when the speaker was situated at the FP, BR, and BL positions. When the speaker is at the FP location, the overall variation in the STI value resulting from the listener's head position remains within 0.024, which is below the just noticeable difference (JND, 0.03) as reported in (Bradley *et al.*, 1999). When the head is at the horizontal coordinate X3, the STI value tends to be higher at the same vertical level compared to other head positions (see Fig. 3a). This is primarily because in the X3 coordinate, the listener's head is positioned nearest to the principal axis direction of the artificial mouth (speaker), as well as is closer to it. Furthermore, the radiation characteristics of the artificial mouth dictate that the radiation intensity peaks in the direction of its principal axis (Liang, Yu, 2023a).

When the speaker is at the BR location, the overall variation in the STI value due to the changes in listener head position is as high as 0.043, exceeding 1 JND. This variation is significantly larger compared to the scenario where the speaker is located at the FP position. The fluctuation in STI is primarily evident in the substantial difference between the heights of H1 and H2. Conversely, the differences among the heights of H2, H3, and H4 remain within 0.01 (as illustrated in Fig. 3b). This observation aligns with the BRIR magnitude results depicted in Fig. 2, which is attributed to the direct obstruction caused by the driver's seatback.

When the speaker is positioned directly behind the listener, specifically in the BL position, the overall variation in the STI value is 0.033, exceeding 1 JND. This variation is slightly lower than that observed in the BR position but slightly higher than that in the FP position. For head positions situated closer to the headrest, such as X1, there is a tendency for larger STI values, particularly at the H1 and H2 heights, as depicted in Fig. 3c. This phenomenon may be attributed to the proximity of the listener to the speaker or the fact that the reflection area generated by the left rear window is situated closer to the headrest. Relevant insights can also be gleaned from previous results for the magnitude spectra, specifically the magnitude observed at the left ear in Fig. 2.

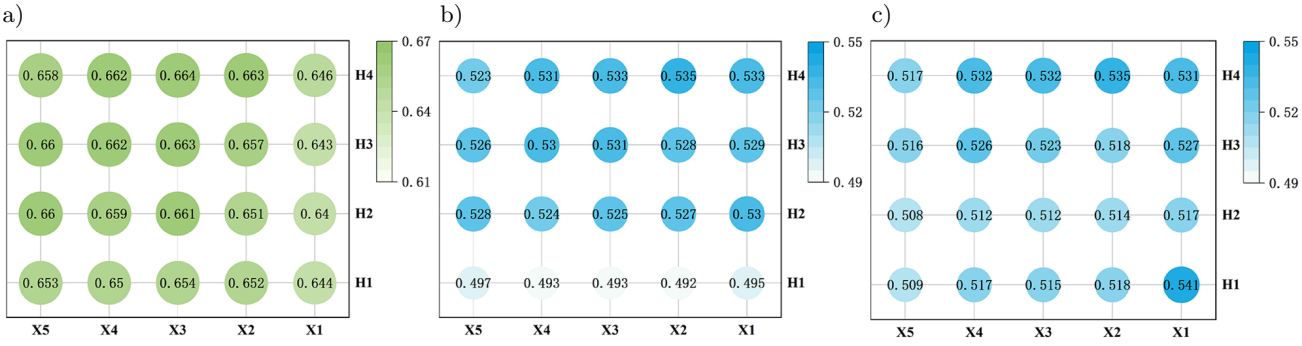

Fig. 3. Speech transmission index (STI) results obtained through a dummy head positioned at various locations when the speaker was situated in three different seats: a) the front passenger (FP); b) the back right (BR); c) the back left (BL) positions.

In general, the STI values observed when the speaker is in the FP location, ranging from 0.64 to 0.664, surpass those observed in the BR position (0.492 to 0.535) and the BL position (0.508 to 0.541). This disparity is predominantly due to the fact that, when the speaker is in the FP position, the radiated sound waves are able to reach the listener's ipsilateral (right) ear without any impediments. Conversely, for speakers in the rear (the BR and BL positions), the sound energy received at both ears is substantially diminished as a result of obstructions posed by seatbacks, as well as the listener's head and external ear (pinna) (Liang *et al.*, 2021).

3.3. Effect of listener head position on SRT results

Figure 4 illustrates the Chinese SRT results obtained when the speaker is at the FP, BR, and BL locations, along with the corresponding average values and the standard error of the mean (SEM). When the speaker occupies the FP position, the SRT value, averaging −14.9 dB, is consistently lower than that of rear-position speakers, which average more than −8.8 dB. Specifically, the SRT typically attains its highest level when the speaker is at the BL location, averaging as high as −6.9 dB, except in instances where it is occasionally lower than the BR position at the H1 height.

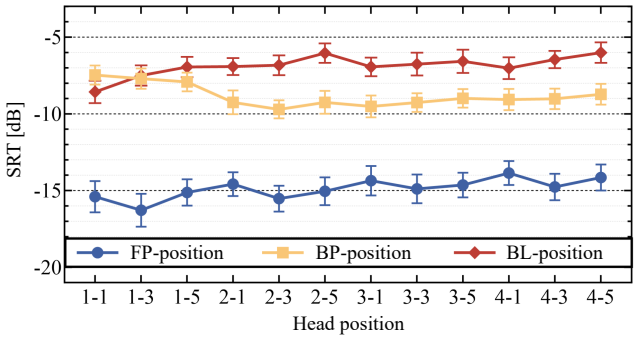

Fig. 4. Speech reception threshold (SRT) results (mean ±standard error of the mean, SEM) from measurements with the dummy head with various head positions when the speaker was in the front passenger (FP), the back right (BR), and the back left (BL) positions.

As depicted in Fig. 4a, when the speaker is located at the FP position, the SRT value fluctuates around −15 dB. Specifically, the SRT reaches a minimum of −16.3 dB at the coordinates (H1, X3) and a maximum of −13.9 dB at the coordinates (H4, X1), yielding a variation range of 2.4 dB. Furthermore, at a constant height, the SRT values recorded at the horizontal coordinate X3 are consistently lower than those at other horizontal coordinates, exemplified by the coordinates (H1, X3), (H2, X3), (H3, X3), and (H4, X3). This observation aligns with the STI results, primarily attributed to the directional characteristics of the speaker (artificial mouth), as illustrated in Fig. 3.

When the speaker is at the BR location, the SRT value ranges from −7 dB to −8 dB at the H1 height, marking a significant increase compared to other head positions. For the remaining head positions, the SRT value fluctuates around −9 dB, with a narrow variation range of approximately 1 dB (refer to Fig. 4b). This finding is in accordance with the previously presented magnitude spectra and STI results (Figs. 2 and 3), which are attributed to seat occlusion. Specifically, the SRT attains its minimum value of −9.7 dB at the coordinates (H2, X3) and its maximum value of −7.4 dB at the coordinates (H1, X1), resulting in an overall variation range of 2.3 dB.

When the speaker is at the BL location, the SRT value attains a minimum of −8.5 dB at the coordinate (H1, X1) and a maximum of −6 dB at the coordinate (H4, X5), encompassing an overall variation range of 2.5 dB (Fig. 4c). Additionally, at a fixed height, the SRT value increases as the head position moves further away from the headrest (or towards the front), suggesting a corresponding decrease in intelligibility. This pattern is consistent with the STI results, primarily due to distance factors and the arrangement of reflection areas generated by the rear side window. Overall, irrespective of the speaker position, certain displacement of the listener's head results in an SRT difference of at least 2 dB, signifying a substantial variation in SI. The SRT distinction for the FP and BL speaker positions is primarily manifested in horizontal displacement, which is influenced by the speaker's directional patterns and the distribution of reflected sound. Conversely, the BR position exhibits a primary difference due to vertical displacement, attributed to seat obstruction.

A two-way analysis of variance (ANOVA) with repeated measures indicated that the SRT results in Mandarin Chinese were significantly influenced by both the speaker position and the head position, with statistical significance demonstrated by $F(2, 22) = 422.9$ and $F(11, 121) = 5.64$, respectively (both $p < 0.0001$). Furthermore, a significant interaction was observed between these two factors for the SRT value, as evidenced by $F(22, 242) = 11.64$ ($p < 0.0001$). This suggests that the impact of head position on SRT varies depending on the speaker's position, and conversely, the influence of speaker position on SRT also varies with changes in head position.

Post-hoc pairwise comparisons, employing Bonferroni corrections, were conducted to assess differences in SRT changes across various head positions. Results indicated that when the speaker was at the FP location, statistically significant differences ($p < 0.05$) were observed between the SRT values at the coordinate (H1, X3) and other head positions, except for the coordinate (H1, X1) and (H2, X3). These differences ranged from 1.16 dB to 2.43 dB. Conversely, at heights H3 and H4, the SRT values across various head positions exhibited nonsignificant differences ($p > 0.05$),

with variations not exceeding 1 dB. When the speaker was at the BR location, a significant difference was observed solely between the H1 height and other heights. When the speaker was at the BL location, the significant differences emerged between the SRT values of the head position at the (H1, X1) coordinate and other coordinates.

3.4. Implications and limitations

From the above results and analysis, there are considerable variations in SI caused by listener's head positions in automotive cabins, especially when the speaker is in the rear. Small head displacements do not cause significant changes in SI. Significant changes in SI occur only when head displacement reaches a critical threshold or is obstructed by the seatback. Differences in SI caused by driver's heights can be ignored unless their height difference exceeds a certain limit.

These insights offer the reference value for follow-up studies and other researchers involved in binaural SI measurement. For the acoustic design of the automotive cabin, the relevant conclusions in this study emphasize the significance of optimizing seat occlusion for verbal communication between front and rear passengers. From the perspective of the front row speaker having a higher SI level for the listener in the driver seat than that of a speaker in the rear, or from the perspective of the impact of seat occlusion on head position changes, a design with a certain gap between the backrest and headrest may be more advantageous.

This study has inherent limitations, including the use of a single vehicle model and a monolingual participant group, which may restrict the generalizability of the results. Different vehicle models feature varying seat structures, which may lead to different results. Additionally, the interior space dimensions and interior materials also affect the in-cabin acoustic environment. To address these limitations, our future research plans include testing multiple vehicle models, conducting cross-lingual studies, and conducting dynamic driving scenario experiments. It is expected that these conclusions will be like those for other car models, as issues such as seat distribution and seat obstruction in cars share strong similarities. Changes in body shape and details should have a minimal impact on the conclusions drawn from this study.

4. Conclusions

To investigate the variations in speech intelligibility (SI) within an automotive cabin as a function of the listener's head position under various speaker locations, this study conducted an exhaustive analysis encompassing BRIR magnitude spectra, STI, and SRT results across various experimental conditions. The research results demonstrate that the SI within the automotive cabin was markedly influenced by the displacement of the listener's head. Across different head positions, notable differences were observed, including octave band magnitudes varying by approximately 7 dB, STI discrepancies exceeding 1 JND, and SRT shifts as substantial as 2.5 dB.

A notable interaction effect on SI was observed between the speaker position and the head position. Specifically, the magnitude of the influence that the head position exerts on SI is contingent on the speaker's position. Conversely, the effect of speaker position on SI also varies in response to alterations in the listener head position. When the speaker was at the FP position, the directivity pattern of the speaker significantly influenced the results. Horizontal coordinates that aligned closely with the speaker's principal axis direction exhibited increased the BRIR magnitude and STI values, coupled with lower SRT values, collectively indicating superior SI. At elevated head heights, the variations in SI were minimal, with SRT changes limited to within 1 dB. In scenarios where the speaker was at the BR position, a substantial decrease in SI was observed when the listener's head was positioned below a certain height threshold. This decrease was primarily attributed to direct obstruction by the seatback, resulting in a decrease in STI by 0.035 and an increase in SRT by more than 2 dB. Conversely, at unobstructed heights, the STI variations remained below 0.01, and SRT changes were generally limited to less than 1 dB. Hence, for speakers at the BR position, the influence of head position on SI was predominantly governed by seat occlusion, particularly in the vertical dimension. When the speaker was located at the BL position, head positions closer to the headrest yielded higher SI, primarily due to the combined effects of distance and reflections from the rear side window. Subjective results revealed insignificant differences among various listener head positions, i.e., the SRT variation did not surpass 1 dB, except for those at the lowest height positions.

Overall, except for head positions at lower heights when the speaker was at the BR position, the differences in STI and SRT values between adjacent measurement points (spaced 4 cm apart) were minor. This suggests that during binaural measurements for SI assessment, minor head displacements do not elicit significant changes. Significant alterations in SI only occur when the head displacement reaches a critical threshold or is obstructed by the seat. This study has systematically analyzed the impact of head position on SI, and its findings offer significant value as a benchmark for future binaural assessments of SI within automotive environments.

Fundings

This work was supported by the Middle-aged and Young Teachers' Basic Ability Promotion Project of

Guangxi (2024KY0028) and the College Student Innovation and Entrepreneurship Training Program Project (S202410593082).

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

Linda Liang designed the overall research framework, reviewed and revised the manuscript, secured research funding, and provided overall guidance. Linghui Liao collected and preprocessed experimental data, conducted field/laboratory work, and organized original data/reports. Jiahui Sun and Lingling Liu led in-depth data analysis and result interpretation, verified research hypotheses, and drafted the Sec. 3. Liuying Ou advised on the research methodology and tool selection, and assisted in manuscript revision (language and citation standardization). Xiaoyue Huang completed literature review and theoretical foundation construction, and summarized research background and existing studies.

Data availability statement

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

Acknowledgments

We would like to extend our heartfelt gratitude to all the participants who contributed to this study. We also thank Prof. Guangzheng Yu of South China University of Technology for providing the Chinese head and torso model.

References

1. Arweiler I., Buchholz J.M. (2011), The influence of spectral characteristics of early reflections on speech intelligibility, *The Journal of the Acoustical Society of America*, **130**(2): 996–1005, https://doi.org/10.1121/1.3609258.
2. Bilzi P., Bozzoli F., Farina A. (2005), Influence of artificial mouth's directivity in determining speech transmission index, *Journal of the Audio Engineering Society*, http://www.aes.org/e-lib/browse.cfm?elib=13345.
3. Biswas R., Nathwani K., Hafiz F., Swain A. (2022), Optimal speech intelligibility improvement for varying car noise characteristics, *Journal of Signal Processing Systems*, **94**: 1429–1446, https://doi.org/10.1007/s11265-022-01815-x.
4. Bradley J.S., Reich R., Norcross S.G. (1999), A just noticeable difference in C_{50} for speech, *Applied Acoustics*, **58**(2): 99–108, https://doi.org/10.1016/S0003-682X(98)00075-9.
5. Bradley J.S., Sato H., Picard M. (2003), On the importance of early reflections for speech in rooms, *The Journal of the Acoustical Society of America*, **113**(6): 3233–3244, https://doi.org/10.1121/1.1570439.
6. Brand T., Kollmeier B. (2002), Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests, *The Journal of the Acoustical Society of America*, **111**(6): 2801–2810, https://doi.org/10.1121/1.1479152.
7. Cao Y. *et al.* (2022), The influence of seat occlusion on driver's binaural signal in automobile based on FEM-RTM simulation, *Journal of Physics: Conference Series*, **2185**(1): 012025, https://doi.org/10.1088/1742-6596/2185/1/012025.
8. Dal Degan N., Prati C. (1988), Acoustic noise analysis and speech enhancement techniques for mobile radio applications, *Signal Processing*, **15**(1): 43–56, https://doi.org/10.1016/0165-1684(88)90027-8.
9. Ebbitt G.L., Remtema T.M. (2015), Automotive speech intelligibility measurements, *SAE International Journal of Passenger Cars – Mechanical Systems*, **8**(3): 1120–1127, https://doi.org/10.4271/2015-01-2337.
10. Ferrari C., Cheer J., Mautone M. (2023), Investigation of an engine order noise cancellation system in a super sports car, *Acta Acustica*, **7**: 1–9, https://doi.org/10.1051/aacus/2022060.
11. GB/T 7347-1987 (1987), *The standard spectrum of Chinese speech*, National Standard of China.
12. Gerrera G.C., Donoso-Garcia P.F., Medeiros E.B. (2016), Intelligibility in low-cost automotive audio systems, *Journal of the Audio Engineering Society*, **64**(5): 320–331, http://www.aes.org/e-lib/browse.cfm?elib=18137.
13. Ghanati G., Azadi S. (2020), Decentralized robust control of a vehicle's interior sound field, *Journal of Vibration and Control*, **26**(19–20): 1815–1823, https://doi.org/10.1177/1077546320907760.
14. Granier E. *et al.* (1996), Experimental auralization of car audio installations, *Journal of the Audio Engineering Society*, **44**(10): 835–849, http://www.aes.org/e-lib/browse.cfm?elib=7882.
15. Houtgast T., Steeneken H.J.M., Plomp R. (1980), Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics, *Acustica*, **46**(1): 60–72.
16. Hu H., Xi X., Wong L.L.N., Hochmuth S., Warzybok A., Kollmeier B. (2018), Construction and evaluation of the Mandarin Chinese matrix (CMNmatrix) sentence test for the assessment of speech recognition in noise, *International Journal of Audiology*, **57**(11): 838–850, https://doi.org/10.1080/14992027.2018.1483083.
17. Huang W., Peng J., Xie T. (2023), Study on Chinese speech intelligibility under different low-frequency

characteristics of reverberation time using a hybrid method, *Archives of Acoustics*, **48**(2): 151–157, https://doi.org/10.24425/aoa.2023.145229.

18. International Electrotechnical Commission (2011), Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index (IEC Standard No. 60268-16:2011), https://webstore.iec.ch/en/publication/1214.
19. Kleiner M., Tichy J. (2014), *Acoustics of Small Rooms*, CRC Applied Mathematics Research Press, Boca Raton, FL.
20. Liang L., Yu G. (2020), Binaural speech transmission index with spatialized virtual speaker in near field: Distance and direction dependence, *The Journal of the Acoustical Society of America*, **148**(2): EL202–EL207, https://doi.org/10.1121/10.0001808.
21. Liang L., Yu G. (2023a), The combined effect of source directivity and binaural listening on near-field binaural speech transmission index evaluation, *Applied Acoustics*, **203**: 109198, https://doi.org/10.1016/j.apacoust.2022.109198.
22. Liang L., Yu G. (2023b), Effect of speaker orientation on speech intelligibility in an automotive environment, *Applied Acoustics*, **205**: 109269, https://doi.org/10.1016/j.apacoust.2023.109269.
23. Liang L., Yu G., Zhou H., Meng Q. (2022), Effect of listener head orientation on speech reception threshold in an automotive environment, *Applied Acoustics*, **193**: 108782, https://doi.org/10.1016/j.apacoust.2022.108782.
24. Liang L., Yu L., Zhao T., Meng Q., Yu G. (2021), Speech intelligibility for various head orientations of a listener in an automobile using the speech transmission index, *The Journal of the Acoustical Society of America*, **149**(4): 2686–2697, https://doi.org/10.1121/10.0004265.
25. 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.
26. Miqueau V., Parizet E., Germes S. (2024), Influence of glazing on sound quality in the car: Validation of auralizations obtained from SEA calculations, *Acoustics Australia*, **52**: 145–159, https://doi.org/10.1007/s40857-024-00315-1.
27. Parizet E. (1992), The influence of speech importance function upon articulation index computation in cars, *Noise Control Engineering Journal*, **38**(2): 73–79, https://www.ingentaconnect.com/content/ince/ncej/1992/00000038/00000002/art00003?crawler=true.
28. Parizet E. (1993), Speech intelligibility mappings in a car compartment, *International Journal of Vehicle Design (IJVD)*, **14**(2/3): 132–144, https://doi.org/10.1504/IJVD.1993.061830.
29. Rife D.D. (1992), Modulation transfer function measurement with maximum length sequences, *Journal of the Audio Engineering Society*, **40**(10): 779–790, http://www.aes.org/e-lib/browse.cfm?elib=7032.
30. Rumsey F. (2016), Automotive audio: They know where you sit, *Journal of the Audio Engineering Society*, **64**(9): 705–708, http://www.aes.org/e-lib/browse.cfm?elib=18378.
31. Samardzic N. (2014), The applicability of the objective speech intelligibility metrics for vehicle interior speech intelligibility evaluation, considering different listening configurations and background noise spectra, *SAE International Journal of Passenger Cars – Mechanical Systems*, **7**(1): 434–438, https://doi.org/10.4271/2014-01-9126.
32. Samardzic N., Novak C. (2011a), In-vehicle speech intelligibility for different driving conditions using the speech transmission index, *Noise Control Engineering Journal*, **59**(4): 397–407, https://doi.org/10.3397/1.3598380.
33. Samardzic N., Novak C. (2011b), In-vehicle application of common speech intelligibility metrics, *International Journal of Vehicle Noise and Vibration (IJVNV)*, **7**(4): 328–346, https://doi.org/10.1504/IJVNV.2011.043193.
34. Samardzic N., Moore B.C.J. (2021), Binaural speech-to-noise loudness ratio at the speech reception threshold in vehicles, *Noise Control Engineering Journal*, **69**(2): 173–179, https://doi.org/10.3397/1/376917.
35. Schroeder M.R. (1981), Modulation transfer functions: Definition and measurement, *Acta Acustica united with Acustica*, **49**(3): 179–182.
36. van Wijngaarden S.J., Drullman R. (2008), Binaural intelligibility prediction based on the speech transmission index, *The Journal of the Acoustical Society of America*, **123**(6): 4514–4523, https://doi.org/10.1121/1.2905245.
37. Visintainer A.P., VanBuskirk J.A. (1997), Effects of sound absorption on speech intelligibility in an automotive environment, *SAE*, Technical Paper 971881, https://doi.org/10.4271/971881.
38. Wang D.F., Tan G.P., Chen S.M., Jiang J.G., Su L.L. (2012), Research on speech intelligibility of sound field inside passenger car, *Advanced Materials Research*, **403–408**: 5214–5219, https://doi.org/10.4028/www.scientific.net/AMR.403-408.5214.
39. Wang Y., Yu G. (2025), Typicality analysis on statistical shape model-based average head and its head-related transfer functions, *The Journal of the Acoustical Society of America*, **157**(1): 57–69, https://doi.org/10.1121/10.0034746.
40. Warzybok A., Rennies J., Brand T., Doclo S., Kollmeier B. (2013), Effects of spatial and temporal integration of a single early reflection on speech intelligibility, *The Journal of the Acoustical Society of America*, **133**(1): 269–282, https://doi.org/10.1121/1.4768880.

Archives of Acoustics Vol. 50, No. 4, pp. 465–476 (2025), https://doi.org/10.24425/aoa.2025.154827

Research Paper

Localization of Virtual Sound Source Reproduced by the Crosstalk Cancellation System Under Different Reflective Conditions

Wei TAN, Guangzheng YU, Jun ZHU, Dan RAO*

Acoustic Laboratory, School of Physics and Optoelectronics, South China University of Technology
Guangzhou, China

*Corresponding Author e-mail: phdrao@scut.edu.cn

Received November 10, 2024; revised July 28, 2025; accepted August 30, 2025; published online October 9, 2025.

This study explores the localization of virtual sound source reproduced by the crosstalk cancellation system under different reflective conditions in virtual rooms and analyzes the localization results with binaural cues. Binaural room impulse responses are generated using the high-order image source method. By modifying the acoustic parameters of the virtual room to manipulate the intensity and temporal structure of the reflection, psychoacoustic experiments were conducted using headphone reproduction. The experimental results indicate that, changes in reflection intensity within a certain range by altering the room reverberation time (RT) do not cause noticeable variations in virtual source localization. Increasing the loudspeaker–listener distance (changing temporal structure of reflections) deteriorates localization performance. The primary distinction between variations in the loudspeaker–listener distance and RT lies in whether the temporal structure of the reflection component changes. Overall, the study highlights the importance of the reflection temporal structure in the virtual source localization. The analysis of binaural cues indicates that, even in reverberant environments, the interaural time difference exhibits greater consistency with localization than the interaural level difference.

Keywords: sound localization; crosstalk cancellation; reflection environment; binaural cues.

1. Introduction

Binaural reproduction attempts to accurately reconstruct the desired auditory events in the listener's ear. Through binaural reproduction, listeners can perceive the spatial impression of acoustic scenes that have been recorded elsewhere or synthesized. This technique is commonly employed in immersive virtual reality (Lentz, 2008; Villegas, 2015). Both headphones and loudspeakers can be used to reproduce binaural signals. For loudspeaker reproduction, the crosstalk phenomenon inevitably occurs between the loudspeaker and the listener's ears. Crosstalk is defined as the sound transmitted from one loudspeaker to the opposite ear, which always deteriorates the sound source localization performance and timbre (Masiero *et al.*, 2011). Crosstalk cancellation (CTC) filters eliminate or reduce the contributions from crosspaths (Gardner, 1998). These filters are typically derived by inverting the head-related transfer function (HRTF) matrices, indicating that the CTC system is best suited to anechoic environments. However, in practical applications, CTC systems are routinely employed in various acoustic environments, such as listening rooms, offices, living rooms, etc. The reflections in actual reproduction environments may disrupt binaural cues, i.e., interaural time difference (ITD) and interaural level difference (ILD).

Regarding the localization performance, researchers have mainly focused on the influence of low-order reflections on the direction localization of the CTC system. By simulating the low-order reflection from an wall with different distances, it is possible to investigate the impact on virtual sources reproduced by the CTC system, with the results explained in terms of binaural cues (Kosmidis *et al.*, 2014; Sæbø, 2001; Tan *et al.*, 2023). Other studies have also investigated the localization performance of reflections on the CTC system through loudspeaker experiments under multiple reflective surfaces (Bahri, 2019; Sæbø, 1999).

In general, existing studies have been restricted to simplified reflection situations, investigating only a limited order of reflections without considering the realistic situation of a full sequence of reflections. On the other hand, despite many studies that have investigated the localization of real sources in reflective environments (Blauert, 1997; Brown *et al.*, 2015; Hartmann, 1983; Rakerd, Hartmann, 2010), differences exist in the sound field generated by virtual and actual sound sources. For a virtual source under reflective conditions, the reflections are determined by the position, input signal intensity, and phase of the loudspeakers reproducing the virtual source (Tan *et al.*, 2023). However, in the case of a real source, the reflections are only determined by the real source itself. Therefore, the binaural signal received by the listener differs significantly in the two cases, potentially resulting in localization disparities. Given this, a systematic study on the localization performance of virtual sources reproduced by CTC systems under different reflective conditions is essential.

This study aims to examine the localization of virtual sources reproduced by the CTC system under varying reflective conditions within enclosed spaces. Although the geometric dimensions of rooms, absorption boundary conditions, and other parameters are complex and varied, reflections can still be characterized by their temporal structure and intensity. Therefore, we explore the effect of reflections with varying temporal structures and intensities on localization of the virtual sound source reproduced by the CTC system, where we manipulate the reflection intensity and temporal structure by changing reverberation time (RT) of the virtual room and the distance between the listener's position and the loudspeakers. Considering that modifying the acoustic parameters in a real room and conducting virtual source localization experiments using loudspeakers are laborious and time-consuming tasks, and implementing such tests poses significant challenges. Thus, the research objectives are achieved using virtual reproduction technology (auralization) based on headphones. The crucial aspect for virtual sound reproduction based on headphones is to produce the correct binaural room impulse response (BRIR). There are two main approaches for obtaining BRIRs in different reflective environments: binaural measurement (Genuit, 1992; Li, Peissig, 2020; Møller, 1992) and simulation (Lehnert, Blauert, 1992; Møller, 1992). The measurements are relatively accurate, but it can be challenging to alter the acoustic parameters of the room. This difficulty can be solved by simulation methods, if the simulation methods are validated by numerical simulations and experimental measurements, such as the validation of the reverberation room model simulated in the ODEON program (Nowoświat, Olechowska, 2022) and room-acoustics diffusion theory (Visentin *et al.*, 2013). The image source method (ISM) is commonly used for acoustic simulations. The ISM is prevalent in architectural acoustics and provides a valuable method for evaluating a room's acoustic quality (Allen, Berkley, 1979; Habets, 2010). The localization performance of sound sources based on the BRIRs generated by the ISM and the stochastic scattering method has been validated via headphones, revealing that it is generally equivalent to the measured BRIR (Rychtáriková *et al.*, 2009).

In this study, the high-order ISM is employed to simulate the spatial room impulse responses (SRIR) in empty rectangular rooms of different sizes under various RTs and loudspeaker distances. The BRIRs under different acoustic conditions are then synthesized by the combination of SRIRs and the corresponding HRTFs. Furthermore, the BRIRs are processed by a series of CTC filters, followed by a synthesis of binaural signals at different target azimuths and conditions. The subjective experiments via headphones are conducted to examine the localization of virtual sound sources generated by the CTC system under the above acoustic conditions. The localization results are analyzed in terms of the ITD and ILD, which are calculated based on a binaural auditory model that accounts for the precedence effect, and discussed from the perspective of psychoacoustics.

The rest of this paper is organized as follows: Sec. 2 introduces the CTC system and the method of generating BRIRs; Sec. 3 conducts the experiment about virtual source localization under different reflective conditions and analyze the experimental results; Sec. 4 analyses the localization cues for experimental results; Sec. 5 conducts a discussion for the results, and finally Sec. 6 presents the conclusions to this study.

2. Simulation of the CTC system in a virtual room

2.1. CTC system

For the two-loudspeaker CTC system in an anechoic room, the transmission of sound signals is shown in Fig. 1. When the loudspeakers of the CTC system emit sound, one of the listener's ears can simultaneously receive signals from both the left and right loudspeakers. To reduce directional distortion caused by crosstalk, binaural signals should be processed through a series of CTC filters before being delivered to the loudspeakers. For the CTC system in the frequency domain, the transmission of sound signals is given by

$$\begin{bmatrix} P_L \\ P_R \end{bmatrix} = \begin{bmatrix} H_{LL} & H_{RL} \\ H_{LR} & H_{RR} \end{bmatrix} \begin{bmatrix} C_{LL} & C_{RL} \\ C_{LR} & C_{RR} \end{bmatrix} \begin{bmatrix} H_L \\ H_R \end{bmatrix} E_0, \quad (1)$$

or

$$\mathbf{P} = \mathbf{H} \cdot \mathbf{C} \cdot \mathbf{E}, \quad (2)$$

where P_L and P_R are binaural pressures in an anechoic room, respectively. H_{IJ} are the elements of $\mathbf{H}$, representing the HRTF of the I-th source to the J-th ear, where I and J denote L or R. The elements C_{IJ} of $\mathbf{C}$ are the corresponding CTC filters. The HRTF of the target virtual source are denoted as H_L and H_R. E_0 is the monaural signal and $\mathbf{E}$ represents the binaural signal.

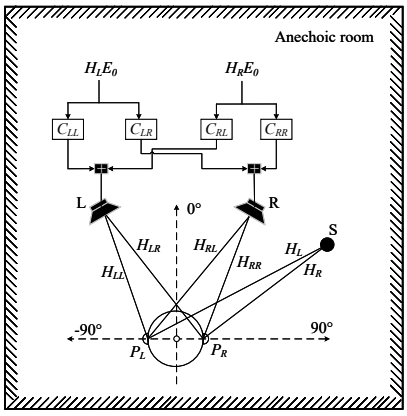

Fig. 1. CTC system in an anechoic room. The coordinate system is represented in the diagram. The virtual source is denoted by 'S'; 'L' and 'R' represent the left and right loudspeakers for reproduction.

To eliminate crosstalk, the product of $\mathbf{H}$ and $\mathbf{C}$ should equal the identity matrix $\mathbf{I}$, that is,

$$\mathbf{HC} = \mathbf{I}. \tag{3}$$

In consequence, $\mathbf{C}$ is the inverse matrix of $\mathbf{H}$. Due to HRTFs are nearly singular and cannot be inverted at some frequencies. To enhance the robustness of the solution, we utilized a regularization method to compute the matrix $\mathbf{C}$ (Kirkeby, Nelson, 1999). Consequently, $\mathbf{C}$ can be computed as the pseudoinverse of $\mathbf{H}$:

$$\mathbf{C} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}+\lambda\mathbf{I}\right)^{-1}\mathbf{H}^{\mathrm{T}}, \tag{4}$$

where the superscript T represents the conjugate transpose, λ is the regularization parameter. In Eq. (4), λ can be adjusted to 0.001 to balance accuracy and the stability of virtual source. The HRTFs used in the present work were obtained from the simulated KEMAR (Knowles Electronics Manikin for Acoustics Research) artificial head HRTF database. The spectral range of the HRTF starts at 50 Hz and extends up to 22.5 kHz with a spatial resolution of 1° and an increment of 50 Hz between each step, which was computed by the boundary element method (Katz, 2001; Rui *et al.*, 2013) as executed in Mesh2HRTF (Ziegelwanger *et al.*, 2015).

2.2. BRIR simulations

To obtain BRIRs in rectangular empty rooms with different reverberations, the high order ISM was used to generate spatial room impulse responses (SRIR). The ISM is based on the principle that a wavefront arriving from a point source and reflections from an infinite plane can be modeled as emanating from an image source. This image source can therefore be visualized as a mirror source. Consider a rectangular room with dimensions of $\{L_x, L_y, L_z\}$ and a sound source positioned at $\{s_x, s_y, s_z\}$. The relative positions of the image sources with respect to the receiver position can be written as

$$(x_i, y_i, z_i) = \begin{pmatrix} (1-2u)s_x + 2nL_x, (1-2v)s_y + 2lL_y, \\ (1-2w)s_z + 2mL_z \end{pmatrix}, \tag{5}$$

where $\{u, v, w\}$ and $\{n, l, m\}$ are integer vector triplets; u, v, and w can take values of 0 or 1, whereas the possible values of n, l, and m are based on the order of the reflections.

For simplification, only omnidirectional sound sources are considered here, and RT are used to replace the variation of sound absorption boundary conditions. Energy absorption by the walls of the room and attenuation over distance for sound propagation (Ocheltree, Frizzel, 1989) are integrated into the calculations of the impulse responses of different order image sources. In this study, the precise materials corresponding to the given absorption coefficients were not specified. For the sake of simplification, a uniform absorption coefficient was assigned to all surfaces in the simulation, thereby enabling a focused investigation of the impact of reverberation time and the delay and intensity of reflected sound.

To incorporate enough reflections in the simulation, the order of the image sources is configured to be sufficiently high, ensuring that the energy attenuation of the image source exceeds 60 dB at that order. After performing the inverse discrete Fourier transform (IDFT) on the corresponding HRTFs, the corresponding head-related impulse responses (HRIRs) are obtained. Next, the corresponding HRIRs were convolved with the impulse represented by the direct source and each image source, and the resulting responses were summed to obtain the BRIR. The process of obtaining BRIR is shown in Fig. 2.

Considering the loudspeaker angles in Fig. 1 in a virtual room, we use the binaural room transfer function (BRTF) to replace the HRTF matrix in Eq. (1). Finally, the binaural sound pressure produced by the CTC system in a virtual room can be expressed as

$$\begin{bmatrix} P'_L \\ P'_R \end{bmatrix} = \begin{bmatrix} B_{LL} & B_{RL} \\ B_{LR} & B_{RR} \end{bmatrix} \begin{bmatrix} C_{LL} & C_{RL} \\ C_{LR} & C_{RR} \end{bmatrix} \begin{bmatrix} H_L \\ H_R \end{bmatrix} E_0, \tag{6}$$

where P'_L and P'_R are the binaural sound pressures at each ear in a virtual room and B_{IJ} is the transfer function for the I-th loudspeaker to the J-th ear.

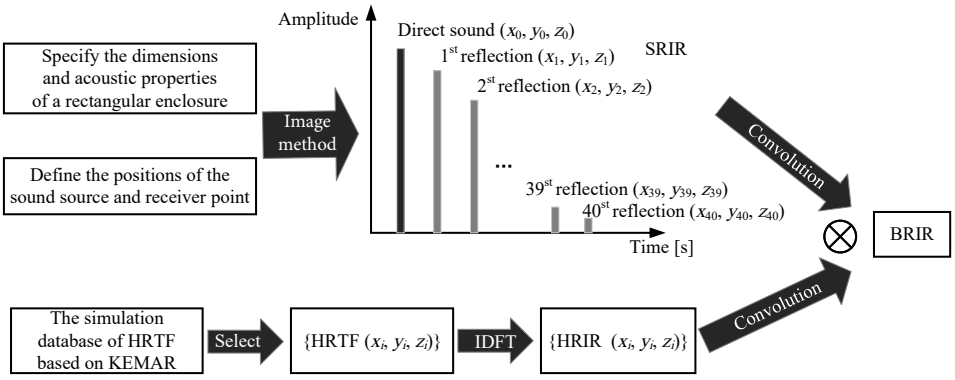

Fig. 2. Process of obtaining BRIRs.

3. Experiment: Different temporal structures and intensities reflections

The change in the reflection environment (condition) is essentially a variation in the temporal structure and intensity of the reflection. Therefore, in this section, we attempt to explore the effect of reflections of different temporal structure and intensity on the localization of virtual source reproduction by the CTC system by varying the acoustic parameters of the room and loudspeaker arrangement.

3.1. Experimental design

3.1.1. Experimental condition

Due to the fact that variations in RT and loudspeaker distance will respectively alter the intensity and temporal structure of the reflections (with intensity changing concurrently), both will also change the direct-to-reverberant energy ratio (DRR), which could potentially affect the localization of virtual sound sources. Therefore, in the present experiment, we consider controlling the RT and the loudspeaker distance to modify the intensity and temporal structure parameters of the reflections.

Although the actual room types, acoustic parameters within the rooms, and other factors are numerous and highly complex, in order to qualitatively analyze the impact of reflection intensity and temporal structure parameters on the localization of virtual sound sources reproduced by the CTC system, we selected two representative acoustical spaces of different scales for the experiments.

The empty room ① 6.4 m (length) × 5.6 m (width) × 2.7 m (height), and the empty room ② measures 8.0 m (length) × 7.0 m (width) × 3.5 m (height). The two rooms and their loudspeaker arrangements are shown in Fig. 3. The center position of the listener's head is set at a nonspecial location in the central area of the room. The initial distance between the sound source and the wall is maintained at 2 m or more to avoid the occurrence of reflections with a small delay. The sound source is positioned at a height of 1.2 m, aligning with the center of the listener's head. The arrangement angle of sound source is 60°.

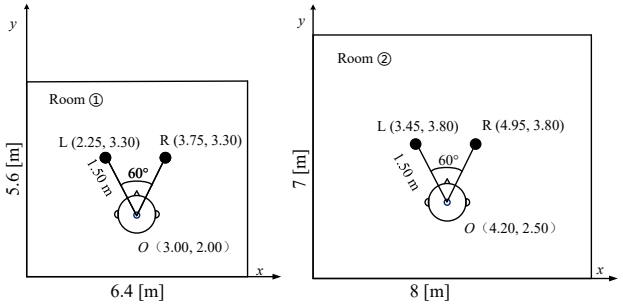

Fig. 3. Room and loudspeaker layouts. The distance between the sound source and the listeners is 1.5 m. Coordinate positions of both the listeners and the loudspeakers are shown for each listening scenario.

Experiment condition 1: different intensities of reflections. In this experiment condition, the listener position and loudspeaker layout are identical to those in the room ①. RT values of 0.3 s, 0.8 s, and 1.2 s are configured, encompassing the typical RTs of the acoustic environments used by CTC systems. Under these conditions, only the intensity of the reflections will change (as shown in Table 1, all reflections parameters are calculated relative to the direct sound). In this scenario, the minimum delay of the reflections remains

Table 1. Variation in reflection parameters due to RT changes.

RT [s]	Minimum delay [ms]	Intensity of the minimum delay reflection [dB]	Total intensity of early reflections [dB]	Total intensity of late reverberation sound [dB]
0.3	3.90	−4.8	1.4	−0.8
0.8	3.90	−3.4	5.5	4.7
1.2	3.90	−3.2	6.4	8.0

Table 2. Variation in reflection parameters due to loudspeaker distance changes.

Loudspeaker distance [m]	Minimum delay [ms]	Intensity of the minimum delay reflection [dB]	Total intensity of early reflections [dB]	Total intensity of late reverberation sound [dB]
1.50	3.90	−6.5	2.1	0.5
2.50	2.80	−3.9	6.3	4.4
3.50	2.20	−2.7	9.4	7.7

around 3.90 ms, which is within the suppression range of the precedence effect. Therefore, the temporal structure of the reflections does not change, while the intensity of the early reflections increases by approximately 5.0 dB and the late reverberation increases by about 8.8 dB.

Experiment condition 2: different temporal structures of reflections. We employ the method of changing loudspeaker distance to control temporal structures of reflections. Under this experimental condition, the size of the virtual room, the loudspeaker arrangement, and the listener's position are consistent with room ② in Fig. 3, and the RT is set to 0.7 s. The loudspeaker distances are set at 1.50 m, 2.50 m, and 3.50 m, respectively, while the loudspeaker span angle remains at 60°. As the distance of the loudspeaker increases, the minimum delay of the reflection decreases from 3.90 ms to 2.20 ms, shifting from the suppression range of the precedence effect (usually greater than 3 ms) to the range where the precedence effect begins to take effect. Additionally, the intensity of the reflection increases accordingly, as shown in Table 2. Unlike changing the RT, altering the loudspeaker distance simultaneously changes both the temporal structure and the intensity of the reflections.

3.1.2. Subjects

The experiment involved eight participants, comprising five males and three females, with ages ranging from 20 to 26 years old. All participants were Master's degree candidates. They self-reported as having typical hearing abilities and had previously engaged in sound localization studies. Compensation was provided for their involvement in the experiment.

3.1.3. Experimental procedure

The BRIRs in the virtual rooms were obtained using the method described in Sec. 2, where the image source order was set to 40. The calculations were implemented in MATLAB on a personal computer. Three stimuli were chosen: music (from Blue Danube), speech (from a Chinese corpus read by a baritone), and a 6-second duration of pink noise processed with fade-in and fade-out. The pink noise was passed through a 10 kHz low-pass finite impulse response filter and reproduced using the Etymotic Research (ER-2) insert earphone. The ER-2 earphones are inserted into the ear canal and bypass the pinna's acoustic effects, their corresponding headphone transfer function does not include pinna coloration. Given that the flat frequency response of the ER-2, no additional headphone equalization was applied. Each stimulus was presented randomly and repeated three times. The average binaural sound pressure level in the condition of the room ① was calibrated to approximately 65 dB(A). The virtual source's target azimuths were categorized into seven distinct directions, ranging from −90° to 90°, with each direction separated by 30° intervals.

Listening tests were performed in an isolated control room. Participants initially engaged in a training session, where they listened to the test stimuli, being clearly informed that the stimuli could emanate from any location within the frontal plane. Feedback on responses was not given throughout the training stage. The azimuth of the virtual source was determined using the Polhemus Fastrak G4™, a portable and mobile wireless electromagnetic tracker that achieves full 6-degrees-of-freedom localization. Each subject held a lightweight carbon fiber rod in their hand with a sensor attached at the end. When the subject heard a stimulus, they pointed the sensor towards the perceived location of the sound source. The sensor recorded the position information and transmitted it to a personal computer. After real-time processing, the subject's perceived angle was determined and recorded. The experiments were divided into three groups, i.e., different room types, different RTs, and different loudspeaker distances. Subjects are required to take a break every 15 to 20 minutes.

3.2. Experimental results

Figure 4 shows the virtual source localization results of the CTC system in rooms with different intensities of reflections (different RTs). At a ±90° target azimuth, the average perceived azimuth (absolute value) under the RT condition of 0.3 s is slightly larger than the average perceived azimuth under other RT conditions. At other target azimuths, there is no significant difference in the average perceived azimuth under different RT conditions (i.e., 0.8 s and 1.2 s). Additionally, the standard deviation of the lateral perceived azimuth under RT conditions of 0.8 s and 1.2 s is slightly higher than that under the RT condition of 0.3 s, with a difference ranging from about 1° to 4°. A multifactor repeated measures analysis of variance (ANOVA) showed that the main effects of RT and signal type

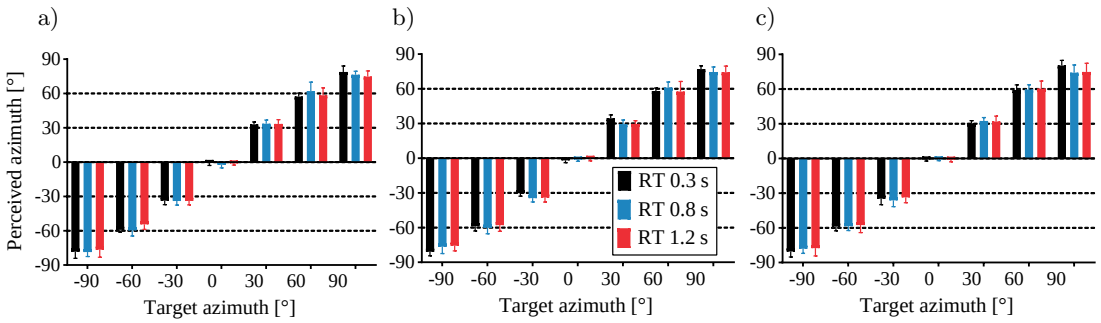

Fig. 4. Localization results at different RTs (different intensities of reflections): a) speech; b) music; c) pink noise.

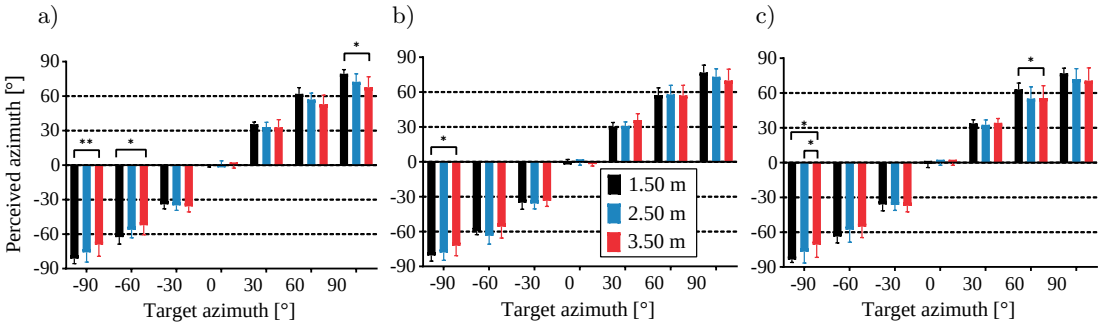

Fig. 5. Localization results of different distance condition (different temporal structures of reflections): a) speech; b) music; c) pink noise.

were not significant. Overall, the localization results indicate that even with a significant increase in the intensity of reflections, when the reproduction loudspeakers are positioned away from the wall (more than 2 m), the subjects can still locate the virtual sound source. The localization accuracy of the CTC system's reproduced virtual sound sources does not significantly decline. Therefore, when the loudspeakers are relatively far from the wall, changes in RT within a certain range, that is, changes in the intensity of the reflections (without altering the temporal structure of the reflections), do not affect the localization of the virtual sound source.

