

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. 1, 2025

WARSAW



ARCHIVES of ACOUSTICS

QUARTERLY, Vol. 50, No. 1, 2025

Research Papers

- M. Yaman, C. Kurtay, G. Ulukavak Harputlugil, *Prediction models with multiple linear regression for improving acoustic performance of textile industry plants* 3
- V.H. Trinh, M. He, *Experimental characterization of sound absorption for composite panel made of perforated plate and membrane foam layer* 17
- A. Szeląg, M. Zastawnik, *Issues in the design and validation of coupled reverberation rooms for testing acoustic insulation of building partitions* 25
- B.J. Kriston, K. Jálcs, *Failure detection of powertrain components in motor vehicles using vibroacoustic methods* 37
- M. Hałucha, J. Bohatkiewicz, P. Mioduszewski, T. Berge, *Tyre labelled noise values in the context of environmental protection: Weaknesses of the method and benefits of silent tyres* 47
- S. Gmyrek, R. Hossa, R. Makowski, *The influence of the amplitude spectrum correction in the HFCC parametrization on the quality of speech signal frame classification* 59
- Y. Luo, J. Peng, L. Ding, Y. Zhang, L. Song, Q. Zhang, H. Chen, *Snoring sounds classification of OSAHS patients based on model fusion* 69
- P. Antoniuk, S.K. Zieliński, *Estimating ensemble location and width in binaural recordings of music with convolutional neural networks* 81
- R. Halama, K. Szklanny, D. Koržinek, *Method for vocal fold paralysis detection based on perceptual and acoustic assessment* 95
- M. Ahangar Darband, E. Najafiaghdam, *Implementation of a cost-effective, accurate photoacoustic imaging system based on high-power LED illumination and FPGA-based circuitry* 107
- X. Yang, G. Zheng, F. Wang, F. Zhu, L. Bai, *Inference of bubble size distribution in sediments based on sounding by chirp signals* 115
- T. Sun, M. Zhou, L. Chen, *An under-sampled line array element signal reconstruction method based on compressed sensing theory* 127

Review Paper

- S. Wang, Y. Yu, J. Meng, *A review of the sonication-assisted exfoliation methods for MoX_2 (X : S, Se, Te) using water and ethanol* 137

Technical Note

- A.K.S. Chauhan, A. Vedrtam, S.J. Pawar, *Experimental and numerical investigations of acoustic variations in a classroom environment* 147

Editorial Board

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

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

Associate Editors

General linear acoustics and physical acoustics

RDZANEK Wojciech P. (University of Rzeszów, Poland)

SNAKOWSKA Anna (AGH University of Krakow, Poland)

Architectural acoustics

KAMISIŃSKI Tadeusz (AGH University of Krakow, Poland)

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

Musical acoustics and psychological acoustics

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

PREIS Anna (Adam Mickiewicz University, Poland)

Underwater acoustics and nonlinear acoustics

MARSZAŁ Jacek (Gdańsk University of Technology, Poland)

Speech, computational acoustics, and signal processing

DRGAS Szymon (Poznan University of Technology)

KOCIŃSKI Jędrzej (Adam Mickiewicz University, Poland)

Ultrasonics, transducers, and instrumentation

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

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

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

Sonochemistry

DZIDA Marzena (University of Silesia in Katowice, Poland)

Electroacoustics

ŻERA Jan (Warsaw University of Technology, Poland)

Vibroacoustics, noise control and environmental acoustics

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

KLEKOT Grzegorz (Warsaw University of Technology, Poland)

KOMPAŁA Janusz (Central Mining Institute, Poland)

LENIOWSKA Lucyna (University of Rzeszów, Poland)

PIECHOWICZ Janusz (AGH University of Krakow, Poland)

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

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

Advisory Editorial Board

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

BATKO Wojciech (AGH University of Krakow, Poland)

BLAUERT Jens (Ruhr University, Germany)

BRADLEY David (The Pennsylvania State University, USA)

CROCKER Malcolm J. (Auburn University, USA)

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

HANSEN Colin (University of Adelaide, Australia)

HESS Wolfgang (University of Bonn, Germany)

LEIGHTON Tim G. (University of Southampton, UK)

LEWIN Peter A. (Drexel University, USA)

MAFFEI Luigi (Second University of Naples SUN, Italy)

PUSTELNY Tadeusz (Silesian University of Technology, Poland)

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

SUNDBERG Johan (Royal Institute of Technology, Sweden)

ŚLIWIŃSKI Antoni (University of Gdańsk, Poland)

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

TORTOLI Piero (University of Florence, Italy)

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

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

Editorial Board Office

Pawińskiego 5B, 02-106 Warsaw, Poland

phone (48) 22 826 12 81 ext. 206

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

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

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

Directory of Open Access Journals (DOAJ)

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




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

PUBLISHED IN POLAND

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

Research Paper

Prediction Models with Multiple Linear Regression for Improving Acoustic Performance of Textile Industry Plants

Muammer YAMAN^{(1)*} , Cüneyt KURTAY⁽²⁾ ,
Gülsu ULUKAVAK HARPUTLUGİL⁽³⁾ 

⁽¹⁾ *Department of Architecture, Faculty of Architecture, Ondokuz Mayıs University
Samsun, Türkiye*

⁽²⁾ *Department of Architecture, Faculty of Fine Arts, Design and Architecture, Başkent University
Ankara, Türkiye; e-mail: cuneytkurtay@baskent.edu.tr*

⁽³⁾ *Department of Architecture, Faculty of Architecture, Çankaya University
Ankara, Türkiye; e-mail: gharputlugil@cankaya.edu.tr*

*Corresponding Author e-mail: muammer.yaman@omu.edu.tr

(received June 6, 2024; accepted September 6, 2024; published online January 9, 2025)

In industrial plants noise is a major threat to the mental and physical health of employees. The risk increases more due to the presence of high noise sources and the presence of too many employees in textile industry plants. This paper aims to predict the consequences of variables that may arise in the plants for acoustic improvement in textile industry plants. For this purpose, scenario plants have been created according to architectural properties and source-transmission path-receiver characteristics. The acoustic analyses of the scenario plants were performed in the ODEON Auditorium, and *A*-weighted sound pressure level (LA), noise reduction (NR), and reverberation time (RT) were determined. From the data, prediction equations were created with a multiple linear regression (MLR) model. To test the prediction equations, acoustic measurements were made, and acoustics improvements were carried out at a textile industry plant located in Türkiye. When the obtained results, the success, validity, and reliability of the prediction method are provided. In conclusion, the effect of architectural properties and the surface absorption on acoustic improvements in the textile industry was revealed. It was emphasized that prediction methods can be used to determine the effectiveness of interventions that can be applied in different facilities and can be improved in future studies.

Keywords: industrial noise control; acoustics simulation; multiple linear regression; prediction methods; textile industry; ODEON Auditorium; noise reduction; reverberation time.



Copyright © 2024 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Noise is one of the physical environmental factors that affect our mental and physical health in today's world. Noise is generally defined as unpleasant sounds that disturb people physically and physiologically and cause environmental pollution by disrupting environmental values (JOB, 1996; KURRA, 2020; DURÁN DEL AMOR *et al.*, 2022). Noise has not only physical and psychological effects on individuals but also many negative effects on employee productivity

(REINHOLD, TINT, 2009; FREDRIKSSON *et al.*, 2015; BAKER, 2015). Industrial plants with intensive working areas pose a risk to many employees as areas with high noise levels. By eliminating the risks, the health of the employees should be created by occupational safety (LEATHER *et al.*, 2003; THEMANN, MASTERSON, 2019; MASULLO *et al.*, 2022). For this purpose, regulations have been made to limit the noise exposure of industry employees in many countries (ARENAS, SUTER, 2014). For example, the Occupational Noise Exposure Regulation in the USA states that the

noise exposure level of employees should be limited to 90 dB(A) for 8 hours ([Occupational Safety and Health Administration, 1995](#)). In Türkiye, in line with the directive of the [European Parliament and the Council of the European Union \(2003\)](#), the exposure limit value is $L_{EX,8h} = 87$ dB(A); $p_{peak} = 200$ Pa. The relevant limit values applied by different countries vary.

The textile industry has developed to a great extent with its close to raw materials and high export rates in Türkiye. Recently, thanks to this development and employment opportunities, the number of employees in textile industry plants has been increasing. High noise in textile industry plants affects employees negatively. Research studies on sound pressure level measurements and noise exposure level measurements are carried out in textile industry plants. [ABBASI *et al.* \(2020\)](#) found that in a textile industry plant, 42.1 % (77) of the employees were exposed to noise below the limit value of 85 dB(A), and 57.9 % (106) of them were exposed to noise above 85 dB(A). In the acoustics measurements they made at the textile industry plant, [YAMAN TURAN and ÖNEY \(2021\)](#) determined that the noise level in the area where the weaving machines are placed varies between 92 dB(A)–97 dB(A), and the noise level in other areas decreases to approximately 82 dB(A). [ZAW *et al.* \(2020\)](#) stated that 66.4 % of the employees in the textile industry plant were exposed to noise above 85 dB(A) and determined the prevalence of hearing loss among the employees as 25.7 % with hearing tests. [ATMACA *et al.* \(2005\)](#) determined that the employees in the textile and cement factory were exposed to very high noise levels with the acoustic measurements they applied in different plants. In particular, they determined that 60 % of those working in the textile industry were exposed to noise at a maximum level of 106 dB(A). [EJIGU \(2019\)](#) determined that the noise exposure level is over 90 dB(A) in the acoustics measurements. Studies have revealed that there are high sound pressure levels in textile industry plants, and this may have negative effects on employees.

Noise, created in textile industry plants, adversely affects the health and task performance of employees. [ALI \(2011\)](#) determined that 47.1 % of the employees of different industrial plants are highly annoyed by noise. It has been determined that there is a significant and positive relationship between noise level and the percentage of employees' noise annoyance. In a study conducted in Pakistan, it was determined that 79 % of textile industry plant employees had hearing loss at levels of 25 dB and above ([SHAHID *et al.*, 2018](#)). Similarly, in the study, hearing loss in employees exposed to high noise levels increases approximately four times compared to normal conditions. Additionally, it has been determined that hearing loss increases as the noise exposure in the plants increases, and the employment time increases ([SHAKHATREH *et al.*, 2000](#)). [AL-DOSKY \(2014\)](#) determined that textile industry

plant employees had a high level of noise annoyance and determined that there was a significant relationship between noise annoyance and employment time. It has been observed in the studies that the employees in the textile industry plants are greatly affected by the noise; and as a result of this, the employees encounter physiological and psychological problems. As a result of the research, it has emerged that the noise problems in the plants should be eliminated, and the appropriate acoustical environments should be created. Various acoustic improvement studies are carried out with computer simulations and models. [MONAZZAM and NAZAFAT \(2007\)](#) used acoustic barriers to reduce spinning machine noise, compared the application and mathematical methods, and obtained effective results in noise reduction (NR). They evaluated the results as related to the high internal absorption. [ILGÜREL \(2013\)](#) investigated the effect of total absorption on NR in all industrial plants by a simulation method. [JAYAWARDANA *et al.* \(2014\)](#) conducted experimental studies on noise control by constructing a mathematical prediction model of the noise determined by measurements. It has been observed that noise can be reduced at high frequencies as a result of the use of suspended ceilings through simulations. The reliability of the model was determined by comparing the results obtained from simulations and prediction models. [MONAZZAM-ESMAEELPOUR *et al.* \(2014\)](#) investigated the effect of the surface absorption on NR by computation in a textile industry plant. Effective results were obtained in NR at high frequencies, and they recommended the use of sound absorption materials with an air gap and increasing the thickness of sound absorption materials for low frequencies. Studies indicate the effective results of noise control measures to reduce noise in textile industry plants.

Reducing noise in textile industry plants is achieved by reducing the sound pressure level and controlling the reverberation time – called RT ([CHATILLON, 2007](#)). For RTs, analysis was performed especially in the mid-frequency bands, and prediction methods were created on 500 Hz ([BISTAFA, BRADLEY, 2000](#); [YAHYA *et al.*, 2010](#); [NOWOŚWIAT, 2023](#)). Determining the interventions that can be made for this purpose and estimating their effectiveness provides practical convenience. Mathematical models, simulations, and prototypes constitute the prediction methods used for this purpose ([BISTAFA, BRADLEY, 2000](#); [PROBST, 2012](#); [FICHERA, 2020](#)). In this paper, acoustic simulations were applied in various textile industry plant scenarios, and prediction models were created for the analysis of acoustical and non-acoustical parameters (independent variables) using multiple linear regression (MLR) analyses. Prediction models include the testing of interventions and analysis of their effectiveness and offer solutions to reduce noise for employees. It also provides a guide for researchers, acousticians, and employers.

2. Materials and methods

2.1. Acoustics scenarios

Scenario plants were created to make acoustic performance evaluations in textile industry plants and compare the effects of interventions. Scenario plants were designed based on the textile industry plants located in the Republic of Türkiye and identified within the scope of the literature review. Independent variables affecting indoor acoustic performance were created through scenario plants. The independent variables were designed as architectural properties (geometry-width-length-height), source characteristics (number of machines, sound power level, frequency spectrum), transmission path characteristics (wall and ceiling sound absorption materials), and receiver characteristics. As a result of the crossover of the independent variables, 480 different textile industry plants were created. The dependent variables investigated were determined as the indoor *A*-weighted sound pressure level (LA), NR, and RT, which are effective acoustic parameters for NR. For this purpose, the effects of different independent variables on the dependent variables were investigated. The improvement of the acoustic performance approach is primarily based on the implementation of engineering. Engineering controls that can be applied in textile industry plants and can provide high efficiency for the purpose are examined, and the effects of the precautions in a virtual environment (ODEON Acoustics) are investigated.

Different scenario plants were created by crossover architectural properties, source, transmission path, and receiver characteristics to control noise distribution and mitigation in textile industrial plants. Architectural properties (K1–D5), source characteristics (Y1, Y2, F1, F2), transmission path characteristics (S1–S12), and receiver characteristics (A1) components were used in the crossover (independent variables). As a result of the crossovers, a total of 480 (240 square plans / 240 rectangular plans) different simulation outputs were obtained (Fig. 1). The LA, NR values, and RT values (dependent variables) were investigated and analyzed in the scenario plants defined as KX/DXYFXSXA1.

Scenario plants, which are analyzed through square (1:1) and rectangular plan schemes (2.5:1) as two basic geometry forms, can also be designed as more complex structures. However, square/rectangular main geometries that can be divided into rational units are prioritized in this research. Variables were created to examine the effects of width, length, and height properties in square and rectangular plans. To compare the square and rectangular plans with each other, their areas [m²] and volumes [m³] were kept at equal values (Table 1). The plants with square and rectangular plans represent five variable plants each. In the analysis of architectural properties in the created scenario plants, evaluations were made depending on the increase in the main area by taking the height constant (K2-K3-K4/D2-D3-D4); with a similar situation, evaluations were made depending on the increase in height within the same main area (K1-K3-K5/D1-D3-D5).

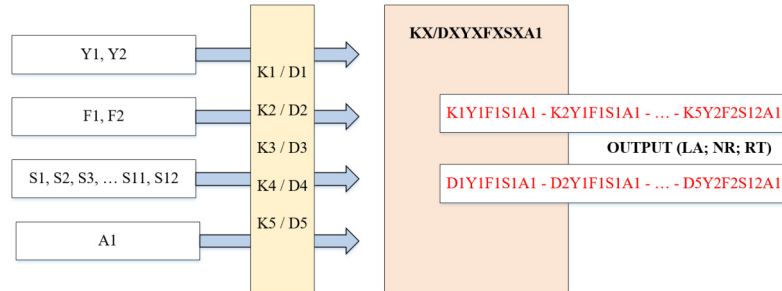


Fig. 1. Formation of different scenario plants.

Table 1. Plan geometries in scenario plants.

Plan geometry code	Length L [m]	Width W [m]	Height h [m]	Area A [m ²]	Volume V [m ³]
K1 (square)	40	40	5	1 600	8 000
K2 (square)	20	20	7	400	2 800
K3 (square)	40	40	7	1 600	11 200
K4 (square)	80	80	7	6 400	44 800
K5 (square)	40	40	9	1 600	14 400
D1 (rectangular)	64	25	5	1 600	8 000
D2 (rectangular)	32	12.5	7	400	2 800
D3 (rectangular)	64	25	7	1 600	11 200
D4 (rectangular)	128	50	7	6 400	44 800
D5 (rectangular)	64	25	9	1 600	14 400

Weaving and spinning machines (open-end and ring spinning) were taken as the basis for examining the acoustic performance in textile industry plants within the scope of source characteristics. Weaving and spinning machines constitute the series of machines that produce the highest noise level in textile industry plants. Two different variables are considered for source characteristics:

- less dense (infrequent) layouts and more dense (frequent) layouts of sources;
- using sources with high frequency and sources with flat frequency distribution in terms of sound power levels.

The layouts of noise sources (more and less dense) in textile industry plants were created based on the number of machines per area of textile industry plants located in the Republic of Türkiye and determined within the scope of the literature review (more dense: approximate values: area/25-frequent placement; less dense: area/50-infrequent placement). From the values, the highest number of machines and the lowest number of machines were analyzed through two variables as machine layout variables (Table 2).

Two different frequency spectrum distributions were accepted in the sound power level distributions of noise sources in scenario plants. These are general hypothetical sound power level spectra obtained from the researched machine catalogues. Two variables were created according to the use of sources with a high frequency spectrum distribution and sources with a flat frequency spectrum distribution (Table 3).

In acoustic performance in textile industry plants, the effect of the surface absorption on dependent variables within the scope of transmission path properties was investigated. In the investigation of the effects of the surface absorption on indoor NR, floor, wall, and ceiling were examined. Due to the industrial floor in the plants, a finish material with high sound reflectivity properties (which cannot be changed) was defined (ODEON Code: 100). A constant sound absorption coefficient was assumed for the floors in all scenario plants (Table 4).

In the analysis of transmission path properties, scenarios allowing the comparison of ceiling and wall were created separately. The effect of sound absorption materials used in the lower (S8) and upper (S9) parts

Table 2. Number of weaving and spinning machines.

Number of sources code	Number of machines			h [m]	A [m ²]	V [m ³]
	Less dense (Y1) – infrequent layout	Mean	More dense (Y2) – frequent layout			
K1, D1	36	50	64	5	1 600	8 000
K2, D2	9	12.5	16	7	400	2 800
K3, D3	36	50	64	7	1 600	11 200
K4, D4	144	192	256	7	6 400	44 800
K5, D5	36	50	64	9	1 600	14 400

Table 3. Frequency distribution of the sound power levels of the sources.

Frequency spectrum code	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	Overall sound power level
F1 Flat frequency	93	93	93	93	93	93	93	93	102
F2 High frequency	78	81	84	87	90	93	96	99	102

Table 4. Weighted sound absorption coefficients of building components in scenario plants*.

Sound absorption coefficients code	Description	Floor	Walls		Ceiling
			Lower part	Upper part	
S1	Live room (high sound reflection)	0.05	0.1	0.1	0.1
S2	Ceiling with medium absorption	0.05	0.1	0.1	0.5
S3	Ceiling with medium absorption (planar)	0.05	0.1	0.1	0.5
S4	Ceiling with medium absorption (baffle)	0.05	0.1	0.1	0.5
S5	Ceiling with medium absorption (canopi)	0.05	0.1	0.1	0.5
S6	Ceiling with high absorption	0.05	0.1	0.1	0.9
S7	Walls with medium absorption	0.05	0.5	0.5	0.1
S8	Walls with medium absorption (lower)	0.05	0.5	0.1	0.1
S9	Walls with medium absorption (upper)	0.05	0.1	0.5	0.1
S10	Walls with high absorption	0.05	0.9	0.9	0.1
S11	Ceiling and walls with medium absorption	0.05	0.5	0.5	0.5
S12	Dead room (high sound absorption)	0.05	0.9	0.9	0.9

*ODEON codes were used to define the sound absorption coefficients of materials: 10 for 0.1; 50 for 0.5; 90 for 0.9.

of the walls on the indoor acoustic performance was analyzed. Additionally, by using the same amount of materials on the ceiling (planar-S3, baffle-S4, canopy-S5 scenarios), the effect of the differentiation of sound absorption materials due to shaping on the indoor acoustic performance was investigated. The fact that the materials are in the same quantities reveals the effect values of the sound absorption materials on the LA, NR, and RT (dependent variables) according to their formal properties.

In the examination of acoustic performance in the textile industry, a homogeneous layout of receivers within the scope of employee characteristics was taken as a basis. Employees have been assigned to each machine to use the machines specified according to Table 2. Analyses were carried out in the form of point receiver calculations to determine the general distribution within the main area in determining sound pressure level distributions and RTs. In point receiver calculations, 150 cm was taken as the ear height of the standing individuals from the floor. Point receiver calculations were based on the homogeneous distribution (A1) to represent individuals standing at different points.

2.2. Prediction models

The relationships between dependent and independent variables in the scenario plants were investigated by regression analysis. Four different plant types were categorized by crossing the components of square and rectangular plan layouts and machine sound power level frequency distributions. The four different plants selected were created using nominal (categorical) variables. The MLR method was used to explain the effects of independent variables on the dependent variables. With the regression equations created to predict the dependent variables, a prediction model for acoustic performance improvement in textile industry plants was created. The recommendations are based on the principle of obtaining appropriate dependent variables by differentiating the independent variables.

The MLR is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The purpose of MLR is to model the linear relationship between independent (explanatory) variables and dependent (response) variables. Since MLR models include more than one independent variable, they use the ordinary least squares (OLS) method as a regression extension (MCINTOSH *et al.*, 2010). Studies on the prediction of variables in acoustic research can be carried out with regression analysis (KUMAR, KUMAR, 2016; BAFFOE, DUKER 2018; TANG *et al.*, 2018; YANG, 2019):

$$\gamma = k + aX_1 + bX_2 + cX_3 \dots + \text{error}, \quad (1)$$

where k is a constant, X_1 , X_2 , X_3 , etc., are the independent variables, a , b , c , etc., are the coefficient of

independent variables, and the error term is taken as the difference between the observed and predicted values of the dependent variable (γ). The lower the error term, the lower the difference between the predicted value and the observed value. Depending on the unit of the estimated dependent variable, the error term may have different numerical magnitudes.

Two different analyses were conducted for the square-plan plants, with flat frequency sound power levels and the plants with high frequency sound power levels. In the analyses, the area, the height, the number of machines, the weighted sound absorption coefficient of the walls, and the weighted sound absorption coefficient of the ceiling were found to be effective for the LA; the height, the weighted sound absorption coefficient of the walls, and the weighted sound absorption coefficient of the ceiling were found to be effective for NR; the width, the height, the weighted sound absorption coefficient of the walls, and the weighted sound absorption coefficient of the ceiling were found to be effective for RT. The coefficient of determination (R^2) values equal to the square of the linear correlation coefficient between the dependent variables and the independent variables were determined (Eqs. (2)–(7)).

Plants with a square plan and flat frequency of machine sound power levels-1:

$$\begin{aligned} \text{LA}_1 &= 96.48 - 0.001A - 0.22h + 0.034n \\ &\quad - 3.65w_{\alpha_w} - 4.97c_{\alpha_w}, \end{aligned} \quad (2)$$

$$R^2 = 0.795 \text{ and } p < 0.01,$$

$$\begin{aligned} \text{NR}_1 &= 2.36 - 0.12h + 2.66w_{\alpha_w} + 3.98c_{\alpha_w}, \\ R^2 &= 0.871 \text{ and } p < 0.01, \end{aligned} \quad (3)$$

$$\begin{aligned} \text{RT}_{500 \text{ Hz}_1} &= 1.86 + 0.015d + 0.25h - 2.52w_{\alpha_{500 \text{ Hz}}} \\ &\quad - 2.07c_{\alpha_{500 \text{ Hz}}}, \end{aligned} \quad (4)$$

$$R^2 = 0.804 \text{ and } p < 0.01.$$

Plants with a square plan and high frequency of machine sound power levels-2:

$$\begin{aligned} \text{LA}_2 &= 96.78 - 0.001A - 0.21h + 0.034n \\ &\quad - 2.92w_{\alpha_w} - 4.14c_{\alpha_w}, \end{aligned} \quad (5)$$

$$R^2 = 0.769 \text{ and } p < 0.01,$$

$$\begin{aligned} \text{NR}_2 &= 2.23 - 0.12h + 2.21w_{\alpha_w} + 3.43c_{\alpha_w}, \\ R^2 &= 0.870 \text{ and } p < 0.01, \end{aligned} \quad (6)$$

$$\begin{aligned} \text{RT}_{500 \text{ Hz}_2} &= 1.83 + 0.015d + 0.25h - 2.52w_{\alpha_{500 \text{ Hz}}} \\ &\quad - 2.07c_{\alpha_{500 \text{ Hz}}}, \end{aligned} \quad (7)$$

$$R^2 = 0.803 \text{ and } p < 0.01.$$

LA_X is the A -weighted sound pressure level [dB] of the plant- x characteristics; NR_X is the NR [dB] of

the plant- x characteristics; $RT_{500\text{Hz}_x}$ is the RT [s] of the plant- x characteristics (500 Hz); A is the plan area [m^2]; d is the width/depth length [m]; h is the height [m], n is the number of machines; w_α is the sound absorption coefficient of the walls (500 Hz at RT); c_α is the sound absorption coefficient of the ceiling (500 Hz at RT).

For rectangular plants, machine sound power levels were analyzed in two different analyses, flat distributed and high frequency plants. In the analyses, the area, the height, the number of machines, the weighted sound absorption coefficient of the walls, and the weighted sound absorption coefficient of the ceiling were found to be effective for LA; the height, the weighted sound absorption coefficient of the walls and the weighted sound absorption coefficient of the ceiling were found to be effective for NR; the width, the height, the weighted sound absorption coefficient of the walls, and the weighted sound absorption coefficient of the ceiling were found to be effective for RT. The R^2 values equal to the square of the linear correlation coefficient between the dependent variables and the independent variables were determined (Eqs. (8)–(13)).

Plants with the rectangular plan and flat frequency of machine sound power levels-3:

$$LA_3 = 94.48 - 0.002A + 0.13d_k - 0.22h + 0.035n - 3.82w_{\alpha_w} - 4.72c_{\alpha_w}, \quad (8)$$

$$R^2 = 0.804 \text{ and } p < 0.01,$$

$$NR_3 = 2.36 - 0.12h + 2.84w_{\alpha_w} + 3.74c_{\alpha_w}, \quad (9)$$

$$R^2 = 0.884 \text{ and } p < 0.01,$$

$$RT_{500\text{Hz}_3} = 1.54 + 0.046d_k + 0.21h - 2.52w_{\alpha_{500\text{Hz}}} - 1.95c_{\alpha_{500\text{Hz}}}, \quad (10)$$

$$R^2 = 0.825 \text{ and } p < 0.01.$$

Plants with the rectangular plan and high frequency of machine sound power levels-4:

$$LA_4 = 95.30 - 0.002A + 0.10d_k - 0.22h + 0.035n - 3.13w_{\alpha_w} - 3.94c_{\alpha_w}, \quad (11)$$

$$R^2 = 0.778 \text{ and } p < 0.01,$$

$$NR_4 = 2.26 - 0.15h + 2.35w_{\alpha_w} + 3.17c_{\alpha_w}, \quad (12)$$

$$R^2 = 0.875 \text{ and } p < 0.01,$$

$$RT_{500\text{Hz}_4} = 1.52 + 0.046d_k + 0.21h - 2.51w_{\alpha_{500\text{Hz}}} - 1.95c_{\alpha_{500\text{Hz}}}, \quad (13)$$

$$R^2 = 0.824 \text{ and } p < 0.01,$$

where d_k is the short side length [m].

The presence of very different production processes in textile industry plants causes very different sound pressure levels in indoor acoustic performance. In this case, it should be known that the constant term in the calculation estimates used to determine the sound pressure levels in regression models can be taken as the sound pressure level of the measured existing situation. In the scenario plants, the LA in the high reflectivity scenarios (KX/DXYXFX'S1'A1) are in line with the sound pressure levels in the textile industry plants before the retrofit. According to the results of MLR analysis, the R^2 ranges between 0.769–0.804 for LA, 0.870–0.884 for NR, and 0.803–0.825 for RT. It is determined that the regression prediction models are at a level that can be applied for the acoustic performance improvement approach in textile industry plants.

3. Case study and test of prediction model

A textile industry plant examined as a case study includes open-end and ring spinning, knitting, and dyed yarn and dyed fabric. The department of open-end yarn spinning has an area of 13 000 m^2 and 12 Schlafhorst (Saurer) open-end machines. The department of ring spinning has 20 Rieter G-33 ring machines in an area of 16 700 m^2 . The textile industry plant is planned as the main production areas, storage units, technical rooms, and administrative departments. Production is carried out in the industrial plant with a daily three-shift system. The surface elements of the production area are formed with a lean concrete floor, partition walls made of metal, glass, and brick, and PVC suspended ceiling. It was observed that the finish materials of components were designed with high sound reflectivity properties. This increases the indoor sound pressure level and creates a noisy environment.

3.1. Acoustics measurements

Investigations were carried out to analyze the indoor acoustic performance in the textile industrial plant selected as a case study. The plant process machines in the indoor environment have high levels of sound power levels. Machines with high sound power levels (spinning ring machine – Rieter G-33 – has 103 dB sound power level) and surface elements with low sound absorption coefficients have caused high sound pressure levels. Testo 816-1 sound level meter (IEC 61672-1 Class 2) and occupational health and safety services dosimeters were used for acoustics measurements. Acoustic measurements in accordance with ISO 9612:2009 standard were carried out in the textile industry plant. The measurements revealed varying minimum-maximum sound pressure levels and the equivalent continuous sound level in different sections (Table 5). It was determined that the highest noise level in the plant was in the ring spinning and

Table 5. Acoustic measurements results in textile industry plant.

Measure no.	Acoustic measurements				Departments of the plant
	L_{\min} dB(A)	L_{\max} dB(A)	$L_{C,peak}$	L_{eq} dB(A)	
1	74.9	90.5	102.7	79.2	Blowroom-carding
2	73.6	85.5	100.3	76.1	Draw frame
3	72.2	82.5	96.0	76.5	Combing
4	76.3	85.8	97.7	80.1	Flyer
5	80.5	91.9	103.8	83.6	Ring yarn
6	77.9	82.3	96.7	79.8	Bobbin
7	71.3	76.2	89.4	74.3	Knitting
8	79.9	91.8	102.9	88.2	Open-end yarn
9	63.4	66.5	81.4	64.6	Sanforizing
10	70.9	80.5	92.3	72.9	Drying
11	71.7	74.1	87.8	72.8	HT 400 Boiler (painting)
12	77.5	84.4	96.8	80.8	Dry reversal
13	76.2	77.7	91.4	77.1	Yarn transfer

open-end spinning production departments. Noise exposure levels were found to be at high levels in parallel with the determined LA. Noise exposure levels in the range of 88.8 dB(A)–90.1 dB(A) ($L_{EX,8h}$) in the department of ring spinning and 86.9 dB(A)–92.8 dB(A) ($L_{EX,8h}$) in the department of open-end spinning were calculated.

There are 20 Rieter G-33 ring machines in the ring-spinning section which is accepted as the cross-sectional area. The cross-sectional area is 53.3 m × 44.4 m, and 4 m in height. The section is approximately 10 412 m³ (Fig. 3). The section has a suspended ceiling covering the air conditioning ducts. Reinforced concrete prefabricated vertical supports divide the working area into two systems. The section area is located after the flyer section. The ring spinning section is sep-

arated from the bobbin, knitting, and control rooms by dividing structural elements and operates independently. There are dividing walls (brick and plaster) on the long sides of the production area. Glass partitions separate the production area from the bobbin section (Fig. 2).

The production area was modeled in three dimensions in the SketchUp. Room acoustic modeling requirements were taken as the basis for modeling the in-plant properties. The necessary surface definitions were made in the 3D model, and the acoustic model was transferred to the acoustic computer simulation program (ODEON Auditorium) via the plugin (SU2Odeon). The acoustic performance of the current situation (digital acoustic twin) was created with the model transferred to ODEON Auditorium. The mate-

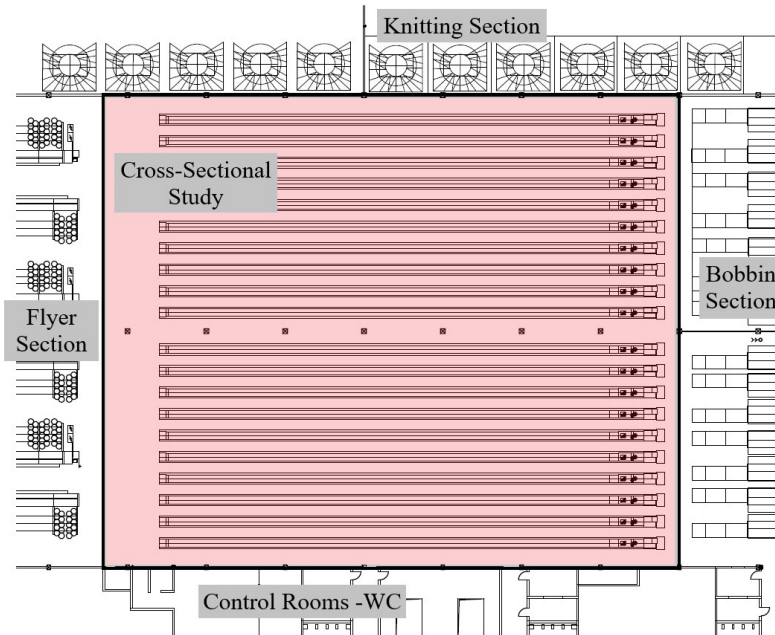


Fig. 2. Ring spinning section of the case study, cross-sectional study.

rials and surface absorptions used in the model were prepared following acoustic measurements. The building component separating the flyer and ring spinning sections obtained by zoning is defined as glass.

While creating the digital acoustic twin of the production area, acoustic calculations made indoors were utilized. The LAs obtained in acoustic measurements were checked, and acoustic performance values were obtained in real situations. In the acoustic measurements, the highest noise level among the ring-spinning machines in the indoor environment was determined as 91.9 dB(A). Digital acoustic twin indoor sound pressure levels were created as a minimum of 91.5 dB(A), maximum of 92.1 dB(A), and average of 91.9 dB(A).

Due to the high levels of noise exposure in the textile industry plant examined, the need for improvement of acoustic performance in the indoor environment has emerged. Indoor LA, NR, and RT were analyzed by changing the parameters affecting indoor acoustic performance on the digital acoustic twin. As a result of the analysis, acoustic improvements that are easy to implement and provide high efficiency were prioritized. The improvements, materials, and applications are presented, and acoustic performance values are determined. In the textile industrial plant, a composite material with a trapezoidal sheet on one side, a perforated sheet on the other side, and a rockwool-filled

core (interlayer) was selected for the suspended ceiling. The fact that the composite material is lightweight, applicable, and cheap has proven to be effective. Additionally, the selected material is non-combustible (A2-s1, d0) and resistant to impacts and pressure. The ceiling material was not used in air conditioning duct lines. On the walls, special sound absorption systems consisting of rock wool panels covered with aluminium-vinyl materials were used. The fact that the materials are lightweight, easy to install and have high sound absorption properties has been effective. Additionally, the special sound absorption system is a non-combustible material (A2-s1, d0) and is resistant to impacts and pressure (Table 6).

As a result of acoustic improvements in the ring-spinning section of the textile industry plant as a case study, the indoor minimum LA was determined as 82.7 dB(A), the maximum LA as 83.4 dB(A), and the average sound pressure level as 83.1 dB(A) (Fig. 3). As a result of acoustic improvements, the indoor RT (T30) was calculated as 0.54 s at 500 Hz, and 0.54 s at 1000 Hz (Fig. 3). The difference between the LA (NR) obtained as a result of acoustic improvements and the existing situation in the ring-spinning section selected for the case study was calculated as 8.7 dB. The values were found following the reference values in the regulation.

Table 6. Sound absorption coefficients of materials used in acoustic improvements.

Materials	63 Hz	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz	α_w
Floor (industrial floor-concrete)	0.02	0.02	0.03	0.03	0.03	0.04	0.07	0.07	0.05
Vertical structural elements (prefabricate concrete)	0.01	0.01	0.01	0.02	0.02	0.02	0.05	0.05	0.05
Walls* (rockwool panel)	0.12	0.47	0.47	0.85	0.84	0.64	0.62	0.62	0.70
Separators between sections* (rockwool panel)	0.12	0.47	0.47	0.85	0.84	0.64	0.62	0.62	0.70
Zoning – separator* (rockwool panel)	0.12	0.47	0.47	0.85	0.84	0.64	0.62	0.62	0.70
Transition between sections (plastic curtain)	0.8	0.8	0.9	0.9	0.9	0.9	0.1	0.1	0.1
Ceiling* (composite panel)	0.3	0.55	0.8	1	1	0.9	0.9	0.9	1
Ceiling (air conditioner ducts)	0.8	0.8	0.9	0.9	0.9	0.9	0.1	0.1	0.1

*It indicates new materials used in acoustic improvement phase.

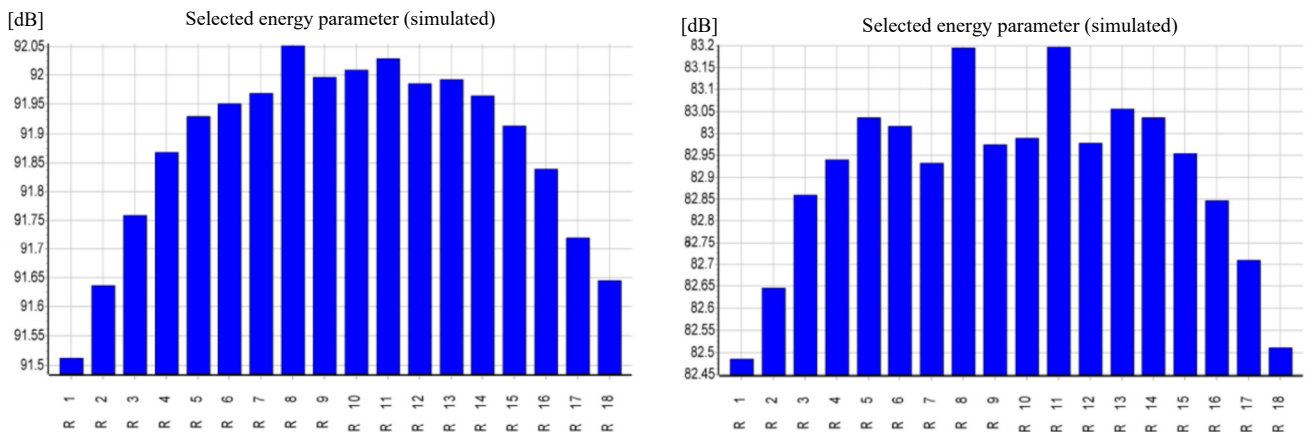


Fig. 3. Average sound pressure level before improvement (left), and after improvement (right).

Table 7. Comparison of acoustic simulation and prediction model results.

Acoustics measurement		Acoustics improvements (acoustics simulation / ODEON)			Acoustics improvements (prediction model)		
Average sound pressure level	Daily noise exposure level	LA	NR	RT (500 Hz)	LA	NR	RT (500 Hz)
LA	$L_{EX,8h}$	LA	NR	$RT_{500\text{ Hz}}$ (T30)	LA_1	NR_1	$RT_{500\text{ Hz}_1}$ (T30)
91.9	90.1	83.4	8.7	0.54	82.64	7.05	0.28

To test the validity of the prediction model, comparative research was carried out on the case study used in problem identification. In the comparative research, in the first phase, the existing situation of the plant was calibrated through the simulation program; a digital acoustic twin was created; and then acoustic improvements were made. In the second phase of the comparative research, acoustic improvements were organized in the existing textile industry plant according to the prediction model created (Eqs. (14)–(16)). For both phases, independent variables were investigated in the same method, and dependent variables were transferred. In the acoustic improvements, the limit values determined by the regulation were taken as the basis. In the prediction model, depending on the characteristics of the case study, the equations for square planned plants with flat frequency of machine sound power levels-1 were used (Table 7);

$$\begin{aligned}
LA_1 &= 91.9 - 0.001A - 0.22h + 0.034n \\
&\quad - 3.65w_{\alpha_w} - 4.97c_{\alpha_w}, \\
LA_1 &= 91.9 - 0.001 \cdot 2.389.63 - 0.22 \cdot 4 \\
&\quad + 0.034 \cdot 20 - 3.65 \cdot 0.6 - 4.97 \cdot 0.9,
\end{aligned} \tag{14}$$

$$LA_1 = 82.64 \text{ dB(A)},$$

$$\begin{aligned}
NR_1 &= 2.36 - 0.12h + 2.66w_{\alpha_w} + 3.98c_{\alpha_w}, \\
NR_1 &= 2.36 - 0.12 \cdot 4 + 2.66 \cdot 0.6 + 3.98 \cdot 0.9, \\
NR_1 &= 7.05 \text{ dB},
\end{aligned} \tag{15}$$

$$\begin{aligned}
RT_{500\text{ Hz}_1} &= 1.86 + 0.015d + 0.25h \\
&\quad - 2.52w_{\alpha_{500\text{ Hz}}} - 2.07c_{\alpha_{500\text{ Hz}}}, \\
RT_{500\text{ Hz}_1} &= 1.86 + 0.015 \cdot 53.34 + 0.25 \cdot 4 \\
&\quad - 2.52 \cdot 0.6 - 2.07 \cdot 0.9,
\end{aligned} \tag{16}$$

$$RT_{500\text{ Hz}_1} = 0.28 \text{ s}.$$

It was found that the difference between the result values of the prediction model prepared to be applied in the textile industry plants and the result values of the digital acoustic twin is at acceptable levels. The differences can be explained by the fact that for the digital acoustic twin, the data can be entered into the computer simulation program in detail, while in the prediction model setup, descriptive data are obtained

by calculations. Additionally, the coefficients of determination (R^2) in the equations used in the calculation estimations for the accepted independent variables also reveal the success of the prediction model. It is envisaged that the prediction model construct can be used in textile industry plants as well as in textile industry plants in the design and planning phase.

4. Results and discussion

In the research, scenario plants were created to analyze acoustic improvements in textile industrial plants. In the scenario plants, architectural properties, and source-transmission path-receiver characteristics were defined as independent variables (input data); LA, NR, and $RT_{500\text{ Hz}}$ were defined as dependent variables (output data). As a result of the research, the findings were obtained through MLR models and comparative analyses of the scenario plants for acoustic improvements.

Textile machines with high sound power levels have been identified in textile industry plants. Due to the identification of sound sources, indoor LAs were obtained at high levels (above 85 dB(A)) following real situations. Moreover, the plan geometry (square or rectangular) of the main production area in the scenario plant did not have a decisive influence on the analysis and improvement of the acoustic performance.

The LAs were found to be low in a relatively small area and volumes provided that the number of machines per area [m^2] in the main production areas remained constant. This situation is considered to be related to the reduction of sound sources. For this purpose, it is necessary to make small divisions within the main space for the function of textile industry plants, and then subdivisions/zoning should be created within the divisions. Approximately 2.5 dB NR in sound pressure levels was achieved with each sub-division ($1/2$ ratio). However, for indoor acoustic performance improvements in textile industry plants, frequency spectrum distributions of sound sources should be determined, and noise control measures should be developed. In the scenario plants, the reverberant sound field is intervened in the NR based on increasing the total absorption of the environment by using the surface absorption, and the A-weighted sound pressure levels of the indoor environment are reduced. In the scenario of textile industry plants, depending on the vari-

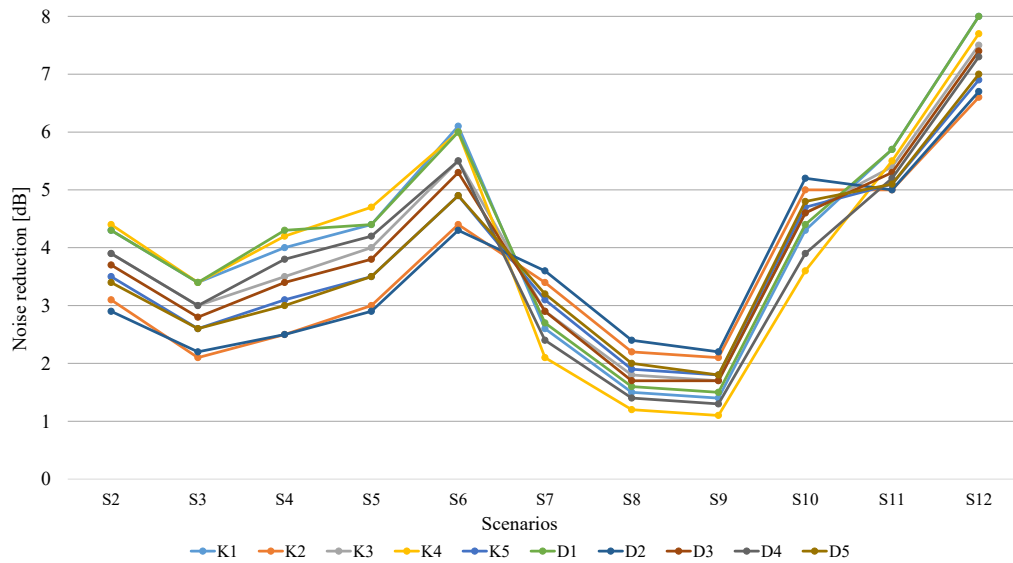


Fig. 4. NR on sound absorption in KX/DXY1F1SXA1 scenarios.

ables, a maximum NR of 8 dB was achieved based on the surface absorption (example of comparative analysis: K1–D5 / difference between S2–S12 and S1 – see Fig. 4).

In NR with the indoor surface absorption, the effect of the ceiling on NR in plants with large areas and volumes is greater than that of the wall (K4–D4 scenarios). In the total absorption, the use of materials with the same weighted sound absorption coefficient in the ceiling (1600 m²) and walls (total of 2240 m²) was investigated separately. In the analysis, based on the medium absorption (α_w : 0.5) in the S2–S7 scenarios, absorption values of 800 m² Sabine for the ceiling, and 1120 m² Sabine for the walls were created separately.

As a result of the analysis, it was concluded that the ceiling is more effective in NR than the walls. It was determined that the difference in NR values in the ceiling and walls was between 1.5 dB–2.5 dB (example of comparative analysis: K4/D4Y2F2SXA1 / difference between S2–S12 and S1 – see Fig. 5). While less sound absorptive material was used in the ceiling than in the walls, ceilings were more effective in total NR. In contrast to this situation, in plants with small areas and volumes, the effect of walls in NR is more effective than the ceiling. While fewer sound-absorptive materials were used on the walls than on the ceiling, the walls were more effective in total NR. With the increase in volume, the distances between the sound

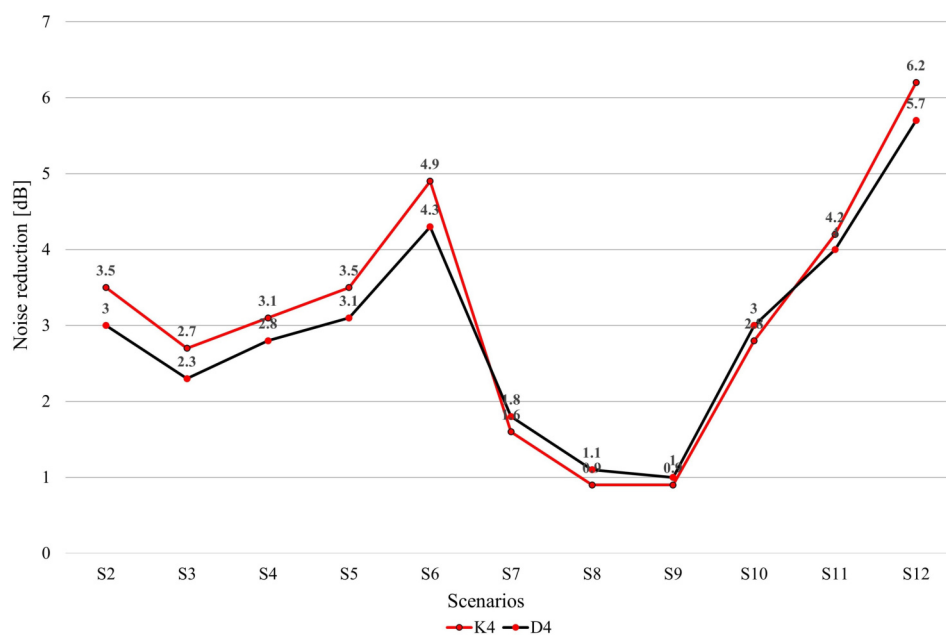


Fig. 5. Effect of ceiling-walls on NR in K4/D4Y2F2SXA1 scenarios.

source and the building components affect the distribution of sound pressure levels and NR. This analysis limits the use of the NR equation based on the sound absorption (Eq. 17):

$$\text{NR} = 10 \log \frac{A_2}{A_1}, \quad (17)$$

where NR is the noise reduction in the room [dB], A_2 is the total volume absorption after improvements (Sabine), and A_1 is the total absorption before improvements (Sabine).

The height as architectural properties in the scenario plants is decisive for the indoor acoustic environment. As the height increases in the plants, indoor LAs decrease. The increase in height allows the sound waves to propagate in a larger area and volume, which leads to a decrease in the sound energy reaching the receiver. Increasing the ceiling height within the scope of acoustic improvements gives effective results in NR. In the scenarios examined (scenarios with medium areas and scenarios with large areas), it was found that the ceiling was more effective than the walls in NR (example of comparative analysis: K1–K3–K5 scenarios / difference between S2–S12 and S1 – see Fig. 6).