Figure 5 displays the localization results for the cases of different loudspeaker distances (different temporal structures of reflections). For front target sound sources (0°, 30°, and −30°), there is little difference in the perceived azimuth under different loudspeaker distance conditions. However, for lateral target sound sources (±60° and ±90°), the perceived azimuth (absolute value) tends to decrease with the increasing loudspeaker distance. For example, in the case of the speech signal and the 90° target azimuth, the perceived azimuths at loudspeaker distances of 1.5 m, 2.5 m, and 3.5 m are 80°, 72°, and 67°, respectively. In addition, as the loudspeaker distance is raised, there is a noticeable increase in the SD of the lateral perceived azimuths (±60° and ±90°). For instance, with a 1.5 m loudspeaker distance, the SD of lateral perceived azimuths ranges from 6° to 8°, whereas at larger loudspeaker distances, this range increases to 8° to 13°. This indicates that participants experience an increase in localization variability when localizing virtual sources at larger distances.

The perceived azimuths were subjected to multifactor repeated measures ANOVA. No significant main effects are found for either distance or signal type. However, pairwise comparisons with Bonferroni corrections show that, for the −90° target azimuth, a significant difference exists between the localization for distances of 1.5 m and 3.5 m (with different stimuli, all $p < 0.05$, refer to the asterisks in Fig. 5 for more details). For 90° and ±60° target azimuths, significant differences exist for some signals between the localization for distances of 1.5 m and 3.5 m, e.g., for speech at 90°, $p = 0.017$.

The ANOVA analysis results confirmed the previously described trends in localization changes. Specifically, at lateral target angles, the perceived azimuths are smaller under conditions of greater loudspeaker distances compared to smaller loudspeaker distances. This indicates that the temporal structure of the reflections affects the localization of virtual sound sources.

4. Localization cues analysis

4.1. Binaural auditory model

To analyze the changes in binaural cues under different reflection conditions, a binaural auditory model was introduced. The model architecture considered throughout this section is shown in Fig. 6. The binaural signal (right and left channels) was obtained by simu-

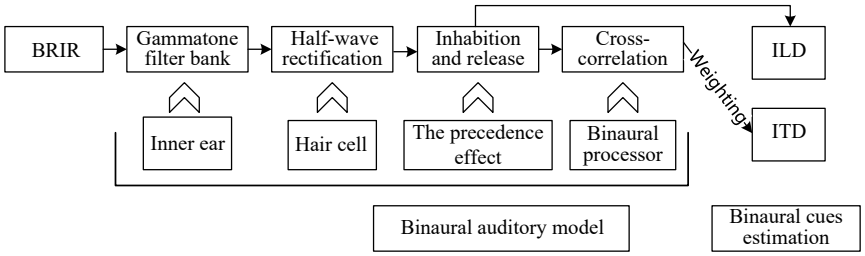

Fig. 6. Binaural auditory model structure of the cross-correlation-based precedence effect.

lation, as described in Sec. 2. The peripheral components contain the middle and inner ear. The influence of the middle ear on the localization is typically omitted, and its effect on the signal is uniform for both ears, thus leaving the ITD and ILD unaffected. The inner ear frequency selectivity was modeled using a gammatone filter bank (Slaney, 1993) of 42 bandpass equivalent rectangular bandwidth (ERB) channels. The center frequencies of the filter bank varied from 100 Hz to 10 kHz, because the main energy of the stimuli was below 10 kHz. A gammatone filter bank is often used as the front end in cochlea simulations, converting intricate sounds into multi-channel activity patterns akin to those observed in the auditory nerve. The nonlinear behavior of the hair cell was then simulated by applying half-wave rectification to the output of the gammatone filters (Braasch, 2013; Cooke, 2005).

To account for the precedence effect, suppression and release mechanisms for reflections were employed. A segmented function was adopted to fit the original function proposed in (Martin, 1997; Yost, Gourevitch, 1987). Figure 7 shows the delay-varying function of the precedence effect on the lag component in localization. In the first few milliseconds, the influence of the delayed sound diminishes as the delay increases. When the delay reaches about 3 ms, the weight is approximately 0, and this value is maintained until the delay is 15 ms. This stage corresponds to the inhibition process. As the delay continues to increase, inhibition slowly releases, and the weight gradually increases until the delay reaches 35 ms. For delays greater than 35 ms, the weights remain unchanged.

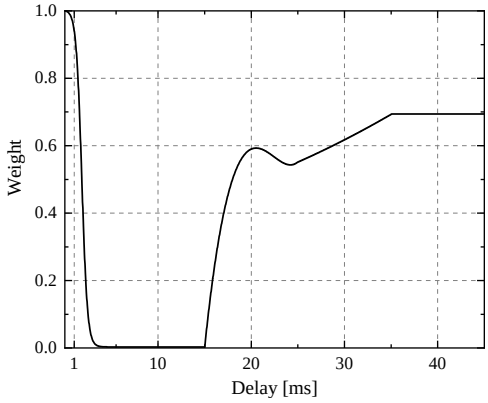

Fig. 7. Function simulating the precedence effect on lag sounds.

In this model, the stage of binaural processing occurs after the precedence effect. The binaural processor was simulated using a cross-correlation model, with the following cross-correlation function employed to obtain the ITD:

$$\Phi_{LR}(\tau) = \frac{\int_{-\infty}^{+\infty} B_{L,N}(t+\tau) B_{R,N}(t)\,\mathrm{d}t}{\left\{\left[\int_{-\infty}^{+\infty} B_{L,N}^2(t)\,\mathrm{d}t\right]\left[\int_{-\infty}^{+\infty} B_{R,N}^2(t)\,\mathrm{d}t\right]\right\}^{1/2}}, \quad (7)$$

where $B_{L,N}$ and $B_{R,N}$ represent the binaural signals of the N-th ERB channel. The range of $\Phi_{LR}(\tau)$ is from 0 to 1. This equation gives the maximum value of $\Phi_{LR}(\tau)$ in the case of $|\tau| \le 1$ ms, which represents the interaural cross-correlation coefficient (IACC). Lower IACC values (greater than 0) typically indicate a larger auditory source width, potentially resulting in an increased localization variability (SD of perceived azimuth) (Blauert, 1997; Morimoto, Iida, 1995).

4.2. Modified binaural localization cues

As described in Eq. (7), under anechoic conditions, $\tau = \tau_{\max}$ corresponding to this maximum value is defined by the ITD (Xie, 2013). Under reflective conditions, the interference between the reflected sound and the direct sound causes severe fluctuations in binaural factors with frequency variations (Kosmidis *et al.*, 2014; Tan *et al.*, 2023). This also leads to apparent multi-peak situations, where the ITD obtained from the peak corresponding to the maximum value usually has difficulty matching the actual perceived direction of the sound source. Therefore, we calculated the delay values corresponding to all peaks of the cross-correlation function and selected the one closest to the ITD value under anechoic conditions as the ITD in the reflective sound environment (i.e., choosing a reasonable ITD value) (Tollin, Henning, 1998).

The ILD is defined as

$$\mathrm{ILD}(f) = 20\log_{10}\left|\frac{P_R(f)}{P_L(f)}\right|, \quad (8)$$

where $P_R(f)$ and $P_L(f)$ represent the binaural sound pressures at frequency f.

Drawing on the auditory system's mechanism of amalgamating spatial cues across different frequency ranges, we calculated the average value and SD of the ITD below 1500 Hz (corresponding to ERB channels 1 to 21) and the ILD from 1.5 kHz to 10 kHz (corresponding to ERB channels 22 to 42). Moreover, the sensitivity of observers to the ITD in the frequency range centered around 700 Hz is widely recognized (Bilsen, 1973; Folkerts, Stecker, 2022; Zwislocki, Feldman, 1956); this frequency band is described as 'the dominance region'. Here, we set up an empirical frequency weighting function to simulate this phenomenon (Stern *et al.*, 1988). For frequencies below 1200 Hz, this function is fitted as a cubic polynomial, and for frequencies above 1200 Hz, the weight coefficients are equal to the value at 1200 Hz. The weighting function is shown in Fig. 8.

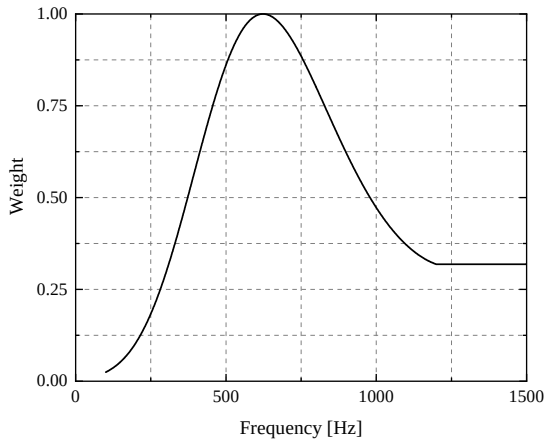

Fig. 8. Empirical frequency weighting function of ITD. The data were gathered by Raatgever (1980).

The weighted average ITDs under the different experimental conditions are shown in Fig. 9. Compared with the localization results of Figs. 4 and 7, the ITDs under these conditions exhibit analogous trends. First, regardless of the experimental conditions, the ITD increases with the target azimuth. Second, as shown in Fig. 9a, the ITD does not change with RT or the intensity of the reflections at most target azimuths. In Fig. 9b, with increasing distance, the delay of the reflection decreases, the intensity of the reflection increases, and the ITD of lateral target azimuths decreases. The above ITD trends generally align with the trends observed in localization results. However, in some cases, the ITD results do not match the localization results (e.g., at −30° under RT conditions of 0.8 s or 1.2 s). This discrepancy may be due to the general binaural auditory model not being applicable to all experimental conditions. Generally, even under larger RT conditions, ITD factors can provide relatively accurate localization information.

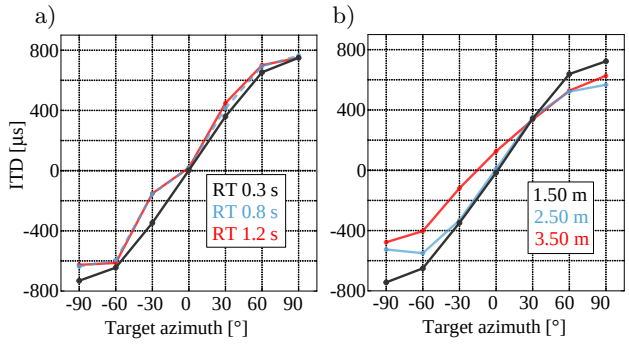

Fig. 9. Weighted average of ITDs under different: a) RTs (reflection intensities); b) distances (temporal structures).

The average ILDs at different azimuths under the different experimental conditions are shown in Fig. 10. The absolute values are significantly smaller in the higher reverberation condition than in the low reverberation condition. For example, the ILDs with RT = 0.3 s are larger than those for RT = 1.2 s or 0.8 s. This is because the late reverberant reflections can come from any direction, causing both ears to receive late reverberant energy of equal intensity. Consequently, ILD (absolute value) decreases towards zero as the DRR decreases, making it less reliable (Shinn-Cunningham *et al.*, 2005). A comparison between the results for the average ILDs and the localization results shows that there are almost no similar trends. This validates the unreliability of ILDs under low-DRR conditions.

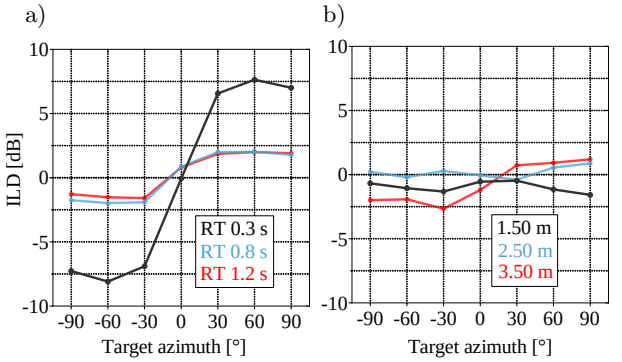

Fig. 10. Average ILD under different: a) RTs (reflection intensities); b) distances (temporal structures).

5. Discussion

5.1. Effects of reflection intensities (change RTs)

An increase in reflection intensities will decrease both the DRR and the ILD. The ILD cues (high-frequency cues) indicate that the perceived direction tends to be biased toward the front as the RT increases. However, the localization results in Sec. 3 show that the perceived azimuths are largely unaffected by changes in RT within the range of our experiments (i.e., 0.3 s–1.2 s). Similar findings have been observed in

the localization of real sound sources, where altering the RT (within the moderate reverberation range) alone did not markedly reduce the subjects' ability to localize sound sources (Rakerd, Hartmann, 2010; Rychtáriková *et al.*, 2009; 2011). This indicates that the ILD is not reliable under low-DRR conditions, as reported in a previous study (Shinn-Cunningham *et al.*, 2005). In contrast, the results in Fig. 9a demonstrate the robustness of the ITD against changes in reflection intensities (RT), which agrees with the localization results. Owing to the large distance from the loudspeaker to the wall, the earliest delay exceeds 11.40 ms (calculated based on geometric distance). In this situation, early reflections are largely suppressed by the precedence effect, resulting in little effect on localization cues. Furthermore, late reverberation adds uncorrelated signals with approximately equal amplitudes into two ears, which decreases the IACC but has little influence on the ITD. The IACC and ITD are calculated using the maximum peak value within a certain delay range of the cross-correlation function and the position at which this maximum peak value occurs, respectively. The cross-correlation function, i.e., Eq. (7), can be rewritten as

$$\Phi_{LR} = \frac{(B_{L,\mathrm{dir}} + B_{L,\mathrm{rev}}) \otimes (B_{R,\mathrm{dir}} + B_{R,\mathrm{rev}})}{|B_{L,\mathrm{dir}} + B_{L,\mathrm{rev}}||B_{R,\mathrm{dir}} + B_{R,\mathrm{rev}}|}, \qquad (9)$$

where $B_{L,\mathrm{dir}}$ and $B_{L,\mathrm{rev}}$ represent the direct sound and reverberation sound of the left impulse response, respectively, and similarly for the right impulse response. The symbol $\otimes$ denotes the correlation operation.

We hypothesize that the role of early reflection in localization is largely suppressed, and the late reverberation creates an ideal diffuse sound field. Hence, the correlation between direct and late reverberation sound, as well as the correlation with binaural late reverberation, is zero. Equation (11) can then be further simplified as

$$\Phi_{LR} = \frac{B_{L,\mathrm{dir}} \otimes B_{R,\mathrm{dir}}}{(B^2_{L,\mathrm{dir}} + B^2_{L,\mathrm{rev}})^{1/2}(B^2_{R,\mathrm{dir}} + B^2_{R,\mathrm{rev}})^{1/2}}. \qquad (10)$$

Considering our experimental conditions, the late reverberation increases with increasing RT, and so the denominator in Eq. (10) becomes larger. Moreover, the maximum peak value of the cross-correlation function decreases, indicating a decrease in the IACC (this implies a slight increase in the SD of the perceived azimuths with increasing RT). However, the position of the maximum peak remains unchanged, resulting in an unchanged ITD.

Based on the above analysis, the possible reason for the slight effect of the reflection intensities (RTs) on the localization of virtual sources are that listeners are more reliant on the ITD (low-frequency cues) than the ILD (high-frequency cues) for the localization in a reverberant environment. Previous studies have shown that subjects struggle to rely on the ITD for localization when stimuli lack transient information (Hartmann, 1983). Although the pink noise in our study was subjected to fade-in and fade-out processing, its localization does not differ significantly from other transient signals. This can be attributed to the fact that pink noise is composed of a series of small impulses, which have random amplitude fluctuations. These fluctuations are transient, meaning that the subjects are still able to utilize the ITD information within it for localization.

5.2. Effects of temporal structures of reflections (change loudspeaker distances)

For a virtual room with constant acoustic parameters, changes in loudspeaker distance will alter the temporal structure and intensity of the reflection. Under the condition of a 3.50 m loudspeaker distance, the minimum delay of the reflection is approximately 2.20 ms. This delay falls within the range where the precedence effect is actively suppressing (below 3 ms to 5 ms). At this point, the relatively high-energy early reflections are not completely suppressed by the precedence effect. A series of partially unsuppressed reflections interfere with the direct sound, causing the ITD to fluctuate with frequency. The average ITD changes with the loudspeaker distance (as shown in Fig. 9), and this interference also leads to a decrease in IACC (Gourevitch, Brette, 2012; Rakerd, Hartmann, 2010; Shinn-Cunningham, Kawakyu, 2003; Tan *et al.*, 2023). Moreover, the localization results in Sec. 3 demonstrate the same trend as the ITDs, that is, as the distance increases, the localization performance (including the SD and deviation of localization) of lateral virtual sources deteriorates. According to the auditory mechanism that merges locational data throughout various frequency bands (Hancock, Delgutte, 2004; Xia, Shinn-Cunningham, 2011), the degraded localization performance of the virtual source may arise from fluctuations in the ITD with frequency and the deviation of the mean ITD.

Variations in both the RT and loudspeaker distance change the DRR (as illustrated in Fig. 11), but only the loudspeaker distance affects the localization of virtual sources (as shown in Figs. 4 and 5). Even when the DRR is similar under different conditions, e.g., an RT of 1.2 s and loudspeaker distance of 3.5 m, there may be significant differences in the localization results. This indicates that the DRR alone may not adequately predict the localization performance in rooms. The temporal structure of reflections, i.e., the time distribution of reflection sequences, may indeed play a crucial role in the localization of the sound source. It is also reasonable to believe that reflections with small delay have a more disruptive effect on the localization of virtual sound sources compared with later reverberations with

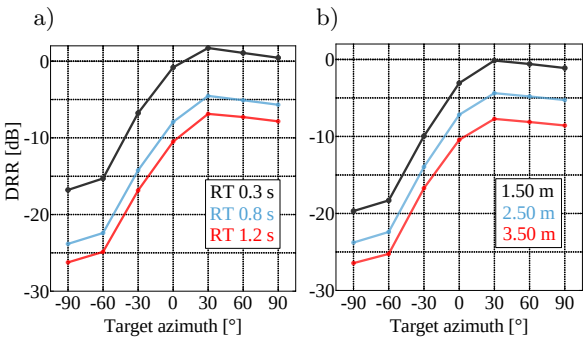

Fig. 11. DRR of the right ear under different: a) RTS (reflection intensities); b) distances (temporal structures). DRR is calculated as the ratio of the sound pressure levels between direct and reflected sound. As the DRR is obtained from the impulse response of the right ear, the DRR of the target azimuth on the right side (30° to 90°) is expected to be greater than that on the left side (−30° to −90°).

higher intensity. These findings provide the following guidance for CTC applications: even in rooms without acoustic decoration, placing the loudspeaker far enough from the wall (ensuring that the earliest delay is much longer than 1 ms) enables good localization performance of the CTC system.

6. Conclusions

This paper has investigated the influence of different temporal structures and intensity of reflections on the localization of virtual sources reproduced by a two-loudspeaker CTC system. The reasons for the variations in localization under different reflective conditions have also been revealed.

The principal conclusions derived from this study can be encapsulated in the following points:

- when the reproduction loudspeaker is located far from the wall (larger than 2 m in this work), in the RT variation range of our experiments (0.3 s to 1.2 s), the increase in the intensity of reflections does not significantly affect the localization performance of virtual sound sources due to the suppression of the precedence effect;
- when the reproduction loudspeaker distance increases (moving away from the listener and closer to the wall), the delay of early reflections decreases, and the temporal structure of the reflection changes. This results in a series of early reflections that are not fully suppressed interfering with the direct sound. This interference causes localization deviation and an increase in the degree of variation in the localization of lateral target angles of the virtual sound source;
- the DRR alone seems inadequate for determining the localization performance of virtual sources in reverberant environments. The temporal structure of reflections may play an important role in sound source localization. Compared to the late reverberation, early reflections with short delays (especially for that not fully suppressed by the precedence effect) have a greater impact on the localization of virtual sound sources;
- the average weighted ITD calculated based on the binaural auditory model accounting for the precedence effect can qualitatively explain the experimental results to some extent, but the average ILD does not.

It should be noted that, in this study, headphone-based binaural reproduction was adopted, and an acoustic simulation based on the ISM was employed. In reality, due to the material properties and geometric irregularities of room surfaces, complex absorption and diffuse reflection occur. While the ISM simplifies the modeling process and improves computational efficiency, it does not fully capture the acoustic response of real environments. Therefore, the results and conclusions presented in this study are limited to the specific experimental conditions (purely specular reflections in the room simulation and headphone reproduction) adopted herein.

Fundings

This research was funded by the National Natural Science Foundation of China with grant number 12074129 and 12474465, as well as by the Natural Science Foundation of Guangdong Province through grant number 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

These authors made equal contributions to this work. All authors reviewed and approved the final manuscript.

Acknowledgments

We express our appreciation to all participants for their contribution to the study.

References

1. Allen J.B., Berkley D.A. (1979), Image method for efficiently simulating small-room acoustics, *The Journal of the Acoustical Society of America*, **65**: 943–950, https://doi.org/10.1121/1.382599.
2. Bahri K. (2019), *Loudspeaker directivity and playback environment in acoustic crosstalk cancelation*, Msc. Thesis, Charles University of Technology, Gothenburg.

3. Bilsen F.A. (1973), Spectral dominance in binaural lateralization, *Acustica*, **28**: 131–132.

4. Blauert J. (1997), *Spatial Hearing: The Psychophysics of Human Sound Localization*, 2nd. ed., The MIT Press, Harvard MA.

5. Braasch J. (2013), A precedence effect model to simulate localization dominance using an adaptive, stimulus parameter-based inhibition process, *The Journal of the Acoustical Society of America*, **134**: 420–435, https://doi.org/10.1121/1.4807829.

6. Brown A.D., Stecker G.C., Tollin D.J. (2015), The precedence effect in sound localization, *Journal of the Association for Research in Otolaryngology*, **16**: 1–28, https://doi.org/10.1007/s10162-014-0496-2.

7. Cooke M. (2005), *Modelling Auditory Processing and Organisation*, Cambridge University Press, London.

8. Folkerts M.L., Stecker G.C. (2022), Spectral weighting functions for lateralization and localization of complex sound, *The Journal of the Acoustical Society of America*, **151**: 3409–3425, https://doi.org/10.1121/10.0011469.

9. Gardner W.G. (1998), *3-D Audio Using Loudspeakers*, Springer Science & Business Media.

10. Genuit K. (1992), Standardization of binaural measurement technique, *Le Journal de Physique IV*, **2**: 405–407, https://doi.org/10.1051/jp4:1992187.

11. Gourevitch B., Brette R. (2012), The impact of early reflections on binaural cues, *The Journal of the Acoustical Society of America*, **132**: 9–27, https://doi.org/10.1121/1.4726052.

12. Habets E.A. (2010), *Room impulse response generator*, Technische Universiteit Eindhoven, Technical Report.

13. Hancock K.E., Delgutte B. (2004), A physiologically based model of interaural time difference discrimination, *Journal of Neuroscience*, **24**: 7110–7117, https://doi.org/10.1523/JNEUROSCI.0762-04.2004.

14. Hartmann W.M. (1983), Localization of sound in rooms, *The Journal of the Acoustical Society of America*, **74**: 1380–1391, https://doi.org/10.1121/1.390163.

15. Katz B.F. (2001), Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation, *The Journal of the Acoustical Society of America*, **110**: 2440–2448, https://doi.org/10.1121/1.1412440.

16. Kirkeby O., Nelson P.A. (1999), Digital filter design for inversion problems in sound reproduction, *Journal of the Audio Engineering Society*, **47**(7/8): 583–595.

17. Kosmidis D., Lacouture-Parodi Y., Habets E.A. (2014), The influence of low order reflections on the interaural time differences in crosstalk cancellation systems, [in:] *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2873–2877, https://doi.org/10.1109/ICASSP.2014.6854125.

18. Lehnert H., Blauert J. (1992), Principles of binaural room simulation, *Applied Acoustics*, **36**(3–4): 259–291, https://doi.org/10.1016/0003-682X(92)90049-X.

19. Lentz T. (2008), Binaural technology for virtual reality, *Journal of the Audio Engineering Society*, **124**(6): 3358–3359, https://doi.org/10.1121/1.3020604.

20. Li S., Peissig J. (2020), Measurement of head-related transfer functions: A review, *Applied Sciences*, **10**(14): 5014, https://doi.org/10.3390/app10145014.

21. Martin K.D. (1997), Echo suppression in a computational model of the precedence effect, [in:] *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, https://doi.org/10.1109/ASPAA.1997.625622.

22. Masiero B., Fels J., Vorländer M. (2011), Review of the crosstalk cancellation filter technique, [in:] *Proceedings of ICSA 2011*.

23. Morimoto M., Iida K. (1995), A practical evaluation method of auditory source width in concert halls, *Journal of the Acoustical Society of Japan (E)*, **16**(2): 59–69, https://doi.org/10.1250/ast.16.59.

24. Møller H. (1992), Fundamentals of binaural technology, *Applied Acoustics*, **36**(3–4): 171–218, https://doi.org/10.1016/0003-682X(92)90046-U.

25. Nowoświat A., Olechowska M. (2022), Experimental validation of the model of reverberation time prediction in a room, *Buildings*, **12**(3): 347, https://doi.org/10.3390/buildings12030347.

26. Ocheltree K.B., Frizzel L.A. (1989), Sound field calculation for rectangular sources, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **36**(2): 242–248, https://doi.org/10.1109/58.19157.

27. Rakerd B., Hartmann W.M. (2010), Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise, *The Journal of the Acoustical Society of America*, **128**(5): 3052–3063, https://doi.org/10.1121/1.3493447.

28. Raatgever J. (1980), *On the binaural processing of stimuli with different interaural phase relations*, Ph.D. Thesis, Technische Universiteit Delft, Netherlands.

29. Rui Y., Yu G., Xie B., Liu Y. (2013), *Calculation of individualized near-field head-related transfer function database using boundary element method*, Presented at the Audio Engineering Society Convention, paper 8901.

30. Rychtáriková M., van den Bogaert T., Vermeir G., Wouters J. (2009), Binaural sound source localization in real and virtual rooms, *Journal of the Audio Engineering Society*, **57**: 205–220.

31. Rychtáriková M., van den Bogaert T., Vermeir G., Wouters J. (2011), Perceptual validation

of virtual room acoustics: Sound localisation and speech understanding, *Applied Acoustics*, **72**(4): 196–204, https://doi.org/10.1016/j.apacoust.2010.11.012.

32. Sæbø A. (1999), *Effect of early reflections in binaural systems with loudspeaker reproduction*, Presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 552–556, New York.

33. Sæbø A. (2001), *Influence of reflections on crosstalk cancelled playback of binaural sound*, Ph.D. Thesis, Norwegian University of Science and Technology, Trondheim.

34. Shinn-Cunningham B., Kawakyu K. (2003), Neural representation of source direction in reverberant space, [in:] *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, pp. 79–82, https://doi.org/10.1109/ASPAA.2003.1285824.

35. Shinn-Cunningham B.G., Kopco N., Martin T.J. (2005), Localizing nearby sound sources in a classroom: Binaural room impulse responses, *The Journal of the Acoustical Society of America*, **117**(5): 3100–3115, https://doi.org/10.1121/1.1872572.

36. Slaney M. (1993), *An efficient implementation of the Patterson–Holdsworth auditory filter bank*, Apple Computer Technical Report #35, Perception Group – Advanced Technology Group.

37. Stern R.M., Zeiberg A.S., Trahiotis C. (1988), Lateralization of complex binaural stimuli: A weighted-image model, *The Journal of the Acoustical Society of America*, **84**(1): 156–165, https://doi.org/10.1121/1.396982.

38. Tan W., Yu G., Rao D. (2023), Influence of first-order lateral reflections on the localization of virtual source reproduced by crosstalk cancellation system, *Applied Acoustics*, **202**: 109165, https://doi.org/10.1016/j.apacoust.2022.109165.

39. Tollin D.J., Henning G.B. (1998), Some aspects of the lateralization of echoed sound in man. I. The classical interaural-delay based precedence effect, *The Journal of the Acoustical Society of America*, **104**(5): 3030–3038, https://doi.org/10.1121/1.423884.

40. Villegas J. (2015), Locating virtual sound sources at arbitrary distances in real-time binaural reproduction, *Virtual Reality*, **19**: 201–212, https://doi.org/10.1007/s10055-015-0278-0.

41. Visentin C., Prodi N., Valeau V., Picaut J. (2013), A numerical and experimental validation of the room acoustics diffusion theory inside long rooms, [in:] *Proceedings of Meetings on Acoustics*, **19**(1): 015024, https://doi.org/10.1121/1.4798976.

42. Xia J., Shinn-Cunningham B. (2011), Isolating mechanisms that influence measures of the precedence effect: Theoretical predictions and behavioral tests, *The Journal of the Acoustical Society of America*, **130**(2): 866–882, https://doi.org/10.1121/1.3605549.

43. Xie B. (2013), *Head-Related Transfer Function and Virtual Auditory Display*, 2nd ed., J. Ross Publishing.

44. Yost W.A., Gourevitch G. (1987), *Directional Hearing*, Springer.

45. Ziegelwanger H., Kreuzer W., Majdak P. (2015), *Mesh2HRTF: An open-source software package for the numerical calculation of head-related transfer functions*, Presented at the 22nd International Congress on Sound and Vibration, https://doi.org/10.13140/RG.2.1.1707.1128.

46. Zwislocki J., Feldman R.S. (1956), Just noticeable differences in dichotic phase, *The Journal of the Acoustical Society of America*, **28**(5): 860–864, https://doi.org/10.1121/1.1908495.

Archives of Acoustics Vol. 50, No. 4, pp. 477–490 (2025), https://doi.org/10.24425/aoa.2025.156930

Research Paper

Subwavelength Underwater Imaging of a Wire Array Metamaterial Based on Fabry–Pérot Resonance

Guo LI[(1)*], FeiLong LI[(1)], LiQing HU[(2)], QunFeng LI[(3)], GuanJun YIN[(4)]

[(1)] *School of Automation, Xi'an Key Laboratory of Advanced Control and Intelligent Processing*
Xi'an University of Posts and Telecommunications
Xi'an, China

[(2)] *Electronic Materials Research Laboratory*
Key Laboratory of the Ministry of Education and International Center for Dielectric Research
School of Electronic Science and Engineering, Xi'an Jiaotong University
Xi'an, China

[(3)] *Jinan University*
Guangzhou, China

[(4)] *Key Laboratory of Ultrasound of Shaanxi Province*
School of Physics and Information Technology, Shaanxi Normal University
Xi'an, China

*Corresponding Author e-mail: liguo@xupt.edu.cn

Received May 20, 2025; revised October 7, 2025; accepted October 8, 2025; published online November 13, 2025.

Metamaterials with Fabry–Pérot (FP) resonance have proven effective for underwater ultrasound imaging. The propagation phenomenon can be understood as a spatial filter with linear dispersion over a finite bandwidth. However, conventional imaging techniques are constrained by the diffraction limit or rely on a strong impedance mismatch between the metamaterial and water. In this paper, we propose a columnar array metamaterial designed for underwater imaging based on FP resonances and validate the proposed design through numerical simulations. The acoustic pressure transmission coefficient, together with the normalized acoustic pressure distribution, is analyzed to quantitatively evaluate imaging quality and verify the physical effectiveness of the model. This novel structure enables deep subwavelength imaging underwater, maintaining excellent and stable imaging performance within a 0.4 kHz bandwidth centered around the operating frequency. We use air-filled metamaterials to create strong acoustic coupling and establish effective sound isolation. This approach significantly enhances imaging resolution, while optimizing energy loss at multiple interfaces, an issue in previous studies. Additionally, in contrast to resonance- or refraction-based approaches such as Helmholtz resonators or hyperlens designs, the proposed FP-resonant metamaterial offers an alternative mechanism for achieving near-field subwavelength imaging through controlled wave transmission and confinement. We also examine the influence of various parameters, such as imaging distance, incidence distance, and array periodicity, on imaging performance. The results demonstrate that the columnar array metamaterial holds great potential for underwater ultrasound imaging applications.

Keywords: Fabry–Pérot (FP) resonance; metamaterial underwater imaging; wire array metamaterial; air-filled metamaterials; finite element simulation.

1. Introduction

In recent years, the design and experimental realization of artificial metamaterials have yielded numerous extraordinary physical properties (Dong *et al.*, 2023; Kawata *et al.*, 2008). The core purpose of acoustic metamaterials is to achieve precise control of sound wave propagation through artificially designed struc-

tures, breaking the physical limitations of natural materials, and thereby enabling acoustic functionalities that are unattainable with conventional materials at specific frequencies or in particular scenarios. A holey-structured metamaterial has demonstrated potential for near-field acoustic imaging beyond the diffraction limit, due to the strong coupling between the evanescent field components of a subwavelength object and the Fabry–Pérot (FP) resonances within the holes (Amireddy *et al.*, 2017). However, the use of holey-structured metamaterial made from metals and polymers in underwater imaging faces significant challenges due to the low acoustic impedance mismatch and high viscous losses (Laureti *et al.*, 2020; Astolfi *et al.*, 2019; Estrada *et al.*, 2008; Pendry, 2000; Christensen *et al.*, 2008; Belov, Silveirinha, 2006).

To address these limitations, we introduce the concept of a 'wire array' metamaterial, fabricated from a polymer with an acoustic impedance closely matching that of water. This design creates FP resonances within the array, while the air-filled gaps between the wires enhance acoustic isolation, enabling more efficient transmission of evanescent waves for deep-subwavelength underwater imaging (Molerón, Daraio, 2015). This approach paves a way for deep-subwavelength imaging using polymer-based acoustic metamaterials underwater (Gulia, Gupta, 2019; Deng *et al.*, 2009).

It is well known that the resolution of traditional acoustic imaging devices is limited by the diffraction limit, which is half the operating wavelength, as these devices are unable to capture evanescent waves (Zhou *et al.*, 2010; Yan, Yuan, 2015; Zhang *et al.*, 2009; Ambati *et al.*, 2008). These evanescent waves carry the fine details of objects but decay exponentially with distance (Christensen, García de Abajo, 2010). To achieve subwavelength resolution beyond the diffraction limit, hyperlenses and superlenses in artificial acoustic metamaterials have garnered significant attention by enhancing the transmission of evanescent waves (Amireddy *et al.*, 2016; Simonetti, 2006). Hyperlenses are non-resonant, strongly anisotropic metamaterials that can convert evanescent waves into propagating waves (Liu *et al.*, 2007; Guenneau *et al.*, 2007; Silveirinha *et al.*, 2008). In contrast, superlenses exhibit either single-negative or double-negative acoustic characteristics, achieved by using membrane-type metamaterials or Helmholtz resonators (Li, Chan, 2004). Superlenses achieve subwavelength resolution by reconstructing evanescent components through negative-index behavior, whereas Helmholtz-resonator designs utilize resonant enhancement of local acoustic fields to improve spatial confinement. Compared to these resonance- or refraction-based mechanisms, FP-resonant metamaterials offer a pathway to realize near-field imaging through controlling wave transmission and confinement within the structure. The effectiveness of holey-structured metamaterials has been demonstrated in air. In some instances, enhanced evanescent wave magnitude has been observed due to highly anisotropic equifrequency contours.

However, traditional metal materials struggle to create a strong impedance mismatch with water. To address this, Astolfi *et al.* (2019) employed tungsten in additive manufacturing to achieve a significant acoustic impedance mismatch with water, thereby improving the propagation of evanescent waves. Nonetheless, tungsten is challenging to process and it is quite heavy, making it inconvenient for underwater applications. Consequently, several holey-structured polymer-based metamaterials utilizing FP resonance have been proposed for subwavelength imaging (Liu *et al.*, 2009). However, for underwater imaging with holey-structured metamaterials at higher frequencies, key challenges arise, including multimode coupling caused by weak acoustic impedance mismatch and high viscous losses between water and the metamaterial (Lafleur, Shields, 1995; Laureti *et al.*, 2014; 2016). Underwater imaging with holey-structured metamaterials presents unique challenges. To address it, Laureti *et al.* (2020) introduced the concept of using trapped air, where the acoustic impedance mismatch between a polymer and water is strongly enhanced when air is confined within the bulk material in a particular way. Additionally, the authors demonstrated that ultrasound imaging of broadband subwavelength apertures in water can be achieved using FP resonance. While these studies reported on polymer-based metamaterials functioning in water, acoustic coupling from the water-filled holes into the polymer is expected to degrade their performance (Laureti *et al.*, 2020).

Recent advances have also explored alternative approaches to achieve subwavelength imaging beyond FP-based metamaterials. For example, neural-network-assisted ultrasonic imaging methods, such as the back propagation neural network-total focusing method (BPNN-TFM), have demonstrated the ability to resolve defects separated by only $0.5\,\lambda$, outperforming several existing super-resolution techniques (Lin *et al.*, 2025). In the optical domain, semiconductor nanophotodetector (NPD) arrays, simulated with the multi-level multi-electron (MLME) finite-difference time-domain (FDTD) method, have achieved detection resolutions of about one-tenth of the operating wavelength, comparable to near-field scanning optical microscopy (Kim *et al.*, 2008). These studies highlight the diversity of subwavelength imaging strategies across different physical platforms and provide a broader context for situating the present metamaterial-based approach, which offers a compact and efficient solution for underwater acoustic applications.

In this paper, we propose the use of polymers with acoustic impedance closely matched to that of water as columns in our metamaterial arrays. These arrays are surrounded by air and sealed with thin cover plates on both the front and back. The proposed design addresses the performance losses typically observed between water and polymers in traditional metamaterials, while it also eliminates efficiency reductions at the water-polymer-air interfaces found in trapped-air configurations. Furthermore, the significant impedance mismatch between the polymer and the surrounding air enhances the FP resonance. This design enables the minimum feature imaging size to be optimized to $0.12\,\lambda$, compared to $0.135\,\lambda$ presented in the prior studies. Our findings demonstrate that these air-filled wire-array metamaterial exhibit outstanding imaging performance at deep subwavelength scales, along with a relatively broad bandwidth.

Through simulations, we optimized imaging frequency, distance, and incident conditions, and we also examined the effects of cover layer thickness and array periodicity on imaging quality. These insights offer valuable guidance for selecting material parameters. With the right configuration, the metamaterials can achieve optimal imaging performance and support potential practical applications. The proposed air-filled metamaterial design holds promise for several real-world applications that benefit from high-resolution underwater acoustic imaging. In the field of marine exploration, such a structure could be deployed for detailed seabed mapping and the detection of small-scale defects in underwater infrastructures. The ability to achieve subwavelength imaging in the near field allows fine structural details to be resolved, which are often blurred by the diffraction limit in conventional sonar systems. In addition, the approach is relevant to biomedical diagnostics in aqueous environments, such as high-frequency ultrasound imaging of tissues or monitoring of microscale biological processes. The strong impedance contrast between the water-like columns and the surrounding air provides efficient FP resonance, enabling enhanced focusing and improved image clarity. These capabilities suggest that the metamaterial design could serve as a compact, low-loss platform for next-generation acoustic microscopes or targeted biomedical sensing devices. Overall, positioning the proposed structure within such application-oriented contexts highlights its potential impact beyond theoretical demonstration.

2. Structural designs

The FP resonance condition describes a scenario in which an acoustic wave undergoes repeated reflections between two parallel boundaries within a cavity, and constructive interference arises when the round-trip propagation distance equals an integer multiple of the wavelength. Under this condition, acoustic energy becomes strongly confined within the cavity, resulting in resonance and enhanced transmission through the structure. In the context of acoustic metamaterials, this mechanism plays a critical role in amplifying evanescent components and thereby sustaining high-resolution imaging performance. By exploiting FP resonances, the metamaterial can overcome part of the diffraction limit and achieve subwavelength focusing or imaging in underwater environments.

Among the various sonic metamaterial designs, holey structures can achieve specific properties, such as extraordinary acoustic transmission or absorption. Previous studies have shown that, when diffraction effects can be neglected, the transmission process is primarily governed by the fundamental propagation modes within the holes. In this case, the zeroth-order transmission coefficient of an acoustic plane wave can be expressed as

$$t(\lambda,k)=\frac{4\,|S_0|^2 Y\exp\left(iq_z h\right)}{\left(1+Y|S_0|^2\right)^2-\left(1-Y|S_0|^2\right)^2\exp\left(2iq_z h\right)}, \quad (1)$$

where the parallel momentum $k=\sqrt{k_x^2+k_y^2}$ and $q_z=k=2\pi/\lambda$ is the propagation constant of the mentioned waveguide mode, $S_0=a/A$ and $Y=k_0/\sqrt{k_0^2-k^2}$ (Zhu *et al.*, 2010). Objects positioned at the input surface of the holey plate can achieve near-perfect acoustic image transfer to the output side, owing to the plate's unique waveguiding properties. In this configuration, the transmission coefficients of both the transmitted and swift waves are unity. A simple analysis of a single hole in a structure with an infinite impedance mismatch with water predicts FP resonances at frequencies f_n, given by

$$f_n=N\frac{c}{2H}, \quad (2)$$

where N is a positive integer representing the harmonic number, c is the speed of sound in water (1480 m/s), and H is the metamaterial thickness (the channel length) (Lorenzo *et al.*, 2021).

Previous studies have also interpreted the efficient transmission of subwavelength details in such structures through the concept of evanescent wave canalization, in which high spatial-frequency components are guided or transformed into propagating modes inside the metamaterial. This mechanism has been widely used to explain near-field image transfer in both acoustic and electromagnetic metamaterials.

However, these analyses typically rely on the examination of spatial frequency spectra, such as Fourier-domain representations or explicit separation between near-field and far-field contributions, to reveal how evanescent components are transmitted through the structure. In contrast, the present work focuses primarily on near-field imaging behavior, characterized by spatial pressure distributions, without performing a di-

rect decomposition of the field into its spatial frequency components. This approach emphasizes the practical imaging performance of the designed metamaterial, rather than the detailed modal evolution inside the structure.

In this paper, we introduce a new class of acoustic metamaterials designed for near-field underwater imaging applications. Figure 1 shows a typical structure, which consists of a soft wire array (easily penetrable by sound waves) with a width of $H = 50\,\text{mm}$. The array forms a periodic structure with a lattice parameter $A = 2\,\text{mm}$ (the distance between the centers of two adjacent arrays), and features deep-subwavelength square wires with a side length of $a = 1\,\text{mm}$. The gaps between wires in the array are filled with air, sealing the whole structure. This facilitates the generation of acoustic isolation and greatly enhances imaging quality (Belov *et al.*, 2008; Astolfi *et al.*, 2019). All the wires are arranged in parallel within a square hollow soft box, with front and rear cover thicknesses of $h = 0.5\,\text{mm}$ and wall thicknesses of $c = 2\,\text{mm}$. Both the soft wire array and the hollow structure are designed to be easily penetrable by underwater sound waves, using materials such as soft polymers. This metamaterial has an acoustic impedance closely matched to that of water, eliminating the traditional coupling losses caused by water-filled holes in polymer substrates, thereby improving imaging efficiency. In the simulations presented in this paper, the array and the metamaterial hollow cubic shell around the array are modeled with Young's modulus of 2400 MPa, a Poisson's ratio of 0.4, and a density of $1100\,\text{kg/m}^3$. This 'wire array' metamaterial fabricated from polymers with acoustic impedance close to that of water, supports the formation of FP resonances inside the array, facilitating the transmission of evanescent waves and thus achieving subwavelength underwater imaging.

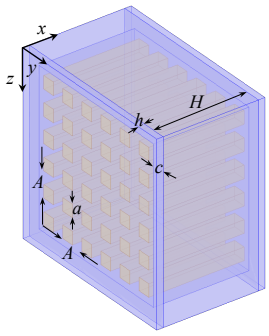

Fig. 1. Schematic diagram of a typical metamaterial structure.

It is worth noting that the acoustic impedance of the wires in array is close to that of water ($Z_{\text{polymer}} \approx Z_{\text{water}} = 1.48 \times 10^6\,\text{Pa}\cdot\text{s/m}$), and the acoustic impedance of air around the wire array ($Z_{\text{air}} = 411.6\,\text{Pa}\cdot\text{s/m}$) is very low. The evanescent waves scattered from the object under- water are confined within each soft polymer wire, enhancing acoustic isolation due to the strong acoustic impedance mismatch between the polymer wire and air $\left(\frac{Z_{\text{polymer}}}{Z_{\text{air}}} \approx 3.6 \times 10^3\right)$. FP resonances occur in each individual polymer wire. Since the polymer arrays exhibit FP resonances under free boundary conditions in air, the significant impedance mismatch between the polymer wires and air ensures that viscous losses at the wire-air interface can be safely neglected during evanescent wave propagation in the metamaterial. The speed of sound and density are shown in Table 1.

Table 1. Material properties.

Material	Velocity [m/s]	Density [kg/m^3]
Water	1500	1000
Air	343	1.2
Polymer	1477.1	1100

3. Assumption of simulation

To enhance clarity and reproducibility, the key modeling assumptions adopted in this work are summarized further.

In this study, the interfaces between the front and back cover plates, the array columns, and the hollow outer shell are assumed to be perfectly bonded, without air gaps or leakage channels. This assumption is supported by practical fabrication processes, where robust bonding techniques generally ensure reliable sealing. Although minor imperfections may exist in practice, their effect on acoustic transmission is expected to be negligible compared to the dominant resonance and cavity–plate interactions.

The array columns are composed of metamaterial whose acoustic impedance closely matches that of water, while the surrounding regions are filled with air. This configuration minimizes interaction between the columns and the ambient medium, thereby suppressing unwanted scattering and allowing FP resonance to govern the system's response.

The side walls are modeled as acoustically hard boundaries, justified by their high stiffness and firm bonding to adjacent components, which render their vibrations negligible. In contrast, the thin front and back cover plates, directly exposed to the surrounding medium, are explicitly treated with acoustic–structure interaction, because their vibration significantly influences transmission.

All acoustic processes are considered linear, as the operating pressure levels are well below the thresholds for nonlinear effects such as harmonic generation, cavitation, or turbulence. Neglecting these effects avoids unnecessary computational complexity while preserving physical fidelity.

These assumptions are commonly adopted in acoustic metamaterial modeling and provide a balanced trade-off between physical realism and compu-

tational efficiency. While relaxing them might slightly alter quantitative metrics such as resonance amplitude or transmission efficiency, the essential FP resonance behavior and the associated imaging performance remain unaffected.

4. Simulation and results

To validate the subwavelength imaging capabilities of the wire array metamaterial, we conducted comprehensive 3D numerical simulations using COMSOL Multiphysics. The simulation was performed in the pressure acoustic-frequency domain, coupled with solid mechanics, to study the problem in detail. In the acoustic domain, we applied plane wave radiation conditions, sound absorption boundaries, and hard sound field boundaries. In the solid mechanics domain, constrained boundary conditions were set. We generated an acoustic-solid coupling boundary that encompasses the surfaces of the front and back covers of the metamaterial, as well as the perimeter of the array columns.