In the scenarios examined (scenarios with medium areas and scenarios with large areas), it was found that the ceiling was more effective than the walls in NR. Additionally, the use of canopy absorbers in the ceiling (S4 scenarios) provides the best performance in NR (S3–S5 scenarios). Moreover, the effect of different positioning of sound absorptive materials used in the walls in textile industrial plants (lower-upper section) on LAs and NR was found to be very low.

In the scenario plants, RT analyses were performed at medium frequencies (500 Hz and 1000 Hz). The RT as a property of the interior space does not depend on the sources (more precisely, the sources have a minimal impact due to their sound absorption and as acoustic barriers). In the RT analyses, the live room (S1 – high sound reflection), the scenario with medium absorption of ceiling and walls (S11), and the dead room (S12 – high sound absorption) were evaluated (Table 4). Very high RTs (in the range of 3 s–6 s) were detected in the live room scenarios. In scenarios where the ceiling and wall planes were designed with medium absorption (α_w : 0.5), RTs were calculated at 0.5 s–2 s levels. Low RTs (0.5 s–1 s) were found in dead room scenarios (Fig. 7). The high RTs lead to an increase in sound pressure levels in the plants.

As a result of MLR analyses, the area and height of the plant, the number of machines, and the weighted sound absorption coefficients of the walls and ceilings were effective in determining the indoor LAs. Additionally, the short edge length of the plant was also effective in determining the LA in rectangular plants. The height of the plant and the average wall and ceiling weighted sound absorption coefficients were effective in NR. In RTs, the depth and height of the plant and the average wall and ceiling weighted sound absorption coefficients were effective. The length of the plant refers to the length of one edge in square-planned plants, while it refers to the length of the short edge in rectangular-planned plants. Acoustic improvement prediction models and acoustic simulations were comparatively tested in the case study, and the prediction model was found to be successful.

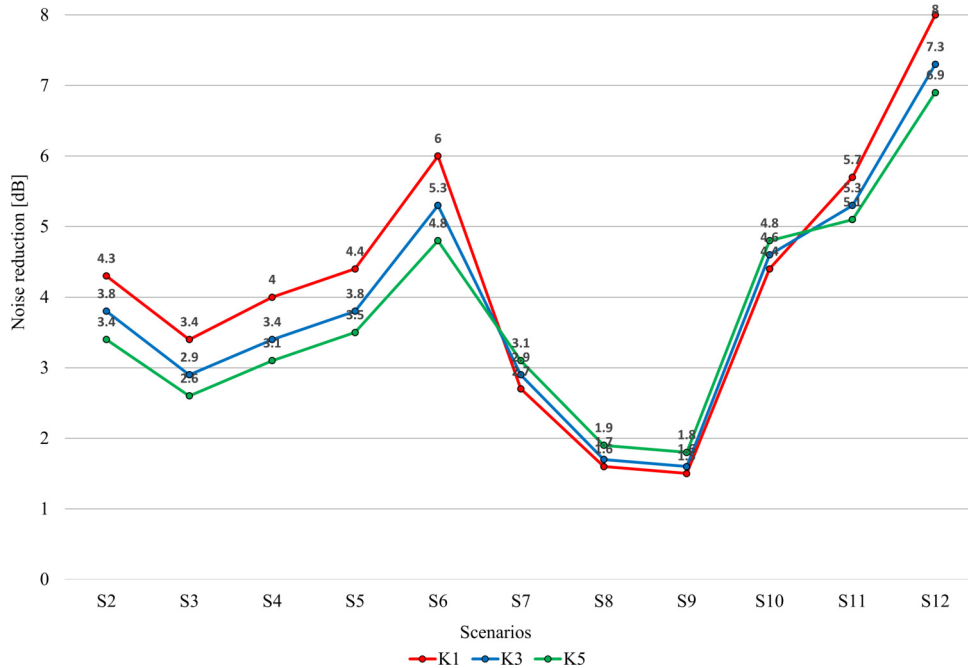


Fig. 6. Effect of height on NR in K1/K3/K5/Y2F2SXA1 scenarios.

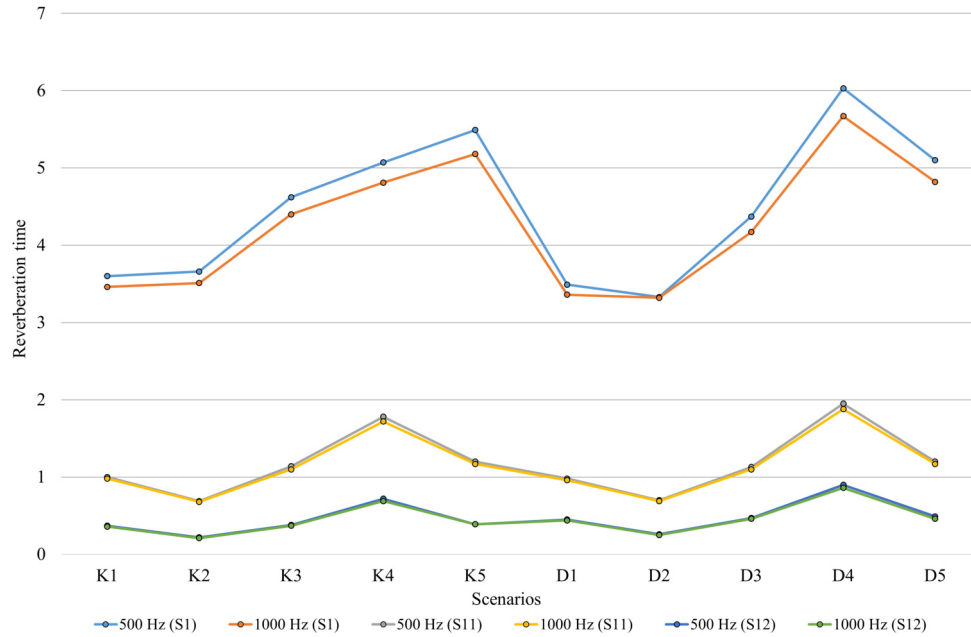


Fig. 7. RT analysis on KX/DXY1F1SXA1 scenarios.

5. Conclusion

This study is part of a wide research involving acoustic improvements for the reduction of high noise levels in textile industry plants. To develop this aim, it should be supported by different noise control mechanisms and detailed with the textile machine design. The study was carried out with scenario plants located in the Republic of Türkiye and determined in the literature review. Different scenario plants created depending on the architecture properties and source-transmission-receiver characteristics were analyzed in the ODEON Auditorium, and LAs, NR, and RTs were analyzed.

In the scenario plant analysis, it has been determined that the plant geometry does not affect *A*-weighted sound pressure levels and NR. Depending on the number of machines per place in the plants, the larger the plant, the more cumulative sound sources, and the higher the indoor sound pressure level. For this purpose, it is necessary to divide the plants into small parts and make zoning. In the acoustic analysis of the plants, a NR of up to 8 dB was achieved by using the surface absorbers. However, the wall and ceiling effectiveness of NR varies. In NR, the ceiling is effective in spaces with a large plan and volume, while the walls are effective in spaces with a relatively small plan and volume. However, as the height increases in the main production area, the decreases are seen in the LA, and as an effect of this situation, effective results are obtained in NR. The lower and upper positioning of the sound absorptive materials used in the walls (facade lighting and ventilation requirement) do not have a decisive variable for the indoor acoustic environment. It is

important to control the RT depending on the surface absorption in textile industry plants. However, it was not found appropriate to be evaluated as an acoustic parameter in industrial plants. As a result of the regression analysis, calculation equations were created to predict the LA, NR, and RT at 500 Hz (dependent variables). The prediction model has been comparatively tested with the application of acoustic simulations and calculations over the case study, and its reliability and validity have been provided. In the model, LAs, NR, and RTS can be estimated with the improvements made in textile industry plants and optimum acoustic comfort conditions are created for employees.

This paper represents a starting point for several future works. It would be appropriate to take noise control precautions for machine designs that are not included in the research, construct vibration isolation, and detail the noise control precautions that can be taken at the design phase in future studies to develop the topic. Moreover, preferring different room acoustics simulation programs, using optimization methods and information technologies, and developing methodological tools based on machine learning will also contribute to the research topic.

References

1. ABBASI M. *et al.* (2020), Noise exposure and job stress – A structural equation model in textile industries, *Archives of Acoustics*, **45**(4): 601–611, <https://doi.org/10.24425/aoa.2020.135248>.
2. AL-DOSKY B.H.M. (2014), Noise level and annoyance of industrial factories in Duhok City, *IOSR Journal of Environmental Science, Toxicology and Food Tech-*

- nology, **8**(5): 01–08, <https://doi.org/10.9790/2402-08510108>.
3. ALI S.A. (2011), Industrial noise level and annoyance in Egypt, *Applied Acoustics*, **72**(4): 221–225, <https://doi.org/10.1016/j.apacoust.2010.11.001>.
 4. ARENAS J.P., SUTER A.H. (2014), Comparison of occupational noise legislation in the Americas: An overview and analysis, *Noise & Health*, **14**(72): 306–319, <https://doi.org/10.4103/1463-1741.140511>.
 5. ATMACA E., PEKER I., ALTIN A. (2005), Industrial noise and its effects on humans, *Polish Journal of Environmental Studies*, **14**(6): 721–726.
 6. BAFFOE P.E., DUKER A.A. (2018), Multiple linear regression approach to predicting noise pollution levels and their spatial patterns for the Tarkwa Mining Community of Ghana, *American Journal of Engineering Research*, **7**(7): 104–112.
 7. BAKER D. (2015), Application of noise guidance to the assessment of industrial noise with character on residential dwellings in the UK, *Applied Acoustics*, **93**: 88–96, <https://doi.org/10.1016/j.apacoust.2015.01.018>.
 8. BISTABA S.R., BRADLEY J.S. (2000), Predicting reverberation times in a simulated classroom, *Journal of the Acoustical Society of America*, **108**: 1721–1731, <https://doi.org/10.1121/1.1310191>.
 9. CHATILLON J. (2007), Influence of source directivity on noise levels in industrial halls: Simulation and experiments, *Applied Acoustics*, **68**(6): 682–698, <https://doi.org/10.1016/j.apacoust.2006.07.010>.
 10. DURÁN DEL AMOR M.d.D., CARACENA A.B., LLORENS M., ESQUEMBRE F. (2022), Tools for evaluation and prediction of industrial noise sources. Application to a wastewater treatment plant, *Journal of Environmental Management*, **319**: 115725, <https://doi.org/10.1016/j.jenvman.2022.115725>.
 11. EJIGU M.A. (2019), Excessive sound noise risk assessment in textile mills of an Ethiopian-Kombolcha textile industry share company, *International Journal of Research in Industrial Engineering*, **8**(2): 105–114, <https://doi.org/10.22105/riej.2019.169138.1071>.
 12. FICHERA I. (2020), The accuracy of reverberation time prediction for general teaching spaces, [in:] *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, pp. 1738–1748.
 13. FREDRIKSSON S., HAMMAR O., TORÉN K., TENENBAUM A., WAYE K.P. (2015), The effect of occupational noise exposure on tinnitus and soundinduced auditory fatigue among obstetrics personnel: A cross-sectional study, *BMJ Open*, **5**(3): e005793, <https://doi.org/10.1136/bmjopen-2014-005793>.
 14. ILGÜREL N. (2013), Effectiveness of the total absorption on noise reduction in industrial plants, *Noise Control Engineering Journal*, **61**(1): 11–25, <https://doi.org/10.3397/1.3702002>.
 15. JAYAWARDANA T.S.S., PERERA M.Y.A., WIJESENA G.H.D. (2014), Analysis and control of noise in a textile factory, *International Journal of Scientific and Research Publications*, **4**(12).
 16. JOB R.F.S. (1996), The influence of subjective reactions to noise on health effects of the noise, *Environment International*, **22**(1): 93–104, [https://doi.org/10.1016/0160-4120\(95\)00107-7](https://doi.org/10.1016/0160-4120(95)00107-7).
 17. KUMAR V., KUMAR S. (2016), A regression model of traffic noise intensity in metropolitan city using artificial neural networks, *International Journal of Research and Engineering*, **3**(12): 27–30.
 18. KURRA S. (2020), *Environmental Noise and Management: Overview From Past to Present*, Hoboken, NJ: John Wiley & Sons, Ltd.
 19. LEATHER P., BEALE D., SULLIVAN L. (2003), Noise, psychosocial stress and their interaction in the workplace, *Journal of Environmental Psychology*, **23**(2): 213–222, [https://doi.org/10.1016/S0272-4944\(02\)00082-8](https://doi.org/10.1016/S0272-4944(02)00082-8).
 20. MASULLO M., TOMA R.A., MAFFEI L. (2022), Effects of industrial noise on physiological responses, *Acoustics*, **4**: 733–745, <https://doi.org/10.3390/acoustics4030044>.
 21. MCINTOSH A.M., SHARPE M., LAWRIE S.M. (2010), Research methods, statistics and evidence-based practice, [in:] *Companion to Psychiatric Studies*, Johnstone E.C., Owens Cunningham D., Lawrie S.M., McIntosh A.M., Sharpe M.D. [Eds.], 8th ed., pp. 157–198, Elsevier Churchill Livingstone.
 22. MONAZZAM M.R., NEZAFAT A. (2007), On the application of partial barriers for spinning machine noise control: A theoretical and experimental model, *Iranian Journal of Environmental Health Science and Engineering*, **4**(2): 113–120.
 23. MONAZZAM-ESMAEELPOUR M.R., HASHEMI Z., GOLMOHAMMADI R., ZAREGAR N. (2014), A passive noise control approach utilizing air gaps with fibrous materials in the textile industry, *Journal of Research in Health Sciences*, **14**(1): 46–51.
 24. NOWOŚWIAT A. (2023), Determination of the reverberation time using the measurement of sound decay curves, *Applied Sciences*, **13**: 8607, <https://doi.org/10.3390/app13158607>.
 25. Occupational Safety and Health Administration (1995), *1910.95 – Occupational noise exposure*, <https://www.osha.gov/laws-regs/regulations/standard-number/1910/1910.95>.
 26. PROBST F. (2012), Prediction of sound pressure levels at workplaces, [in:] *Acoustics 2012*, pp. 23–27.
 27. REINHOLD K., TINT P. (2009), Hazard profile in manufacturing: determination of risk levels towards enhancing the workplace safety, *Journal of Environmental Engineering and Landscape Management*, **17**(22), 69–80, <https://doi.org/10.3846/1648-6897.2009.17.69-80>.

28. SHAHID A., JAMALI T., KADIR M.M. (2018), Noise induced hearing loss among an occupational group of textile employees in Karachi, Pakistan, *Occupational Medicine & Health Affairs*, **6**(4): 282, <https://doi.org/10.4172/2329-6879.1000282>.
29. SHAKHATREH F.M., ABDUL-BAQI K.J., TURK M.M. (2000), Hearing loss in a textile factory, *Saudi Medical Journal*, **21**: 58–60.
30. TANG X., KONG D., YAN X. (2018), Multiple regression analysis of a woven fabric sound absorber, *Textile Research Journal*, **89**(5): 855–866, <https://doi.org/10.1177/0040517518758001>.
31. The European Parliament and the Council of the European Union (2003), Directive 2003/10/EC of the European Parliament and of the Council of 6 February 2003 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise), Official Journal of the European Union.
32. THEMANN C.L., MASTERSON E.A. (2019), Occupational Noise exposure: A review of its effects, epidemiology, and impact with recommendations for reducing its burden, *Journal of the Acoustical Society of America*, **146**(5): 3879–3905, <https://doi.org/10.1121/1.5134465>.
33. YAHYA M.N., OTSURU T., TOMIKU R., OKOZONO T. (2010), Investigation the capability of neural network in predicting reverberation time on classroom, *International Journal of Sustainable Construction Engineering and Technology*, **1**(1): 1–13.
34. YAMAN TURAN N., ÖNEY O. (2021), Investigation of the noise exposure in weaving workplaces in Western Turkey, *Textile and Apparel*, **31**(1): 27–33, <https://doi.org/10.32710/tekstilvekonfeksiyon.762195>.
35. YANG M. (2019), A review of regression analysis methods: Establishing the quantitative relationships between subjective soundscape assessment and multiple factors, [in:] *Proceedings of the 23rd International Congress on Acoustics*, <https://doi.org/10.18154/RWTH-CONV-239497>.
36. ZAW A.K. *et al.* (2020), Assessment of noise exposure and hearing loss among employees in textile mill (Thamine), Myanmar: A cross-sectional study, *Safety and Health at Work*, **11**(2): 199–206, <https://doi.org/10.1016/j.shaw.2020.04.002>.

Research Paper

Experimental Characterization of Sound Absorption for Composite Panel Made of Perforated Plate and Membrane Foam Layer

Van-Hai TRINH⁽¹⁾, Mu HE^{(2)*}

⁽¹⁾ *Institute of Vehicle and Energy Engineering, Le Quy Don Technical University*
Hoang Quoc Viet, Hanoi, Vietnam

⁽²⁾ *School of Mechanical Science and Engineering, Huazhong University of Science and Technology*
Wuhan, Hubei, China

*Corresponding Author e-mail: mu.he@foxmail.com

(received May 14, 2024; accepted January 15, 2025; published online March 4, 2025)

A recent key challenge in noise engineering is the development of structures or materials that achieve desirable acoustic performance in practical settings. Combinations of porous layers and perforated plates offer potential composite absorbers for various acoustic applications. The present work conducts experimental characterizations of sound absorption performance of absorbers based on membrane foams combined with perforated plates. Membrane foams with the well-controlled cell size and porosity are fabricated by milli-fluidic tools, whereas perforated plates are made within a tuned perforation ratio. The three-microphone method is used to perform the acoustic measurements. The results obtained from ten combination samples reveal that the sound absorption behavior of the foam-based layers can be successfully tailored and improved by a thin perforated plate within a reasonable hole diameter and spacing while maintaining the total thickness of the composite absorber.

Keywords: membrane foam; monodisperse; perforated plate; composite absorber; sound absorption.



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

1. Introduction

Undesirable or harmful outside sounds, produced primarily by mechanical equipment, daily activities, and industrial processes, have a significant impact on both human and equipment performance. Designing sound-absorbing materials for real-world applications is one of the most frequent issues faced by acoustic engineers (ATTENBOROUGH, VÉR, 2006). Along this path, man-made materials (e.g., cellular foams, fibrous structures, particle-packed media) are showing their great potential for various acoustic applications in civil, automotive and aerospace engineering (ARENAS, CROCKER, 2010). Due to the small size of the interconnected pores in porous media, the sound absorption performance of these materials is governed by the thermal and viscous dissipations occurring inside the pores (ALLARD, ATALLA, 2009). The relationship between the microstructure and the properties of porous

absorbers can be characterized by different approaches (SAGARTZAZU *et al.*, 2008), which can guide the design of the required sound absorption coefficients (SAC).

The most popular models for characterizing sound-absorbing materials fall into three main groups: semi-empirical, semi-phenomenological, and phenomenological ones (SAGARTZAZU *et al.*, 2008; ALLARD, ATALLA, 2009). With the help of analytical, numerical and experimental advances, our understanding of the material behavior is improving. A porous medium with a rigid skeleton is represented by two frequency-independent factors, namely the complex density and complex bulk modulus (known as the equivalent fluid method (ALLARD, ATALLA, 2009)). Based on this powerful framework, the functional properties of the acoustic materials can be well modeled and characterized. The effective macro-scale properties are then numerically determined by finite element analysis using three alternative methods (ZIELIŃSKI *et al.*, 2020):

direct numerical simulations, direct multiscale homogenization, and hybrid multiscale homogenization. In the first framework, acoustic properties can be estimated from the solution of the uncoupled (thermoviscous) linearised Navier–Stokes equations. In the second technique, the macro-scale complex characteristics are defined from dynamic viscous and thermal permeability functions computed directly by a multiscale model (see (GASSER *et al.*, 2005; LEE *et al.*, 2009)), while the third approach allows computing a set of transport properties (i.e., characteristic lengths, permeabilities, tortuosities) to derive the final acoustic absorption, (see (PARK *et al.*, 2017; TRINH *et al.*, 2022b)). From these known macroscopic transports, the above complex factors can also be calculated by the semi-phenomenological models, such as the Johnson–Champoux–Allard–Pride–Lafarge (JCAPL) model, known as the 8-parameter model. Based on the standard tube testing, acoustical and non-acoustical parameters of sound absorbing materials can be determined directly or indirectly (PANNETON, OLYN, 2006; OLYN, PANNETON, 2008; SALISOU, PANNETON, 2010). With the help of advanced computing tools, the development of optimized properties of sound-absorbing materials can now be done through machine learning and artificial intelligence approaches, where the computational cost can be significantly reduced by generating new data from the limited computational or experimental data (ZHANG *et al.*, 2021; TRINH *et al.*, 2022a).

Owing to the high sound absorption, foam-based absorbers have been widely developed based on the theoretical understanding, simulation knowledge or experimental evidence (YANG *et al.*, 2015; PARK *et al.*, 2017; LANGLOIS *et al.*, 2020). For single foam layers at different pore scales (TRINH *et al.*, 2019; LANGLOIS *et al.*, 2020), various local morphologies ranging from open-cell (LANGLOIS *et al.*, 2020; TRINH *et al.*, 2022b) to membrane (TRINH *et al.*, 2019) structures have been designed by either the typical foaming process (PARK *et al.*, 2017; TRINH *et al.*, 2019) or the 3D-printing technique (ZIELIŃSKI *et al.*, 2022). On the other hand, several references have demonstrated a great improvement (e.g., low-frequency or high average absorption) in the absorption capacity by using multi-layer (BOULVERT *et al.*, 2019) or composite (TRINH *et al.*, 2022b; BORELLI, SCHENONE, 2021) absorbers filled with foams or fibers, and perforated plates can be employed as the potential sub-layers in composite absorbers for tuning the overall system response (LIU *et al.*, 2017; DUAN *et al.*, 2019).

The noise control engineering often requires specific acoustic properties for a wide frequency range (BOULVERT *et al.*, 2019) and other functions, e.g., anti-flaming, high strength, and high heat conduction (GASSER *et al.*, 2005; JAFARI *et al.*, 2020; KOSALA, 2024). For this reason, a systematic investigation of

the sound absorption of composite absorbers based on solid foam and perforated plate should be addressed. In this respect, the inner components of the composite absorbers (CA) considered in this study are membrane foams with controlled cell size and some configurations of perforated facings, and the absorption peaks as the quarter-wavelength resonances of the foam layer are modified by the presence of perforated facing.

2. Materials and experiments

Figure 1 illustrates the structure of the composite absorber, which includes a membrane foam layer and a facing perforated plate.

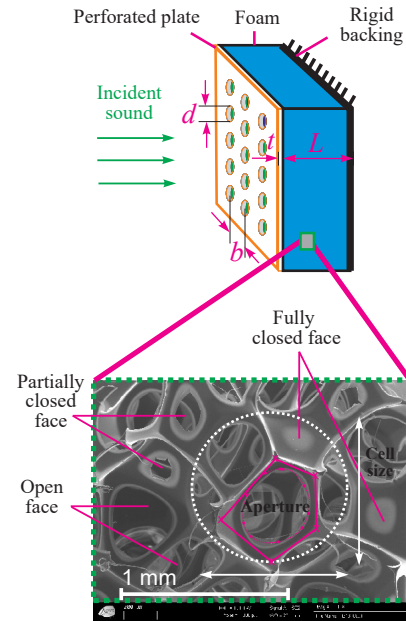


Fig. 1. Illustration of the sound absorber configuration and the foam characteristics.

Table 1. Geometrical parameters of the perforated plates.

Plate	Hole diameter d [mm]	Hole spacing b [mm]	Perforation ratio p [-]
PP1	0.5	4.0	0.012
PP2	1.0	4.0	0.049
PP3	0.5	2.0	0.049
PP4	1.5	4.0	0.110
PP5	1.0	2.0	0.196

The monodisperse foam material is fabricated as follows (TRINH *et al.*, 2019):

- 1) a precursor aqueous foam and a gelatin solution are prepared: the precursor foam within a controlled bubble size $\sim 810 (\pm 30) \mu\text{m}$ and a constant liquid fraction of 0.99 is generated in a glass column by tuning the flow rates of nitrogen and foaming liquid (i.e., Tetradecyltrimethylammonium bromide (TTAB) at 3 g/L). On the other

hand, the aqueous gelatin solution, within a tuned mass concentration from 12 % to 18 %, is prepared and maintained at $T \sim 60^\circ\text{C}$ (above the sol-gel transition temperature $\sim 30^\circ\text{C}$);

- 2) then, the precursor foam is mixed with the hot gelatin solution, and their flow rates are adjusted to get the gas fraction of 0.8. The foaming mixture is filled into a 40 mm-diameter cylindrical cell with a length of 40 mm. To avoid gravity effects during the decreasing temperature process, the material cell is rotated (~ 50 rpm) around its axis;
- 3) the cell is stored in a climatic chamber for one hour at 5°C then one week at $T = 20^\circ\text{C}$ and $\text{RH} = 30\%$ for water evaporating. Finally, after unmolding, a 20 mm-thick specimen for acoustic tests is cut from the central region along the cell axis.

The density and the air flow resistance of the membrane foam samples are defined as follows. For the density, with the specific gravity of the dried gelatin measured to be $g_g = 1.36$, the density of the cut foam sample (diameter – $D = 40$ mm, and thickness – $L = 20$ mm) was calculated from the sample mass m_s as $\rho_s = 10^{-6}m_s/V_s$ with $V_s = \frac{\pi D^2}{4}L$. This gives the density and the open porosity of the foam samples as $\rho_s = 27.1 (\pm 2.3) \text{ kg/m}^3$ and $\phi = 0.98 (\pm 0.003)$. The air flow resistance can be estimated from the air flow resistivity σ of the sample through the formula $R_s = \sigma L$. For the foam with low air flow resistivity, we have $\sigma = A\Delta P/Q$, where A is the sample cross-section area and ΔP is the measured pressure drop, and Q is the air flow rate. For the sample with high air flow resistivity, the value of σ can be inversely characterized as $\sigma = \lim_{\omega \rightarrow 0} [\mathcal{I}(\omega \rho_{eq})]$, where ρ_{eq} is the effective density measured from the impedance tube test (see (PANNETON, OLNy, 2006)). Among the two test foam samples, only sample *F1* allows for direct measurement of resistivity $\sigma = 10700 \text{ Nsm}^{-4}$ which is very similar to the characterized value shown in Table 2.

The cell size of the final foam layers measured from SEM images is $810 \mu\text{m}$ within the monodisperse structure. The membranes range from open to closed cells depending on the gelatin concentration used. Two specimens are selected, namely *F1* and *F2*, within a moderate membrane fraction. As illustrated in the bottom part of Fig. 1, morphological characterizations can be undertaken to measure the membrane fractions of fully open or fully closed faces and the ratio of closure membrane (i.e., aperture/face area) in partially open ones, a detailed description of the

foam characterization can be found in (TRINH *et al.*, 2019).

Five stainless steel perforated plates (PP) with different configurations are manufactured. The geometrical parameters of the PP (Fig. 1, see the top part) are detailed in Table 1. For a square array, the perforation ratio is given as $p = \pi d^2/(4b^2)$. These plates have a thickness of $t = 1$ mm and a diameter of 40 mm. It should be noted that this diameter matches the size of the cut cylindrical foam samples to fit the inner diameter of the impedance tube for acoustic experiments.

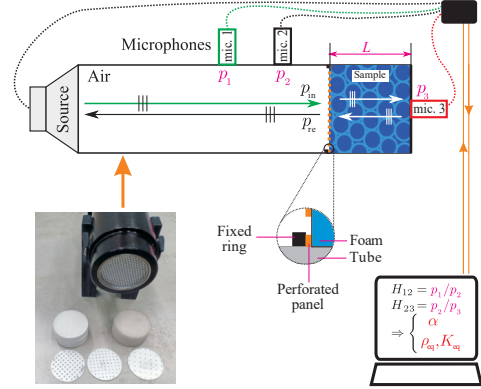


Fig. 2. Experimental setup of three-microphone impedance tube (length – 1 m, diameter – 40 mm).

Acoustic properties are measured using a three-microphone impedance tube (SALISSOU, PANNETON, 2010) in the frequency range of $f \in [4, 4500] \text{ Hz}$ with a step of 4 Hz; see Fig. 2. Note that the perforated plate is placed adjacent to the foam layer without any bonding layers or membranes. Here, a steel ring (with an internal diameter of 39 mm and a square wire size of 1 mm) is used to hold the two material layers in the horizontal test tube. The SAC at normal incidence α is measured through the pressure transfer function H_{12} between microphones 1 and 2. Another function H_{23} between microphones 2 and 3 is used for direct evaluations of the equivalent properties (i.e., density ρ_{eq} and bulk modulus K_{eq}) and inverse estimations of the macroscopic transports (i.e., thermal characteristic length Λ' , viscous characteristic length Λ , static air flow resistivity σ , thermal permeability k'_0 , and tortuosity α_∞ (PANNETON, OLNy, 2006; OLNy, PANNETON, 2008)). Based on the data obtained from the impedance tube experiment, the characterized transport properties of the two foam samples are estimated, see Table 2. In the next section, the results of the absorption properties of the foam layers and the composite absorbers are evaluated.

Table 2. Characterized macroscopic transport parameters of the foam samples.

Sample	$\Lambda' [\mu\text{m}]$	$\Lambda [\mu\text{m}]$	$\sigma [\text{Nsm}^{-4}]$	$k'_0 [\times 10^{-10} \text{ m}^2]$	$\alpha_\infty [-]$
<i>F1</i>	220 (± 36)	73 (± 14)	11560 (± 750)	109 (± 25)	2.48 (± 0.26)
<i>F2</i>	180 (± 30)	55 (± 8)	17500 (± 1200)	93 (± 19)	4.05 (± 0.33)

3. Results and discussion

The SAC curves of the single-layer foams and the perforated plates backed by an air gap of 7 mm are provided in Fig. 3. In terms of the acoustic properties of porous materials with membrane structures, a comparison between the measured data (solid line) the characterized absorption values (dashed line with circle markers) shows a high degree of agreement, as depicted in Fig. 3a. Note that computed sound absorption curves are defined from the semi-phenomenological model (i.e., Johnson–Champoux–Allard–Lafarge model). The results of the transport properties demonstrate consistency with previously obtained results for monodisperse foams with a thin solid membrane (TRINH *et al.*, 2019) as well as foam materials with a high polydispersity of the pore size (NGUYEN *et al.*, 2024). Both foam layers show a quarter-wavelength resonance behavior with $\alpha \sim 1$ at the central frequency of the first peak, \hat{f}_1 (Fig. 3a). The results $\hat{f}_1 = 2132$ Hz ($\Delta f = 1888$ Hz) and $\hat{f}_1 = 1392$ Hz ($\Delta f = 884$ Hz) are, respectively, for $F1$ and $F2$, where Δf is the peak width at $\alpha = 0.8$. It is clear that the high membrane foam $F2$ provides broadband performance (i.e., an absorption peak at lower frequencies) compared with the foam $F1$. However, when the membrane ratio in the foam sample is too high, causing the cell faces to be nearly closed, the absorption capability of the foam layer decreases because airborne waves cannot easily propagate into the foam structure (TRINH *et al.*, 2019). The foam thickness of 20 mm is much smaller than the optimal thickness of granular-packed layers (>100 mm) to

achieve the above peaks, see Eq. (42) in (VIET DUNG *et al.*, 2019). As depicted in Fig. 3b, the effect of perforation ratios on the sound absorption indicates that reducing the ratio p leads to an increase in the absorption, which is consistent with the findings in (LIU *et al.*, 2017), while the plates PP2 and PP3 have the same perforation rate, they have different airflow resistivity (i.e., viscous permeability) and viscous characteristic lengths (due to different hole diameters) and these properties are responsible for different sound absorption performances. It should be noted that to clearly illustrate the absorption characteristics of perforated plates within different perforation ratios in the test frequency range, an air gap of 7 mm was chosen as an example within the range of 2 mm to 8 mm, as used in (LIU *et al.*, 2017).

As shown in Fig. 4, thin perforated plates change significantly the acoustic behavior of the base foams. Herein, the configuration CA ij denotes the combination of the foam F_i with the perforated plate PP j with $i = \{1, 2\}$ and $j = \{1, \dots, 5\}$. The original absorption curves are generally shifted towards lower frequencies by combining PPs within a low ratio p , and the shift distance depends on the original peak or complex wavelength $\lambda_e = \sqrt{K_{eq}/\rho_{eq}}/f$. In detail, the frequency \hat{f}_1 of the foam $F1$ is significantly reduced to 1316 Hz (e.g., CA11 in Fig. 4a), whereas it can be challenging to reduce that of the high-membrane foam $F2$ (i.e., $\hat{f}_1 = 1252$ Hz for CA23, Fig. 4b). In contrast, the use of PPs with a high perforation ratio can improve the absorption capacity of the composite panels

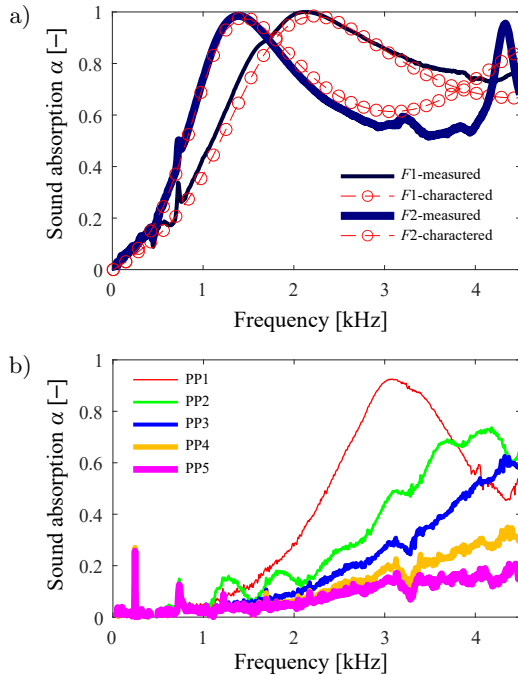


Fig. 3. Normal incidence sound absorption coefficients of (a) foam layers and (b) perforated plates backed by an air gap of 7 mm.

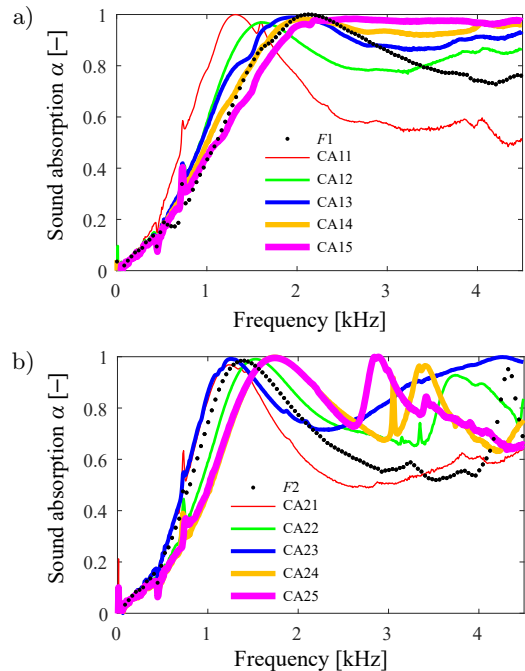


Fig. 4. Normal incidence sound absorption coefficients of composite absorbers based on foam $F1$ (a) and foam $F2$ (b).

in the high frequency range (see CA i 3 to CA i 5 for both foams). These observations confirm the link between the absorption property of CA and the imposed PP structure described in (DUAN *et al.*, 2019). Furthermore, the stably high absorption of CA14 and CA15 (Fig. 4a) behaves like a thick fibrous layer (SOLTANI, NOROUZI, 2020). In terms of modeling the structure studied (i.e., the perforated plate combined with an air layer or a foam layer), the assumption of rigid-frame porous models can be used with the necessary tortuosity correction (ATALLA, SGARD, 2007).

In order to rate the sound absorption performance of the test absorbers, the sound absorption coefficients on a set of $1/3$ octaves from 200 Hz to 2500 Hz are used for evaluation. According to the (ASTM C423-23, 2023), two rating index numbers (i.e., the sound absorption average (SAA) and the noise reduction coefficient (NRC)) are calculated. Noted that the SAA and NRC are, respectively, calculated over the twelve $1/3$ octave bands (from 200 Hz to 2500 Hz) and four frequencies (i.e., 250 Hz, 500 Hz, 1000 Hz, and 2000 Hz), and two rating results are approximately estimated from the field induced by normal incidence. As shown in Table 3, most of the composite absorbers based on foam $F1$ show a clear improvement in the rating index number (i.e., SAA = 0.49 and NRC = 0.50), while only the configuration CA23 shows the same behavior due to the peak occurring at the frequency of 1275 Hz. It can be said that by using a foam layer with a low membrane ratio (i.e., foam $F1$), we can easily shift the peak of the absorption curve to a lower frequency band.

Table 3. Rating of sound absorption of the test samples.

Test absorbers	Rating index	
	SAA [-]	NRC [-]
$F1$	0.41	0.45
CA11	0.49	0.50
CA12	0.44	0.45
CA13	0.45	0.45
CA14	0.43	0.45
CA15	0.41	0.45
$F2$	0.47	0.50
CA21	0.45	0.45
CA22	0.44	0.45
CA23	0.49	0.50
CA24	0.42	0.45
CA25	0.42	0.45

The absorption coefficients are next averaged as $\bar{\alpha} = (1/N) \sum_{i=1}^N \alpha(f_i)$ over N discrete frequencies f_i in [200, 1500] Hz for the low range and [1500, 4000] Hz for the high range (BOULVERT *et al.*, 2019; TRINH *et al.*, 2021). With a test frequency step of 4 Hz, N takes the values of 325 and 625, corresponding to the low-frequency range and high-frequency range, re-

spectively. By lowering the value of \hat{f}_1 , the average sound absorption of CA11 to CA14 (Fig. 5a) shows an improvement in the low frequency range. The absorption $\bar{\alpha}$ of CA11 increases approximately 1.5 times to reach ~ 0.6 (see the highest bar in low group in Fig. 5a), which could be the limit for all composite panels based on the foam $F2$ (low group in Fig. 5b). In high groups, the values $\bar{\alpha}$ averaging from configurations CA i 3 to CA i 5 are 0.934 for $i = 1$ (foam $F1$) and 0.840 for $i = 2$ (foam $F2$); the ratios between the value $\bar{\alpha}$ of the two-layer composite panels and that of the foam layer are calculated as 1.06 and 1.29, respectively. These observations provide quantitative evidence of the absorption performance of composite panels. Based on PP3 with a medium perforation ratio and small holes, both CA i 3 configurations exhibit improved sound absorption over the whole frequency range of interest.

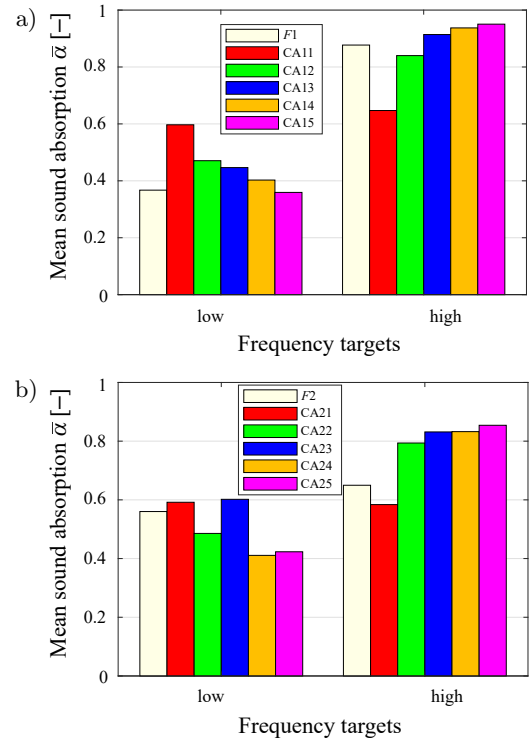


Fig. 5. Bar graphs of the average sound absorption of (a) foam $F1$ + plates PP j and (b) foam $F2$ + plates PP j .

4. Conclusions

In this paper, the sound absorption of foam layers covered by perforated plates has been experimentally characterized. The experimental evidence reveals the effects of the membrane level and the perforation parameters on the local absorption resonances (i.e., modified quarter-wavelength resonances of the foam layer within the influence of the facing perforated plate). The absorption behavior of a given foam material can be effectively tailored to the desired performance by adding appropriate perforated facings. Perforated

plates with a low perforation ratio are advantageous for low-frequency sound absorption applications and vice versa. In addition, good sound absorption over the full frequency range can be achieved by using composite layers with a fixed thickness of about ~20 mm. Based on the present framework, further works can be designed for the systematic characterization of composite absorbers developed for real applications.

Acknowledgments

The authors wish to thank C. Perrot, V. Langlois, and O. Pitois from the Gustave Eiffel University for the foam materials and experiment device.

References

- ALLARD J.F., ATALLA N. (2009), *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials*, 2nd ed., JohnWiley & Sons.
- ARENAS J.P., CROCKER M.J. (2010), Recent trends in porous sound-absorbing materials, *Sound & Vibration*, **44**(7): 12–18.
- ASTM C423-23 (2023), *Standard test method for sound absorption and sound absorption coefficients by the reverberation room method*, ASTM International, <https://doi.org/10.1520/C0423-22>.
- ATALLA N., SGARD F. (2007), Modeling of perforated plates and screens using rigid frame porous models, *Journal of Sound and Vibration*, **303**(1–2): 195–208, <https://doi.org/10.1016/j.jsv.2007.01.012>.
- ATTENBOROUGH K., VÉR I.L. (2006), Sound-absorbing materials and sound absorbers, [in:] *Noise and Vibration Control Engineering: Principles and Applications*, VÉR I.L., VERANEK L.L. [Eds.], 2nd ed., John Wiley & Sons, <https://doi.org/10.1002/9780470172568.ch8>.
- BORELLI D., SCHENONE C. (2021), On the acoustic transparency of perforated metal plates facing a porous fibrous material, *Noise Mapping*, **8**(1): 185–203, <https://doi.org/10.1515/noise-2021-0014>.
- BOULVERT J. *et al.* (2019), Optimally graded porous material for broadband perfect absorption of sound, *Journal of Applied Physics*, **126**(17): 175101, <https://doi.org/10.1063/1.5119715>.
- DUAN H., SHEN X., YANG F., BAI P., LOU X., LI Z. (2019), Parameter optimization for composite structures of microperforated panel and porous metal for optimal sound absorption performance, *Applied Sciences*, **9**(22): 4798, <https://doi.org/10.3390/app9224798>.
- GASSER S., PAUN F., BRÉCHET Y. (2005), Absorptive properties of rigid porous media: Application to face centered cubic sphere packing, *The Journal of the Acoustical Society of America*, **117**(4): 2090–2099, <https://doi.org/10.1121/1.1863052>.
- JAFARI M.J., KHAVANIN A., EBADZADEH T., FAZ-LALI M., SHARAK M.N., MADVARI R.F. (2020), Optimization of the morphological parameters of a metal foam for the highest sound absorption coefficient using local search algorithm, *Archives of Acoustics*, **45**(3): 487–497, <https://doi.org/10.24425/aoa.2020.134066>.
- KOSALA K. (2024), Modelling the acoustic properties of baffles made of porous and fibrous materials, *Archives of Acoustics*, **49**(3): 345–357, <https://doi.org/10.24425/aoa.2024.148792>.
- LANGLOIS V., KADDAMI A., PITOIS O., PERROT C. (2020), Acoustics of monodisperse open-cell foam: An experimental and numerical parametric study, *The Journal of the Acoustical Society of America*, **148**(3): 1767–1778, <https://doi.org/10.1121/10.0001995>.
- LEE C.-Y., LEAMY M.J., NADLER J.H. (2009), Acoustic absorption calculation in irreducible porous media: A unified computational approach, *The Journal of the Acoustical Society of America*, **126**(4): 1862–1870, <https://doi.org/10.1121/1.3205399>.
- LIU Z., ZHAN J., FARD M., DAVY J.L. (2017), Acoustic properties of multilayer sound absorbers with a 3D printed micro-perforated panel, *Applied Acoustics*, **121**: 25–32, <https://doi.org/10.1016/j.apacoust.2017.01.032>.
- NGUYEN C.T., LANGLOIS V., GUILLEMINOT J., DUVAL A., PERROT C. (2024), Effect of pore size polydispersity on the acoustic properties of high-porosity solid foams, *Physics of Fluids*, **36**(4): 047101, <https://doi.org/10.1063/5.0191517>.
- OLNY X., PANNETON R. (2008), Acoustical determination of the parameters governing thermal dissipation in porous media, *The Journal of the Acoustical Society of America*, **123**(2): 814–824, <https://doi.org/10.1121/1.2828066>.
- PANNETON R., OLNY X. (2006), Acoustical determination of the parameters governing viscous dissipation in porous media, *The Journal of the Acoustical Society of America*, **119**(4): 2027–2040, <https://doi.org/10.1121/1.2169923>.
- PARK J.H. *et al.* (2017), Optimization of low frequency sound absorption by cell size control and multiscale poroacoustics modeling, *Journal of Sound and Vibration*, **397**(9): 17–30, <https://doi.org/10.1016/j.jsv.2017.03.004>.
- SAGARTZAZU X., HERVELLA-NIETO L., PAGALDAY J.M. (2008), Review in sound absorbing materials, *Archives of Computational Methods in Engineering*, **15**(3): 311–342, <https://doi.org/10.1007/s11831-008-9022-1>.
- SALISSOU Y., PANNETON R. (2010), Wideband characterization of the complex wave number and characteristic impedance of sound absorbers, *The Journal of the Acoustical Society of America*, **128**(5): 2868–2876, <https://doi.org/10.1121/1.3488307>.

21. SOLTANI P., NOROUZI M. (2020), Prediction of the sound absorption behavior of nonwoven fabrics: Computational study and experimental validation, *Journal of Sound and Vibration*, **485**: 115607, <https://doi.org/10.1016/j.jsv.2020.115607>.
22. TRINH V.H., GUILLEMINOT J., PERROT C. (2021), On the sensitivity of the design of composite sound absorbing structures, *Materials & Design*, **210**: 110058, <https://doi.org/10.1016/j.matdes.2021.110058>.
23. TRINH V.H., GUILLEMINOT J., PERROT C., VU V.D. (2022a), Learning acoustic responses from experiments: A multiscale-informed transfer learning approach, *The Journal of the Acoustical Society of America*, **151**(4): 2587–2601, <https://doi.org/10.1121/10.0010187>.
24. TRINH V.H., LANGLOIS V., GUILLEMINOT J., PERROT C., KHIDAS Y., PITOIS O. (2019), Tuning membrane content of sound absorbing cellular foams: Fabrication, experimental evidence and multiscale numerical simulations, *Materials & Design*, **162**: 345–361, <https://doi.org/10.1016/j.matdes.2018.11.023>.
25. TRINH V.-H., NGUYEN T.-V., NGUYEN T.-H.-N., NGUYEN M.-T. (2022b), Design of sound absorbers based on open-cell foams via microstructure-based modeling, *Archives of Acoustics*, **47**(4): 501–512, <https://doi.org/10.24425/aoa.2022.142894>.
26. VIET DUNG V., PANNETON R., GAGNÉ R. (2019), Prediction of effective properties and sound absorption of random close packings of monodisperse spherical particles: Multiscale approach, *The Journal of the Acoustical Society of America*, **145**(6): 3606–3624, <https://doi.org/10.1121/1.5111753>.
27. YANG X., REN S., WANG W., LIU X., XIN F., LU T. (2015), A simplistic unit cell model for sound absorption of cellular foams with fully/semi-open cells, *Composites Science and Technology*, **118**: 276–283, <https://doi.org/10.1016/j.compscitech.2015.09.009>.
28. ZHANG H., WANG Y., LU K., ZHAO H., YU D., WEN J. (2021), SAP-Net: Deep learning to predict sound absorption performance of metaporous materials, *Materials & Design*, **212**: 110156, <https://doi.org/10.1016/j.matdes.2021.110156>.
29. ZIELIŃSKI T.G. *et al.* (2022), Taking advantage of a 3D printing imperfection in the development of sound-absorbing materials, *Applied Acoustics*, **197**: 108941, <https://doi.org/10.1016/j.apacoust.2022.108941>.
30. ZIELIŃSKI T.G., VENEGAS R., PERROT C., ČERVENKA M., CHEVILLOTTE F., ATTENBOROUGH K. (2020), Benchmarks for microstructure-based modelling of sound absorbing rigid-frame porous media, *Journal of Sound and Vibration*, **483**: 115441, <https://doi.org/10.1016/j.jsv.2020.115441>.

Research Paper

Issues in the Design and Validation of Coupled Reverberation Rooms for Testing Acoustic Insulation of Building Partitions

Agata SZELĄG^{(1)*} , Marcin ZASTAWNIK⁽²⁾ ⁽¹⁾ *Tadeusz Kościuszko Cracow University of Technology*
Kraków, Poland⁽²⁾ *Jan Długosz University in Częstochowa*
Częstochowa, Poland; e-mail: m.zastawnik@ujd.edu.pl*Corresponding Author e-mail: aszelag@pk.edu.pl*(received February 26, 2024; accepted September 24, 2024; published online January 20, 2025)*

The paper presents the characteristics of the sound field in two pairs of coupled reverberation rooms, designed in accordance with International Organization for Standardization [ISO] (2021c). The analyses are based on the results of the following studies. Firstly, the acoustic airborne sound insulation of selected test samples was measured in the reverberation rooms without using any sound diffusing nor sound absorbing elements. In the second step, the tests were repeated successively with an increasing number of diffusers installed in the rooms. The last stage of the research involved measurements with additional absorbers mounted in the rooms. The results show that although the geometry and construction of the reverberation rooms are in line with the standard guidelines, in most situations it was necessary to use diffusing and absorbing elements to improve the acoustic field in the rooms. Such elements, however, are very undesirable as they significantly limit the usable space of the rooms, making it more difficult to assemble samples and distribute sources and measurement points in the measurement space. Later in the article, the authors prove that even using typically available design tools, i.e., 1st and 2nd Bonello criteria, numerical simulations with the image-source method and the finite element method, or more advanced research methods, such as measurements using scaled samples, it seems impossible to prevent at the design stage the future necessity of using additional diffusing and absorbing elements in the reverberation rooms. Only via verification by measurements performed in the completed rooms provides the assessment if such additional elements are required.

Keywords: reverberation chambers; transmission loss; acoustic field; small scale model.



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

1. Introduction

The test bench for measuring the airborne sounds insulation of building partitions consists of two coupled reverberation rooms (according to (ISO, 2021a)). In order to achieve adequate repeatability and reproducibility of measurements (based on (ISO, 2014)), the test stand must follow strict guidelines. These guidelines are applicable to the geometry and construction of the reverberation rooms described in (ISO, 2021c) as well as to the measurement equipment and procedure described in (ISO, 2021b). While the requirements for the selection of appropriate equipment and the imple-

mentation of the correct measurement procedure are precise and unambiguous, the guidelines for the construction and, in particular, the geometry of the reverberation rooms are very general (they relate only to the volume of the rooms and the area of the measurement window). Therefore, coupled reverberation rooms in various laboratories may be constructed differently (for example: (URIS *et al.*, 2007; ZHU, 2022; OLIAZADEH *et al.*, 2022)). As a consequence, the distribution of the sound field in these rooms and its influence on the measurement results will also vary as shown in (DIJCKMANS, VERMEIR, 2013). According to (ISO, 2021c), the sound field in reverberation

rooms should be as diffused as possible. If sufficient sound diffusion is not ensured by the interior geometry alone, additional diffusing elements are required. To quote the standard: “the position and number of diffusing elements should be arranged in such a way that the sound reduction index is not influenced when further diffusion elements are installed” (ISO, 2021c, p. 2). However, at the design stage, it is difficult to accurately model the acoustic field inside the reverberation rooms which was studied by CHAZOT *et al.* (2016), SCHMAL *et al.* (2021) or BORK (2000), let alone its effect on the measurement results. In practice, the qualification procedure is carried out only after the rooms have been constructed. For qualification, sound diffusers (ZHU, 2022; BRADLEY *et al.*, 2014; MLECZKO, WSZOLEK, 2019) as well as sound absorbing elements (FUCHS *et al.*, 2000; YAO *et al.*, 2020) are installed in the rooms to unify the sound field inside. Unfortunately, such measures involve additional costs and also obstruct work in the laboratory until the rooms are adapted for testing. Furthermore, additional sound absorbing and diffusing elements significantly limit the usable space of the rooms, making it more difficult to assemble samples and distribute sources and measurement points in the measurement space.

This paper presents the characteristics of the sound field in two pairs of coupled reverberation rooms, designed following the guidelines and the requirements of (ISO, 2021c). The need for additional design guidelines to achieve satisfactory acoustic field characteristics in reverberation rooms is demonstrated. Such procedures would target the spaces used to measure the acoustic insulation of samples without the need to install any sound diffusing and absorbing elements in their interiors.

2. Subject of study

The research presented in this paper was carried out on two original test benches. The former was an available in the laboratory coupled reverberation rooms made in a small scale (hereinafter named: small reverberation rooms), the latter was a full-size room designed in accordance with the restrictions of the future user (hereinafter named: large reverberation rooms). The first stand, allowed for pilot studies to be carried out on smaller samples. This approach, which is often used in scientific research (BALMORI *et al.*, 2024; DJAMBOVA *et al.*, 2022) was far more economical and quicker to implement at the initial stage of research. The pilot studies were aimed at verifying the adopted research methodology and determining preliminary conclusions regarding the impact of sound diffusing and sound absorbing elements on the acoustic field in exemplary reverberation rooms. At the second stand, the target case was studied, i.e., the acoustic field inside the individually designed full-size rooms.

These results directly reflected reality without the risk of scale influence on the results obtained. A detailed description of these two stands is provided in Subsecs. 2.1 and 2.2, respectively.

2.1. Small reverberation rooms

Tests on small-scale samples were conducted in small, coupled reverberation rooms (see Fig. 1) which replicated the full-size reverberation rooms located at the Department of Mechanics and Vibroacoustics of AGH University of Science and Technology. Quoting SZELAŁ *et al.* (2021), Fig. 2 shows the detailed dimensions of this measurement stand. Both rooms, source and receiving, had a volume of about 0.35 m^3 (which is almost 180 m^3 at 1:1 scale). As described in the aforementioned article, the rooms were effectively vibration-isolated from each other and from the ground. Moreover, due to the fact that for the purposes of acoustic insulation tests there is no need to scale the parameters of the gas filling the rooms, the interiors could remain filled with atmospheric air. During individual measure-

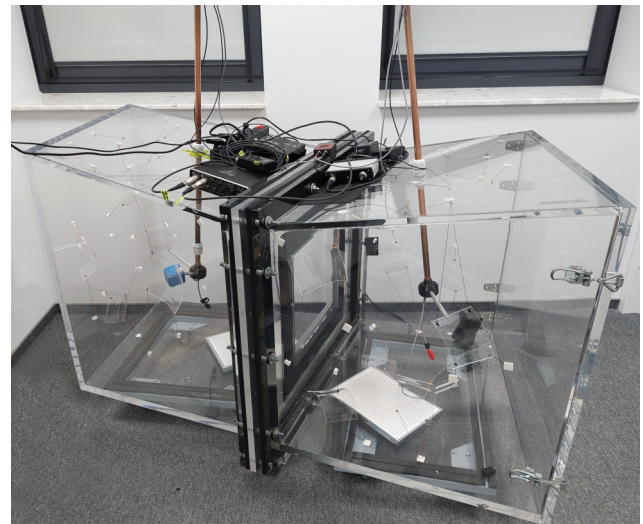


Fig. 1. Small reverberation rooms made of 20 mm-thick plexiglass panels.

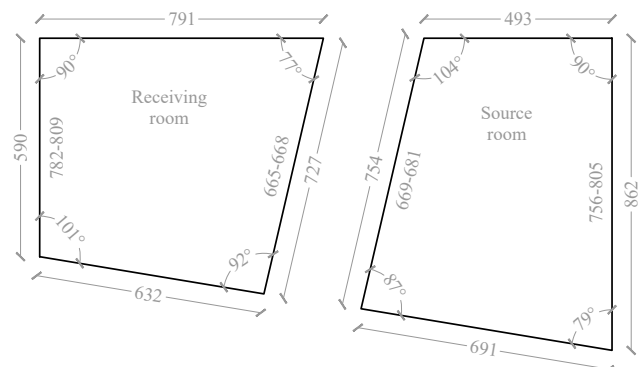


Fig. 2. Dimensions [mm] of the small reverberation rooms; the intervals define the walls heights that vary along the width (SZELAŁ *et al.*, 2021).

ment sessions, only the consistency of air parameters such as pressure, temperature and air humidity was monitored, and finally, based on the results of reverberation time (RT) measurement in receiving room, the influence of acoustic absorption of the interior on measured sound pressure levels (SPLs) was removed.

During the subsequent test stages shown in this paper, in the reverberation rooms, diffusors made of sound-reflecting plexiglass panels were installed. In the source room, eight pieces with dimensions of $100\text{ m} \times 150\text{ m} \times 2\text{ mm}$ and two with dimensions of $150\text{ m} \times 150\text{ m} \times 2\text{ mm}$ were ultimately mounted. In the receiving room, seven pieces with dimensions of $100\text{ m} \times 150\text{ m} \times 2\text{ mm}$ and three with dimensions of $150\text{ m} \times 150\text{ m} \times 2\text{ mm}$ were ultimately mounted. The plexiglass panels were pre-curved to provide better sound diffusion properties. In addition, on the floor of both reverberation rooms, one slotted sound absorbing structure was placed. This absorber was made of 15 mm-thick foam covered with 3 mm-thick aluminum plate with 1 mm-wide slots incised at 10 mm intervals. The overall dimensions of each absorber were $310\text{ m} \times 220\text{ mm}$.

The measurement stand consisted of the following components: two custom made high-frequency sound sources, two $1/4''$ 46BE G.R.A.S. microphone sets, two 12AL G.R.A.S. amplifiers, UMC204HD BEHRINGER U-PHORIA measurement card and a dedicated computer script in the MATLAB environment for processing measurement results (for the detail description of the measurement stand see (SZELAQ *et al.*, 2021)). This article also proves that both the scaled measurement stand, and the measurement methodology meet the requirements of (ISO, 2021a; 2021b; 2021c) adapted to the scale factor as well as that the uncertainty of measurements on the tested stand meets the requirements of (ISO, 2014) for maximum uncertainty values. Therefore, the reliability and repeatability of measurement results obtained on this stand was confirmed.

2.2. Large reverberation rooms

For the full-size tests, two coupled reverberation rooms (see Fig. 3) located in the laboratory Mobilne Laboratorium Techniki Budowlanej Sp. z o.o. in Wałbrzych were used. These rooms were designed and made in accordance with the standard requirements (ISO, 2021b; 2021c), taking into account certain architectural limitations. The detailed dimensions of the source and receiving rooms are shown in Fig. 4. The volumes of the source and receiving rooms were 77 m^3 and 57 m^3 , respectively. The reverberation rooms were constructed of reinforced concrete structure with a wall thickness of 30 cm. The rooms were divided by a reinforced concrete frame with a cross section of $100\text{ cm} \times 100\text{ cm}$. The rooms and the frame were decoupled and vibration-isolated from each

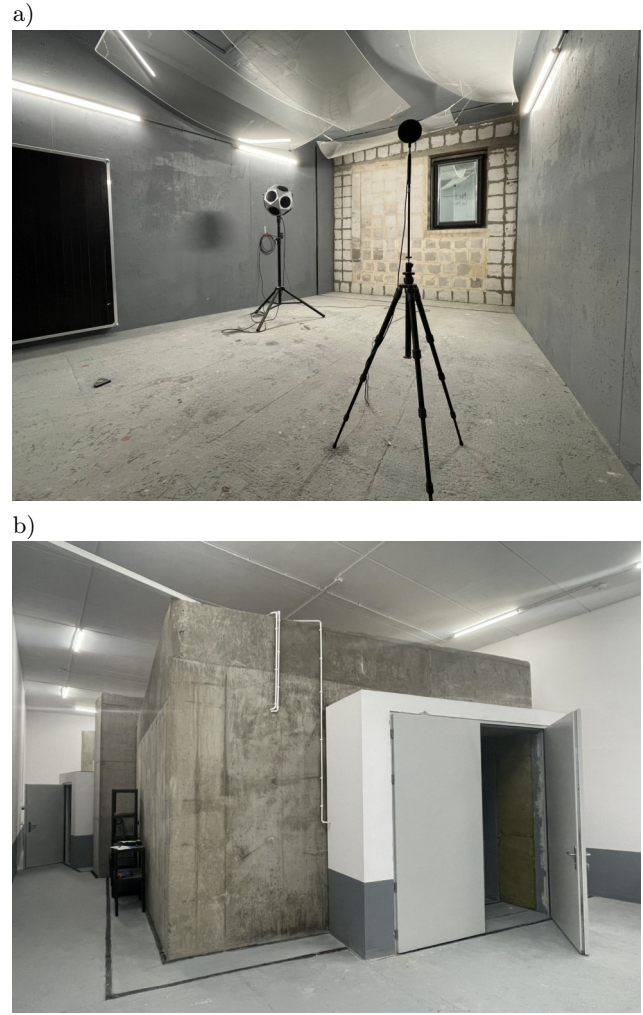


Fig. 3. Exterior view of the large reverberation rooms from the side of the source room (a) and interior view of the source room (b).

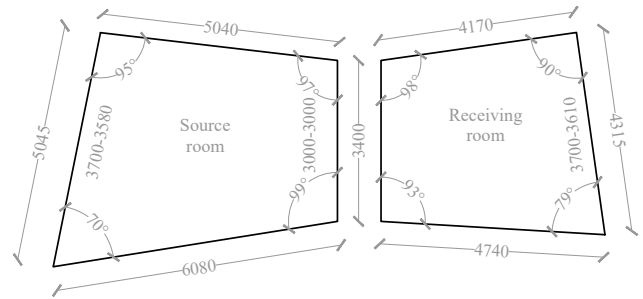


Fig. 4. Dimensions [mm] of the large reverberation rooms; the intervals define the walls heights that vary along the width.

other and from the surroundings. Each room was accessed via a dedicated acoustic sluice equipped with two doors. The acoustic sluice structure was decoupled and vibration-isolated from the rooms and from the surroundings. A single-wing door was fitted in the receiving room, while a double-wing door was used in the source room for technological reasons, i.e., to allow large measurement samples to be brought in.

During the subsequent test stages shown in this paper, in both reverberation rooms, diffusers made of sound-reflecting plexiglass panels were installed. Ultimately, three diffusers of dimensions $3000\text{ mm} \times 1000\text{ mm} \times 6\text{ mm}$ and three of dimensions $2000\text{ mm} \times 1000\text{ mm} \times 6\text{ mm}$ were installed in the source room as well as in the receiving room. The panels suspended from the ceiling and on the walls were bent due to their own weight, resulting in an even better sound diffusion effect. In addition, in both rooms, one slotted sound absorbing structure was mounted on the wall. This absorber was made of 100 mm-thick wool covered with 21 mm-thick board with 4 mm-wide slots incised at 65 mm intervals. The overall dimensions of each absorber were $1850\text{ mm} \times 1050\text{ mm}$.

The measurement stand consisted of the following components: an omni-directional sound source B&K 4292-L-001, a power amplifier B&K 2734, two measurement microphones B&K 4189 together with preamplifiers B&K ZC0032, a two-channel sound analyser B&K 2270A and a computer program for building acoustics B&K 7830.

3. Methodology

The studies presented in this paper were carried out in three stages at each of the measurement stand. In stage 1, the acoustic airborne sound insulation of selected test samples was measured in reverberation rooms, without using any additional sound diffusing nor sound absorbing elements. In the small reverberation rooms, a plexiglass sample with dimensions of $12.5\text{ mm} \times 25.0\text{ mm}$ and a thickness of 1 mm was tested, while in the large reverberation rooms, a door with an area of 2.47 m^2 was tested. The following criteria guided the selection of measurement samples. Firstly, the scale and full-size samples were supposed to have similar dimensions after taking into account their scaling, and this was achieved. Secondly, the samples had to have low sound insulation so that the test results were not dependent on the flanking sound transmission. At this point it is worth noting that it is not important whether the scale sample has a full-size equivalent or the samples tested at both measurement stands are the same. The aim of the research was to determine the acoustic field in the rooms and its impact on the measurement results, and not to verify the insulation of the samples themselves or to check the measurement capabilities and validate the test stands.

Measurements were taken in accordance with the guidelines of (ISO, 2021a). In both source and receiving rooms, the SPL was recorded at ten different measurement points, five for each of the two sound source positions. The averaging time for a single measurement was 15 seconds. The RT was measured in both the source and receiving rooms. For the measurement in small reverberation rooms the impulse response in-

tegration method based on the swept sine signal was used, while in the large rooms the intermittent noise method was adopted. The tests carried out at the subsequent stages followed the same path as in stage 1, except that in stage 2 in the reverberation rooms additional sound diffusing elements were installed in batches, while in stage 3, a sound absorbing structure was placed in each reverberation room. All measured sound insulation indicators were supplemented with measurement uncertainty values U_{95} determined in accordance with ISO (2020a) assuming the measurement situations C (standard uncertainty of measurement repeatability) – appropriate values read from Tables 2 and 3 in the standard (WITTSTOCK, 2015).

4. Measurement results and discussion

4.1. Measurements in small reverberation rooms

Figure 5 shows the results of the acoustic insulation measurement for a sample tested in the small reverberation rooms in four different variants of interior acoustic adaptation, i.e., for different numbers of sound diffusing elements and with or without the sound absorbing structure. The results, after scaling them to actual measurement frequencies (SONIN, 2001) are presented in the full frequency range typical for such tests, i.e., 50 Hz–5000 Hz. The graph also provides information on the values of the sound insulation single-number quantities of the sample, R_w , $R_w + C$, $R_w + C_{50-3150}$, $R_w + C_{tr}$, and $R_w + C_{tr,50-3150}$ calculated according to (ISO, 2020b), for each of the alternatives tested. All indicators presented in Fig. 5 are supplemented with measurement uncertainty values U_{95} determined in accordance with (ISO, 2020a). The conclusions from the analysis of the data contained in Fig. 5 are as follows. After introducing a large number of sound diffusing elements, that is 10 pieces into each reverberation room, a decrease in R -values in the low-frequency bands (50 Hz–125 Hz) can be observed. In addition, the acoustic insulation characteristics in the 160 Hz band evened out after the installation of sound absorbing structures in the rooms. The observed deviations between test results for individual measurement variants are of statistical significance, as in most bands in the indicated frequency range they are higher than standardised values. In the other frequency bands, i.e., from 200 Hz upwards, the diffusing and sound absorbing elements had no significant impact on the sound insulation characteristics. The noticeable decrease in sound insulation in the 3150 Hz–5000 Hz bands for the variant with sound absorbing elements occurred due to a reduction in the SPL in the source room because of the interior damping, and consequently an insufficient separation between the signal and the background noise in the receiving room. The sound insulation values in these frequency bands are

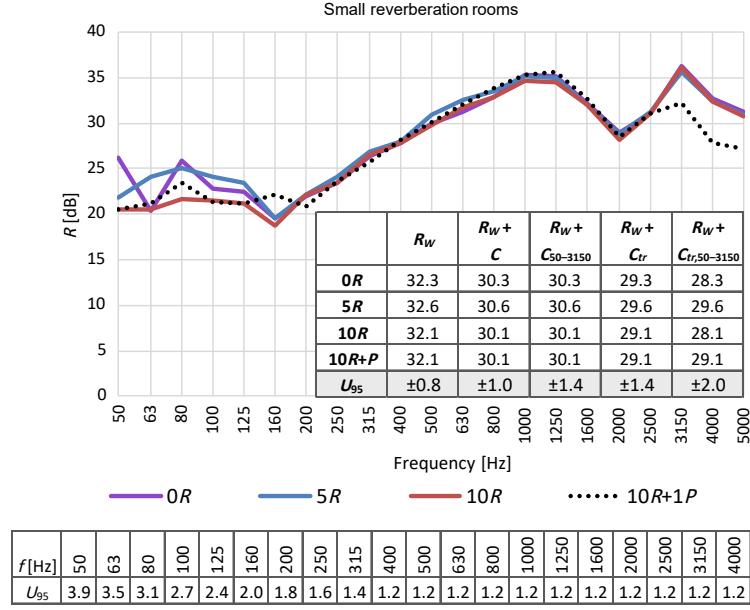


Fig. 5. Acoustic insulation of the sample tested in the small reverberation rooms in four different variants of interior acoustic adaptation: 0R – no diffusing elements in the reverberation rooms; 5R – five diffusing elements in each room; 10R – ten diffusing elements in each room; 10R+1P – ten diffusing elements and one sound absorbing element in each room. The results from the small rooms are scaled to actual measurement frequencies.

therefore underestimated. It is worth mentioning that the variation in the sound insulation values in the low-frequency bands were not strongly reflected in the values of the single-number quantities. Differences in the values of individual indicators are smaller than their measurement uncertainty.

In order to verify the acoustic field in the source and receiving rooms, the SPL spectra in the rooms were plotted in Fig. 6 for all analysed measurement variants. The graphs also show the scatter of the results as a difference of the maximum and minimum SPL obtained in a given frequency band between in-

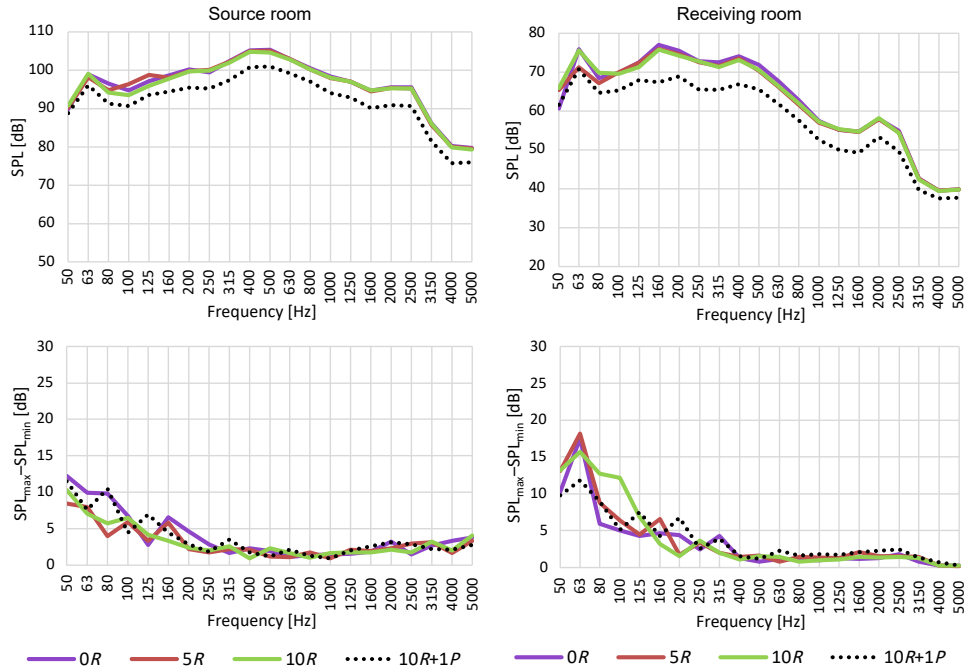


Fig. 6. Equivalent SPL and scatter of the results between individual measurement points in the small source and receiving rooms in four different variants of interior acoustic adaptation: 0R – no diffusing elements in the reverberation rooms; 5R – five diffusing elements in each room; 10R – ten diffusing elements in each room; 10R+1P – ten diffusing elements and one sound absorbing element in each room. The results from the small rooms are scaled to actual measurement frequencies.

dividual measurement points. An analogous comparisons are presented in Fig. 7 for the values of RT in the rooms.

Based on Fig. 6, it can be stated that the equivalent SPLs in both the source and receiving rooms do not differ significantly for interiors with different numbers of diffusing elements. Only the addition of sound absorbing structures reduces the SPL in the rooms, which is obviously due to the partial absorption of sound by such elements. Moreover, no significant trend can be observed in the variation of SPLs values for the different measurement points depending on the number of diffusing and absorbing elements in the rooms. At most, an improved homogeneity of the results in the 63 Hz band may be noticed in the receiving room after installing the sound absorber. This band was characterised previously by the greatest inhomogeneity of the sound field. It can also be added that slightly greater scatter of the results in the low-frequency bands is obtained for the receiving room. However, for the highest frequency bands the results scatter in this room decreases due to the overlap between the signal value and the background sound level generated by the measurement path itself.

Figure 7 shows that with the increasing number of sound diffusing elements and adding a sound absorbing structure, the RT in both source and receiving rooms decreases. In the case of the diffusing elements, it should be noted that this is not a result of sound absorption by this type of elements, as they

were made of sound-reflecting plexiglass. This occurs due to the improved diffusion of the sound field in the rooms, the shortening of the path between reflections and the increase in the number of reflecting planes. Importantly, the use of sound diffusing elements only is not sufficient to achieve the RT recommended by ISO (2021c). Additional sound absorbing elements are required. Such structures installed in the tested rooms made it possible to meet the standard requirements in the basic frequency range of 100 Hz–3150 Hz, except for 100 Hz in the receiving room. In order to meet the standard requirement in the full frequency range (from 50 Hz), it would be necessary to add a low-frequency sound absorbing structure tuned to frequency 80 Hz, for which the measured values of RT are the highest. Based on the plots showing the scatter of the RT values between the individual measurement points, it can be concluded that the increase of the number of sound diffusing elements and addition of a sound absorbing structure reduces this scatter, however, some deviations from this rule are noticeable in selected frequency bands. Nevertheless, the obtained scatter of the results is not high in all measurement cases, which indicates a quite good diffusion of the sound field inside the small reverberation rooms.

In summary, the following conclusions can be drawn from the tests carried out in the small reverberation rooms. The sound fields in terms of spatial uniformity are similar in both reverberation rooms. Even without diffusing and absorbing elements, a quite good

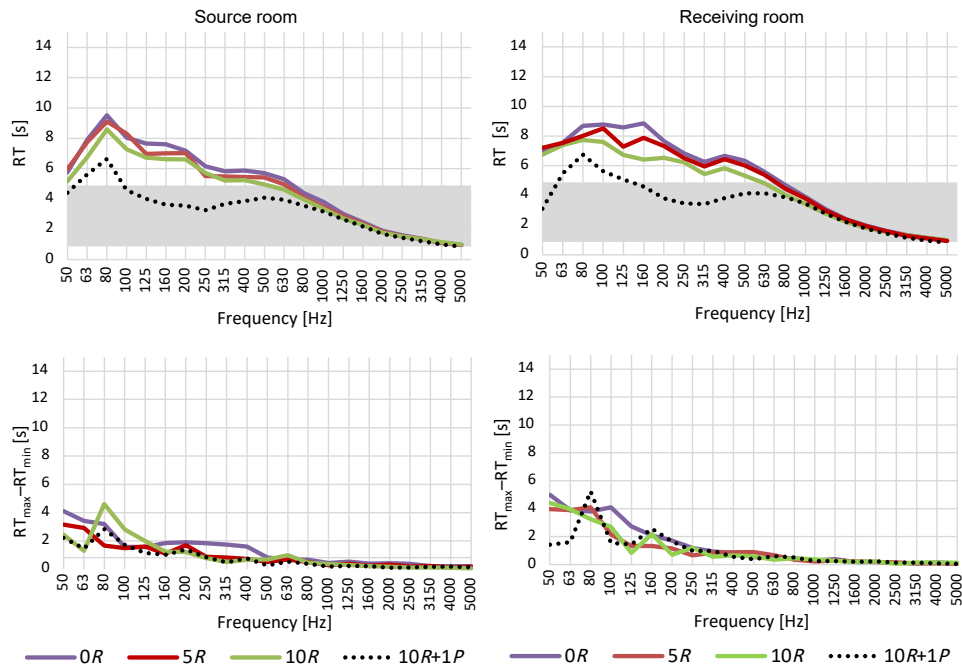


Fig. 7. RT and the scatter of the results between individual measurement points in the small source and receiving rooms in four different variants of interior acoustic adaptation: 0R – no diffusing elements in the reverberation rooms; 5R – five diffusing elements in each room; 10R – ten diffusing elements in each room; 10R+1P – ten diffusing elements and one sound absorbing element in each room. In the RT diagrams, the grey colour indicates the RT ranges recommended by ISO (2021c) standard for the respective room. The results from the small rooms are scaled to actual measurement frequencies.

homogeneity of the results for both the rooms was obtained, i.e., the scatter in SPLs and values of RT in individual measurement points did not deviate from typical values obtained in other laboratories (compared to (NUTTER *et al.*, 2007) and (VALLIS *et al.*, 2015)). Nevertheless, in order to achieve the recommended RT in the rooms, it was necessary to add sound diffusing and sound absorbing elements. However, the results presented in Fig. 5 show that the use of sound diffusing elements in the context of the correct value of the sample sound insulation was necessary only in the low-frequency bands. Further reduction of the RT to the recommended values by installing the absorber had no effect on the sound insulation value of the samples. In conclusion, the analysed small reverberation rooms are characterised by a quite good spatial homogeneity of the sound field, nonetheless they require the use of additional diffusers in order to obtain the correct sound insulation values.

4.2. Measurements in large reverberation rooms

Figure 8 shows the results of the acoustic insulation measurement for a sample tested in the large reverberation rooms in five different variants of interior acoustic adaptation, i.e., for different numbers of sound diffusing elements and with or without the sound absorbing structure. The graph also provides information on the values of the sound insulation single-number quantities of the sample, R_w , $R_w + C$, $R_w + C_{50-3150}$, $R_w + C_{tr}$, and $R_w + C_{tr,50-3150}$ calculated according to (ISO, 2020b),

for each of the alternatives tested. All indicators presented in Fig. 8 are supplemented with measurement uncertainty values U_{95} determined in accordance with ISO (2020a). The conclusions from the analysis of the data contained in Fig. 8 are as follows: a decreasing trend of R -values in the low-frequency bands (50 Hz–315 Hz) can be observed with more sound diffusing elements being introduced into the rooms. The observed deviations between test results for individual measurement variants are of statistical significance, as in all bands in the indicated frequency range they are higher than standardised values. In the other frequency bands, i.e., above 315 Hz, sound diffusing and sound absorbing elements had no significant effect on the sound insulation characteristics of the sample. It is worth mentioning that the variation in the sound insulation values in the low-frequency bands were also reflected in the values of the single-number quantities. Differences in the values of most indicators, only except R_w , are larger than their measurement uncertainty, so they are of statistical significance.

Similarly, as for the case of the small reverberation rooms, in order to verify the acoustic field in the source and receiving rooms, the SPL spectra in the rooms were plotted in Fig. 9 for all analysed measurement variants. The graphs also show the scatter of the results as a difference of the maximum and minimum SPL obtained in a given frequency band between individual measurement points. An analogous comparisons are presented in Fig. 10 for the values of RT in the rooms.

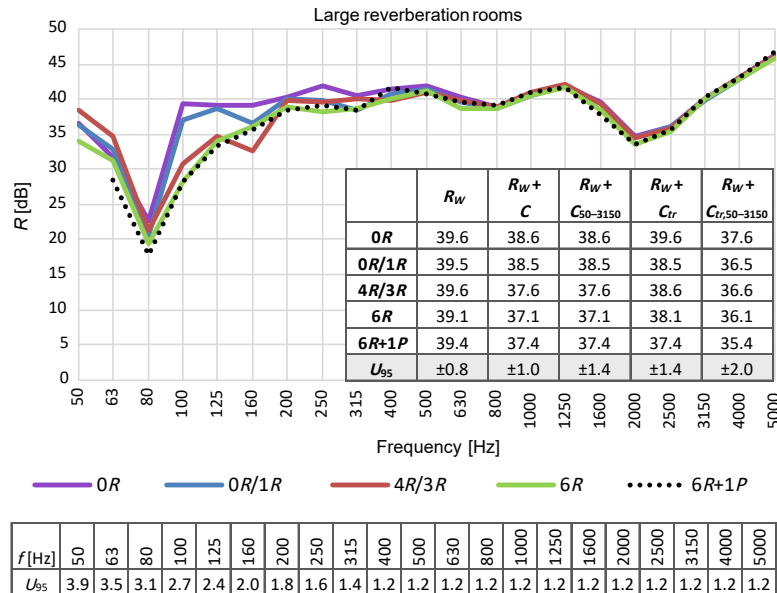


Fig. 8. Acoustic insulation of the sample tested in the large reverberation rooms in five different variants of interior acoustic adaptation: 0R – no diffusing elements in the reverberation rooms; 0R/1R – no diffusing elements in the source room and one diffusing element in the receiving room; 4R/3R – four diffusing elements in the source room and three diffusing elements in the receiving room; 6R – six diffusing elements in each room; 6R+1P – six diffusing elements and one sound absorption element in each room.

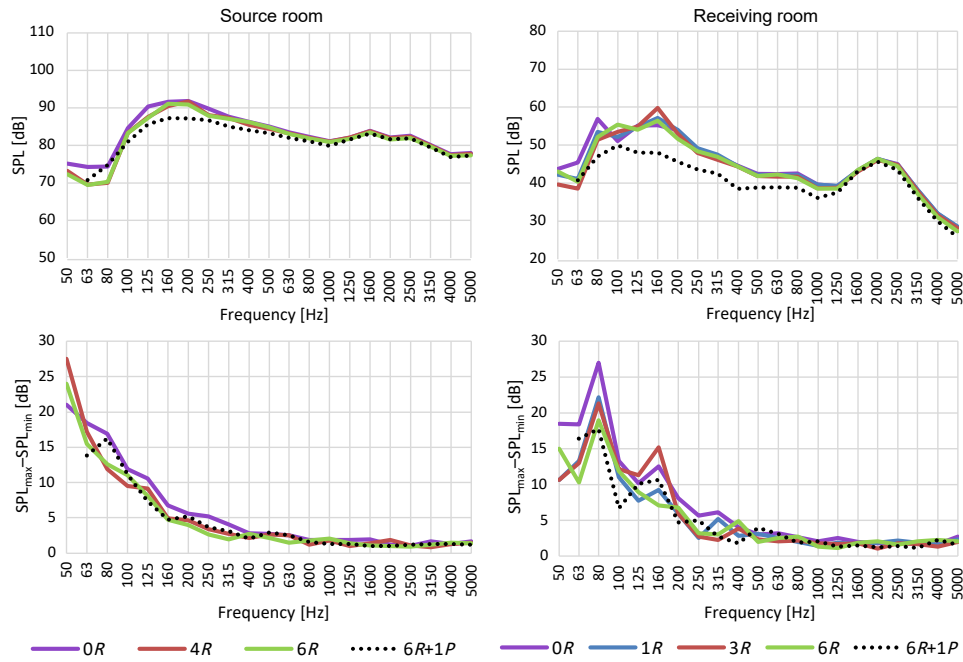


Fig. 9. Equivalent SPL and scatter of the results between individual measurement points in the large reverberation rooms in five different variants of interior acoustic adaptation: 0R – no diffusing elements in the rooms; 0R/1R – no diffusing elements in the source room and one diffusing element in the receiving room; 4R/3R – four diffusing elements in the source room and three diffusing elements in the receiving room; 6R – six diffusing elements in each room; 6R+1P – six diffusing elements and one sound absorbing system in each room.

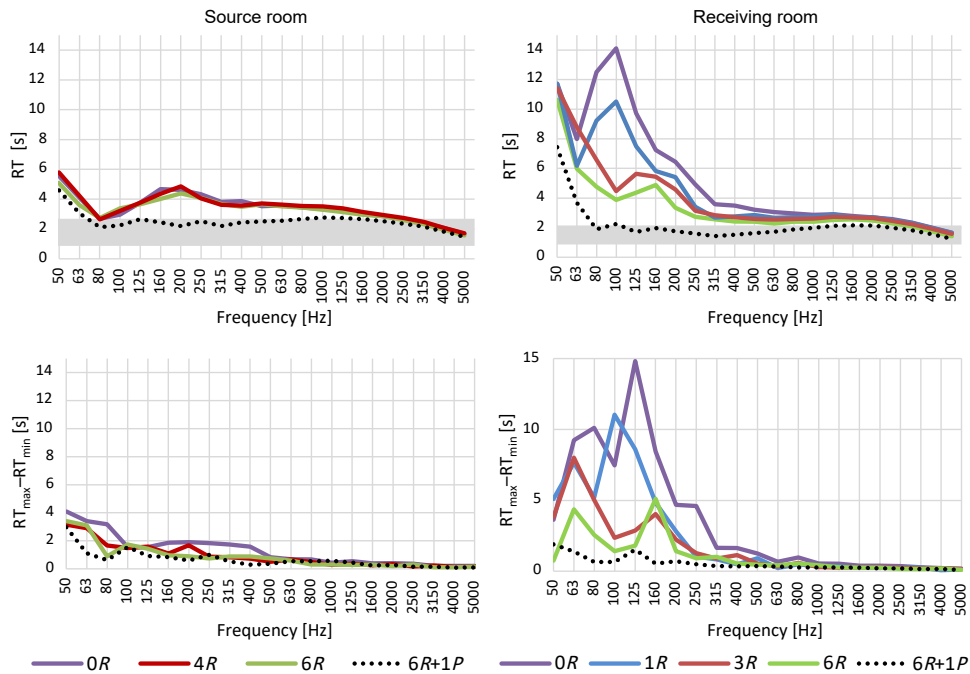


Fig. 10. RT and the scatter of the results between individual measurement points in the large source and receiving rooms in five different variants of interior acoustic adaptation: 0R – no diffusing elements in the rooms, 0R/1R – no diffusing elements in the source room and one diffusing element in the receiving room, 4R/3R – four diffusing elements in the source room and three diffusing elements in the receiving room, 6R – six diffusing elements in each room, 6R+1P – six diffusing elements in each room and additionally one sound absorbing system. In the RT diagrams, the grey colour indicates the RT ranges recommended by ISO (2021c) standard for the respective room.

Based on Fig. 9, it can be stated that the equivalent sound levels in both the source and receiving rooms do not differ significantly for interiors with different numbers of sound diffusing elements, except when there are no such elements in the rooms. In the latter case, the sound level in the low-frequency bands (below 100 Hz) is slightly higher. Through the addition of sound absorbing structures, the sound level is reduced in the rooms. Moreover, a certain dependency can be observed between the number of sound diffusing and absorbing elements and the scatter of measured values in individual points. It is the most evident in the case of receiving room at the frequency 80 Hz, for which the scatter of the values is the highest.

The graphs shown in Fig. 10 illustrate a very interesting phenomenon. On the one hand, in the source room the RT does not depend on the number of diffusing elements used, and the scatter in the results between the individual measurement points only slightly decreases as the number of such elements increases. The installation of the absorbing element in the source room ultimately reduces the RT, allowing the standard requirements to be met in the bands from 80 Hz upwards. In the receiving room, on the other hand, the RT is extremely dependent on the number of sound diffusing elements, especially in the low-frequency bands, where the difference in values reaches up to 10 s in the 100 Hz band. The situation is analogous for the scatter in the results between the individual measurement points. With a larger number of diffusing elements, these values decrease significantly. Of course, even better results are obtained with the introduction of the sound absorbing structure, both in terms of RT values, where the standard requirements are met from as low as 80 Hz, and in terms of scatter, which is rather small for this situation. Interestingly, the initial values of the RT for the situation where there were no diffusing and absorbing elements in the rooms were significantly higher in the receiving room than in the source room, even though the receiving room has a smaller volume than the source room, so theoretically the situation should be the opposite. In the source room, in principle, the use of diffusing elements was unnecessary, as the initial results demonstrate the homogeneity of the sound field. Alternatively, a sound absorbing structure could have been used to reduce the RT to the value recommended by [ISO \(2021c\)](#) standard. However, this was not necessary, as the standard recommends reducing the RT only if it can have a significant effect on the sound insulation results, which is not relevant to the analysed situation. The situation is quite different in the case of the receiving room. Here, the use of diffusing elements was necessary to control the sound field inside the room. These elements significantly reduced the RT in the room, but not because they had sound absorbing properties, but because they scattered the sound waves in the room and ensured

that the sound field was uniform. The additional sound absorbing structure further improved the situation, especially in terms of the scatter of measurement results.

In summary, the results of measurements carried out in the receiving room were extremely surprising. In the absence of diffusing elements, the room was virtually unsuitable for testing. The falsely inflated RT values (significantly higher than in the larger source room) significantly affected the final sound insulation of the sample (see Fig. 8). A completely different situation concerns the source room. From the point of view of the accuracy of the results, no additional sound diffusing and absorbing elements could actually be used in the source room. The presented measurement results raise the question as to why there are such unfavorable acoustic conditions in the receiving room and if this could have been avoided at the design stage. As mentioned at the beginning of the article, the [ISO \(2021b\)](#) standard gives quite a lot of freedom in choosing the geometry of reverberation rooms and does not impose the need for any procedure to verify the effect of the geometry design on the acoustic parameters of the interior at the design stage. It is only at the post-construction stage of the reverberation rooms that the acoustic field inside is verified and, if necessary, additional sound diffusing or sound absorbing elements are installed. The authors therefore intend to verify whether it was possible to predict at the design stage that the interior acoustic parameters of the receiving room would not be satisfactory and thus introduced a modification of the room geometry to avoid the need to install sound diffusing or sound absorbing elements undesirable by users.

The basic tools used in modeling of interior acoustics are computer programs based on the image-source method, such as: CATT Acoustic, ODEON, EASE. However, according to ([KUTTRUFF, 2000](#)) such a method is reliable only in the frequency range above the so-called Schroeder frequency. In the case of the analysed receiving room, the Schroeder frequency is 496 Hz, and for the source room it is 430 Hz. It should therefore be concluded that this is not a suitable method for the present design case, as well as for the design of other typical reverberation rooms. The above conclusion is illustrated by the graph presented in Fig. 11 which presents a comparison of the measured and simulated in CATT-Acoustic RT curves for the studied receiving room. As can be seen, the simulated RT values coincide from 500 Hz onwards with the measured values. Below 500 Hz, the curves diverge, and the measured RT takes on significantly higher values than the simulated one.

In the next step, the correctness of the design of the reverberation rooms was verified using the Bonello criteria ([BONELLO, 1981](#)). These criteria relate to the distribution of the room's intrinsic moduli, and their fulfilment is intended to ensure the uniformity of the

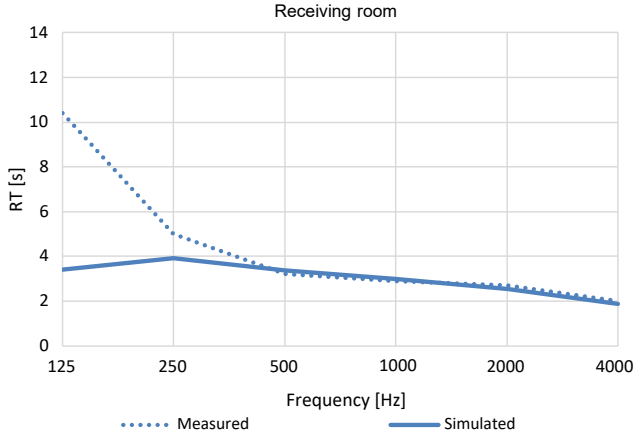


Fig. 11. RT measured and simulated in CATT-Acoustic software in the large receiving reverberation room. The simulation parameters were as follows: 50 000 rays, model 1, consideration of air sound absorption, ray tracing for 4500 ms, three source positions and ten microphone positions.

acoustic field in the interior and the minimisation of wave phenomena. The first criterion requires that the number of modes per $1/3$ octave frequency band is to be a non-decreasing function. The second criterion requires that there are no modes of overlapping frequencies. Alternatively, overlapping modes are allowed in these $1/3$ octave bands where the number of modes is minimum 5. In the analyses presented in this paper, a distance between modes of less than 1 Hz was adopted as the criterion for overlapping modes. The number of reverberation room eigenmodes were determined in two ways. The first way assumed analytical calculations using the equation proposed by MORSE and BOLT (1994):

$$N = \frac{4\pi f^3 V}{3c^3} + \frac{\pi f^2 S}{4c^2} + \frac{fP}{8c}, \quad (1)$$

where N is the number of modes from 0 Hz up to f Hz, f is the frequency [Hz], V is the room volume [m^3], S is the room surface area [m^2], and P is the total room perimeter [m]. In the second method a finite element method (FEM) modal analysis was carried out in the ANSYS environment. In the simulations, a mesh division into 10 cm finite elements was adopted. Figure 12 presents the results of the analyses for the 1st Bonello criterion carried out by both the analytical method and using computer simulations. Firstly, there is a very poor agreement between the results obtained by the analytical method and the FEM simulation results. Nevertheless, all results show that the 1st Bonello criterion is met in both the source and receiving rooms. Next, the overlap of eigenmodes in the different $1/3$ bands was compared, as indicated by the 2nd Bonello criterion. Although the overlapping modes were identified, all of them occurred in the $1/3$ bands with a minimum number of modes of 5, which

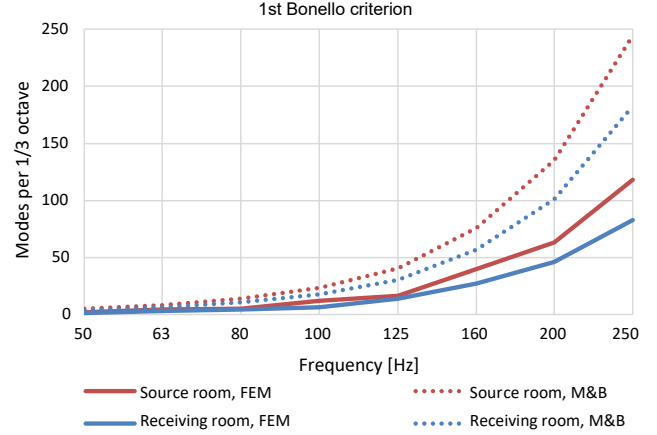


Fig. 12. Results according to the 1st Bonello criterion: eigenmodes of the source and receiving reverberation room determined according to Morse and Bolt equation (M&B) and modal analysis using FEM.

is permissible according to the given criterion. In summary, the Bonello criterion did not identify any irregularities in the receiving room geometry that could cause such a large irregularity in the sound field inside.

Analysing the results of the research presented above, it should be stated that a typical design approach based on theoretical criteria or computer simulations using the image-source method did not allow for the detection of the problem of a very high irregularity of the acoustic field in the receiving room, which became apparent at the stage of experimental research. Therefore, in the next step, the authors decided to take more advanced actions, i.e., they conducted research on a 1:7 scale model of the problematic receiving room (Fig. 13). A 38 mm-thick chipboard was used to build this model. The measurement stand was the same that was used in earlier scale studies (see Subsec. 2.1).

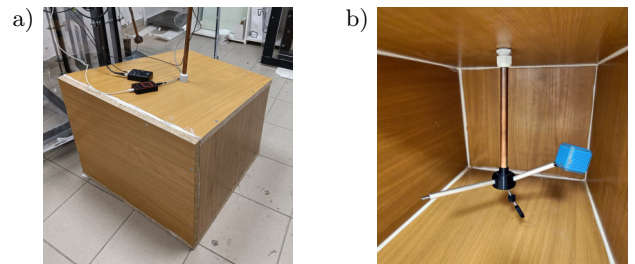


Fig. 13. 1:7 scale model of the problematic receiving room along with measurement equipment: outside view (a), inside view (b).

Figure 14 shows the comparison of the measured RT values and their scatter between individual measurement points in the full-size receiving room (1:1 scale room) and its 1:7 scale equivalent. Unfortunately, the scale tests do not identify the problem of inhomogeneous sound field in the low-frequency bands (below 250 Hz). The RT at these frequencies does not tend to be as high as it was in the full-sized room.

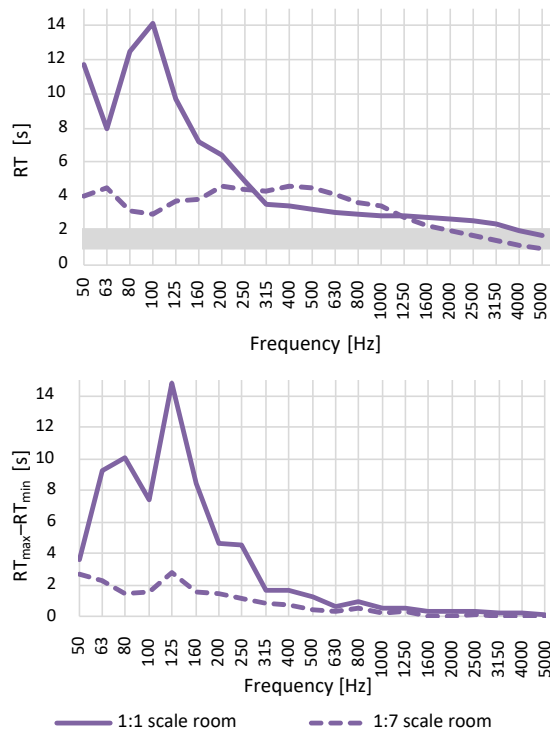


Fig. 14. RT and the scatter of the results between individual measurement points in the full-size receiving room (1:1 scale room) and its 1:7 scale equivalent. In the RT diagram, the grey colour indicates the RT ranges recommended by ISO (2021c) standard.

The scatter of the results is also small in the case of the 1:7 scale room. In the higher frequency bands, the measurement results are much more similar for both rooms. Small differences in the 315 Hz–1000 Hz bands are probably due to the mismatch of the surface sound absorption coefficients between scale and full-size rooms. In the bands above 1600 Hz, the RT in the scaled room is slightly understated because of the significant absorption of sound by the air. It should be remembered that in reality measurements were performed in a frequency range seven times higher.

5. Summary

This paper presents the characteristics of the sound field in the two pairs of coupled reverberation rooms, designed following the guidelines and the requirements of (ISO, 2021c). The results showed that only in one room, i.e., the large source reverberation room, the initial sound field was sufficiently homogeneous such that the room did not require the use of any additional sound diffusing or absorbing elements. These elements, however, were strongly recommended in the other tested rooms. Moreover, in the large receiving reverberation room they were indispensable. The lack of such elements resulted in large discrepancies between measured quantities at individual points, and above all, the recorded RT was significantly overestimated

in the low-frequency bands, where unfavourable wave phenomena occurred. This had an impact on the values of sample sound insulation. The obtained values were falsely inflated. As expected, the situation was greatly improved after introducing sound diffusing and absorbing elements in accordance with the ISO (2021b) standard. Nevertheless, diffusing and absorbing elements are not always the preferred option, since they significantly limit the usable space in the rooms and make the installation of samples, sources and measurement points more difficult. Therefore, a situation where the presence of additional diffusing and absorbing elements would not be necessary is desired. Unfortunately, following the design procedures described in the standards or using the typically available design tools, i.e., 1st and 2nd Bonello criteria, numerical simulations with the image-source method and the FEM, it seems impossible to prevent at the design stage the future necessity of using additional diffusing and absorbing elements in the reverberation rooms. Even more advanced research methods, such as measurements using scaled samples, turned out to be unhelpful. Only via verification by measurements performed in the completed rooms provides the assessment if such additional elements are required.