Our overall model construction is mainly divided into four parts, consisting of front-end water, metamaterial, back-end water, and perfectly matched layer (PML) (Fig. 2).

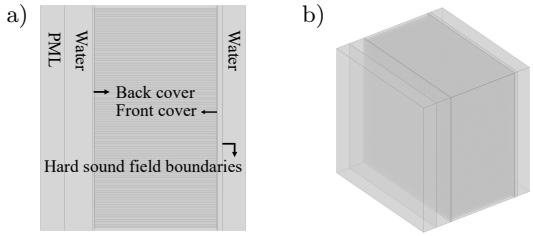

Fig. 2. a) Schematic of the simulation domain division; b) schematic of the full simulation model.

As described previously, the metamaterial array column is encapsulated within a hollow cubic shell, with both ends of the cube sealed by front and rear covers. The internal cavity of the cubic enclosure, excluding the space occupied by the array structure, is filled with air and maintained in a hermetically sealed condition. The front and rear cover thicknesses are defined as $0.5\,A$, where A is the period of the array. The side length of each square array column is denoted as a, the lateral (vertical) length of the array column is H, and the wall thickness of the surrounding hollow cubic shell is c.

We constructed an accurate simulation model to replicate a underwater environment for various test scenarios using COMSOL Multiphysics (Fig. 2), establishing the necessary theoretical conditions for the simulation. A plane wave is emitted, passes through the water, propagates to the metamaterial, generates resonance, and then forms an image on the opposite side. To ensure the simulation's accuracy, we included a PML (Fig. 2), beyond the water, with a wave speed identical to that of water (1480 m/s) to simulate an infinite domain. Additionally, impedance matching was applied around the metamaterial to ensure consistency and physical realism in the simulation.

To simulate plane wave emission for the image source, our approach is to first establish plane wave radiation conditions using a plane of the same size as the cover plate. As the sound wave propagates, we create a working plane of the same size as the plane wave emitting surface, referred to as 'E'. Outside the 'E' region, we apply hard acoustic field boundary conditions, while the area within the 'E' is left as a hollow space. This allows the acoustic wave to pass through the working plane and propagate toward the metamaterial surface. The 'E' structure consists of three horizontal rectangles, each measuring 46 mm in length and 6 mm in width, and one vertical rectangle measuring 62 mm in length and 6 mm in width, together forming the shape of the letter 'E' (hereafter referred to collectively as 'E').

This well-structured model is also facilitates precise mesh generation, resulting in concise calculations that meet the accuracy requirements for this work. We first applied swept meshing to the water domains and the PML, ensuring that the mesh size in the water region is less than $\lambda/6$, with the PML consisting of 20 layers, which complies with the established meshing criteria. Next, we applied mesh sweeps to the columns, the hollow metamaterial cubic shell, and the air domains within the metamaterials. The remaining connections between the front and rear cover plates were constructed using a free tetrahedral mesh. We conducted a grid convergence study by refining the mesh until the resonant frequency variation was below 0.05 %. In addition, we verified the numerical stability by adjusting solver tolerance and frequency step size, both of which showed negligible influence on the results.

Additionally, to better evaluate the imaging performance on the receiving surface, we analyzed the sound pressure distribution along a 3D cut line (Fig. 3). This allowed us to observe the variation trends in sound pressure. As illustrated, the 3D cut line lies along the yz-plane at the intended focal distance, positioned at the center of the hollow 'E' structure on the working plane, thereby encompassing its three horizontal edges. The vertical axis spans from 0 m (near the lower edge) to 0.092 m (near the upper edge). By examining the sound pressure distribution along this line, we can assess the imaging quality. An optimal focusing effect should exhibit the following characteristics:

1) a smooth and continuous pressure profile,
2) peak sound pressure at the three edges of the 'E' structure,
3) minimum sound pressure in the regions between the adjacent edges,
4) approximately uniform sound pressure magnitudes across all three edges.

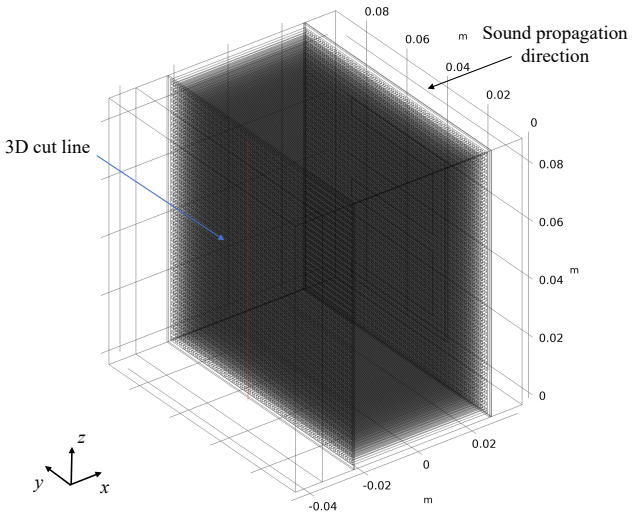

Fig. 3. 3D cut line.

In this model, since the sound wave propagates in the negative x-direction, the sound pressure value obtained on the imaging plane is inherently negative. Therefore, the absolute sound pressure value should be used for comparative analysis of the sound pressure magnitude.

Before initiating the simulation test groups, we measured the acoustic pressure transmission coefficient of the material. The sound source setup is identical to the one presented the previous section: a plane wave is emitted from the rightmost plane, passing through the hollow 'E' to reach the front cover of the metamaterial. The output face is defined as the end face of the cover farthest from the sound source, from which we extract the total acoustic pressure (P_{out}). The input face, closer to the sound source, is used to extract the total acoustic pressure (P_{in}). The sound pressure transmission coefficient is calculated as the ratio of P_{out} to P_{in}. We selected a frequency range of 27 kHz to 34 kHz and plotted the sound pressure transmission coefficient curve, as shown in Fig. 4. From the curve, we observe that the transmission coefficient exceeds 0.82 within the frequency range of 30.6 kHz to 31.6 kHz, maintaining a broad bandwidth. This information is crucial for identifying the optimal incidence frequency. However, there may be a slight deviation between this frequency range and the actual frequency that yields the highest imaging quality.

The preceding analysis of the acoustic pressure transmission coefficient ($P_{\text{out}}/P_{\text{in}}$) provides a quantitative, physics-based measure of the metamaterial's ability to transmit both propagating and evanescent components. In principle, more detailed metrics such as the point-spread function (PSF) or modulation transfer function (MTF) could be extracted from full-wave simulations using tools such as the finite-difference time-domain (FDTD) or transfer matrix method (TMM), which compute the response to different transverse wave vectors ($\mathbf{k}_x, \mathbf{k}_y$) (Zhu *et al.*, 2018). The $P_{\text{out}}/P_{\text{in}}$ analysis employed here captures the essential physics of energy transmission and validates the effectiveness of the overall model, providing a simplified yet rigorous basis for subsequent visual evaluation of near-field subwavelength imaging performance.

4.1. Optimal imaging frequency comparison

To determine the optimal imaging frequency, the incident frequency was first varied, revealing a range between 27.5 kHz and 30 kHz in which clear imaging was achievable. A step size of 0.1 kHz was used for a frequency sweep. For these simulations, the imaging distance of 1.5 A and the incident distance of 0.1 A were tentatively set. The corresponding visual imaging results are shown in Fig. 5, and the quantitative sound pressure distributions across the image plane for selected frequencies are shown in Fig. 7.

As shown in Figs. 5 and 6, we observed that imaging quality is poor between 27.5 kHz to 28.5 kHz,

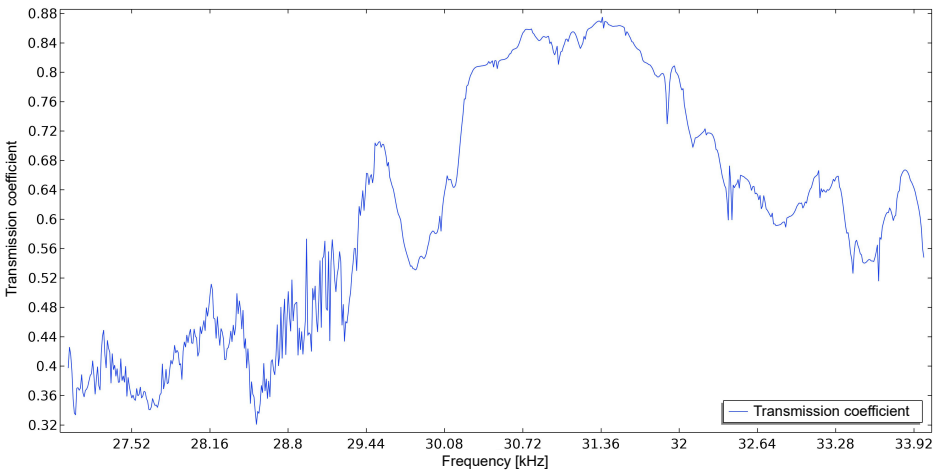

Fig. 4. Transmission coefficient range from 27 kHz to 34.0 kHz.

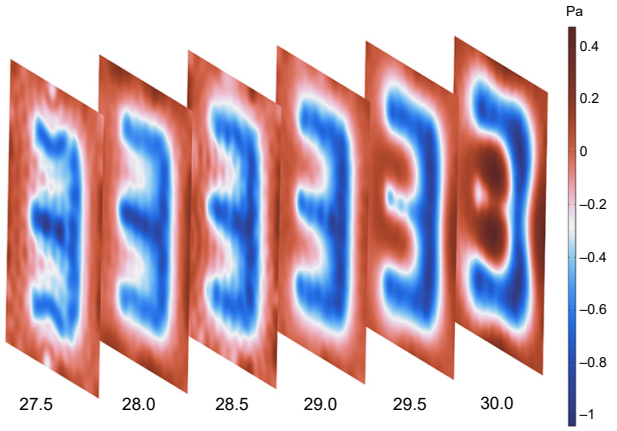

Fig. 5. Overall imaging frequency range from 27.5 kHz to 30.0 kHz.

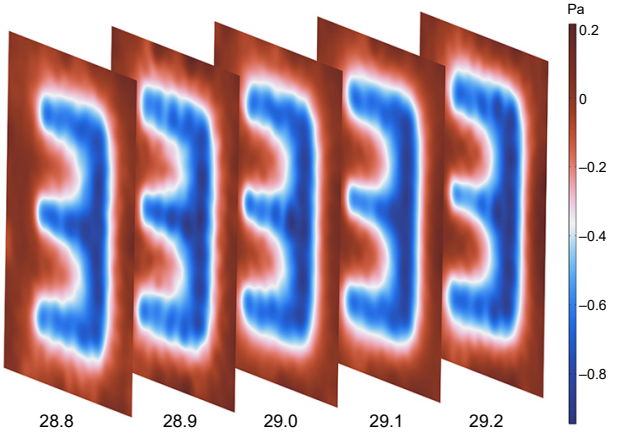

Fig. 6. Optimal imaging frequency range from 28.8 kHz to 29.2 kHz.

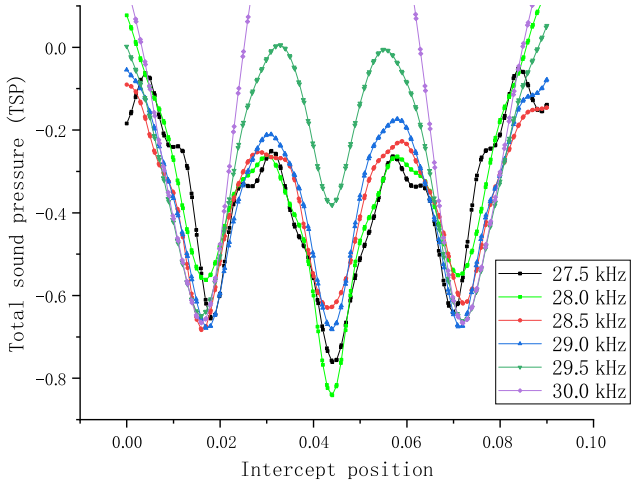

Fig. 7. Sound pressure curves in the frequency range of 27.5 kHz to 30.0 kHz.

the sound pressure distribution along the three boundaries of the 'E' is uneven, and fine details are not well presented. Additionally, at the vertical junctions, defocusing occurs, resulting in an unclear outline of the object. When the frequency exceeds 29.2 kHz, a noticeable thinning occurs at the center of the image 'E', which does not satisfy our imaging quality requirements. In contrast, Fig. 6 shows the optimal imaging frequency range from 28.8 kHz to 29.2 kHz, where these issues are resolved.

In the 29.5 kHz–30 kHz image, the second side of the 'E' is not well imaged, the outline of the image 'E' improves, though some intermittent areas remain. It can be seen from the sound pressure curve in Fig. 7 that, consistent with our imaging effect, the 27.5 kHz sound pressure curve is not smooth near the 0.03 mm and 0.06 mm positions. The reason is that there is defocusing at the boundary, and the background field sound pressure and the transmitted sound pressure cannot be distinguished well. This situation results in poor presentation of image detail information. The sound pressure transmission at the lowest peak of the 28 kHz sound pressure curve, which is the second side of 'E', is good, but the sound pressure transmission at the other two sides (the first and third lowest peaks) differs and cannot be transmitted very evenly, which caused the center of the 'E' to be clearly imaged while the surrounding areas appear blurred.

At 28.5 kHz, the curve is very smooth as a whole, smooth at the position of 0.03 mm and 0.06 mm positions of the sound pressure curve section, and the sound pressure at the three lowest peaks is nearly equal. However, the 29 kHz curve shows higher transmission sound pressure and a larger difference between the maximum and minimum peaks, resulting in better imaging quality. At 29.5 kHz curve, the absolute value of the sound pressure at the second minimum point is smaller than the transmission sound pressure at the other two sides, that is, the resolution effect of the second side of the 'E' is poor. At 30 kHz, the sound pressure on the second side is not less than 0, so the second side of the 'E' is almost invisible. After comparison, we determined that the optimal imaging frequency lies between 28.8 kHz and 29.2 kHz (Fig. 6). This frequency range demonstrates good imaging quality and robustness.

Comparing the sound pressure curves in Fig. 8, it is found that within the frequency range of 28.8 kHz to 29.2 kHz, the sound pressure curves are very smooth and the imaging effect is good, but there is a slight difference between the minimum and maximum peaks. At 29 kHz, the transmission sound pressure values at the three minimum peaks of the curve are closer. In addition, there is a large difference between them and the maximum peak, enabling better distinction of the image details. The imaging effect at 29.2 kHz is inferior to that at 29 kHz because the absolute value of the sound pressure of the second side of the 'E' is lower. Consequently, we selected 29 kHz as the optimal frequency for subsequent analyses. The imaging quality across the frequency range is summarized in Table 2.

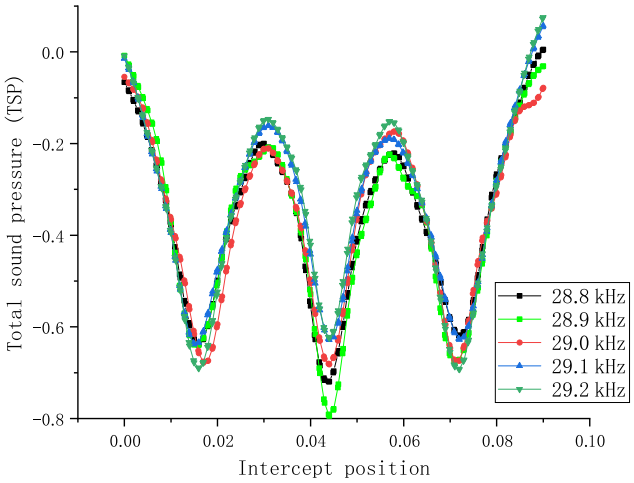

Fig. 8. Sound pressure curves in the frequency range of 28.8 kHz to 29.2 kHz.

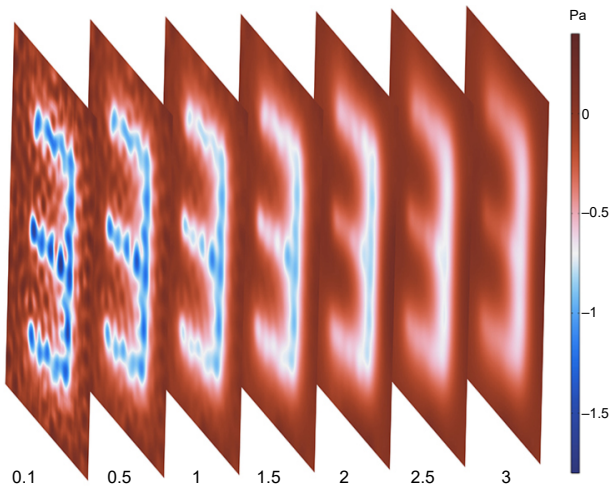

Fig. 9. Imaging performance comparison from 0.1 A to 3 A for optimal distance selection.

Table 2. Imaging quality under different incident frequencies.

Frequency range [kHz]	Imaging quality	Main features / issues
27.5–28.5	Poor	Uneven sound pressure; defocusing at junctions; poor details
28.8–29.2	Good	Smooth curves; balanced transmission across edges; robust imaging
29.2–30.0	Degraded	Central thinning, edge degradation; second edge fades at 30 kHz

4.2. Comparison of optimal imaging distance

After establishing the incident frequency at 29 kHz, we proceeded to determine the optimal imaging distance. During this phase, the distance from the 'E' sound-emitting surface to the front cover plate of the metamaterial was fixed at 0.1 A, while all other parameters were kept constant. Cross-sections were generated at intervals ranging from from 0.1 A to 3 A to adjust the receiving surface and obtain the corresponding images (Fig. 9).

As shown in Fig. 9, the imaging distance has a significant impact on image quality. At 0.1 A, the imaging surface is closest to the cover plate, where the sound pressure distribution is relatively uniform and the outline of the 'E' is clear, with minimal edge defocusing. This results in a more realistic reconstruction. As the imaging distance increases, the absolute sound pressure on the three sides of the 'E' decreases markedly, the contrast with the background weakens, and junction details become blurred. This degradation arises from evanescent wave decay: high-frequency spatial Fourier components ($k_z > k_0$) diminish exponentially with distance $e^{-|k_z|z}$, leading to the loss of subwavelength information.

The sound pressure curves in Fig. 10 further validate these observations. At 0.1 A, the transmitted sound pressure of the three sides of the 'E' reaches its maximum, with noticeable peaks and stronger contrast between maxima and minima, which enhances detail resolution despite some extra oscillations. With increasing distance, the curves become smoother and extra peaks disappear, but the difference between the highest and lowest values diminishes, reducing image sharpness. Moreover, the peak distribution indicates that the second side of the 'E' is imaged more clearly than the upper and lower sides, leading to uneven reconstruction. Although small oscillations remain at 0.1 A, these can be mitigated using filtering or curve-fitting algorithms (Allen, Vlahopoulos, 2002). Based on these results, the optimal imaging distance was determined to be 0.1 A. The imaging performance at different distances is summarized in Table 3.

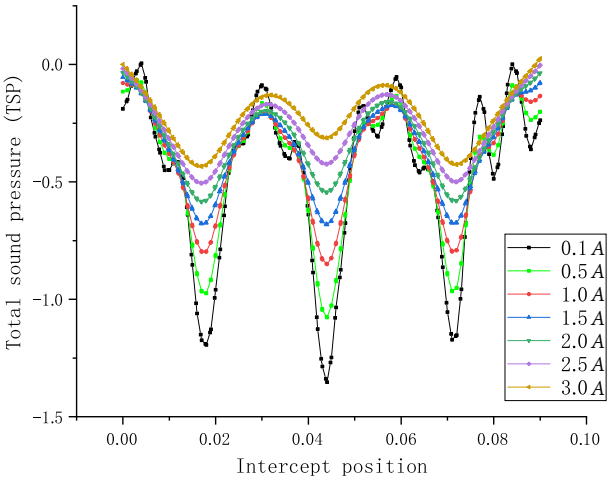

Fig. 10. Sound pressure curves at imaging distances ranging from 0.1 A to 3 A.

Table 3. Summary of imaging performance at different distances.

Imaging distance (A)	Imaging quality	Main features / issues
0.1	Good	Clear outline; uniform sound pressure; minimal defocusing
0.5–1	Medium	Reduced contrast; partial blurring at junctions; uneven side imaging
>1	Poor	Significant loss of detail; weak contrast with background due to evanescent wave decay

4.3. Comparison of optimal incidence distances

Through comparative simulations, we established that the optimal incident frequency is 29 kHz and the best receiving distance is 0.1 A. Based on these parameters, we further analyzed the influence of incident distance to determine a more suitable configuration. During this process, the receiving plane was fixed at 0.1 A from the back cover plate of the metamaterial, while the incident plane was adjusted from 0.1 A to 3 A (Fig. 11). The model parameters and meshing were independently recalculated for each distance, while the acoustic boundary conditions remained consistent with the default boundary around the 'E'.

As shown in Fig. 11, the imaging quality degrades progressively as the incident distance increases. When the incident distance is 0.1 A, the 'E' image is sharp and continuous, with a uniform sound pressure distribution and minimal edge distortion. The short propagation distance enhances evanescent wave coupling, resulting in strong transmitted sound pressure and a clear reconstruction of all edges. As the incident distance increases to 0.3 A–0.5 A, the transmitted sound pressure along the three edges of the 'E' decreases, accompanied by the emergence of stray peaks and mild distortion in the image. At 1 A, the central region of the sound field becomes dominant, while the upper and lower edges weaken, causing the 'E' to appear blurred. When the incident distance further increases to 2 A–3 A, the image of the 'E' becomes indistinct and nearly disappears.

The sound pressure curves in Fig. 12 confirm this trend. The decrease in transmitted sound pressure amplitude and the growing smoothness of the curves reflect the attenuation of high-frequency evanescent components. Additionally, slight impedance mismatches

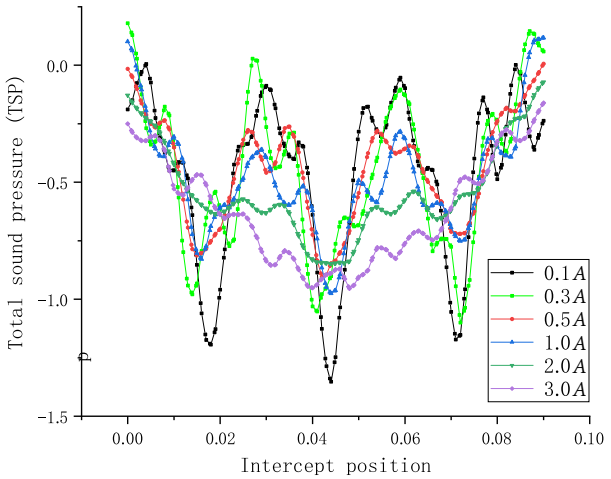

Fig. 12. Sound pressure curve for incidence distances from 0.1 A to 3 A.

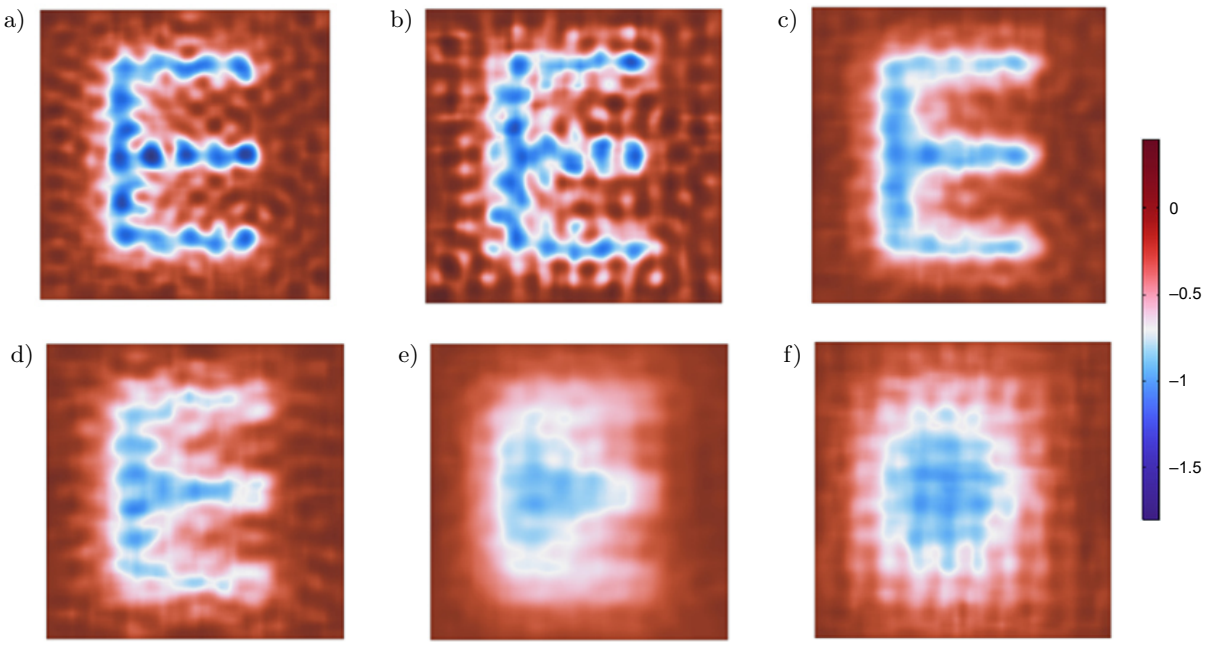

Fig. 11. Comparison of optimal incidence distances: a) 0.1 A; b) 0.3 A; c) 0.5 A; d) 1 A; e) 2 A; f) 3 A.

between the metamaterial covers and water cause multiple reflections between the 'E' baffle and the front cover, introducing stray peaks and positional shifts in the sound pressure extrema. These effects lead to degradation of subwavelength imaging performance. To preserve high-frequency information and minimize wave interference, the incident plane should be positioned as close as possible to the metamaterial surface. Consequently, 0.1 A is determined to be the optimal incident distance. The imaging performance at different incident distances is summarized in Table 4.

Table 4. Summary of imaging performance at different incident distances.

Imaging distance (A)	Imaging quality	Main features / issues
0.1	Good	Strong evanescent coupling; uniform sound pressure; minimal edge distortion
0.3–0.5	Medium	Slight loss of edge sharpness; reduced sound pressure; appearance of stray peaks
1.0	Poor	Central sound field dominates; weakened upper/lower edges; image blurring
2.0–3.0	Very Poor	Strong reflection effects; severe evanescent decay; 'E' shape almost invisible

4.4. Comparison of imaging quality after changing cover thickness

As previously established, the optimal incident frequency is 29 kHz, and both the optimal receiving and incident distances are set to 0.1 A. In the following analysis, the thickness of the metamaterial's front and back cover plates was varied simultaneously from 0.5 mm to 3.0 mm to examine its influence on imaging quality, while maintaining the optimal geometric and acoustic conditions. To ensure accurate comparison, the water region length in the model was kept constant during all simulations.

As illustrated in Fig. 13, increasing the cover plate thickness leads to a gradual degradation of imaging quality. When the thickness is 0.5 mm, the imaging of the letter 'E' is clear and continuous, indicating good acoustic transmission and minimal phase distortion. As the thickness increases to around 1.0 mm–1.5 mm, slight blurring and burrs appear along the middle horizontal stroke of the 'E', and local discontinuities emerge due to partial phase mismatching between the transmitted and reflected sound waves. At 2.0 mm, the central line of the 'E' exhibits breakpoints, and at 3.0 mm, both the upper and lower horizontal edges begin to curve and distort, with the overall image becoming defocused and noisy. This degradation is primarily attributed to the multiple reflections within the thicker cover layers, which induce phase interference and attenuate the effective transmission of evanescent components.

The sound pressure trends in Fig. 14 further validate this observation. When the cover plate is thin (0.5 mm), the transmitted acoustic pressure along the three edges of the 'E' reaches higher absolute values and shows clear separation between peaks and troughs, corresponding to sharp and distinct image boundaries. With increasing thickness, the sound pressure

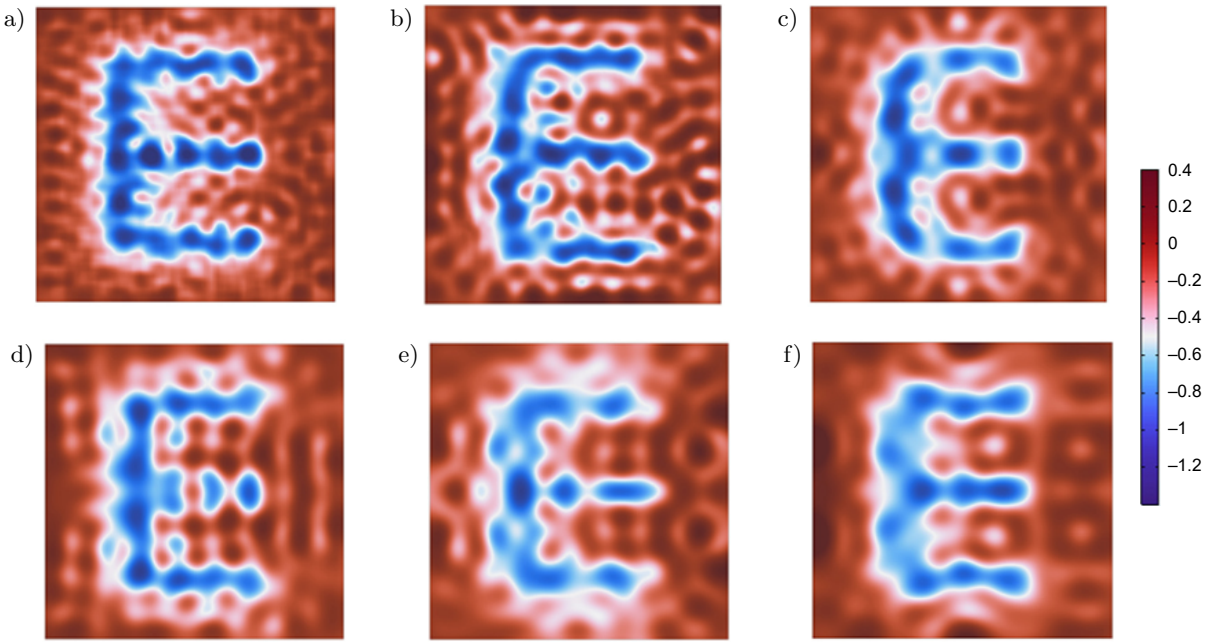

Fig. 13. Comparison of imaging quality for different cover thickness: a) 0.5 mm; b) 1.0 mm; c) 1.5 mm; d) 2.0 mm; e) 2.5 mm; f) 3.0 mm.

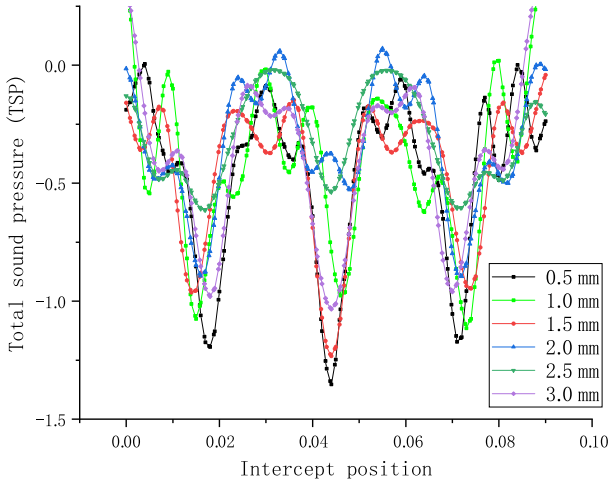

Fig. 14. Sound pressure curves for cover thicknesses ranging from 0.5 mm to 3.0 mm.

curves become progressively irregular, with more spurious peaks and reduced amplitude differences, indicating uneven transmission and increased scattering within the covers. At 2.0 mm and beyond, the pressure at the second edge of the 'E' weakens sharply, while background pressure fluctuations intensify, causing image details to blur or disappear.

In summary, increasing the cover plate thickness results in stronger internal reflection and enhanced evanescent decay, leading to phase distortion and reduced subwavelength imaging fidelity. The 0.5 mm cover thickness provides the most stable and clear imaging performance under the given conditions.

4.5. Comparison of imaging quality with changing array cycles

The aim of this experiment is to maintain the size and position of the letter 'E' while proportionally reducing both the column bottom edge length (a) and the array period, ensuring that their ratio to the original model remains constant. In the original configuration, each edge of the 'E' corresponds to the orthocenter between two arrays. After scaling down, it is crucial to preserve the alignment of the 'E' with this orthocenter. However, if the dimensions of the 'E' remain unchanged, precise alignment of all three edges with the intended array positions cannot be guaranteed. Therefore, the dimensions of the 'E' are fine-tuned to ensure full alignment with the adjusted array configuration. Throughout this process, both the incidence and reception distances are maintained at their optimal values, and the cover plate thickness is fixed at 0.5 mm to achieve the best imaging performance.

As shown in Fig. 15, the comparison across six sets of experiments reveals that when only a single column exists within the gap of the 'E', the resulting image appears blurred, and the 'E' is indistinguishable at the optimal imaging distance. With two columns, imaging quality improves but remains suboptimal. When the number of columns increases to three, the 'E' becomes distinctly visible, and its outline more closely resembles that of the model at the incident plane. At four columns, the image contours are clearer, with straighter sides and nearly perpendicular intersections, further enhancing fidelity to the original 'E'. However, as the number of columns increases to five and six,

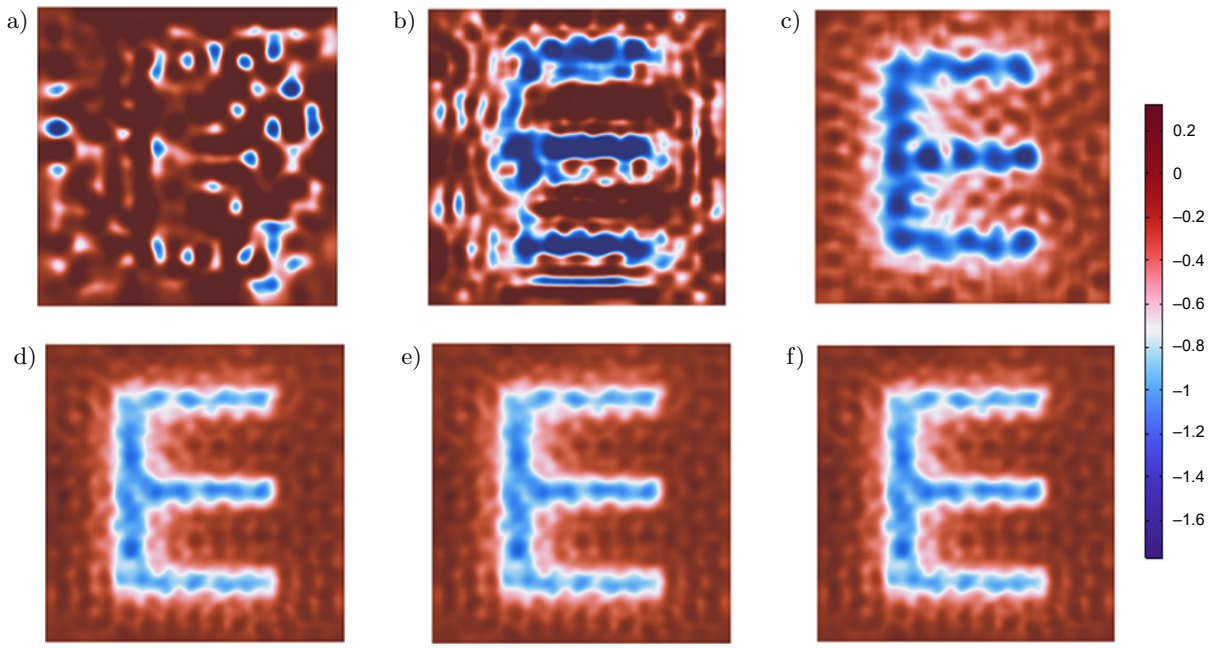

Fig. 15. Comparison of imaging quality with varying number of array columns: a) 1; b) 2; c) 3; d) 4; e) 5; f) 6.

no substantial improvement is observed, indicating that image quality reaches saturation at four columns.

From the sound pressure distributions shown in Fig. 16, imaging with a single column is ineffective; hence, analysis begins with two columns. When two array columns are used, the absolute acoustic pressure along the three edges of the 'E' is high, and the contrast with the background field is pronounced. However, numerous spurious peaks appear, and the background field is irregular. When the number of columns increases to three, the absolute pressure at the 'E' edges slightly decreases, but the background field becomes more uniform. With four or more columns, the sound pressure curves show minimal further change. Although the absolute pressure at the edges continues to decrease slightly, the background field remains evenly distributed, resulting in a stable and clearly defined image.

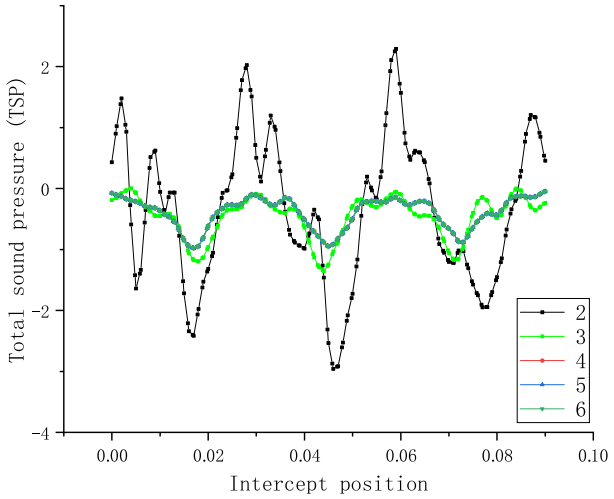

Fig. 16. Sound pressure curves for array periods corresponding to 2 to 6 columns.

It is worth noting that these imaging results were obtained under idealized simulation conditions, without considering real-world disturbances such as background acoustic noise, medium inhomogeneity, or object motion, which are common in underwater environments. Although this simplification enables a clear evaluation of the intrinsic imaging performance of the proposed metamaterial, future work will incorporate these factors to assess the robustness and practical applicability of the system under realistic underwater conditions.

5. Conclusions

This work demonstrated that by placing the image plane at a specific distance from the output plane, a faithful representation of the 'E' pattern underwater can be achieved. The imaging quality is influenced by several factors, including incident frequency, incident distance, imaging distance, cover thickness, and array period. By adjusting the frequency, we can achieve high-quality imaging within the range of 28.8 kHz to 29.2 kHz, with an optimum frequency of 29 kHz, indicating that the metamaterial exhibits a broad bandwidth response. We determined the optimal incidence and imaging distances to be 0.2 mm from both the front and rear cover plates.

Additionally, we found that excessive cover thickness negatively impacts imaging quality, with the optimal thickness being 0.5 mm. Furthermore, we confirmed that the array period plays a significant role in enhancing imaging quality. As the number of arrays passing through the 'E' increases, the imaging quality improves; however, when more than four array columns are present, the quality tends to saturate and does not significantly change with the addition of more columns. These findings confirm that a wire array metamaterial functions effectively as a near-field acoustic imaging device capable of operating at very deep subwavelength scales underwater. This imaging capability and the associated principles provide some theoretical support for applications. including medical ultrasonography, micro-device flaw detection, and ultrasonic non-destructive evaluation.

Fundings

This work was supported by the National Natural Science Foundation of China (no. 12104369, 12174004), the Postgraduate Innovation Project Fund of Xi'an University of Posts and Telecommunications (CXJJZL2023032), the State Key Laboratory for Manufacturing Systems Engineering open fund (no. sklms2022001), the general grant of Shaanxi Province in China (no. 2023-JC-YB-521), the Zhuhai City Philosophy and Social Sciences Planning 2023 Annual Planning Project (2023YBB054, GD23XTY35), and the Guangdong Province Philosophy and Social Sciences Planning 2023 Discipline Co-construction Project (GD23XTY35).

Competing 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 was responsible for conceptualization, funding acquisition, resources, supervision, writing original draft; FeiLong Li was responsible for data curation, investigation, methodology, formal analysis, finite element simulation, writing original draft; LiQing Hu was responsible for finite element simulation, investigation; QunFeng Li was responsible for resources, supervision;

GuanJun Yin was responsible for finite element simulation, validation. All authors reviewed and approved the final manuscript.

References

1. Allen M.J., Vlahopoulos N. (2002), Numerical probabilistic analysis of structural/acoustic systems*, *Mechanics of Structures and Machines*, **30**(3): 353–380, https://doi.org/10.1081/SME-120004422.
2. Ambati M., Fang N., Sun C., Zhang X., (2008), Surface resonant states and superlensing in acoustic metamaterials, *Physical Review B*, **75**(19): 195447, https://doi.org/10.1103/PhysRevB.75.195447.
3. Amireddy K.K., Balasubramaniam K., Rajagopal P. (2016), Holey-structured metamaterial lens for subwavelength resolution in ultrasonic characterization of metallic components, *Applied Physics Letters*, **108**(22): 224101, https://doi.org/10.1063/1.4950967.
4. Amireddy K.K., Balasubramaniam K., Rajagopal P. (2017), Deep subwavelength ultrasonic imaging using optimized holey structured metamaterials, *Scientific Reports*, **7**: 7777, https://doi.org/10.1038/s41598-017-08036-4.
5. Astolfi L. *et al.* (2019), Negative refraction in conventional and additively manufactured phononic crystals, [in:] *2019 IEEE International Ultrasonics Symposium (IUS)*, pp. 2529–2532, https://doi.org/10.1109/ULTSYM.2019.8926236.
6. Belov P.A., Silveirinha M.G. (2006), Resolution of subwavelength transmission devices formed by a wire medium, *Physical Review E*, **73**(5): 056607, https://doi.org/10.1103/PhysRevE.73.056607.
7. Belov P.A., Silveirinha M.G., Simovski C.R., Hao Y., Parini C. (2008), Comment on "Guiding, Focusing, and Sensing on the Subwavelength Scale Using Metallic Wire Arrays", arXiv, https://doi.org/10.48550/arXiv.0804.3670.
8. Christensen J., Martin-Moreno L., Garcia-Vidal F.J. (2008), Theory of resonant acoustic transmission through subwavelength apertures, *Physical Review Letters*, **101**(1): 014301, https://doi.org/10.1103/PhysRevLett.101.014301.
9. Christensen J., García de Abajo F.J. (2010), Acoustic field enhancement and subwavelength imaging by coupling to slab waveguide modes, *Aplied Physics Letters*, **97**(16): 164103, https://doi.org/10.1063/1.3504700.
10. Deng K., Ding Y., He Z., Zhao H., Shi J., Liu Z. (2009), Theoretical study of subwavelength imaging by acoustic metamaterial slabs, *Journal of Applied Physics*, **105**(12): 124909, https://doi.org/10.1063/1.3153976.
11. Dong E., Cao P., Zhang J., Zhang S., Fang N.X., Zhang Y. (2023), Underwater acoustic metamaterials, *National Science Review*, **10**(6): 258–280, https://doi.org/10.1093/nsr/nwac246.
12. Estrada H., Candelas P., Uris A., Belmar F., García de Abajo F.J., Meseguer F. (2008), Extraordinary sound screening in perforated plates, *Physical Review Letters*, **101**(8): 084302, https://doi.org/10.1103/PhysRevLett.101.08430.
13. Guenneau S., Movchan A., Pétursson G., Ramakrishna S.A. (2007), Acoustic metamaterials for sound focusing and confinement, *New Journal of Physics*, **9**: 399, https://doi.org/10.1088/1367-2630/9/11/399.
14. Gulia P., Gupta A. (2019), Sound attenuation in triple panel using locally resonant sonic crystal and porous material, *Applied Acoustics*, **156**: 113–119, https://doi.org/10.1016/j.apacoust.2019.07.012.
15. Kawata S., Ono A., Verma P. (2008), Subwavelength colour imaging with a metallic nanolens, *Nature Photonics*, **2**(7): 438–442, https://doi.org/10.1038/nphoton.2008.103.
16. Kim K.Y., Liu B., Huang Y., Ho S.-T. (2008), Simulation of photodetection using finite-difference time-domain method with application to near-field subwavelength imaging based on nanoscale semiconductor photodetector array, *Optical and Quantum Electronics*, **40**(5): 343–347, https://doi.org/10.1007/s11082-008-9190-0.
17. Lafleur L.D., Shields F.D. (1995), Low-frequency propagation modes in a liquid-filled elastic tube waveguide, *The Journal of the Acoustical Society of America*, **97**(3): 1435–1445, https://doi.org/10.1121/1.412981.
18. Laureti S., Davis L.A.J., Ricci M., Hutchins D.A. (2014), The study of broadband acoustic metamaterials in air, [in:] *2014 IEEE International Ultrasonics Symposium*, pp. 1344–1347, https://doi.org/10.1109/ULTSYM.2014.0332.
19. Laureti S. *et al.* (2020), Trapped air metamaterial concept for ultrasonic subwavelength imaging in water, *Scientific Reports*, **10**(1): 10601, https://doi.org/10.1038/s41598-020-67454-z.
20. Laureti S., Hutchins D.A., Davis L.A.J., Leigh S.J., Ricci M. (2016), High-resolution acoustic imaging at low frequencies using 3D-printed metamaterials, *American Institute of Physics Advances*, **6**(12): 121701, https://doi.org/10.1063/1.4968606.
21. Li J., Chan C.T. (2004), Double-negative acoustic metamaterial, *Physics Review E*, **70**(5): 055602, https://doi.org/10.1103/PhysRevE.70.055602.
22. Lin L., Shen H., Shi S., Zhang D., Fu D., Ma Z. (2025), Subwavelength resolution imaging of ultrasonic total focusing method by decoupling overlapped signals through back propagation neural network, *Mechanical Systems and Signal Processing*, **231**: 112724, https://doi.org/10.1016/j.ymssp.2025.112724.
23. Liu F., Cai F., Peng S., Hao R., Ke M., Liu Z. (2009), Parallel acoustic near-field microscope: A steel slab with a periodic array of slits, *Physics Review E*, **80**(2): 026603, https://doi.org/10.1103/PhysRevE.80.026603.

24. Liu Z., Lee H., Xiong Y., Sun C., Zhang X. (2007), Far-field optical hyperlens magnifying sub-diffraction-limited objects, *Science*, **315**(5819): 1686, https://doi.org/10.1126/science.1137368.

25. Lorenzo A. *et al.* (2021), Holey-structured tungsten metamaterials for broadband ultrasonic subwavelength imaging in water, *The Journal of the Acoustical Society of America*, **150**(1): 74–81, https://doi.org/10.1121/10.0005483.

26. Molerón M., Daraio C. (2015), Acoustic metamaterial for subwavelength edge detection, *Nature Communications*, **6**(8): 8037, https://doi.org/10.1038/ncomms9037.

27. Pendry J.B. (2000), Negative refraction makes a perfect lens, *Physical Review Letters*, **85**(18): 3966–3969, https://doi.org/10.1103/PhysRevLett.85.3966.

28. Silveirinha M.G., Belov P.A., Simovski C.R. (2008), Ultimate limit of resolution of subwavelength imaging devices formed by metallic rods, *Optics Letters*, **33**(15): 1726–1728, https://doi.org/10.1364/OL.33.001726.

29. Simonetti F. (2006), Multiple scattering: The key to unravel the subwavelength world from the farfield pattern of a scattered wave, *Physical Review E*, **73**(3): 036619, https://doi.org/10.1103/PhysRevE.73.036619.

30. Yan X., Yuan F.-G. (2015), Conversion of evanescent Lamb waves into propagating waves via a narrow aperture edge, *The Journal of the Acoustical Society of America*, **137**(6): 3523–3533, https://doi.org/10.1121/1.4921599.

31. Zhang S., Yin L., Fang N. (2009), Focusing ultrasound with an acoustic metamaterial network, *Physical Review Letters*, **102**(19): 194301, https://doi.org/10.1103/PhysRevLett.102.194301.

32. Zhou Y. *et al.* (2010), Acoustic surface evanescent wave and its dominant contribution to extraordinary acoustic transmission and collimation of sound, *Physical Review Letters*, **104**(16): 164301, https://doi.org/10.1103/PhysRevLett.104.164301.

33. Zhu J. *et al.* (2010), A holey-structured metamaterial for acoustic deep-subwavelength imaging, *Nature Physics*, **7**(1): 52–55, https://doi.org/10.1038/nphys1804.

34. Zhu X.-F., Wei Q., Wu D.-J., Liu X.-J. (2018), Broadband acoustic subwavelength imaging by rapidly modulated stratified media, *Scientific Reports*, **8**(1): 4934, https://doi.org/10.1038/s41598-018-23411-5.

Archives of Acoustics Vol. 50, No. 4, pp. 491–499 (2025), https://doi.org/10.24425/aoa.2025.154829

Research Paper

Single-Sensor Passive Ranging of Underwater Monopoles Using Near-Field/Far-Field Energy Contrasts

Saeir MAHMOUD*, Louay SALEH, Ibrahim CHOUAIB

Department of Electronic and Mechanical Systems
Higher Institute for Applied Sciences and Technology
Damascus, Syria

*Corresponding Author e-mail: saeir.mahmoud@hiast.edu.sy

Received June 26, 2025; accepted September 26, 2025;
published online October 27, 2025.

While acoustic vector sensors (AVS) are well-established for detection and direction-of-arrival (DOA) estimation using co-located pressure and particle motion (PM) measurements, their potential for passive range estimation remains largely unexplored. This paper introduces a novel single-AVS method for passive range estimation to an acoustic monopole source by exploiting the fundamental near-field dominance of PM energy. We derive the frequency and the distance dependent ratio (ξ) of kinetic to potential acoustic energy density – a key near-field signature inaccessible to conventional hydrophones. By leveraging simultaneous AVS pressure and PM velocity measurements, our method estimates ξ, inverts the monopole near-field model to obtain the Helmholtz number, and directly computes the range. Crucially, we demonstrate that PM sensors offer a potential signal-to-noise ratio (SNR) advantage over pressure sensors within the near-field (>7.8 dB). Validation under simulated noise conditions shows accurate range estimation (RMSE <10 %) for low-frequency sources (<100 Hz) within 8 m–25 m ranges at 0 dB SNRs, with performance degrading as frequency increases or SNR decreases. Critically, robustness is confirmed using recorded basin noise profiles, overcoming the isotropic Gaussian noise assumption. This technique extends AVS functionality beyond DOA, enabling single-sensor passive ranging without arrays, environmental priors, or reference signals where conventional methods fail.

Keywords: monopole source; passive ranging; acoustic vector sensor (AVS); particle motion (PM); near-field acoustics; underwater acoustics; energy ratio; single-sensor localization.

1. Introduction

In underwater environment, the use of electromagnetic waves in detection systems faces significant challenges due to attenuation, scattering, and dispersion (Kaushal, Kaddoum, 2016). However, acoustic detection systems have proven to be more effective. The widely used SONAR system, which is based on acoustic waves, provides a larger coverage area compared to electromagnetic wave-based systems such as radio frequency (RF) and optical systems (Kaushal, Kaddoum, 2016; Eleftherakis, Vicen-Bueno, 2020) (Table 1). Other techniques, such as magnetic detection systems, may also be employed; however, they are

Table 1. Comparison of different wireless underwater technologies (Kaushal, Kaddoum, 2016).

Parameter	Acoustic	RF	Optical
Attenuation	0.1 dB/km–4 dB/km	3.5 dB/m–5 dB/m	0.39 dB/m (ocean) 11 dB/m (turbid)
Speed	1500 m/s	$2.3\,\text{m/s} \times 10^8$ m/s	$2.3\,\text{m/s} \times 10^8$ m/s
Distance	Up to km	≤10 m	≈ 10 m–100 m
Frequency band	10 kHz–15 kHz	30 MHz–300 MHz	$5\,\text{Hz} \times 10^{14}$ Hz

limited by the low signature of certain underwater objects (Soldani *et al.*, 2022).

SONAR systems are categorized as active or passive (Abraham, 2019). Active systems emit high-energy pulses for echo analysis, enabling precise ranging, at the cost of high-power consumption, ecological impact, and operational expense (Hari *et al.*, 2015; Jin, Xu, 2020); conversely, passive systems listen to ambient sounds, providing low-cost, energy-efficient, and environmentally benign surveillance (Jin, Xu, 2020).

Acoustic fields arise from pressure fluctuations, modeled as monopoles (pulsating spheres), dipoles (out-of-phase monopole pairs), or higher-order multipoles (Kalmijn, 1988). These generate two measurable components: scalar pressure and the vector PM, aligned with the wave direction in the free far-field (Jansen *et al.*, 2017). The pressure-PM relationship, defined by specific acoustic impedance, is real-valued in the far-field but complex in the near-field (Lin *et al.*, 2021). It is a critical distinction for ranging. Sensors diverge in capturing these: hydrophones measure pressure, while acoustic vector sensors (AVS) capture PM (velocity/acceleration) and optionally pressure (Tichavsky *et al.*, 2001).

Single-hydrophone systems detect divers (Cole, 2019; Tu *et al.*, 2020; Korenbaum *et al.*, 2020), ships, and biological sources (Ferguson *et al.*, 2010) but fail at passive ranging without environmental priors. Hydrophone array enable direction-of-arrival (DOA) estimation via beamforming (Krishnaveni *et al.*, 2013) or cross-correlation (Sutin *et al.*, 2013) but incur prohibitive cost and deployment complexity. While single AVS advances support DOA estimation (Zhao *et al.*, 2018) and detect sources (Yuan *et al.*, 2022), such as divers (Mahmoud *et al.*, 2025), air gun or boats (Jansen *et al.*, 2017; 2019), they remain prohibitively expensive (Jansen *et al.*, 2017), and research overwhelmingly focuses on DOA – neglecting passive ranging. Existing ranging techniques such as triangulation (Abraham, 2019), multipath delays (Abraham, 2019; Lohrasbipeydeh *et al.*, 2013; Ferguson *et al.*, 2010), dispersion curves (Li *et al.*, 2023), or matched filter (Liang *et al.*, 2022) require arrays, environmental knowledge, shallow-water constraints, or reference signals.

The fundamental near-field characteristic of PM (exhibiting $1/r^2$ decay versus pressure's $1/r$ decay) remains unexploited for passive monopole ranging. We introduce a novel, the unified AVS framework that mathematically models monopole near-field/far-field signatures and fuses pressure energy, particle kinematics, and frequency-dependent decay profiles to jointly estimate the range and DOA using a single sensor. Our key contribution enables single-sensor passive ranging without arrays, environmental priors, or reference signals.

The structure of this paper is arranged as follows: Sec. 2 outlines the fundamental equations governing the propagation of acoustic signals in the underwater environment; Sec. 3 presents an overview of the sensors employed in the detection and localization process, along with the challenges associated with their utilization. The concepts of near-field and far-field, as well as the relationship between pressure signal and PM signals within each field, are discussed in Sec. 4. Section 5 presents and evaluates our proposed methodology for monopole source ranging. Finally, Sec. 6 concludes the paper by summarizing the key findings and their implications.

2. Underwater acoustic wave propagation

Acoustic wave propagation in underwater environments originates from pressure disturbances at the source, governed by the wave equation under assumptions of a homogeneous, lossless, dispersionless, and unbounded medium (Abraham, 2019):

$$\Delta^2 p - \frac{1}{c}\frac{\partial^2}{\partial t^2} p = 0, \tag{1}$$

where p is the acoustic pressure, c is the sound speed in water, and t is time.

For a monopole point source (this study's model), the pressure at a distance r is

$$p(r,t) = \frac{p_0}{r}\cos\left(2\pi f t - kr\right), \tag{2}$$

where p_0 is the pressure magnitude at 1 m, f is the frequency, $k = 2\pi f/c$ is the wavenumber.

As the wave propagates, it induces oscillatory motion in water particles. The relationship between pressure and PM is defined by Euler's equation which is given as (Lin *et al.*, 2021):

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \mathbf{a} = -\frac{\nabla p}{\rho}, \tag{3}$$

where ρ is the water density, $\mathbf{v}$ is the particle velocity, and $\mathbf{a}$ is the particle acceleration.

Substituting Eq. (2) in Eq. (3) yields:

$$\mathbf{a}(r) = -i2\pi f \frac{p(r,t)}{\rho c}\left(1 + \frac{i}{kr}\right)\mathbf{u}, \tag{4}$$

where $\mathbf{u}$ is the radial unit vector in spherical coordinates. And the velocity formula $\mathbf{v}(r,t)$ is given as

$$\mathbf{v}(r) = -\frac{p(r,t)}{\rho c}\left(1 + \frac{i}{kr}\right)\mathbf{u}. \tag{5}$$

Another important term is the intensity $\mathbf{I}$, representing the power flux per unit area, is the time-average product of pressure and particle velocity

(Abraham, 2019). It is given by the following equation (Nedelec *et al.*, 2021; Hovem, 2007):

$$\mathbf{I}(r) = \frac{1}{2\rho c}\frac{p_0^2}{r^2}\mathbf{u}. \tag{6}$$

This energy propagates as potential and kinetic energy. While the first corresponds to pressure and is more likely to be measured by hydrophones, the second corresponds to the PM and is more likely to be measured by PM sensors. The formula of potential energy density $\overline{E}_{\text{pot}}$ is given as (Nedelec *et al.*, 2021):

$$\overline{E}_{\text{pot}} = \frac{1}{2\rho c^2}p_{\text{rms}}^2 = \frac{1}{4\rho c^2}\frac{p_0^2}{r^2}, \tag{7}$$

where $p_{\text{rms}} = \frac{p_0^2}{\sqrt{2}r}$ is the root-mean-square pressure.

And the formula of kinetic energy density $\overline{E}_{\text{kin}}$ is given as (Nedelec *et al.*, 2021):

$$\overline{E}_{\text{kin}} = \frac{\rho}{2}v_{\text{rms}}^2 = \frac{1}{4\rho c^2}\frac{p_0^2}{r^2}\left(1+\frac{1}{(kr)^2}\right), \tag{8}$$

where v_{rms} is the root-mean-square PM velocity.

Critically $\overline{E}_{\text{pot}}$ decays solely with distance ($\propto \frac{1}{r^2}$), while $\overline{E}_{\text{kin}}$ exhibits frequency-dependent and distance-dependent decay. This fundamental contrast in energy decay profiles underpins our proposed range-estimation method exploiting the near-field PM dominance.

3. Sensors employed for acoustic source detection and localization

Underwater acoustic systems utilize two primary sensor types for detection and localization: pressure sensors (hydrophones) and acoustic vector sensors (AVS). These may be deployed singly or in arrays, with the selection driven by application-specific requirements for precision, cost and environmental constraints.

3.1. Pressure sensors (hydrophones)

Hydrophones convert incident acoustic pressure waves into electrical signals via piezoelectric elements (Nedelec *et al.*, 2021). Under plane-wave conditions, the pressure p and the particle velocity v relate through the specific acoustic impedance $z_0 = \rho c$ as following:

$$p = z_0 v. \tag{9}$$

Hydrophones exhibit an omni-directional response when their size is small relative to the wavelength of the acoustic signal of interest. In practice, their frequency response typically ranges from a few hertz to several hundred kilohertz (Abraham, 2019; Saheban, Kordrostami, 2021), making them widely used in underwater detection systems.

3.2. Acoustic vector sensor AVS

AVS captures both pressure and vector PM (velocity/acceleration), enabling DOA estimation. Two implementation approaches exist: the inertial method and the pressure gradient method. The first method utilizes accelerometers or geophones to directly measure the particle acceleration or velocity. This approach contends with practical challenges including suspension system, geometry, and buoyancy (Gray *et al.*, 2016).

Alternatively, the pressure gradient method derives the particle velocity from spatial pressure differences. For the x-component the Euler equation yields:

$$\begin{aligned} a_x(0,t) &= \int_{\tau=0}^{t} v_x(t)\,\mathrm{d}\tau \\ &\approx \frac{1}{\rho}\frac{p\left(x+\frac{\Delta x}{2},t\right)-p\left(x-\frac{\Delta x}{2},t\right)}{\Delta x}, \end{aligned} \tag{10}$$

where Δx is the spacing between the two hydrophones.

Multi-axis particle measurements require additional hydrophones (e.g., Silvia *et al.* (2002) used six sensors). Challenges include optimal spacing, calibration, and bandwidth limitation (Nedelec *et al.*, 2021; Gray *et al.*, 2016).

The PM velocity or acceleration is an oscillatory directional quantity that exhibits 180-degree ambiguity. This ambiguity can be resolved by measuring the acoustic intensity, a non-oscillatory quantity that aligns with the direction of wave propagation (Nedelec *et al.*, 2021). Consequently, incorporating a pressure sensor with a multi-axis velocity or acceleration sensor results in an intensity vector sensor commonly referred to as an intensity probe or is a key component constituting the complete AVS system. Furthermore, the dipole directivity pattern (figure-of-eight response) inherent to PM sensors (Yuan *et al.*, 2022) provides a 4.8 reduction in isotropic ambient noise compared to omnidirectional (Levin *et al.*, 2012).

4. Near-field and far-field contrast

The PM equation, described by Eq. (5), governs acoustic wave propagation and reveals a fundamental contrast between the near-field and far-field regions surrounding a source. This equation comprises two primary terms: the first term $\frac{p(r,t)}{\rho c}$ represents the propagating acoustic wave (far-field component), while the second term $\frac{ip(r,t)}{\rho ckr}$ represents the local hydrodynamic flow (near-field component) (Kalmijn, 1988). Regions surrounding a source can be divided into three zones as:

– far-field ($kr \gg 1$): the local flow component becomes negligible compared to the propagating

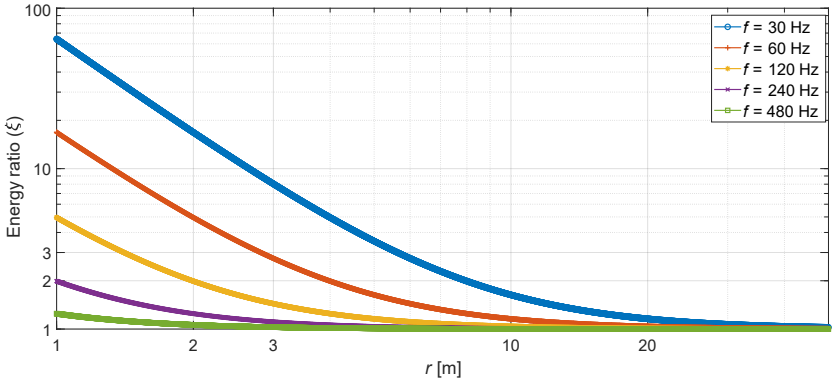

Fig. 1. Variation of energy ratio ξ with frequency and source distance for an acoustic monopole (logarithmic scale).

wave. Pressure and PM velocity are in phase and related by specific acoustic impedance $z_0 = \rho c$ as in Eq. (9) and it is real and constant;

– near-field ($kr \ll 1$): the local flow dominates over the propagation wave component. Pressure and PM velocity exhibit a quadrature relationship (90-degree phase difference), and the acoustic impedance becomes complex, frequency-dependent, and varies with distance r and is given as $z = (p(rt))/(v(rt)) = -i\rho ckr$. Characteristically, particle velocity decays faster with distance than pressure;

– transition zone ($kr \approx 1$): between these distinct regions lies a transition zone (intermediate zone) where neither component dominates completely.

The Helmholtz number (He = kr), representing the number of wavelengths within the distance r (Jansen *et al.*, 2017), is the key parameter distinguishing these regimes. Since $k = \frac{2\pi f}{c}$, He is not solely dependent on the distance r but also on the frequency of the signal. The value of r becomes particularly significant for low frequencies. For example, a 20 Hz source, the near-field is bounded by a distance r of approximately 12 m (considering $c = 1500$ m/s), whereas it is bounded by approximately 1 m for a frequency of 240 Hz.

The magnitude of PM velocity relative to pressure increases as the Helmholtz number decreases (He $\to 0$). Consequently, the contribution of kinetic energy to the total energy also increases. This relationship can be observed in the energy equations represented by Eqs. (7) and (8). To quantify this relationship, we defined the energy ratio ξ as the ratio of time-average kinetic energy density to potential energy density:

$$\xi = \frac{\overline{E}_{\text{kin}}}{\overline{E}_{\text{pot}}} = \rho^2 c^2 \frac{v_{\text{rms}}^2}{p_{\text{rms}}^2} = \left(z_0 \frac{v_{\text{rms}}}{p_{\text{rms}}}\right)^2 = 1 + \frac{1}{(kr)^2}. \quad (11)$$

In the far-field, $p_{\text{rms}} = z_0 v_{\text{rms}}$, leading to $\xi \approx 1$, indicating equipartition of energy. As He decreases, ξ increases significantly, reflecting a greater dominance of kinetic energy over potential energy as shown in Fig. 1. This figure graphically represents Eq. (11), it plots ξ against a distance r for selected frequencies on logarithmic axes, clearly showing this increase within the near-field. For instance, at $r \approx 8$ m and $f = 30$ Hz ($kr \approx 1$), $\xi \approx 2$, meaning that the kinetic energy is nearly twice the potential energy.

This energy distribution difference has implications for the sensor SNR (signal-to-noise ratio). Consider a source producing the potential energy E_{pot} and the kinetic energy $E_{\text{kin}} = \xi E_{\text{pot}}$ at the sensor location. Under isotropic ambient noise conditions (Levin *et al.*, 2012), the kinematic noise energy $E_{n_{\text{kin}}}$ and potential noise energy $E_{n_{\text{pot}}}$ satisfy $E_{n_{\text{kin}}} = \frac{1}{3} E_{n_{\text{pot}}}$. Under these assumptions, the SNR at the input of a PM sensor SNR_v and a pressure sensor SNR_p satisfy the following equation:

$$\text{SNR}_v = \frac{E_{\text{kin}}}{E_{n_{\text{kin}}}} = 3 \cdot \xi \frac{E_{\text{pot}}}{E_{n_{\text{pot}}}} = 3 \cdot \text{SNR}_p \left(1 + \frac{1}{(kr)^2}\right). \quad (12)$$

This yields a substantial near-field SNR gain for PM sensors (>7.8 dB at $kr < 1$). While theoretically significant, practical limitations such as bandwidth constraints in pressure-gradient AVS implementations may mitigate this advantage.

5. Estimating source distance using energy ratio

While conventional AVS applications focus on detection and DOA estimation, this work proposes a novel method for estimating the distance to an acoustic source using a single AVS. This approach exploits the fundamental near-field energy relationship characterized by the ratio ξ (Eq. (11)), leveraging simultaneous pressure p and PM velocity $\mathbf{v}(t) = [v_x(t)\ v_y(t)\ v_z(t)]^{\text{T}}$ measurements intrinsic to the AVS.

5.1. Methodological framework

The processing chain (Fig. 2) follows these steps:

1) the AVS outputs four time-domain signals: pressure $p(t)$ and orthogonal velocity components $v_x(t)$, $v_y(t)$, $v_z(t)$, related by:

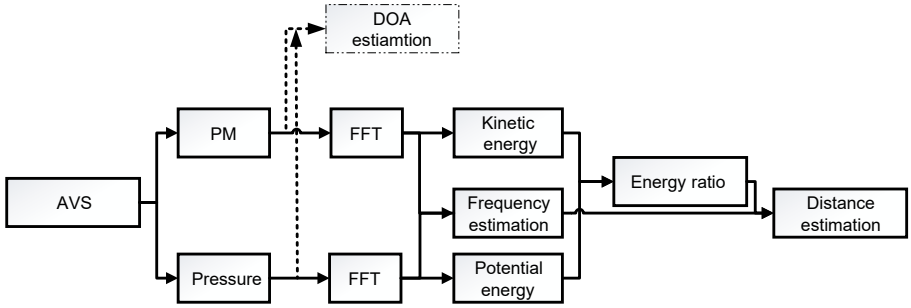

Fig. 2. Proposed processing chain for monopole range estimation using single AVS.

$$\mathbf{v}(t) = [v_x(t)\ \ v_y(t)\ \ v_z(t)]^{\mathrm{T}} = -\frac{p(t)}{\rho c}\left(1+\frac{i}{k}\right) \cdot [\cos\theta\cos\phi \quad \sin\theta\cos\phi \quad \sin\phi]^{\mathrm{T}}, \tag{13}$$

where ϕ is the elevation angle and θ is the azimuth angle. In the DOA task the estimation of these two angles are done;

2) for a tonal source at frequency f_s, spectrum estimation is performed using the fast Fourier transform (FFT) which also serves in frequency estimation $\widehat{f_s}$. The potential energy density estimate is calculated as

$$\widehat{E}_{\mathrm{pot}} = \frac{1}{2\rho c^2}\widehat{p}^2_{\mathrm{rms}} = \frac{1}{4\rho c^2}\left|P\left(f_s\right)\right|^2, \tag{14}$$

where $P(f_s)$ denotes the FFT coefficient of $p(t)$ at f_s.

The kinetic energy density estimate follows as

$$\widehat{E}_{\mathrm{kin}} = \frac{\rho}{2}\widehat{v}^2_{\mathrm{rms}} = \frac{\rho}{4}\left(|V_x(f_s)|^2 + |V_y(f_s)|^2 + |V_z(f_s)|^2\right), \tag{15}$$

with $V_i(f_s)$ representing FFT coefficients of velocity components $v_i(t)$, where i corresponds to the Cartesian coordinates x, y, or z, presents this method;

3) the energy ratio ξ is computed as:

$$\widehat{\xi} = \frac{\widehat{E}_{\mathrm{kin}}}{\widehat{E}_{\mathrm{pot}}}; \tag{16}$$

4) the Helmholtz number He = kr is estimated by inverting the monopole near-field relationship as following:

$$\widehat{\mathrm{He}} = \frac{1}{\sqrt{\widehat{\xi}-1}}; \tag{17}$$

5) finally, the range is derived:

$$\widehat{r} = \frac{c\widehat{\mathrm{He}}}{2\pi\widehat{f_s}}. \tag{18}$$

For M independent monopole sources emitting distinct, non-overlapping frequencies $\{f_{s,1}, ..., f_{s,M}\}$, the method estimates $\widehat{E}_k(f_{s,m})$ and $\widehat{E}_p(f_{s,m})$ across frequencies. Energy at each $f_{s,m}$ are isolated via frequency-bin selection, and Eqs. (16)–(18) are applied per source to estimate individual ranges $\widehat{r}_m$.

5.2. Performance validation

To validate the proposed method, we first performed numerical simulation of monopole radiation in a homogeneous medium. The following assumptions and configurations were adopted:

- sensor model: the AVs modeled as a co-located unit consisting of one omnidirectional pressure sensor and three orthogonal particle velocity sensors. The pressure and velocity components were assumed to be spatially collocated, consistent with an analytical model in Eq. (13);
- medium parameters: a homogeneous, isotropic medium with sound speed $c = 1500\,\mathrm{m\cdot s^{-1}}$ and density $\rho = 1000\,\mathrm{kg\cdot m^{-3}}$;
- source model: a monopole source emitting tonal signals at frequencies {30, 60, 120, 240, 480} Hz. The source is placed at $(\boldsymbol{\theta}, \boldsymbol{\phi})$ which are random generated, separated from AVS by a distance $r \in [1,\ 100]\,\mathrm{m}$, as shown in Fig. 3. Multipath and depth-related effects were neglected;
- noise model: independent additive white Gaussian noise (AWGN) was applied to each channel. SNR values tested were −6 dB, 0 dB, and 6 dB per channel;
- computational environment: simulation was implemented in Matlab. 1-second analysis window was used. For each configuration, 1000 Monte Carlo trials were run;
- performance metric: the range estimation error δ_{err} was quantified using the relative error defined in Eq. (19):

$$\delta_{\mathrm{err}}(\%) = \begin{cases} 100\dfrac{|r-\widehat{r}|}{r} & \xi \geq 1, \\ 100 & \xi < 1. \end{cases} \tag{19}$$

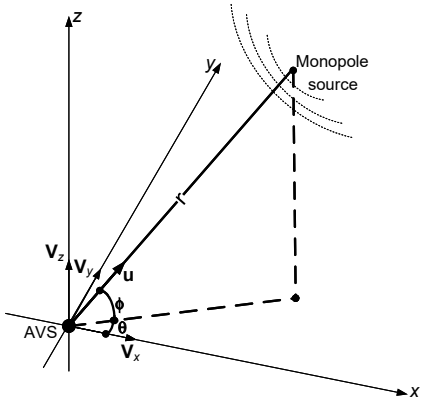

Fig. 3. Geometry of the monopole source relative to AVS. The Cartesian unit vectors ($\mathbf{V}_x$, $\mathbf{V}_y$, $\mathbf{V}_z$) define the sensor's coordinate frame. The source direction is described by the azimuth ($\boldsymbol{\theta}$) and elevation ($\boldsymbol{\phi}$) angles, the radial unit vector ($\mathbf{u}$), and the radial distance (r).

The resulting RMSE (root mean square error) of δ_{err} is calculated and logarithmically presented in Figs. 4–6, demonstrating the method's range-frequency dependence. Figure 5 shows that estimation with a 10 % error, for SNR = 0 dB, is achieved for distances up to 25 m when dealing with a source emitting 30 Hz frequency signal. This distance decreases to 6.6 m when the source frequency increases to 120 Hz. These results exhibit an enhancement when SNR increases: in Fig. 6, with SNR = 6 dB, the 10 % error is extended to 37 m at 30 Hz. In contrast, when SNR decrease to −6 dB (Fig. 4), the maximum distance decreases to 17 m. Overall, all curves in Fig. 5 will rightward shift with increasing the SNR (Fig. 6), while decreasing SNR causes the curves to shift towards the left (Fig. 4), confirming a strong SNR-frequency dependence.

5.3. Limitations and operational guidelines

The method achieves the highest accuracy where $kr \lesssim 1$ ($x \gtrsim 2$) exemplified by <6 % error at 30 Hz within 20 m. However, the far-field operation $kr \gg 1$ ($\xi \approx 1$) requires impractical SNR (SNR $\gg$ 0 dB).

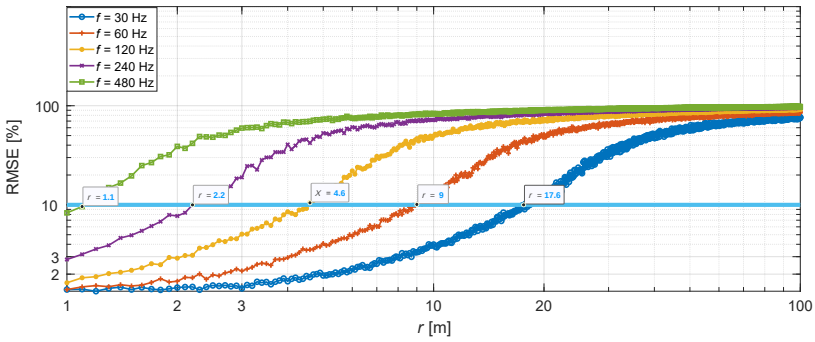

Fig. 4. Range estimation RMSE at SNR = −6 dB.

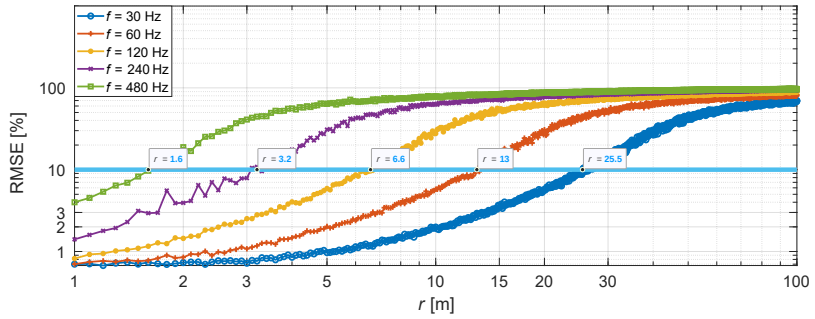

Fig. 5. Range estimation RMSE at SNR = 0 dB.

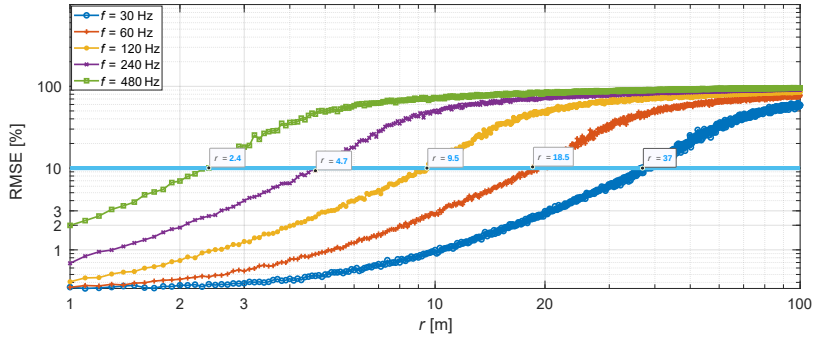

Fig. 6. Range estimation RMSE at SNR = 6 dB.

In high-frequency (≥ 480 Hz) or noisy (SNR $\ll 0$ dB) environments, it primarily functions as a proximity indicator. Accuracy assumes monopole-like radiation; dipoles/quadrupoles alter the ξ vs. kr relationship. Additionally, at low frequencies, large wavelengths yield multipath effects in bounded environments, degrading ranging affects this approach.

This technique extends AVS functionality beyond DOA, enabling single-sensor ranging where conventional methods fail – particularly valuable for near-field targets in constrained applications. Operational effectiveness peaks for low-frequency sources (<100 Hz) within 8 m–25 m ranges at 0 dB SNR.

5.4. Experimental validation with realistic noise profiles

To validate robustness beyond the isotropic additive Gaussian noise assumption used in simulations, experiments employed authentic ambient noise recorded from an operational test basin (25 m × 15 m × 2 m) using the AVS configuration characterized in (Mahmoud *et al.*, 2025) (see Fig. 4 for time-series and spectrogram representations).

Analysis of the realistic noise (Fig. 7) revealed the following key characteristics:

- pressure vs. velocity noise: noise in the pressure channel exceeds that in the velocity channels, consistent with its omnidirectional sensitivity;
- distinct self-noise profiles: the inherent self-noise characteristics differ between the pressure sensor and velocity sensor. The isotropic conditions are not satisfied (when calculating the pressure power and velocity power);
- velocity channel coherence: the three orthogonal velocity channels exhibit the same levels and waveforms noise;
- spectral tilt: noise energy decreased significantly with increasing frequency;
- tonal interference: prominent tonal interference was present.

To evaluate range estimation performance, we injected a directional tonal signal into the recorded noise. The amplitude was calibrated to achieve SNR = 0 dB when added to the AVS pressure channel noise. The same tonal signal, respecting its DOA, was injected into the velocity channel noise signals. We applied

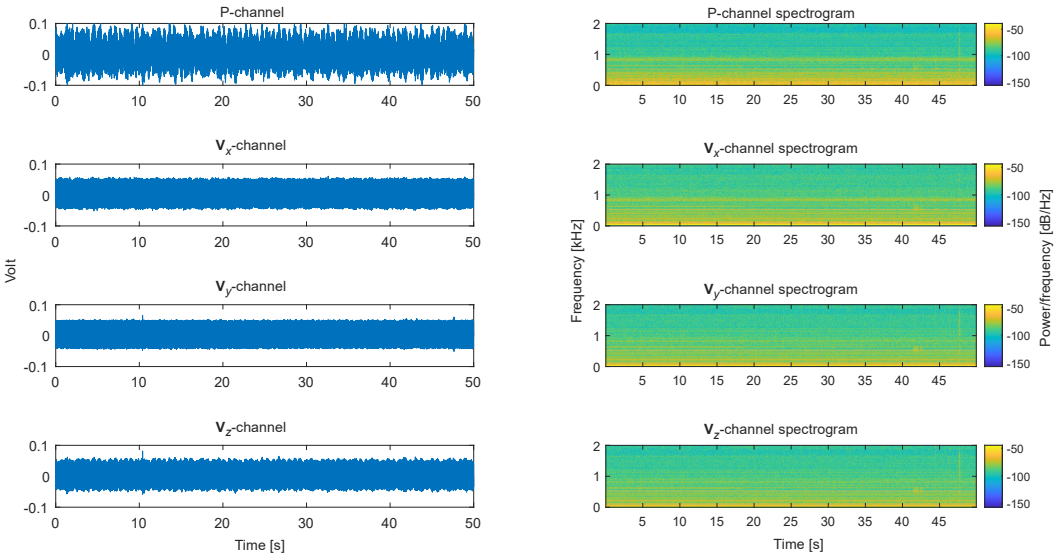

Fig. 7. Measured AVS noise signal.

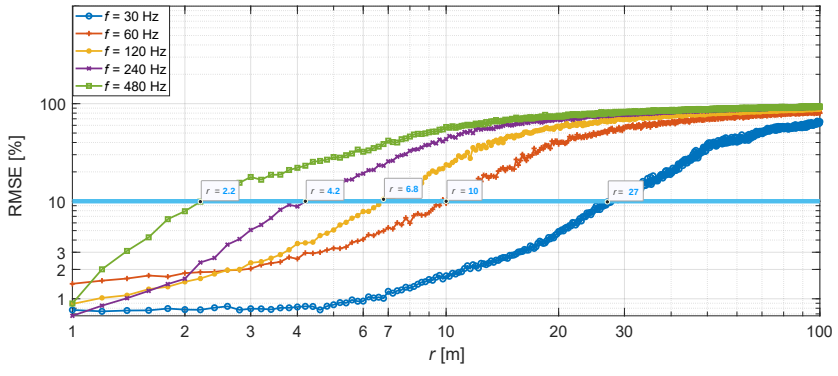

Fig. 8. Range estimation RMSE under realistic ambient noise (SNR = 0 dB).

the proposed algorithm to this combined signal-plus-noise data.

Figure 8 shows estimated range vs true range under realistic noise (SNR = 0 dB) for representative frequencies. Performance was assessed using 1000 Monte Carlo trials, each employing a different 1-second segment of the recorded noise. The results demonstrate that the proposed algorithm is significantly less affected by realistic noise compared to simulated AWGN conditions. For a 30 Hz tonal signal at a range of 27 m, the RMSE corresponds to less than 10 % relative error. This confirms the method's viability and robustness in non-ideal, real-world noise environments, extending beyond the limitations of theoretical AWGN assumptions.

6. Conclusion

This study has established a novel framework for passive monopole source ranging in underwater acoustics using a single AVS. By exploiting the fundamental near-field dominance of PM energy – quantified by the kinetic-to-potential energy density ratio (ξ) – we demonstrate that AVS measurements enable single-sensor range estimation where conventional hydrophone arrays fail. Key findings reveal:

1) PM SNR advantage: PM sensors achieve higher SNR than pressure sensors in the near field ($kr \lesssim 1$), validating the theoretical foundation for our approach;
2) accurate passive ranging: the proposed energy-decay method enables passive ranging up to 25 m for 30 Hz sources at 0 dB SNR with <10 % error;
3) real-noise robustness: validation using recorded basin noise profiles confirms the method viability despite violating the isotropic noise assumption.

While effective for near-field monopoles, limitations exist: performance degrades at high frequencies due to near-field contraction and in bounded environments where a low-frequency multipath distorts wave propagation. Future work will extend this framework to broadband sources and experimental validation in complex channels. This technique significantly advances passive sonar capabilities, enabling compact, cost-effective solutions for close-range surveillance.

Fundings

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

Conflict of interest

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

Data availability statement

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

Authors' contributions

Saier Mahmoud – methodology, software, hardware, writing (primary); Louay Saleh – hardware, supervision, conceptualization, revision; Ibrahim Chouaib – supervision, conceptualization, revising.

References

1. Abraham D.A. (2019), *Underwater Acoustic Signal Processing: Modeling, Detection, and Estimation*, Springer.
2. Cole A.M. (2019), *Automated open circuit scuba diver detection with low cost passive sonar and machine learning*, Msc. Thesis, Massachusetts Institute of Technology, Woods Hole Oceanographic Institution, United States.
3. Eleftherakis D., Vicen-Bueno R. (2020), Sensors to increase the security of underwater communication cables: A review of underwater monitoring sensors, *Sensors*, **20**(3): 737, https://doi.org/10.3390/s20030737.
4. Ferguson B.G., Lo K.W., Rodgers J.D. (2010), Sensing the underwater acoustic environment with a single hydrophone onboard an undersea glider, [in:] *OCEANS'10 IEEE Sydney*, https://doi.org/10.1109/OCEANSSYD.2010.5603889.
5. Gray M., Rogers P.H., Zeddies D.G. (2016), Acoustic particle motion measurement for bioacousticians: Principles and pitfalls, [in:] *Proceedings of Meetings on Acoustics*, **27**(1): 010022, https://doi.org/10.1121/2.0000290.
6. Hari V.N., Chitre M., Too Y.M., Pallayil V. (2015), Robust passive diver detection in shallow ocean, [in:] *OCEANS 2015 – Genova*, https://doi.org/10.1109/OCEANS-Genova.2015.7271656.
7. Hovem J.M. (2007), Underwater acoustics: Propagation, devices and systems, *Journal of Electroceramics*, **19**: 339–347, https://doi.org/10.1007/s10832-007-9059-9.
8. Jansen H.W., Brouns E., Prior M.K. (2017), *Vector sensors and acoustic calibration procedures*, TNO report (TNO 2017 R11589).
9. Jansen H.W., Prior M.K., Brouns E. (2019), On the conversion between sound pressure and particle motion, [in:] *Proceedings of Meetings on Acoustics*, **37**(1): 070012, https://doi.org/10.1121/2.0001280.
10. Jin B., Xu G. (2020), A passive detection method of divers based on deep learning, [in:] *IEEE 3rd International Conference on Electronics Technology*, pp. 650–655, https://doi.org/10.1109/ICET49382.2020.9119556.

11. Kalmijn Ad.J. (1988), Hydrodynamic and acoustic field detection, [in:] *Sensory Biology of Aquatic Animals*, pp. 83–130, Springer, https://doi.org/10.1007/978-1-4612-3714-3_4.
12. Kaushal H., Kaddoum G. (2016), Underwater optical wireless communication, IEEE Access, **4**: 1518–1547, https://doi.org/10.1109/ACCESS.2016.2552538.
13. Korenbaum V., Kostiv A., Gorovoy S., Dorozhko V., Shiryaev A. (2020), Underwater noises of open-circuit scuba diver, *Archives of Acoustics*, **45**(2): 349–357, https://doi.org/10.24425/aoa.2020.133155.
14. Krishnaveni V., Kesavamurthy T., B. Aparna (2013), Beamforming for direction-of-arrival (DOA) estimation – A survey, *International Journal of Computer Applications,* **61**(11): 4–11, https://doi.org/10.5120/9970-4758.
15. Levin D., Habets E.A.P., Gannot S. (2012), Maximum likelihood estimation of direction of arrival using an acoustic vector-sensor, *The Journal of the Acoustical Society of America*, **131**(2): 1240–1248, https://doi.org/10.1121/1.3676699.
16. Li X., Chen H., Lu H., Bi X., Mo Y. (2023), A method of underwater sound source range estimation without prior knowledge based on single sensor in shallow water, *Frontiers in Physics*, **11**: 1109220, https://doi.org/10.3389/fphy.2023.1109220.
17. Liang N., Zhou J., Yang Y. (2022), Single hydrophone passive source range estimation using phase-matched filter, *Journal of Marine Science and Engineering*, **10**(7): 866, https://doi.org/10.3390/jmse10070866.
18. Lin H., Bengisu T., Mourelatos Z.P. (2021), *Lecture Notes on Acoustics and Noise Control*, Springer.
19. Lohrasbipeydeh H., Gulliver T.A., Zielinski A., Dakin T. (2013), Single hydrophone passive source range and depth estimation in shallow water, [in:] *OCEANS 2013 MTS/IEEE Bergen*, https://doi.org/10.1109/OCEANS-Bergen.2013.6608042.
20. Mahmoud S., Saleh L., Chouaib I. (2025), Experimental results of diver detection in harbor environments using single acoustic vector sensor, *Archives of Acoustics*, **50**(2): 173–185, https://doi.org/10.24425/aoa.2025.153663.
21. Nedelec S.L. *et al.* (2021), *Best practice guide for underwater particle motion measurement for biological applications*, Technical report by the University of Exeter for the IOGP Marine Sound and Life Joint Industry Programme, https://www.researchgate.net/publication/356911609.
22. Saheban H., Kordrostami Z. (2021), Hydrophones, fundamental features, design considerations, and various structures: A review, *Sensors and Actuators A: Physical*, **329**: 112790, https://doi.org/10.1016/j.sna.2021.112790.
23. Silvia M.T., Richards R.T. (2002), A theoretical and experimental investigation of low-frequency acoustic vector sensors, [in:] *OCEANS '02 MTS/IEEE*, **3**: 1887–1898, https://doi.org/10.1109/oceans.2002.1191918.
24. Soldani M., Faggioni O., Zunino R., Carbone A., Gemma M. (2022), The LAMA system: A "smart" magnetometer network for harbour protection, *Journal of Applied Geophysics*, **204**: 104743, https://doi.org/10.1016/j.jappgeo.2022.104743.
25. Sutin A., Salloum H., Delorme M., Sedunov N., Sedunov A., Tsionskiy M. (2013), Stevens passive acoustic system for surface and underwater threat detection, [in:] *2013 IEEE International Conference on Technologies for Homeland Security* (*HST*), pp. 195–200, https://doi.org/10.1109/THS.2013.6698999.
26. Tichavsky P., Wong K.T., Zoltowski M.D. (2001), Near-field/far-field azimuth and elevation angle estimation using a single vector hydrophone, [in:] *IEEE Transactions on Signal Processing*, **49**(11): 2498–2510, https://doi.org/10.1109/78.960397.
27. Tu Q., Yuan F., Yang W., Cheng E. (2020), An approach for diver passive detection based on the established model of breathing sound emission, *Journal of Marine Science and Engineering*, **8**(1): 44, https://doi.org/10.3390/JMSE8010044.
28. Yuan M., Wang C., Da L., Li Q. (2022), Signal detection method using a single vector hydrophone in ocean acoustics, *The Journal of the Acoustical Society of America*, **152**(2): 789–798. https://doi.org/10.1121/10.0013219.
29. Zhao A., Ma L., Hui J., Zeng C., Bi X. (2018), Open-lake experimental investigation of azimuth angle estimation using a single acoustic vector sensor, *Journal of Sensors*, **2018**: 4324902, https://doi.org/10.1155/2018/4324902.

Archives of Acoustics Vol. 50, No. 4, pp. 501–512 (2025), https://doi.org/10.24425/aoa.2025.156931

Research Paper

Simulation Analysis of Beam Intensity Attenuation Patterns and Source Depth Estimation Using a Vertical Long Line Array

Hao WANG, Guangying ZHENG*, Fangwei ZHU, Xiaohong YANG, Shuaishuai ZHANG, Xiaowei GUO

Hangzhou Applied Acoustics Research Institution
Hangzhou, China

*Corresponding Author e-mail: 276454158@qq.com

Received May 29, 2025; revised September 30, 2025; accepted October 8, 2025; published online November 17, 2025.

In the deep-water reliable acoustic path (RAP), when estimating target depth using a vertical array, a large-aperture array can enhance the extraction of the acoustic field interference structure under low signal-to-noise ratio (SNR). However, this operation introduces slow envelope modulation (the envelope amplitude of peak beam intensity decreases with frequency) to the broadband acoustic field interference pattern, significantly degrading the performance of estimating the source depth. The Kraken normal-mode model can accurately calculate low-frequency sound fields in deep-water environments. This paper uses this tool to find that, in the deep-water direct arrival zone (DAZ), the peak beam intensity output of a vertical linear array varies across a broadband frequency range, exhibiting a pattern combining periodic changes of Lloyd's mirror interference and inherent envelope attenuation changes. The physical mechanism of envelope attenuation is explained through both theoretical derivation and simulation analysis, key factors affecting the envelope-attenuation pattern are clarified, and the impact of beam-intensity envelope attenuation on the depth-estimation method based on matched beam intensity processing (MBIP) is pointed out. Based on this, a modified target depth estimation method of matched beam intensity processing (M-MBIP) that contains an attenuation coefficient is proposed, and its effectiveness is verified through simulated data.

Keywords: deep-water direct arrival zone (DAZ); Lloyd's mirror interference; broadband attenuation pattern; source depth estimation.

1. Introduction

Three-dimensional acoustic target localization involves the estimation of azimuth, range, and depth, with target depth being a key indicator for surface and underwater target identification (Gaul *et al.*, 2007). Recently, underwater target depth estimation has gained significant attention from acousticians.

As a typical sound propagation mode in deep water, the reliable acoustic path (RAP) propagation mode is widely used for target detection in the upper water column (typically within 200 m from the surface). RAP-based target localization has two main advantages. First, the grazing angle of the received signal measured by a vertical array can be used to estimate the target range. Second, the acoustic signal radiated by a near-sea surface source propagates to a near-seabed receiver through the reliable acoustic path. The acoustic signal of the receiver mainly comes from the superposition of the direct acoustic signal and the sea surface-reflected acoustic signal, forming a typical Lloyd's mirror interference effect that produces distinct interference fringes in the acoustic field. These fringes are highly sensitive to changes in source depth (Worcester *et al.*, 2013). Due to these advantages, using Lloyd's mirror interference for estimating source depth has attracted extensive research (McCargar, Zurk, 2013; Li *et al.*, 2022; Duan *et al.*, 2012; Wei *et al.*, 2020).