The authors believe that it is necessary to define additional procedures and design guidelines to improve the reverberation rooms design process. Ideally, the resulting acoustic field in the reverberation rooms should be satisfactory without installation of diffusing and absorbing elements. Firstly, the authors intend to carry out more advanced finite element simulations as basic simulations based on modal analysis failed to identify the field problem experienced in a large receiving reverberation room. Secondly, it is planned to expand the scope of research on scaled samples. The lack of convergence of measurement results between a full-size room and its 1:7 scale equivalent is very surprising and requires further verification.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. However, the authors acknowledge the support of the company Mobilne Laboratorium Techniki Budowlanej Sp. z o.o. with headquarters in Wałbrzych for making available the results of acoustic measurements conducted in their laboratories.

References

- BALMORI J.-A., CASADO-SANZ M., MACHIMBARRENA M., QUIRÓS-ALPERA S., MOSTAZA R., ACUÑA L. (2024), The use of waste tyre rubber recycled products in lightweight timber frame systems as acoustic insulation: A comparative analysis of acoustic performance, *Buildings*, **14**(1): 35, <https://doi.org/10.3390/buildings14010035>.

2. BONELLO O. (1981), A new criterion for the distribution of normal room modes, *Journal of the Audio Engineering Society*, **29**(9): 597–606.
3. BORK I. (2000), A comparison of room simulation software – The 2nd round robin on room acoustical computer simulation, *Acta Acustica united with Acustica*, **86**(6): 943–956.
4. BRADLEY D.T., MÜLLER-TRAPET M., ADELGREN J., VORLÄNDER M. (2014), Effect of boundary diffusers in a reverberation chamber: Standardized diffuse field quantifiers, *The Journal of the Acoustical Society of America*, **135**: 1898–1906, <https://doi.org/10.1121/1.4866291>.
5. CHAZOT J.D., ROBIN O., GUYADER J.L., ATALLA N. (2016), Diffuse acoustic field produced in reverberant rooms: A boundary diffuse field index, *Acta Acustica united with Acustica*, **102**(3): 503–316, <https://doi.org/10.3813/AAA.918968>.
6. DIJCKMANS A., VERMEIR G. (2013), Numerical investigation of the repeatability and reproducibility of laboratory sound insulation measurements, *Acta Acustica united with Acustica*, **99**(3): 421–432, <https://doi.org/10.3813/AAA.918623>.
7. DJAMBOVA S.T., IVANOVA N.B., PLESHKOVA-BEKIARSKA S.G. (2022), Comparative measurements of sound insulation of materials placed in small size acoustic chamber, [in:] *2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*, <https://doi.org/10.1109/ICEST55168.2022.9828622>.
8. FUCHS H.V., ZHA X., POMMERER M. (2000), Qualifying freefield and reverberation rooms for frequencies below 100 Hz, *Applied Acoustics*, **59**(4): 302–322, [https://doi.org/10.1016/S0003-682X\(99\)00038-9](https://doi.org/10.1016/S0003-682X(99)00038-9).
9. International Organization for Standardization (2020a), *Acoustics – Determination and application of measurement uncertainties in building acoustics. Part 1: Sound insulation* (ISO Standard No. ISO 12999-1:2020), <https://www.iso.org/standard/73930.html>.
10. International Organization for Standardization (2020b), *Acoustics – Rating of sound insulation in buildings and of building elements. Part 1: Airborne sound insulation* (ISO Standard No. ISO 717-1:2020), <https://www.iso.org/standard/77435.html>.
11. International Organization for Standardization (2021a), *Acoustics – Laboratory measurement of sound insulation of building elements. Part 2: Measurement of airborne sound insulation* (ISO Standard No. ISO 10140-2: 2021), <https://www.iso.org/standard/79487.html>.
12. International Organization for Standardization (2021b), *Acoustics – Laboratory measurement of sound insulation of building elements. Part 4: Measurement procedures and requirements* (ISO Standard No. ISO 10140-4: 2021), <https://www.iso.org/standard/73911.html>.
13. International Organization for Standardization (2021c), *Acoustics – Laboratory measurement of sound insulation of building elements. Part 5: Requirements for test facilities and equipment* (ISO Standard No. ISO 10140-5: 2021), <https://www.iso.org/standard/79482.html>.
14. KUTTRUFF H. (2000), *Room Acoustics*, 4th ed., Spon Press, London.
15. MLECZKO D., WSZOLEK T. (2019), Effect of diffusing elements in a reverberation room on the results of airborne sound insulation laboratory measurements, *Archives of Acoustics*, **44**(4): 739–746, <https://doi.org/10.24425/aoa.2019.129729>.
16. MORSE P.M., BOLT R.H. (1944), Sound waves in rooms, *Reviews of Modern Physics*, **16**(2): 69–150, <https://doi.org/10.1103/RevModPhys.16.69>.
17. NUTTER D.B., LEISHMAN T.W., SOMMERFELDT S.D., BLOTTER J.D. (2007), Measurement of sound power and absorption in reverberation chambers using energy density, *The Journal of the Acoustical Society of America*, **121**: 2700–2710, <https://doi.org/10.1121/1.2713667>.
18. OLIAZADEH P., FARSHIDIANFAR A., CROCKER M.J. (2022), Experimental study and analytical modeling of sound transmission through honeycomb sandwich panels using SEA method, *Composite Structures*, **280**: 114927, <https://doi.org/10.1016/j.compstruct.2021.114927>.
19. SCHMAL J., HERRIN D., SHAW J., MORITZ Ch., TALBOT A., GHASIAS N. (2021), Using simulation to predict reverberation room performance: Validation and parameter study, [in:] *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, pp. 4903–4912, <https://doi.org/10.3397/IN-2021-2879>.
20. SONIN A.A. (2001), *The Physical Basis of Dimensional Analysis*, 2nd ed., Department of Mechanical Engineering, MIT, Cambridge.
21. SZELĄG A., BARUCH-MAZUR K., BRAWATA K., PRZY-SUCHA B., MLECZKO D. (2021), Validation of a 1:8 scale measurement stand for testing airborne sound insulation, *Sensors*, **21**(19): 6663, <https://doi.org/10.3390/s21196663>.
22. URIS A., BRAVO J.M., LLINARES J., ESTELLES H. (2007), Influence of plastic electrical outlet boxes on sound insulation of gypsum board walls, *Building and Environment*, **42**(2): 722–729, <https://doi.org/10.1016/j.buildenv.2005.10.025>.
23. VALLIS J., HAYNE M., MEE D., DEVEREUX R., STEEL A. (2015), Improving sound diffusion in a reverberation chamber, [in:] *Proceedings of Acoustics 2015*.
24. WITTSTOCK V. (2015), Determination of measurement uncertainties in building acoustics by interlaboratory tests. Part 1: Airborne sound insulation, *Acta Acustica united with Acustica*, **101**: 88–98, <https://doi.org/10.3813/AAA.918807>.
25. YAO D., ZHANG J., WANG R., XIAO X. (2020), Effects of mounting positions and boundary conditions on the sound transmission loss of panels in a niche, *Journal of Zhejiang University – SCIENCE A*, **21**: 129–146, <https://doi.org/10.1631/jzus.A1900494>.
26. ZHU Q. (2022), A case study on the transmission loss suite in the University of Technology Sydney, [in:] *Proceedings of the Annual Conference of the Australian Acoustical Society, Acoustics 2021*.

Research Paper

Failure Detection of Powertrain Components in Motor Vehicles
Using Vibroacoustic MethodsBalázs József KRISTON[✉], Károly JÁLICS*[✉]*Institute of Machine and Product Design, University of Miskolc*
Miskolc-Egyetemváros, Hungary*Corresponding Author e-mail: karoly.jalics@uni-miskolc.hu

(received February 3, 2024; accepted December 19, 2024; published online March 4, 2025)

Although noise and vibration measurements are widespread in the machine diagnostics, they are not used in the diagnostics of the powertrain of motor vehicles. Our research aims to investigate the possibilities, advantages, and drawbacks of using noise and vibration diagnostics performed for motor vehicles. In this paper, we attempt to use vibroacoustic signals from a motor vehicle for diagnostic purposes. Ordinary audible malfunctions, for example, misfiring in a passenger car, were artificially created. The differences between the normal and faulty operating conditions were examined to identify evidence of failure in the vibration signal. Primarily, evaluation through Fourier transformation was performed to provide a visual correlation between the fault and the vibration behavior of the car. Detailed conclusions from the measurements and future research plans are discussed.

Keywords: vibration; acoustics; diagnostics; misfire; vehicle; analysis; internal combustion engine; malfunction.



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

1. Introduction

In the automotive industry, vehicle noise and vibration performance have become an important design parameter, as in other technical fields. Sound quality is one of the main factors that define the product itself, making vibroacoustic control of motor vehicles a key activity for automotive engineers. Furthermore, noise and vibration pollution are regulated by standards, making noise refinement during the predevelopment stage essential to protect users from health problems and other adverse effects. Malfunctions in the car's motor and powertrain can increase overall noise levels, and consequently, reduce good sound quality, leading to a noisier and less refined auditory experience.

One of the most common problems in internal combustion engines (ICEs) is an aged spark plug, which causes weak ignition and results in misfiring. This issue is particularly prevalent in older vehicles and results in reduced fuel efficiency, and potentially causing serious long-term engine damage. Nowadays, misfire detection methods are built-in in every vehicle to comply with

environmental protection regulations. Some common detection strategies include measuring cylinder pressure or monitoring speed fluctuation in the crankshaft. Unusual noises from a car can induce stress and feelings of insecurity in drivers. However, this unwanted phenomenon can also be utilized, since the vibration is sensitive to all faults, whereas other physical parameters, such as those monitored by onboard diagnostics system (OBD), are sensitive only to specific faults. This means that monitoring a vehicle's vibration behavior can identify potential failures.

Nevertheless, vibration diagnostics has its limitations, as they depend on the product's complexity, operation mode, and the severity of the fault. Solely relying on the overall sound pressure level (SPL) and averaged vibration spectra does not give a sufficient representation of sound quality. That is why the introduction of psychoacoustic measurements is necessary to gain a more comprehensive insight into human sound perception. On the other hand, psychoacoustic analysis can serve as a diagnostic tool for identifying vehicle malfunctions since experienced mechanics can

often diagnose certain malfunctions in cars by sound alone. For example, at idle speed, a knocking noise from beneath the valve cover may be clearly audible. As the engine's rotational speed increases the noise frequency also increases, which indicates that the valve clearance is too large.

At the Aachen University in Germany (BRECHER *et al.*, 2011), a correlation analysis was conducted between gear parameters and psychoacoustic values based on noise measurements from different gear sets. For the research, gear sets with different surface microstructure and pitch deviation were selected. The study found that both the loudness and sharpness of the noise increased with rotational speed. According to the study's findings, roughness proved to be the most valuable parameter for identifying pitch deviation failures in gears. BABU DEVAENAPATI *et al.* (2010) analyzed a four-stroke four-cylinder petrol engine with a misfire problem. For problem identification, they used statistical parameters of vibration signals, such as kurtosis, standard deviation, mean, etc. A decision tree was developed that could extract the most appropriate parameters for failure detection and to classify various ICE misfire problems with 95 % accuracy. FIRMINO *et al.* (2012) collected vibration and acoustic data from a four-stroke spark ignition engine with a misfire in one cylinder. After performing feature extraction using the fast Fourier transform (FFT) algorithm, the data was used to feed different artificial neural network (ANN) systems in order to detect the misfire failure. Both networks demonstrated great results, achieving accuracy of around 99 % in misfire detection. DELVECCHIO *et al.* (2018) reviewed the existing state-of-the-art vibroacoustic techniques for diagnosing failures in ICEs, including misfires. According to this study, the most commonly used techniques for ICE malfunctions are joint time-frequency methods. However, these methods are mainly applied to failure detection rather than condition monitoring purposes.

WOJNAR and MADEJ (2009) tested ICEs using vibroacoustic methods and concluded that relying only on the FFT does not deliver sufficient results. They emphasized the advantages of joint time-frequency methods, particularly wavelet analysis. WOJNAR and STANIK (2010) compared vibration and acoustic signals for diagnosing car wheel bearings. Their investigation revealed that bearing wear can be determined through vibroacoustic methods. SZABÓ and DÖMÖTÖR (2022) also investigated the wheel bearings of a passenger vehicle with vibroacoustic methods, and confirmed that these methods are effective for detecting bearing faults. WOJNAR *et al.* (2011) further investigated roller bearing defects, focusing on non-dimensional factors (e.g., impulse factor, crest factor, etc.). Their findings showed that these parameters are sufficient for detecting bearing faults.

Psychoacoustic quantities are not currently involved in detection or monitoring actions. Analysis acoustic data such as SPL obtained from a microphone, is rarely used due to the masking effect of background noise, making it unsuitable for detecting assembly faults. However, joint techniques based on acoustic signals remain useful for capturing and localizing transient events in the time or angular domain, especially when the noise characteristics cover a wide frequency range and originate from different areas of the engine. Such events in ICE could be knocking, misfires, or injection problems. Using these methods, more mechanical events that influence the vibroacoustic behavior of the engine can be captured in a single measurement. On the other hand, misfires produce structure-borne noise, which means that vibration signals are effective for detecting such failures as well. For purely airborne noise, the SPL signal is relevant for: turbocharger, ventilation fan, or exhaust system; however, for mechanical malfunctions, which are structure-borne transmitted, the fault must be in advanced stage to be detectable by acoustic signals. Additionally, the use of transducers allows for targeted examination of sub-components of the ICE, depending on their positioning.

Time domain analysis focuses on observing the shape of the time signal. The information that the time domain contains can be described by the above mentioned statistical single values. While these values are sufficient for detecting malfunctions, they are not effective for localizing failures. To use these values as decision-making criteria in automated diagnostic system, the time signal must be insensitive to background noise and should not contain unnecessary information. The signal-to-noise ratio can be maximized by applying frequency band filters to the time signal.

The analysis can also be performed in the frequency domain, where distinct frequency peaks and harmonics correspond to different components. For this purpose, FFT is applied, revealing the frequencies of various events with different energy content. This algorithm is effective only for cyclo-stationary signals, helping to understand the cause of failure and providing reliable information for condition monitoring and diagnostic activities.

As a summary, the authors recommend performing time-frequency analysis when the nature of the fault is impulsive, with the consideration of the level of investigation and computational efforts required. For condition monitoring and failure detection, it is common practice to combine scalar parameters with 2D analysis. In this case, the scalar parameter serves as input for the decision-making algorithm, while the latter is a visual representation for the user. It is important to note that the scalar value must contain all the information stored in the 2D map to ensure accurate diagnostics.

2. Measurement arrangement

Based on our experience and the suggestions of the above-mentioned authors, we performed a test series on a real vehicle. The vehicle was a first-generation Ford Focus passenger car (1998 model; front-wheel drive, 5-speed manual transmission) with a 1.6 liter, 4-cylinder, four-stroke naturally aspirated petrol engine. For data collection, a 4-channel Brüel & Kjaer Photon+ DAQ system was connected through USB to a notebook. The notebook itself was powered by its built-in battery, which helped eliminate the potential interference from the 50 Hz AC mains.

During the measurements, an easy installation of the sensors (accelerometers and a microphone) without dismantling the car was a key requirement. This was based on the general requirement of workshop repair personnel, to avoid excessive disassembly for a simple test. To this end, one uniaxial acceleration sensor was placed on the right front side of the car body, and another was positioned on the connection bolt head between engine block and the gearbox housing (Fig. 1). Additionally, a condenser microphone was placed at the front passenger's head level. The measurements were repeated several times at idle speed with engine speeds of 1000 rpm, 2000 rpm, 3000 rpm, and 4000 rpm, all without load. Furthermore, noise and vibration were measured in accelerated mode under partial open throttle (POT) conditions during a run-up and run-down cycle from 1000 rpm to 5000 rpm and back to 1000 rpm. The length of the run-up and run-down time was controlled by the driver using the gas pedal. During the measurements, the coolant temperature was monitored via the onboard coolant temperature gauge, and it was kept around 90 °C operating temperature during the tests. The acquired raw time signal was later post-processed with the help of Artemis Suite noise evaluation software.

To create a faulty condition in the engine, the operation of one of the four cylinders was eliminated by

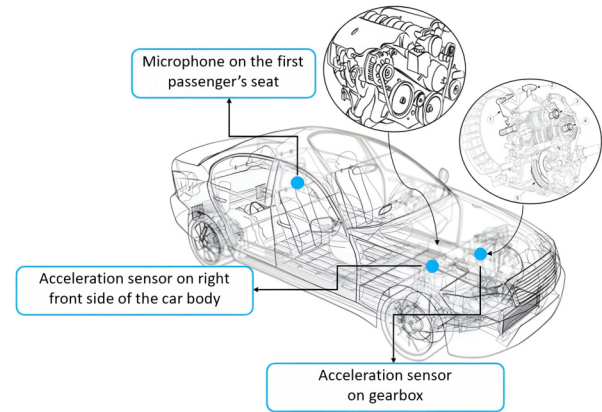


Fig. 1. Sensor positions.

disabling the ignition in cylinder 1 (on the side of the timing drive). The effect of the misfire was clearly noticeable by ear in the immediate vicinity of the car. The goal of the measurements was to analyze the vibration behavior of the engine in the presence of a misfire fault. Based on this analysis, the potential for detecting and localizing failures should be investigated.

3. Analysis

In the course of the analysis, the raw time signals were post-processed by the FFT algorithm. The purpose of the analysis was to find acoustic patterns which may refer to a malfunctioned part in spectrums and spectrograms.

Initially, the time signals were analyzed. We can state that the microphone signal recorded during the run-up tests provided more promising outputs from a diagnostic perspective, since time domain signal obtained from the microphone's measurement showed better separation (Fig. 3) in sound pressure between healthy (blue) and faulty (red) conditions, compared to the acceleration signal recorded on the gearbox (Fig. 2).

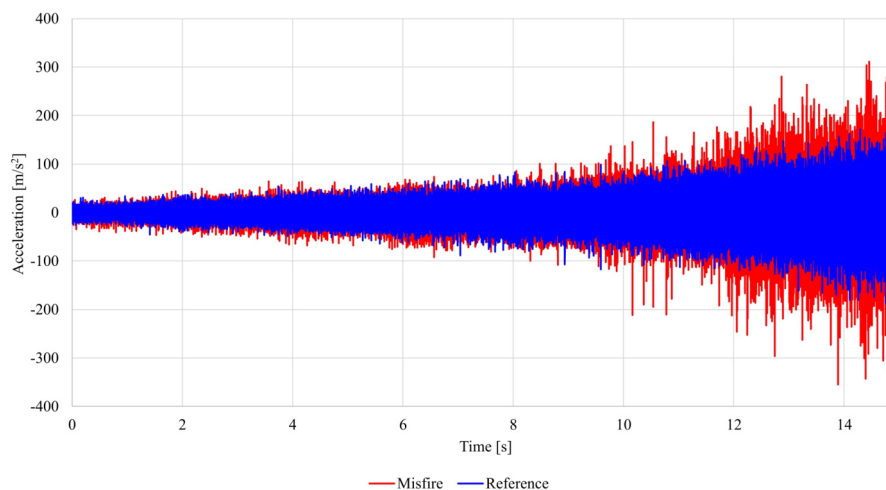


Fig. 2. Gearbox time history at ramp speed of 1500 rpm–4200 rpm.

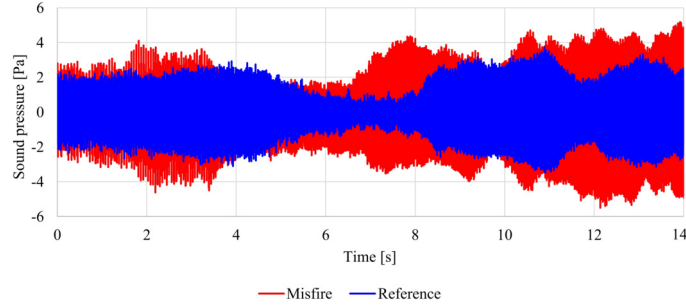


Fig. 3. Microphone time history at ramp speed of 1500 rpm–5000 rpm (POT).

In Fig. 3, the form of the acoustic signal is characterized by the components directed toward the passenger's seat, with different frequency-dependent damping properties. The constant-speed measurements do not seem to be very useful for distinguishing failure modes. However, an interesting effect is observable, especially at higher rotational speed and is evident only in the microphone signal, see Fig. 4.

The shape of the time signal shows a very slow, pure sinusoidal, strong modulation (1.5 Hz–2 Hz). This modulation effect becomes stronger when the engine is misfiring. In our opinion, it is caused by fluctuations in the engine crankshaft's rotational speed, as one can see in Fig. 5. This effect can be explained by different cylinder pressures caused by the misfire. However, it is important to note that combustion engines have a certain speed fluctuation, unlike electric motors. The rpm signal (Fig. 5) was created with an rpm generator, which is a built-in function in the noise evaluator software.

The time interval between the distinct peaks is around 0.0075 seconds, which corresponds to a calculated frequency of 133.33 Hz. This is the ignition frequency at 4000 rpm (Fig. 4), which can be calculated for a 4-stroke internal combustion engine using the following formula:

$$f_{\text{ignition}} = \frac{1}{2} \cdot \frac{\text{rpm}}{60} \cdot \text{cylinders} [\text{Hz}], \quad (1)$$

where rpm is the motor crankshaft speed in [1/min], and “cylinders” refers to the number of cylinders (four in this case) and it is divided by two, since two ignition is required to rotate the crankshaft 360° as two cylinders move together at the same time. As the engine construction is fixed, this frequency depends only on the rotational speed.

The motor frequency can be easily calculated as

$$f_{\text{motor}} = \frac{\text{rpm}}{60} [\text{Hz}]. \quad (2)$$

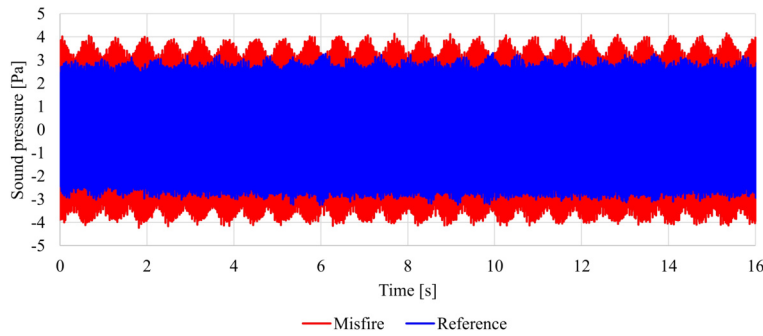


Fig. 4. Microphone time history at constant speed of 4000 rpm.

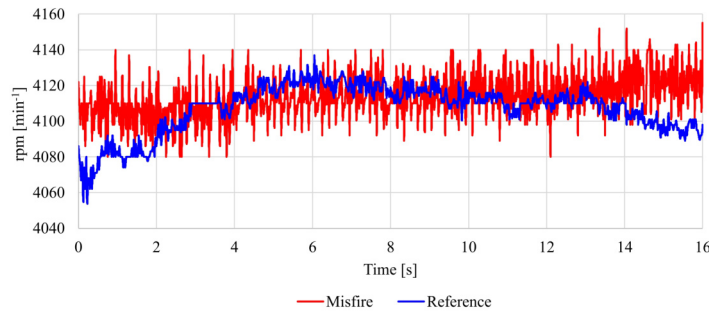


Fig. 5. Rpm curves at constant 4000 rpm derived from the microphone's signal.

The camshaft is connected to the crankshaft through a belt drive. This shaft activates the cylinders' intake and exhaust valves, thus controlling the combustion process. The transmission ratio between the shafts is usually 2:1, which means that the camshaft's rotation speed is half of the motor shaft. Therefore, the camshaft frequency is

$$f_{\text{camshaft}} = \frac{\text{rpm}/2}{60} [\text{Hz}], \quad (3)$$

which is the same frequency as the motor 0.5th order. It is observable in the spectrogram of every sensor, but the best representation of the failure can be derived from the microphone's signal. Table 1 presents the first-order fundamental frequencies of the motor at different speeds for comparison in the analysis.

Table 1. Fundamental frequencies of the engine [Hz].

	2000 rpm	3000 rpm	5000 rpm
Crank frequency	33.33	50	83.33
Camshaft frequency	16.66	25	41.66
Ignition frequency	66.67	100	166.67

The ignition frequency is recognizable in both healthy and faulty cases. It means that even under normal conditions, the ignition phenomena characterize the vibration behavior of the motor. Since the time signal during the ramp speed measurements can visualize the problem, single statistical values – such as root mean square (RMS), crest factor, standard deviation, kurtosis, etc., should show a high deviation factor between the two conditions. The RMS value of a given set of discrete data points can be calculated by the following formula:

$$\text{RMS} = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2}{n}}. \quad (4)$$

First, the data points are squared, then the average of all the squared values is taken. After that, the square root of the average is calculated. This process tells us how much energy is contained in the waveform.

The skewness shows the asymmetry of a distribution. If the skewness value is zero, the distribution is symmetrical. A normal distribution has a zero skew. The easiest method to check the skewness is to plot the data on a histogram. If the distribution has right (positive) skew, it means the distribution is shifted to the right relative to the axis of symmetry. Conversely, in the case of left (negative) skew, the distribution is longer on the opposite side (TURNERY, 2022). The skewness values obtained from the gearbox acceleration sensor and the microphone signal show that the skewness value is negative, while the sensors on the car body yield positive values. The equation for skewness is as follows:

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3. \quad (5)$$

The mean value was calculated with the following equation:

$$\text{Mean} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}. \quad (6)$$

The standard deviation is a measure of the spread around the mean value. A low standard deviation means the data are clustered around the mean, while a high standard deviation indicates data are more spread out. The formula used to calculate standard deviation is

$$\text{Standard deviation} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-2}}. \quad (7)$$

The peak amplitude derived from the RMS is given by:

$$\text{Peak} = \frac{2}{\sqrt{2}} \text{RMS}. \quad (8)$$

The peak-to-peak amplitude is the difference between the highest positive and the lowest negative amplitude in the waveform:

$$\text{Peak to peak amp.} = \max \{x_i\} - \min \{x_i\}. \quad (9)$$

The crest factor gives the ratio of the peak values to the effective value, showing how prominent the peaks are in the waveform. A crest factor of 1 indicates no peaks, while a higher crest factor indicates peaks. The crest factor is calculated as

$$\text{Crest factor} = \frac{\text{Peak}}{\text{RMS}}. \quad (10)$$

The statistical parameter called kurtosis is a measure of the “peakedness” of a random signal:

$$\text{Kurtosis} = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}. \quad (11)$$

Unfortunately, the statistical single values of the microphone's time signal do not provide adequate difference between the bad and good conditions (Fig. 6).

The same statement is true for the crest factor values: the failure shows no separation in this parameter compared to the original condition (Fig. 7). However, in certain engine speed ranges (around 2700 rpm and above 4000 rpm) the kurtosis parameter indicates a small deviation between the conditions (Fig. 8).

Nevertheless, the result is not conclusive due to the low distinction of the individual overall values. To better understand the malfunction, joint time-frequency (FFT vs. time) analysis was performed at both constant and ramp speeds. Joint analysis is the representation of series of Fourier transformations over different time periods (or at different rotation speeds),

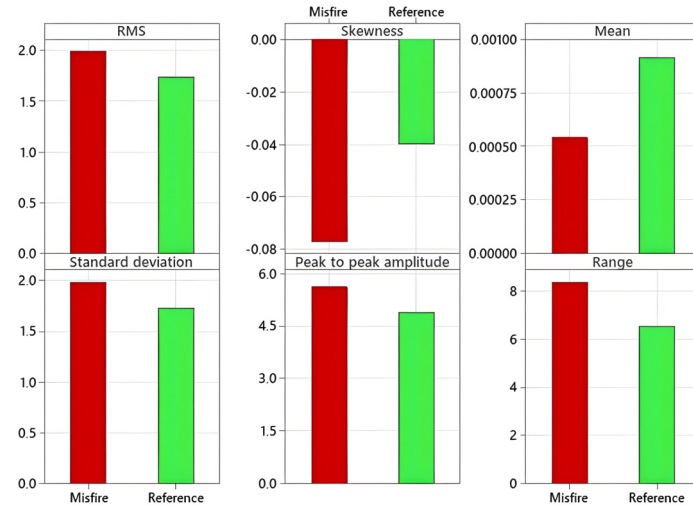


Fig. 6. Time domain statistical single values derived from the microphone's signal on 4000 rpm.

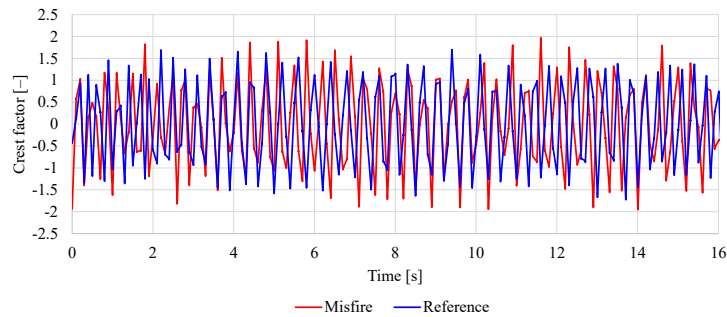


Fig. 7. Crest factor in the function of time derived from the microphone's signal at 4000 rpm.

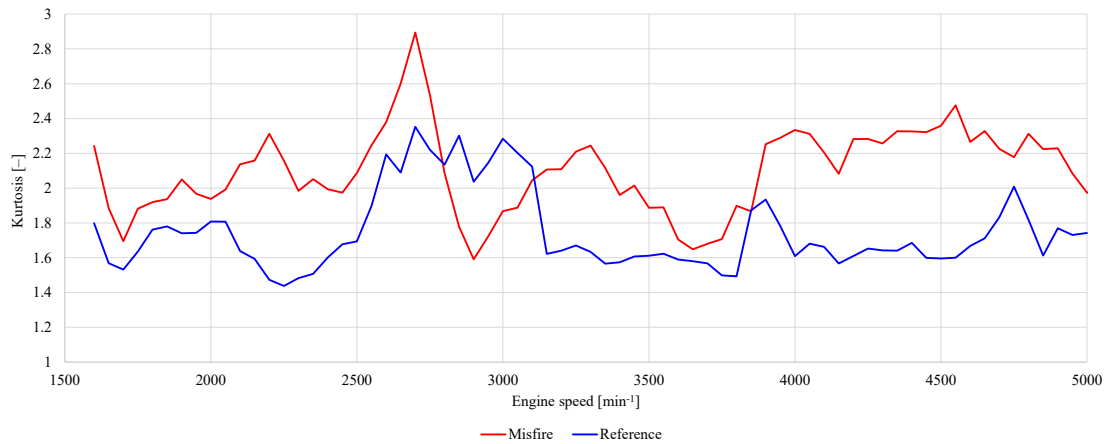


Fig. 8. Kurtosis as a function of engine rotation speed derived from the microphone's ramp signal.

mapping a 1D time domain into a 2D diagram that shows energy (color scale) versus time (x -axis) and frequency (y -axis). This analysis helps to understand how the energy content of frequencies varies over time or as a function of rotation speed. As shown in Fig. 9, it is clear that the sound pressure level increases at specific motor frequencies. The sound pressure at the ignition frequency is a dominant contributor to the overall sound pressure level inside the car, even in healthy con-

dition, where only vibrations below 2000 Hz are significant. The dominance of the ignition frequency is observable at the other measurement points as well. The motor subharmonics create abnormal colormap picture in the case of a misfire issue. Among the topological integer motor frequencies, the motor half-orders (subharmonics) appear with higher energy. This leads to the assumption that the motor 0.5th order (17.58 Hz) causes modulation in the signal. One possible reason

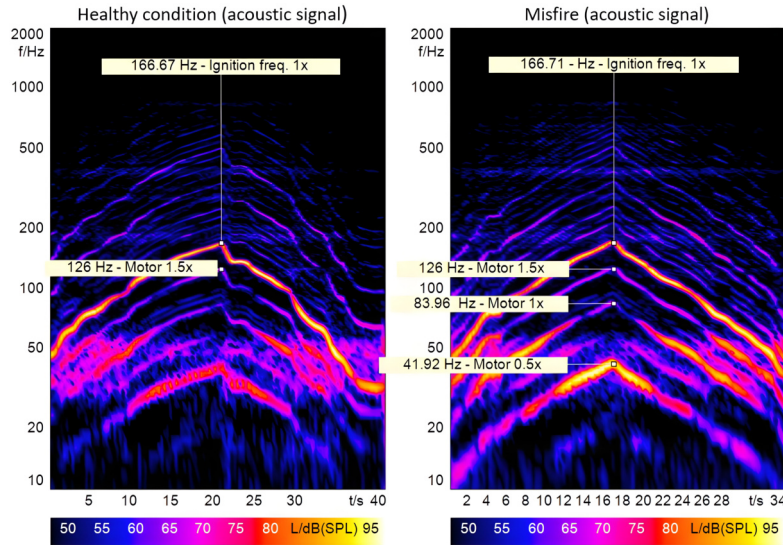


Fig. 9. Microphone spectrogram (10 Hz–2000 Hz) in healthy (left) and faulty (right) conditions during ramp speed (1500 rpm–5000 rpm).

for this is that the engine crankshaft rotation becomes, let us say, more unbalanced due to the misfire in cylinder 1.

Rather than stating that the crankshaft itself is unbalanced, it is more accurate to say that, as a consequence of the misfire, the shaft rotation speed fluctuates, causing uneven running. This hypothesis is supported by CAVINA *et al.* (2002), who claim that misfire results in a sudden lack of torque on the crankshaft, leading to damped torsional vibrations at representative frequencies of the engine.

The joint analysis of the acoustic signal made possible to determine the location of the malfunction, as we were able to identify frequencies that correspond to the engine crankshaft 0.5th, 1st, 1.5th, 2.5th, and 3rd, as well as other higher-orders. However, one can de-

tect with a high degree of certainty that the failure is coming from the motor by simply listening to the sound of the car. Unfortunately, resonance appears in the joint time-frequency analysis in a similar manner to harmonic frequencies at constant speed. Due to this fact, it is worth considering the spectrogram when the rotation speed varies over time, e.g., in ramped speed measurements.

One can see that there is a resonance at 50 Hz, which increases the sound pressure level of the motor's first order, when it operates between 2400 rpm–3100 rpm (Fig. 10). The order shapes demonstrate how the motor speed changes over time: the motor accelerates over 30 seconds, reaching a maximum speed of 5000 rpm during the run-up, and then slows down to 1500 rpm during the run-down phase. This

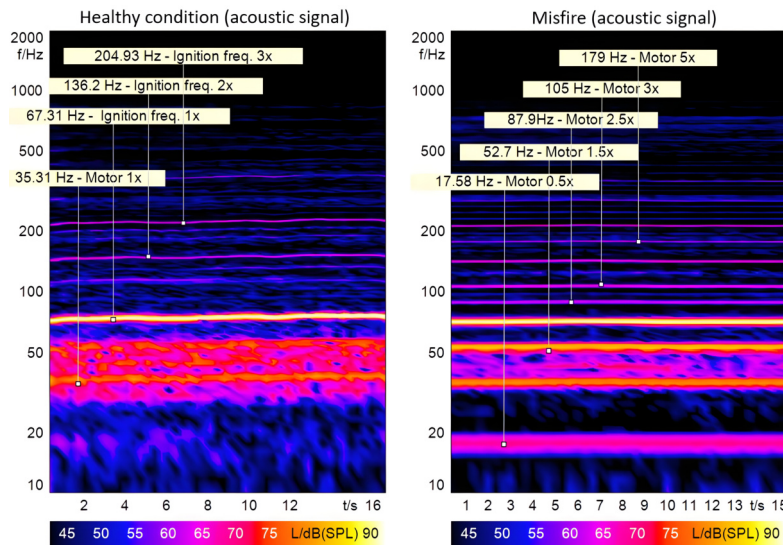


Fig. 10. Microphone spectrogram (10 Hz–2000 Hz) in healthy (left) and faulty (right) conditions at 2000 rpm.

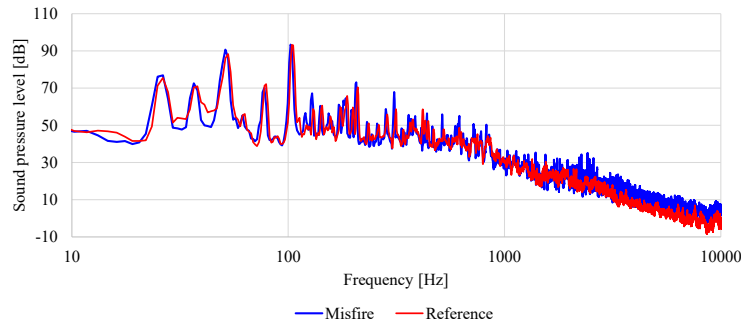


Fig. 11. Microphone's spectrum comparison at 3000 rpm (logarithmic abscissa).

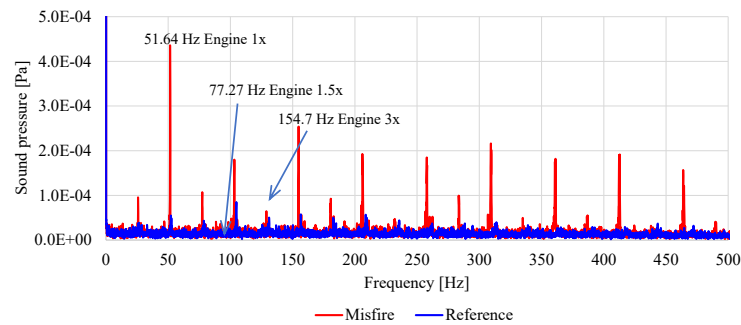


Fig. 12. Microphone's modulation spectrum comparison at 3000 rpm.

method reveals the resonance frequencies without mistake based on excitation and helps to avoid misunderstandings during analysis. The time domain can be transformed into the frequency domain with FFT. The energy content of the microphone signal in terms of frequencies is represented in the spectrum at 3000 rpm motor speed, as shown in Fig. 11. This gives a slight correlation with the fault, though the correlation is even weaker at lower speeds below 3000 rpm. Based on the spectrum, it is difficult to identify the problem. There is a deviation in the frequency range of 1 kHz–10 kHz, due to assumed amplitude modulations. While the operation of the misfire is visible in the spectrum, it is challenging to identify a specific frequency component related to a particular part of the engine. Based on the aforementioned analysis, we reasonably assume that – based on the FFT vs. time analysis as well – that there is amplitude modulation in the signal. Since the FFT vs. time diagram shows that a wide frequency range of the signal is affected, it makes sense to check the modulation spectrum.

The modulation spectrum provides overview of the modulation frequencies across the entire or a selected frequency range. The modulation spectrum shown in Fig. 12 includes the frequency range of 2.8 kHz–5.6 kHz. The envelope low-pass frequency is 1000 Hz, so the frequencies that modulate the signal appear up to 1000 Hz. The analysis reveals that the half-order motor frequency plays a significant role in the modulation. Specifically, the modulation frequency is 25 Hz, which is half of the crankshaft's rotation frequency.

4. Discussion

In this paper, the misfire event in a motor vehicle was studied with vibroacoustic methods. The misfire caused an unbalanced, or more accurately, uneven rotation of the crankshaft. By analyzing the microphone's time-domain signal, one can make a clear distinction between healthy and faulty conditions of the engine. A short frequency calculation analysis showed that the ignition plays a main role in the vibration behavior of both the car body and passenger area. The FFT spectrum also indicates the presence of the failure, similarly to the time-domain signal, but tracking the frequency components in the spectrum does not allow for precise localization of the failure.

The most useful method was the FFT vs. time analysis, where the topological integer and odd-order engine showed increased energy in the faulty condition. The outcome of the modulation spectrum confirmed that there is subharmonic motor order modulation in the spectra. This result allowed us to localize the place of the noise problem inside the car. However, even without advanced analysis, a trained ear could identify that the issue likely originates from the engine.

In summary, with the help of vibroacoustic methods the noise problem could be spotted inside a vehicle. However, with the current measurement points and tools, it is possible to determine in which cylinder the misfire occurs. This could be potentially achieved by placing more acceleration sensors on the car body, for example, on the left front side. Based on the re-

sults, the possible location of the noise problem can be narrowed down; however, the type of the malfunction is not clearly identified.

The reason behind this is that we cannot be sure that only a misfire failure causes the observed acoustic patterns and changes in the spectrograms. It is not the misfire itself, but rather its consequences or, more precisely, the complete absence of the stroke in cylinder 1 (resulting in uneven running of the crankshaft) that determines the vibration behavior of the engine. The shafts are statically and dynamically balanced during manufacturing to account for the moving masses in the crank mechanism, ensuring they do not generate significant radial vibrations. Due to the uneven running of the shaft, torsional vibration occurs, but these were not measured. However, vertical and horizontal vibrations can originate from gas forces and mass forces, although the cylinders were not modified. At this point, the unevenness of the gas forces must be considered, because in the first cylinder only the maximum pressure (2 MPa–3 MPa), corresponding to the compression cycle, prevails at the end of the combustion cycle. In contrast, in the other three cylinders, a higher pressure (8 MPa–10 MPa) derived from ignition, is present at the beginning of work cycle.

This difference in cylinder pressure causes abnormal torque behavior, which is why orders with odd numbers and subharmonic orders are present in the result. Practically, the tested engine operates as a 3-cylinder engine where odd orders such as the 1.5th, 3rd, etc., and subharmonics appear. However, despite of this, the four pistons are moving, so integer order numbers (1st, 2nd, 4th, etc.) are also present in the spectrograms. The acoustical pattern of this failure is not unique; other malfunctions that affect crankshaft rotation can trigger the same vibroacoustic behavior. This means that joint analysis alone is not capable to identify the misfire; other non-vibroacoustic measurements are essential for exclusively detecting the problem.

5. Further plans





The low-frequency motor modulation can be linked to psychoacoustical parameters such as fluctuation strength and roughness. These parameters could possibly serve as good indicators of this type of failure, but to justify the relevance of this idea further investigation is necessary. As a continuation of the research, it would be worth to examine how the vibration behavior of the vehicle changes when more than one cylinder is misfiring. Furthermore, we are interested in examining

other malfunctions, e.g., valve clearance defect. The ultimate goal is to pinpoint the misfiring cylinder and distinguish this failure mode from other malfunctions using only vibroacoustic tools.

References

1. BABU DEVASENAPATI S., SUGUMARAN V., RAMACHANDRAN K.I. (2010), Misfire identification in a four-stroke four-cylinder petrol engine using decision tree, *Expert Systems with Applications*, **37**(3): 2150–2160, <https://doi.org/10.1016/j.eswa.2009.07.061>.
2. BRECHER C., GORGELS C., CARL C., BRUMM M. (2011), Benefit of psychoacoustic analysing methods for gear noise investigation, *Gear Technology*, **28**(5): 49–55.
3. CAVINA N., CORTI E., MINELLI G., SERRA G. (2002), Misfire detection based on engine speed time-frequency analysis, *SAE Transactions*, **111**: 1011–1018, <http://www.jstor.org/stable/44743127>.
4. DELVECCHIO S., BONFIGLIO P., POMPOLI F. (2018), Vibro-acoustic condition monitoring of internal combustion engines: A critical review of existing techniques, *Mechanical Systems and Signal Processing*, **99**: 661–683, <https://doi.org/10.1016/j.ymssp.2017.06.033>.
5. FIRMINO J.L., NETO J.M., OLIVEIRA A.G., SILVA J.C., MISCHINA K.V., RODRIGUES M.C. (2012), Misfire detection of an internal combustion engine based on vibration and acoustic analysis, *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, **43**: 336, <https://doi.org/10.1007/s40430-021-03052-y>.
6. SZABÓ J.Z., DÖMÖTÖR F. (2022), Comparative testing of vibrations in vehicles driven by electric motor and internal combustion engine (ICE), [in:] *Vehicle and Automotive Engineering 4*, Jármai K., Cservedák Á. [Eds.], pp. 871–879, https://doi.org/10.1007/978-3-031-15211-5_72.
7. TURNEY S. (2022), Skewness. Definition, examples & formula, <https://www.scribbr.com/statistics/skewness/> (access: 05.01.2024).
8. WOJNAR G., CZECH P., STANIK Z. (2011), Use of amplitude estimates and nondimensional discriminants of vibroacoustic signal for detection of operational wear of rolling bearings, *Scientific Journal of Silesian University of Technology. Series Transport*, **72**: 109–114.
9. WOJNAR G., MADEJ H. (2009), Averaged wavelet power spectrum as a method of piston – Skirt clearance detection, *Diagnostyka*, **2**(50): 93–98.
10. WOJNAR G., STANIK Z. (2010), Influence of wear of bearings carriageable wheels on acoustics pressure, *Scientific Journal of Silesian University of Technology. Series Transport*, **66**: 117–122.

Research Paper

Tyre Labelled Noise Values in the Context of Environmental Protection:
Weaknesses of the Method and Benefits of Silent TyresMaciej HAŁUCHA^{(1)*} , Janusz BOHATKIEWICZ⁽²⁾ ,
Piotr MIODUSZEWSKI⁽³⁾ , Truls BERGE⁽⁴⁾ ⁽¹⁾ *EKKOM Sp. z o.o.*
Cracow, Poland⁽²⁾ *Road and Bridge Research Institute*
Warsaw, Poland; e-mail: jbohatkiewicz@ibdim.edu.pl⁽³⁾ *Faculty of Mechanical Engineering and Ship Technology, Gdańsk University of Technology*
Gdańsk, Poland; e-mail: pmiodusz@pg.edu.pl⁽⁴⁾ *SINTEF Digital, Torgarden*
Trondheim, Norway; e-mail: truls.berge@sintef.no*Corresponding Author e-mail: maciej.halucha@ek-kom.com

(received May 7, 2024; accepted November 26, 2024; published online February 24, 2025)

The purpose of this work was to examine the impact of the inadequacies in the current procedure for car tyre labelling, specifically in the context of environmental noise, and to present the benefits of adopting more realistic procedure with the use of low-noise tyres. This was done using two approaches: an impact analysis and a cost-benefit analysis. The calculations were performed to show this impact on environmental noise. This was done using the common noise assessment methods in Europe (CNOSSOS-EU) model (recommended for strategic noise mapping of EU countries), which was validated using test results from sound exposure level measurements on both ISO test track and on real road sections. Using the noise calculation results, a cost-benefit analysis was performed, incorporating financial analyses of both the current and projected situation under different strategies to reduce tyre/road noise.

Keywords: tyres; tyre noise; road noise; environmental protection; EU tyre label; tyre labelling procedure; traffic noise calculation.



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

1. Introduction

Vehicle noise is generated by three main sources: powertrain noise, tyre/road noise, and aerodynamic noise. The first source depends on factors related to the load and speed of the engine. The noise level varies with the road gradient, vehicle, speed, and the type of vehicle. Driving style also has a significant influence. Similarly, tyre/road noise is influenced by different factors. In this case, the noise level depends mainly on the type of road surface and tyres. Tyre/road noise increases with vehicle speed (BERGE, 2023; SANDBERG, EJSMONT, 2002). It is the dominant source of noise at higher speeds, but can still be heard at lower speeds.

This is demonstrated by research on the Swiss model sonROAD18 (HEUTSCHI, LOCHER, 2018), as presented in Table 1. Aerodynamic noise, created by airflow dis-

Table 1. Contribution of tyre/road noise at different vehicle speeds (HEUTSCHI, LOCHER, 2018).

Speed [km/h]	Percentage of tyre/road noise [%]
30	62.5
40	78.5
50	86.6
60	90.9
80	94.8
100	96.1

turbance, is also a significant component of traffic road noise at higher speeds.

Tyre/road noise is the most significant contributor to traffic noise, making low-noise surfaces one of the most effective noise reduction measures (BOHATKIEWICZ, HAŁUCHA, 2017; BOHATKIEWICZ *et al.*, 2022). This noise can also be effectively reduced by using low-noise tyres on vehicles. The proportion of tyre/road noise will increase as the number of electric vehicles increases in traffic flow, as powertrain noise is extremely low at low speeds in electric cars, making tyre/road noise the dominant source. This will be especially important in urban conditions (HAŁUCHA *et al.*, 2023).

The combination of quieter tyres and quieter pavements is the most effective measure to reduce noise in road surroundings (BERGE *et al.*, 2022; BERGE, 2023). To make such solutions feasible, it is necessary to ensure that consumers have access to information on the noise levels of car tyres. Tyre labels could serve as a valuable tool for this purpose. The European Parliament and the Council introduced Directive on tyre labelling (European Union, 2009) aimed at increasing consumer awareness of car tyres in terms of three main parameters: wet grip, rolling resistance and rolling sound. The new directive (European Union, 2020) introduced several changes, including the current form of the label. The method used to determine the noise level subsequently put on the tyre label is described in Regulation No. 117 (United Nations Economic Commission for Europe [UN/ECE], 2011). This method involves measuring noise during a controlled pass-by of a test vehicle equipped with the test tyres. These tests are conducted on a specially designed surface defined in (International Standard Organisation [ISO], 2021).

Although tyre labels have been on the market for several years, there remains significant uncertainty in the results of tyre labelling (SANDBERG, MIODUSZEWSKI, 2022). This uncertainty is mainly influenced by the test tyres themselves, variations in the noise properties of ISO surfaces, and the influence of the test vehicle and meteorological conditions, among

others. This issue is described in the STEER project (strengthening the effect of quieter tyres on European roads), which was commissioned by CEDR in 2020 and finalised in 2022 (BÜHLMANN *et al.*, 2022). The project estimated that the uncertainty for C1 (passenger car tyres) and C2 (van and light truck tyres) ranges from 1.4 dB to 2.0 dB, expressed as standard deviations. Such large uncertainties make the labelled data unreliable.