McCargar and Zurk (2012) were the first to explore the use of Lloyd's mirror interference for estimating source depth, showing that for narrowband sig-

nals, acoustic intensity, as a function of range, is modulated by source depth. They proposed the generalized Fourier transform (GFT) method for depth estimation. Kniffin *et al.* (2016) later provided a theoretical analysis of the GFT method's performance and introduced a more straightforward depth-estimation technique based on the spacing of beam-intensity nulls. Lei *et al.* (2016) presented a passive source localization method that uses deep-water multipath RAP and cross-correlation matching for localizing source. Xu *et al.* (2023), addressing the performance degradation in GFT implementation in real-world deep-water environments, designed a preprocessing resampling scheme that enhances the periodicity of beam intensity in the grazing angle sine domain and improves depth-estimation accuracy when applied to GFT.

Zheng *et al.* (2020) pointed out that GFT is a typical non-perfect match from the generalized matched-field processing perspective. They proposed the matched beam intensity processing method (MBIP), an incoherent processing technique that matches data-beam intensity variations with those of assumed source depth, achieving better accuracy for near-surface source. Based on the research of Zheng *et al.* (2020), Zhou *et al.* (2022) proposed a depth estimation method that matches the interference structure in the frequency domain for narrowband source-depth estimation. This method can be used for real-time or semi-real-time source-depth estimation and classification. Wang *et al.* (2021) presented a broadband source-depth estimation method using the frequency-grazing angle interference structure to distinguish multiple underwater targets, validated by both simulation and experimental data.

The aforementioned methods and experiments were conducted using a pressure hydrophone array, while a vector hydrophone can simultaneously measure both acoustic pressure and particle velocity at the same point in the acoustic field. Zhang *et al.* (2025) addressed the passive detection problem using deep-water vector vertical arrays in a RAP environment, and proposed a coherent matched broadband complex acoustic intensity interference pattern (CM-BCAIIP) method for shallow-target depth estimation with high real-time capability. Sun *et al.* (2016) studied the distribution characteristics of the RAP vector acoustic field and estimated the range using derived from angle-of-arrival information from the horizontal and vertical components of complex acoustic intensity (see also, Zhu, Sun, 2023).

Whether using pressure-field or vector-field broadband interference structures for target- depth estimation in RAP, a low signal-to-noise ratio (SNR) causes large errors in the acoustic-field broadband interference structure extraction. This leads to poor performance in target-depth estimation. Although increasing the array aperture can increase array processing gains and improve the SNR of tracking beams, thereby enhancing the extraction of the broadband interference structure, this approach introduces slow envelope modulation. This causes the peak beam intensity to decay with frequency, which can significantly degrade the performance of traditional target-depth estimation methods.

To address this, this paper reviews Lloyd's mirror interference theory, pointed out the fast calculation equation for peak beam intensity, the attenuation law of peak beam intensity under vertical long array was briefly analyzed (which will be verified later), and based on this law proposes a M-MBIP method based on the MBIP approach. It then utilizes the Kraken normal-mode model to accurately compute the acoustic field at low-frequency (usually below 500 Hz) in the deep sea. This enables to analyze the key factors and patterns causing broadband attenuation of beam intensity through theory and simulation. Finally, simulation results are used to confirm that the proposed M-MBIP method is superior to the conventional MBIP method.

1.1. Lloyd's mirror interference theory

Duan *et al.* (2012) presented the conventional beamforming (CBF) output of a near-bottom vertical line array (VLA) arranged as shown in Fig. 1 (see also, Zhu *et al.*, 2021), under the assumption of a constant sound speed:

$$P(\omega, z_s, z_j) \approx -2iS(\omega)\frac{e^{ikR}}{R_s}\sin(kz_s \sin\theta_S), \tag{1}$$

$$B(\omega, \sin\theta, z_s) = \left|\sum_{n=1}^{N} e^{jk(nd-\overline{z})\sin\theta} P(\omega, z_s, z_j)\right|^2, \tag{2}$$

where N represents the number of array elements, d represents the spacing between array elements, $\overline{z}$ is the depth of the VLA center, θ is the grazing angle of the sound signal, $S(\omega)$ denotes the source strength, $P(\omega, z_s, z_j)$ represents the complex sound pressure received by the j-th hydrophone, $B(\omega, \sin\theta, z_s)$ is the beam intensity obtained after applying CBF to the complex sound pressure field recorded by the VLA, $\omega = 2\pi f$ is the angular frequency of the acoustic wave,

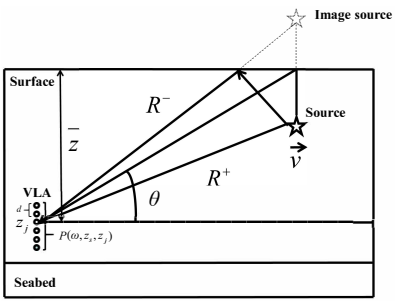

Fig. 1. Diagram of the source and vertical line array (VLA) geometry.

$k = k(\omega)$ is the wavenumber of the acoustic wave, z_j denotes the receiving depth of the j-th hydrophone, z_s is the source depth, $R_s = \sqrt{R^2 + z_j^2}$, R is the horizontal range between the source and the VLA, and $\sin\theta_S = \frac{\bar{z}}{\sqrt{\bar{z}^2+R^2}}$.

When the signal's detected grazing angle is θ_s, the peak beam intensity B can be expressed as (Zheng *et al.*, 2020):

$$B(\omega, \sin\theta_s, z_s) = 2\frac{|S(\omega)|^2}{\bar{z}^2}\sin^2\theta_s \cdot [1 - \cos(2kz_s\sin\theta_s)]. \quad (3)$$

Equation (3) represents the beam intensity under a constant sound speed. In the actual process of sound propagation, the change in the sound speed gradient causes acoustic ray refraction. The equation of peak beam intensity considering the bending of acoustic rays is as follows:

$$B(\omega, \sin\theta_s, z_s) = 2\frac{|S(\omega)|^2}{\bar{z}^2}\sin^2\theta_s \cdot \left[1 - \cos\left(2kz_s\sqrt{c_r^2/\bar{c}^2 + \sin\theta_s^2 - 1}\right)\right], \quad (4)$$

where c_r is the sound speed at the receiving depth, and $\bar{c}$ is the equivalent sound speed from the sea surface to the source depth, expressed as

$$\bar{c} = \sqrt{\left(z_s \Big/ \int_0^{z_s} 1/c^2(z)\,\mathrm{d}z\right)}. \quad (5)$$

Then, the frequency interference period of the peak beam intensity, considering the refraction of sound rays, can be expressed as

$$\Delta f_{\mathrm{PD}} \approx \frac{c_r}{2z_s\sqrt{c_r^2/\bar{c}^2 + \sin\theta_s^2 - 1}}. \quad (6)$$

It is clear from this formulation that the change of beam intensity with frequency is periodic whether under constant sound speed or varying sound speed. Therefore, the behaviour of beam intensity variation under broadband conditions can be studied based on either case.

1.2. Broadband modified MBIP target depth estimation method (M-MBIP)

Under the Lloyd's mirror interference theory, the MBIP method proposed in (Zheng *et al.*, 2020) is based on a small-aperture VLA. It constructs replica beam intensity time series (referred to as replica envelopes) at different depths and matches them with the actual output beam intensity time series from the array (referred to as data envelopes) to estimate the source depth. This process is completed through a fuzziness function similar to Eq. (10). The replica envelopes are calculated using Eq. (3). However, for large-aperture VLAs, the peak beam intensity attenuates with frequency increases after beamforming. In this case, the replica envelope calculated by Eq. (3) does not match the actual value, and using the MBIP method can lead to erroneous depth estimates. To solve this, this paper proposes a modified target depth estimation method based on the MBIP method, as detailed further.

When the target source is within 5 km of the VLA, the Kraken program (Porter, 1991) can be used. In the simulated marine environment, attenuated replica envelopes for different source depths can be calculated. By matching these attenuated replica envelopes with the data envelopes, the target depth can be estimated. For target sources at a range of 5 km–15 km from the array, the peak attenuation of its replica envelope is close to a constant value. By using Kraken to calculate the envelope attenuation coefficient of the replica at any of the above ranges and substituting it into Eq. (3), an approximate attenuated replica envelope can be obtained. Matching this approximate replica with the data envelope can quickly provide the target source depth while reducing computation time. The main steps of the proposed method are as follows:

1) estimate the target range r_e based on the VLA measurement of signal's arrival grazing angle θ_r;
2) for the given deep water environment, assuming a frequency band $\omega \in [\omega_l, \omega_h]$, array depth $z_r \in [z_{r1}, z_{rN}]$, target range r_e, and target depth $z_s \in [z_{s1}, z_{sN}]$, generate the broadband sound field $p(\omega; r_e, z_s, z_r)$ at a certain array element based on Kraken;
3) when the target depth $z_s = z$, the sound field matrix of the entire array can be represented as

$$\mathbf{p} = \left[p\left(\omega; r_e, z, z_r^{(1)}\right), p\left(\omega; r_e, z, z_r^{(2)}\right), ..., p\left(\omega; r_e, z, z_r^{(N)}\right)\right]^{\mathrm{T}}; \quad (7)$$

4) when the arrival grazing angle is θ_r, the peak beam intensity I of the array can be calculated as

$$I(\omega; z; \sin\theta_r) = \mathbf{w}'\mathbf{p}\mathbf{p}'\mathbf{w}, \quad (8)$$

where w is the steering vector for beamforming, which incorporates the spacing d between array components. The steering vector w (incorporating the spacing d between array components) is defined as

$$\mathbf{w} = \left[1, e^{jkd\sin\theta_r}, ..., e^{jk(N-1)d\sin\theta_r}\right]^{\mathrm{T}}; \quad (9)$$

5) calculate the ambiguity function of the broadband modified MBIP target depth estimation method by matching the peak beam intensity time series measured from data against a replica peak beam intensity time series evaluated for an assumed source depth, where its peak can be regarded as the real depth of the source:

$$M_{M-\mathrm{MBIP}}(z)=\begin{cases}\dfrac{\int_{\omega_l}^{\omega_h} I_{rp1}(\omega;z;\sin\theta_r)I_{\mathrm{data}}(\omega;z;\sin\theta_r)\,\mathrm{d}\omega}{\sqrt{\int_{\omega_l}^{\omega_h}|I_{rp1}(\omega;z;\sin\theta_r)|^2\,\mathrm{d}\omega}\sqrt{\int_{\omega_l}^{\omega_h}|I_{\mathrm{data}}(\omega;z;\sin\theta_r)|^2\,\mathrm{d}\omega}} & (r<5\,\mathrm{km}),\\ \dfrac{\int_{\omega_l}^{\omega_h}(1+\mu*\Delta f)I_{rp1}(\omega;z;\sin\theta_r)I_{\mathrm{data}}(\omega;z;\sin\theta_r)\,\mathrm{d}\omega}{\sqrt{\int_{\omega_l}^{\omega_h}|(1+\mu*\Delta f)I_{rp1}(\omega;z;\sin\theta_r)|^2\,\mathrm{d}\omega}\sqrt{\int_{\omega_l}^{\omega_h}|I_{\mathrm{data}}(\omega;z;\sin\theta_r)|^2\,\mathrm{d}\omega}} & (5\,\mathrm{km}\le r<5\,\mathrm{km}),\end{cases} \tag{10}$$

where I_{rp1} denotes the replica envelope, I_{data} denotes the data envelope, μ is the attenuation coefficient of the envelope, and Δf is the frequency interval.

Similar to MBIP, the depth corresponding to the peak of the ambiguity function is the estimated target depth.

2. Research and analysis of beam intensity broadband attenuation pattern

This section studies the mechanisms responsible for the attenuation of peak beam intensity in deep-water, large-aperture VLAs. It also analyzes the patterns of beam intensity attenuation under variations in array aperture, source depth, and other related factors.

2.1. Sound field interference structure for a long VLA

To reasonably analyze the factors influencing the extracted envelope of the sound field interference structure from a VLA, simulations are conducted using the Kraken normal-mode acoustic field calculation program. The simulation adopts a Munk sound speed profile typical of deep water, as shown in Fig. 2, with a critical depth of 4800 m. A 128-element VLA is laid near the seabed, with an element spacing of 5 m, the first element located at a depth of 4315 m and the last element at 4950 m, giving a total array aperture of 635 m.

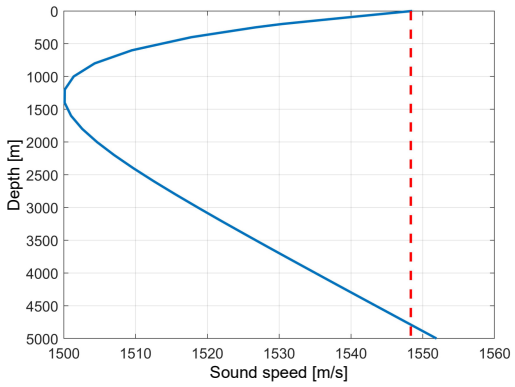

Fig. 2. Deep-water Munk sound speed profile.

In the simulation, the source is set at a depth of 50 m and a horizontal range of 7 km. Broadband array data spanning from 50 Hz to 200 Hz is generated using the Kraken program. CBF is applied to the received VLA data. The results are shown in Fig. 3, where display peaks (red stripes) correspond to the grazing angles of signal arrivals. Positive and negative values correspond to waves arriving from the sea surface and seabed directions, respectively. The peak beam exhibits pronounced interference in the frequency dimension.

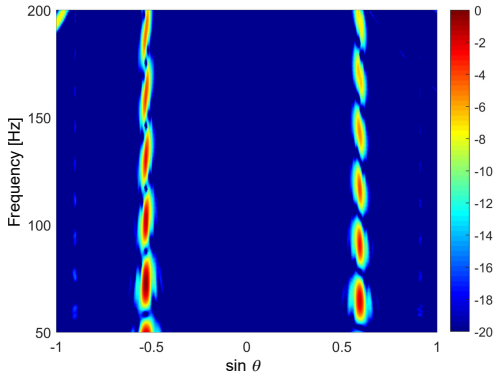

Fig. 3. Conventional beamforming output as a function of grazing angle and frequency.

The peak beam intensity at a grazing angle sine (+0.59) is extracted from Fig. 3 and shown in Fig. 4. It can be seen that the peak beam intensity changes periodically with frequency, while the envelope of the peak shows an almost linear attenuation. This attenuation pattern can lead to incorrect target depth estimation in MBIP, which could cause the omission of information necessary for target depth identification.

To explore the cause of the peak beam intensity attenuation, Fig. 5 shows the broadband transmission loss at different receiver depths corresponding to the source.

As shown in Fig. 5, the energy peaks correspond to two frequency points: 190 Hz (high-frequency) and

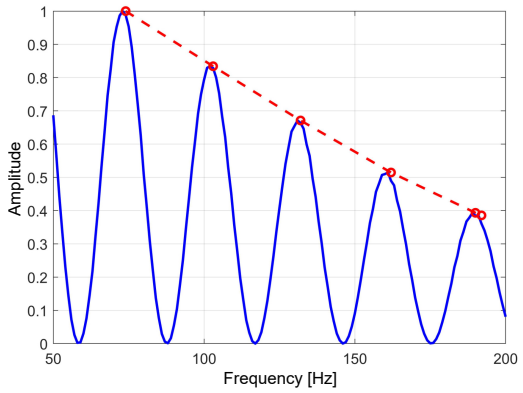

Fig. 4. Peak beam intensity variation with frequency (red dashed line in the figure shows the attenuation trend of the peak).

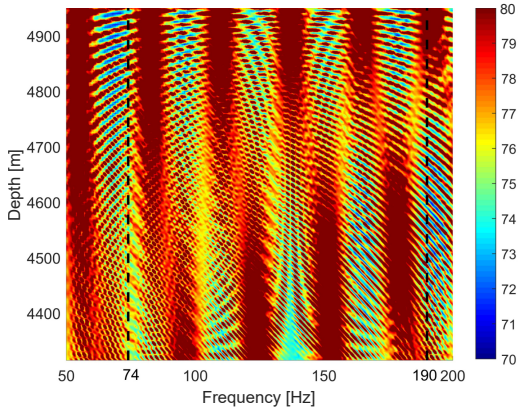

Fig. 5. Transmission loss at different depths.

74 Hz (low-frequency). Figure 6 further illustrates the variation of transmission loss with receiver depth. It is observed that the transmission loss at 74 Hz remains relatively stable across depths, whereas at 190 Hz, the transmission loss progressively increases with depth, leading to a gradual attenuation of beam energy. This phenomenon results in a reduction of peak beam intensity as frequency rises.

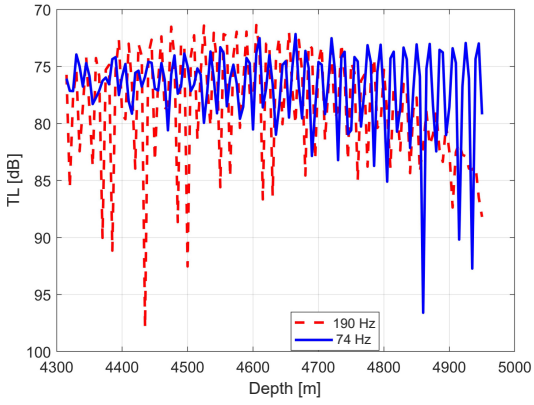

Fig. 6. Transmission loss of sound propagation at different frequencies.

It is evident that the transmission loss of high-frequency acoustic signals increases with depth. This is the main reason for the attenuation of peak beam intensity with frequency after beamforming using a long VLA.

2.2. Influence of array aperture on beam intensity attenuation pattern

In the simulation environment described in Subsec. 2.1, with the array element spacing and the depth of the first array element kept constant, CBF is performed for different array apertures: 32, 64, 96, and 128 elements. The resulting variation of peak beam intensity with range and frequency is shown in Fig. 7.

Observing the energy variation from low to high frequency in Fig. 7, it can be seen that at the same range, as the frequency increases, the peak beam intensity fluctuates periodically. Additionally, as the array aperture increases, the energy attenuation with frequency becomes faster. Figure 8 shows the variation of peak beam intensity with frequency at different ranges for 64 and 128 array elements. Comparing the two subfigures in Fig. 8, it can be seen that after 5 km, the peak beam intensity at high frequencies for 128 array elements is significantly lower than that for 64 array elements, and the peak attenuation is close to linear, as shown by the red dashed curve, while the red circles indicate the extremes.

2.2.1. Linear attenuation coefficient

To quantify the attenuation pattern of beam intensity across broadband frequencies, an in-band linear attenuation coefficient is defined, with the calculation method as follows (the μ in this section corresponds to the same variable previously defined in Subsec. 1.2):

$$\mu = \left(A_{h_p} - A_{l_p}\right) / \left(f_{h_p} - f_{l_p}\right), \tag{11}$$

where, assuming there are multiple extreme points in a frequency band of 50 Hz–200 Hz, A_{h_p} is the value of the last extreme point, A_{l_p} is the value of the first extreme point, f_{h_p} is the frequency corresponding to the last extreme point, and f_{l_p} is the frequency corresponding to the first extreme point.

The linear attenuation coefficients for the 64-element and 128-element arrays are calculated and shown in Fig. 9. It can be seen that the linear attenuation coefficient μ is related to the target range, with around 5 km acting as a critical point. When the range is less than 5 km, μ fluctuates greatly. When the range is greater than 5 km, it varies within a certain range. Moreover, the larger the array aperture, the greater the absolute value of μ, indicating that the beam intensity attenuates more rapidly with frequency change. This observation also confirms the effect of depth-dimension extension on peak beam intensity, as mentioned in Subsec. 2.1.

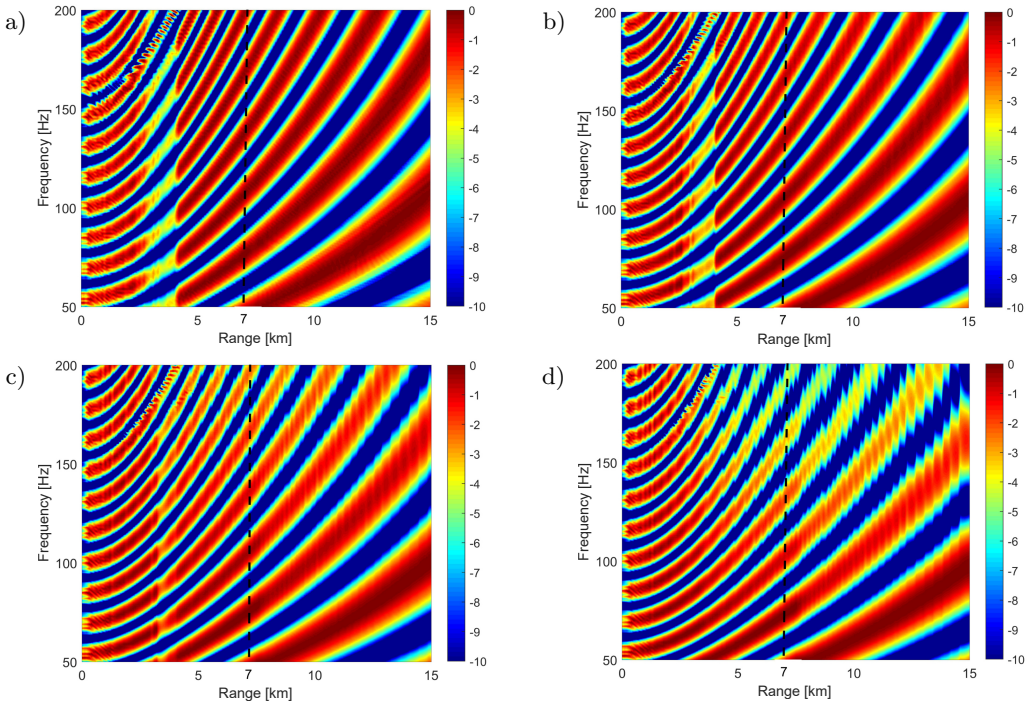

Fig. 7. Variation of peak beam intensity with range and frequency for different array apertures: a) 32 array elements; b) 64 array elements; c) 96 array elements; d) 128 array elements.

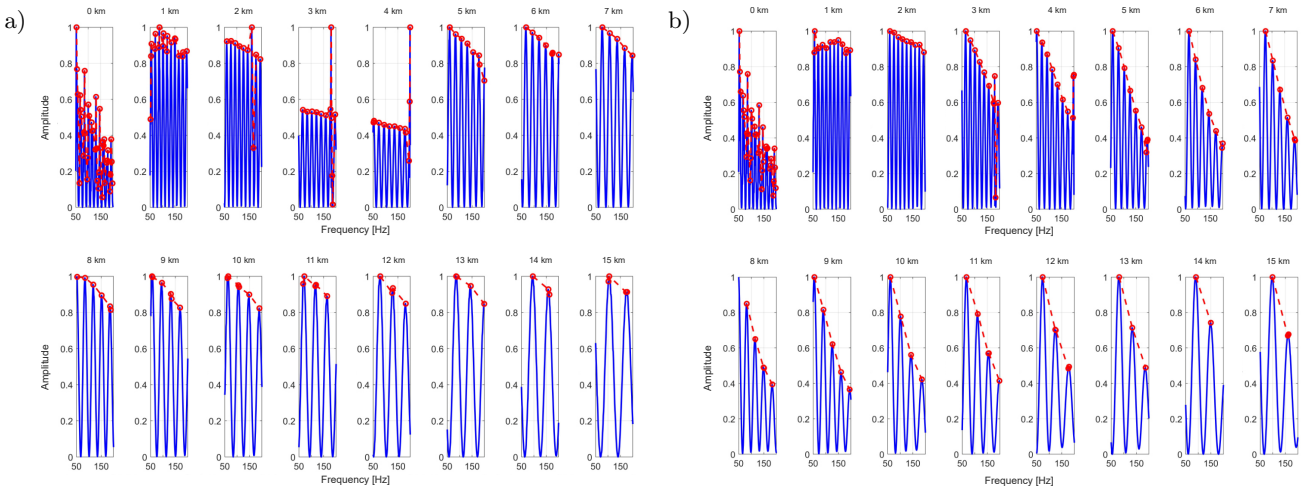

Fig. 8. Variation of peak beam intensity with frequency at different array apertures and horizontal ranges: a) 64 array elements; b) 128 array elements.

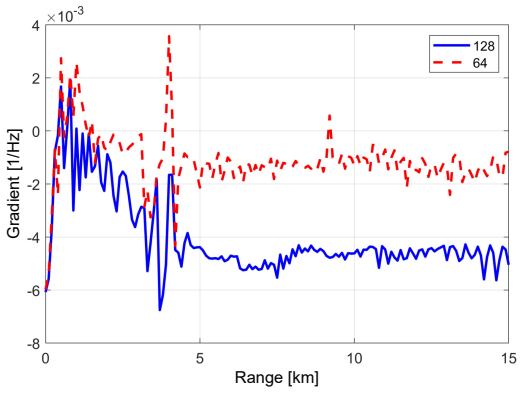

Fig. 9. Variation of linear attenuation coefficient with target range.

2.3. Influence of source depth on beam intensity attenuation pattern

Under the same simulation conditions and procedures as described in Subsec. 2.1, the source depth was varied at 50 m, 100 m, and 200 m. The variations of peak beam intensity as functions of range and frequency are presented in Fig. 10.

From Fig. 10, it can be observed that, at the same range, as the frequency increases, the peak beam intensity fluctuates periodically, the number of interference fringes increases significantly, and the amplitude of the energy gradually decreases.

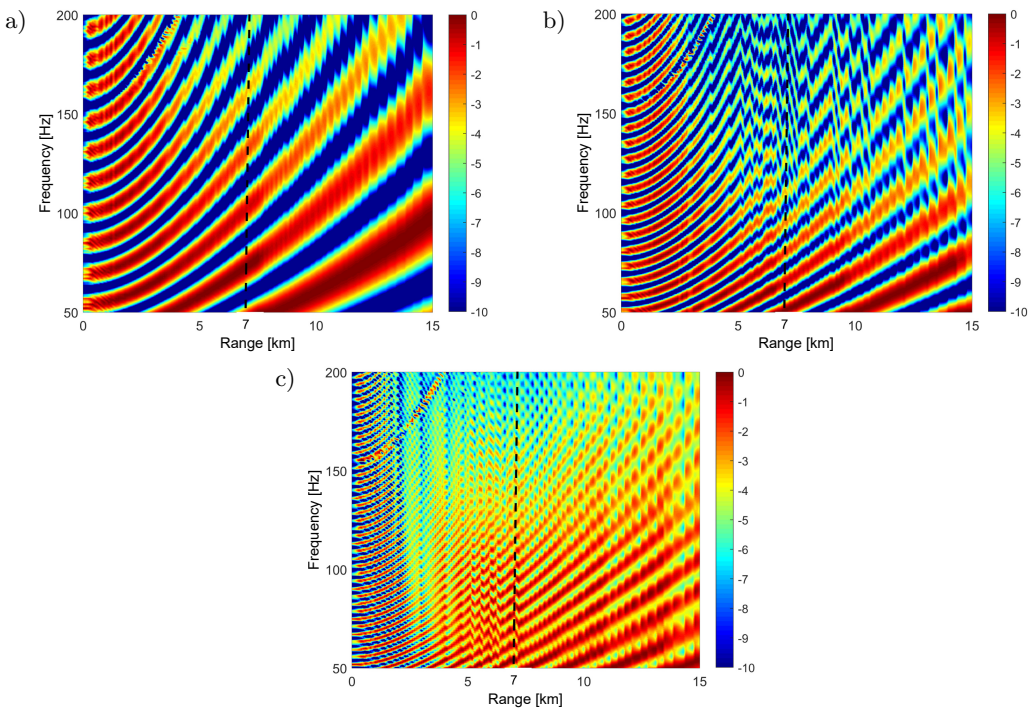

Fig. 10. Variation of peak beam intensity with range and frequency at target depths of: a) 50 m; b) 100 m; c) 200 m.

Furthermore, the peak beam intensity at a range of 7 km under different source depth conditions is obtained, and its variation with frequency is shown in Fig. 11.

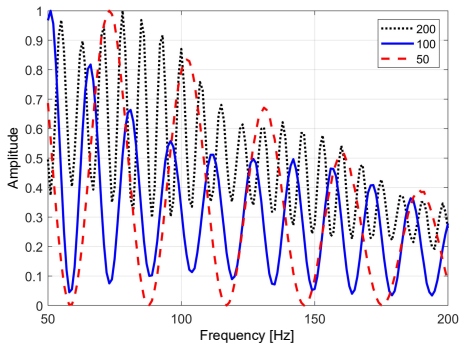

Fig. 11. Variation of peak beam intensity with frequency at 7 km range for different source depths.

It can be seen from Fig. 11 that as the source depth increases, the period of the sound field interference shortens and the number of interference fringes increases, which is consistent with the analysis shown in Fig. 10. The linear attenuation coefficients corresponding to different source depths are calculated, and the results are presented in Fig. 12. It can be seen that as the source depth increases, the absolute value of the attenuation coefficient decreases slightly, while the attenuation trends of different source depths are basically the same. The absolute value of the attenuation coefficient gradually increases within a range of 5 km, and beyond 5 km, it tends to a constant value.

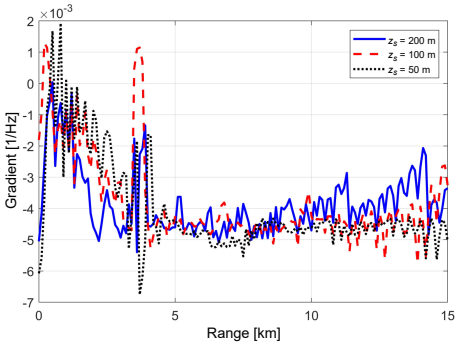

Fig. 12. Variation of linear attenuation coefficient with target for different target depths.

2.4. Influence of array depth on beam intensity attenuation pattern

This section analyzes the broadband attenuation pattern of beam intensity when the array deployment depth varies. Using the same simulation conditions as in Subsec. 2.1, the deployment depth of the array is varied by changing the depth of the 128-th element to 3950 m, 4450 m, and 4950 m. The variation of the peak beam intensity with range and frequency is shown in Fig. 13.

From Fig. 13, it can be seen that at the same range, the peak beam intensity fluctuates periodically from low to high frequency. However, as the array deployment depth increases, there is no obvious trend in the energy attenuation rate. Figure 14 further shows the variation of peak beam intensity with frequency at a range of 7 km for different array deployment depths.

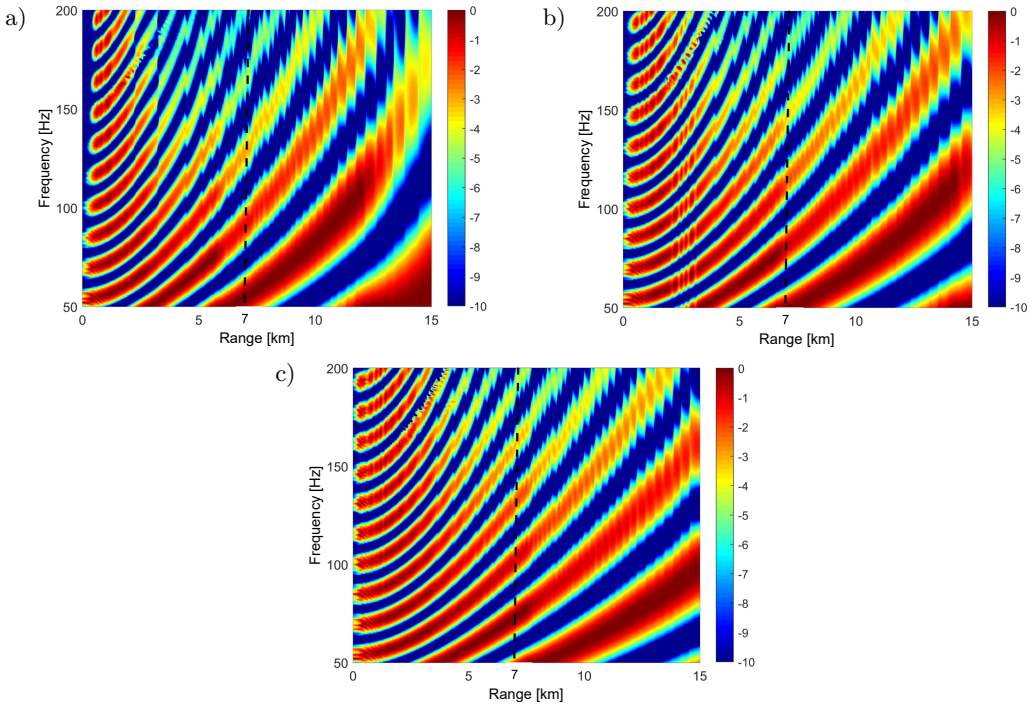

Fig. 13. Variation of peak beam intensity with range and frequency at different array deployment depths: a) 3950 m; b) 4450 m; c) 4950 m.

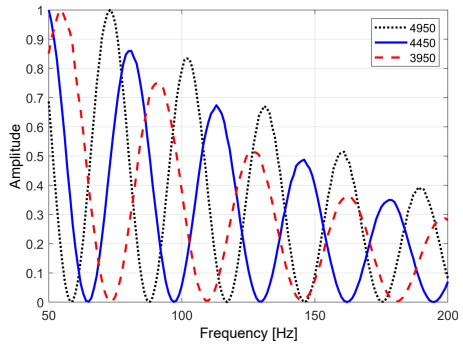

Fig. 14. Variation of peak beam intensity with frequency at 7 km range for different array deployment depths.

It can be seen from Fig. 14 that the attenuation trends of the peak beam intensities at 7 km under different source depths are basically the same. The linear attenuation coefficients corresponding to different array deployment depths are calculated, as shown in Fig. 15. It can be seen that the attenuation trends of the beam intensity under different array deployment depths remain basically the same, although the absolute value of the linear attenuation coefficient decreases as the deployment depth of the array increases.

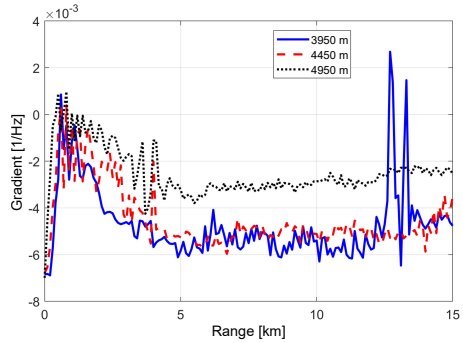

Fig. 15. Variation of linear attenuation coefficient with target range for different array deployment depths.

2.5. Influence of sound speed profile on beam intensity attenuation pattern

Given the unique characteristics of the deep-water DAZ, this study compares isovelocity (1510 m/s) with Munk sound speed profiles to illustrate the influence of the sound-speed profile on the broadband beam intensity attenuation pattern.

Except for setting the sound-speed gradient as a constant sound speed, the same simulation conditions are adopted as those in Subsec. 2.1. After generating broadband array data with Kraken and performing conventional beamforming, the peak beam intensity envelopes at different ranges are shown in Fig. 16.

Compared with Fig. 8b, when the range between the source and array is greater than 5 km, the beam intensity envelope shows nearly linear attenuation within 50 Hz–200 Hz.

Figure 17 further illustrates the variation of linear attenuation coefficients with range for both isovelocity and Munk profiles, showing similar trends in both cases.

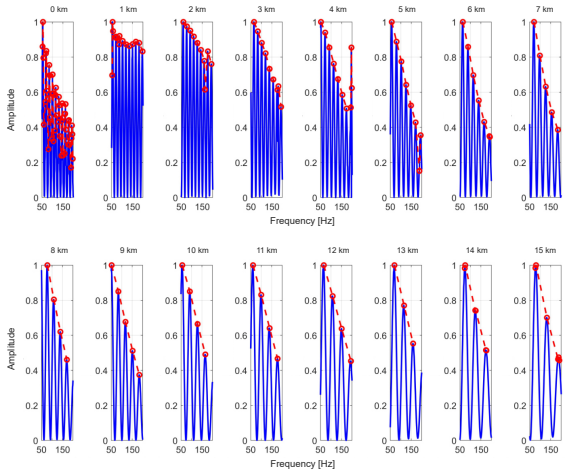

Fig. 16. Variation of peak beam intensity with range under constant sound speed.

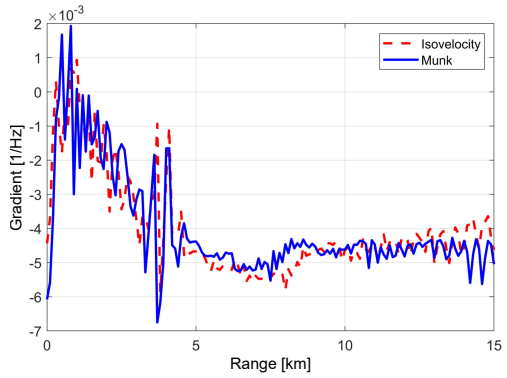

Fig. 17. Variation of linear attenuation coefficient with target range under profiles – isovelocity and Munk sound speed.

3. Simulation results for target depth estimation

To verify whether the depth-estimation performance of the M-MBIP method under a long VLA is better than that of the MBIP method, the Kraken program is used for simulation. The simulation selects a typical Munk sound channel, sets the source depths to 50 m, 100 m, and 150 m. The receiving array is a 128-element VLA, the depth of the first element is at 4315 m and an element spacing is 5 m. The corresponding center frequency is at 125 Hz, the receiving range is at 7 km, and Gaussian white noise with an SNR of −10 dB is added to the VLA data. After conventional beamforming, the results are presented in Fig. 18.

As known from Subsec. 2.1, the peak beam intensity at the sine of the grazing angle in Fig. 18 corresponds to waves incoming from the sea surface, which is used as the replica envelope of the M-MBIP method. At the same time, the replica envelope of the MBIP method is generated using Eq. (3), and both are compared with the data envelope of the array output (see Fig. 19).

It can be seen in Fig. 19 that the data envelope is basically covered within the envelopes of both M-MBIP and MBIP, but the envelope of M-MBIP has a higher degree of coincidence with the data envelope.

The source depth is estimated using the M-MBIP method and the MBIP method, as shown in Fig. 20.

It can be seen from Fig. 20 that the depth-estimation ambiguity function of the M-MBIP and

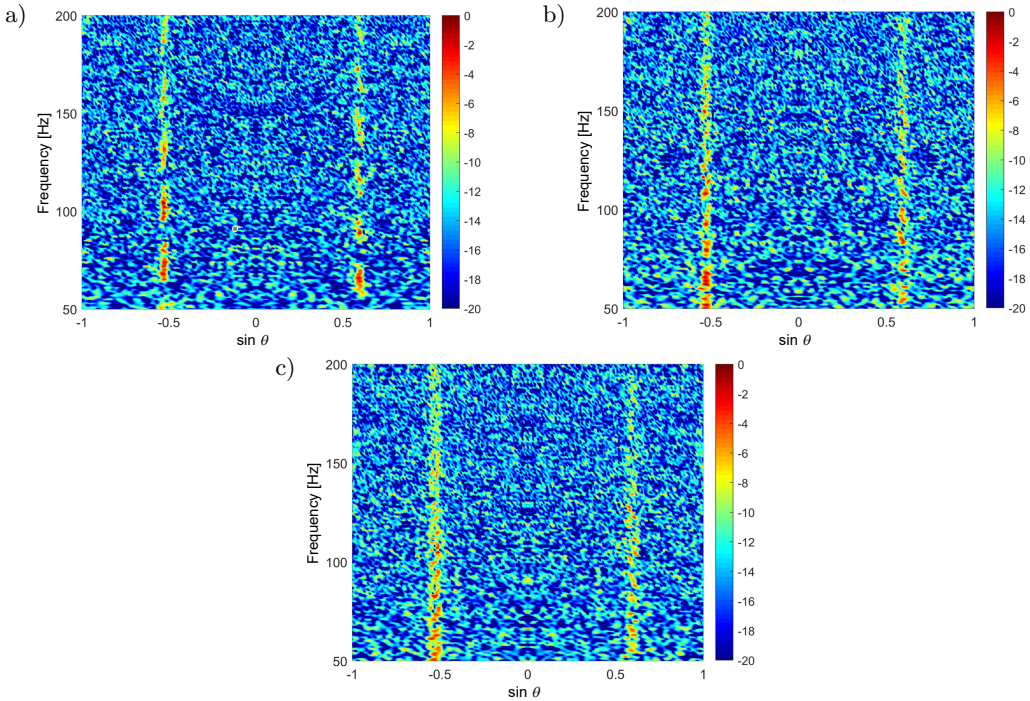

Fig. 18. Conventional beamforming output with an SNR of −10 dB at different source depths: a) z_s = 50 m; b) z_s = 100 m; c) z_s = 150 m.

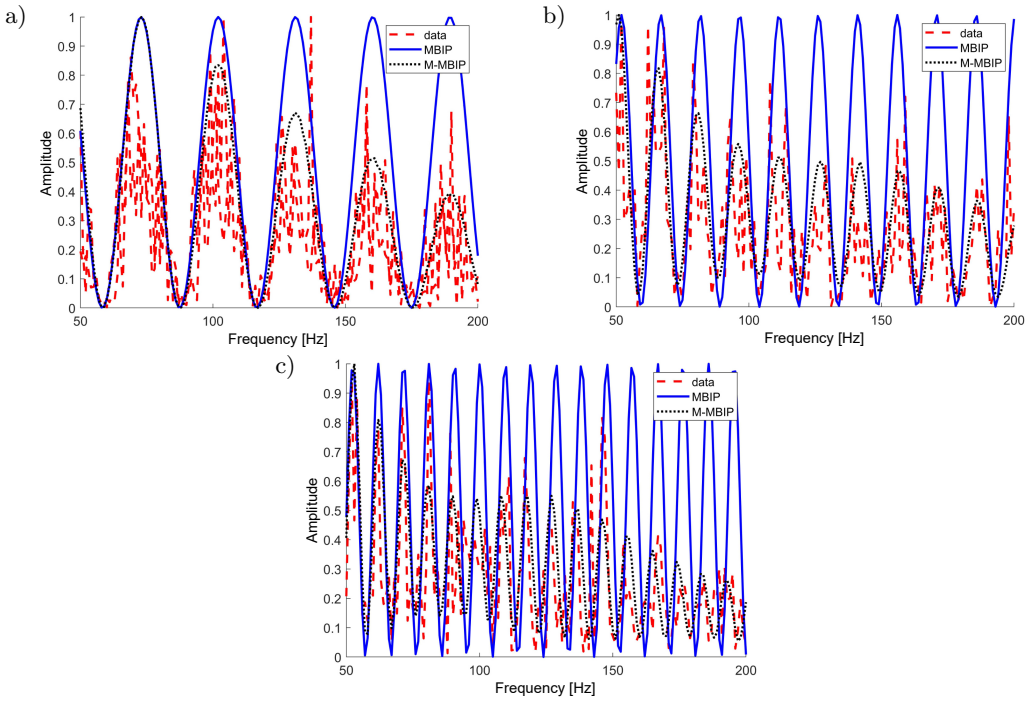

Fig. 19. Interference envelopes of beam energy for different source depths: a) z_s = 50 m; b) z_s = 100 m; c) z_s = 150 m.

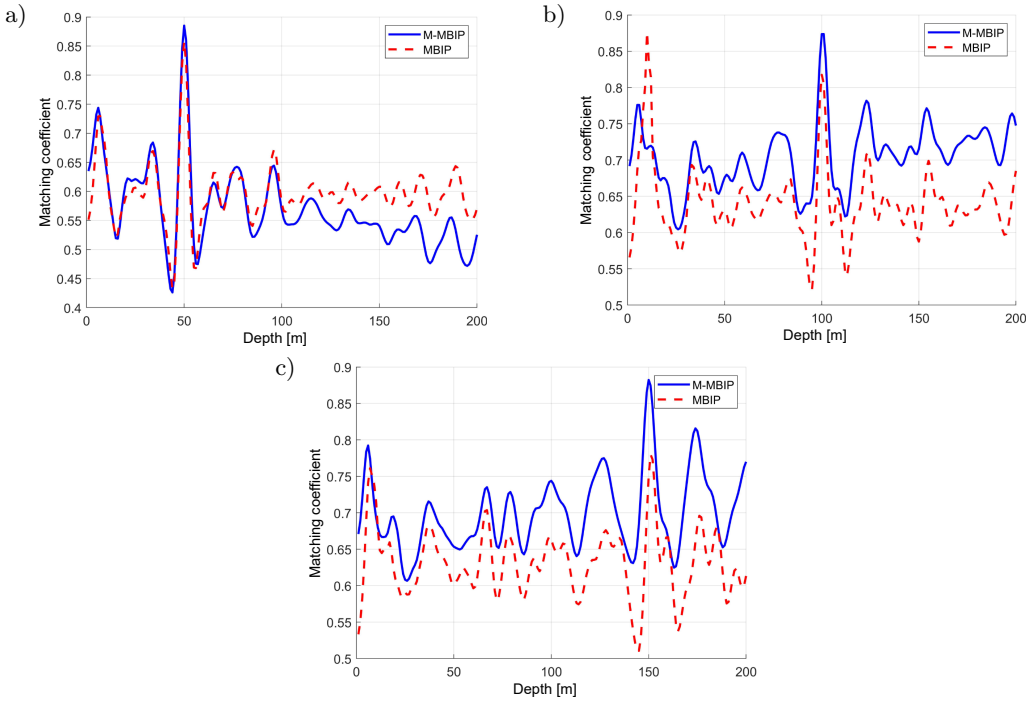

Fig. 20. Comparison of M-MBIP and MBIP methods for different source depths: a) z_s = 50 m; b) z_s = 100 m; c) z_s = 150 m.

MBIP methods have similar estimation capabilities when the source depth is 50 m. The peak of the M-MBIP method at the true depth is slightly higher than that of the MBIP method; when the source depth is 100 m or 150 m, the MBIP method performs poorly, with false peaks appearing at shallow depths, resulting in depth misjudgment problems. However, the peak of the M-MBIP method at the true source depth is always higher than that of the MBIP method, and no false peaks appear. Therefore, the depth-estimation performance of the M-MBIP method under a long VLA is better than that of the MBIP method.

4. Conclusion

This paper examined the issue of slow envelope modulation in the broadband interference structure of large-aperture VLA in deep water DAZ, which degrades source depth estimation. Through theoretical and simulation analyses, the key factors affecting the attenuation of peak beam intensity were identified, and a modified M-MBIP method based on MBIP was proposed. There are several conclusions that can be drawn:

1) the increase in transmission loss of high-frequency sound signal with depth is the key reason for the frequency-dependent attenuation of peak beams after beamforming in a VLA;
2) the attenuation rate of beam intensity is proportional to the array aperture, inversely proportional to the array deployment depth and source depth, and largely independent of the sound speed profile;
3) within a certain range, the M-MBIP method significantly outperforms the MBIP method in estimating source depth using a large-aperture VLA.

During data processing, it was observed that the replica envelope may exhibit a frequency shift relative to the data envelope. This phenomenon has the potential to impact the accuracy of source depth estimation methods. Additionally, the M-MBIP cannot be validated due to the lack of experimental data. Therefore, future research will focus on exploring the feasibility of target depth estimation under conditions of undersampling in the frequency domain of the sound field and verifying the effectiveness of the method with sea trial data.

Fundings

This work was supported by the National Natural Science Foundation of China, grant no. 12304501.

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

Hao Wang conceptualized the study, performed the analysis, and wrote the original draft. Guangying Zheng performed the analysis and contributed to simulation data interpretation. Fangwei Zhu conceptualized the study and performed the analysis. Xiaohong Yang contributed to simulation data interpretation. Shuaishuai Zhang wrote the original draft. Xiaowei Guo performed the analysis. All authors reviewed and approved the final manuscript.

References

1. Duan R., Yang K., Li H., Ma Y. (2017), Acoustic–intensity striations below the critical depth: Interpretation and modeling, *The Journal of the Acoustical Society of America,* **142**(3): EL245 https://doi.org/10.1121/1.5000325.
2. Duan R., Yang K.D., Ma Y.L., Lei B. (2012), A reliable acoustic path: Physical properties and a source localization method, *Chinese Physics B*, **21**(12): 276–289, https://doi.org/10.1088/1674-1056/21/12/124301.
3. Gaul R.D., Knobles D.P., Shooter J.A., Wittenborn A.F. (2007), Ambient noise analysis of deep-ocean measurements in the northeast pacific, *IEEE Journal of Oceanic Engineering*, **32**(2): 497–512, https://doi.org/10.1109/JOE.2007.891885.
4. Kniffin G.P., Boyle J.K., Zurk L.M., Siderius M. (2016), Performance metrics for depth-based signal separation using deep vertical line arrays, *The Journal of the Acoustical Society of America*, **139**(1): 418–425, https://doi.org/10.1121/1.4939740.
5. Li H. *et al.* (2022), A multi-step method for passive broadband source localisation using a single vector sensor, *IET Radar, Sonar & Navigation*, **16**(10): 1656–1669, https://doi.org/10.1049/rsn2.12287.
6. Lei Z., Yang K., Ma Y. (2016), Passive localization in the deep ocean based on cross-correlation function matching, *The Journal of the Acoustical Society of America*, **139**(6): EL196, https://doi.org/10.1121/1.4954053.
7. McCargar R.K., Zurk L.M. (2012), Depth-based suppression of moving interference with vertical line arrays in the deep ocean, *The Journal of the Acoustical Society of America*, **132**(3_Supplement): 2081, https://doi.org/10.1121/1.4755682.
8. McCargar R., Zurk L.M. (2013), Depth-based signal separation with vertical line arrays in the deep ocean, *The Journal of the Acoustical Society of America*, **133**(4): EL320, https://doi.org/10.1121/1.4795241.
9. Porter M.B. (1991), *The KRAKEN Normal Mode Program*, ACLANT Undersea Research Centre, La Spezia, Italy.
10. Sun M., Zhou S.-H., Li Z.-L. (2016), Near-surface source localization in the direct-arrival zone in deep water using a deep-located vector sensor, *Chinese Science: Physics, Mechanics, Astronomy*, **46**(9): 094309, https://doi.org/10.1360/SSPMA2016-00080.
11. Wang W.B. *et al.* (2021), A broadband source depth estimation based on frequency domain interference pattern structure of vertical array beam output in direct zone of deep water [in Chinese], *Acta Acustica*, **46**(2): 161–170, https://doi.org/10.15949/j.cnki.0371-0025.2021.02.001.
12. Wei R., Ma X., Li X. (2020), Depth estimation of deep water moving source based on ray separation,

Applied Acoustics, **174**(2): 107739, https://doi.org/10.1016/j.apacoust.2020.107739.

13. Worcester P.F. *et al.* (2013), The North Pacific Acoustic Laboratory deep-water acoustic propagation experiments in the Philippine Sea, *The Journal of the Acoustical Society of America*, **134**(4): 3359–3375, https://doi.org/10.1121/1.4818887.

14. Xu Z., Li H., Lu D., Duan R., Yang K. (2023), Beam intensity resampling-based source depth estimation by using a vertical line array in deep water, *Applied Acoustics*, **211**: 109495, https://doi.org/10.1016/j.apacoust.2023.109495.

15. Zhang J.N., Fang E.Z., Wang H., Gui C. (2025), Target depth estimation based on matched broadband complex acoustic intensity by deep water vector vertical array [in Chinese], *Acta Acustica*, **50**(3): 677–692, https://doi.org/10.12395/0371-0025.2024001.

16. Zheng G.Y., Yang T.C., Ma Q., Du S. (2020), Matched beam intensity processing for a deep vertical line array, *The Journal of the Acoustical Society of America*, **148**(1): 347–358, https://doi.org/10.1121/10.0001583.

17. Zhou L., Zheng G.Y., Yang T.C. (2022), Target depth estimation by frequency interference matching for a deep vertical array, *Applied Acoustics*, **186**: 108493, https://doi.org/10.1016/j.apacoust.2021.108493.

18. Zhu F.W., Zheng G.Y., Liu F.C. (2021), Matched arrival pattern-based method for estimating source depth in deep water bottom bounce areas [in Chinese], *Journal of Harbin Engineering University*, **42**(10): 1510–1517, https://doi.org/10.11990/jheu.202007001.

19. Zhu Q., Sun C. (2023), Underwater source localisation utilising interference pattern under low SNR conditions, *IET Radar, Sonar & Navigation*, **17**(5): 876–887, https://doi.org/10.1049/rsn2.12384.

Archives of Acoustics Vol. 50, No. 4, pp. 513–524 (2025), https://doi.org/10.24425/aoa.2025.156934

OSA 2025

Potential Applications of Ultrasonic Parametric Array Loudspeakers (PALs) in Room Acoustic Measurements

Filip WĘGRZYN*, Adam PILCH

AGH University of Krakow
Kraków, Poland

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

Received September 8, 2025; revised November 5, 2025; accepted November 12, 2025; published online November 20, 2025.

In this paper, the potential use of parametric array loudspeakers (PALs) in acoustic measurements of the room is analysed, especially in the assessment of the effectiveness of reflective panels and intentionally angled surfaces. PALs are sound sources capable of emitting highly directional acoustic beams within the audible frequency range. Their operation is based on the emission of a high-frequency (ultrasonic) carrier modulated so that, through nonlinear demodulation in air, audible sound is generated. This process results in a narrow, focused sound beam, enabling precise acoustic emission. To explore PALs potential for acoustic measurement applications, the propagation behaviour of PAL-generated signals is first investigated under free-field conditions, focusing on how different surface types influence sound reflection. Subsequent experiments are carried out in a controlled indoor space, where impulse responses are recorded for various beam incidence angles and receiver positions. The collected data are used to generate sound-level distribution maps, allowing for the visualization and quantification of reflected sound coverage areas. The results show that PALs produce beams with substantially reduced lateral dispersion compared to conventional loudspeakers, enabling precise identification of reflection points and incidence angles. This directional precision makes it possible to accurately assess how effectively the reflective acoustic elements and structures shape the sound field within the room. Overall, these findings may contribute to optimising sound design in acoustically complex environments.

Keywords: parametric array loudspeakers (PALs); room acoustics; ultrasonic; reflection; directivity.

1. Introduction

Localising acoustic flaws in rooms is one of the key problems in modern acoustics. Most measurements use omnidirectional speakers to record the room's impulse response, which can later be processed and analysed. However, the omnidirectionality of such a speaker causes the measurement to include reflections from every element in the room. As a result, the influence of certain elements may be masked in the recording (Gallien *et al.*, 2024). Using a speaker with high directivity may allow the user to examine only a chosen structure, such as a reflective panel or a diffusor. Additionally, high directivity enables easier tracking of the first reflection, which is, if not properly managed, one of the most common causes of acoustic flaws in rooms.

In order to obtain a sufficiently narrow beam that can generate sound waves in one direction only, some speakers utilise the parametric array effect. This effect was discovered in the early 1960s by Westervelt (1963). He demonstrated that, in theory, an end-fire array of virtual sources at the difference frequency can be produced by the interaction of two intense, collimated beams with slightly different high frequencies. These virtual sources arise because the instantaneous sound speed, which is one of the physical parameters depends inherently on the sound pressure or particle velocity. The resulting virtual end-fire array generated by this nonlinear interaction is therefore referred to as a parametric acoustic array, or simply a parametric array (Gan *et al.*, 2012a). When two primary waves of frequencies f_1 and f_2 ($f_2 > f_1$) are fully confined

beams, the angle at which the sound intensity of the difference frequency $f = f_2 - f_1$ is reduced by one-half (3 dB), is approximately given by

$$\theta_h \approx \sqrt{2\alpha_T/k}, \tag{1}$$

where k is the wavenumber of the difference-frequency wave, and α_T is the total sound absorption coefficient of the primary waves (Gan *et al.*, 2012b).

Amplitude modulation of ultrasonic carrier waves was introduced by Berktay (1965), who managed to substitute the tonal difference-frequency with a full frequency spectrum. Later, after Bennett and Blackstock (1975) successfully carried out the parametric array experiment in air (Gan *et al.*, 2012b), the phenomenon was utilised in audio applications by Yoneyama *et al.* (1983). When used as a parametric array loudspeaker (PAL), audible sound can be generated through the self-demodulation of the carrier's ultrasound and with the high directivity inherited from the parametric array (Ju, Kim, 2010).

Previous research in the field of room acoustics employed high-directivity speakers for tracing reflection paths or obtaining spatial impulse responses (Tervo *et al.*, 2009). However, while multiple studies analyse reflective elements mostly under laboratory conditions, in-situ measurements usually focus on tracing reflection paths in general, without assessing the performance of a singular reflector or diffusor. This paper studies potential applications of said speakers, focusing on the analysis of modulated sound reflections generated by PALs from different surfaces of varying sizes and materials.

The aim of this study is to determine whether the PALs can serve as a tool for identifying sources of acoustic flaws in rooms, such as determining the effectiveness of reflective panels or intentionally angled surfaces.

2. Laboratory measurements

All measurements were performed using a Videotel Digital HyperSound HSS 3000 speaker (Videotel Digital, 2014). First, the frequency response and directivity index were measured using a Klippel GmbH near-field scanner (NFS). The NFS performs holographic measurements of the near-field sound pressure to obtain a set of coefficients that precisely characterise the sound pressure at any point within the three-dimensional field outside the scanning surface. By leveraging the benefits of near-field measurements and applying spherical harmonic wave expansion, the near-field data can be extrapolated into the far field. This expansion enables high spatial resolution with fewer measurement points, and the post-processed results are both faster to obtain and more comprehensive than those from conventional directivity-measurement techniques (Login, 2015). The Klippel setup with the PAL mounted is shown in Fig. 1.

Fig. 1. Klippel GmbH near-field scanner measurement setup.

Secondly, the directivity of the sound reflected from different surfaces was measured in an anechoic chamber. The test surfaces included three plates: a wooden plate (140 cm × 140 cm), a wooden plate (40 cm × 40 cm), and an acrylic glass plate (40 cm × 40 cm). To test directivity, the PAL was placed at a 45° angle, 2.5 m from the plate placed in the middle of the chamber. The microphone was positioned on a crane-style arm on the opposite side of the chamber, oriented to record sound reflected from the plate at the chosen angle. The distance between the plate and the microphone was also equal to 2.5 m. In this setup, the crane-style arm was automatically repositioned after each measurement, maintaining the same distance while simultaneously adjusting the angle. Consequently, a directivity pattern of the reflected sound was achieved, with a resolution of 2° over the range of 15°–75°. A wider range was not necessary as signal levels outside this interval reached the noise floor values. The complete setup is shown in Fig. 2. All sounds were recorded using a GRAS 46AE 1/2″ CCP free-field standard microphone.

Fig. 2. Directivity of reflected sound, measurement setup in an anechoic chamber.

3. Results of laboratory measurements

For comprehensiveness, the obtained results are subdivided into two distinct segments. The initial seg-

ment shows speaker parameters acquired from the Klippel measurements, regarding the PAL's beamwidth and directivity, while the second segment presents polar plots of the reflected sound, generated from measurements performed in the anechoic chamber.

3.1. Directivity of a speaker

The Klippel NFS measurement was performed with a 5° resolution in a full 360° sphere around the speaker, with a frequency resolution of six points per octave. The frequency range was limited to 300 Hz–8000 Hz to shorten measurement time and match the frequency range of this PAL model, which starts at 300 Hz (Videotel Digital, 2014). Figure 3 shows the acquired directivity index (DI).

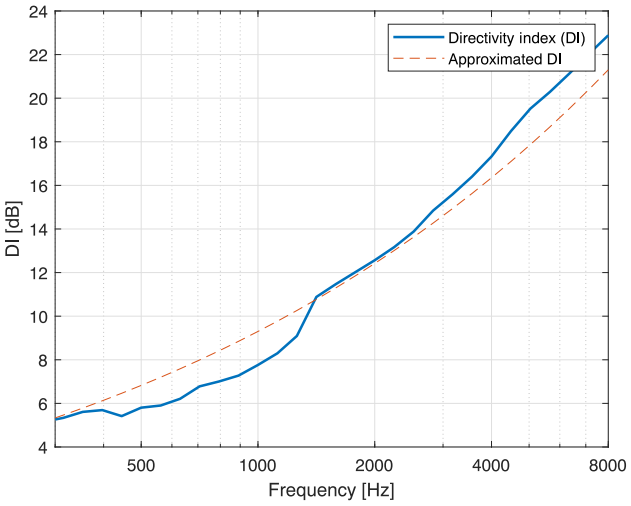

Fig. 3. Directivity index of the PAL (blue line) and approximated DI (red, dashed line).

The directivity index of the loudspeaker (Fig. 3) starts at around 5 dB at 300 Hz, then it reaches 10 dB at 1.5 kHz, and finally goes up to 23 dB at 8 kHz. The measured values were approximated with the function $12f^{1/3} - 2.7$, where f is the frequency in Hz. This approximation was based on the value of root mean square (RMS) difference between the measured and approximated values, which resulted in the lowest RMS value of 1.08 for the function given above.

A perfect theoretical cardioid has a directivity index of 4.8 dB; therefore, a directivity index of around 10 dB or more indicates a significantly directional loudspeaker (Vuine, 2024). Consequently, the analysed speaker achieves a highly concentrated beam only above 1.5 kHz. To study the precise beamwidth, a corresponding plot showing the PAL's beamwidth is presented in Fig. 4.

As illustrated in Fig. 4, for frequencies between 500 Hz–1000 Hz, the sound beam generated by the PAL reaches its maximum width of approximately 70° at a −6 dB sound pressure level (SPL) decrease. This

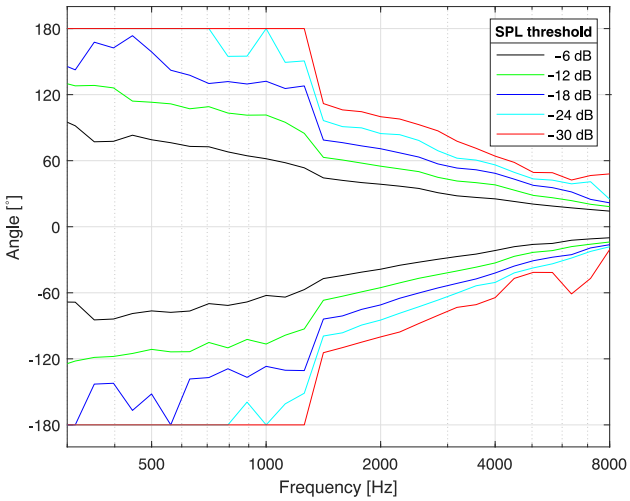

Fig. 4. Directivity (beamwidth) of the PAL.

threshold is usually taken as the beamwidth limit, since it is a value at which the signal power is halved (Keele, 2016). Afterwords, the beam slowly narrows, from 50° at 1.5 kHz to almost 20° at 5 kHz. Additionally, the SPL drop increases to −30 dB at just 40°. Sudden changes in beamwidth at lower SPL thresholds, visible around 1.5 kHz, are most likely caused by Klippel's internal algorithm switching to a different computation method.

3.2. Directivity of reflected sound

During the measurements in the anechoic chamber, directivity patterns of the reflected sound were obtained, each one for a different plate. The results are represented in the form of polar plots in Fig. 5, with six patterns corresponding to centre frequencies of octave bands ranging from 500 Hz to 16 kHz.

The widths of the reflected sound beam for a given frequency are summarised in Table 1. The limits for each beamwidth were assumed to be −6 dB on both sides, as in the Klippel measurement. In Fig. 5a, we can observe the directivity characteristic of the sound reflected from the large wooden plate. The strongest directivity is obtained at 16 kHz and 8 kHz. For the smaller wooden plate, the reflection pattern visible in Fig. 5b is nearly identical to the one obtained from the acrylic glass plate (Fig. 5c). Although the strongest directivity is also observed at the highest frequencies, at 500 Hz both the acrylic and small wooden plates have a reflected beamwidth of only 2°. This is most likely caused by the small size of the plates. Due to the larger area of the 140 cm × 140 cm wooden plate, the sound level measured in the 500 Hz bandwidth is substantially higher in comparison with two other plates. Additionally, there are no irregularities caused by an insufficient surface area size. Such irregularities would likely explain the unreasonably narrow angle observed for the small plates, since only a small

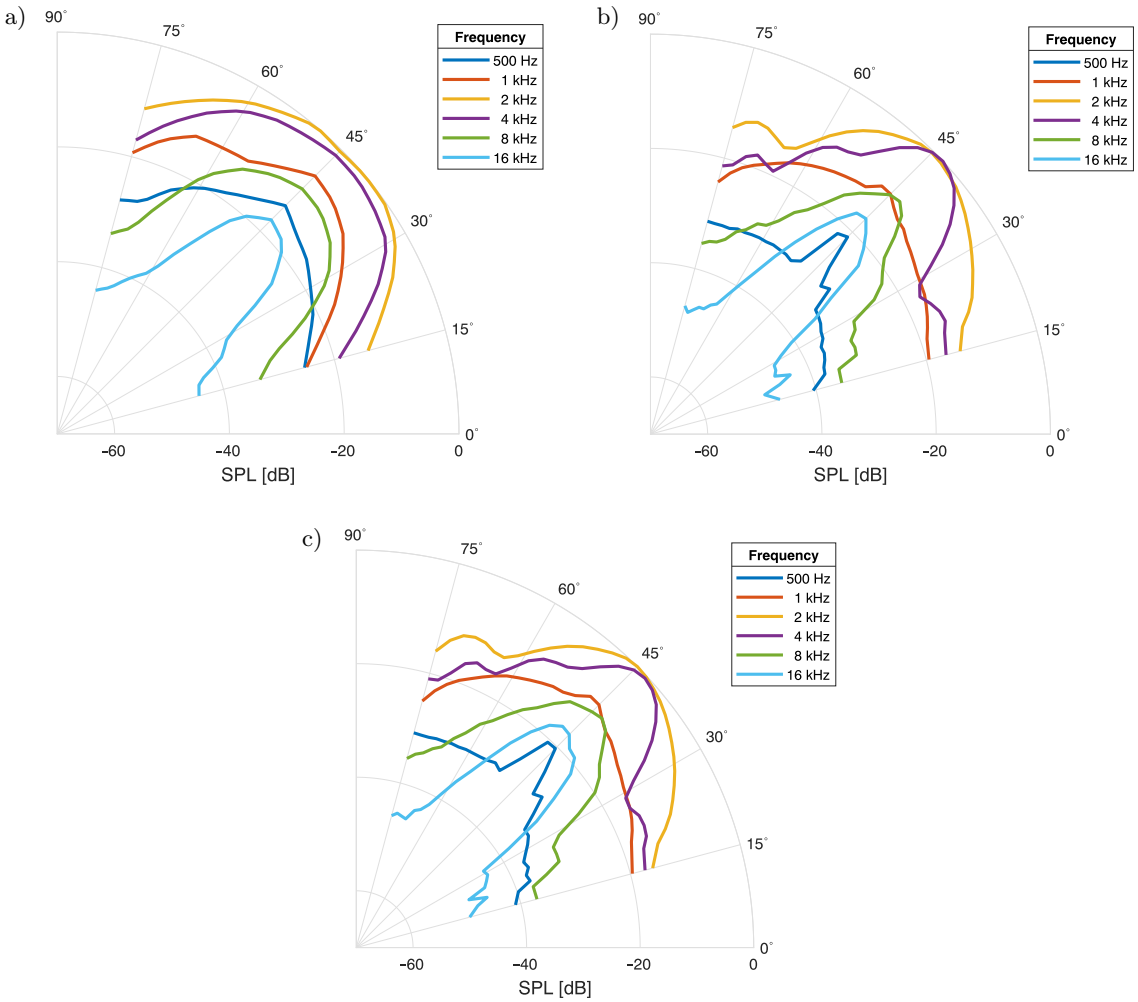

Fig. 5. Speaker angle −45°, polar plots of the PAL's directivity patterns reflected from: a) wooden plate (140 cm × 140 cm); b) wooden plate (40 cm × 40 cm); c) acrylic glass (40 cm × 40 cm).

Table 1. Width of the reflected beam for given frequencies and plate types.

Plate type	Reflected beamwidth angle					
	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	16 kHz
Wood – 140 cm × 140 cm	14°	14°	18°	18°	14°	10°
Wood – 40 cm × 40 cm	2°	18°	18°	10°	6°	6°
Acrylic – 40 cm × 40 cm	2°	18°	18°	10°	6°	6°

portion of the wave is reflected at the angle of incidence. It is apparent that the size of the sample affects the width of the reflected beam, which explains the substantial narrowing of the reflected angle compared to the beamwidth of the direct sound acquired with the Klippel system.

In summary, as long as the area of an analysed sample has a sufficiently large reflective area, there are no differences in the shapes of the polar patterns of the reflected sound. However, if the sample is too small, some frequencies will not be properly reflected, resulting in lower sound pressure levels and more irregular patterns. Furthermore, the obtained angles are significantly narrower than those acquired from the Klippel analysis. Larger sizes of the reflector result in a wider reflected sound beam.

4. Measurement of sound reflections from an angled ceiling

To verify the capabilities of parametric speakers in room acoustic analysis, a measurement of sound reflections from an angled ceiling was conducted. This measurement was performed in the WA3 classroom, inside the D1 building of AGH University of Krakow. The classroom has a part of ceiling angled at 5.25° to direct reflections from the lecturer to the students. The aim of the analysis was to determine whether the

angle of the ceiling is correct and whether the size of the plate is sufficient to distribute reflections from it evenly across the room.

To measure the sound reflected from the ceiling with a PAL, eight microphones were placed at random positions throughout the room. To avoid standing waves and reflections from the walls, all microphones were positioned off the main axis of the room and at least 1 m away from the walls. The height of each microphone was set to 1.2 m, which is the average height of a seated person (Rakerd, 2018). The parametric speaker was placed on a stand and angled so that the PAL aimed at the lower edge of the angled ceiling. The setup is visible in Fig. 6, while the theoretical model is shown in Fig. 7 (side view) and Fig. 8 (top view). For both source positions, five measurements of impulse responses were taken, each at a different PAL angle. The fifth measurement for position S1 was ignored due to obstruction from a projector in the path of the sound beam. A sine sweep from 300 Hz to 18 kHz was used as the excitation signal to match the bandwidth of this parametric speaker model (Videotel Digital, 2014).

Fig. 6. Experimental setup for measuring reflections from the angled ceiling in the classroom.

4.1. Results of measurements

From all recorded impulse responses, heatmaps of SPLs for each octave band were interpolated in the MATLAB programming environment. The maps are shown in Figs. 9–16. The horizontal line in each map marks the end of the angled ceiling. The interpolation area starts at the first row of desks. To limit the number of figures, results for the 8 kHz and 16 kHz bands were omitted, since these frequency bands are rarely used in room acoustic analysis.

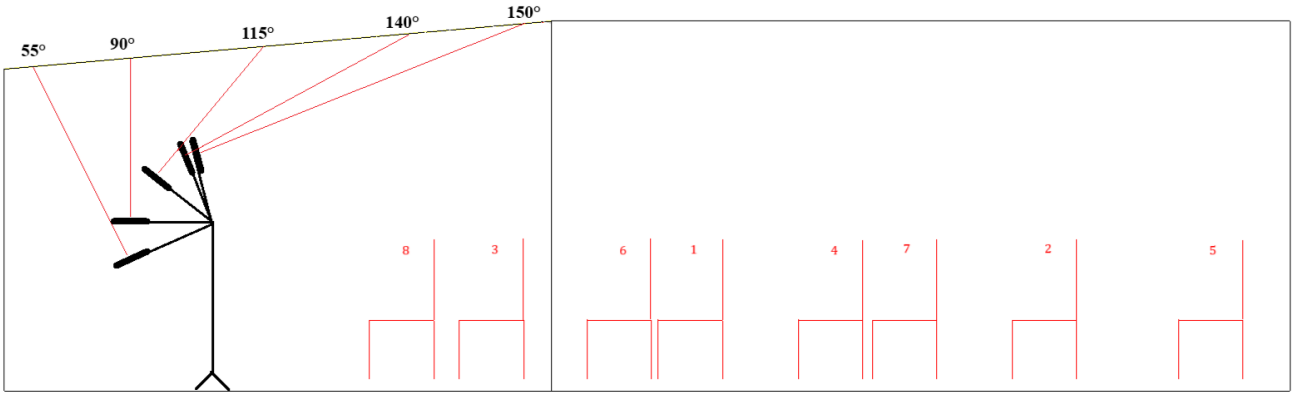

Fig. 7. Angled ceiling measurements, all microphone positions and speaker angles (side view).

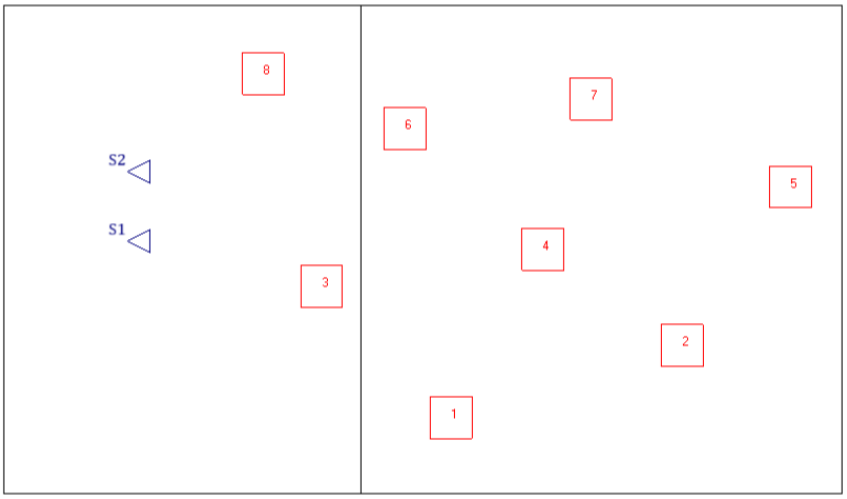

Fig. 8. Angled ceiling measurement, all microphone and speaker positions (top view).

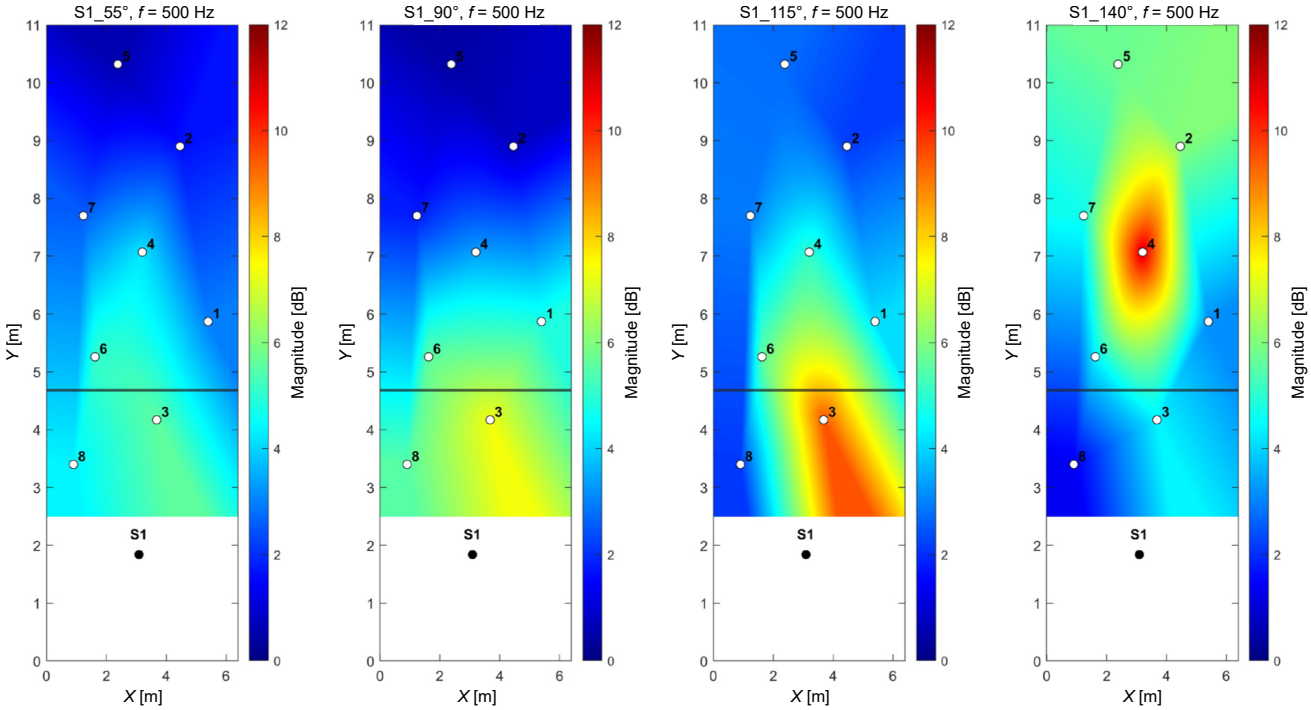

Fig. 9. SPL heatmaps for all PAL angles, 500 Hz octave band, source position – S1.

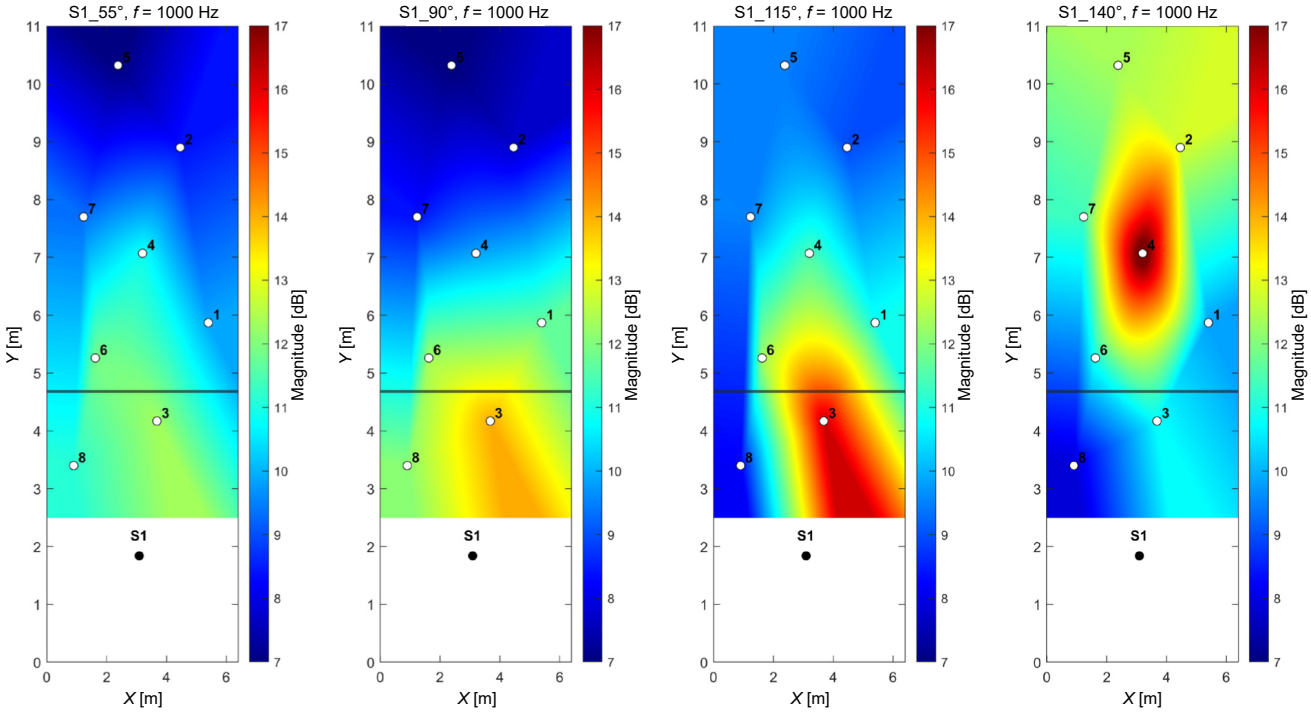

Fig. 10. SPL heatmaps for all PAL angles, 1 kHz octave band, source position – S1.

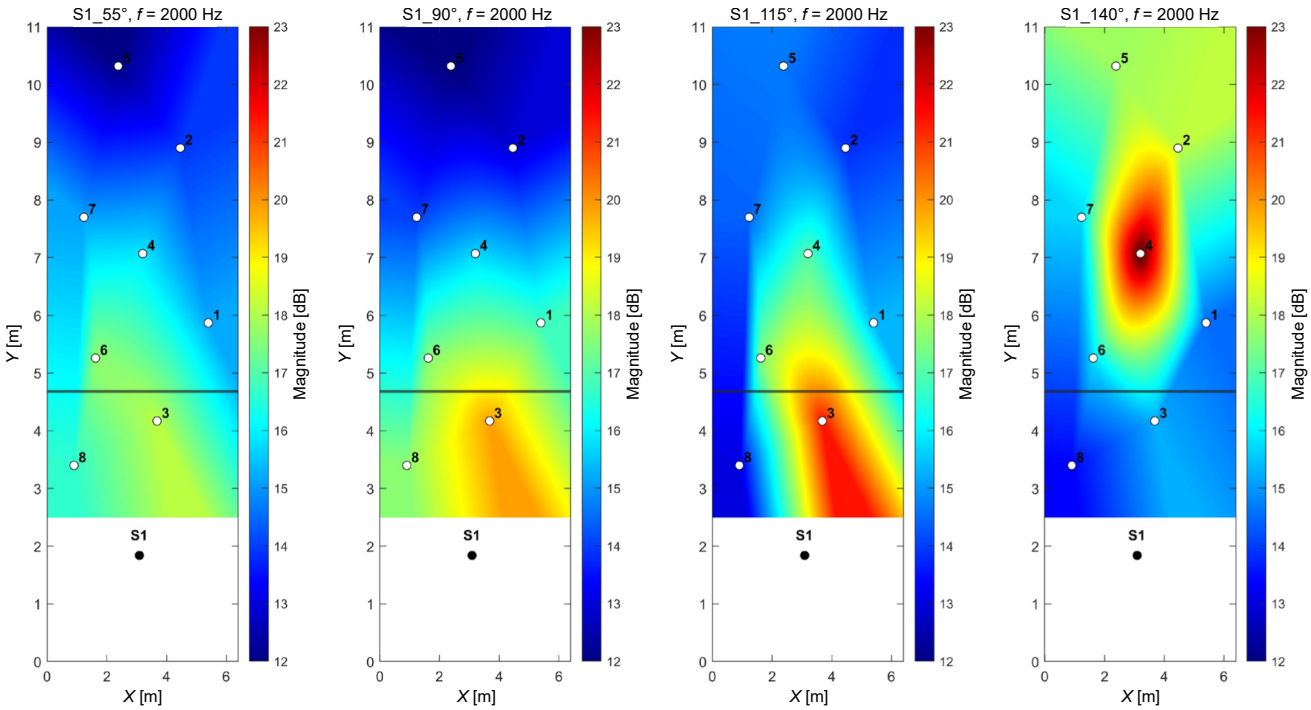

Fig. 11. SPL heatmaps for all PAL angles, 2 kHz octave band, source position – S1.

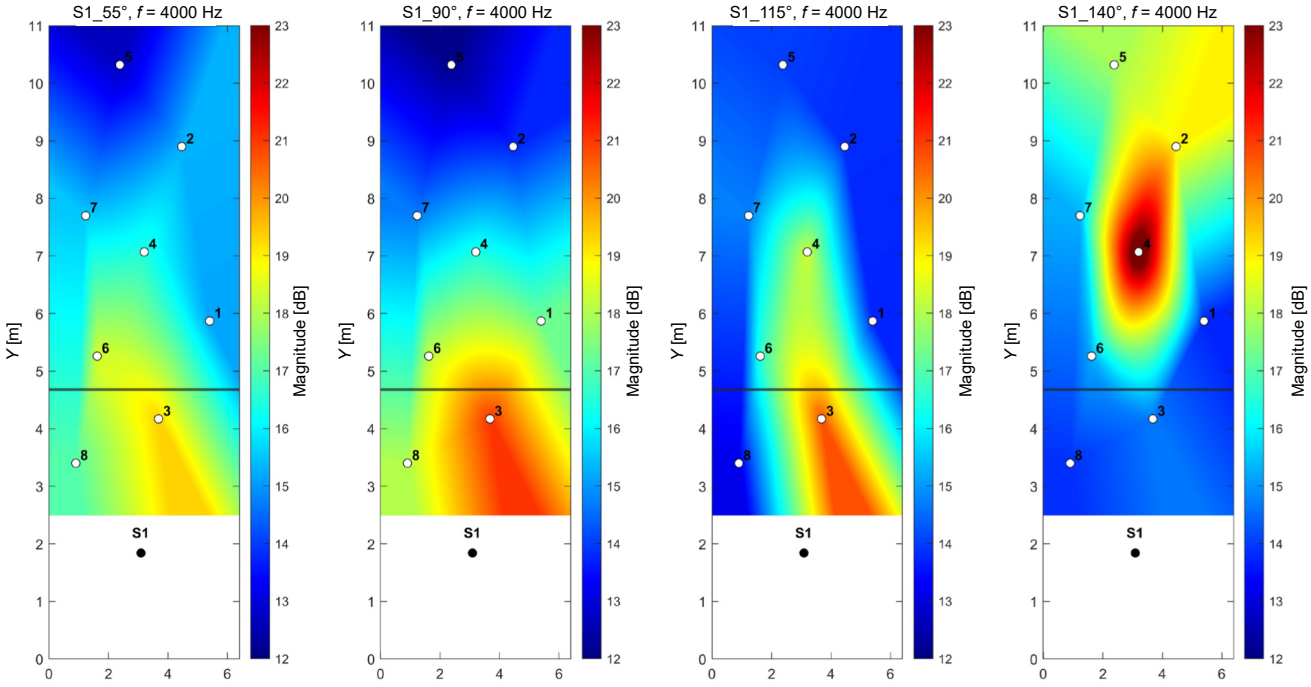

Fig. 12. SPL heatmaps for all PAL angles, 4 kHz octave band, source position – S1.

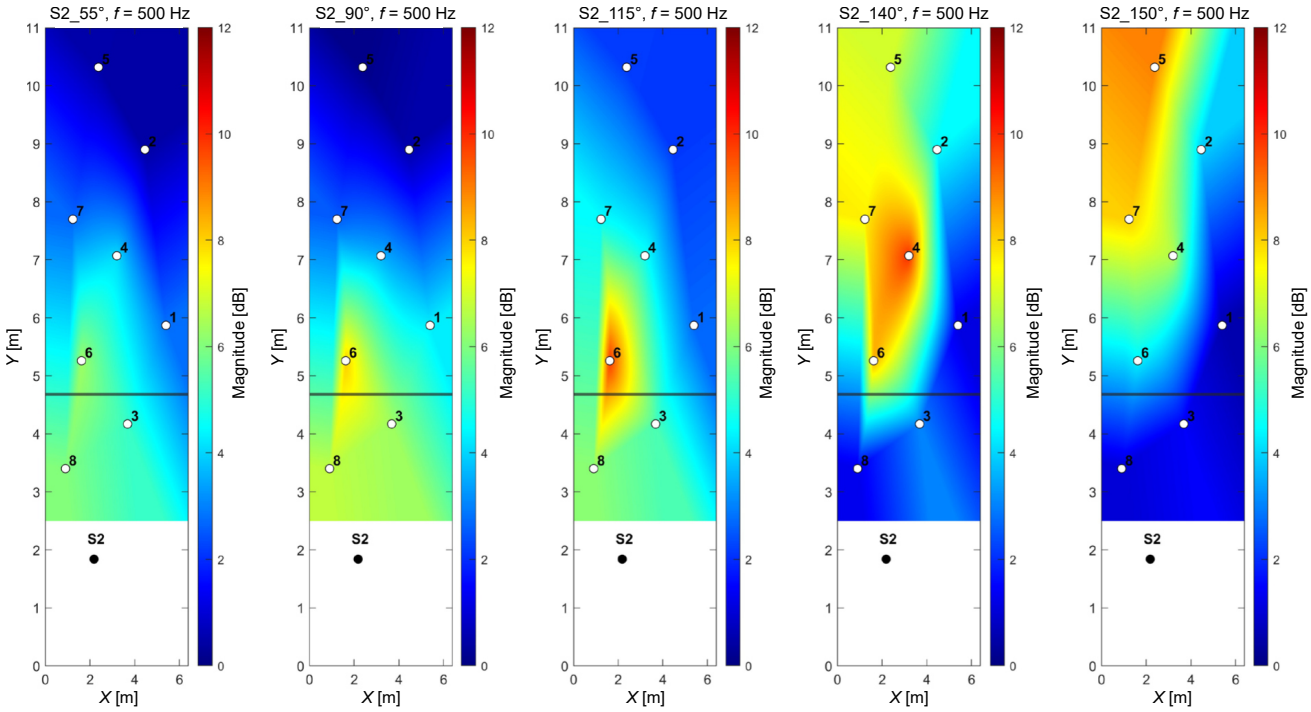

Fig. 13. SPL heatmaps for all PAL angles, 500 Hz octave band, source position – S2.

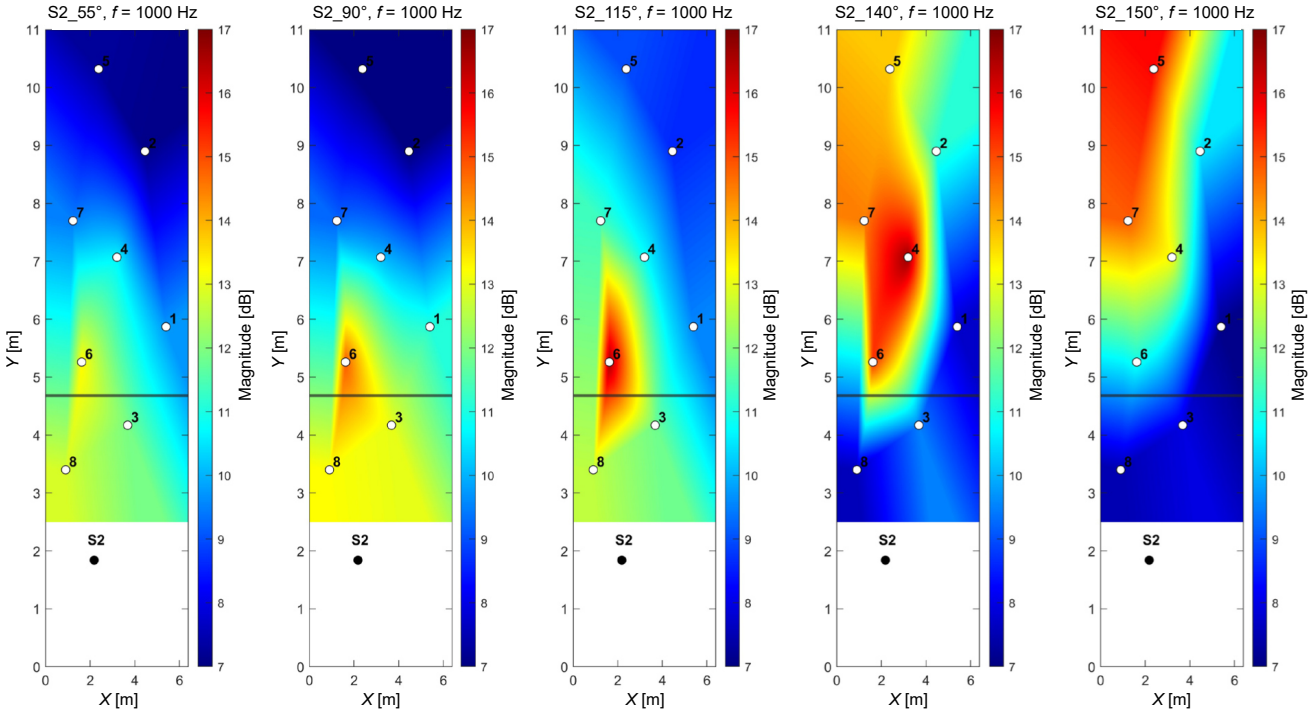

Fig. 14. SPL heatmaps for all PAL angles, 1 kHz octave band, source position – S2.