Despite these uncertainties, the tyre labelling system remains an important tool for consumers to select the best tyres. It should be emphasised that external noise is not the decisive criterion for drivers, but it is one of the factors considered (BÜHLMANN *et al.*, 2022). A survey conducted among consumers in Finland, France, Germany, Italy, Sweden, and the UK (VIEGAND, 2016) confirmed this fact. The results of this survey are shown in Fig. 1.

Rolling noise is the fourth most important criterion for consumers. The most important criteria for them are wet grip and price. This is also confirmed by the results of survey conducted by SANDBERG (2008), in which consumers indicated that wet grip was the most important factor in selecting tyres. It is also worth noting that the price of tyres is not correlated with their noise level (DITTRICH *et al.*, 2015; SANDBERG, 2008). Therefore, the decision to choose quieter tyres does not directly involve additional costs for consumers. This is an important argument in favour of selecting lower-noise tyres. Additionally, quieter tyres contribute to lower noise levels inside the vehicle, although the correlation in this case is not so high (BÜHLMANN *et al.*, 2022).

Reducing traffic noise through the use of low-noise tyres can be an effective protection measure. However, this requires ongoing and consistent awareness of the harmful impact of tyre/road noise on the population of the European Union. This awareness is closely linked to the efforts of non-governmental organizations (NGOs) and legislative actions taken by governments and road authorities. These measures could include: reduction or elimination of taxes on the purchase of the quietest tyres, allowing only vehicles equipped with

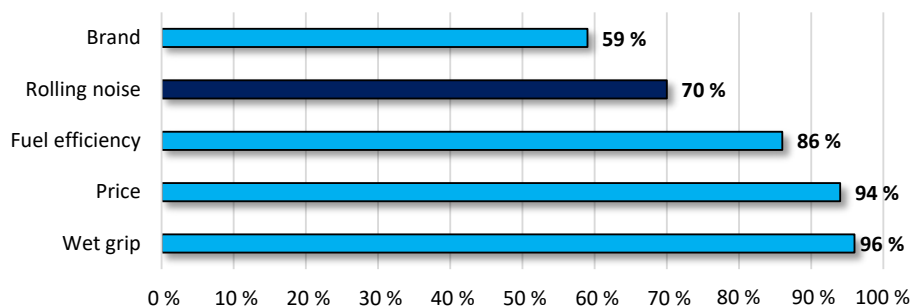


Fig. 1. Importance of specific information on tyre labels for consumers – percentage of respondents who consider the information very important or important (BÜHLMANN *et al.*, 2022; VIEGAND, 2016).

quieter tyres to enter selected urban areas (using appropriate chips) or requiring the use of quiet tyres in public administration fleets (BÜHLMANN *et al.*, 2022).

The tyre industry is also one of the major stakeholders in influencing the use of quieter tyres by consumers. Achieving this would require car tyre manufacturers to enter into an agreement or letter of intent to promote the sale of increasingly quieter tyres, while gradually withdrawing noisier tyres from sale. The STEER project (BÜHLMANN *et al.*, 2022) proposes that such an agreement should aim to ensure that the total noise level of all tyres sold does not exceed a predetermined threshold noise limit. Additionally, possible scenarios for reducing environmental noise were proposed in the ELANORE project (BOHATKIEWICZ *et al.*, 2024).

2. Methodology and input data

First, sound exposure level (SEL) measurements were conducted on both the ISO test track and trafficked sections of roads. A class 1 sound level meter was used, with the FAST time constant and a type A-weighting filter. Test results were stored in the instrument's memory at 1 s intervals. The sound level meter was calibrated with a class 1 acoustic calibrator before and after the measurements. The range of measurements covered four selected car tyres with theoretically different noise levels – their label data were: 67 dB(A), 69 dB(A), 71 dB(A), and 74 dB(A).

By comparing these values and the results obtained on the ISO tracks and trafficked roads, it was possible to identify the weaknesses of the procedure described in Regulation No. 117 (UN/ECE, 2011), in relation to environmental noise. To visualise these variabilities, the equivalent sound level (L_{eq}) for a sample road section was calculated.

The next step was to calculate the traffic noise level, with a focus on tyre/road noise. This was done using the CNOSSOS-EU model, which was calibrated using the measurement results, as described in detail in the later part of this section (see Eqs. (3) and (4)). Subsequently, traffic noise was calculated for different types of roads and road surroundings.

To determine the environmental impact of tyre noise (based on labelled data), calculations were made for selected traffic scenarios, using the information provided in the Nordic calculation model NORD 2000 (KRAGH *et al.*, 2006). Three vehicle categories are assumed in this model: light – cat. 1, medium – cat. 2, and heavy – cat. 3. Six scenarios were selected for further analysis, as shown in Table 2.

Then, an attempt was made to estimate the noise levels of the tyres currently used by drivers. For this purpose, the data presented in the STEER project report (BÜHLMANN *et al.*, 2022) were used, with the permission of the Swiss Federal Office for the Environment (FOEN). This is a database containing the C1 tyres approved for sale in Switzerland in 2021. Table 3 shows the number of tyres with a given sound level on the

Table 2. Traffic volume, composition, and vehicle speed on various types of roads based on NORD 2000 model assumptions (KRAGH *et al.*, 2006).

Traffic scenario	Description	Traffic volume [V/d]	Composition [%]			Speed [km/h]		
			Cat. 1	Cat. 2	Cat. 3	Cat. 1	Cat. 2	Cat. 3
A	Motorway	20 000	85	5	10	120	90	90
B	Urban motorway	30 000	85	5	10	90	85	85
C	Main road	15 000	85	10	5	85	75	75
D	Urban road	20 000	90	5	5	70	65	65
E	Feeder road in residential area	10 000	95	5	0	50	50	50
F	Residential road	5000	100	0	0	35	35	35

Table 3. Ranking of summer tyres approved for sale in 2021 in Switzerland (BÜHLMANN *et al.*, 2022).

Group	Sound level range [dB(A)]	Percentage share of tyres on the market [%]	Noise level on the label [dB(A)]	Total number of tyres (<i>n</i>)
Group 1	66–67	2.6	66	11
			67	168
Group 2	68–69	19.7	68	425
			69	967
Group 3	70–72	69.4	70	1881
			71	1695
			72	1326
Group 4	73–75	8.3	73	451
			74	57
			75	81

Source of data: Touring Club Switzerland, financed by the FOEN, <https://www.bafu.admin.ch/bafu/en/home.html>.

label that were approved for sale. It should be emphasised that these figures relate to summer tyres only. The data were aggregated into four groups with different noise levels.

With the data presented in Table 3, the average sound level was calculated using weighted logarithmic averaging:

$$L_{\text{avg}} = 10 \cdot \log \frac{\sum_{i=1}^n (10^{0.1 \cdot L_i} \cdot n_i)}{n} \text{ [dB(A)]}, \quad (1)$$

where L_{avg} – weighted average sound level [dB(A)], i – sound level value marked on the tyre label [dB(A)], L_i – sound level determined for a tyre with noise value i on the label [dB(A)], n_i – number of tyres with noise values i on the label, n – total number of tyres.

Under these assumptions, the calculated average sound level was 70.8 dB(A). This value was used as the reference level. Then, four different scenarios for improving the acoustic conditions in the road surroundings were identified. One of these scenarios involves withdrawing the noisiest tyres from the market. It is worth noting that some tyres currently available have sound levels that are above or equal to the limits (BÜHLMANN *et al.*, 2022). To determine the impact of this measure, the weighted average sound level was recalculated, considering only those tyres with a sound level that does not exceed the permissible limits. In this case, the sound level is reduced from 70.8 dB(A) to 70.3 dB(A).

A greater reduction in traffic road noise could be achieved if tyres with noise levels equal to the existing limits were also withdrawn from the market. However, this could be resisted by manufacturers and the automotive industry. After recalculating the weighted average sound level, a value of 69.7 dB(A) was obtained, indicating a noise reduction of 1.1 dB compared to the current situation.

To achieve a greater reduction, it is necessary to take measures to promote quieter tyres among vehicle owners. It was assumed that tyres with noise levels above or equal to the limits would be withdrawn from sale, and the percentage of quieter tyres would increase at the expense of noisier tyres. Two scenarios were assumed. The first was referred to as the sustainable scenario, and the second, the optimistic scenario. The percentages of the individual tyre groups in these scenarios are shown in Table 4.

In the first scenario (sustainable), the weighted average sound level was 69.1 dB(A), resulting in a noise reduction of 1.7 dB. In the optimistic scenario, the average level was 68.5 dB(A). In this case, a reduction in noise level was 2.4 dB.

It should be emphasised that these results were based on sound level calculations, which show the effect of the noise reduction, but do not account for the variability in traffic parameters (such as traffic vol-

Table 4. Percentage of tyres for each group under the sustainable and optimistic scenarios.

Group	Percentage share of tyres in sustainable scenario [%]	Percentage share of tyres in optimistic scenario [%]
Group 1 [66 dB(A) – 67 dB(A)]	10	15
Group 2 [68 dB(A) – 69 dB(A)]	55	65
Group 3 [70 dB(A) – 72 dB(A)]	35	20
Group 4 [73 dB(A) – 75 dB(A)]	0	0

ume, vehicle speeds, and traffic composition), which affect noise levels. The impact of these parameters was considered in the noise modeling carried out with the CNOSSOS-EU model. In the first step, a calibration of the model was performed for light vehicles (cat. 1) by incorporating an additional factor. Calibration was not conducted for the other vehicle categories (medium and heavy vehicles), because they were not the object of the study.

To calibrate the model to account for the influence of tyre noise, the CNOSSOS-EU relationship for rolling sound power level calculations was used. For this purpose, light vehicles were assumed to move at a speed v_m of 80 km/h (the reference speed for determining the labelled sound level for C1 tyres). An additional correction factor ΔL_{tyre} was included in the equation, which determines the effect of the noise level of the car tyres, as shown in the equation:

$$L_{WR,i,m} = A_{R,i,m} + B_{R,i,m} \cdot \log \left(\frac{v_m}{v_{\text{ref}}} \right) + \Delta L_{WR,i,m} + \Delta L_{\text{tyre}} \text{ [dB(A)]}, \quad (2)$$

where $L_{WR,i,m}$ – rolling sound power level [dB(A)], $A_{R,i,m}$ and $B_{R,i,m}$ – coefficients given in the frequency bands for each vehicle category and reference speed [–], v_m – average speed of vehicles in category m (equal to 80 km/h) [km/h], v_{ref} – reference speed, equal to 70 km/h, $\Delta L_{WR,i,m}$ – sum of the correction factors for rolling noise emissions in specific road conditions or for specific vehicles (influence of road surface, studded tyres, traffic lights or junction, temperature) [dB(A)], ΔL_{tyre} – correction factor for the impact of tyre noise [dB(A)].

The ΔL_{tyre} factor in the CNOSSOS-EU model can be assumed for each $1/1$ octave frequency band separately (from 63 Hz to 8000 Hz). In this study, the same value is used for each sound frequency. This assumption does not significantly affect the calculation results.

Tyre noise measurements (made using the procedure defined in Regulation No. 117 (UN/ECE, 2011)) and CNOSSOS-EU algorithms consider two sources of noise: rolling noise and powertrain noise. At speeds

of 70 km/h to 90 km/h, at which the C1 tyre tests are conducted, the contribution from powertrain noise is small (see Table 1), but it is still present. Therefore, the measurement results include both tyre/road noise and powertrain noise. Similarly, the CNOSSOS-EU model includes both sound sources, as expressed in the following model algorithm:

$$L_{W,i,m}(v_m) = 10 \cdot \log \left(10^{\frac{L_{WR,i,m}(v_m)}{10}} + 10^{\frac{L_{WP,i,m}(v_m)}{10}} \right) \text{ [dB(A)/m]}, \quad (3)$$

where $L_{W,i,m}$ – directional sound power of one vehicle in category m in the frequency range i (125 Hz to 4 kHz) [dB(A)], $L_{WR,i,m}$ – rolling sound power level [dB(A)], $L_{WP,i,m}$ – sound power level of the propulsion unit noise [dB(A)], v_m – average speed of vehicles in category m [km/h].

The calibration of the CNOSSOS-EU model involved adjusting the correction factor ΔL_{tyre} in such a way that the directional sound power for cat. 1 vehicles across the entire frequency range changed by exactly the amount indicated by the results of the Regulation No. 117 tests (UN/ECE, 2011). This relationship was calculated by regression analysis and is as follows:

$$\Delta L_{W,1}(v_{m=80 \text{ km/h}}) = 0.70 \cdot \Delta L_{\text{tyre}} + 0.06 \text{ [dB(A)]}, \quad (4)$$

where $\Delta L_{W,1}$ – variation in the sound power of cat. 1 vehicles across the entire frequency range [dB(A)], v_m – average speed of cat. 1 vehicles, equal to 80 km/h, ΔL_{tyre} – correction factor for the impact of tyre noise [dB(A)].

These relationships were obtained using the CNOSSOS-EU method, but they can also be calculated using other methods. The results obtained with contemporary models do not differ significantly (HALUCHA, 2023), so the ΔL_{tyre} factor from the CNOSSOS-EU model can also be used directly for other models.

Next, a cost-benefit analysis for selected EU countries was conducted. Noise exposure data for the population, derived from the strategic noise maps, were used for the analyses. These data were taken from (European Environment Agency [EEA], 2024).

First, the number of people exposed to day-evening-night noise (L_{DEN}) levels greater than 55 dB(A) was calculated. The data reported by EU member states after the 2016 strategic noise mapping was used as the baseline scenario. Next, it was calculated how many people would be exposed to the same noise level after the introduction of the previously described scenarios. It should be noted that the data provided by the EEA is divided into 5 dB intervals. The first interval identifies the number of people exposed to noise levels between 55 dB(A)–59 dB(A), and the

last interval to noise levels greater than 75 dB(A). To calculate the number of people exposed to noise within each range after implementing the successive scenarios, it was necessary to approximate the data to narrower 0.1 dB intervals. This approximation was done as accurately as possible, however, the lack of knowledge about the original distribution of people across the 0.1 dB ranges introduces additional uncertainty into the analyses. Nevertheless, this uncertainty is assumed to be negligible.

The number of people exposed to L_{DEN} levels greater than 55 dB(A) was calculated, and the financial benefits of reducing the population exposed to noise were then determined. For this purpose, the environmental costs described in the Handbook on the External Costs of Transport (European Commission, 2020) were used. These costs are related to the annoyance experienced by people exposed to specific noise ranges and the associated health effects. The costs were estimated for 2016, so it is expected that the financial results will be slightly underestimated considering the current situation (2024), particularly due to the high inflation experienced in most EU countries.

3. Impact of surface on tyre labelling in the environmental noise context

One of the main sources of uncertainty in the results of tyre labelling (and often a reason the data on labels may be unrealistic) is the surface on which the tests are conducted as specified in accordance with Regulation No. 117 (SANDBERG, MIODUSZEWSKI, 2022). It is a specific surface (very smooth) meeting the requirements of the ISO (2021) standard. This issue becomes evident when comparing measurement results for four selected car tyres. First, the results of tests on the ISO test tracks are presented and compared with the data on the labels, which is shown in Table 5.

Measurements were taken on four different test tracks, with tyres 1 and 2 being tested on only two tracks due to unfavourable meteorological conditions that prevented additional tests. The procedure used was described in Regulation No. 117 (UN/ECE, 2011), with all requirements met. The test car was driven at speeds ranging from 70 km/h to 90 km/h. All pass-by noise levels were measured using a sound analyser, two microphones with preamplifiers, a laptop computer, an external radar and a light barrier, all of which held valid calibration certificates.

The average sound level calculated using the data on the labels differs from the sound level derived from real measurements on the ISO test tracks by just 0.3 dB, which is not significant. However, the variability between individual tyres is much more substantial, with differences of up to 3.0 dB for tyres 1, 2, and 4. This shows how unrealistic the data on the labels are.

Table 5. Comparison of the *A*-weighted average sound level calculated from the label data and the results of measurements on the ISO test tracks.

Tyre	Label values [dB(A)]	Sound level measured on the ISO test tracks [dB(A)]	Calculated label value [dB(A)]	Difference between label and calculated values [dB]
Tyre 1	67	71.4	70	3
Tyre 2	69	73.6	72	3
Tyre 3	71	73.3	72	1
Tyre 4	74	72.9	71	–3
Weighted average sound level	71.0	72.9	71.3	0.3

When using this labelled data for acoustic calculations, it is important to be aware of the significant inaccuracies. This is shown in Fig. 2, which illustrates the results of calculations based on both label data and test data. The calculations were made for an example motorway section (traffic scenario A) and expressed by an equivalent sound level of 60 dB(A).

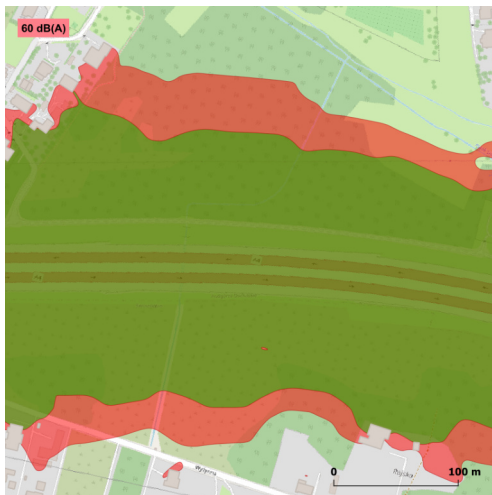


Fig. 2. Results of acoustic calculations using label data (green) and measurement values (red).

These differences reflect the results of measurements conducted strictly according to the Regulation No. 117 procedure on a surface that meets the requirements of ISO 10844 (UN/ECE, 2011). This surface has significantly different acoustic characteristics from those found on trafficked roads. As a result, this differences also impact the sound levels in the environment. This can be observed by comparing the results of measurements made for the same tyres on ISO tracks and typical road surfaces used on trafficked roads (MA11, SMA8, SMA11, SMA16, EACC). These data are shown in Table 6.

The variability range of weighted average sound level from 2.1 dB to 4.2 dB is very high. This can be also seen in Fig. 3, which shows the results of equivalent sound level calculations for the same section of motorway.

Table 6. Comparison of noise levels measured on ISO and typical road surfaces.

Tyre	Sound level measured according to Regulation No. 117 [dB(A)]					
	ISO	MA11	SMA8	SMA11	SMA16	EACC
Tyre 1	71.4	74.6	75.0	76.5	77.7	76.7
Tyre 2	73.6	75.2	76.0	77.0	76.5	75.8
Tyre 3	73.3	75.3	76.4	77.0	76.4	76.3
Tyre 4	72.9	75.0	75.6	76.9	77.8	77.0
Weighted average sound level	72.9	75.0	75.8	76.9	77.1	76.5

Explanations:

- ISO: surface meeting the requirements of the ISO 10844 (UN/ECE, 2011);
- MA11: a Norwegian term for a “soft asphalt” / dense surface with an 11 mm maximum chipping size designed for low traffic volume;
- SMA8, SMA11, SMA16: stone mastic asphalt with maximum chipping sizes of 8 mm, 11 mm, and 16 mm, respectively;
- EACC: exposed aggregate cement concrete.

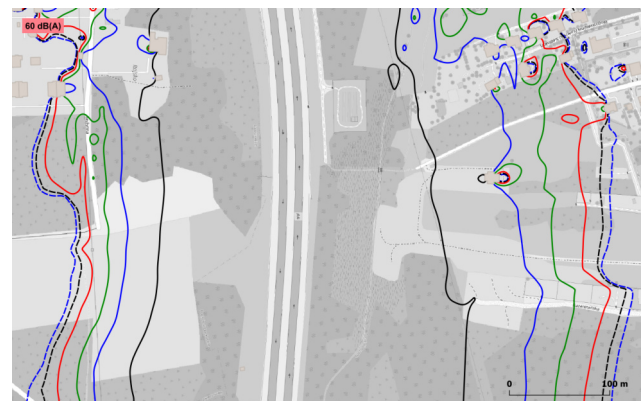


Fig. 3. Comparison of noise levels measured on ISO surfaces and typical road surfaces (black line – ISO, blue line – MA11, green line – SMA8, red line – EACC, black dashed line – SMA11, blue dashed line – SMA16).

The sound level calculated for the ISO surface is significantly lower than that for all other real surfaces. The lowest variability is observed for MA11, though it is not widely used (it is used in Norway on roads with

very low traffic). Noise levels for the other surfaces, especially for the SMA surfaces (used in many European countries), are much higher than those of the ISO surface currently used for tyre labelling.

An additional problem is the varying ranking of tyres depending on the road surface on which the tests are conducted. For example, tyre 1 is quieter than tyre 2 on the smoother surfaces (such as MA11, SMA8, and SMA11), but noisier on rougher ones (such as SMA16 and EACC) – see Table 6. For this reason, it can be very difficult to choose tyres that consistently produce the lowest noise levels on all surfaces. This challenge would also arise if the reference surface for tyre labelling were changed from the current ISO surface to one of the real-world surfaces.

4. Results of acoustic calculations and cost-benefit analyses for different noise mitigation scenarios

First, it was calculated how the noise levels in the surroundings of different road sections (A – motorways, B – urban motorways, C – main roads, D – urban roads, E – feeder roads, F – residential roads) would

be affected by the withdrawal of tyres with noise levels above the legal limit. The noise reduction varied from 0.1 dB up to 0.4 dB depending on the traffic scenario (with the greatest reduction observed on motorways). A greater improvement (from 0.3 dB up to 0.9 dB) was found when tyres with noise levels equal to or above the limit were withdrawn from the market. The results of the calculations are shown in Fig. 4.

Further noise calculations considered the effects of promotional activities aimed at encouraging vehicle owners to choose quieter tyres (see Table 4). The results of these calculations are presented in Fig. 5.

These findings are also illustrated in Fig. 6, which show the results of calculations for individual traffic scenarios on selected road sections in Poland. It shows the differences between the most optimistic scenario (in purple) and the current situation where no actions have been taken (depicted in red). For graphical representation, an isophone of 60 dB(A) was used for traffic scenarios A–E and 55 dB(A) for traffic scenarios F, where the noise level in the road vicinity was below 60 dB(A).

The greatest reduction in noise is observed on motorways, where vehicles travel at the highest speeds. For other types of roads, the improvement is smaller

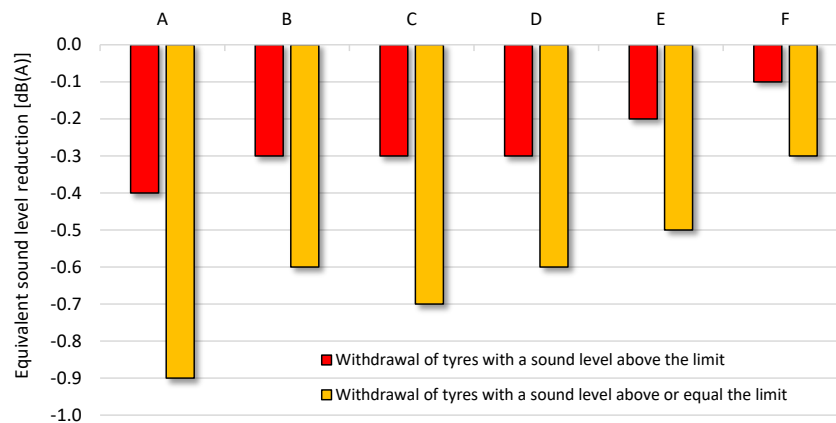


Fig. 4. Reduction in the equivalent sound level after the withdrawal of tyres with noise levels equal to or above the limit.

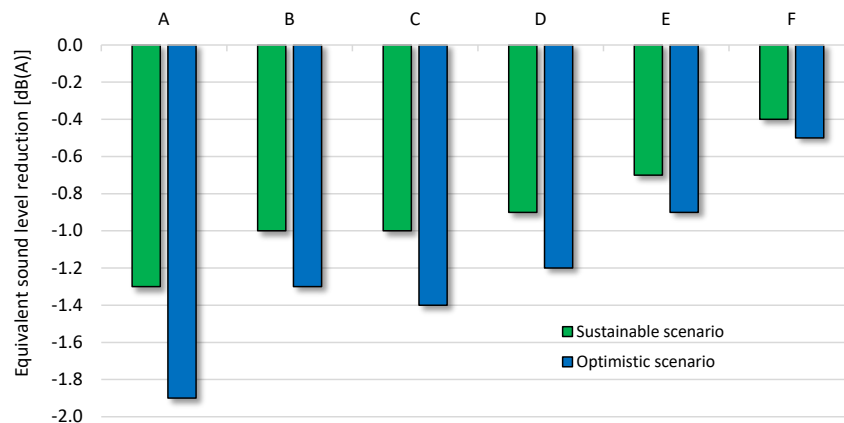


Fig. 5. Reduction in the equivalent sound level considering the promotion of quiet tyres in both the sustainable and optimistic scenarios.

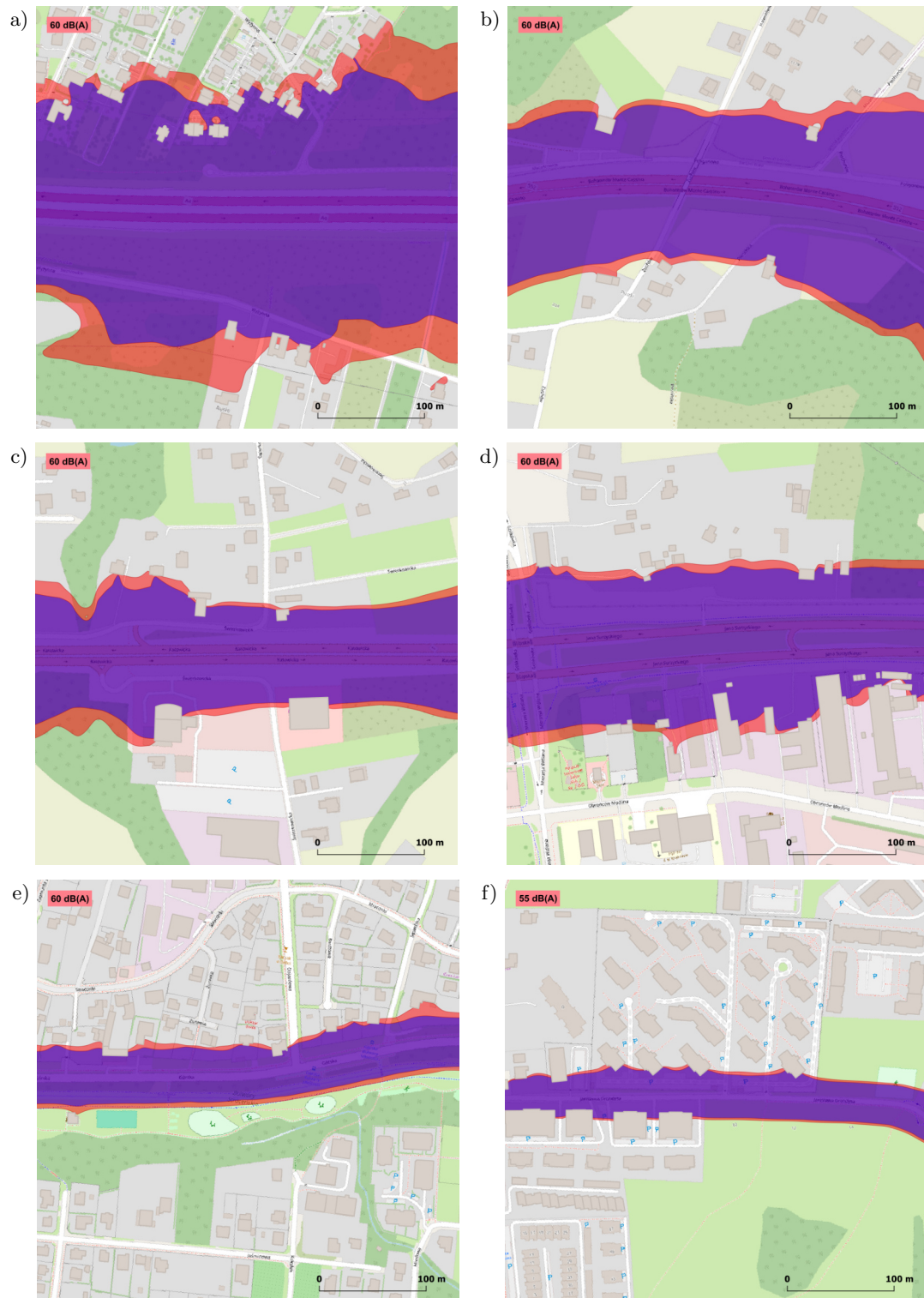


Fig. 6. Reduction in the equivalent sound level for traffic: a) case A – motorway; b) case B – urban motorway; c) case C – main road; d) case D – urban road; e) case E – feeder road; f) case F – residential road.

and it depends on the speed of light vehicles and traffic composition. While the reductions are generally smaller than the measurement uncertainty of ± 1.2 dB, they still demonstrate the potential impact these measures can have on environmental noise.

A greater improvement is observed when tyres with sound levels exceeding or equal to the permissible limits are withdrawn from the market. In the case of motorways, this reduction was almost 1 dB. From an environmental point of view, this is a noticeable im-

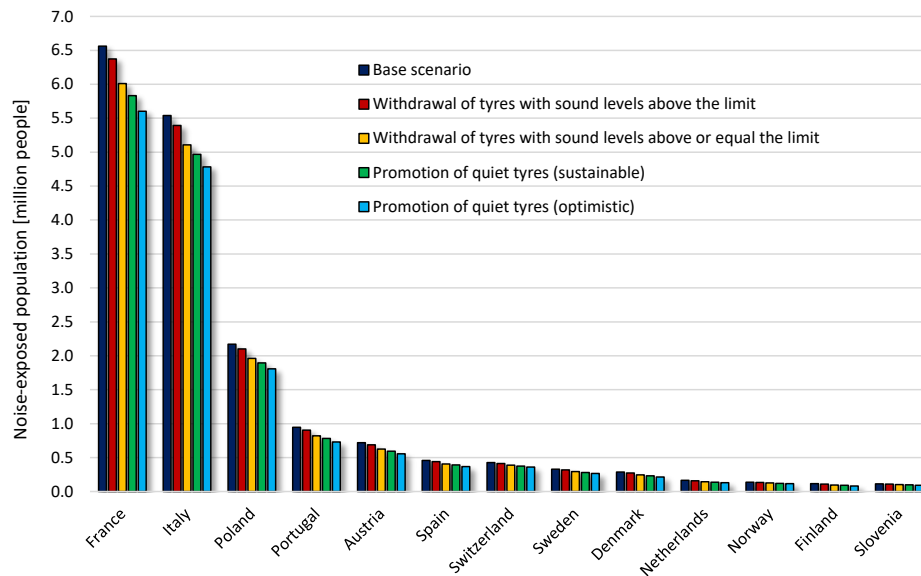


Fig. 7. Comparison of the population exposed to noise above 55 dB(A) in the baseline and noise reduction scenarios (main roads in selected EU countries).

provement. For other types of roads, excluding residential roads, the noise reduction ranges from 0.5 dB to 0.7 dB.

To achieve better results in reducing noise in the road vicinity, further efforts are needed to promote the use of quiet tyres by consumers. In an optimistic scenario, the noise reduction could be significant (over 1.8 dB for motorways). For other roads, the noise reduction is significant, but still noticeable for those living nearby. In all cases, except residential roads, the noise reduction would be greater than 1.0 dB.

Based on the results of noise calculations and the population exposed to noise levels greater than 55 dB(A), it was calculated how the tyre/road noise reduction scenarios would improve the acoustic con-

ditions in the road environment. These improvements are shown in Fig. 7 for main roads in selected EU countries.

The most effective measures are those outlined in the strategies, which include the withdrawal of the noisiest tyres and the promotion of the quietest tyres. In these cases, the reduction in the number of people exposed to noise is significant and noticeable. The introduction of the other strategies also yields a desired effect, although not so high, but still measurable.

The financial benefits were calculated based on the variability of the environmental costs in the baseline scenario and the noise reduction scenarios. These benefits are presented in Fig. 8 showing the gains for the country concerned over a one-year period.

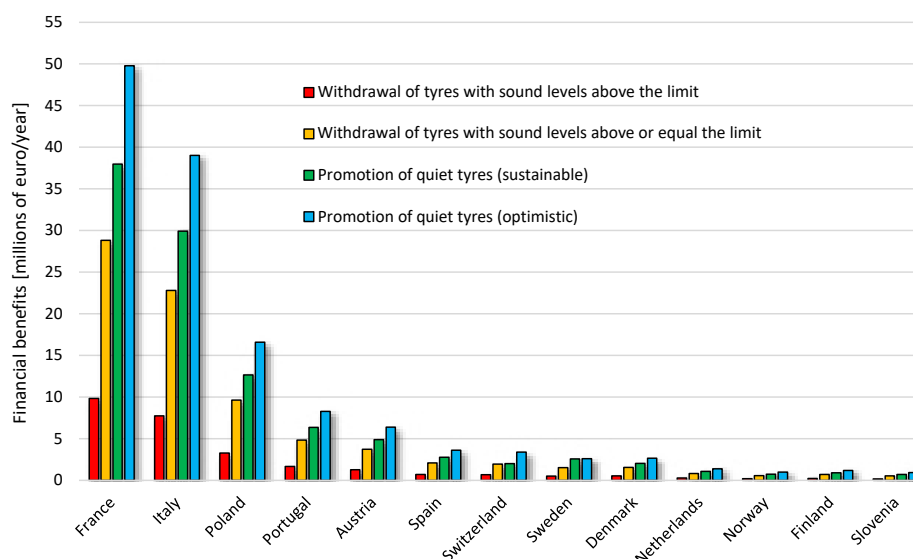


Fig. 8. Financial benefits for one year after implementation of noise reduction scenarios.

The introduction of the analysed scenarios can bring significant financial benefits. For the selected countries, these benefits could amount, in optimistic scenario, to almost €50 million for France, almost €40 million for Italy and more than €15 million for Poland. It should be highlighted that these are benefits for a one-year period, which will be proportionally multiplied in the long term.

The financial benefits were calculated for major roads outside urban agglomerations. No less important are the roads within cities, which were not included in these analyses. In these cases, the noise reduction associated with the use of quiet tyres will be lower due to the lower speeds of cars. However, an improvement in acoustic conditions will still be observed in the surroundings of main roads and motorways in cities. In the ELANORE technical report (BOHATKIEWICZ *et al.*, 2024), financial benefits were also estimated for selected cities. For example, the annual benefit for Rome is almost €6 million, for Budapest it is more than €4.5 million, and for Prague it is more than €4 million per one year.

Promoting quiet tyres to consumers also incurs costs. At present, it is not possible to make a precise estimate of these costs, because measures to promote low-noise tyres can be implemented on different scales. The necessary financial effort will depend on the scale of the measures taken; however, the costs will certainly be far lower than the financial benefits.

5. Summary

Decreasing the noise level of vehicle tyres is an effective measure to improve environmental acoustic conditions. This is especially important because there is the increasing number of electric vehicles on the road, for which tyre/road noise is the most important source of sound. Encouraging consumers to use low-noise tyres can lead to a considerable reduction in environmental noise. However, it is essential that the data on the labels must be accurate and reflect the noise characteristics of tyres on surfaces commonly used on roads.

The procedure described in Regulation No. 117 (UN/ECE, 2011) is currently used for tyre labelling. However, it is characterised by high uncertainties due to, e.g., the influence of the road surface on which the tyres are tested (along with other factors not studied in the article, including variations in test tyres, the influence of the test vehicle, meteorological conditions, and more). The results of testing four selected car tyres using this procedure indicated differences between the label data and the calculated values based on measurements from the ISO test track. The variability of the weighted average sound level was 0.3 dB, which is not a large difference. More importantly, the differences for individual tyres, in some cases, reached

up to 3.0 dB. This shows the inaccuracy of the current label data, which very often fail to reflect the real noise level of the tyres.

The results of measurements and calculations show that tyre noise levels vary according to the road surface. First, it should be emphasised that the ISO surface used for the labelling has acoustic characteristics that differ significantly from those of other surfaces used on trafficked roads. The weighted average sound level calculated for the four tested tyres tested on the ISO surface differs from that on the other pavements by from 2.1 dB to 4.2 dB. In each case, the sound level measured on the test track is lower than that measured on the real road sections. The smallest variability was observed for the MA11 pavement (a very smooth asphalt surface), which is not widely used on roads in European countries. The variability between the sound level measured on the ISO and rougher pavements (e.g., SMA11 or SMA16) is more than 4.0 dB. From an environmental perspective, this is a very large discrepancy.

More important is the fact that the same tyres produce different noise levels on different real surfaces. The maximum variability of the weighted average sound level is 2.1 dB (between MA11 and SMA16). The ranking of tyres also varies depending on the road pavement. For example, tyre 1 is quieter than tyre 2 on smoother surfaces (MA11, SMA8, and SMA11) but noisier on rougher ones (SMA16 and EACC). This has a direct impact on the precision of environmental noise calculations. These results indicate that vehicle tyre labels are biased by additional inaccuracies due to the varying characteristics of typical road surfaces. The same tyre may be quieter on one road surface and noisier on another.

It is not possible to eliminate most of the uncertainty components of the current procedure. Therefore, replacing it with another measurement method should be considered. For example, a laboratory method using drums equipped with a replica of the road surface appears to be a promising direction. Similar methods are already used to measure tyre rolling resistance. Consideration should also be given to equipping these drums with replicas of real pavements (e.g., SMA11 or AC11), which are widely used in most EU countries.

Despite these differences, efforts should be made to reduce the noise level of tyres and to promote those with low noise on the most widely used surfaces. The results of equivalent sound level calculations for selected road sections, varying traffic parameters (from motorway to residential road), showed that this is an effective noise reduction measure. Withdrawing vehicle tyres from the market with sound levels above the permissible limits can reduce noise by 0.4 dB on motorways to 0.1 dB on residential roads. This reduction could be significantly increased by lowering the permissible limits and promoting low-noise tyres to con-

sumers. In this case (optimistic scenario) environmental noise could be reduced by 1.8 dB on motorways to 0.5 dB on residential roads. For all other road categories in this scenario, the noise reduction is greater than 1.0 dB. This is a significant improvement in the acoustic conditions around roads. In addition, it is a source-based action, which is always characterised by high efficiency.

Decreasing environmental noise exposure also results in a reduction in the number of affected people. Based on data taken from the strategic noise maps, it was calculated how many fewer people would be exposed to noise levels in the 55 dB(A) noise range after the introduction of the measures described in the article. For the countries with the largest populations exposed to adverse noise impacts (among those selected for the analyses), highly beneficial effects were observed with the implementation of the different strategies. The most prominent examples are France and Italy, where the number of people exposed to noise above 55 dB(A) can be reduced by almost one million people. Reducing the exposure of the population to excessive noise brings significant financial benefits. These are estimated at nearly €50 million for France, almost €40 million for Italy, and more than €15 million for Poland. These benefits are for a one-year period, which will be multiplied proportionally in the long term. Reducing the noise of car tyres is thus justified from an economic point of view as well.

The use of low-noise tyres is very important in terms of environmental protection. Withdrawing the noisiest tyres from the market and promoting low-noise tyres can significantly reduce environmental noise. A necessary condition for achieving this is to improve the labelling system for car tyres so that the data presented on the labels are as realistic as possible.

Acknowledgments

The work for the ELANORE project was conducted under the programme “Applied Research under the Norwegian Financial Mechanisms of 2014–2019 between Norway Grants and NCB”, under the contract no. NOR/POLNOR/ELANORE/0001/2019-00. The project was conducted by a consortium consisting of the Gdańsk University of Technology, SINTEF, and EKKOM.

References

1. BERGE T. (2023), Technical report on CPX, CPB and SEL measurements. Discussion on selected noise results for improvement of the test method, *Elanore*, <https://elanore.mech.pg.gda.pl/en/documents/open-documents>.
2. BERGE T., MIODUSZEWSKI P., BOHATKIEWICZ J., HAŁUCHA M. (2022), Final report on the noise measurement on ISO reference surface and on conventional pavements, *Elanore*, <https://elanore.mech.pg.gda.pl/en/documents/open-documents>.
3. BOHATKIEWICZ J., HAŁUCHA M. (2017), The impact of quiet pavements' usage on traffic noise on people in loosely built-up areas, [in:] *Traffic Noise: Exposure, Health Effects and Mitigation*, Łucjan C., Gerard D. [Eds], pp. 105–205, Nova Science Publishers, New York.
4. BOHATKIEWICZ J., HAŁUCHA M., DĘBIŃSKI M., JUKOWSKI M., TABOR Z. (2022), Investigation of acoustic properties of different types of low-noise road surfacers under in situ and laboratory conditions, *Materials*, **15**(2): 480, <https://doi.org/10.3390/ma15020480>.
5. BOHATKIEWICZ J., HAŁUCHA M., ŚWIĄTEK Ł., DRACH M., TYKSIŃSKA G., WOŹNIAK Ł. (2024), Impact of the new tyre labelling method on the analyses of environmental noise in the vicinity of a road (Technical Report TR19-ELANORE-EKKOM-01), *Elanore*, <https://elanore.mech.pg.gda.pl/en/documents/open-documents>.
6. BÜHLMANN E., SANDBERG U., BERGE T., GOUBERT L., SCHLATTER F. (2022), *Call 2018 Noise and Nuisance STEER Project Final Report, CEDR Contractor Report 2022-07*, <https://www.cedr.eu/docs/view/6373a6fec0dc7-en> (access: 15.04.2024).
7. DITTRICH M.G., ROO F. de, ZYL S. van, JANSEN S.T.H., GRAAFF E. de (2015), Triple A tyres for cost-effective noise reduction in Europe, [in:] *Proceedings of Euro-Noise 2015*, pp. 2607–2612.
8. European Commission (2020), Handbook on the external costs of transport, Version 2019 1.1, <https://op.europa.eu/en/publication-detail/-/publication/9781f65f-8448-11ea-bf12-01aa75ed71a1/language-en> (access: 15.04.2024).
9. European Environment Agency [EEA] (2024), Noise data reported under Environmental Noise Directive (END), <https://www.eea.europa.eu/en/datahub/datahubitem-view/c952f520-8d71-42c9-b74c-b7eb002f939b> (access: 27.02.2024).
10. European Union (2009), Regulation (EC) No. 1222/2009 of the European Parliament and of the Council of 25 November 2009 on the labelling of tyres with respect to fuel efficiency and other essential parameters, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32009R1222> (access: 15.04.2024).
11. European Union (2020), Regulation (EU) 2020/740 of the European Parliament and of the Council of 25 May 2020 on the labelling of tyres with respect to fuel efficiency and other parameters, amending Regulation (EU) 2017/1369 and repealing Regulation (EC) No. 1222/2009, <https://eur-lex.europa.eu/eli/reg/2020/740/oj> (access: 15.04.2024).
12. HAŁUCHA M. (2023), Differences in the results of road noise calculations made using different models used in European countries, [in:] *The Proceedings from the 27th World Road Congress*, Prague, Czech Republic.

13. HALUCHA M., BOHATKIEWICZ J., MIODUSZEWSKI P. (2023), Modelling the effect of electric vehicles on noise levels in the vicinity of rural road sections, *Archives of Civil Engineering*, **69**(3): 573–586, <https://doi.org/10.24425/ace.2023.146098>.
14. HEUTSCHI K., LOCHER B. (2018), *SonROAD18 – Calculation model for road noise* [in German: SonROAD18 – Berechnungsmodell für Strassenlärm].
15. International Standard Organisation [ISO] (2021), *Acoustics – Specification of test tracks for measuring sound emitted by road vehicles and their tyres* (ISO Standard No. 10844:2021), <https://www.iso.org/standard/80557.html>.
16. KRAGH J., JONASSON H., PLOVSING B., SARINEN A., STOREHEIER S.Å., TARALDSEN G. (2006), User's Guide Nord2000 Road, <https://forcetechnology.com/-/media/force-technology-media/pdf-files/projects/nord2000/nord2000-users-guide-road.pdf> (access: 15.04.2024).
17. SANDBERG U. (2008), Consumer label for tyres in Europe (Report), Swedish National Road and Transport Institute (VTI), Linköping, Sweden, <https://www.transportenvironment.org/discover/consumer-label-tyres-europe/>.
18. SANDBERG U., EJSMONT J.A. (2002), *Tire/Road Noise Reference Book*, Informex, Sweden.
19. SANDBERG U., MIODUSZEWSKI P. (2022), The EU tyre noise label: The problem with measuring the noise level of only a few of all tyre variants, [in:] *Proceedings of Inter-Noise 2022*.
20. United Nations Economic Commission for Europe [UN/ECE] (2011), Provisions concerning the approval of tyres with regard to rolling sound emissions and to adhesion on wet surfaces and/or to rolling resistance, Regulation No. 117 of the Economic Commission for Europe of the United Nations.
21. VIEGAND M. (2016), Final report-review study on the Regulation (EC) No. 1222/2009 on the labelling of tyres, https://energy.ec.europa.eu/publications/review-study-regulation-ec-no-12222009-labelling-tyres_en.

Research Paper

The Influence of the Amplitude Spectrum Correction in the HFCC Parametrization on the Quality of Speech Signal Frame Classification

Stanisław GMYREK*, Robert HOSSA, Ryszard MAKOWSKI

*Faculty of Electronics, Photonics and Microsystems, Department of Acoustics, Multimedia and Signal Processing
Wrocław University of Science and Technology*

Wrocław, Poland; e-mails: robert.hossa@pwr.edu.pl, ryszard.makowski@pwr.edu.pl

*Corresponding Author e-mail: stanislaw.gmyrek@pwr.edu.pl

(received July 8, 2022; accepted December 11, 2024; published online February 28, 2025)

The voiced parts of the speech signal are shaped by glottal pulse excitation, the vocal tract, and the speaker's lips. Semantic information contained in speech is shaped mainly by the vocal tract. Unfortunately, the quasiperiodicity of the glottal excitation, in the case of the HFCC parameterization, is one of the factors affecting the significant scatter of the feature vector values by introducing ripples into the amplitude spectrum. This paper proposes a method to reduce the effect of quasiperiodicity of the excitation on the feature vector. For this purpose, blind deconvolution was used to determine the vocal tract transfer function estimator and the corrective function of the amplitude spectrum. Subsequently, on the basis of the obtained HFCC parameters, statistical models of individual Polish speech phonemes were developed in the form of mixtures of Gaussian distributions, and the influence of the correction on the quality of classification of speech frames containing Polish vowels was considered in details. The aim of the introduced solution was to narrow the GMM distributions, which clearly, according to the detection theory, reduces classification errors. The results obtained confirm the effectiveness of the proposed method.

Keywords: automatic speech recognition; robust parametrization; amplitude spectrum correction; inverse filtering; GMM model; distance between GMM distributions.



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

1. Introduction

In automatic speech recognition (ASR) systems, there is a need to compensate for the influence of many factors, such as recording conditions, interpersonal variability, contextuality, etc., which negatively affect the performance of the system. The most widely used compensation methods are (MAKOWSKI, 2011):

- 1) clustering with developing independent statistical models for speakers with similar personal characteristics (HOSSA, MAKOWSKI, 2016);
- 2) normalisation, which involves modifying the values of parametrization coefficients (PRASAD, UMESH, 2013);
- 3) adaptation, involving changing the parameter values of statistical models (ZAMBRZYCKA, 2021);
- 4) robust parametrization (MRÓWKA, MAKOWSKI, 2008), which should make the parameter vector

robust to the factors mentioned above or at least reduce their impact.

The present work stands for the robust parametrization.

Among at least a dozen different parametrization methods available in the literature (SHARMA *et al.*, 2020), the most commonly used and effective solutions in practical applications include methods that use short time-frequency transformations and cepstral representations of the resulting coefficients. To this group of solutions we can include the algorithms:

- Mel-frequency cepstral coefficients, MFCC (DAVIS, MERMELSTEIN, 1980);
- human factor cepstral coefficients, HFCC (SKOWRONSKI, HARRIS, 2003);
- the basilar-membrane frequency-band cepstral coefficient, BFCC (KUAN *et al.*, 2016);

- the gammatone cepstral coefficient, GTCC (YIN *et al.*, 2011).

On the other hand, the second group of solutions are algorithms using linear prediction methods and examples of their implementations are the parametrizations:

- linear prediction cepstral coefficients, LPCC (RABINER JUANG, 1993);
- the perceptual linear prediction, PLP (HERMAN-SKY, 1990).

Most of the aforementioned parametrizations naturally have mechanisms for robustness against small noise interference, which can be further enhanced by supplementing the method with the relative spectral (RASTA) algorithm to suppress those of the components that are not related to speech articulation. Based on such an idea, the RASTA-PLP hybrid algorithm (KOEHLER *et al.*, 1994) and the multi-resolution RASTA filtering solution (HERMAN-SKY, FOUSEK, 2005) were developed. Another equivalent representation in the form of the amplitude modulation filter bank (AMFB) has been considered in (MORITZ, KOLLMEIER, 2015). Among the robust parametrization algorithms, we can also distinguish algorithms based on the minimum variance distortionless response (MVDR) the estimator proposed in (MURTHI, RAO, 2000) and further developed into the MVDR-MFCC algorithm in (DHARANIPRAGADA, RAO, 2001).

In general, the voiced parts of the speech signal are shaped by linear cascade without interactions of the glottal pulse excitation, the vocal tract, and the speaker's lips (QUATIERI, 2002). Hence a widely accepted source-filter model of speech production is of the form

$$s(n) = x(n) \star h(n) \star r(n), \quad (1)$$

where $x(n)$ is the excitation, $h(n)$ is the impulse response of the vocal tract, $r(n)$ is the impulse response characterizing the sound emission by the lips, n is the discrete time, and \star is the discrete time convolution operator.