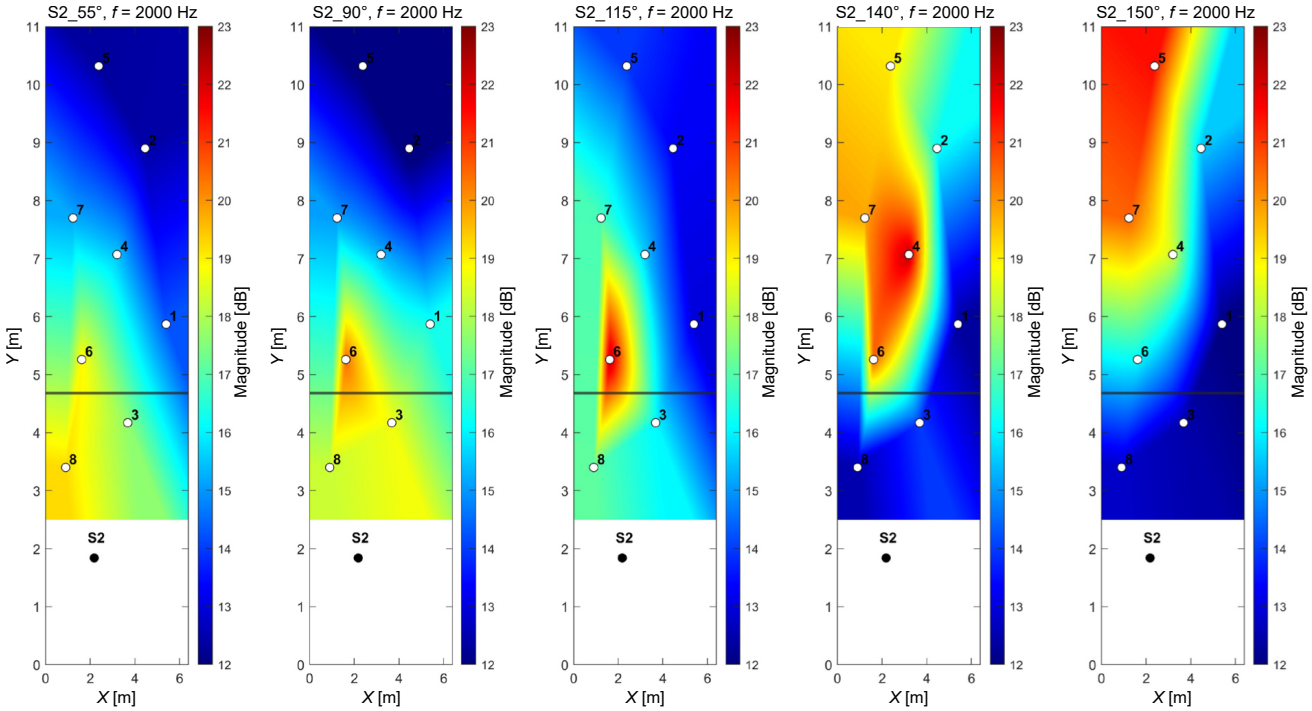

Fig. 15. SPL heatmaps for all PAL angles, 2 kHz octave band, source position – S2.

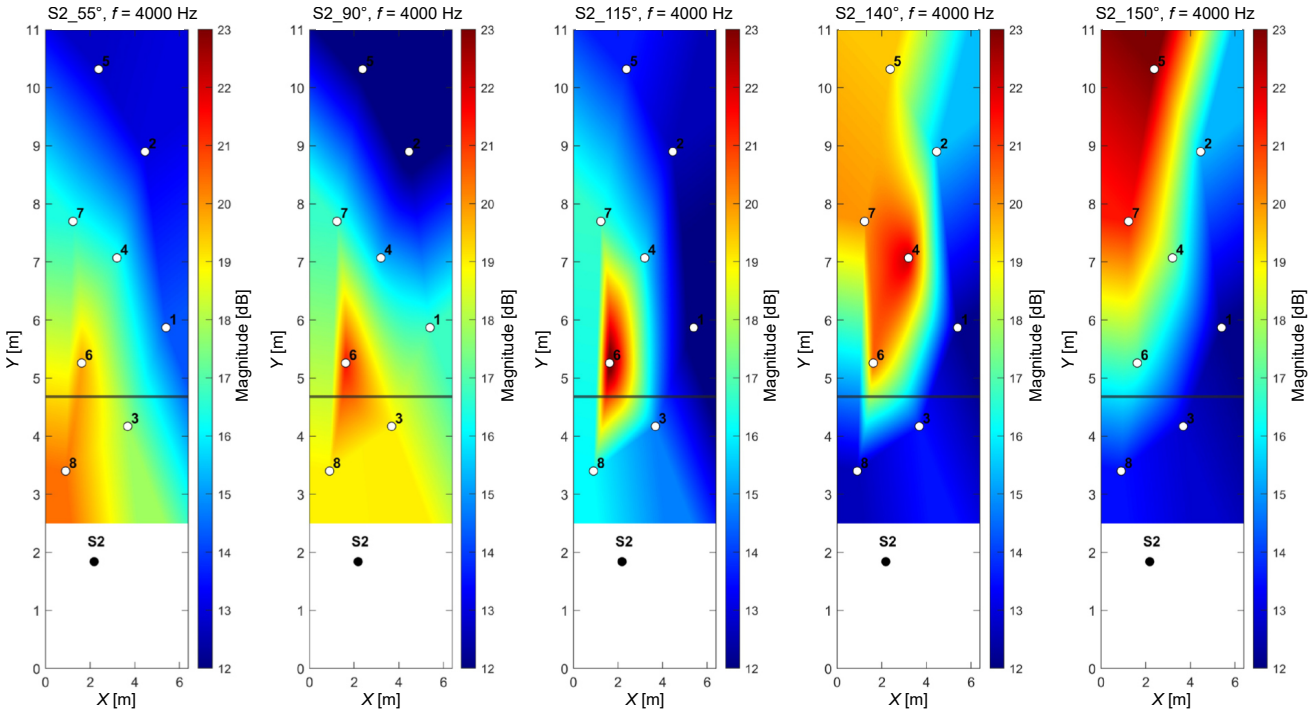

Fig. 16. SPL heatmaps for all PAL angles, 4 kHz octave band, source position – S2.

All of the presented maps show a strong concentration of sound at measurement points closest to the reflection path. Lower frequencies are more spread out than the higher frequencies, which proves that the presented method allows analysis of both geometric and wave phenomena, and that the obtained results match those acquired from laboratory experiments. Since the angle of 150° indicated that the PAL was pointed at the furthest edge of the angled ceiling, the values presented on the S2_150° heatmaps in Figs. 13–16 show the effective range of the said ceiling. High levels of sound pressure visible near the back wall of the room prove that its length is appropriate for the current room dimensions. As shown in the laboratory measurements, the area of a reflector needs to be large enough to reflect sound from lower frequency bands. Since the 500 Hz frequencies are visibly concentrated on the acquired maps, the size of the angled portion is also sufficient.

4.2. Simulation verification

To verify the results, the measured room was modelled and analysed in EASE Acoustic 4.4 using the raytracing module. The angled ceiling was included in the model, along with the measurement points. As shown in Fig. 17, reflections from the furthest edge of the angled portion of the ceiling are directed toward the last row of desks in the room. These results match those obtained from the in-situ measurements and confirm the effectiveness of parametric speakers in room acoustic analysis. Simulation outcomes for the S2 source were nearly identical to those acquired for S1; therefore, only the results for the S1 source are presented in this paper. It is important to emphasise that this method of analysis does not account for wave effects and was therefore used only to verify reflections at the incident angle. More advanced simulations will be carried out in future work.

4.3. Leave-one-out cross-validation of interpolation

To statistically verify the accuracy of the interpolation, the leave-one-out cross-validation (LOOCV) method was used, as it has the advantage of being applicable even with small samples (Geroldinger *et al.*, 2023). In this approach, a single observation is used for validation while the rest of the data forms the training dataset. This process is repeated so that each observation in the entire dataset is used only once for validation (Lumumba *et al.*, 2024). In this method, assume that we have the dataset D, where

$$D = \{(x_1, y_1)\,(x_2, y_2)\,, ..., (x_i, y_i)\}. \tag{2}$$

In this approach, x_i represents the features, in this case, the coordinates, and y_i represents the corresponding label of the outcome for each observation i (where $i = 1, 2, ..., n$), which in this case is the SPL for a given octave at one of the measurement points. Model training is conducted on $n-1$ observations and only one observation is used as the validation set, giving the classification error that can be expressed with the following mathematical equation (Lumumba *et al.*, 2024):

$$\text{LOOCV}_{\text{ERROR}} = \frac{1}{n}\sum_{i=1}^{n} L(y_i, \widehat{y}_i), \tag{3}$$

where L is the loss function and $\widehat{y}_i$ is the expected value for point i from a model trained without point i. The common loss function is the root mean squared error (RMSE), expressed as

$$L\,(y_i, \widehat{y}_i) = \sqrt{(y_i, \widehat{y}_i)}^{\,2}. \tag{4}$$

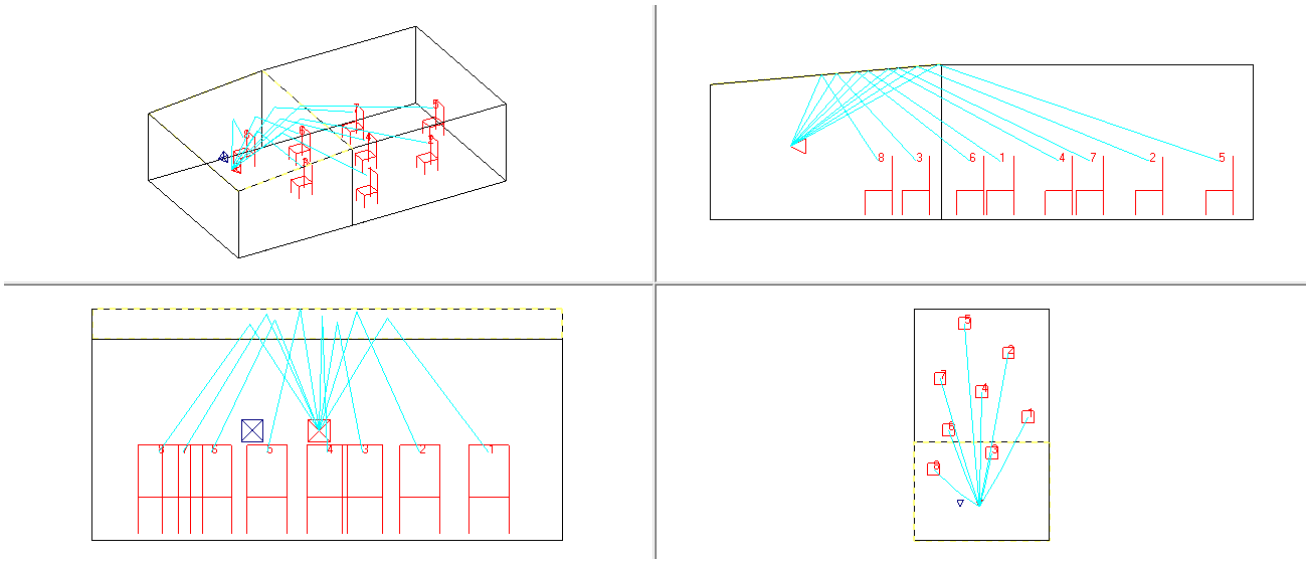

Fig. 17. Ray-tracing results of the measured room modelled in EASE Acoustic – S1 source.

This loss function is sensitive to large deviations, which makes it suitable for SPL measurements, as SPL varies significantly over short distances. The values of the loss function L calculated for all measurements are shown in Table 2.

Table 2. Loss function values calculated from LOOCV.

Source_angle	Loss function values in each bandwidth [dB]			
	500 Hz	1 kHz	2 kHz	4 kHz
S1_55°	1.2	1.2	1.5	2.0
S1_90°	1.1	1.1	1.2	1.5
S1_115°	2.9	3.1	3.4	3.6
S1_140°	2.7	2.8	3.0	3.4
S2_55°	1.0	1.0	1.0	1.1
S2_90°	1.4	1.5	1.8	2.1
S2_115°	2.1	2.2	2.8	3.8
S2_140°	3.5	3.7	3.9	3.9
S2_150°	2.1	2.2	2.3	2.5

As shown in Table 2, all loss function values are in the range of 1 dB–3.9 dB. This level of accuracy is not significant; however, since the differences between high and low sound pressure can vary drastically over short distances, even up to 10 dB, it is sufficient for this experiment. Since the measurements were preliminary, for further study, the density of the receiver grid should be increased to minimise the error.

5. Summary

This paper presented research results concerning the use of PALs in room acoustic analysis, especially for tracing reflection from highly reflective surfaces. In summary, it was found that certain models of PALs possess sufficiently narrow beamwidths to allow first-reflection analysis without interference from other elements in the room. The reflected sound exhibited a directivity pattern dependent on the panel size, with larger panels producing wider beamwidths and reflecting more energy at lower frequencies, which is consistent with wave phenomena typically observed in room acoustic measurements. Therefore, PALs can potentially be used interchangeably with omnidirectional speakers, without losing essential information about wave behaviour in a given sound field. The characteristics of the reflected sound were also found to be largely independent of the surface material, provided the material is sufficiently reflective. Consequently, panels made from different materials, such as wood or acrylic glass, should yield comparable results when measured with a PAL. Finally, it was demonstrated that PALs can be effectively utilised for assessing reflective elements in rooms and can reliably generate reflected sound distribution maps, allowing for precise evaluation of an element's acoustic performance.

Fundings

This research was funded by a research subvention supported by the Polish Ministry of Science and Higher Education (grant 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

Filip Węgrzyn conceptualized the study, wrote the original draft, conducted the measurements and performed the analysis. Adam Pilch conducted the measurements, contributed to data interpretation, edited the manuscript. All authors reviewed and approved the final manuscript.

References

1. Bennett M.B., Blackstock D.T. (1975), Parametric array in air, *Journal of the Acoustical Society of America*, **57**(3): 562–568, https://doi.org/10.1121/1.380484.
2. Berktay H.O. (1965), Possible exploitation of nonlinear acoustics in underwater transmitting applications, *Journal of Sound and Vibration*, **2**(4): 435–461, https://doi.org/10.1016/0022-460X(65)90122-7.
3. Gallien A., Prawda K., Schlecht S. (2024), Matching early reflections of simulated and measured RIRs by applying sound-source directivity filters, [in:] *Audio Engineering Society Conference: AES 2024 International Acoustics & Sound Reinforcement Conference*, https://aes2.org/publications/elibrary-page/?id=22373.
4. Gan W.-S., Yang J., Kamakura T. (2012a), Parametric acoustic array: Theory, advancement, and applications, *Applied Acoustics*, **73**(12): 1209–1210, https://doi.org/10.1016/j.apacoust.2012.06.016.
5. Gan W.-S., Yang J., Kamakura T. (2012b), A review of parametric acoustic array in air, *Applied Acoustics*, **73**(12): 1211–1219, https://doi.org/10.1016/j.apacoust.2012.04.001.
6. Geroldinger A., Lusa L., Nold M., Heinze G. (2023), Leave-one-out cross-validation, penalization, and differential bias of some prediction model performance measures – A simulation study, *Diagnostic and Prognostic Research*, **7**: 9, https://doi.org/10.1186/s41512-023-00146-0.
7. Ju H.S., Kim Y.-H. (2010), Near-field characteristics of the parametric loudspeaker using ultrasonic transducers, *Applied Acoustics*, **71**(9): 793–800, https://doi.org/10.1016/j.apacoust.2010.04.004.
8. Keele Jr. D.B. (2016), Design of free-standing constant beamwidth transducer (CBT) loudspeaker line arrays for sound reinforcement, *Journal of the Audio*

Engineering Society, https://aes2.org/publications/elibrary-page/?id=18428.

9. Login D. (2015), *A new approach to loudspeaker measurements*, Klippel GmbH, https://www.klippel.de/uploads/media/Logan_Klippel_Near_Field_Scanner_2015.pdf (access: 20.07.2025).
10. Lumumba V.W., Kiprotich D., Mpaine M.L., Makena N.G., Kavita M.D. (2024), Comparative analysis of cross-validation techniques: LOOCV, k-folds cross-validation, and repeated k-folds cross-validation in machine learning models, *American Journal of Theoretical and Applied Statistics*, **13**(5): 127–137, https://doi.org/10.11648/j.ajtas.20241305.13.
11. Rakerd B., Hunter E.J., Berardi M., Bottalico P. (2018), Assessing the acoustic characteristics of rooms: A tutorial with examples, *Perspectives of the ASHA Special Interest Groups*, **3**(19): 8–24, https://doi.org/10.1044/persp3.SIG19.8.
12. Tervo S., Pätynen J., Lokki T. (2009), Acoustic reflection path tracing using a highly directional loudspeaker, [in:] *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 245–248, https://doi.org/10.1109/ASPAA.2009.5346530.
13. Videotel Digital (2014), HyperSound® HSS 3000 mono/stereo system directional audio speakers, Turtle Beach Corporation, https://cdn.shopify.com/s/files/1/0260/5894/8671/files/HyperSound_Owners_Manual.pdf?v=1588116815 (access: 20.07.2025).
14. Vuine F. (2024), Loudspeaker, European Patent Office, Patent number EP4122215B1.
15. Westervelt P.J. (1963), Parametric acoustic array, *Journal of the Acoustical Society of America*, **35**(4): 535–537, https://doi.org/10.1121/1.1918525.
16. Yoneyama M., Fujimoto J.I., Kawamo Y., Sasabe Y. (1983), The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design, *Journal of the Acoustical Society of America*, **73**(5): 1532–1536, https://doi.org/10.1121/1.389414.

Archives of Acoustics Vol. 50, No. 4, pp. 525–541 (2025), https://doi.org/10.24425/aoa.2025.154830

Review Paper

Review of Microphone-Based Contactless Vital Signs Monitoring Systems

Abiodun Ernest AMORAN*, Dariusz BISMOR

Department of Measurements and Control Systems, Silesian University of Technology
Gliwice, Poland

*Corresponding Author e-mail: abiodun.amoran@polsl.pl

Received October 23, 2024; revised June 10, 2025; accepted September 16, 2025; published online November 6, 2025.

Microphones are sensors common to a variety of the Internet of Things (IoT) and healthcare applications. Many examples have proved that microphones can be useful in detecting, e.g., abnormal breathing rates. There are already applications that serve this purpose, e.g., respiratory acoustic monitoring, ResApp, etc. Breath signal was studied using a range of technologies and sensors, including the most common: radar, accelerometer, wearables, and so on. The majority of these sensors are attached to the body of a monitored person. However, the emergence of COVID-19 has drawn particular attention to the importance of using non-contact technologies for monitoring breath signals and other vital signs. This paper presents a comprehensive review of microphone-based non-contact vital sign monitoring, including the methodologies and concepts, while identifying new research gaps and opportunities for the future studies.

Keywords: beamforming; microphone; machine learning; vital signs.

Acronyms

CIR – channel impulse response,
DOA – direction of arrival
DYCTNN – dynamic convolution-transformer neural network,
FFT – fast Fourier transform,
FMCW – frequency-modulated continuous wave,
FPGA – field-programmable gate array,
IDRes – identity-based respiration monitoring system for digital twins enabled healthcare,
IoT – Internet of Things,
MEMS – microelectromechanical microphone,
NCVS – non-contact vital signs,
RIP – respiratory inductance plethysmography,
RMSE – root-mean-square error,
SNR – signal-to-noise ratio,
STFT – short time Fourier transform,
TDOA – time difference of arrival.

1. Introduction

The vulnerability of today's healthcare system was evident during COVID-19 pandemic, a serious global concern in which the number of patients outweighed available equipment. It was a common practice that the respiratory apparatus, known as ventilators, was shared between two patients (Garzotto *et al.*, 2020). According to (Branson, Rodriquez, 2023; Tsai *et al.*, 2022), the use of ventilators increased by 30 % in case of adults and 15 % in case of children following COVID-19. This indicated that persons with normal breathing problems have been neglected during this period. This group of people includes both the younger and older generations.

The application of digital processing to a microphone signal makes it suitable for various research applications. Microphones, commonly used for recording audio, have now evolved into sophisticated non-contact monitoring sensors. By using advanced signal processing techniques, microphones can be used to sense and analyze vital signs such as heart rate and respiration rate without making contact with the body of the target. This new approach offers numerous advantages over traditional methods, resulting in microphone-based non-contact monitoring systems being a promising technology for remote health monitoring and wellness applications.

A microphone is a transducer that converts sound waves to electrical signals. It detects slight changes in air pressure induced by sound and generates an

electrical signal representing the acoustic sound waveform. This electrical signal can subsequently be amplified, processed, recorded, or transmitted for numerous purposes such as telecommunications, audio recording, active noise control and speech recognition. Microphones vary in kind, design, and technology, but they always work on the principle of converting acoustic energy (sound pressure) into electrical energy. Common types of microphone include dynamic, condenser, ribbon, piezoelectric microphones and microelectromechanical microphone (MEMS); each comes with different properties that make it best suited for different applications.

Numerous studies have been conducted in the area of contact and non-contact vital signal monitoring. Among the research results, a few include thermal imaging cameras (Savazzi *et al.*, 2020), photoplethysmography (PPG) (Ryu *et al.*, 2021; Artemyev *et al.*, 2020; Boccignone *et al.*, 2023; Hashim *et al.*, 2023; Khong, Mariappan, 2019), doppler radar (Islam *et al.*, 2019; Joshi *et al.*, 2023; Edanami *et al.*, 2022; Zhang *et al.*, 2023a; Wahyu *et al.*, 2022; Mercuri *et al.*, 2018), microwave sensors (Katoh *et al.*, 2023; Celik *et al.*, 2011; Dei *et al.*, 2009), and acoustic sensors (Okamoto *et al.*, 2023; Xiao, Yu, 2021; Liu *et al.*, 2022; Jahanshahi *et al.*, 2018; Smithard *et al.*, 2017). Other popular choices are video cameras (Huang *et al.*, 2021; Sabokrou *et al.*, 2021; Artemyev *et al.*, 2020; Hsu *et al.*, 2020; Shokouhmand *et al.*, 2022) and fiber cable (Xu *et al.*, 2020; 2021; Liang *et al.*, 2023; Zhao *et al.*, 2023; Lyu *et al.*, 2022). However, using microphones for non-contact recording offers several advantages, including robustness, the ability to capture detailed information (Kranjec *et al.*, 2014; Fang *et al.*, 2016) and their sensitivity across a wide range of coverage, making them adaptable to different scenarios. Microphones are also useful for making respiratory sounds accessible via phones, laptops, and other portable devices (Massaroni *et al.*, 2021), although this approach has its own drawbacks. This review focuses on microphones for non-contact vital sign monitoring and it is divided into sections discussing various methods that have been developed in this field. These methods include beamforming techniques, smartphone-based solutions, hardware and artificial intelligence (AI) based approaches.

2. Microphone

The advent of the Internet of Things (IoT) has made the use of microphones more relevant, increasing their usefulness by 17 % per year (Beckmann, 2017). This may be a result of microphones changing from just a device for voice reception to their adaptation to mobile applications. Modern applications of microphones include mobile phones and tablets, cameras, wearables, bluetooth speakers, and security cameras. They can act as a sensor for detecting the respiration or heart rate of humans. Different microphones are being used in sound analysis due to their unique capabilities and features (Balgemann *et al.*, 2023). Moreover, some of them are equipped with a digital signal processor that enables them to modify the audio signal based on the distance and direction to the sound source. The pulse-density modulated microphone has been recently gaining attention due to its ability to delivering audio to digital processors, but its high-order decimation filter for pulse code modulation increases the cost and power consumption when used as a beamformer (Ipenza, Masiero, 2018). Table 1 shows a summary of microphone applications found in the literature.

2.1. Types of microphones used in audio signal analysis

2.1.1. Dynamic microphones

Dynamic microphones have the advantage of providing balanced sound recording. They are also durable,

Table 1. Applications of microphones from the literature.

Paper	Microphone type	Placement	Contact/Non-contact
(Doyle, 2019)	Electret	Attached to trachea, lungs	Contact
(Valipour, Abbasi-Kesbi, 2017)	Capacitor	Chest region	Contact
(Kavsaoğlu, Sehirli, 2023)	Stethoscope	Chest region	Contact
(Zhang *et al.*, 2023b)	MEMS	–	Non-contact
(Shih *et al.*, 2019)	Smartphone	Mouth/Chest	Contact
(LoMauro *et al.*, 2022)	–	Chest wall and lungs	–
(Dafna *et al.*, 2015)	Rode NTG-1 directional	–	Non-contact
(Islam *et al.*, 2021)	Wearable and smartphone	Chest region	Contact
(Chauhan *et al.*, 2017)	Smartphone and wearable		Contact
(Khodaie *et al.*, 2021)	MEMS	Mouth region	Contact
(Khatkhate *et al.*, 2022)	Pressure sensors	Ribcage	Contact
(Fang *et al.*, 2023)	Circular microphone array	–	Non-contact
(Xu *et al.*, 2022)	Smartphone	–	Non-contact
(Xie *et al.*, 2020)	–	Modelling of chest region	Contact

portable, and capable of producing high-quality sound. These microphones work on the principle of electromagnetic induction, the movement of a wire in a magnetic field creates an electromotive force (EMF) in the wire, which forces the current to flow. When sound waves hit the diaphragm, it moves either the magnet or the coil, creating a small current that can capture sounds from up to one meter away.

2.1.2. Condenser microphone

This kind of microphone functions as a capacitor consisting of two plates near each other, one of which acts as the diaphragm. When sound reaches the diaphragm, it vibrates, generating changes in capacitance, resulting in an electrical representation of the acoustic signal. Condenser microphones have a standard diaphragm diameter: large and small; the small having the advantage of being more compact and sensitive to picking up higher frequency sound (PreSonus, 2022). Its high fidelity, excellent frequency response, low noise levels, and sensitivity make it appropriate for acoustic research (Todorović *et al.*, 2015).

2.1.3. Electret microphone

An electret microphone is a type of condenser microphone that eliminates the need for a high-voltage power supply by using a permanently charged material called an electret. Like most microphones, it consists of a diaphragm placed near a metal backplate, forming a capacitor. When sound waves impinge on the diaphragm, it vibrates and changes the capacitance, generating an electrical signal corresponding to the sound. They are commonly used in devices such as mobile phones, hearing aids, and voice recorders because they are compact, less expensive, and they require a small power source for their in-built preamplifier (Open Music Lab, 2022).

2.1.4. Microelectromechanical microphone

The MEMS microphone operates by using a tiny mechanical system etched onto a silicon chip to convert sound waves into electrical signals. It is made of a flexible diaphragm and a fixed backplate which forms a variable capacitor. The capacitance of the capacitor changes as sound waves hit the diaphragm, and this change/variation is then converted into an electrical signal by an integrated circuit. MEMS microphones are gradually replacing electret microphones due to their smaller size and greater suitability for smartphones. They have the advantage of picking up signals equally from all directions, making these microphone omnidirectional. They are also tiny in size and consume low amount of power. This implies they can be used to determine the direction of sound in a microphone array (Wang *et al.*, 2020). However, when MEMS microphone recordings are converted to electrical signals, some noise is introduced (Rose, 2022). The audio data used in smartphones is generated digitally as a result of current movement in a very small mechanical sound diaphragm. MEMS microphones are employed in mobile devices because of its tiny footprint and good performance (Picchio *et al.*, 2019).

Table 2. Comparison of non-contact health monitoring technologies.

Technology	Strengths	Weaknesses	Applications	References
Microphone-based	Low cost, high accuracy in detecting physiological sounds, easy integration with existing devices, non-invasive, versatile, low power consumption	Sensitive to noise, privacy concerns, limited range 0.5 m–1 m for breath sounds	Respiratory monitoring, heartbeat detection, speech recognition	Fukuda *et al.*, 2018; Aarts, 2019; Genova, 1997; Sharma *et al.*, 2019
Radar-based	Accurate for motion detection, capable of detecting chest movements for breathing rate monitoring, non-contact, works in the dark	Expensive hardware, limited in detecting internal physiological sounds, consumes more power than other methods, long range between 10 m and 50 m	Breathing rate monitoring, heart rate monitoring, motion detection	Sakamoto, Yamashita, 2019; Zakrzewski, 2015; Lv *et al.*, 2021
Infrared sensors	Effective for detecting body temperature changes, non-contact, can detect presence or absence based on heat signatures, variable power consumption	Requires line-of-sight, affected by ambient temperature variations, limited to surface-level observations, calibration needed, range between 0.1 m to 4 m	Body temperature monitoring, motion detection, sleep studies	Thundat *et al.*, 2000; Fraden, 2014; Yang *et al.*, 2022
Ultrasonic	Good for distance measurement and obstacle detection, non-contact, safe to use, non-invasive, low power consumption	Limited resolution for detecting fine physiological details, requires direct path for sound waves, affected by material properties, range between 0.3 m to over 10 m	Fall detection, obstruction detection, motion monitoring	Hoctor *et al.*, 2008; Barany, 1993; Toa, Whitehead, 2019

2.2. Comparison of non-contact health monitoring technologies

A microphone-based non-contact monitoring system is a better alternative to other non-contact monitoring methods such as radar, infrared and ultrasonic sensor. It is capable of monitoring breath signal, heartbeat detection and identifying vocal patterns. These capabilities are possible due to its ability to leverage on the properties of sound waves to monitor physiological features. Table 2 outlines the advantages and disadvantages of these methods.

2.3. Microphone array

A microphone array (MA) is an arrangement of several microphones positioned to gather signal from different spatial locations. The main goal of MA is a robust representation of the signal. It works on the principle of sound propagation that several inputs are able to either attenuate or enhance by processing signals from specific directions even in the presence of noise (Dey, Ashour, 2018; Levy *et al.*, 2010; Doclo *et al.*, 2015). MAs are essential in non-contact measurement of signals, leveraging on the combined power and sensitivity of the connected microphones. The spatial arrangement of MA consists of several configurations which include linear arrays, circular arrays or spherical arrays, depending on the purpose an array is intended (Alexandridis, Mouchtaris, 2017). The configurations also determine the spacing between the connected microphones (Dey, Ashour, 2018). Exemplary array arrangements are shown in Fig. 1. In this configuration, the microphone may be replaced with a smartphone or a beamforming method. The difference between a single microphone and an array arrangement is that a single microphone cannot provide the direction of a sound source and reduction of reverberation without the need for post-processing. An array arrangement, on the other hand, can improve the speech signal quality using the received radiation pattern from the direction of a desired signal, thus improving the signal-to-noise ratio (SNR) (Dey, Ashour, 2018).

Two important terms associated with array arrangement is beamforming and the direction of arrival (DOA). Beamforming is the procedure of estimating DOA and can be defined as a process of changing the phase and amplitude of signals received by an array of sensors (in this case microphones). The goal of beamforming is to enhance the signals from one direction while suppressing the other directions, to make the received signal specific to a direction. There are two major types of beamforming: data-dependent and data-independent. Data dependent methods usually change parameters based on the received signal example are adaptive or optimal, phase-shift frequency beamforming. Data-independent (or fixed) beamforming have fixed parameters; examples include delay-and-sum, filter-and-sum, subband, and minimum variance distortionless response beamforming (Mathworks, n.d.). The DOA, on the other hand, is a process of determining the direction (for example, in degrees) in which a received signal was transmitted. The degree of accuracy of the estimated DOA is affected by the performance of beamforming, thereby making beamforming and DOA interdependent on each other.

3. Principles of microphone-based monitoring

Microphone-based health monitoring systems utilize the body's natural sounds (signals), such as breathing, heartbeats, and coughing, to obtain vital physiological data. By detecting these acoustic signals, these systems can constantly and non-invasively monitor an individual's health, as a suitable alternative to contact-based devices. Microphones are sensitive to the vibrations caused by physiological events like airflow during respiration, heart valve closures, or even vocal cord vibrations. The vital signals monitored by the microphones include heart rate, respiratory rate, snoring and coughing.

3.1. Signal processing techniques

The accuracy and effectiveness of a microphone in monitoring vital signs depend largely on the parameters of the microphone itself and possibly on the preamplifier working with it. These factors should be well supported by the signal processing method adopted. Recorded information contains noise and other unnecessary data, necessitating the use of filtering techniques to extract important signals.

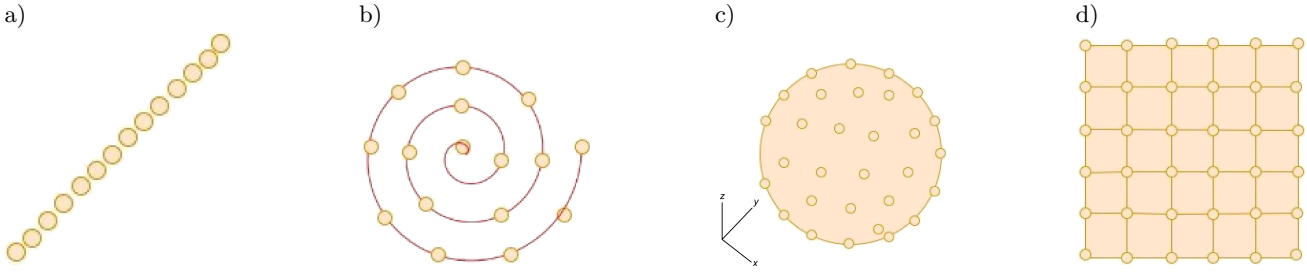

Fig. 1. Different array arrangements: a) linear array; b) spiral array; c) circular array; d) planar array.

3.1.1. Noise filtering and amplification

A significant challenges in microphone-based monitoring is the capture of background noise (Llorca-Bofi *et al.*, 2024; Paul *et al.*, 2023). This noise can degrade the quality of the acquired signal, making it difficult to the isolate desired signal. Noise can be addressed using numerous filtering techniques. For example, band-pass filter can clearly separate breath signals from noise, as the frequency of the breath signal is between 100 Hz and 1000 Hz while heart sound ranges between 20 Hz to 200 Hz (Han *et al.*, 2023). Other filtering methods such as low pass and high pass filters can also be applied to recover a desired signal. Amplification is essential for enhancing low-amplitude signals, such as shallow breath signal.

3.1.2. Adaptive filtering

Adaptive filtering is a signal processing technique commonly used in noise cancellation, system identification, channel equalization and control systems. The major difference between an adaptive filter and other types of filters is its ability to dynamically adjust its coefficients in response to changes in the signal environment (Arenas-Garcia *et al.*, 2021). This dynamic adjustment makes it suitable for processing a non-stationary signal, such as the breath signal. One common example of a adaptive filter is the adaptive line enhancement (ALE). ALE uses adaptive filters with dual roles: predicting the narrowband component of a noisy signal and enhancing them while eliminating broadband noise. ALE assumes that the narrowband signal is either sinusoidal or periodic, allowing it to exploit the time correlation in the narrowband signal to distinguish between the original or desired signal from the uncorrelation broadband noise. To improve the quality of the desired signal, ALE uses the previous input to separate the narrowband components from the broadband noise. The basic components of ALE include the input signal, the delay lines, the adaptive filter, and the computed error signal. The use of ALE in microphone-based, non-contact health monitoring is important, as one of the challenges associated with microphones is their tendency to pick up background noise along with the desired signal. ALE can be applied to solve this problem (Atkins *et al.*, 2021).

3.1.3. Time-domain and frequency-domain analysis

Signal features can be extracted using either time-domain or frequency-domain analysis. The time domain describes changes in a signal amplitude with respect to time and is useful for detecting breath cycles or heartbeats. On the other hand, frequency domain analysis, examines the signal energy's distribution across a range of frequencies, which helps identifying specific physiological signals characterized by specific frequencies (Rangayyan, 2015). A common example used in frequency domain analysis is the fast Fourier transform (FFT), which coverts a time domain to the frequency domain for more detailed analysis of its frequency components (Henry, 2023). Furthermore, the time-frequency distribution (TFD) combines both time and frequency domain information, providing a more comprehensive analysis when both time and frequency domain information are needed simultaneously.

3.1.4. Machine learning and AI integration

Recent advancements in microphone-based non-contact health monitoring system focus on integrating machine learning (ML) and AI. These algorithms enable the model to identify, classify and interpret physiological signals. For example, deep learning models such as convolutional neural network (CNN) and long short term memory (LSTM) networks are used to distinguish normal and abnormal breath or heartbeat patterns (Li, Qian, 2024; Roseline *et al.*, 2024). Additionally, these algorithms can handle large datasets and learn from previously collected physiological signals, improving accuracy. Numerous ML- and AI-based methods have been used for the identification and classification of different types of coughs, wheezes or heart signals (Ferrante *et al.*, 2020; Orlandic *et al.*, 2021; Pramono *et al.*, 2019; Renjini *et al.*, 2021). In a case when microphone records patient's respiratory signal, the raw data serves as an input to the AI-powered system, which filters out (remove) noises, identifies key features, and classifies the data based on the trained model. This facilitates real time diagnosis of diseases associated with breath and heartbeat signals.

3.1.5. Pattern recognition and feature extraction

Pattern recognition plays a crucial role in identifying acoustic signal. The algorithm detect repetitive patterns in the signal, such as peaks in the amplitude or the periodicity of heartbeats and breathing cycles. The wavelet transform is a commonly used feature extraction method, and it decomposes complex signals into simpler components, allowing unique characteristics that may indicate the presence or absence of diseases to be clearly identified (Taghavirashidizadeh *et al.*, 2022).

4. Beamforming based methods

Beamforming can be defined as the process of combining multiple signals from microphones in an array to amplify sound in a specific direction. Beamforming can be combined with other approaches, such as radar systems and cameras, to locate targets (Xiong *et al.*, 2023; Wang *et al.*, 2023). To detect a signal in a specific direction, the beamformer controls the phase and amplitude at the transmission end. In non-contact vital

sign monitoring, beamformers ensure accuracy measurement of vital sounds and allow monitoring of multiple subjects at the same time. Frequent body movementa and noise are among the main factors hindering beamforming. This section reviews literature that has adopted beamforming.

A dynamic convolution-transformer neural network (DYCTNN) for sound source localization using functional beamforming was proposed by Zhang *et al.* (2024). Dynamic convolution and self-attention techniques were used to capture the spatial distribution of sound sources. The model was trained and tested using a dataset generated via acoustic simulation on a 2 m × 2 m plane with a 60-channel spiral microphone array and one to five monopole sources producing sound fields at various frequencies. Xiong *et al.* (2023) utilized beamforming by combining a phased array of an antenna and a double-phase shifter (DPS) to adjust the magnitude and phase of the transmitted signal. Beamforming allowed for simultaneous monitoring of numerous people with minimal target interference. Actuators were utilised to simulate human chest movement, while an omni-antenna was employed to generate and receive signals. This method worked well; however, adding antennas make this system too complex.

Sun *et al.* (2022) used a phase-shifting technique for transmitting beam formation and digital beamforming for optimal spatial filtering at the receiving end. The method utilized a frequency-modulated continuous wave (FMCW) radar with 9 transmitting and 16 receiving channels. Digital beamforming was designed to obtain optimal spatial filtering at the receiving end, enhancing the capability of multi-person detection. The arctangent demodulation method was used for phase estimation, and phase unwrapping was thereafter applied to address phase ambiguity. The proposed method was able to detect targets within the range of 1.8 m to 12 m. Hall *et al.* (2015) developed the phased array non-contact vital sign (NCVS) sensor system with an autonomous beam steering algorithm, implemented in LabVIEW. The selected phased array arrangements were tested, and data samples were gathered to assess the performance of the autonomous beam steering algorithm. The results showed that heart rate measurement accuracy was approximately 95 % within 5 bpm, and the automatic beam steering algorithm achieved an accuracy of 94.36 % within 5 bpm with a 2.82 bpm standard deviation.

Wang *et al.* (2023) introduced the dualforming-based method that combined both spatial and frequency domain beamforming to improve the signal-to-noise ratio (SNR) across multiple subject locations. The multiple subtle signal classification (MUS2IC) approach was used to separate subjects with subtle movements from static objects. Empirical mode decomposition (EMD) was used to extract heartbeat patterns by decomposing the cardiac frequency response (CFR) streams into separate intrinsic mode functions (IMFs). The method measured heart rate within a 10 m range, allowing the monitoring of heartbeats of six subjects at the same time. Tashev and Acero (2006) presented a post-processing a microphone array's beamformer output. The algorithm estimated the spatial probability of sound source presence and applied a spatio-temporal filter. Experimental results showed that the directivity index improved up to 8 dB and jammer suppression up to 17 dB at the angle of 40° from the sound source.

5. Microphone sensor based method

A microphone can detect vital signals in both contact and non-contact modes. The latter produces less noise since the sensors are not in direct contact with the subject's body. This section explores various studies conducted in this area. A simple diagram of the microphone-based method is shown in Fig. 2.

Chen *et al.* (2015) presented a microphone position calibration approach to distribution microphone arrays, combining an acoustic energy decay model with the time difference of arrival (TDOA) method. The method first estimates the coarse distance between the microphone and the sound source, followed by TDOA to find the accurate distance within a specific range near the coarse distance. The microphone's position is determined using the least mean square error estimate approach, which yields high positioning accuracy, steady calibration performance, and low processing complexity. Qian *et al.* (2018) employed FMCW sonar to send a chirp signal and calculated the spectrogram of the baseband signal to extract vital signals such as breath rate, heart rate, and individual

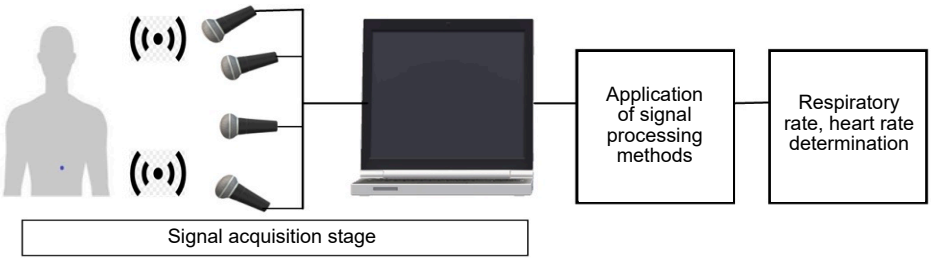

Fig. 2. Typical diagrammatic representation of vital signs monitoring system setup using microphone.

heartbeat from the acoustic signal phase. The method downsampled the FMCW signal to baseband and continually monitored the signal phase in the spatial bin containing vital motions. The use of dual microphone enhanced the performance of the system.

Valipour and Abbasi-Kesbi (2017) monitored heartbeat and respiration rate using a phonocardiogram based miniature wireless acoustic sensor with two capacitor microphones, a microprocessor, and a transceiver operating at 2.54 GHz in the industrial, scientific, and medical (ISM) band. The sensor was placed on the volunteer's chest, and ECG signals were acquired. The findings showed root-mean-square errors (RMSEs) of less than 2.27 bpm for heartbeat and 0.92 bpm for respiration rate, with standard deviation of less than 1.26 for heartbeat and 0.63 for respiration rate. Overall, the developed approach is contact-based. Tran *et al.* (2014) used a hybrid hardware-software technique to detect an infant's vital signs, using an infrared non-contact temperature sensor and a microphone-based breathing sensor. The system was designed in a hardware description language (HDL) and implemented on an field-programmable gate array (FPGA) board. The developed device identified the infant's vital signs when tested on the Altera DE2-115 FPGA board.

Taniguchi *et al.* (2023) presented a vital sign monitoring system for dogs based on the MEMS microphone and the Raspberry Pi wireless system. To extract the heart rate, they first removed the DC offset from the obtained data, then transformed it using the short time Fourier transform (STFT), and finally applied the fifth-order Butterworth bandpass filter. The filtered data was then normalised, and the heart rate was calculated by counting amplitude peaks within a specific time frame. The heart rate extraction technique includes calculating the number of data points and amplitude thresholds, as well as computing the distance between peaks. The heart rates acquired during the surgery were monitorable every second, with an average heart rate of 110 bpm.

Dafna *et al.* (2015) proposed a non-contact microphone-based polysomnography (PSG) to measure breathing noises and estimate breath rate during sleeping. Adaptive noise reduction techniques was used to suppress background noise and non-periodic spectrum components were filtered out by a periodicity augmentation module. The BR module was the final stage, and it estimated BR based on the filtered signal. The system was tested on 204 individuals who participated in an in-laboratory in the study. The Pearson correlation coefficient between the two techniques was $R = 0.97$, showing a strong relationship. An epoch-by-epoch BR comparison revealed a mean relative error of 2.44 % and Pearson correlation of 0.68, demonstrating good agreement between the audio-based BR estimation and the gold-standard respiratory belts.

Wang *et al.* (2021) presented a low-cost, contactless heartbeat monitoring device based on a commodity speaker and a microphone array. Acoustic impulses are transmitted by a speaker and received by a microphone array to estimate the human heartbeat. Passive beamforming and frequency domain filtering were used to improve the quality of the signal accuracy. A wideband time-delay approach was also used to predict the DOA of the target-reflected signal. The prototype monitors heart rate at a distance of 1.7 m, with an estimation error of 0.5 bpm. A wearable microphone sensor based on the adaptive windowing technique was employed by Zhang *et al.* (2024) to estimate heart rate. The method used a spectrogram to derive an initial estimate and calculate the optimal window length based on frequency resolution and physiological constraints. A one-step autoregressive model was used to correct estimates, thereby improving the heart rate measurement accuracy by ±2.8 bpm. The developed method was tested on a group of 26 healthy subjects. Ashraf and Moussavi (2024) designed a piezoelectric surface microphone placed at the suprasternal notch to capture tracheal breathing sounds. This device produced clear respiratory waveforms with minimal sensitivity to ambient noise. A wearable accelerometer microphone (Gupta *et al.*, 2021) captured lung sounds and chest-wall motion to derive respiratory patterns in hospitalized patients with COPD, pneumonia, etc. These contact sensors can measure both breath sounds and rate with high fidelity, even amid patient motion or background noise; for example, the piezo sensor showed negligible degradation across frequency bands when noise was present.

6. Smartphone and contact based methods

Smartphone technology started in 1992 (Tocci, 2024), and it has surpassed expectations, particularly in the development of applications that can run on smartphones. A significant contributor to this success is the microphone, which has helped acquisition of data for various applications, including those focused on vital sign monitoring. Smartphones are now capable of monitoring vital signs such as heart rate, respiratory rate, blood pressure, and blood oxygen saturation, whether through contact-based or non-contact methods. The section focuses on the literature that has used microphones installed in smartphone for vital sign monitoring.

Kavsaoğlu and Sehirli (2023) captured audio signals from the heart and trachea, resulting in a dataset for detecting inhalation and exhalation circumstances. Two methods were used to obtain these signals: one involving heart sound and the other involving trachea sounds. The audio signals were classified into inhalation and exhalation phases using ML models. The highest accuracy and performance

were achieved using a majority voting strategy with k-nearest neighbour, random forest, and support vector machines. Doyle (2019) used a flat adhesive acoustic sensor and the TASCAM DR-40 Digital Recorder to record bioacoustic data. Recordings were taken from multiple of locations, including the neck, external ear canal, oxygen mask, as well as a leak-free microphone attached to a laryngeal mask airway. Audacity, an open-source digital audio editor and recording programme, was used to analyse breath sounds and apply digital filters.

Lee *et al.* (2023) used an array of MEMS microphones to record lung sound waves, which were converted into acoustic images. The system's performance was assessed using waterbags to stimulate airway blockages, and its accuracy was compared to that of digital stethoscopes. The proposed method demonstrated better detection of lung conditions, with a room square error of 0.28 and SNR of 7 dB. Lo-Mauro *et al.* (2022) introduced a semi-automatic, robust pre-processing for respiratory data analysis using functional data analysis (FDA) techniques. The approach involved separating, detecting outliers based on time-duration, amplitude, and shape, and clustering breaths using K-medoids for different breathing patterns. The proposed methodology showed an error rate of less than 5 % for minimum detection and outlier removal.