The semantic information contained in speech is mainly shaped by the vocal tract. Unfortunately, the quasiperiodicity of the glottal excitation, in the case of parametrizations based on different time-frequency representations, e.g., MFCC or HFCC, is one of the factors affecting the significant scatter of the feature vector values, by introducing ripples into the amplitude spectrum (see Sec. 2). Furthermore, in (SKOWRONSKI, HARRIS, 2003) it was shown that the HFCC parametrization is characterized by greater robustness to noise than the MFCC and studies have shown differences in recognition performance of up to 30 %. As a result, the classical solution, i.e., the HFCC parametrization, was selected as the representative for further research on ripple reduction.

The paper proposes an algorithm to reduce the impact of glottal flow excitation through its filtering op-

eration. The first step is to estimate the glottal excitation signal $x(n)$ and then determine the HFCC coefficients based on the magnitude of the vocal tract transfer function. The estimation of the excitation signal is one of the most important problems in speech signal processing, and in practical applications it is used, among others, for speaker recognition (PLUMPE *et al.*, 1999), analysis of the speaker's emotional state (WAARAMA *et al.*, 2010) or speech synthesis (RAITIO *et al.*, 2011). Inverse filtering algorithms are most commonly used in the literature to filter out the influence of the components $h(n)$ and $r(n)$ of the speech signal model form (Eq. (1)) based on their parametric models determined by the LPC analysis. In this approach, it is important to determine a reliable vocal tract model, which is possible in several ways (WALKER, MURPHY, 2005). Among them, it is worth mentioning:

- 1) closed phase inverse filtering, CPIF, the algorithm (WONG *et al.*, 1979) with the closed phase of the vocal cord vibration cycle analysis only;
- 2) algorithms that use an iterative approach and synchronization mechanisms, e.g., iterative adaptive inverse filtering – IAIF (ALKU, 1991; RAITIO *et al.*, 2011), and pitch synchronous iterative adaptive inverse filtering – PSIAIF (Alku, 1992).

In addition to inverse filtering, there are also parametric methods (QUERESHI, SYED, 2011) and algorithms based on a mixed-phase model of the speech signal. They assume that the impulse response of the vocal tract and the part of the excitation corresponding to the return phase are treated as causal components, while the part of the excitation representing the opening phase in the vocal cord cycle is treated as a non-causal component. Separation of these components can be done using the zeros of the Z-transform (ZZT) algorithm (BOZKURT *et al.*, 2005) or the complex cepstrum decomposition (CCD) algorithm (DRUGMAN *et al.*, 2009). In the present work, as starting point in our research, the IAIF algorithm was used. The elimination of excitation influence are performed for each of the speech frames containing vowels. The HFCC parametrization is then performed, resulting in the cepstral coefficient vectors $c(t, m)$, that is

$$c(t, m) = \sum_{j=1}^J Y_l(t, j) \cos \left(m \left(j - \frac{1}{2} \right) \frac{\pi}{J} \right), \quad m = 1, \dots, M, \quad (2)$$

where $Y_l(t, j)$ is the logarithm of the ERB-scaled spectrum $Y(t, j)$ obtained from the amplitude spectrum $S(t, f)$ under correction multiplied by a bank of Mel filters whose widths were determined according to the equivalent rectangular bandwidth (ERB) scale, t is the frame number, j is the Mel band number, J is the number of Mel bands, and M is the number of HFCC coefficients. The use of a Mel scale of frequencies and a nonlinear function on the values of the spectrum allows a better representation of the performance

of the human auditory system by taking into account the nonlinearity of the perception of frequency and intensity of sound. The expected purpose of the amplitude spectrum correction was to narrow the GMM distributions and reduce classification errors. The effectiveness of the proposed solution was evaluated on the basis of the distance between individual GMM distributions and FER measure before and after the correction.

2. The influence of fundamental frequency on HFCC coefficients

Figures 1 shows the amplitude spectra of consecutive frames of phoneme *a* selected from longer utterances by the same speaker, recorded under identical

conditions, differing in fundamental frequencies (frequency f_0), e.g., for Fig. 1a this is $f_0 \approx 130$ Hz, and for Fig. 1b – $f_0 \approx 195$ Hz.

The main difference between these spectra is in the other positions of the local maxima, which are multiples of the frequency f_0 . Furthermore due to the presence of ripples, the formants are not clearly visible, although their frequencies are approximately: 800 Hz, 1.3 kHz, 2.4 kHz, and 4.0 kHz. In these figures, filters with centre frequencies corresponding to the Mel scale (as in the HFCC parametrization) are also indicated by dotted lines. The consequence of the different positions of the local maxima of the spectrum is the different energy per successive Mel filter band, which leads to different ERB-scaled spectra at different f_0 . This can be observed on the plots presented in Fig. 2. Especially large differences are found for the fourth band.

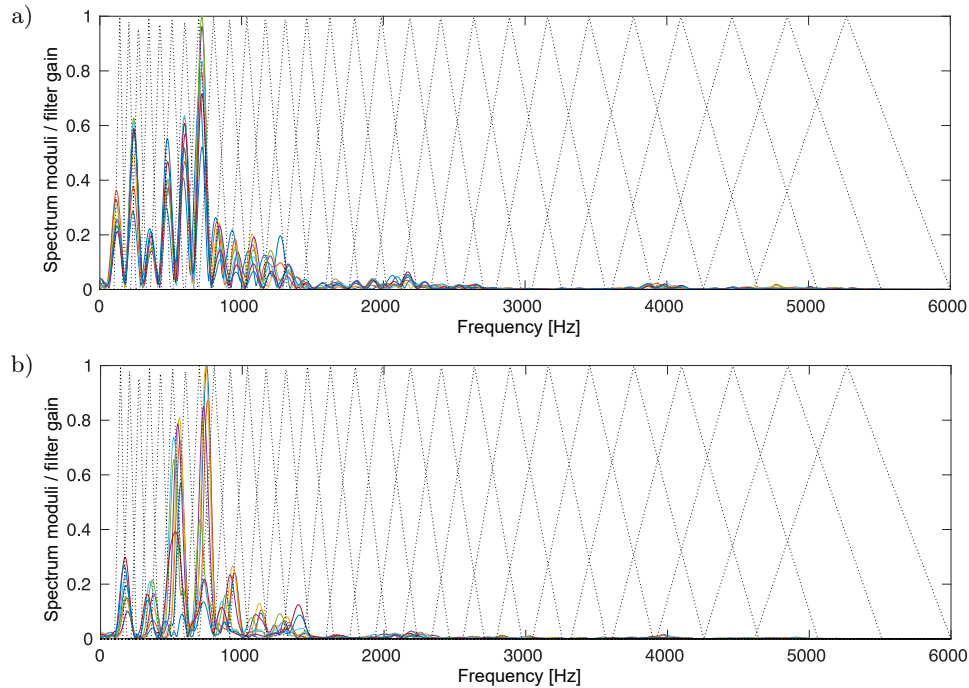


Fig. 1. Amplitude spectra of consecutive frames of phoneme *a* with applied ERB-scale filterbank; the fundamental frequency is about 130 Hz (a) and about 195 Hz (b).

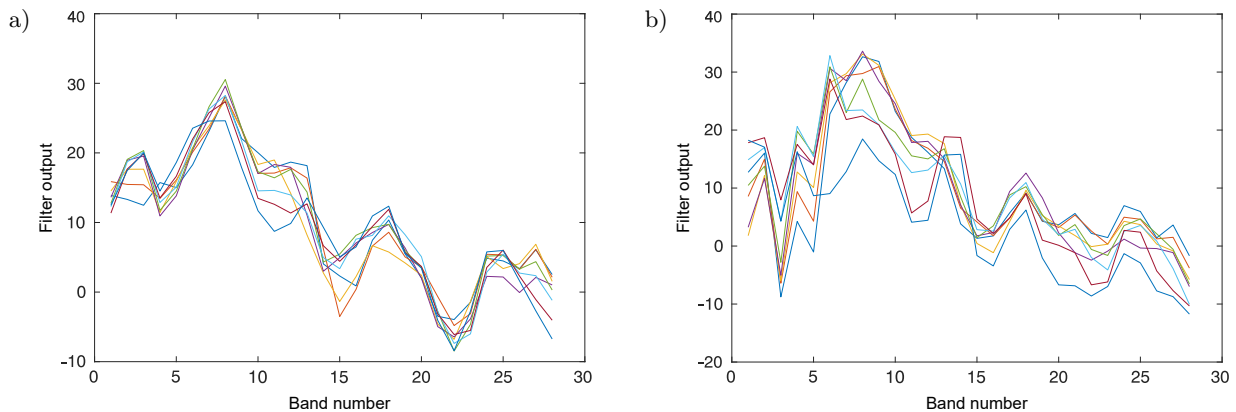


Fig. 2. Spectra of consecutive ERB-scaled frames of the phoneme *a*; the fundamental frequency is about 130 Hz (a) and about 195 Hz (b).

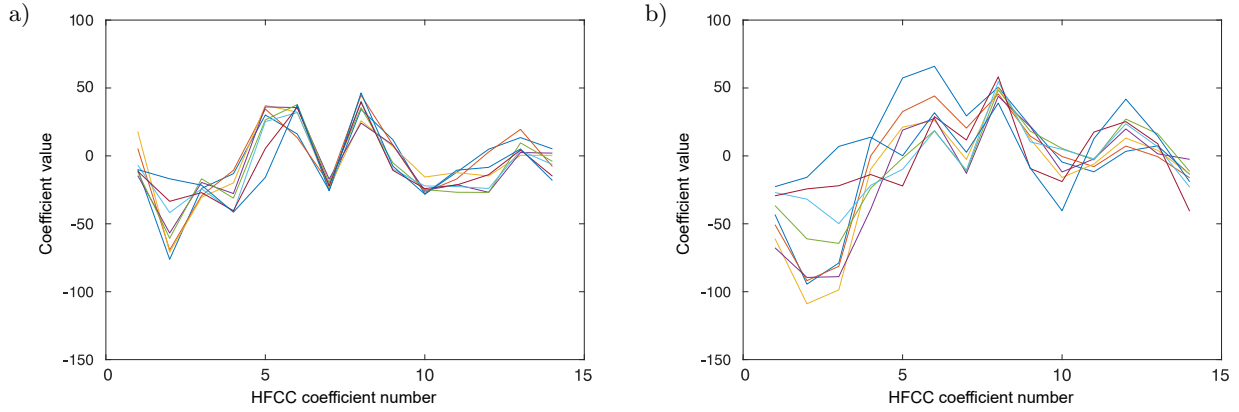


Fig. 3. HFCC coefficients of the phoneme *a* frames; the fundamental frequency is about 130 Hz (a) and about 195 Hz (b).

In turn, Fig. 3 shows plots of HFCC coefficient values for the amplitude spectra presented in Fig. 1. Significant differences can be observed in these figures and the presented examples show the strong influence of the frequency f_0 on the final values of the HFCC coefficients.

3. Glottal excitation signal estimation, correction implementation

In consequence of the experiments analyzed in detail in Sec. 2, the aim of the proposed method is to minimize the effect of excitation signal periodicity on the values of the HFCC coefficients. Theoretically, the excitation signal, for each voiced frame, can be determined using the IAIF (RAITIO, 2011; DRUGMAN *et al.*, 2011), i.e.:

$$x(n) = s(n) \star (h(n) \star r(n))^{-1}, \quad (3)$$

where $(\cdot)^{-1}$ denotes the inverse in the convolution sense. Introducing $w(n) = x(n) \star r(n)$, i.e., as the convolution of the excitation signal $x(n)$ and the function $r(n)$ describing the lips radiation, the quantity $w(n)$ can be determined from the equation

$$\tilde{w}(n) = s(n) \star \tilde{h}(n)^{-1}. \quad (4)$$

Equation (4) presents a case of the blind deconvolution problem. This operation requires the estimation

of the $h(n)$ and then the determination of its inverse in the convolution sense. In the considered situation, the problem of stability can arise, but, this property is guaranteed if the $h(n)$ is a minimum phase or an algorithm, enforcing this minimum phase property, is used. The most popular solution in the case is mean-square filtering (QUATIERI, 2002) and is used in the applied pitch synchronized IAIF (PS-IAIF) filtering.

The PS-IAIF block diagram, modified for the purposes of the work, is presented in Fig. 4. In the preprocessing step the estimator YIN (CHEVEIGNÉ, KAWAHARA, 2002) for the fundamental frequency f_0 of the input voiced speech is calculated. This algorithm is widely applied in the literature and is known as an effective solution. An input signal $s(n)$ is partitioned, based on the YIN estimator, into frames with length equal to current values of the fundamental period $T_0 = 1/f_0$. Next, for each input frame, in the first step of PS-IAIF, a preliminary estimator of the filter is determined that models the combination of glottal excitation and the lip radiation using an LPC filter of the order the 1. In the second step, after compensating for the influence of $G_1(z)$ on the signal $s(n)$, a preliminary estimator $H_{v1}(z)$ of the vocal tract is determined with LPC filter of the order 10. The resulting estimator $H_{v1}(z)$, in the third step, is used to filter out the influence of the vocal tract from the signal $s(n)$. In this step, the influence of the lip emission properties is also eliminated by integration, and a more

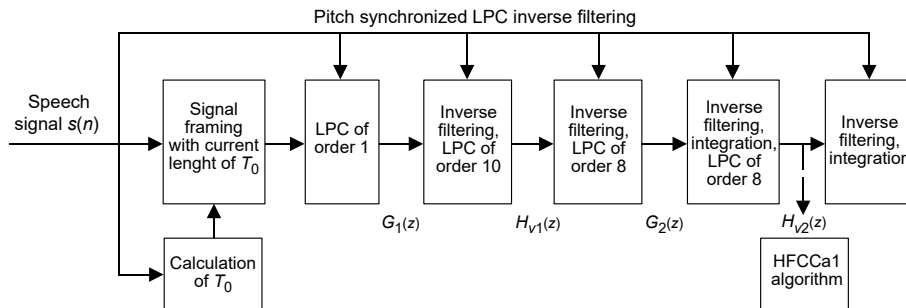


Fig. 4. Block diagram of the applied inverse filtering algorithm (PS-IAIF).

accurate parametric model $G_2(z)$ is determined with the LPC filter of the order 8. In the fourth step, using $G_2(z)$, by means of inverse filtering, integration, and LPC analysis, the parameters of the $H_{v2}(z)$ model of the vocal tract of the order 8 are determined. Given $H_{v2}(z)$, the frequency domain transfer function is of the form

$$H_{v2}(f) = \frac{1}{1 - \sum_{p=1}^7 a_p e^{-j2\pi f p / f_s}}. \quad (5)$$

The result of this operation is used to determine the HFCC coefficients after compensating for the influence of the glottal excitation (the HFCCa1 algorithm). Since the phase of the signal spectrum is not taken into account in the HFCC parametrization, we assume here that modelling using the LPC technique will yield minimum phase property of all elements of Eq. (1).

4. Correction quality measures

In order to evaluate the effectiveness of the proposed methods of modifying the HFCC parametrization, numerical tests were carried out on Polish speech vowels occurring in the recording database described in Subsec. 5.1. Performing experiments required the prior development of acoustic models of these vowels in the form of GMM probability distributions, two measures were used to evaluate the effectiveness of the compensation:

- 1) the Kullback–Leibler distance between the probability distributions (KULLBACK, 1968) – the $\text{KL}(\cdot)$ measure;
- 2) the single frame error recognition – the FER measure (MAKOWSKI, 2011).

4.1. Probabilistic acoustic model of phonemes

The acoustic GMM models used in the frame recognition process are a mixture of $K = 7$ multidimensional normal probability distributions with a diagonal covariance matrices $\Sigma_{p,i}$ determined based on the expectation-maximization (EM) algorithm (DEMPSTER et al., 1977), i.e.:

$$p_f(\mathbf{o}) = \sum_{i=1}^K w_{fi} \mathcal{N}(\mathbf{o}, \mathbf{m}_{f,i}, \Sigma_{f,i}), \quad (6)$$

where

$$\mathcal{N}(\mathbf{o}, \mathbf{m}_{f,i}, \Sigma_{f,i}) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma_{f,i,n}^2}} e^{-\frac{1}{2\sigma_{f,i,n}^2} [o_n - m_{f,i,n}]^2}. \quad (7)$$

4.2. Distances between GMM distributions

In general, a typical measure to calculate the distance between two probability density distributions

$p_h(\mathbf{o})$ and $p_g(\mathbf{o})$ for a N -dimensional vector of random variables \mathbf{o} is the Kullback–Leibler divergence defined as follows (KULLBACK, 1968):

$$\text{KL}(p_h \parallel p_g) = \int_{\mathcal{O}} p_h(\mathbf{o}) \log \left(\frac{p_h(\mathbf{o})}{p_g(\mathbf{o})} \right) d\mathbf{o}. \quad (8)$$

Unfortunately, for the case of distributions represented by a mixture of Gaussian GMM distributions of the form

$$p_h(\mathbf{o}) = \sum_{i=1}^K w_{h,i} \mathcal{N}(\mathbf{o}, \mathbf{m}_{h,i}, \Sigma_{h,i}) = \sum_{i=1}^K w_{h,i} p_{h,i}(\mathbf{o}), \quad (9)$$

$$p_g(\mathbf{o}) = \sum_{i=1}^K w_{g,i} \mathcal{N}(\mathbf{o}, \mathbf{m}_{g,i}, \Sigma_{g,i}) = \sum_{i=1}^K w_{g,i} p_{g,i}(\mathbf{o}),$$

where $\mathbf{m}_{h,i}$ and $\mathbf{m}_{g,i}$ are the mean value vectors and $\Sigma_{h,i}$ and $\Sigma_{g,i}$ the autocovariance matrices of the components of the Gaussian distributions in the mixtures, there is no closed form formula of the $\text{KL}(\cdot)$ measure determination. However, we can use a deterministic approximation of Eq. (8) based on the unscented transform (UT) transformation (JULIER, UHLMANN, 2004). Under the assumption that the distributions $p_h(\mathbf{o})$ and $p_g(\mathbf{o})$ are of the GMM form (Eq. (9)) with diagonal covariance matrices, i.e., $\Sigma_{h,i} = \text{diag}\{\sigma_{h,i,k}^2\}$ and $\Sigma_{g,i} = \text{diag}\{\sigma_{g,i,k}^2\}$ for $k = 1, 2, \dots, N$, we can write that

$$\begin{aligned} \text{KL}(p_h \parallel p_g) &= \int_{\mathcal{O}} p_h(\mathbf{o}) \log \left(\frac{p_h(\mathbf{o})}{p_g(\mathbf{o})} \right) d\mathbf{o} \\ &= E_{p_h}[\log p_h(\mathbf{o})] - E_{p_h}[\log p_g(\mathbf{o})] \\ &= \sum_{i=1}^K w_{h,i} E_{p_{h,i}}[\log p_h(\mathbf{o})] \\ &\quad - \sum_{i=1}^K w_{h,i} E_{p_{h,i}}[\log p_g(\mathbf{o})]. \end{aligned} \quad (10)$$

According to the UT method, for each of the K component distributions of the GMM mixture $p_{h,i}(\mathbf{o}) = \mathcal{N}(\mathbf{o}, \mathbf{m}_{h,i}, \Sigma_{h,i})$ with diagonal matrices $\Sigma_{h,i} = \text{diag}\{\sigma_{h,i,k}^2\}$, we generate a set of $2N$ “sigma” points of the form

$$\mathbf{o}_{i,k} = \mathbf{m}_{h,i} - \sqrt{N\sigma_{h,i,k}^2} \mathbf{e}_k, \quad (11)$$

$$\mathbf{o}_{i,k+N} = \mathbf{m}_{h,i} + \sqrt{N\sigma_{h,i,k}^2} \mathbf{e}_k,$$

where \mathbf{e}_k for $k = 1, 2, \dots, N$ are basis vectors in the N dimensional Cartesian coordinate system and we determine the approximation of the integral $E_{p_{h,i}}[\log p_g(\mathbf{o})]$ based on the formula (GOLDBERGER, ARONOWITZ, 2005)

$$\begin{aligned} E_{p_{h,i}}[\log p_g(\mathbf{o})] &= \int_{\mathcal{O}} p_{h,i}(\mathbf{o}) \log p_g(\mathbf{o}) d\mathbf{o} \\ &\approx \frac{1}{2N} \sum_{k=1}^{2N} \log p_g(\mathbf{o}_{i,k}). \end{aligned} \quad (12)$$

We include all the partial results of the calculations into Eq. (10) and obtain the approximation of the distance value $KL(\cdot)$ between the considered distributions. To satisfy the symmetry property of the distance measure $KL(\cdot)$ between the GMM distributions $p_g(\mathbf{o})$ and $p_h(\mathbf{o})$, the final form

$$d(p_g, p_h) = \frac{1}{2}(KL(p_h \parallel p_g) + KL(p_g \parallel p_h)) \quad (13)$$

was applied in numerical experiments.

4.3. Frame error rate

The frame error rate (FER) is typically used to evaluate the quality of speech recognition at the individual frame level and is defined as

$$m = \frac{T_{\text{err}}}{T} \cdot 100 \%, \quad (14)$$

where T is the number of all frames to be recognised and T_{err} is the number of frames incorrectly recognised.

5. Correction results

5.1. Speech recordings

The set of recordings that constitute the database for the experiments consists of 36 male adult voices recorded in different Polish cities. For each speaker, 150 words of Polish were recorded and speech fragments containing vowels from preliminary chosen 43 words were used in the experiment. The sampling

rate of the signals was 12 kHz. The results obtained from numerical experiments are for noisy signals with a signal-to-noise ratio of 35 dB. All of these recordings were manually segmented and labelled, and the phonetic unit in the labelling process was the phoneme. The frame length was 30 ms with the frame shift 10 ms and the number of cepstral coefficients was $N = 14$.

5.2. Examples of algorithm results

The section presents example results of the HFCCa1 algorithm for three consecutive frames of the *a* phoneme, whose statistics are presented in Figs. 1–3. Figure 5 presents successively:

- the magnitude of the preliminary estimator $G_1(f)$;
- the magnitude of the transfer function of the preliminary estimator $H_{v1}(f)$;
- the magnitude of the estimator $G_2(f)$;
- the magnitude of a transfer function of the estimator $H_{v2}(f)$;
- the amplitude spectra of the signal frames;
- the amplitude spectra of the frames after correction.

The cepstral coefficients in the HFCCa1 method are calculated based on the results, examples of which are presented in Fig. 5d. Comparison of plots from Figs. 5d and 5e shows the effectiveness of the proposed algorithms to eliminate ripples caused by the quasiperiodicity of the glottal excitation.

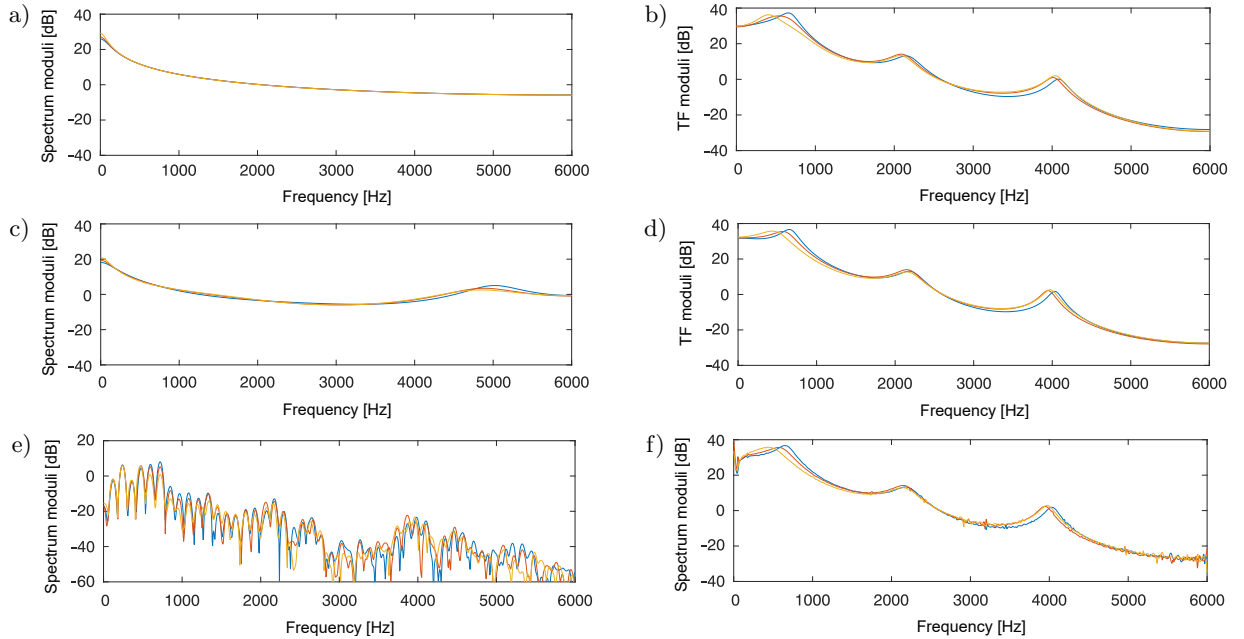


Fig. 5. Example results of the HFCCa1 algorithm for three consecutive frames of the phoneme *a*: a) moduli of the preliminary estimator $G_1(f)$; b) transfer function moduli of the preliminary estimator $H_{v1}(f)$; c) moduli of the estimator $G_2(f)$; d) transfer function moduli of the estimator $H_{v2}(f)$; e) amplitude spectra of the signal frames; f) amplitude spectra of the frames after correction.

5.3. Global results of compensation quality assessment

In Fig. 6, in the form of a table, the KL differences after and before correction between the GMM distributions of the six Polish vowels are presented. Furthermore, the red colour indicates a decrease in the distance after correction and the green colour an increase.

	<i>i</i>	<i>y</i>	<i>e</i>	<i>a</i>	<i>o</i>	<i>u</i>
<i>i</i>		13.46	15.98	1.94	2.931	26.68
<i>y</i>	13.56		19.65	-1.44	3.52	0.47
<i>e</i>	15.98	19.65		0.88	5.89	2.11
<i>a</i>	1.94	-1.44	0.89		23.45	13.69
<i>o</i>	2.31	3.52	5.89	23.45		11.09
<i>u</i>	26.68	0.47	2.11	13.69	11.09	

Fig. 6. Differences in KLD distances after and before correction between the six vowels of Polish speech. The red colour indicates a decrease in distance and the green colour an increase.

It is easily observed that in most cases of comparisons an increase in these distances is observed, and the differences are largest for the phonemes *i* and *u*. Simultaneously, significant decreases in distance are noticed between the phonemes *y* and *a*. Presenting the results more synthetically, by summing the distances between a given GMM distribution and the other distributions, i.e., determining the values

$$D_f = \sum_{i=1}^F d(p_f, p_i) \quad (15)$$

we obtain global KLD distances for individual phonemes before and after correction. These measures are presented in Fig. 7, where it can be seen that an increase in KLD distances occurred for all vowels.

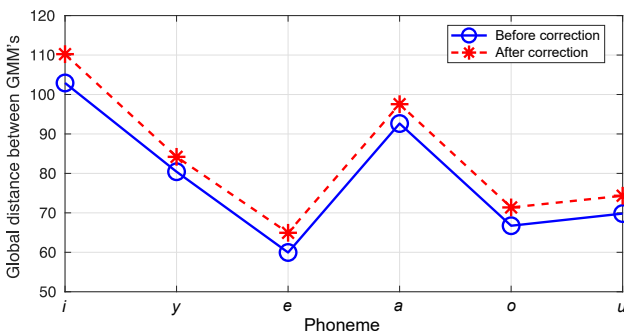


Fig. 7. Global KLD distances for vowels.

In turn, the results of the FER measure in one-to-one recognition for Polish speech vowels are presented

in the form of a table in Fig. 8. The upper values in the table elements indicate the FER before correction and the lower values after correction. Furthermore, the green colour indicates situations for which there was a decrease in FER, and the red colour indicates an increase.

	<i>i</i>	<i>y</i>	<i>e</i>	<i>a</i>	<i>o</i>	<i>u</i>
<i>i</i>		2.24 2.22	0.35 1.40	0.00 0.00	0.15 0.04	1.07 1.66
<i>y</i>	1.16 1.07		9.29 7.22	0.06 0.28	0.00 0.00	0.72 0.66
<i>e</i>	1.02 1.35	14.41 13.03		5.43 3.89	0.63 0.98	0.49 0.39
<i>a</i>	0.00 0.00	0.19 0.19	7.11 6.74		2.94 3.07	0.01 0.00
<i>o</i>	0.00 0.00	0.33 0.66	1.39 1.53	4.35 3.02		4.32 4.32
<i>u</i>	1.12 0.68	0.69 1.21	0.08 0.00	0.28 0.25	4.36 3.07	

Fig. 8. FER values for Polish speech vowels.

The results presented in Fig. 8 imply that in most cases there was a reduction in single frame recognition errors. On the other hand, Fig. 9 shows plots of the FER sum following the table rows of Fig. 8.

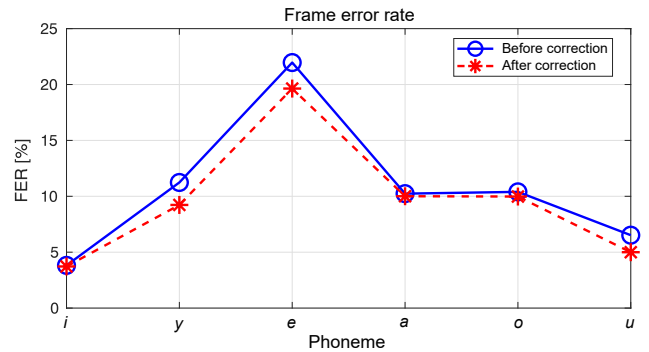


Fig. 9. Global FER values for Polish speech vowels.

This form of obtained data analysis shown in Fig. 9 also confirms that the proposed correction results in a reduction in FER errors.

6. Conclusions

The modification of the HFCC parametrization proposed in this paper meets the predicted expectations. Through estimation and inverse filtering it is possible to minimise the influence of the quasiperiodicity of the source of voiced speech, in the function of the amplitude spectrum $|H_{v_2}(f)|$ used to determine the HFCC coefficients. Consequently, the area of fluctuations of the feature vector values is reduced. This form of the conclusion is confirmed by the obtained results of the Kullback–Leibler distances between the GMM distributions of Polish speech vowels, which are

larger after the correction. Simultaneously, the classification errors of individual frames evaluated by the frame-error-rate measure are also reduced. As a result, the proposed modification of the HFCC parametrization should result in an increase in the efficiency of the complete ASR system. Finally, it should be kept in mind that, in general, the variability of the components of the feature vector, in addition to the considered influence of the quasiperiodicity of the glottal excitation, is affected by a number of other factors such as inter- and intrapersonal variability, contextual variability, influence of recording conditions, etc.

Acknowledgments

Calculations have been carried out using resources provided by the Wrocław Centre for Networking and Supercomputing (grant no. 376).

References

1. ALKU P. (1991), Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering, [in:] *Proceedings 2nd European Conference on Speech Communication and Technology (Eurospeech 1991)*, pp. 1081–1084, <https://doi.org/10.21437/Eurospeech.1991-257>.
2. ALKU P. (1992), Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering, *Speech Communication*, **11**(2–3): 109–118, [https://doi.org/10.1016/0167-6393\(92\)90005-R](https://doi.org/10.1016/0167-6393(92)90005-R).
3. BOZKURT B., DOVAL B., D'ALESSANDRO C., DUTOIT T. (2005), Zeros of Z-transform representation with application to source-filter separation in speech, *IEEE Signal Processing Letters*, **12**(4): 344–347, <https://doi.org/10.1109/LSP.2005.843770>.
4. CHEVEIGNÉ A., KAWAHARA H. (2002), YIN, a fundamental frequency estimator for speech and music, *The Journal of the Acoustical Society of America*, **111**(4): 1917–1930, <https://doi.org/10.1121/1.1458024>.
5. DAVIS S., MERMELSTEIN P. (1980), Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE Transactions on Acoustics, Speech and Signal Processing*, **28**(4): 357–366, <https://doi.org/10.1109/TASSP.1980.1163420>.
6. DEMPSTER A.P., LAIRD N.M., RUBIN D.B. (1977), Maximum-likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1): 1–38.
7. DHARANIPRAGADA S., RAO B.D. (2001), MCDR based feature extraction for robust speech recognition, [in:] *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 309–312, <https://doi.org/10.1109/ICASSP.2001.940829>.
8. DRUGMAN T., BOZKURT B., DUTOIT T. (2009), Complex cepstrum-based decomposition of speech for glottal source estimation, [in:] *Proceedings of the Annual Conference of the International Speech Communication Association, InterSpeech*, <https://doi.org/10.21437/Interspeech.2009-27>.
9. DRUGMAN T., BOZKURT B., DUTOIT T. (2011), A comparative study of glottal source estimation techniques, *ArXiv*, <https://doi.org/10.48550/arXiv.2001.00840>.
10. GOLDBERGER J., ARONOWITZ H. (2005), A distance measure between GMMs based on the unscented transform and its application to speaker recognition, [in:] *9th European Conference on Speech Communication and Technology, InterSpeech*, pp. 1985–1988, <https://doi.org/10.21437/Interspeech.2005-624>.
11. HERMANSKY H. (1990), Perceptual linear predictive (PLP) analysis of speech, *The Journal of the Acoustical Society of America*, **87**(4): 1738–1752, <https://doi.org/10.1121/1.399423>.
12. HERMANSKY H., FOUSEK P. (2005), Multi-resolution RASTA filtering for TANDEM-based ASR, [in:] *Proceedings of the Annual Conference of the International Speech Communication Association, InterSpeech*, pp. 361–364, <https://doi.org/10.21437/Interspeech.2005-184>.
13. HOSSA R., MAKOWSKI R. (2016), An effective speaker clustering method using UBM and ultra-short training utterances, *Archives of Acoustics*, **41**(1): 107–118, <https://doi.org/10.1515/aoa-2016-0011>.
14. JULIER S.J., UHLMANN J.K. (2004), Unscented filtering and nonlinear estimation, [in:] *Proceedings of the IEEE*, **92**(3): 401–422, <https://doi.org/10.1109/JPROC.2003.823141>.
15. KOEHLER J., MORGAN N., HERMANSKY H., HIRSCH H.G., TONG G. (1994), Integrating RASTA-PLP into speech recognition, [in:] *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, <https://doi.org/10.1109/ICASSP.1994.389266>.
16. KUAN T.-W., TSAI A.-C., SUNG P.-H., WANG J.-F., KUO H.-S. (2016), A robust BFCC feature extraction for ASR system, *Artificial Intelligence Research*, **5**(2), <https://doi.org/10.5430/air.v5n2p14>.
17. KULLBACK S. (1968), *Information Theory and Statistics*, Dover Publications, New York.
18. MAKOWSKI R. (2011), *Automatic Speech Recognition – Selected Problems* [in Polish: *Automatyczne Rozpoznawanie Mowy – Wybrane Zagadnienia*], Oficyna Wydawnicza Politechniki Wrocławskiej.
19. MORITZ N., ANEMULLER J., KOLLMEIER B. (2015), An auditory inspired amplitude modulation filter bank for robust feature extraction in automatic speech recognition, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **23**(11): 1926–1937, <https://doi.org/10.1109/TASLP.2015.2456420>.
20. MRÓWKA P., MAKOWSKI R. (2008), Normalization of speaker individual characteristics and compensation of linear transmission distortions in command recognition systems, *Archives of Acoustics*, **33**(2): 221–242.
21. MURTHI M.N., RAO B.D. (2000), All-pole modeling of speech based on the minimum variance distortionless response spectrum, *IEEE Transactions on Speech*

- and Audio Processing, **8**(3): 221–239, <https://doi.org/10.1109/89.841206>.
22. PLUMPE M.D., QUATIERI T.F., REYNOLDS D.A. (1999), Modeling of the glottal flow derivative waveform with application to speaker identification, *IEEE Transactions on Speech and Audio Processing*, **7**(5): 569–586, <https://doi.org/10.1109/89.784109>.
 23. PRASAD N.V., UMESH S. (2013), Improved cepstral mean and variance normalization using Bayesian framework, [in:] *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 156–161, <https://doi.org/10.1109/ASRU.2013.6707722>.
 24. QUATIERI T.F. (2002), *Discrete-Time Speech Signal Processing: Principles and Practice*, Pearson Education.
 25. QUERESHI T.M., SYED K.S. (2011) A new approach to parametric modeling of glottal flow, *Archives of Acoustics*, **36**(4): 695–712, <https://doi.org/10.2478/v10168-011-0047-3>.
 26. RABINER L., JUANG B.-H. (1993), *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs.
 27. RAITIO T. et al. (2011), HMM-Based speech synthesis utilizing glottal inverse filtering, *IEEE Transactions on Audio, Speech and Language Processing*, **19**(1): 1530–165, <https://doi.org/10.1109/TASL.2010.2045239>.
 28. SHARMA G., UMAPATHY K., KRISHNAN S. (2020), Trends in audio signal feature extraction methods, *Applied Acoustics*, **158**: 107020, <https://doi.org/10.1016/j.apacoust.2019.107020>.
 29. SKOWRONSKI M., HARRIS J.G. (2003) Improving the filter bank of a classic speech feature extraction algorithm, [in:] *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS '03*, pp. 281–284, <https://doi.org/10.1109/ISCAS.2003.1205828>.
 30. WAARAMAA T., LAUKKANEN A.M., AIRAS M., ALKU P. (2010), Perception of emotional valences and activity levels from vowel segments of continuous speech, *Journal of Voice*, **24**(1): 8–30, <https://doi.org/10.1016/j.jvoice.2008.04.004>.
 31. WALKER J., MURPHY P. (2005), A review of glottal waveform analysis [in:] *Progress in Nonlinear Speech Processing, Workshop on Nonlinear Speech Processing, Lecture Notes in Computer Science*.
 32. WONG D., MARKEL J., GRAY A. (1979), Least squares glottal inverse filtering from the acoustic speech waveform, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **27**(4): 350–355, <https://doi.org/10.1109/TASSP.1979.1163260>.
 33. YIN H., HOHMANN V., NADEU C. (2011), Acoustic features for speech recognition based on Gammatone filterbank and instantaneous frequency, *Speech Communication*, **53**(5): 707–715, <https://doi.org/10.1016/j.specom.2010.04.008>.
 34. ZAMBRZYCKA A. (2021), *Adaptation in automatic speech recognition systems* [in Polish: *Adaptacja w systemach automatycznego rozpoznawania mowy*], Ph.D. Thesis, Wrocław University of Science and Technology.

Research Paper

Snoring Sounds Classification of OSAHS Patients Based on Model Fusion

Yexin LUO⁽¹⁾, Jianxin PENG^{(1)*}, Li DING^{(2)*}, Yikai ZHANG⁽¹⁾, Lijuan SONG⁽³⁾,
Qianfan ZHANG⁽¹⁾, Houpeng CHEN⁽¹⁾

⁽¹⁾ *School of Physics and Optoelectronics, South China University of Technology
Guangzhou, China*

⁽²⁾ *School of Advanced Manufacturing Engineering, Hefei University
Hefei, China*

⁽³⁾ *State Key Laboratory of Respiratory Disease, Department of Otolaryngology-Head and Neck Surgery,
Laboratory of ENT-HNS Disease, First Affiliated Hospital, Guangzhou Medical University
Guangzhou, China*

*Corresponding Authors e-mails: phjxpeng@163.com (Jianxin Peng); gtxydingli@163.com (Li Ding)

(received April 19, 2024; accepted November 25, 2024; published online February 19, 2025)

Obstructive sleep apnea hypopnea syndrome (OSAHS) is a prevalent and detrimental chronic condition. The conventional diagnostic approach for OSAHS is intricate and costly. Snoring is one of the most typical and easily obtained symptom of OSAHS patients. In this study, a series of acoustic features are extracted from snoring sounds. A fused model that integrates a deep neural network, K-nearest neighbors (KNN), and a random under sampling boost algorithm is proposed to classify snoring sounds of simple snorers (SSSS), simple snoring sounds of OSAHS patients (SSSP), and apnea-hypopnea snoring sounds of OSAHS patients (APSP). The ReliefF algorithm is employed to select features with high relevance in each classification model. A hard voting strategy is implemented to obtain an optimal fused model. Results show that the proposed fused model achieves commendable performance with an accuracy rate of 85.76 %. It demonstrates the effectiveness and validity of assisting in diagnosing OSAHS patients based on the analysis of snoring sounds.

Keywords: obstructive sleep apnea hypopnea syndrome; snoring sounds; deep neural network; model fusion.



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

1. Introduction

Obstructive sleep apnea hypopnea syndrome (OSAHS) is a chronic sleep-related disease with the high incidence and great harm that is characterized by partial or complete collapse of the upper airway during sleep (ECKERT *et al.*, 2007; FRIEDMAN *et al.*, 2004; IZCI, DOUGLAS, 2012; OSMAN *et al.*, 2018). There are many contributors to the collapse, including an ineffective pharyngeal dilator muscle function during sleep, a low threshold for arousal to airway narrowing during sleep, and unstable control of breathing, which may be caused by a narrow, crowded, or collapsible upper airway of OSAHS patients (OSMAN *et al.*, 2018). OSAHS not only adversely influences the sleep qual-

ity of patients, but also leads to hypertension, coronary heart disease, diabetes, cerebrovascular disease, other complications, and even causes sudden death at night (REDLINE *et al.*, 2010; WHITE, 2005). The recent epidemiological survey has found that the prevalence of OSAHS among the global population ranged from 9 % to 38 % (CARON *et al.*, 2017). The elderly are the high incidence group that the prevalence rate of OSAHS is as high as 90 % for older males and 78 % for older females (CASTILLO-ESCARIO *et al.*, 2019). Polysomnography (PSG) is the gold standard for diagnosing OSAHS by detecting respiratory disturbance events that mainly include apnea and hypopnea events (MINARITZOGLOU *et al.*, 2008). The apnea-hypopnea-index (AHI) is obtained by PSG to measure the aver-

age number of respiratory disturbance events per hour during sleep. According to the American Academy of Sleep Medicine (AASM), subjects can be diagnosed as a simple snorer, mild, moderate, and severe OSAHS patient based on $AHI \leq 5$, $5 < AHI \leq 15$, $15 < AHI \leq 30$, and $AHI > 30$, respectively (BERRY *et al.*, 2012). The PSG requires more than 15 sensors connected to the patients that needs to be operated and checked by professional doctors in the hospital to monitor multiple biological signals of the test subject during sleep. The expensive cost, inconvenient device, and complex process limit the wide use of PSG that cause OSAHS to be a serious disease with a low diagnostic rate (GOTTLIEB, PUNJABI, 2020; OSMAN *et al.*, 2018). The high prevalence and low diagnostic rate make the OSAHS be a public health problem that greatly influences the life quality of patients. With the increasing concern about sleep problems, researchers have been focused on studying various physiological signals during sleeping to assist in monitoring apnea and hypopnea events. The AASM indicates that one or more physiological signals, including oxygen, nasal airflow, electrocardiogram, electroencephalography, and snoring sound can be applied to detect apnea and hypopnea events to diagnose OSAHS (BERRY *et al.*, 2012).

Snoring is the most prominent symptom of OSAHS patients that caused by the vibration of the upper airway (GISLASON, BENEDIKTSDDOTTIR, 1995; PEVERNAGIE *et al.*, 2010; SOWHO *et al.*, 2020; ULUALP, 2010). The acoustic features of snoring sounds can reflect the specific structure of the upper airway (LUGARES *et al.*, 1988). Studies have indicated that there are obviously anatomical and non-anatomical structural differences of the upper airway between simple snorers and OSAHS patients (AZARBARZIN, MOUSAVI, 2013; FIZ *et al.*, 1996; MARKANDEYA *et al.*, 2018; HERZOG *et al.*, 2008). Early studies have indicated that palatal snoring mainly occurs in simple snorers without any obstruction of the upper airways, while non-palatal snoring can be an indicator for OSAHS patients (QIAN *et al.*, 2021). Recent work by SUN *et al.* (2023) has revealed that snoring sounds of OSAHS patients exhibit higher formant frequencies. PEREZ-PADILLA *et al.* (1993) found that there was different energy distribution around 800 Hz of snoring sounds between simple snoring and those of OSAHS patients. Based on this condition, studies have been focused on identifying simple snorers and OSAHS patients. SOLÀ-SOLER *et al.* (2007) classified simple snorers and OSAHS patients based on $AHI = 10$, which yielded 93 % precision. SUN *et al.* (2023) applied two Gaussian mixture models to explore the acoustic characteristics of snoring sounds throughout the whole night to classify simple snorers and OSAHS patients with 90.0 % accuracy. DING *et al.* (2024) applied a fused model obtained from different domain to classify snoring sounds during the whole night of simple snorers

and OSAHS patients, which could exactly identify OSAHS patients. Furthermore, researchers (LEE, EL-LIS, 2012; HOU *et al.*, 2019; ALSHAER *et al.*, 2019; CHENG *et al.*, 2022; DING *et al.*, 2023) have explored the characteristics of snoring sounds obtained by different sleep stages during the whole sleep to diagnose the severity of OSAHS patients. LEE *et al.* (2012) showed that there was different energy distribution of snoring sounds during apnea-hypopnea events and simple sleeping. DING *et al.* (2023) proposed VGG19-LSTM model to classify snoring sounds of simple snorers and OSAHS patients with 99.31 % accuracy and 99.13 % sensitivity. A long short-term memory (LSTM) neural network was proposed to classify three-category snoring sounds related to the severity of OSAHS with 81.6 % accuracy (CHENG *et al.*, 2022). These studies have demonstrated the effectiveness and convenience of diagnosing OSAHS patients based on analysis of snoring sound.

The aforementioned classification results of snoring sounds have clearly demonstrated that the structure of the upper airway of OSAHS patients is obviously different from that of simple snorer. The abnormal structure could cause the occurrence of apnea and hypopnea respiratory events, as well as abnormal snoring sounds, which provided a strong basis for the diagnosis of OSAHS based on snoring sounds. Few studies (CHENG *et al.*, 2022; SONG *et al.*, 2023; SUN *et al.*, 2023) focused on whether the abnormal upper airway may influence the normal sleep process of OSAHS patients. Since the characteristic of snoring sounds could reflect the structure of the upper airway, intuitively classifying snoring sounds of simple snorers (SSSS), apnea-hypopnea snoring sounds of OSAHS patients (APSP), and simple snoring sounds of OSAHS patients (SSSP) could explore the characteristics of the upper airway in the different stages of sleep for simple snorers and OSAHS patients, respectively. The classification results could indicate that whether the abnormal upper airway can be reflected by snoring sounds and whether the abnormal upper airway influence the normal sleep for OSAHS patients. The existing studies about snoring sound classification are based on a single classification model, which had limited classification accuracy and robustness. On this condition, the snoring sound classification tasks based on a fusion strategy might help to diagnose OSAHS patients more accurately.

In this study, a fused model is proposed to classify three kinds of snoring sounds, including SSSS, APSP, and SSSP. A series of acoustic features were extracted from snoring sounds. Three classifiers were first used to classify these three kinds of snoring sounds based on extracted acoustic features. Then a hard voting model fusion strategy was applied to integrate these basic models to obtain a model with relatively better classification performance and higher robustness.

2. Material and methods

2.1. Dataset

The 46 subjects selected from the PSG-Audio dataset are applied to validate the proposed method, including 8 simple snorers and 38 OASHS patients with different severities (KOROMPILI *et al.*, 2021). All snoring sounds are collected clinically. When a subject undergoes the PSG (Alice 6), an ultra-linear measurement condenser microphone (Berringer ECM800) is placed approximately 1 m above the subject's bed to record snoring sounds during the whole night. Sound signals are sampled at 48 kHz with 24-bit resolution and saved as WAV. All recorded signals are enhanced and segmented by the noise reduction algorithm (WANG *et al.*, 2017). These enhanced snoring segments are labeled by ear-nose-throat (ENT) experts as SSSS, SSSP, and APSP. In the experiment, there are 73 373 effective snoring segments extracted from all 46 subjects, including 12 967 SSSS, 44 748 simple SSSP, and 15 658 APSP. These snoring sounds are divided into a training set and a validation set by the ratio of 4:1.

2.2. Proposed fused model

In the work, a fused model is proposed to classify SSSS, SSSP, and APSP to explore structures of the upper airway of simple snorers and OSAHS patients during sleeping. The overall structure of the proposed model is shown in Fig. 1. A series of acoustic features are firstly extracted to express snoring sounds. Three basic classifiers, including the deep neural network – DNN (JANIESCH *et al.*, 2021), K-nearest neighbors – KNN (ZHANG *et al.*, 2017), and random under sampling boost algorithm – RUSBoost (SEIFFERT *et al.*, 2010) are applied to classify these three types of snor-

ing sounds. To adequately integrate these basic classifiers from different domains, a model fusion strategy based on hard voting is used to fuse these classifiers. That is to say, the final classification results of snoring sounds were obtained by averaging the probability of these three basic models.

The three basic classifiers used in this work are DNN, KNN, RUSBoost. KNN is one of the most mature and simplest machine learning classification algorithms with relatively high performance in different domains. The basic idea of KNN is to calculate the distance between the test sample and all training samples to obtain its nearest neighbors and then conduct KNN classification. Choosing the proper K -value is an important part for training the KNN model. The RUSBoost algorithm is an effective ensemble method for the classification task with the unbalanced sample distribution. RUSBoost incorporates random under-sampling technology to remove samples from the majority class at each boosting iteration of the Adaboost.M2 algorithm. Based on this strategy, RUSBoost could adequately apply samples of the majority class and solve the problem of unbalanced sample distribution. The parameters that RUSBoost needs to be trained are mainly concentrated on Adaboost.M2, including a base estimator, the learning rate, n -estimators, and so on. DNN is a useful technology for the classification task with large samples. In this work, a DNN structure with two hidden layers is constructed to classify snoring sounds. There are 100 neurons in the first hidden layer and 5 neurons in the second hidden layer. The loss function and the activation function used in the DNN are the logistic cost regression function and the sigmoid function, respectively. The optimizer used for training is Adam. The batch size and the learning rate are set as 64 and 0.05, respectively.

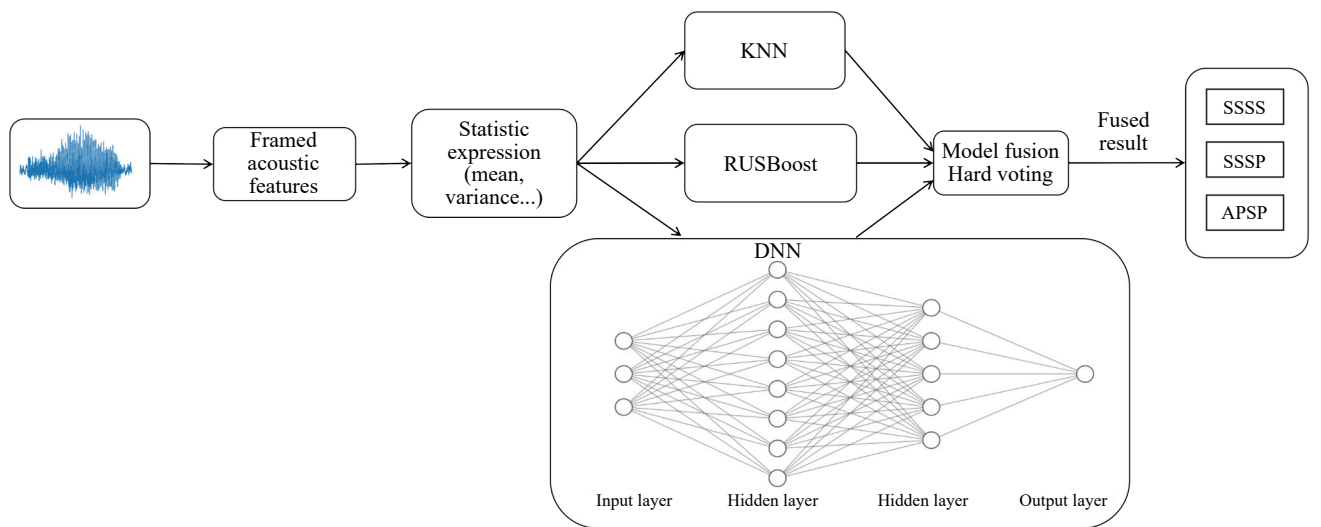


Fig. 1. Overall structure of the proposed system.

2.3. Feature extraction

In the work, a series of acoustic features from the time and frequency domains are extracted to express snoring sounds. There are 16 features with 45 dimensions, including the Mel-frequency cepstral coefficient (MFCC), the linear prediction coefficient (LPC), 800 Hz power ratio (PR800), the crest factor (CF), the fundamental frequency (F_0), the pitch, formants, and a series of spectrum related features. Since the generation process of snoring sounds has a significant effect on its high frequency band and a smaller effect on the low frequency band, all snoring sounds are conducted pre-emphasizing that aims to compensate for the loss of high frequency components before the feature extraction. These pre-emphasized snoring sounds are framed by a hamming window with length of 20 ms and 50 % overlap. All features are firstly extracted for each frame. Statistic functions, including mean, minimum, maximum, and variance, are calculated by frames for each snoring segment to describe the feature distribution for each snoring segment.

2.3.1. Mel-frequency cepstral coefficient

The extraction of MFCC can be divided into five parts (ZHENG *et al.*, 2001). Firstly, preprocessing, including pre-emphasis, and framing aims to compensate for the loss of high-value components. Then, performing fast Fourier transform on each frame signal to transform the time-domain signal into a frequency-domain signal. The spectral energy of each frame is calculated. Finally, the Mel filter is applied to transform frequency-domain signal into Mel-frequency scale to describe the human ear perception of frequency. The Mel-frequency (f_{mel}) could be obtained from the real liner frequency (f_{real}) by the equation:

$$f_{\text{mel}} = 2595 \cdot \log \left(1 + \frac{f_{\text{mel}}}{700} \right). \quad (1)$$

In this study, the average of all frames of an audio segment are taken as features. MFCCs with dimension of 13 were extracted.

2.3.2. Linear prediction coefficient

The basic concept of a linear prediction is that the current sampling value of audio can be approximately replaced by a linear combination of several past sampling values (SUN *et al.*, 2022). A unique set of prediction coefficients can be obtained by approximating the minimum mean square error of the actual audio sampling value and the linear prediction sampling value. LPC have the advantages of fast calculation and effective prediction. The 12-element LPC parameters of each sound segment were extracted, and the average value for each frame of every segment is calculated as the feature vector.

2.3.3. Power ratio

The PR is the ratio of power below and above a certain frequency f_0 . It can roughly reflect the power distribution of audio signals divided by a certain frequency (SUN *et al.*, 2023). The PR can be expressed by:

$$\text{PR}_{f_0} = \log \left(\frac{\sum_{f_i=0}^{f_0} (Y_i)^2}{\sum_{f_i=f_0}^{f_C} (Y_i)^2} \right), \quad (2)$$

where f_C and Y are the cutoff frequency and spectrum of the audio signal, respectively. In this work, f_0 is set as 800 Hz. Four statistic features, including PR_{mean} , PR_{min} , PR_{max} , PR_{var} are calculated to express PR.

2.3.4. Fundamental frequency

The definition of F_0 is the lowest oscillation frequency in a free oscillation system or the lowest frequency in a composite wave. It can reflect the opening and closing time of the vocal cords. In this work, the normalized autocorrelation function is applied to calculate F_0 values for each frame audio signal. The average of all frames of an audio segment are taken as features.

2.3.5. Pitch

The tone is related to the fundamental frequency of the sound, reflecting the information of pitch. The average, minimum, maximum, and variance of all frames of an audio segment are taken as features, which are expressed as $\text{Pitch}_{\text{mean}}$, $\text{Pitch}_{\text{min}}$, $\text{Pitch}_{\text{max}}$, and $\text{Pitch}_{\text{var}}$, respectively.

2.3.6. Crest factor

The CF is defined as the ratio of the waveform peak to the effective value (QIAN *et al.*, 2016):

$$\text{CF} = \frac{V_m}{V_e}, \quad (3)$$

where V_m is the maximum absolute value of an audio signal amplitude, and V_e is the root mean square value of the audio signal amplitude absolute value. It reflects the amplitude of changes in the audio signal in the time domain. The mean value of the peak factor of each frame of the signal is taken as a feature.

2.3.7. Spectrum related features

Spectrum related features are widely used in the analysis of snoring sounds. It can reflect important details of snoring sounds with different types. In this work, spectral cut-off frequency, spectral skewness, spectral slope, spectral variance, spectral kurtosis, spectral entropy, and spectral flux are extracted for further analysis (SUN *et al.*, 2023).

Spectral skewness is a measure of the direction and degree of skewness in the distribution of statistical data, which is a numerical characteristic of the degree of asymmetry in the distribution of statistical data. It is defined as the third-order standard moment of the sample, and the calculation formula is as follows:

$$\text{Skewness}(X) = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right] = \frac{k_3}{\sigma^3} = \frac{k_3}{k_2^{3/2}}, \quad (4)$$

where k_2 and k_3 represent the second- and third-order central moment, respectively.

Spectral slope is a measure of the speed at which the spectrum of an audio signal tilts towards high frequencies, typically calculated using linear regression. Spectral variance is used to measure the degree of dispersion of a sound signal. This can be expressed as follows:

$$\text{Var} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5)$$

For spectral variance, which can reflect the interference of noise on data, this paper uses a noise power function of the carrier frequency, and the spectral variance of the signal can be obtained by the Fourier transform of its autocorrelation function:

$$V(\Omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i2\pi\Omega\tau} \langle y(t)y(t+\tau) \rangle d\tau. \quad (6)$$

Spectral kurtosis can be used to measure the steepness of the probability distribution of random variables. Take the average of the obtained results to obtain the average kurtosis in this work. In this work, the sample entropy is calculated for the entire effective snoring signal. Spectral traffic records the sum of squares of the normalized amplitude differences between two frames, which can describe the changes in adjacent frames. Its definition is

$$Fl_{i,i-1} = \sum_{k=1}^{Wl} (E_i(k) - E_{i-1}(k))^2, \quad (7)$$

$$E_i(k) = \frac{x_i(k)}{\sum_{n=1}^{Wl} x_i(n)}, \quad (8)$$

where $E_i(k)$ is the normalized amplitude, and Wl is the sampling window length.

2.3.8. Formants

Formants are areas in the spectrum of audio signals where energy is relatively concentrated. It reflects the physical characteristics of the vocal tract, namely the degree of contraction of the throat. The first three formant frequencies of snoring sounds are extracted in this work, including the first formant ($F1$), the second

formant ($F2$), and the third formant ($F3$). The average value of all frames is applied to express a piece of snoring sound.

2.4. Feature selection

Studies have indicated that extracted features not only determine the performance of a classification model, but also determine the complexity of the model and influence its computation cost (KURSA, RUDNICKI, 2010; LI *et al.*, 2017). Selecting effective features with high discriminability and low complexity is an important step for machine learning. It can reduce the dimension of features and the complexity of the proposed classification model. In this work, the ReliefF algorithm is applied to select features by calculating the contribution of each feature to the classification task (WU *et al.*, 2020).

The idea of ReliefF algorithm can be simply expressed as: if a feature has the same category to its nearest neighbor (with similar numerical values), the feature weight will be reduced; if the feature is different from its nearest neighbor category, increase its weight. The specific calculation method for the weight W is as follows. Firstly, setting the weights of all features W to 0. When calculating the weight of the j -th feature, an observation value x_o is randomly selected from the feature and the k -observation values are found in the dataset of each category of the feature that are closest in value to the observation value. Updating the weight of the feature parameter by the relationship between each nearest neighbor (x_n) and the observed value (x_o). Then repeating the iterative calculation until all parameters of the feature are traversed. The specific calculation formula is as follows:

- 1) when the observed value x_o is of the same category as the nearest neighbor x_n :

$$W_j^i = W_j^{i-1} - \frac{\Delta j(x_o, x_n)}{m} \cdot don; \quad (9)$$

- 2) when the observed value x_o is different from the category of the nearest neighbor x_n :

$$W_j^i = W_j^{i-1} + \frac{p_{y_n}}{1 - p_{y_o}} * \frac{\Delta j(x_o, x_n)}{m} \cdot don, \quad (10)$$

where W_j^i is the weight of the i -th iteration of the j -th feature; $\Delta j(x_o, x_n)$ is the relative difference between x_o and x_n , where F_j represents the set of the j -th feature parameter, then the expression for $\Delta j(x_o, x_n)$ is

$$\Delta j(x_o, x_n) = \frac{|x_o - x_n|}{\max(F_j) - \min(F_j)}, \quad (11)$$

where don is the formal distance function between x_o and x_n :

$$don = \frac{\tilde{d}_{on}}{\sum_{r=1}^k \tilde{d}_{or}}, \quad (12)$$

$$\tilde{d}_{on} = \exp \left[- \left(\frac{\text{rank}(o, n)}{\text{sigma}} \right)^2 \right], \quad (13)$$

where $\text{rank}(o, n)$ is the corresponding position of a certain nearest neighbor x_n in the total nearest neighbor sorting table of x_o after sorting KNN by distance. Calculate sigma in Eq. (13) to change the scaling ratio, p_{y_o} is the prior probability of the category to which the observed value x_o belongs, p_{y_n} is the prior probability of the category to which the nearest neighbor x_n belongs.

3. Result

3.1. Feature selection

A strategy of feature selection based on the ReliefF algorithm is applied to select features with high robustness and low redundancy. Figure 2 displays the normalized weight value of each feature to the related label. Most features make significant contributions to this classification task, especially for MFCC and pitch features. The importance weights of MFCC1 to MFCC13 are higher than 0.1, which indicates that there is evidently different energy distribution on each frequency band divided by the Mel filter. The MFCC5 to MFCC8

yield the highest weights more than 0.14. These results show that the differences of snoring sounds of simple snorers, normal snoring sounds of OSAHS patients, and abnormal snoring sounds of OSAHS patients mainly concentrated on the low and middle frequency bands. Pitch_{var} also has relatively high weight values, which means that the three kinds of snoring sound have different pitches.

Furthermore, the relationship between the dimension of selected features and the classification results is explored to select optimal features. Figure 3 shows the relationship between the dimension of selected features and the accuracy based on the KNN, RUSBoost, and DNN classifiers. The dimension of features has great influence on the classification results for all classification models. With the increase of the dimension of selecting features, the accuracy of classifiers gradually increases and tends to be stable. When the feature dimension exceeds the optimal one, the classification result will not increase with the increase of the feature dimension. The redundant features not only cannot improve the model classification performance, but also increase the computational complexity of the model. For different classification models, there are significant differences in the degree of influence of features and the dimension of optimal features. The optimal feature dimension is 16, 18, and 37 for KNN, RUSBoost, and DNN classifier, respectively. The related accuracy of KNN, RUSBoost, and DNN model with optimal features are 85.44 %, 84.45 %, and 83.91 %, respectively.

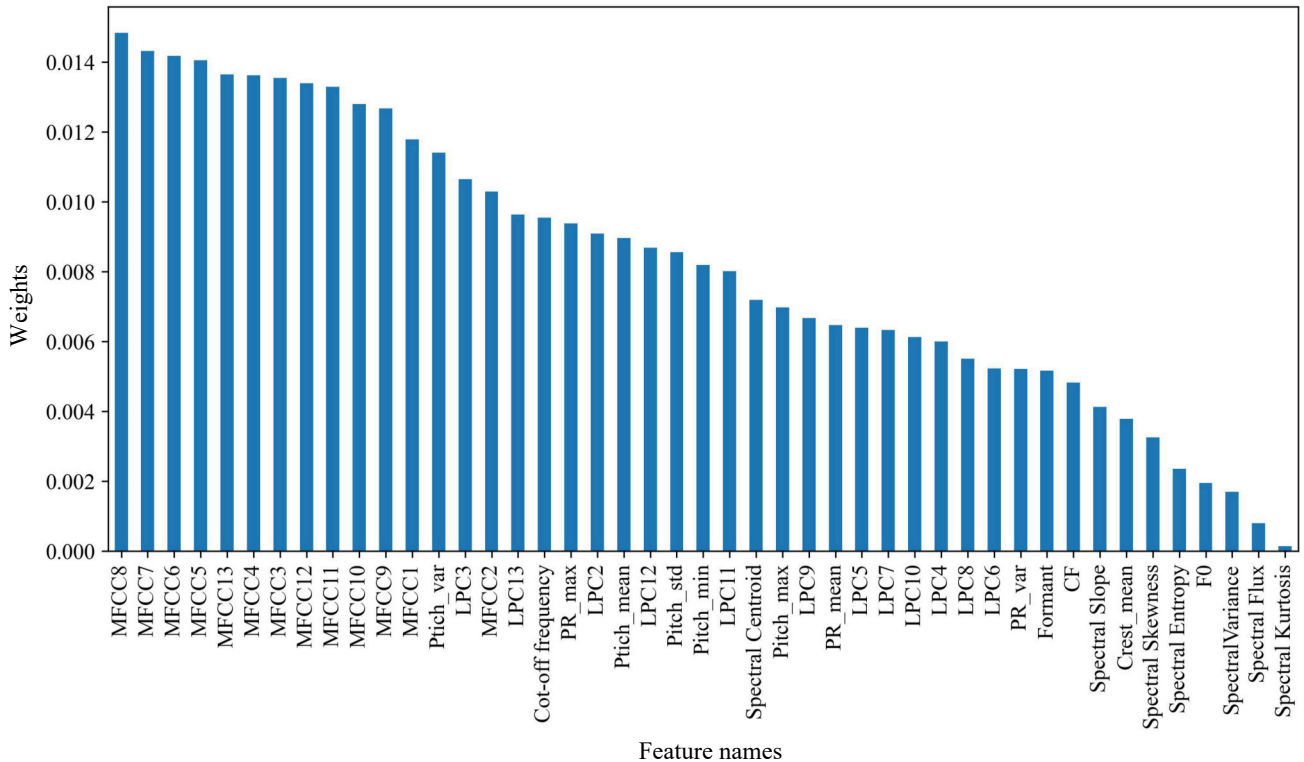


Fig. 2. Normalized weights of each feature obtained by ReliefF algorithm.

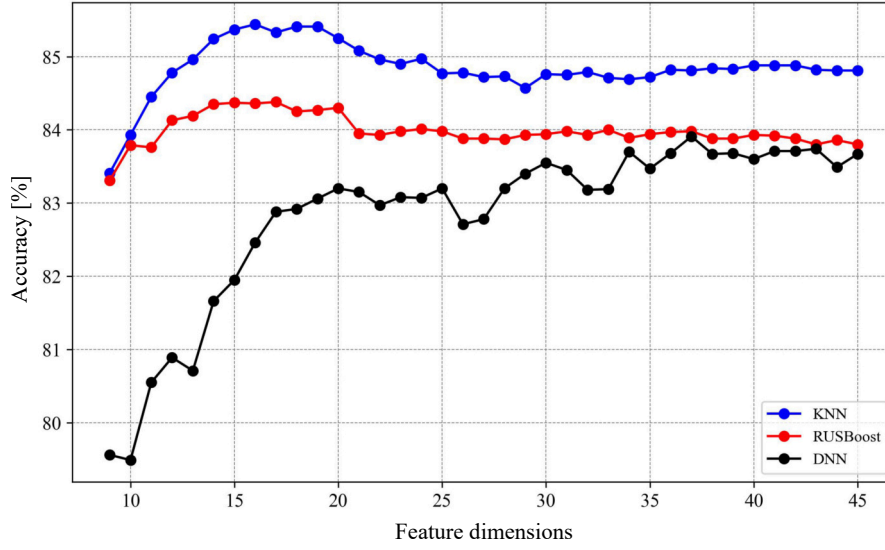


Fig. 3. Relationship of cross-validation average accuracy of KNN, RUSBoost, and DNN with the selected feature dimensions.

3.2. Classification results

Table 1 shows the classification results of SSS, SSSP, and ASSP based on KNN, RUSBoost, DNN classifiers under the original feature set. The accuracy obtained by KNN, RUSBoost, and DNN are 84.81 %, 83.80 %, and 83.67 %, respectively. Under the same feature set, different classifiers may have different emphases. KNN and DNN achieve much higher recall for SSSS and SSSP and lower recall for ASSP than RUSBoost. Specifically, the recall of SSSS, SSSP, and ASSP obtained by DNN are 97.53 %, 91.94 %, and 48.54 %, respectively. The recall of SSSS, SSSP, and ASSP obtained by KNN are 99.14 %, 91.26 %, and 54.50 %, respectively. The recall of SSSS, SSSP, and ASSP obtained by KNN are 97.99 %, 84.60 %, and 69.75 %, respectively.

Table 1. Classification results of SSS, SSSP, and ASSP based on different classifiers under the original feature set.

Snoring type	Evaluation	KNN	RUSBoost	DNN	Fused model
SSSS	Accuracy	0.8481	0.8380	0.8367	0.8556
	Recall	0.9914	0.9799	0.9753	0.9954
	Precision	0.9799	0.9704	0.9900	0.9847
	F1	0.9856	0.9751	0.9826	0.9900
SSSP	Recall	0.9126	0.8460	0.9194	0.9083
	Precision	0.8515	0.8886	0.8314	0.8655
	F1	0.8810	0.8667	0.8732	0.8864
	F1	0.8810	0.8667	0.8732	0.8864
ASSP	Recall	0.5450	0.6975	0.4854	0.5938
	Precision	0.6939	0.6179	0.6839	0.6987
	F1	0.6105	0.6553	0.5678	0.6420

To obtain classification results with higher robustness and stableness, the three basic models KNN,

RUSBoost, and DNN are further fused by the voting strategy. The fused model adequately fuses the advantage of the three basic models. It achieves 85.56 % accuracy, which increases nearly 2 % compared with RUSBoost and DNN. The fused model not only maintains the relatively high recall for SSSS, but also significantly increases the recall of ASSP. The recalls obtained by the fused model are 10.84 % and 4.88 % higher than DNN and KNN, respectively. The recalls of SSSS and SSSP of fused model are 99.57 % and 90.21 %, respectively, which indicates that there are evident differences between SSSS and SSSP. The classification results imply that the upper airway structure of OSAHS patients on the normal sleep is different from that of simple snorers.

To obtain model with lower complexity and high performance, the feature selection strategy is applied in the model. Table 2 shows the classification results of SSS, SSSP, and ASSP based on KNN, RUSBoost, and DNN classifiers under the selected feature set with

Table 2. Classification results of SSS, SSSP, and ASSP based on different classifiers under the selected feature set.

Snoring type	Evaluation	KNN	RUSBoost	DNN	Fused model
SSSS	Accuracy	0.8544	0.8445	0.8391	0.8576
	Recall	0.9926	0.9840	0.9775	0.9957
	Precision	0.9859	0.9830	0.9909	0.9856
	F1	0.9892	0.9835	0.9842	0.9906
SSSP	Recall	0.9090	0.8521	0.9127	0.9021
	Precision	0.8617	0.8930	0.8385	0.8709
	F1	0.8847	0.8721	0.8740	0.8862
ASSP	Recall	0.5841	0.7072	0.5143	0.6162
	Precision	0.6974	0.6257	0.6782	0.6931
	F1	0.6357	0.6639	0.5850	0.6524

the dimension of 16, 18, and 37, respectively. Comparing Tables 1 and 2, the progress of feature selection not only reduces the complexity of the proposed fused model, but also improves the classification of SSSS, SSSP, and ASSP. Compared with the original feature set, the recall of ASSP obtained by the fused model conducting the feature selection improves value of 2.24 %.

Tables 3 and 4 illustrate the confusion matrices of KNN, DNN, RUSBoost, and its related fused model under the original feature set and selected feature set. There is a substantial distinction between snoring sounds of simple snorers and snoring sounds of OSAHS patients. For all classification models, recalls of SSSS are higher than 98 %. Under all test conditions, a certain amount of ASSP and SSSP are mislabeled, resulting in relatively lower recall and precision. The results of Tables 3d and 4d indicate that the proposed fused method could effectively merge the advantage of different classifiers and different features to relatively accurate SSSS, SSSP, and ASSP.

Table 3. Confusion matrices of SSS, SSSP, and ASSP based on different classifiers under the original feature set.

Real label	Predict label			Recall [%]
	SSSS	SSSP	ASSP	
a) KNN-under the original feature set				
SSSS	3213	23	6	99.1
SSSP	43	10 209	935	91.3
ASSP	23	1758	2133	54.5
Precision [%]	98	85.2	69.4	—
b) RUSBoost-under the original feature set				
SSSS	3189	24	28	98.4
SSSP	27	9532	1628	85.2
ASSP	28	1118	2768	70.7
Precision [%]	98.3	89.3	62.6	—
c) DNN-under the original feature set				
SSSS	3161	78	2	97.5
SSSP	26	10 285	876	91.9
ASSP	6	2008	1900	48.5
Precision [%]	99	83.1	68.4	—
d) Fusion-under the original feature set				
SSSS	3226	11	4	99.5
SSSP	28	10 161	998	90.8
ASSP	22	1568	2324	59.4
Precision [%]	98.5	86.6	69.9	—

Table 5. Literature reviews about snoring sounds classification of OSAHS patients.

Author	Subjects	Feature	Validation method	Accuracy [%]
CHENG <i>et al.</i> (2022)	44	MFCC, LPC, Fbanks	LSTM	81.60
DING <i>et al.</i> (2023)	50	Mel-spectrogram	VGG19+LSTM	85.21
SONG <i>et al.</i> (2023)	40	Mel-spectrogram	CNN, ResNet, and XGBoost fused model	83.44
Shen <i>et al.</i> (2020)	32	MFCC, LPCC, and LPMFCC	LSTM	87.00
Hou <i>et al.</i> (2019)	120	MFCC	GMMs	80.00
This work	40	A series of acoustic features	KNN, RUSBoost, and DNN fused model	85.76

Table 4. Confusion matrices of SSS, SSSP, and ASSP based on different classifiers under the selected feature set.

Real label	Predict label			Recall [%]
	SSSS	SSSP	ASSP	
a) KNN-under the selected feature set				
SSSS	3217	20	4	99.3
SSSP	30	10 169	988	90.9
ASSP	16	1612	2286	58.4
Precision [%]	98.6	86.2	69.7	–
b) RUSBoost-under the selected feature set				
SSSS	3176	37	28	98
SSSP	63	9464	1660	84.6
ASSP	34	1150	2730	69.8
Precision [%]	97	88.9	61.8	–
c) DNN-under the selected feature set				
SSSS	3168	70	3	97.8
SSSP	25	10 210	952	91.3
ASSP	4	1897	2013	51.4
Precision [%]	99.1	83.9	67.8	–
d) Fusion-under the selected feature set				
SSSS	3227	12	2	99.6
SSSP	29	10 092	1066	90.2
ASSP	18	1484	2412	61.6
Precision [%]	98.6	87.1	69.3	–

4. Discussion

In this study, a fused model based on KNN, RUSBoost, and DNN is proposed to classify SSSS, SSSP, and APSP. The ReliefF algorithm is applied to select optimal features in each basic model. The hard voting strategy is employed to fuse the three basic models. The feature selection and model fusion strategies evidently improve the classification performance of the proposed model. Experiment results show that the proposed model achieves 85.76 % accuracy. The recall and precision of SSSS are 99.57 % and 98.56 %, respectively. The recall and precision of SSSP are 90.21 % and 87.09 %, respectively. The recall and precision of ASSP are 61.62 % and 69.31 %, respectively.

Table 5 displays details of studies on the identification of APSP. Since there is no open snoring dataset with label, studies of analysis of snoring sounds are based on dataset collected and labeled by their own labs. The unavoidable situation makes it impossible to compare the performance of different classifica-

tion models directly. As Table 5 shows, these studies are capable of classifying snoring sounds with apnea-hypopnea events or without apnea-hypopnea events. Specifically, CHENG *et al.* (2022) extracted acoustic features including MFCC, LPC and used LSTM to classify SSSS, normal snoring sounds of OSAHS patients, and post-apnea snoring sounds of OSAHS patients with accuracy of 81.6 %. Their work had high recall for SSSS and normal snoring sounds of OSAHS patients and low recall for post-snoring sounds of OSAHS patients with value of 88.1 %, 93.4 %, and 63.5 %, respectively. The classification model proposed by this work achieved recall with values of 99.87 %, 90.21 %, and 61.26 % for SSSS, SSSP, and APSP, which are relatively better than the mentioned studies. The comparison demonstrates that the fused model yields higher classification result and better robustness. Since the snoring sound is generated by the vibration of the upper airway, the classification results of SSSS, SSSP, and APSP demonstrate that the structure of the upper airway is evidently different from that of OSAHS patients. Obesity, smoking, and other pathological reasons cause the upper airway of OSAHS patients gets narrow (GHOSH *et al.*, 2021). The narrow upper airway is the main reason for the occurrence of apnea and hypopnea events of OSAHS patients. The classification results of simple snoring sounds of OSAHS patients and apnea-hypopnea snoring sounds of OSAHS patients indicate that OSAHS patients snore continually throughout the whole night, which is caused by the narrow upper airway. It can be said that the narrow upper airway not only induces hypopnea and apnea events during sleep, but also negatively influences the normal sleep qualities and frequently causes snoring sounds. Furthermore, the high recall and precision of SSSS show solid experimental verification for identifying simple snorers and OSAHS patients based on analysis of snoring sounds. These studies mentioned in Table 5 are concentrated on distinguishing simple snoring sounds of OSAHS patients and apnea-hypopnea snoring sounds of OSAHS patients (CHENG *et al.*, 2022; DING *et al.*, 2023; SHEN *et al.*, 2020; SONG *et al.*, 2023). The accuracies of all classification model are higher than 80 %. These results indicate that there are evident differences among snoring sounds occurred in different sleep stages for OSAHS patients.

SONG *et al.* (2023) proposed a CNN, ResNet, and XGBoost fused model to classify snoring sounds occurred in different sleep stages and achieved 83.44 % accuracy. The classification model may be only concentrated on differences at one latitude and achieve limited classification results. The model fusion strategy based on different fusion methods is proposed to fuse basic classification models that has been widely used in different kinds of classification tasks. In this work, a hard voting fusion strategy is applied to fuse

KNN, RUSBoost, and DNN classifiers. This method significantly increases classification recall and precision of SSSS, SSSP, and ASSP. It also improves the effectiveness and robustness of the proposed model. Experiment results show promising foreground for diagnosing severities of OSAHS patients based on analysis of snoring sounds.

There are also some limitations of the proposed model. Firstly, validation experiments of this work are conducted based on subject dependence. It mainly focuses on exploring differences among these types of snoring sounds. Further subject independent experiments should be conducted to validate the generation error and robustness of the proposed model. Moreover, the proposed model just focuses on exploring differences among snoring sounds occurred in different sleep stages. The relationship between apnea-hypopnea snoring sounds and apnea-hypopnea events should be studied to identify apnea-hypopnea events and estimate AHI values of OSAHS patients.

5. Conclusion

In this work, a fused model based on KNN, RUSBoost, and DNN is proposed to classify SSSS, SSSP, and APSP. Firstly, a series of acoustic features are extracted to express snoring sounds. Three classifiers KNN, RUSBoost, and DNN are independently trained. The ReliefF algorithm is applied to select features in each classification model. A hard voting strategy is used to obtain an optimal fused model. Experiment results show that the proposed fused model achieves high performance with accuracy of 85.76 %. The recalls of SSSS, SSSP, and APSP obtained by the proposed model are 99.87 %, 90.21 %, and 61.26 %, respectively. It demonstrates the effectiveness and validity of assisting in diagnosing OSAHS patients based on analysis of snoring sounds.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 11974121) and National Youth Foundation of China (grant no. 81900927).

Data availability statement

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

Ethical approval

This study was approved by the Ethics Committee of Guangzhou Medical University and an informed consent was obtained from each participant.

Conflicts of interest

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

References

1. ALSHAER H., HUMMEL R., MENDELSON M., MARSHAL T., BRADLEY T.D. (2019), Objective relationship between sleep apnea and frequency of snoring assessed by machine learning, *Journal of Clinical Sleep Medicine*, **15**(3): 463–470, <https://doi.org/10.5664/jcsm.7676>.
2. AZARBARZIN A., MOUSSAVI Z. (2013), Snoring sounds variability as a signature of obstructive sleep apnea, *Medical Engineering and Physics*, **35**(4): 479–485, <https://doi.org/10.1016/j.medengphy.2012.06.013>.
3. BERRY R.B. *et al.* (2012), Rules for scoring respiratory events in sleep: Update of the 2007 AASM manual for the scoring of sleep and associated events, *Journal of Clinical Sleep Medicine*, **8**(5): 597–619, <https://doi.org/10.5664/jcsm.2172>.
4. CARON C.J.J.M. *et al.* (2017), Obstructive sleep apnoea in craniofacial microsomia: Analysis of 755 patients, *International Journal of Oral and Maxillofacial Surgery*, **46**(10): 1330–1337, <https://doi.org/10.1016/j.ijom.2017.05.020>.
5. CASTILLO-ESCARIO Y., FERRER-LLUIS I., MONTSERRAT J.M., JANE R. (2019), Entropy analysis of acoustic signals recorded with a smartphone for detecting apneas and hypopneas: A comparison with a commercial system for home sleep apnea diagnosis, *IEEE Access*, **7**: 128224–128241, <https://doi.org/10.1109/ACCESS.2019.2939749>.
6. CHENG S. *et al.* (2022), Automated sleep apnea detection in snoring signal using long short-term memory neural networks, *Biomedical Signal Processing and Control*, **71**(Part B): 103238, <https://doi.org/10.1016/j.bspc.2021.103238>.
7. DING L., PENG J., SONG L., ZHANG X. (2023), Automatically detecting apnea-hypopnea snoring signal based on VGG19 + LSTM, *Biomedical Signal Processing and Control*, **80**(Part 2): 104351, <https://doi.org/10.1016/j.bspc.2022.104351>.
8. DING L., PENG J., SONG L., ZHANG X. (2024), Automatically detecting OSAHS patients based on transfer learning and model fusion, *Physiological Measurement*, **45**(5): 055013, <https://doi.org/10.1088/1361-6579/ad4953>.
9. ECKERT D.J., JORDAN A.S., MERCHIA P., MALHOTRA A. (2007), Central sleep apnea: Pathophysiology and treatment, *Chest*, **131**(2): 595–607, <https://doi.org/10.1378/chest.06.2287>.
10. FIZ J.A. *et al.* (1996), Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea, *European Respiratory Journal*, **9**(11): 2365–2370, <https://doi.org/10.1183/09031936.96.09112365>.
11. FRIEDMAN M., IBRAHIM H., JOSEPH N.J. (2004), Staging of obstructive sleep apnea/hypopnea syndrome: A guide to appropriate treatment, *Laryngoscope*, **114**(3): 454–459, <https://doi.org/10.1097/00005537-200403000-00013>.
12. GHOSH P., VARMA N.K.S., AJITH V.V., SURESH A. (2021), Upper airway and its association with neck circumference and hyoid position in OSA subjects – A cephalometric study, *International Journal of Current Research and Review*, **13**(6): 167–171, <https://doi.org/10.31782/IJCRR.2021.13610>.
13. GISLASON T., BENEDIKTSÐÓTTIR B. (1995), Snoring, apneic episodes, and nocturnal hypoxemia among children 6 months to 6 years old: An epidemiologic study of lower limit of prevalence, *Chest*, **107**(4): 963–966, <https://doi.org/10.1378/chest.107.4.963>.
14. GOTTLIEB D.J., PUNJABI N.M. (2020), Diagnosis and management of obstructive sleep apnea: A review, *Journal of the American Medical Association*, **323**(14): 1389–1400, <https://doi.org/10.1001/jama.2020.3514>.
15. HERZOG M., SCHMIDT A., BREMERT T., HERZOG B., HOSEMAN W., KAFTAN H. (2008), Analysed snoring sounds correlate to obstructive sleep disordered breathing, *European Archives of Oto-Rhino-Laryngology*, **265**(1): 105–113, <https://doi.org/10.1007/s00405-007-0408-8>.
16. HOU L., ZHANG W., SHI D., LIU H. (2019), Estimation of apnea hypopnea index based on acoustic features of snoring, *Journal of Shanghai University (Natural Science)*, **25**(4): 435–444, <https://doi.org/10.12066/j.issn.1007-2861.1942>.
17. IZCI B., DOUGLAS N.J. (2012), Obstructive sleep apnea-hypopnea syndrome, [in:] *Obstructive Sleep Apnea: Causes, Treatment and Health Implications*, Sacchetti L.M., Mangiardi P. [Eds.], pp. 129–182, Nova Science Publishers.
18. JANIESCH C., ZSCHECH P., HEINRICH K. (2021), Machine learning and deep learning, *Electronic Markets*, **31**(3): 685–695, <https://doi.org/10.1007/s12525-021-00475-2>.
19. KARUNAJEEWA A.S., ABEYRATNE U.R., HUKINS C. (2011), Multi-feature snore sound analysis in obstructive sleep apnea-hypopnea syndrome, *Physiological Measurement*, **32**(1): 83, <https://doi.org/10.1088/0967-3334/32/1/006>.
20. KOROMPILI G. *et al.* (2021), PSG-Audio, a scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies, *Scientific Data*, **8**(1): 197, <https://doi.org/10.1038/s41597-021-00977-w>.
21. KURSA M.B., RUDNICKI W.R. (2010), Feature selection with the Boruta package, *Journal of Statistical Software*, **36**(11): 1–13, <https://doi.org/10.18637/jss.v036.i11>.
22. LEE B.S., ELLIS D.P.W. (2012), Noise robust pitch tracking by subband autocorrelation classification, [in:] *13th Annual Conference of the International Speech Communication Association 2012*, <https://doi.org/10.21437/interspeech.2012-221>.

23. LI J. *et al.* (2017), Feature selection: A data perspective, *ACM Computing Surveys*, **50**(6): 94, <https://doi.org/10.1145/3136625>.
24. LUGARESI E., CIRIGNOTTA F., MONTAGNA P. (1988), Pathogenic aspects of snoring and obstructive apnea syndrome, *Schweizerische Medizinische Wochenschrift*, **118**(38).
25. MARKANDEYA M.N., ABEYRATNE U.R., HUKINS C. (2018), Characterisation of upper airway obstructions using wide-band snoring sounds, *Biomedical Signal Processing and Control*, **46**: 201–211, <https://doi.org/10.1016/j.bspc.2018.07.013>.
26. MINARITZOGLOU A., VAGIAKIS E. (2008), Polysomnography: Recent data on procedure and analysis, *Pneumonia*, **21**(4).
27. OSMAN A.M., CARTER S.G., CARBERRY J.C., ECKERT D.J. (2018), Obstructive sleep apnea: Current perspectives, *Nature and Science of Sleep*, **10**: 21–34, <https://doi.org/10.2147/NSS.S124657>.
28. PEREZ-PADILLA J.R., SLAWINSKI E., DIFRANCESCO L.M., FEIGE R.R., REMMERS J.E., WHITELAW W.A. (1993), Characteristics of the snoring noise in patients with and without occlusive sleep apnea, *American Review of Respiratory Disease*, **147**(3), <https://doi.org/10.1164/ajrccm/147.3.635>.
29. PEVERNAGIE D., AARTS R.M., DE MEYER M. (2010), The acoustics of snoring, *Sleep Medicine Reviews*, **14**(2): 131–144, <https://doi.org/10.1016/j.smrv.2009.06.002>.
30. QIAN K. *et al.* (2021), Can machine learning assist locating the excitation of snore sound? A review, *IEEE Journal of Biomedical and Health Informatics*, **25**(4): 1233–1246, <https://doi.org/10.1109/JBHI.2020.3012666>.
31. QIAN K., JANOTT C., ZHANG Z., HEISER C., SCHULLER B. (2016), Wavelet features for classification of vote snore sounds, [in:] *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 221–225, <https://doi.org/10.1109/ICASSP.2016.7471669>.
32. REDLINE S. *et al.* (2010), Obstructive sleep apnea–hypopnea and incident stroke, *American Journal of Respiratory and Critical Care Medicine*, **182**(2), <https://doi.org/10.1164/rccm.200911-1746oc>.
33. SEIFFERT C., KHOSHGOFTAAAR T.M., VAN HULSE J., NAPOLITANO A. (2010), RUSBoost: A hybrid approach to alleviating class imbalance, [in:] *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, **40**(1): 185–197, <https://doi.org/10.1109/TSMCA.2009.2029559>.
34. SHEN F., CHENG S., LI Z., YUE K., LI W., DAI L. (2020), Detection of snore from OSAHS patients based on deep learning, *Journal of Healthcare Engineering*, <https://doi.org/10.1155/2020/8864863>.
35. SOLÀ-SOLER J., JANÉ R., FIZ J.A., MORERA J. (2007), Automatic classification of subjects with and without Sleep Apnea through snoring analysis, [in:] *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, <https://doi.org/10.1109/IEMBS.2007.4353739>.
36. SONG Y., SUN X., DING L., PENG J., SONG L., ZHANG X. (2023), AHI estimation of OSAHS patients based on snoring classification and fusion model, *American Journal of Otolaryngology*, **44**(5): 103964, <https://doi.org/10.1016/j.amjoto.2023.103964>.
37. SOWHO M., SGAMBATI F., GUZMAN M., SCHNEIDER H., SCHWARTZ A. (2020), Snoring: A source of noise pollution and sleep apnea predictor, *Sleep*, **43**(6), <https://doi.org/10.1093/sleep/zsz305>.
38. SUN X., DING L., SONG Y., PENG J., SONG L., ZHANG X. (2023), Automatic identifying OSAHS patients and simple snorers based on Gaussian mixture models, *Physiological Measurement*, **44**(4): 045003, <https://doi.org/10.1088/1361-6579/accd43>.
39. SUN X., PENG J., ZHANG X., SONG L. (2022), Effective feature selection based on Fisher Ratio for snoring recognition using different validation methods, *Applied Acoustics*, **185**: 108429, <https://doi.org/10.1016/j.apacoust.2021.108429>.
40. ULUALP S.O. (2010), Snoring and obstructive sleep apnea, *Medical Clinics of North America*, **94**(5): 1047–1055, <https://doi.org/10.1016/j.mcna.2010.05.002>.
41. WANG C., PENG J., SONG L., ZHANG X. (2017), Automatic snoring sounds detection from sleep sounds via multi-features analysis, *Australasian Physical and Engineering Sciences in Medicine*, **40**(1): 127–135, <https://doi.org/10.1007/s13246-016-0507-1>.
42. WHITE D.P. (2005), Pathogenesis of obstructive and central sleep apnea, *American Journal of Respiratory and Critical Care Medicine*, **172**(11), <https://doi.org/10.1164/rccm.200412-1631SO>.
43. WU Z., WANG X., JIANG B. (2020), Fault diagnosis for wind turbines based on ReliefF and eXtreme gradient boosting, *Applied Sciences*, **10**(9): 3258, <https://doi.org/10.3390/app10093258>.
44. ZHANG S., LI X., ZONG M., ZHU X., CHENG D. (2017), Learning k for kNN Classification, *ACM Transactions on Intelligent Systems and Technology*, **8**(3): 43, <https://doi.org/10.1145/2990508>.
45. ZHENG F., ZHANG G., SONG Z. (2001), Comparison of different implementations of MFCC, *Journal of Computer Science and Technology*, **16**(6): 582–589, <https://doi.org/10.1007/BF02943243>.

Research Paper

Estimating Ensemble Location and Width in Binaural Recordings of Music with Convolutional Neural Networks

Paweł ANTONIUK^{*}, Sławomir K. ZIELIŃSKI¹*Faculty of Computer Science, Białystok University of Technology*
Białystok, Poland^{*}Corresponding Author e-mail: pawel.antonik@sd.pb.edu.pl*(received July 7, 2024; accepted October 31, 2024; published online February 10, 2025)*

Binaural audio technology has been in existence for many years. However, its popularity has significantly increased over the past decade as a consequence of advancements in virtual reality and streaming techniques. Along with its growing popularity, the quantity of publicly accessible binaural audio recordings has also expanded. Consequently, there is now a need for automated and objective retrieval of spatial content information, with ensemble location and width being the most prominent. This study presents a novel method for estimating these ensemble parameters in binaural recordings of music. For this purpose, a dataset of 23 040 binaural recordings was synthesized from 192 publicly-available music recordings using 30 head-related transfer functions. The synthesized excerpts were then used to train a multi-task spectrogram-based convolutional neural network model, aiming to estimate the ensemble location and width for unseen recordings. The results indicate that a model for estimating ensemble parameters can be successfully constructed with low prediction errors: 4.76° ($\pm 0.10^\circ$) for ensemble location and 8.57° ($\pm 0.19^\circ$) for ensemble width. The method developed in this study outperforms previous spatiogram-based techniques recently published in the literature and shows promise for future development as part of a novel tool for binaural audio recordings analysis.

Keywords: ensemble width; ensemble location; binaural; spatial audio; localization; convolutional neural network; head-related transfer function; angle of arrival.



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

1. Introduction

The human auditory system demonstrates exceptional proficiency in segregating, localizing, and interpreting diverse auditory signals, despite being limited to two ears. This is possible, among other factors, by internal examination of interaural differences in time, loudness, and frequency, known as binaural hearing (BLAUERT, 1996), which enables precise localization of sound sources in complex auditory environments. A notable advantage of binaural hearing is exemplified by the “cocktail party effect”, highlighting humans’ capability to concentrate on foreground sound sources while suppressing background noise (CHERRY, 1953). Understanding the auditory system is essential for comprehending its limits but also for leveraging these insights to create more immersive binaural experiences for entertainment purposes (ZHANG *et al.*, 2017). It is

also important for enhancing auditory signal reception in hearing aid devices (HIRSH, 1950; THIEMANN *et al.*, 2016).

The advance of sophisticated machine learning techniques, especially deep learning networks, has initiated an interesting exploration of their potential to emulate the human auditory system. Recently emerged studies have demonstrated that relying on the advanced spatial audio feature engineering is not necessary in computational audio source localization (PANG *et al.*, 2019; VERA-DIAZ *et al.*, 2018; YANG, ZHENG, 2022). While applying convolutional neural networks – CNNs (LECUN *et al.*, 1989) to audio signals is well-established, often in conjunction with spectrograms (ESPI *et al.*, 2015; HAN *et al.*, 2017; THOMAS *et al.*, 2014) or other feature engineering techniques (ABDEL-HAMID *et al.*, 2012; SAINATH *et al.*, 2013), these approaches continue to be refined and adapted for au-

dio processing. Building on these foundations, this study develops an audio localization method using a spectrogram-based multi-task CNN model.

Humans tend to localize groups of sound sources rather than individual ones (BREGMAN, 1994; RUMSEY, 2002). Inspired by this fact, the objective of the proposed model is to estimate the location and width of these groups, termed “ensembles”, instead of the positions of individual sources. This study is unique as it not only developed the method but also tested it on a relatively large, realistic music corpus. The corpus comprised 23 040 binaural excerpts synthesized using 192 multi-track music recordings (from a repository provided by SENIOR (2023)) and 30 sets of publicly available head-related transfer functions (HRTFs) acquired from various sources (see Table 1 in Appendix for a detailed list). The music recordings covered many different genres, including rock, jazz, pop, and classical music.

The findings demonstrate that this method is effective in accurately estimating the spatial characteristics of groups of sound sources in near-real-world scenarios. This paper also demonstrates an experimental framework that facilitates the objective measurement of a binaural localization technique, employing a large-scale dataset synthesized from real-world music signals (for applications of similar frameworks, see studies conducted by ANTONIUK and ZIELIŃSKI (2023) and ZIELIŃSKI *et al.* (2020; 2022a; 2022b)). One of the key advantages of the proposed method is that it does not assume the number of audio sources. However, significant limitations of this study include the absence of reverberation in the synthesized recordings and the method’s inapplicability to real-time scenarios – both are critical areas for future research.

The developed method has the potential to be highly beneficial in automated information retrieval tasks, where a significant number of binaural recordings must be analyzed or labeled in terms of their spatial content information. This could be utilized in the development of a hypothetical autonomous “web-crawler bot” that will collect binaural recordings from publicly accessible repositories and label them according to the spatial properties of the sound sources, such as the location of the music ensemble or the sparsity of audio source positions. This method may also assist audio engineers in objectively assessing and segregating binaural audio recordings with regard to their spatial content.

This paper is structured as follows: Sec. 2 presents related studies. The description of the method developed for this study is provided in Sec. 3, which also includes detailed definitions of ensemble location and width, along with a description of the experiments used to evaluate this method. Section 4 presents and discusses the performance of the proposed method as well as the results of the experiments conducted in this

study. Finally, Sec. 5 offers concluding remarks and suggestions for future research.

2. Related studies

Most existing literature on computational sound source localization reports techniques that take advantage of multiple microphone arrays with more than two channels (CHUNG *et al.*, 2022; HAHMANN *et al.*, 2022; KAVEH, BARABELL, 1986; LIU *et al.*, 2022; PAN *et al.*, 2021; PAVLIDI *et al.*, 2012). Although these methods can improve localization precision by providing additional spatial information, they do not utilize binaural hearing, rendering them ineffective for binaural recordings. In the context of sound source localization in binaural signals, the focus of research is put on the identification of individual sound sources, rather than groups of sounds (BENAROYA *et al.*, 2018; DIETZ *et al.*, 2011; MA, BROWN, 2016; MA *et al.*, 2017; MAY *et al.*, 2011; 2012; 2015; WOODRUFF, WANG, 2012).

Considering source direction of arrival (DoA) methods, the majority of research assumes a fixed number of sound sources (ARTHI, SREENIVAS, 2021; MA *et al.*, 2017; PANG *et al.*, 2019; VERA-DIAZ *et al.*, 2018; WOODRUFF, WANG, 2012), which limits its practical applications as this information is rarely known in real-life binaural recordings. Moreover, the majority of studies have focused on relatively homogeneous signals, namely speech (BENAROYA *et al.*, 2018; DIETZ *et al.*, 2011; LIU *et al.*, 2018; MA, BROWN, 2016; MA *et al.*, 2017; 2018; MAY *et al.*, 2011; 2012; 2015; WANG *et al.*, 2020; WOODRUFF, WANG, 2012; YANG, ZHENG, 2022).

In contrast to the aforementioned studies, the proposed method is not constrained by the number of sources. Moreover, the approach is not narrowed to speech and has been applied to a wide range of musical datasets, including instruments and vocals. In contrast to studies that primarily focused on individual sources, the proposed method does not aim to separate them, but rather considers them as a group, or in this case – a musical ensemble – similar to how real musical ensembles are arranged on stage. To the authors’ knowledge, this is one of the first methods to localize ensemble width (see (ANTONIUK, ZIELIŃSKI, 2023) for the previous ensemble-width-related study), and the first to localize both ensemble position and width simultaneously using a multi-task model.