Chauhan *et al.* (2017) developed a framework that combines smartphone acoustic sensors to identify breathing phases and estimate biomarkers. Breathing data was collected from pulmonary patients and healthy individuals using Samsung Galaxy Note 8 smartphones, chest bands, and spirometers. The system achieved 77.33 % accuracy and over 90 % accuracy in estimating respiratory rate and other biomarkers. Shih *et al.* (2019) developed a real-time breathing detection algorithm with low latency, running on a smartphone. To train and evaluate the developed system over 2.76 million breathing sounds from 43 participants was captured, and the system achieved 75.5 % accuracy in detecting breathing phases using a combination of attention-based LSTM models and CNN-based extraction modules. Wang *et al.* (2018) used a correlation-based frequency-modulated continuous wave (C-FMCW) approach for monitoring human breathing via audio signals. The common speaker and microphone components found in most homes were used. The system accurately identified subjects' respiration in a variety of environments, including different rooms and subject sleep positions. Khodaie *et al.* (2021) developed a system that records respiratory sounds from the upper airways using microphones implanted in a breathing mask. The study discovered a strong correlation (coefficient of 0.9) between acoustic features of respiratory sounds and respiratory metrics such as the peak flow and average flow.

Fang *et al.* (2023) proposed the identity-based respiration monitoring system for digital twins enabled healthcare (IDRes). The respiration rate was estimated by tracing the changes in the phase of the sonar signal and detecting the doppler frequency shift to capture chest motion characteristics. Experimental results showed 93.3 % recognition accuracy and the mean detection error of 0.49 bpm.

Xu *et al.* (2020) proposed the BreathListener, a system that monitors breathing in driving scenarios using audio devices on smartphones. The method captured fine-grained breathing waveforms in driving scenarios. The device used the energy spectrum density (ESD) of acoustic waves to record breathing processes in driving conditions. BreathListener used background removal and variational mode decomposition (VMD) to remove interference from driving settings while extracting the breathing pattern from the ESD signals. The retrieved breathing pattern was then translated into the Hilbert spectrum, and the fine-grained breathing waveform was generated using a deep learning architecture, based on generative adversarial networks. Chara *et al.* (2023) developed an FMCW-based acoustic system on a smartphone by emitting and receiving high-frequency chirps, the phone tracks tiny chest displacements. In trials this approach achieved extremely high accuracy – a median breathing-rate error below 0.15 breaths per minutes across various conditions. A smartphone-based contact method. Phokela *et al.* (2020) used a headset microphone under the nose to record nasal airflow sounds: it achieved respiration-rate errors less than 10 % even in noisy environments, demonstrating feasibility for home use.

Nemcova *et al.* (2020) estimated the heart rate, blood oxygen saturation (SpO2), and blood pressure (BP) using smartphone sensors. HR and SpO2 were determined by generating a photoplethysmogram (PPG) from the camera data, while BP was measured by calculating the pulse transit time value from the PPG and recording a phonocardiogram (PCG) via the microphone. The results showed mean absolute errors (MAE) of 1.1 % for SpO2 and 1.4 bpm for heart rate. Vincent *et al.* (2023) presented a multi-target blind source separation technique based on a single sonar. The use of the freqency hopping (FH) technique within the ULCW (Ultra-CW) scheme helped to minimize the effects of frequency-selective fading (FSF) and intersymbol interference (ISI) in the baseband, thereby improving the accuracy of acoustic signal transmission. The combination of continuous wave (CW) and FMCW signals in the ULCW scheme enhanced the transmission of energy from the smartphone, enabling accurate acoustic signal propagation over long distances. Doheny *et al.* (2023) developed a method to predict respiratory rate and exhale length from smartphone captured audio data. The method required cal-

culation of the audio signal's basic frequency and detection of individual exhales with adaptive thresholding. Exhale boundary timings were optimised with adaptive physiological thresholds. The respiratory rate was determined by identifying peaks and troughs in the respiratory inductance plethysmography (RIP) signal, and exhale durations were calculated as the time between each peak and the next dip in the RIP signal. The RIP respiratory rate was utilised as the standard against which the audio respiratory rate was measured. The fundamental frequency of the respiration envelope was found as the frequency corresponding to the first peak in the harmonic product spectrum above 0.09 Hz. Other active acoustic methods use smart speakers to monitor breathing or heart rate, though these are mostly prototypes or proof-of-concept. The advantage of these non-contact methods is comfort and convenience and suitability for home or telehealth. However, they require a device (smartphone or speaker) close to the subject and can be sensitive to environmental noise or interference. Thick clothing, bedding, or a distance beyond 2 m–3 m can degrade signal, so practical use often means limiting the scenario.

7. Hardware based methods

This section examines approaches that incorporated hardware components, whether handheld or not, with the capability to determine the respiration rate of subjects.

Al-Ali and Lee (2012) patented a physiological acoustic monitoring system that collects physiological data from an acoustic sensor and generates respiration-related parameters in both real-time and non-real time. The system processes data by downsampling to provide raw audio of breathing sounds, and compresses it for futher analysis. Wang *et al.* (2023) presented the MultiResp, a multi-user respiration monitoring system that detects chest movement using acoustic signals. The system captures acoustic signals reflected from participants' chests, allowing for robust respiration monitoring even when subjects are facing away from the transceiver or blocked by barriers. MultiResp extracted fine-grained breath rate and phase differences between participants to differentiate breath waves with similar rates and adjust to dynamic variations in the number of monitored subjects. However, MultiResp fails when the sound pressure is less than 55 dB or when there is body movement which causes significant alterations in the multipath signals, causing erratic fluctuation of the channel impulse response (CIR).

Abbasi-Kesbi *et al.* (2018) presented a wireless acoustic sensor that used a phonocardiogram to detect heartbeat and respiratory rate. The system comprises a processor, transceiver, and two capacitor microphones for capturing heartbeat and respiration rate. The technology also measures breathing rate with a capacitor microphone placed near the mouth. The wireless acoustic sensor demonstrated high accuracy in predicting heartbeat and breathing rate, with RMSEs of less then 2.27 beats/min and 0.92 breaths/min, and standard deviations of less then 1.26 and 0.63, respectively. The system's sensitivity and specificity in recognizing PCG sounds ranged for S1 to S4 at 98.1 % and 98.3 %, respectively, representing a 3 % improvement over earlier work. This method accurately recorded heart and respiration rate in a variety of circumstances, including resting and breath-holding, with consistent results across numerous volunteers.

Wan *et al.* (2023) introduced a continuous multi-user respiratory tracking system designed for household settings using acoustic based commercial off-the-shelf (COTS) sensors. The system employed multi-stage algorithm to isolate and recombine respiration data from different paths to calculate the respiration rate of several moving persons. By utilizing features from multiple dimensions to distinguish between users in the same region, and applying Zadoff–Chu (ZC) sequences with optimal auto-correlation, it differentiates user pathways. The system transmits the ZC sequence modulated by a sinusoidal carrier as the transmitted sound signal, with its detection range and bandwidth determined by the length of the ZC sequence and frame length. The experimental results showed that RespTracker's two-stage algorithm can differentiate the respiratory pattern of at least four subjects over a three-meter distance.

8. Artificial intelligence based method

This section reviews studies that have adopted ML techniques using a microphone as the primary signal acquisition method. Figure 3 illustrates ML-based method.

Xie *et al.* (2023) utilised an autoencoder (AE) neural network to quantify the residual between the original and reconstructed signals, which can increase the end-to-end (e2e) respiration monitoring accuracy by a factor of 2.75 when compared to the baseline. Their approach employed deep learning techniques, combining an autoencoder neural network and a self-supervised learning to quantify signal quality. The use of radio frequency quality (RF-Q) further enhanced respiration monitoring accuracy. However, large volumes of training data are required for deep learning algorithms and the need for manual labelling, as training datasets for DL techniques is typically not publicly available.

Liu *et al.* (2021) proposed a reverberation aware network (RAN) algorithm for improving the robustness of DOA estimation. The algorithm used the beam cross-correlation (BCC) as an input to a deep neural network (DNN), explicitly characterizing reverber-

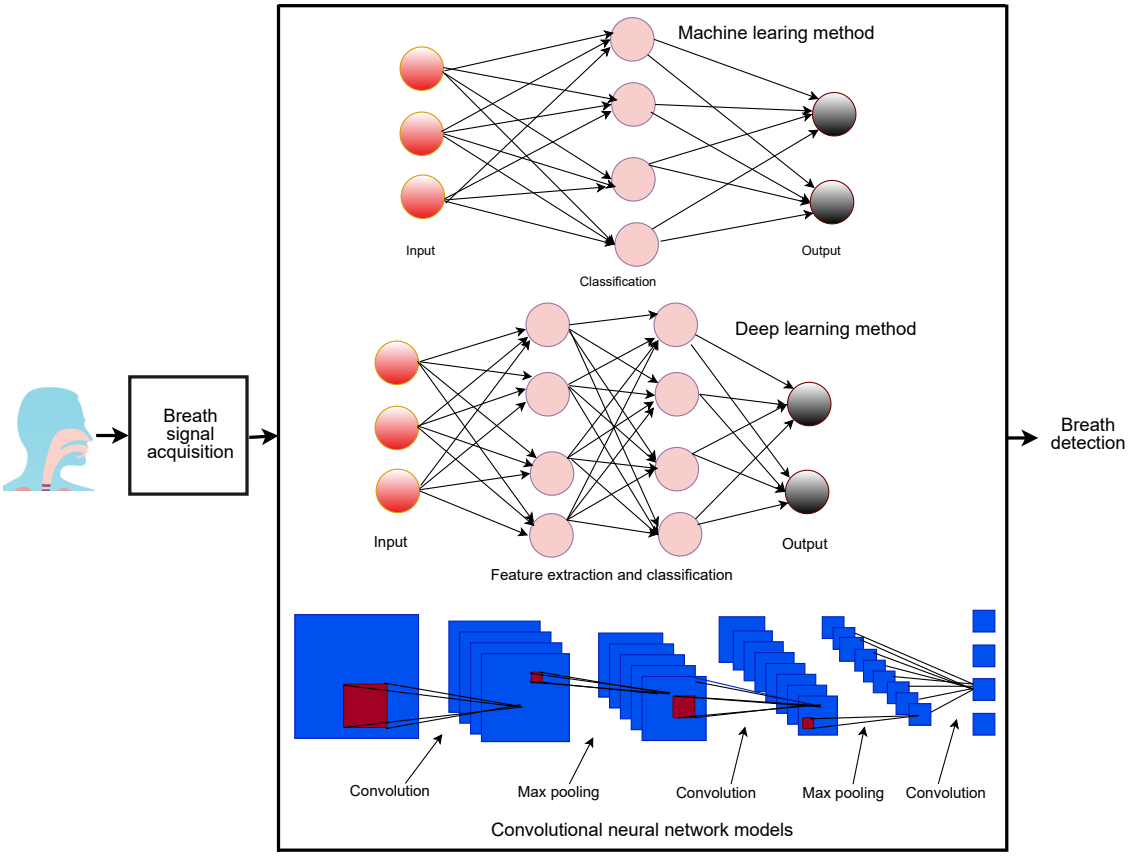

Fig. 3. Machine learning based method.

ation in the captured speech signal. The classic beamforming algorithm was used to generate beamforming outputs, the observed signals, which was then used as a reference for reverberation identification. The filter-sum (FS) beamforming algorithm was adopted for beamforming processing. Numerical simulations were based on virtual room environments generated with a reverberation model, as well as practical experiments under physical room environments, to evaluate the performance of the proposed method. The impact of different environments on the performance was evaluated by conducting experiments with different noise levels and source distances. In addition to the aforementioned research, some studies have also combined two methods, such as beamforming with ML (Zhang *et al.*, 2024) and smartphone with ML (Kavsaoğlu, Sehirli, 2023; Shih *et al.*, 2019; Xu *et al.*, 2020). Despite the promising application of ML in different fields, this area is underexplored especially when using a microphone as the non-contact health monitoring method.

9. Challenges and solution

Using microphones for signal acquisition in medical applications presents several challenges, with noise and interference being the most significant. To address these issues, some techniques have been proposed, including the use of adaptive noise reduction algorithms (Abed *et al.*, 2022; Thomsen, Du, 2020; Meyer *et al.*, 2020; Wu *et al.*, 2020; Wang, Qiu, 2020), directional microphones (Fischer, Puder, 2012; Kanamori, Terada, 2016; Nongpiur, 2018; Park *et al.*, 2020), and the application of ML (Jain, Hera, 2019; Shiozawa *et al.*, 2020; Takenaka, Ozawa, 2022). While these three methods have been independently used in the literature, this review suggests an integrated method that combines these approaches. In this proposed solution, adaptive noise reduction reduces inherent noise from recordings, directional microphones capture signals from a single direction or a patient, and ML processes the signals to minimize noise interference more effectively. Another challenge is the urgency with which some respiratory data are required to make informed decisions. High latency or processing delays can be problematic, this issue can be addressed by using edge computing, which processes data locally, or by employing optimized algorithms for real-time data processing. These suggested methods can significantly improve the responsiveness and reliability of microphone-based non-contact monitoring systems in medical applications. Apart from the above, other challenges with microphone include privacy, data security and technical implementation. Although data

privacy and security were not mentioned by some of the articles reviewed. However, they remain one of the challenges associated with the use of microphones and other audio-based signal acquisition methods. Audio recordings should be treated with utmost security, as they reveal sensitive health information. Alqudaihi *et al.* (2021) suggested that only numeric features should be extracted from heart and respiratory signals and that data transmission should be anonymized summaries. To further enhance data privacy and security. Alqudaihi *et al.* (2021) also recommended implementing blockchain-based audit logs or federated learning techniques. Future directions could focus on the development of specialized contact microphones and the adoption of AI-based denoising and data encryption algorithms to improve the reliability and security of microphone-based monitoring systems. Moreover, privacy-preserving hardware innovations, such as MicPro proposed by Xiao *et al.* (2023) alongside the end-to-end encryption protocol used in some social media messaging applications – can address these issues adequately. Overall, research into the suggested solutions could enhance the performance of microphones as vital sign monitoring systems.

10. Conclusion

This review has presented the potential of microphone-based systems for non-contact sign monitoring. The transition from simple acoustic sensors to the adoption of intelligent health monitoring was made possible by technologies such as beamforming, ML, and smartphone integration. These systems have evolved from simple signal-capturing devices to sophisticated devices capable of detecting complex physiological patterns. Non-contact health monitoring systems can leverage these innovations, such as the integration of deep learning algorithms like CNN, RNN, or LSTM (Acharya, Basu, 2020; Thakur *et al.*, 2022). Although some research has been done in this area, the accuracy and real-time application of ML-based methods can be further improved through enhanced data collection processes, hybrid deep learning models, better feature extraction methods, and the use of microphone arrays instead of single microphones. Future directions could also focus on leveraging smartphone-based applications and cloud-based platforms to improve access, accuracy, and reliability while addressing other challenges associated with microphone-based systems.

Fundings

This work was supported by the state budget for science in Poland in 2025 under grant no. 02/050/BKM_25/0048.

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

Abiodun Ernest Amoran conceptualized the study and wrote the original draft; Dariusz Bismor supervised the research, provided guidance, and reviewed the manuscript. All authors reviewed and approved the final manuscript.

References

1. Aarts R.M. (2019), Acoustical patient monitoring using a sound classifier and a microphone, US Patent 10,269,228.
2. Abbasi-Kesbi R., Valipour A., Imani K. (2018), Cardiorespiratory system monitoring using a developed acoustic sensor, *Healthcare Technology Letters*, **5**(1): 7–12, https://doi.org/10.1049/htl.2017.0012.
3. Abed A., Cordon E., Warren N., Sanders D., Patel M., Liya D. (2022), Digital microphone with reduced processing noise, US Patent 11,438,682.
4. Acharya J., Basu A. (2020), Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning, *IEEE Transactions on Biomedical Circuits and Systems*, **14**(3): 535–544, https://doi.org/10.1109/TBCAS.2020.2981172.
5. Al-Ali A., Lee S.U.K. (2012), Physiological acoustic monitoring system, US Patent App. 13/675,014, https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2013056141.
6. Alexandridis A., Mouchtaris A. (2017), Multiple sound source location estimation in wireless acoustic sensor networks using DOA estimates: The data-association problem, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **26**(2): 342–356, https://doi.org/10.1109/TASLP.2017.2772831.
7. Alqudaihi K.S. *et al.* (2021), Cough sound detection and diagnosis using artificial intelligence techniques: Challenges and opportunities, *IEEE Access*, **9**: 102327–102344, https://doi.org/10.1109/ACCESS.2021.3097559.
8. Arenas-Garcia J., Azpicueta-Ruiz L.A., Silva M.T.M., Nascimento V.H., Sayed A.H. (2021), Combinations of adaptive filters: Performance and convergence properties, *IEEE Signal Processing Magazine*, **33**: 120–140, https://doi.org/10.1109/MSP.2015.2481746.
9. Artemyev M., Churikova M., Grinenko M., Perepelkina O. (2020), Robust algorithm for remote photoplethysmography in realistic conditions, *Digital Signal Processing*, **104**: 102737, https://doi.org/10.1016/j.dsp.2020.102737.

10. Ashraf W., Moussavi Z. (2024), Design and analysis of a contact piezo microphone for recording tracheal breathing sounds, *Sensors*, **24**(17): 5511, https://doi.org/10.3390/s24175511.
11. Atkins A., Cohen I., Benesty J. (2021), Adaptive line enhancer for nonstationary harmonic noise reduction, *Computer Speech and Language*, **70**: 101245, https://doi.org/10.1016/j.csl.2021.101245.
12. Barany L.P. (1993), Ultrasonic non-contact motion monitoring system, US Patent 5,220,922.
13. Balgemann T.W., Pedersen S.C., Pessin J.M., Shumard B.R., Sutter L.L., Perkofski R.J. (2023), Microphone with adjustable signal processing, US Patent App. 18/197,416.
14. Beckmann P. (2017), How digital signal processing can enhance the utility and performance of microphones, https://21088554.fs1.hubspotusercontent-na1.net/hubfs/21088554/digital_microphone_processing_paper%20(1).pdf (access: 29.05.2024).
15. Boccignone G., D'Amelio A., Ghezzi O., Grossi G., Lanzarotti R. (2023), An evaluation of non-contact photoplethysmography-based methods for remote respiratory rate estimation, *Sensors*, **23**(7): 3387, https://doi.org/10.3390/s23073387.
16. Branson R.D., Rodriquez D. Jr. (2023), COVID-19 lessons learned: Response to the anticipated ventilator shortage, *Respiratory Care*, **68**(1): 129–150, https://doi.org/10.4187/respcare.10676.
17. Celik N., Gagarin R., Youn H., Iskander M.F. (2011), A noninvasive microwave sensor and signal processing technique for continuous monitoring of vital signs, *IEEE Antennas and Wireless Propagation Letters*, **10**: 286–289, https://doi.org/10.1109/LAWP.2011.2132690.
18. Chara A., Zhao T., Wang X., Mao S. (2023), Respiratory biofeedback using acoustic sensing with smartphones, *Smart Health*, **28**: 100387, https://doi.org/10.1016/j.smhl.2023.100387.
19. Chauhan J., Hu Y., Seneviratne S., Misra A., Seneviratne A., Lee Y. (2017), BreathPrint: Breathing acoustics-based user authentication, [in:] *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 278–291, https://doi.org/10.1145/3081333.3081355.
20. Chen Z., Li Z., Wang S., Yin F. (2015), A microphone position calibration method based on combination of acoustic energy decay model and TDOA for distributed microphone array, *Applied Acoustics*, **95**: 13–19, https://doi.org/10.1016/j.apacoust.2015.02.013.
21. Dafna E., Rosenwein T., Tarasiuk A., Zigel Y. (2015), Breathing rate estimation during sleep using audio signal analysis, [in:] *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5981–5984, https://doi.org/10.1109/EMBC.2015.7319754.
22. Dei D. *et al.* (2009), Non-contact detection of breathing using a microwave sensor, *Sensors*, **9**(4): 2574–2585, https://doi.org/10.3390/s90402574.
23. Dey N., Ashour A.S. (2018), *Direction of Arrival Estimation and Localization of Multi-Speech Sources*, Springer, https://doi.org/10.1007/978-3-319-73059-2.
24. Doclo S., Kellermann W., Makino S., Nordholm S.E. (2015), Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones, *IEEE Signal Processing Magazine*, **32**(2): 18–30, https://doi.org/10.1109/MSP.2014.2366780.
25. Doheny E.P. *et al.* (2023), Estimation of respiratory rate and exhale duration using audio signals recorded by smartphone microphones, *Biomedical Signal Processing and Control*, **80**(Part 1): 104318, https://doi.org/10.1016/j.bspc.2022.104318.
26. Doyle D.J. (2019), Acoustical respiratory monitoring in the time domain, *The Open Anesthesia Journal*, **13**, https://doi.org/10.2174/2589645801913010144.
27. Edanami K. *et al.* (2022), Remote sensing of vital signs by medical radar time-series signal using cardiac peak extraction and adaptive peak detection algorithm: Performance validation on healthy adults and application to neonatal monitoring at an NICU, *Computer Methods and Programs in Biomedicine*, **226**: 107163, https://doi.org/10.1016/j.cmpb.2022.107163.
28. Fang B., Lane N.D., Zhang M., Boran A., Kawsar F. (2016), BodyScan: Enabling radio-based sensing on wearable devices for contactless activity and vital sign monitoring, [in:] *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pp. 97–110, http://doi.org/10.1145/2906388.2906411.
29. Fang K. *et al.* (2023), IDRes: Identity-based respiration monitoring system for digital twins enabled healthcare, *IEEE Journal on Selected Areas in Communications*, **41**(10): 3333–3348, https://doi.org/10.1109/JSAC.2023.3310095.
30. Ferrante G., Licari A., Marseglia G., La Grutta S. (2020), Artificial intelligence as an emerging diagnostic approach in paediatric pulmonology, *Respirology*, **25**(10): 1029–1030, https://doi.org/10.1111/resp.13842.
31. Fischer E., Puder H. (2012), Method for reducing interferences of a directional microphone, US Patent 8,135,142.
32. Fraden J. (2014), Non-contact medical thermometer with stray radiation shielding, US Patent 8,834,019.
33. Fukuda T., Watanabe S., Matsumoto H., Tsuji H. (2018), Microphone array, monitoring system, and sound pickup setting method, US Patent 9,860,635.
34. Garzotto F. *et al.* (2020), COVID-19: Ensuring our medical equipment can meet the challenge, *Expert Review of Medical Devices*, **17**(6): 483–489, https://doi.org/10.1080/17434440.2020.1772757.
35. Genova J.J. (1997), Non-invasive medical monitor system, US Patent 5,590,650.

36. Gupta P., Wen H., Di Francesco L., Ayazi F. (2021), Detection of pathological mechano-acoustic signatures using precision accelerometer contact microphones in patients with pulmonary disorders, *Scientific Reports*, **11**(1): 13427, https://doi.org/10.1038/s41598-021-92666-2.

37. Haase W.C., Haase Z.S., Young R.F., Monin P.E., McCarthy C.W., Sarles F.W. (2011), Non-contact biometric monitor, US Patent App. 13/034,394.

38. Hall T. *et al.* (2015), A phased array non-contact vital signs sensor with automatic beam steering, [in:] *2015 IEEE MTT-S International Microwave Symposium*, pp. 1–4, https://doi.org/10.1109/MWSYM.2015.7166973.

39. Han L. *et al.* (2023), Health monitoring via heart, breath, and Korotkoff sounds by wearable piezoelectret patches, *Advanced Science*, **10**(28): 2301180, https://doi.org/10.1002/advs.202301180.

40. Hashim H.A., Ahmed N.M., Shabeeb A.G. (2023), Infant heart rate estimation based on non-contact UV photoplethysmography, *Indonesian Journal of Electrical Engineering and Computer Science*, **31**(1): 180–188, http://doi.org/10.11591/ijeecs.v31.i1.pp180-188.

41. Henry M. (2023), An ultra-precise fast Fourier transform, *Measurement*, **220**: 113372, https://doi.org/10.1016/j.measurement.2023.113372.

42. Hoctor R.T., Thomenius K.E., Dentinger A.M. (2008), Method and apparatus for ultrasonic continuous, non-invasive blood pressure monitoring, US Patent 7,425,199.

43. Hsu G.-S.J., Xie R.-C., Ambikapathi A.M., Chou K.-J. (2020), A deep learning framework for heart rate estimation from facial videos, *Neurocomputing*, **417**: 155–166, https://doi.org/10.1016/j.neucom.2020.07.012.

44. Huang B., Lin C.-L., Chen W., Juang C.-F., Wu X. (2021), A novel one-stage framework for visual pulse rate estimation using deep neural networks, *Biomedical Signal Processing and Control*, **66**: 102387, https://doi.org/10.1016/j.bspc.2020.102387.

45. Ipenza S.C., Masiero B.S. (2018), Microphone array processing of pulse-density modulated bitstreams, [in:] *28th Meeting of the Brazilian Acoustics Society October 3 to 5, 2018, Porto Alegre – RS*, **6**: 7, https://doi.org/10.17648/sobrac-87156.

46. Islam B. *et al.* (2021), BreathTrack: Detecting regular breathing phases from unannotated acoustic data captured by a smartphone, [in:] *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **5**(3): 1–22, https://doi.org/10.1145/3478123.

47. Islam S.M.M., Yavari E., Rahman A., Lubecke V.M., Boric-Lubecke O. (2019), Multiple subject respiratory pattern recognition and estimation of direction of arrival using phase-comparison monopulse radar, [in:] *2019 IEEE Radio and Wireless Symposium (RWS)*, pp. 1–4, https://doi.org/10.1109/RWS.2019.8714272.

48. Jahanshahi P., Wei Q., Jie Z., Zalnezhad E. (2018), Designing a non-invasive surface acoustic resonator for ultra-high sensitive ethanol detection for an on-the-spot health monitoring system, *Biotechnology and Bioprocess Engineering*, **23**: 394–404, https://doi.org/10.1007/s12257-017-0432-5.

49. Jain A.D., Hera C.M. (2019), Noise reduction using machine learning, Patent No. 10,580,430, https://patents.google.com/patent/WO2019079713A1/fr.

50. Joshi M., Moadi A.-K., Theilmann P., Fathy A.E. (2023), Compact millimeter wave radar for vital sign detection: A comprehensive study, [in:] *2023 16th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS)*, pp. 114–117, https://doi.org/10.1109/TELSIKS57806.2023.10316015.

51. Kanamori T., Terada Y. (2016), Directional microphone device, acoustic signal processing method, and program, US Patent 9,264,797.

52. Katoh M., Kanazawa T., Abe Y., Sun G., Matsui T. (2023), Development of a non-contact 15-second paediatric respiratory rate monitor using microwave radar and its clinical application, *Acta Paediatrica*, **112**(3): 493–495, https://doi.org/10.1111/apa.16585.

53. Kavsaoğlu A.R., Sehirli E. (2023), A novel study to classify breath inhalation and breath exhalation using audio signals from heart and trachea, *Biomedical Signal Processing and Control*, **80**(Part 1): 104220, https://doi.org/10.1016/j.bspc.2022.104220.

54. Khatkhate A.M., Raut V., Jadhav M., Alva S., Vichare K., Nadkarni A. (2022), Identification of basic respiratory patterns for disease-related symptoms through a microphone device, *Journal of Engineering Research and Sciences*, **1**(6): 36–44, https://doi.org/10.55708/js0106005.

55. Khodaie M., Nafisi V.R., Moghadam F.F. (2021), Design and implementation of an apparatus for respiratory parameters estimation based on acoustic methods, [in:] *2021 28th National and 6th International Iranian Conference on Biomedical Engineering (ICBME)*, pp. 15–21, https://doi.org/10.1109/ICBME54433.2021.9750286.

56. Khong W.L., Mariappan M. (2019), The evolution of heart beat rate measurement techniques from contact based photoplethysmography to non-contact based photoplethysmography imaging, [in:] *2019 IEEE International Circuits and Systems Symposium (ICSyS)*, pp. 1–4, https://doi.org/10.1109/ICSyS47076.2019.8982534.

57. Kranjec J., Beguš S., Geršak G., Drnovšek J. (2014), Non-contact heart rate and heart rate variability measurements: A review, *Biomedical Signal Processing and Control*, **13**: 102–112, https://doi.org/10.1016/j.bspc.2014.03.004.

58. Lee C.S., Li M., Lou Y., Abbasi Q.H., Imran M. (2023), An acoustic system of sound acquisition and

image generation for frequent and reliable lung function assessment, *IEEE Sensors Journal*, **24**(3): 3731–3747, https://doi.org/10.1109/JSEN.2023.3344136.

59. Levy A., Gannot S., Habets E.A.P. (2010), Multiple-hypothesis extended particle filter for acoustic source localization in reverberant environments, *IEEE Transactions on Audio, Speech, and Language Processing*, **19**(6): 1540–1555, https://doi.org/10.1109/TASL.2010.2093517.
60. Li C., Qian Q. (2024), A deep learning-based animation video image data anomaly detection and recognition algorithm, *Journal of Organizational and End User Computing (JOEUC)*, **36**(1): 1–25, https://doi.org/10.4018/JOEUC.345929.
61. Liang H. *et al.* (2023), Wearable and multifunctional self-mixing microfiber sensor for human health monitoring, *IEEE Sensors Journal*, **23**(3): 2122–2127, https://doi.org/10.1109/JSEN.2022.3225196.
62. Liu J., Li D., Wang L., Zhang F., Xiong J. (2022), Enabling contact-free acoustic sensing under device motion, [in:] *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **6**(3): 128, https://doi.org/10.1145/3550329.
63. Liu Y., Tong F., Zhong S., Hong Q., Li L. (2021), Reverberation aware deep learning for environment tolerant microphone array DOA estimation, *Applied Acoustics*, **184**: 108337, https://doi.org/10.1016/j.apacoust.2021.108337.
64. Llorca-Bofi J., Heck J., Dreier C., Vorlaender M. (2024), Urban background sound recordings for virtual acoustics under various weather conditions, *The Journal of the Acoustical Society of America*, **155**(3_Supplement): A178, https://doi.org/10.1121/10.0027230.
65. LoMauro A., Colli A., Colombo L., Aliverti A. (2022), Breathing patterns recognition: A functional data analysis approach, *Computer Methods and Programs in Biomedicine*, **217**: 106670, https://doi.org/10.1016/j.cmpb.2022.106670.
66. Lv W., He W., Lin X., Miao J. (2021), Non-contact monitoring of human vital signs using FMCW millimeter wave radar in the 120 GHz band, *Sensors*, **21**(8): 2732, https://doi.org/10.3390/s21082732.
67. Lyu W., Chen S., Tan F., Yu C. (2022), Vital signs monitoring based on interferometric fiber optic sensors, *Photonics*, **9**(2): 50, https://doi.org/10.3390/photonics9020050.
68. Massaroni C., Nicolò A., Sacchetti M., Schena E. (2021), Contactless methods for measuring respiratory rate: A review, *IEEE Sensors Journal*, **21**(11): 12821–12839, https://doi.org/10.1109/JSEN.2020.3023486.
69. Mathworks (n.d.), *Conventional and adaptive beamformer*, https://www.mathworks.com/help/phased/ug/conventional-and-adaptive-beamformers.html (access: 4.06.2024).
70. Mercuri M. *et al.* (2018), A direct phase-tracking Doppler radar using wavelet independent component analysis for non-contact respiratory and heart rate monitoring, *IEEE Transactions on Biomedical Circuits and Systems*, **12**(3): 632–643, https://doi.org/10.1109/TBCAS.2018.2813013.
71. Meyer P., Elshamy S., Fingscheidt T. (2020), Multichannel speaker interference reduction using frequency domain adaptive filtering, *EURASIP Journal on Audio, Speech, and Music Processing*, **2020**: 14, https://doi.org/10.1186/S13636-020-00180-6.
72. Nemcova A. *et al.* (2020), Monitoring of heart rate, blood oxygen saturation, and blood pressure using a smartphone, *Biomedical Signal Processing and Control*, **59**: 101928, https://doi.org/10.1016/j.bspc.2020.101928.
73. Nongpiur R.C. (2018), Directional microphone device and signal processing techniques, Patent.
74. Okamoto Y., Nguyen T.-V., Takahashi H., Takei Y., Okada H., Ichiki M. (2023), Highly sensitive low-frequency-detectable acoustic sensor using a piezoresistive cantilever for health monitoring applications, *Scientific Reports*, **13**(1): 6503, https://doi.org/10.1038/s41598-023-33568-3.
75. Open Music Lab (2022), *Electret microphones*, http://www.openmusiclabs.com/learning/sensors/electret-microphones/index.html (access: 1.06.2024).
76. Orlandic L., Teijeiro T., Alonso D.A. (2021), The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms, *Scientific Data*, **8**: 156, https://doi.org/10.1038/s41597-021-00937-4.
77. Park D.Y., Lim H.Y., Min C.K. (2020), Directional microphone device, Patent.
78. Paul V.-S., Hahn N., Hollebon J. (2023), Extraction of ambience sound from microphone array recordings for spatialisation, [in:] *2023 Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, https://doi.org/10.1109/I3DA57090.2023.10289397.
79. Phokela K.K., Naik V. (2020), Use of smartphone's headset microphone to estimate the rate of respiration, [in:] *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pp. 64–69, https://doi.org/10.1109/COMSNETS48256.2020.9027297.
80. Picchio R. *et al.* (2019), Comparing accuracy of three methods based on the GIS environment for determining winching areas, *Electronics*, **8**(1): 53, https://doi.org/10.3390/electronics8010053.
81. Pramono R.X.A., Imtiaz S., Rodriguez-Villegas E. (2019), Evaluation of features for classification of wheezes and normal respiratory sounds, *PLoS ONE*, **14**(3): e0213659, https://doi.org/10.1371/journal.pone.0213659.
82. PreSonus (2022), *What is a condenser microphone*, https://www.presonus.com/blogs/technical/what-is-a-condenser-microphone? (access: 1.06.2024).
83. Qian K., Wu C., Xiao F., Zheng Y., Zhang Y., Yang Z., Liu Y. (2018), Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart

devices, [in:] *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pp. 1574–1582, https://doi.org/10.1109/INFOCOM.2018.8485978.

84. Rangayyan R.M. [Ed.] (2015), Frequency domain characterization of signals and systems, [in:] *Biomedical Signal Analysis*, Wiley, pp. 349–395, https://doi.org/10.1002/9781119068129.ch6.
85. Renjini A., Swapna M.S., Raj V., Sankararaman S. (2021), Graph-based feature extraction and classification of wet and dry cough signals: A machine learning approach, *Journal of Complex Networks*, **9**(6): cnab039, https://doi.org/10.1093/comnet/cnab039.
86. Rose B. (2022), *An introduction to MEMS microphone arrays*, https://www.sameskydevices.com/blog/an-introduction-to-mems-microphone-arrays (access: 31.05.2024).
87. Roseline S.A., Saraf K., Sruti I.N.V.D. (2024), Intelligent human anomaly detection using LSTM autoencoders, [in:] *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, pp. 1–7, https://doi.org/10.1109/ACCAI61061.2024.10602454.
88. Ryu J., Hong S., Liang S., Pak S., Chen Q., Yan S. (2021), A measurement of illumination variation-resistant noncontact heart rate based on the combination of singular spectrum analysis and subband method, *Computer Methods and Programs in Biomedicine*, **200**: 105824, https://doi.org/10.1016/j.cmpb.2020.105824.
89. Sabokrou M., Pourreza M., Li X., Fathy M., Zhao G. (2021), Deep-HR: Fast heart rate estimation from face video under realistic conditions, *Expert Systems with Applications*, **186**: 115596, https://doi.org/10.1016/j.eswa.2021.115596.
90. Sakamoto T., Yamashita K. (2019), Noncontact measurement of autonomic nervous system activities based on heart rate variability using ultra-wideband array radar, *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, **4**(3): 208–215, https://doi.org/10.1109/JERM.2019.2948827.
91. Savazzi S., Rampa V., Costa L., Kianoush S., Tolochenko D. (2020), Processing of body-induced thermal signatures for physical distancing and temperature screening, *IEEE Sensors Journal*, **21**(13): 14168–14179, https://doi.org/10.1109/JSEN.2020.30 47143.
92. Sharma P., Imtiaz S.A., Rodriguez-Villegas E. (2019), Acoustic sensing as a novel wearable approach for cardiac monitoring at the wrist, *Scientific Reports*, **9**: 20079, https://doi.org/10.1038/s41598-019-55599-5.
93. Shih C., Tomita N., Lukic Y.X., Reguera Á.H., Fleisch E., Kowatsch T. (2019), Breeze: Smartphone-based acoustic real-time detection of breathing phases for a gamified biofeedback breathing training, [in:] *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **3**(4): 1–30, https://doi.org/10.1145/3369835.
94. Shiozawa K., Ozawa K., Ise T. (2020), Noise suppression using a differential-type microphone array and two-dimensional amplitude and phase spectra, [in:] *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 46–51.
95. Shokouhmand A., Eckstrom S., Gholami B., Tavassolian N. (2022), Camera-augmented non-contact vital sign monitoring in real time, *IEEE Sensors Journal*, **22**(12): 11965–11978, https://doi.org/10.1109/JSEN.2022.3172559.
96. Smithard J. *et al.* (2017), An advanced multi-sensor acousto-ultrasonic structural health monitoring system: Development and aerospace demonstration, *Materials*, **10**(7): 832, https://doi.org/10.3390/ma10070832.
97. Sun L., Luo C., Bai G., Gu C., Zhu X. (2022), Remote measurement of human breathing and heartbeat in large scene based on TX-RX joint beamforming, [in:] *2022 16th ICME International Conference on Complex Medical Engineering (CME)*, pp. 12–15, https://doi.org/10.1109/CME55444.2022.10063285.
98. Taghavirashidizadeh A., Sharifi F., Vahabi S.A., Hejazi A., SaghabTorbati M., Mohammed A.S. (2022), WTD-PSD: Presentation of novel feature extraction method based on discrete wavelet transformation and time-dependent power spectrum descriptors for diagnosis of Alzheimer's disease, *Computational Intelligence and Neuroscience*, **2022**: 9554768, https://doi.org/10.1155/2022/9554768.
99. Takenaka K., Ozawa K. (2022), Noise suppression system using deep learning for smart devices, [in:] *2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech)*, pp. 559–560, https://doi.org/10.1109/LifeTech53646.2022.9754759.
100. Taniguchi M., Kageyama T., Okamoto Y., Matsunaga T., Lee S.S. (2023), A vital sign monitoring system using a MEMS microphone for dog, [in:] *2023 IEEE 18th International Conference on Nano/Micro Engineered and Molecular Systems (NEMS)*, pp. 70–73, https://doi.org/10.1109/NEMS57332.2023.10190878.
101. Tashev I., Acero A. (2006), Microphone array post-processor using instantaneous direction of arrival, [in:] *Proceedings of International Workshop on Acoustic, Echo and Noise Control IWAENC 2006*, https://www.microsoft.com/en-us/research/publication/microphone-array-post-processor-using-instantaneous-direction-of-arrival/.
102. Thakur D., Biswas S., Ho E.S.L., Chattopadhyay S. (2022), ConvAE-LSTM: Convolutional autoencoder long short-term memory network for smartphone-based human activity recognition, *IEEE Access*, **10**: 4137–4156, https://doi.org/10.1109/ACCESS.2022.3140373.
103. Thomsen H., Du Y. (2020), Digital microphone noise attenuation, US Patent 10,559,293.
104. Thundat T.G., Oden P.I., Datskos P.G. (2000), Non-contact passive temperature measuring system and method of operation using micro-mechanical sensors, US Patent 6,050,722.

105. Toa M., Whitehead A. (2019), Ultrasonic sensing basics. Application note, *Texas Instruments*, https://www.ti.com/lit/an/slaa907d/slaa907d.pdf (access: 30.05.2025).
106. Tocci M. (2024), *Smartphone history and evolution*, Simple Texting, https://simpletexting.com/blog/where-have-we-come-since-the-first-smartphone/ (access: 3.04.2024).
107. Todorović D. *et al.* (2015), Multilayer graphene condenser microphone, *2D Materials*, **2**(4): 045013, https://doi.org/10.1088/2053-1583/2/4/045013.
108. Tran D., Duong K., Bhowmik U.K. (2014), A VHDL based controller design for non-contact temperature and breathing sensors suitable for crib, [in:] *2014 IEEE International Conference on Bioinformatics and Bioengineering*, 126–133, https://doi.org/10.1109/BIBE.2014.11.
109. Tsai T.C., Orav E.J., Jha A.K., Figueroa J.F. (2022), National estimates of increase in US mechanical ventilator supply during the COVID-19 pandemic, *JAMA Network Open*, **5**(8): e2224853, https://doi.org/10.1001/jamanetworkopen.2022.24853.
110. Valipour A., Abbasi-Kesbi R. (2017), A heartbeat and respiration rate sensor based on phonocardiogram for healthcare applications, [in:] *2017 Iranian Conference on Electrical Engineering (ICEE)*, pp. 45–48, https://doi.org/10.1109/IranianCEE.2017.7985502.
111. Vincent K., Wang L., Liu W. (2023), Towards one-to-many respiration monitoring via dual-mode acoustic signals, *IEEE Transactions on Consumer Electronics*, **70**(1): 2970–2978, https://doi.org/10.1109/TCE.2023.3321683.
112. Wahyu Y. *et al.* (2022), 24 GHz FMCW radar for non-contact respiratory detection, [in:] *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, pp. 752–755, https://doi.org/ 10.1109/ICITISEE57756.2022.10057704.
113. Wan H., Shi S., Cao W., Wang W., Chen G. (2023), Multi-user room-scale respiration tracking using COTS acoustic devices, *ACM Transactions on Sensor Networks*, **19**(4): 85, https://doi.org/10.1145/3594220.
114. Wang G., Zhang C., Chen X., Ji X., Xue J., Wang H. (2020), Bi-stream pose-guided region ensemble network for fingertip localization from stereo images, *IEEE Transactions on Neural Networks and Learning Systems*, **31**(12): 5153–5165, https://doi.org/10.1109/TNNLS.2020.2964037.
115. Wang T. *et al.* (2023), *MultiResp*: Robust respiration monitoring for multiple users using acoustic signal, *IEEE Transactions on Mobile Computing*, **23**(5): 3785–3801, https://doi.org/10.1109/TMC.2023.3279976.
116. Wang T., Zhang D., Zheng Y., Gu T., Zhou X., Dorizzi B. (2018), C-FMCW based contactless respiration detection using acoustic signal, [in:] *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, **1**(4): 1–20, https://doi.org/10.1145/3161188.
117. Wang Y., Wang Z., Zhang J.A., Zhang H., Xu M. (2023), Vital sign monitoring in dynamic environment via mmWave radar and camera fusion, *IEEE Transactions on Mobile Computing*, **23**(5): 4163–4180, https://doi.org/10.1109/TMC.2023.3288850.
118. Wang Z., Qiu F. (2020), Voice signal noise reduction processing method, microphone and electronic equipment, Unified Patents, Patent No. CN-109065067-A.
119. Wang Z., Zhang F., Li S., Jin B. (2021), Exploiting passive beamforming of smart speakers to monitor human heartbeat in real time, [in:] *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, https://doi.org/10.1109/GLOBECOM46510.2021.9685922.
120. Wu S., Wang W., Bian Y., Xu Z. (2020), Wind noise processing method, device, and system employing multiple microphones, and storage medium, Patent.
121. Xiao S., Ji X., Yan C., Zheng Z., Xu W. (2023), MicPro: Microphone-based voice privacy protection, [in:] *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1302–1316, https://doi.org/10.1145/3576915.3616616.
122. Xiao W., Yu L. (2021), Non-contact passive sensing of acoustic emission signal using the air-coupled transducer, *Health Monitoring of Structural and Biological Systems XV*, **11593**: 412–419, https://doi.org/10.1117/12.2583218.
123. Xie W., Tian R., Zhang J., Zhang Q. (2020), Noncontact respiration detection leveraging music and broadcast signals, *IEEE Internet of Things Journal*, **8**(4): 2931–2942, https://doi.org/10.1109/JIOT.2020.3021915.
124. Xie Z., Nederlander A., Park I., Ye F. (2023), Poster: Quantifying signal quality using autoencoder for robust RF-based respiration monitoring, [in:] *Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies*, pp. 187–188, https://doi.org/10.1145/3580252.3589999.
125. Xiong J., Hong H., Xiao L., Wang E., Zhu X. (2023), Vital signs detection with difference beamforming and orthogonal projection filter based on SIMO-FMCW radar, *IEEE Transactions on Microwave Theory and Techniques*, **71**(1): 83–92, https://doi.org/10.1109/TMTT.2022.3181129.
126. Xu W. *et al.* (2021), Unobtrusive vital signs and activity monitoring based on dual mode fiber, *Optical Fiber Technology*, **64**: 102530, https://doi.org/10.1016/j.yofte.2021.102530.
127. Xu W., Shen Y., Yu C., Dong B., Zhao W., Wang Y. (2020), Long modal interference in multimode fiber and its application in vital signs monitoring, *Optics Communications*, **474**: 126100, https://doi.org/10.1016/j.optcom.2020.126100.
128. Xu X., Yu J., Chen Y. (2022), Leveraging acoustic signals for fine-grained breathing monitoring in driving environments, *IEEE Transactions on Mobile Computing*, **21**(3): 1018–1033, https://doi.org/10.1109/TMC.2020.3015828.

129. Yang F., He S., Sadanand S., Yusuf A., Bolic M. (2022), Contactless measurement of vital signs using thermal and RGB cameras: A study of COVID-19-related health monitoring, *Sensors*, **22**(2): 627, https://doi.org/10.3390/s22020627.

130. Zakrzewski M. (2015), *Methods for Doppler Radar Monitoring of Physiological Signals*, Ph.D. Thesis, Tampere University of Technology, Finland, https://trepo.tuni.fi//handle/10024/114695.

131. Zhang G., Geng L., Xie F., He C.-D. (2024), A dynamic convolution-transformer neural network for multiple sound source localization based on functional beamforming, *Mechanical Systems and Signal Processing*, **211**: 111272, https://doi.org/10.1016/j.ymssp.2024.111272.

132. Zhang H. *et al.* (2023a), Radar-Beat: Contactless beat-by-beat heart rate monitoring for life scenes, *Biomedical Signal Processing and Control*, **86**(Part C): 105360, https://doi.org/10.1016/j.bspc.2023.105360.

133. Zhang X., Bao Z., Yin Y., Yang X., Xu X., Niu Q. (2023b), Finding potential pneumoconiosis patients with commercial acoustic device, [in:] *2023 IEEE Symposium on Computers and Communications (ISCC)*, pp. 310–315, https://doi.org/10.1109/ISCC58397.2023.10217963.

134. Zhao T., Fu X., Zhou Y., Zhan J., Chen K., Li Z. (2023), Noncontact monitoring of heart rate variability using a fiber optic sensor, *IEEE Internet of Things Journal,* **10**(17): 14988–14994, https://doi.org/10.1109/JIOT.2023.3262634.