Sound localization methods can be classified into two categories based on the implementation of their underlying algorithms, termed as glass-box and black-box techniques. Glass-box methods could be considered as more traditional in the literature. They rely on manually designed algorithms that mimic the auditory system to explicitly extract key features for the localization estimation, such as interaural level differences, interaural time differences, interaural coherence,

or interaural phase differences (BLAUERT (1996) provides detailed descriptions of these features). Examples of glass-box methods can be found in numerous studies, including those conducted by DIETZ *et al.* (2011), MA, BROWN (2016), MA *et al.* (2017; 2018), MAY *et al.* (2011; 2012; 2015), WOODRUFF, WANG (2012), and ZIELIŃSKI *et al.* (2022b). These features are typically extracted using an auditory model. An advanced implementation capable of extracting these features was developed as part of the Two!Ears project (RAAKE, 2016).

Black-box methods use a minimal degree of feature engineering, depending on deep neural networks to both extract features and make estimations. While effective, these methods do not necessarily consistently mimic human hearing, rendering them less suitable for objective measurement tasks (e.g., VERA-DIAZ *et al.* (2018), YANG and ZHENG (2022)). Additionally, it is challenging to reveal their internally extracted features. Due to their opacity, unpredictable results, and numerous learning parameters, these methods should be treated more carefully. Moreover, they require large datasets for their development and evaluation. These datasets often contain thousands of examples, such as the TIMIT corpus (GAROFALO *et al.*, 1993) used in multiple studies (BENAROYA *et al.*, 2018; MA *et al.*, 2017; 2018; MAY *et al.*, 2015; PANG *et al.*, 2019; VERA-DIAZ *et al.*, 2018; WANG *et al.*, 2020; YANG, ZHENG, 2022). Some researchers have even created custom corpora with hundreds of thousands of recordings (ANTONIUK, ZIELIŃSKI, 2023; ZIELIŃSKI *et al.*, 2020; 2022a; 2022b).

The necessity of having a large corpus to train deep learning models poses a significant challenge in gathering a sufficiently large and diverse collection of labeled binaural recordings. However, this challenge can be addressed through the synthesis of binaural sounds, as demonstrated in various studies (ANTONIUK, ZIELIŃSKI, 2023; MA *et al.*, 2018; YANG, ZHENG, 2022; ZIELIŃSKI *et al.*, 2020; 2022a; 2022b) and discussed further in Subsec. 3.2.

3. Methodology

This part of the paper presents a detailed description of the model developed in this study, as outlined in Subsec. 3.1. It also describes the audio dataset used for training and evaluating the model, as detailed in Subsec. 3.2. In Subsec. 3.3., the spectrogram calculation procedure is presented. Subsection 3.4. describes the model topology, whereas Subsec. 3.5 addresses model training and evaluation.

3.1. Ensemble location and width definition

The objective of the model developed in this study is to estimate the ensemble location (θ) and width (ω),

as illustrated in Fig. 1. An ensemble is defined as a group of audio point sources positioned on a circle around the listener on a virtual acoustic scene with an equal distance to the listener. The location of source i is denoted by θ_i . The ensemble width (ω) is defined as the angular distance between two extreme point sources ($\max_i(\theta_i) - \min_i(\theta_i)$), while the ensemble location, designated by θ , represents the middle angle between two extreme sound sources ($\frac{(\max_i(\theta_i) + \min_i(\theta_i))}{2}$). For the purposes of this study, the locations of the sources were limited to the frontal hemisphere only, i.e., $\theta \in [-45^\circ, 45^\circ]$, $\omega \in [0^\circ, 90^\circ]$, as this range encompasses the majority of real-world recording scenarios. It should be noted that although humans possess some limited abilities to localize sound sources in the vertical plane, in this study all sources are placed in the horizontal plane, at the ear-level of the listener. This covers the majority of cases for real-world recordings (see (MA *et al.*, 2018; ZIELIŃSKI *et al.*, 2022a) for related studies that cover top-down discrimination).

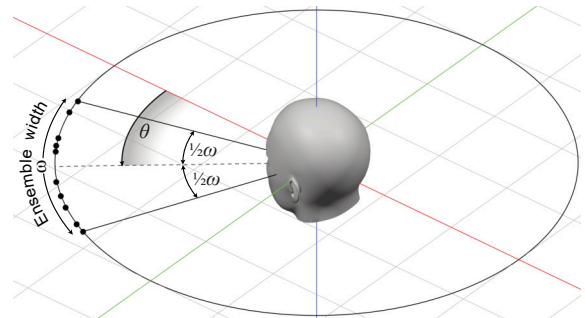


Fig. 1. Illustration of ensemble width (ω) and ensemble location (θ) relative to the direction of the head orientation. Black dots represent the positions of audio sources θ_i . The ensemble location (θ) is the angular position of the center of the ensemble relative to the direction the head is facing. The ensemble width (ω) is the angular distance between the two most extreme audio sources in the ensemble.

3.2. Synthesis of binaural music recordings

The experiments conducted in this study involved 23040 binaural recordings of music. These recordings were synthesized using 192 publicly-available multi-track music recordings (SENIOR, 2023) and 30 HRTF databases (see Table 1 in Appendix for a detailed list). The large number of HRTF databases was necessary to make the model as generalizable as possible. In real-world scenarios, the HRTF used for binaural synthesis is often unknown, so constructing a model for a single HRTF would have limited practical utility. The aim was to predict ensemble parameters regardless of the specific HRTF function used. Additionally, the large number of HRTF functions increased the amount of data available for model training, which is particularly beneficial in the context of deep neural networks. The number of HRTF databases (30) was determined using

heuristics from previous study conducted by [ZIELIŃSKI *et al.* \(2022b\)](#), which suggested this number should be sufficient for the task.

The number of tracks in multi-track recording ranged from 5 to 62, with median of 9. For each pair of a multi-track recording and an HRTF database, four binaural recordings were synthesized with different random ensemble parameters, namely location θ and width ω , as defined in Subsec. 3.1. Both parameters were drawn from a uniform random distribution. Furthermore, the tracks of the input multi-track recordings were randomly assigned to sound source positions (θ_i) to enhance the diversity of the final binaural corpora. Before the synthesis, the signals in each track were equalized to -23 LKFS, in accordance with ([ITU, 2023](#)) recommendation.

The binaural recordings were obtained in this study using the binaural synthesis procedure, known as binauralization, whose aim was to simulate the positions of sound sources within a virtual acoustic environment ([BLAUERT, 1996](#)). This was achieved by convolving multi-track signals with head-related impulse responses from a specified HRTF database. The resulting binaural output signal $y_c[n]$ for each stereo channel c (left or right) at a sample n is given by the equation:

$$y_c[n] = \sum_{i=1}^N \sum_{k=0}^{K-1} x_i[k] \times h_{c,\theta_i}[n-k], \quad (1)$$

where x_i represents the signal of an individual sound source i from the input music recording and h_{c,θ_i} denotes the head-related impulse response for channel c at location θ_i of source track i .

After the binauralization procedure, the synthesized recordings were truncated to a duration of seven seconds, with sine-squared fade-in and fade-out effects of 0.01 seconds applied. The recordings were then RMS-normalized, scaled by a factor of 0.9, and DC-offset corrected. They were stored as uncompressed files at 48 000 samples per second and with a 32-bit resolution.

Due to copyright restrictions, the music corpus utilized in this study was not published. However, the corpus can be provided upon reasonable request from the authors of this paper.

3.3. Calculation of spectrograms

Prior being input into the model, the binaural recordings of music were transformed into magnitude spectrograms. Although spectrograms do not directly provide information that can be translated into ensemble features, especially the ensemble width, the goal of this task was to reduce the number of independent variables compared to the raw audio signal by extracting more compressed and informative data in the frequency domain. This step was also necessary to decrease the likelihood of overfitting, reducing the num-

ber of examples needed to train the model, and thereby lower the overall computational power requirements. It is worth mentioning, however, that recently published studies have shown that CNNs are suitable for end-to-end audio localization without the spectrogram extraction step, as demonstrated by [VECCHIOTTI *et al.* \(2019\)](#) and [VERA-DIAZ *et al.* \(2018\)](#).

To prepare the input for the model, a Hamming window of 40 ms with an overlap of 20 ms was applied to each frame of the signal, resulting in a total of 349 time frames. From each frame, spectrograms were extracted using the fast Fourier transform (FFT) algorithm, with 150 frequency bands spaced linearly from 100 Hz to 16 kHz. This procedure was conducted for both the left and right channels, yielding two spectrograms for each binaural sample. Consequently, each sample was represented by the 32-bit floating-point precision matrix of dimensions $2 \times 349 \times 150$. This method parallels the procedure presented by [ZIELIŃSKI *et al.* \(2022b\)](#).

3.4. Network topology

The network topology employed in this study was strongly influenced by the AlexNet convolutional neural network introduced by [KRIZHEVSKY *et al.* \(2012\)](#). While AlexNet was originally designed for image classification, in this study it was adapted for the audio analysis task by converting binaural recordings into magnitude spectrograms, as described in Subsec. 3.3. This conversion allowed the spectrograms to be treated as visual data, enabling them to be used in an image-recognition-like task.

As illustrated in Fig. 2, the network architecture consists of an input layer accepting a pair of spectrograms, followed by a series of convolutional units and classification units, culminating in two outputs predicting ensemble location and width, respectively. This design employs a multi-task approach, enabling a single network to estimate both ensemble parameters simultaneously.

The topology finalized in this study was chosen, among many alternative architectures, based on the highest prediction quality observed on the validation dataset. Despite the existence of numerous algorithms for automatic topology selection ([BRANKE, 1995](#); [MIKKULAINEN *et al.*, 2017](#); [SHAFIEE *et al.*, 2016](#); [STANLEY, MIKKULAINEN, 2002](#); [ZHANG *et al.*, 2018](#)), the final topology was determined manually, primarily due to the high computational demands relative to the available resources.

Various architectural configurations were assessed, with key parameters being varied such as the number of convolutional units (from 1 to 5), the number of classification units (from 1 to 5), the inclusion or exclusion of max pooling layers after each convolution layer, the number of filters within the convolutional layers, the dimensions of these filters (2×2 , 2×3 , or 3×3),

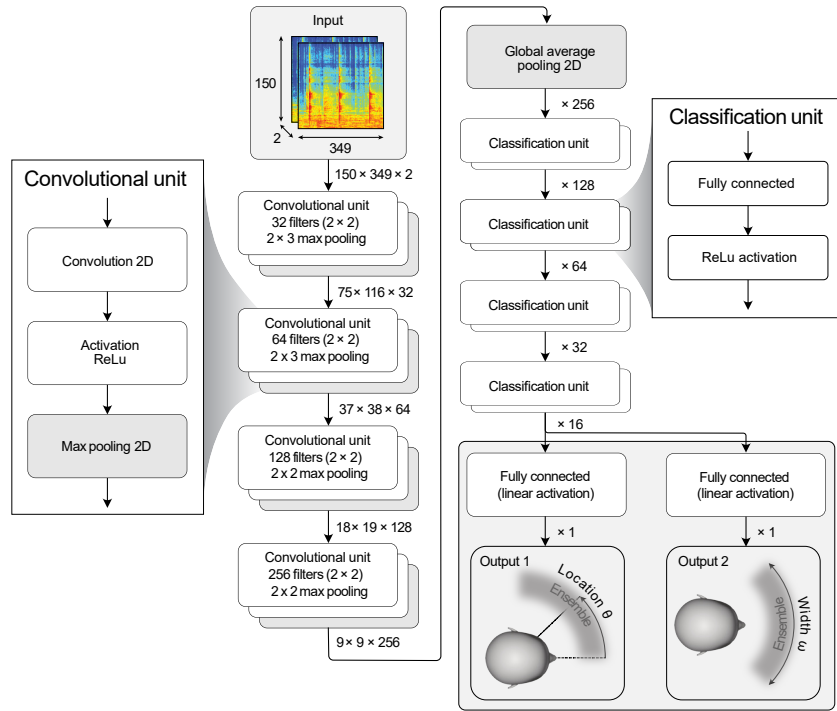


Fig. 2. Topology of the CNN used for estimating ensemble location and width, illustrating the layers (grouped in “convolutional” and “classification” units) and connections of the network architecture.

the stride size, and the dimensions of the max pooling layers (2×2 , 2×3 , or 3×3). Based on this procedure, it was concluded that the model is robust against variations in the assessed topologies. The differences in mean prediction error among the configurations were minimal, typically less than 1° for most configurations. Among the many tested topologies that yielded similar errors, the simplest one was selected to optimize both the performance efficiency and model simplicity.

Despite the availability of widely used techniques for addressing an overfitting effect, such as the dropout layer (SRIVASTAVA *et al.*, 2014), and for accelerating training, such as batch normalization (IOFFE, SZEGEDY, 2015), neither technique was employed in this study as they were observed to be ineffective for the specific estimation task being undertaken. Instead, a global average pooling layer was utilized, known for its capabilities in reducing overfitting (LIN *et al.*, 2013). This was confirmed in this particular task, as the inclusion of this layer significantly reduced overfitting, lowering the final mean absolute error (MAE) score by 0.83° (average across 10 trials) compared to configurations where a simple flattening layer was used instead.

3.5. Model training and evaluation

The topology described in the previous section resulted in a model with 216 562 learning parameters. The model training procedure was repeated 10 times, employing the Monte Carlo cross-validation method, as described by KUHN and JOHNSON (2013). For each

repetition, the entire dataset was randomly divided into two parts: a development set containing two-thirds of the dataset (15 360 recordings) for model construction, and a test set consisting of the remaining one-third of the dataset (7680 recordings) for its evaluation. This repetition procedure was employed to ensure more reliable and generalizable results by assessing the model’s performance across different subsets of the data. Additionally, it helped to account for the inherent variability in neural network training, where slight changes in initial conditions or optimization paths can lead to different model outcomes. While a large and diverse dataset could mitigate this issue, the binaural excerpts used in this study were generated from only 196 multi-track music recordings. This limited source material raised concerns by these authors about potential significant variations between the development and test sets in each repetition. In hindsight, these concerns were valid, as the maximum observed difference in MAE between repetitions reached up to 0.85° for ensemble width.

To ensure that the evaluation process was unbiased, the data split was done in such a way that no original multi-track recordings used for synthesis were included in both the development and test sets simultaneously. However, this rule was not applied to HRTFs databases, allowing for the possibility of HRTF information leaking between the development and test sets. This could be seen as a significant limitation of the study. However, it is known that a human auditory system uses a single HRTF represented by ears, head,

and torso, only slightly changing throughout the entire life, mainly during infancy (CLIFTON *et al.*, 1988; KING *et al.*, 2001). Therefore, this limitation could be considered in pair how the human auditory system behaves in real life. Nevertheless, it is worth noting that some studies implement HRTF-independent testing for binaural localization models, as demonstrated by ANTONIUK and ZIELIŃSKI (2023) and ZIELIŃSKI *et al.* (2022a; 2022b).

The development set was divided into training and validation subsets at a 7:1 ratio, with 13 440 recordings in the training subset and 1920 recordings in the validation subset. The training subset was used to update the model's learning parameters, while the validation subset was solely used for early stopping (MORGAN, BOURLARD, 1989; POCKOCK, HUGHES, 1989) and model checkpointing (EISENMAN *et al.*, 2020). These techniques were employed to select the model with the best generalization capabilities and prevent overfitting. The test subset, which included data not seen during the training or validation phases, was used solely for performance assessment once per a repetition. This divide-train-and-evaluate process was repeated to collect 10 MAEs, from which the final model error was determined.

For each sample, the model received two spectrograms as input: one for the left channel and one for the right channel. The rationale behind the application of CNNs to this task was to automatically extract local features from the spectrograms and use these features to estimate two contiguous ensemble parameters: ensemble location and width, both measured in degrees. For model training, the Adam algorithm (KINGMA, BA, 2014) was used. The algorithm minimized prediction errors, calculated as the difference between the actual ensemble parameters (known a priori from the binaural synthesis described in Subsec. 3.2) and the predicted values.

The optimizer was configured with the following hyperparameters: an initial learning rate of 10^{-3} , a decay rate of 10^{-6} , and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Training was conducted using a batch size of 8, with a maximum of 256 epochs set. An early stopping technique was implemented to prevent overfitting, terminating the process if no improvement was observed on the validation set for 20 consecutive epochs. Consequently, the maximum number of epochs was never reached; instead, training concluded after 25 to 36 epochs, with a median of 27.5 epochs. During the training process, the losses for both outputs were combined additively, ensuring equal weighting of both ensemble features.

The computational work for this study was conducted on a workstation equipped with an RTX Nvidia GeForce 4090 GPU and a 48-core AMD Ryzen ThreadRipper processor (up to 4.5 GHz). On the software side, MATLAB (The MathWorks Inc., 2022b) with

the Audio Toolbox (The MathWorks Inc., 2022a) was used for the binaural recording synthesis, while Python (VAN ROSSUM, DRAKE, 2009) with the SciPy package (VIRTANEN *et al.*, 2020) was used for feature extraction and Keras (CHOLLET *et al.*, 2015) for training the CNN model. The complete source code for all the experimental stages is publicly available on the GitHub repository (ANTONIUK, 2024). The spectrogram calculation phase required 21 minutes, data partitioning took 34 minutes, and the total training time for all iterations amounted to 40 minutes, making the entire training and evaluation process 95 minutes long.

4. Results and discussion

The overall model performance measured across 10 experiment iterations, expressed as MAE, was equal to $8.57^\circ (\pm 0.19^\circ)$ for ensemble width and $4.76^\circ (\pm 0.10^\circ)$ for ensemble location. As both ensemble parameters were constrained within the same range of 90° , the results demonstrate that the model exhibits a 44 % lower error for ensemble location compared to ensemble width. This outcome is not unexpected, given that ensemble location is a less complex parameter. Essentially, it represents the average location of all sources. Therefore, it is more resistant to temporal fluctuations in individual audio sources than ensemble width, which is dependent on the two most extreme sound sources that vary over time. Furthermore, estimating ensemble width necessitates the identification of these two extreme sources, a process that is inherently more complex than estimating a single average location.

Figure 3 compares the actual and predicted ensemble widths for each sample, showing a heteroscedastic relationship between them, with a slight bias towards predicting lower ensemble width values for higher actual widths. This relationship exhibits a strong positive correlation, with the Pearson coefficient r of 0.90. Additionally, the results indicate that the model provides more precise predictions for narrower ensemble widths, with an average MAE of 5.65° for $\omega < 30^\circ$. However, performance deteriorates as the ensemble width increases, resulting in an MAE of 12.44° for $\omega > 80^\circ$. This

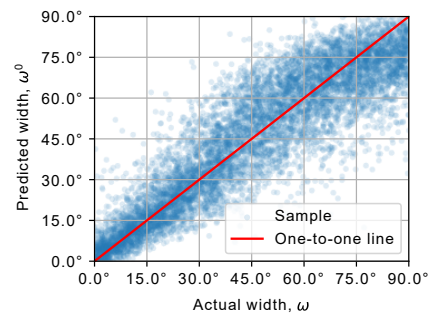


Fig. 3. Comparison between the actual ensemble width ω and the predicted ensemble width ω' for a single iteration (of the total ten).

effect is more visible in Fig. 4, which highlights the impact of the actual ensemble width on the precision of prediction. The correlation between the actual ensemble width and prediction error shows a weak positive relationship, with the Pearson coefficient r of 0.27.

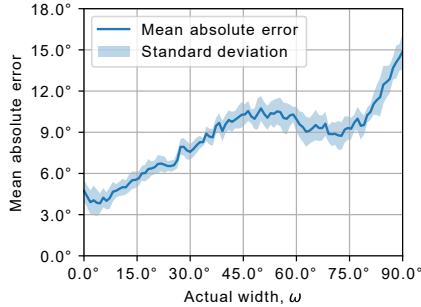


Fig. 4. Impact of the actual ensemble width ω on the mean absolute prediction error, averaged across all ten iterations, with indicated standard deviation.

The reduced accuracy in the width prediction can be attributed to the sparse distribution of audio sources in wider ensembles, which amplifies the influence of extreme sound sources on prediction errors, resulting in lower precision as the ensemble width increases. Moreover, Fig. 4 reveals that the relationship between the ensemble width and the error is nonlinear, displaying a notable decrease in error between 60° and 75° . The reason for this nonlinearity is currently unclear and requires further investigation.

The correlation between the actual and predicted ensemble location values exhibits a very high degree of correlation, as illustrated in Fig. 5. In this case, the Pearson correlation coefficient r is equal to as much as 0.97. In contrast to the ensemble width, no significant relationship is observed between actual location and its prediction error. This finding suggests that the model's ability to localize the center of the ensemble is robust, unaffected by the actual spatial positioning of the ensemble, including lateral locations. Figure 6 corroborates this observation, demonstrating the relatively consistent location error across most positions, with minor increases at extreme locations. The negligible correlation ($r = -0.03$) between the absolute lo-

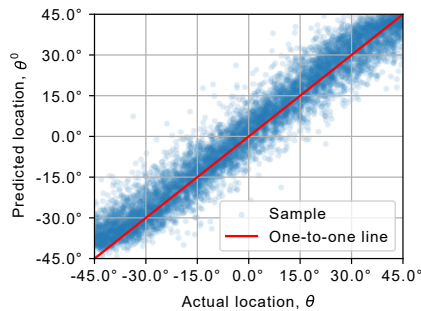


Fig. 5. Comparison between the actual ensemble location θ and the predicted ensemble location θ' for a single iteration (of the total ten).

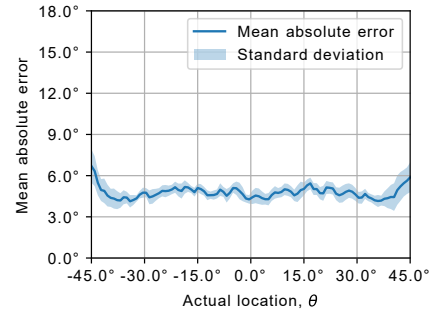


Fig. 6. Impact of the actual ensemble width ω on the mean absolute prediction error, averaged across all ten iterations, with indicated standard deviation.

cation value and prediction error further supports the model's spatial invariance in its performance.

Figure 7 illustrates the influence of both the ensemble location and width on the mean absolute error for an ensemble location, providing a detailed perspective complementing the results presented in Fig. 6. Notably, the figure highlights asymmetric anomalies, particularly within the $\theta \in [15^\circ, 30^\circ]$ range compared to the $\theta \in [-30^\circ, -15^\circ]$ range, which can be attributed to the sparsity of sample result data across specific regions of this heatmap. While the figure suggests that ensemble location does not significantly affect the model's precision in predicting location, it clearly demonstrates that ensemble width has a substantial impact. Specifically, there is a positive correlation between the width of the ensemble and the error in its location prediction, with error magnitude increasing as the width expands.

Figure 8 reveals a characteristic performance depression in $\omega \in [30^\circ, 60^\circ]$ previously shown from a different perspective in Fig. 4. This heatmap highlights another interesting phenomenon in its upper corners as the error in these areas is considerably higher. This indicates that the model's performance for estimating ensemble width is substantially worse at extreme widths and locations, i.e., when both the width and locations are near their maximum investigated values ($|\theta| \approx 45^\circ$, $\omega \approx 90^\circ$).

The model presented in this study demonstrates a significant improvement in ensemble-width performance compared to the spatiogram-based model, first introduced by ARTHI and SREENIVAS (2021) and further investigated by ANTONIUK and ZIELIŃSKI (2023), under similar evaluation conditions. While the dataset used in this study was expanded with 40 additional multi-track recordings and 10 HRTF databases, ANTONIUK and ZIELIŃSKI (2023) showed that the spatiogram model's performance does not improve with further increases in dataset size. This finding enables a direct comparison of results between the two models in terms of the precision of the ensemble width estimation, despite the differences in dataset composition. Our model achieved a MAE of 8.57°, outperforming

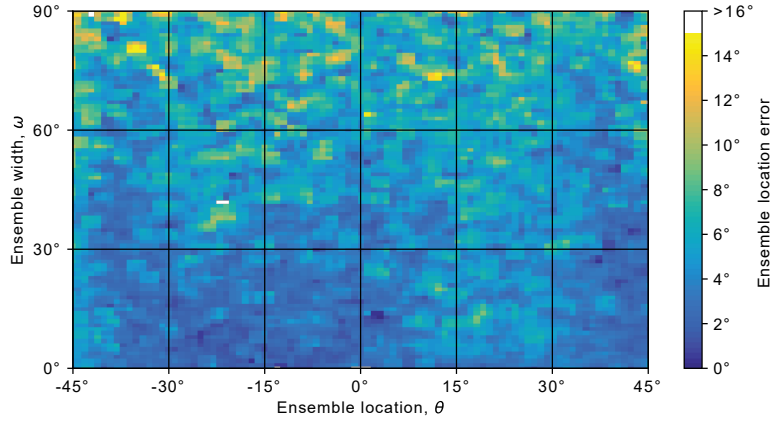


Fig. 7. Heatmap illustrating the MAE of ensemble location distribution across different ensemble locations (x -axis) and ensemble widths (y -axis). The color intensity corresponds to the MAE values, with lighter areas indicating higher errors.

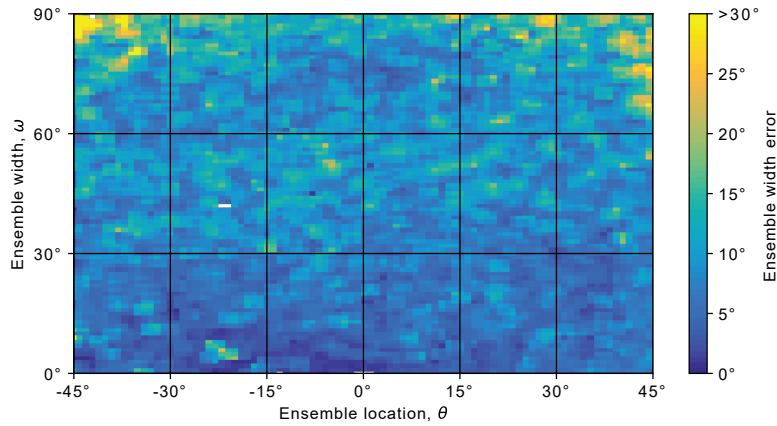


Fig. 8. Heatmap illustrating the MAE of ensemble width distribution across different ensemble locations (x -axis) and ensemble widths (y -axis). The color intensity corresponds to the MAE values, with lighter areas indicating higher errors.

the spatiogram-based model’s result of 13.62° by 5.05° . This substantial improvement is further enhanced by the current model’s ability to estimate ensemble location, a feature absent in the previous model.

Assuming terms of ensemble location prediction, the novelty of the proposed method makes direct comparison with existing literature challenging. However, its efficacy can be only evaluated indirectly against state-of-the-art individual-source localization techniques. The ensemble location prediction precision ($\text{MAE} = 4.76^\circ$) of the proposed method can be contextualized with the leading-edge binaural localization DeepEar model introduced by YANG and ZHENG (2022). Their model reported MAEs of 7.4° and 2.3° for multi-source and single-source angle of arrival (AoA) estimation, respectively. As another promising example, the WaveLoc-CONV model developed by VECCHIOTTI *et al.* (2019) demonstrated errors of 0° in anechoic conditions and 1.7° – 2.4° in multi-condition scenarios. However, these results are limited to the single-source speech localization, a substantially less complex task than the ensemble location prediction addressed by the proposed method. These experiments, while differing in objectives and datasets, provide valuable

context for the proposed method’s performance within current DoA and AoA estimation research.

5. Conclusions

This paper introduces a novel approach to locating audio sources in binaural recordings. Unlike traditional methods that predict the locations of individual audio sources, this study focuses on estimating “ensemble parameters” of audio sources, thus allowing the audio scene to be described using two parameters only: ensemble location and width. This approach makes it possible to avoid making restrictive assumptions about the number of audio sources, rendering the proposed method more suitable for real-world applications. The study also explores the use of CNN in conjunction with spectrograms applied to their inputs. According to the obtained results, the networks show exceptionally good performance, demonstrating their suitability for the investigated scenario.

The method was developed using 23 040 synthesized binaural excerpts intended to mimic real-world music recordings. The results show its outstanding performance, with the model achieving MAE

of $4.76^\circ (\pm 0.10^\circ)$ and $8.57^\circ (\pm 0.19^\circ)$ for the estimation of ensemble location and width, respectively. While the model is resilient to lateral ensemble locations, it is sensitive to the actual ensemble width, lowering the model accuracy as the width increases. The proposed method demonstrates a significant improvement over the previous technique based on spatiograms (ANTONIUK, ZIELIŃSKI, 2023), lowering the MAE by 5.05° .

Despite its high precision, the method exhibits certain limitations. Since it has been developed using the binaural excerpts synthesized with the head-related impulse responses being inherently anechoic in their characteristics, the method's performance under reverberant conditions has not been validated. Moreover, the proposed method is incapable of operating in real-time scenarios. Validating the method under reverberant conditions as well as optimizing its architecture for practical real-time scenarios constitute

the topics for future research. Other minor limitations include the lack of HRTF independence between the development and test sets, and the absence of vertical variations in audio source placement, as all sources were positioned on the horizontal plane. Additionally, the proposed approach requires substantial computational resources, particularly GPU usage, which was not necessary for the previously used spatiogram-based method.

These limitations, however, present opportunities for future research. Despite the current constraints, this study introduces a novel method for characterizing acoustic scenes in binaural recordings of music, demonstrating substantial potential for advancing binaural audio analysis. The method offers promising prospects for developing innovative tools that can objectively analyze large repositories of binaural audio recordings, focusing on spatial content.

Appendix

Table 1. List of HRTF sets used to synthesize binaural audio excerpts.

No.	Type	Head	Radius [m]	Source	Acronym
1.	Human	Human subject	1.2	RWTH Aachen University (BRAREN, FELS, 2020)	AACHEN
2.	Artificial	GRAS 45BB-4 KEMAR	1		
3.	Human	Subject 2	1.2		
4.	Human	Subject 4	1.2	Austrian Academy of Science (2014)	ARI
5.	Human	Subject 10	1.2		
6.	Artificial	ARI printed head	1.2		
7.	Human	Subject 012	1	CIPIC Interface Laboratory, University of California (ALGAZI <i>et al.</i> , 2001)	CIPIC
8.	Human	Subject 015	1		
9.	Human	Subject 020	1		
10.	Artificial	Neumann KU 100	0.9	NASA (ANDREOPOULOU <i>et al.</i> , 2015)	CLUBFRITZ
11.	Artificial	Neumann KU 100	1.5	Helsinki University of Technology (ANDREOPOULOU <i>et al.</i> , 2015)	
12.	Artificial	FABIAN	1.47	Technical University Berlin, Huawei Technologies, Munich Research Centre, Sennheiser Electronic (BRINKMANN <i>et al.</i> , 2019)	HUTUBS
13.	Human	Subject pp2	1.47		
14.	Human	Subject pp3	1.47		
15.	Human	Subject 1003	1.95	IRCAM, AKG (Listen HRTF Database, n.d.)	LISTEN
16.	Human	Subject 1002	1.95		
17.	Artificial	KEMAR DB-4004 (DB-061)	1.4	MIT (GARDNER, MARTIN, 1994)	MIT
18.	Artificial	KEMAR DB-4004 (DB-065)	1.4		
19.	Human	Subject 001	1.5	Tohoku University (WATANABE <i>et al.</i> , 2014)	RIEC
20.	Human	Subject 002	1.5		
21.	Artificial	Koken SAMRAI	1.5		
22.	Artificial	Neumann KU 100	1.2	University of York (ARMSTRONG <i>et al.</i> , 2018)	SADIE II
23.	Human	Subject H3	1.2		
24.	Human	Subject H4	1.2		
25.	Artificial	KEMAR	1	South China University of Technology (YU <i>et al.</i> , 2018)	SSCUT
26.	Artificial	Neumann KU 100	1	TH Köln (PÖRSCHMANN <i>et al.</i> , 2017)	STH Köln
27.	Artificial	FABIAN	1.7	TU Berlin (BRINKMANN <i>et al.</i> , 2017; WIERSTORF <i>et al.</i> , 2011)	TU Berlin
28.	Artificial	GRAS 45BA KEMAR	1		
29.	Artificial	GRAS 45BB-4 KEMAR – subject A attachment	1	Aalborg University; University of Iceland (SPAGNOL <i>et al.</i> , 2019; 2020)	VIKING
30.	Artificial	GRAS 45BB-4 KEMAR – subject B attachments	1		

Acknowledgments

The work was supported by the grants from the Białystok University of Technology (WI/WI-IIT/3/2022 and WZ/WI-IIT/5/2023) and funded with resources for research by the Ministry of Science and Higher Education in Poland.

References

1. ABDEL-HAMID O., MOHAMED A.-I., JIANG H., PENN G. (2012), Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition, [in:] *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4277–4280, <https://doi.org/10.1109/ICASSP.2012.6288864>.
2. ALGAZI V.R., DUDA R.O., THOMPSON D.M., AVENDANO C. (2001), The CIPIC HRTF database, [in:] *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*, pp. 99–102, <https://doi.org/10.1109/ASPAA.2001.969552>.
3. ANDREOPOULOU A., BEGAULT D.R., KATZ B.F.G. (2015), Inter-laboratory round robin HRTF measurement comparison, [in:] *IEEE Journal of Selected Topics in Signal Processing*, **9**(5): 895–906, <https://doi.org/10.1109/JSTSP.2015.2400417>.
4. ANTONIUK P. (2024), Software repository: Estimating ensemble location and width in binaural recordings of music with convolutional neural networks, GitHub, <https://github.com/pawel-antoniuk/ensemble-width-cnn> (access: 07.01.2024).
5. ANTONIUK P., ZIELIŃSKI S.K. (2023), Blind estimation of ensemble width in binaural music recordings using ‘spatiograms’ under simulated anechoic conditions, [in:] *Audio Engineering Society Conference: AES 2023 International Conference on Spatial and Immersive Audio*.
6. ARMSTRONG C., THRESH L., MURPHY D., KEARNEY G. (2018), A perceptual evaluation of individual and non-individual HRTFs: A case study of the SADIE II database, *Applied Sciences*, **8**(11): 2029, <https://doi.org/10.3390/app8112029>.
7. ARTHI S., SREENIVAS T.V. (2021), Spatiogram: A phase based directional angular measure and perceptual weighting for ensemble source width, ArXiv, <https://doi.org/10.48550/arXiv.2112.07216>.
8. Austrian Academy of Sciences (2014), HRTF-Database, <https://www.oeaw.ac.at/en/ari/das-institut/software/hrtf-database>.
9. BENAROYA E.L., OBIN N., LIUNI M., ROEBEL A., RAUMEL W., ARGENTIERI S. (2018), Binaural localization of multiple sound sources by non-negative tensor factorization, [in:] *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **26**(6): 1072–1082, <https://doi.org/10.1109/TASLP.2018.2806745>.
10. BLAUERT J. (1996), *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, <https://doi.org/10.7551/mitpress/6391.001.0001>.
11. BRANKE J. (1995), Evolutionary algorithms for neural network design and training, [in:] *Proceedings of the First Nordic Workshop on Genetic Algorithms and its Application*, pp. 145–163.
12. BRAREN H.S., FELS J. (2020), A high-resolution individual 3D adult head and torso model for HRTF simulation and validation: HRTF measurement, *RWTH Publications*, <https://doi.org/10.18154/RWTH-2020-06761>.
13. BREGMAN A. (1994), Auditory scene analysis: The perceptual organization of sound, *The Journal of the Acoustical Society of America*, **95**(2): 1177–1178, <https://doi.org/10.1121/1.408434>.
14. BRINKMANN F., DINAKARAN M., PELZER R., GROSCHE P., VOSS D., WEINZIERL S. (2019), A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses, *Journal of the Audio Engineering Society*, **67**(9): 705–718, <https://doi.org/10.17743/jaes.2019.0024>.
15. BRINKMANN F. *et al.* (2017), A high resolution and full-spherical head-related transfer function database for different head-above-torso orientations, *Journal of the Audio Engineering Society*, **65**(10): 841–848, <https://doi.org/10.17743/jaes.2017.0033>.
16. CHERRY E.C. (1953), Some experiments on the recognition of speech, with one and with two ears, *The Journal of the Acoustical Society of America*, **25**(5): 975–979, <https://doi.org/10.1121/1.1907229>.
17. CHOLLET F. *et al.* (2015), Keras, GitHub, <https://github.com/fchollet/keras> (access: 07.01.2024).
18. CHUNG M.-A., CHOU H.-C., LIN C.-W. (2022), Sound localization based on acoustic source using multiple microphone array in an indoor environment, *Electronics*, **11**(6): 890, <https://doi.org/10.3390/electronics11060890>.
19. CLIFTON R.K., GWIAZDA J., BAUER J.A., CLARKSON M.G., HELD R.M. (1988), Growth in head size during infancy: Implications for sound localization, *Developmental Psychology*, **24**(4): 477–483, <https://doi.org/10.1037/0012-1649.24.4.477>.
20. DIETZ M., EWERT S.D., HOHMANN V. (2011), Auditory model based direction estimation of concurrent speakers from binaural signals, *Speech Communication*, **53**(5): 592–605, <https://doi.org/10.1016/j.specom.2010.05.006>.
21. EISENMAN A. *et al.* (2020), Check-N-Run: A checkpointing system for training recommendation models, ArXiv.
22. ESPI M., FUJIMOTO M., KINOSHITA K., NAKATANI T. (2015), Exploiting spectro-temporal locality in deep learning based acoustic event detection, *EURASIP Journal on Audio, Speech, and Music Processing*, **2015**: 26, <https://doi.org/10.1186/s13636-015-0069-2>.

23. GARDNER B., MARTIN K. (1994), HRTF Measurements of a KEMAR dummy-head microphone, <https://sound.media.mit.edu/resources/KEMAR.html> (access: 06.19.2024).
24. GAROFOLO J.S., LAMEL L., FISHER W.M., FISCUS J.G., PALLETT D.S., DAHLGREN N.L. (1993), *DARPA TIMIT: Acoustic-Phonetic Continuous Speech Corpus CD-ROM, NIST Speech Disc 1-1.1*, NIST Publications, <https://doi.org/10.6028/NIST.IR.4930>.
25. HAHMANN M., FERNANDEZ-GRANDE E., GUNAWAN H., GERSTOFT P. (2022), Sound source localization using multiple ad hoc distributed microphone arrays, *JASA Express Letters*, **2**(7): 074801, <https://doi.org/10.1121/10.0011811>.
26. HAN Y., PARK J., LEE K. (2017), Convolutional neural networks with binaural representations and background subtraction for acoustic scene classification, [in:] *Workshop on Detection and Classification of Acoustic Scenes and Events*.
27. HIRSH I.J. (1950), Binaural hearing aids: A review of some experiments, *Journal of Speech and Hearing Disorders*, **15**(2): 114–123, <https://doi.org/10.1044/jshd.1502.114>.
28. IOFFE S., SZEGEDY C. (2015), Batch normalization: Accelerating deep network training by reducing internal covariate shift, [in:] *Proceedings of the 32nd International Conference on Machine Learning*, pp. 448–456.
29. ITU (2023), *BS.1770: Algorithms to measure audio programme loudness and true-peak audio level*, International Communications Union, Geneva, Switzerland.
30. KAVEH M., BARABELL A. (1986), The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise, [in:] *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **34**(2): 331–341, <https://doi.org/10.1109/TASSP.1986.1164815>.
31. KING A.J., KACELNIK O., MRSIC-FLOGEL T.D., SCHNUPP J.W., PARSONS C.H., MOORE D.R. (2001), How plastic is spatial hearing?, *Audiology and Neurotology*, **6**(4): 182–186, <https://doi.org/10.1159/000046829>.
32. KINGMA D.P., BA J. (2014), Adam: A method for stochastic optimization, [in:] *International Conference on Learning Representations*.
33. KRIZHEVSKY A., SUTSKEVER I., HINTON G.E. (2012), ImageNet classification with deep convolutional neural networks, [in:] *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, **25**.
34. KUHN M., JOHNSON K. (2013), *Applied Predictive Modeling*, Springer, New York, <https://doi.org/10.1007/978-1-4614-6849-3>.
35. LECUN Y. *et al.* (1989), Handwritten digit recognition with a back-propagation network, [in:] *Advances in Neural Information Processing Systems 2 (NIPS 1989)*, **2**.
36. LIN M., CHEN Q., YAN S. (2013), Network in network, [in:] *International Conference on Learning Representations*.
37. Listen HRTF Database (n.d.), <http://recherche.ircam.fr/equipes/salles/listen/> (access: 06.19.2024).
38. LIU M., HU J., ZENG Q., JIAN Z., NIE L. (2022), Sound source localization based on multi-channel cross-correlation weighted beamforming, *Micromachines*, **13**(7): 1010, <https://doi.org/10.3390/mi13071010>.
39. LIU Q., WANG W., DE CAMPOS T., JACKSON P.J.B., HILTON A. (2018), Multiple speaker tracking in spatial audio via PHD filtering and depth-audio fusion, [in:] *IEEE Transactions on Multimedia*, **20**(7): 1767–1780, <https://doi.org/10.1109/TMM.2017.2777671>.
40. MA N., BROWN G.J. (2016), Speech localisation in a multitalker mixture by humans and machines, [in:] *Interspeech 2016*, pp. 3359–3363, <https://doi.org/10.21437/Interspeech.2016-1149>.
41. MA N., GONZALEZ J.A., BROWN G.J. (2018), Robust binaural localization of a target sound source by combining spectral source models and deep neural networks, [in:] *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **26**(11): 2122–2131, <https://doi.org/10.1109/TASLP.2018.2855960>.
42. MA N., MAY T., BROWN G.J. (2017), Exploiting deep neural networks and head movements for robust binaural localisation of multiple sources in reverberant environments, [in:] *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **25**(12): 2444–2453, <https://doi.org/10.1109/TASLP.2017.2750760>.
43. MAY T., MA N., BROWN G.J. (2015), Robust localisation of multiple speakers exploiting head movements and multi-conditional training of binaural cues, [in:] *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2679–2683, <https://doi.org/10.1109/ICASSP.2015.7178457>.
44. MAY T., VAN DE PAR S., KOHLRAUSCH A. (2011), A probabilistic model for robust localization based on a binaural auditory front-end, [in:] *IEEE Transactions on Audio, Speech, and Language Processing*, **19**(1): 1–13, <https://doi.org/10.1109/TASL.2010.2042128>.
45. MAY T., VAN DE PAR S., KOHLRAUSCH A. (2012), A binaural scene analyzer for joint localization and recognition of speakers in the presence of interfering noise sources and reverberation, [in:] *IEEE Transactions on Audio, Speech, and Language Processing*, **20**(7): 2016–2030, <https://doi.org/10.1109/TASL.2012.2193391>.
46. MIKKULAINEN R. *et al.* (2017), Evolving deep neural networks, ArXiv.
47. MORGAN N., BOURLARD H. (1989), Generalization and parameter estimation in feedforward nets: Some experiments, [in:] *Advances in Neural Information Processing Systems 2 (NIPS 1989)*.
48. PAN Z., ZHANG M., WU J., WANG J., LI H. (2021), Multi-tone phase coding of interaural time difference for sound source localization with spiking neural networks, [in:] *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **29**: 2656–2670, <https://doi.org/10.1109/TASLP.2021.3100684>.

49. PANG C., LIU H., LI X. (2019), Multitask learning of time-frequency CNN for sound source localization, [in:] *IEEE Access*, **7**: 40725–40737, <https://doi.org/10.1109/ACCESS.2019.2905617>.
50. PAVLIDI D., PUIGT M., GRIFFIN A., MOUCHTARIS A. (2012), Real-time multiple sound source localization using a circular microphone array based on single-source confidence measures, [in:] *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2625–2628, <https://doi.org/10.1109/ICASSP.2012.6288455>.
51. POCOCK S.J., HUGHES M.D. (1989), Practical problems in interim analyses, with particular regard to estimation, *Controlled Clinical Trials*, **10**(4): 209–221, [https://doi.org/10.1016/0197-2456\(89\)90059-7](https://doi.org/10.1016/0197-2456(89)90059-7).
52. PÖRSCHMANN C., AREND J., NEIDHARDT A. (2017), A spherical near-field HRTF set for auralization and psychoacoustic research, [in:] *Proceedings of the 142nd AES Convention*.
53. RAAKE A. (2016), A computational framework for modelling active exploratory listening that assigns meaning to auditory scenes – Reading the world with two ears, *Two!Ears*, <http://twoears.eu> (access: 06.11.2024).
54. RUMSEY F. (2002), Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm, *Journal of the Audio Engineering Society*, **50**(9): 651–666.
55. SAINATH T.N., MOHAMED A.-r., KINGSBURY B., RAMABHADRAN B. (2013), Deep convolutional neural networks for LVCSR, [in:] *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8614–8618, <https://doi.org/10.1109/ICASSP.2013.6639347>.
56. SENIOR M. (2023), The 'Mixing Secrets' Free Multi-track Download Library, Cambridge Music Technology, <https://cambridge-mt.com/ms/mtk/> (06.10.2024).
57. SHAFIEE M.J., MISHRA A., WONG A. (2016), Deep learning with Darwin: Evolutionary synthesis of deep neural networks, *Neural Processing Letters*, **48**: 603–613, <https://doi.org/10.1007/s11063-017-9733-0>.
58. SPAGNOL S., MICCINI R., UNNTHÓRSSON R. (2020), The Viking HRTF Dataset v2.
59. SPAGNOL S., PURKHÚS K.B., UNNTHÓRSSON R., BJÖRNSSON S.K. (2019), The Viking HRTF Dataset.
60. SRIVASTAVA N., HINTON G., KRIZHEVSKY A., SUTSKEVER I., SALAKHUTDINOV R. (2014), Dropout: A simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, **15**(56): 1929–1958, <http://jmlr.org/papers/v15/srivastava14a.html>.
61. STANLEY K.O., MIIKKULAINEN R. (2002), Evolving neural networks through augmenting topologies, *Evolutionary Computation*, **10**(2): 99–127, <https://doi.org/10.1162/106365602320169811>.
62. The MathWorks Inc. (2022a), Audio Toolbox, Version: 9.13.0 (R2022b), Natick, Massachusetts, United States, <https://www.mathworks.com>.
63. The MathWorks Inc. (2022b), MATLAB, Version: 9.13.0 (R2022b), Natick, Massachusetts, United States, <https://www.mathworks.com>.
64. THIEMANN J., MÜLLER M., MARQUARDT D., DO-CLO S., VAN DE PAR S. (2016), Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene, [in:] *EURASIP Journal on Advances in Signal Processing*, **2016**(1), <https://doi.org/10.1186/s13634-016-0314-6>.
65. THOMAS S., GANAPATHY S., SAON G., SOLTAU H. (2014), Analyzing convolutional neural networks for speech activity detection in mismatched acoustic conditions, [in:] *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2519–2523, <https://doi.org/10.1109/ICASSP.2014.6854054>.
66. VAN ROSSUM G., DRAKE F.L. (2009), *Python 3 Reference Manual*, Scotts Valley, CA: CreateSpace.
67. VECCHIOTTI P., MA N., SQUARTINI S., BROWN G.J. (2019), End-to-end binaural sound localisation from the raw waveform, [in:] *ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 451–455, <https://doi.org/10.1109/ICASSP.2019.8683732>.
68. VERA-DIAZ J.M., PIZARRO D., MACIAS-GUARASA J. (2018), Towards end-to-end acoustic localization using deep learning: From audio signals to source position coordinates, *Sensors*, **18**(10): 3418, <https://doi.org/10.3390/s18103418>.
69. VIRTANEN P. *et al.* (2020), SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nature Methods*, **17**: 261–272, <https://doi.org/10.1038/s41592-019-0686-2>.
70. WANG J., WANG J., QIAN K., XIE X., KUANG J. (2020), Binaural sound localization based on deep neural network and affinity propagation clustering in mismatched HRTF condition, *EURASIP Journal on Audio, Speech, and Music Processing*, **2020**, <https://doi.org/10.1186/s13636-020-0171-y>.
71. WATANABE K., IWAYA Y., SUZUKI Y., TAKANE S., SATO S. (2014), Dataset of head-related transfer functions measured with a circular loudspeaker array, *Acoustical Science and Technology*, **35**(3): 159–165, <https://doi.org/10.1250/ast.35.159>.
72. WIERSTORF H., GEIER M., RAAKE A., SPORS S. (2011), A free database of head-related impulse response measurements in the horizontal plane with multiple distances, [in:] *130th Convention. Engineering Brief. Audio Engineering Society*.
73. WOODRUFF J., WANG D. (2012), Binaural localization of multiple sources in reverberant and noisy environment, [in:] *IEEE Transactions on Audio, Speech, and*

- Language Processing*, **20**(5): 1503–1512, <https://doi.org/10.1109/TASL.2012.2183869>.
74. YANG Q., ZHENG Y. (2022), DeepEar: Sound localization with binaural microphones, [in:] *IEEE INFOCOM 2022 – IEEE Conference on Computer Communications*, pp. 960–969, <https://doi.org/10.1109/INFOCOM48880.2022.9796850>.
 75. YU G., WU R., LIU Y., XIE B. (2018), Near-field head-related transfer-function measurement and database of human subjects, *The Journal of the Acoustical Society of America*, **143**(3): EL194–EL198, <https://doi.org/10.1121/1.5027019>.
 76. ZHANG H., KIRANYAZ S., GABBOUJ M. (2018), Finding better topologies for deep convolutional neural networks by evolution, ArXiv, <https://doi.org/10.48550/arXiv.1809.03242>.
 77. ZHANG W., SAMARASINGHE P.N., CHEN H., ABHAYAPALA T.D. (2017), Surround by sound: A review of spatial audio recording and reproduction, *Applied Sciences*, **7**(5): 532, <https://doi.org/10.3390/app7050532>.
 78. ZIELIŃSKI S.K., ANTONIUK P., LEE H. (2022a), Spatial audio scene characterization (SASC): Automatic localization of front-, back-, up-, and down-positioned music ensembles in binaural recordings, *Applied Sciences*, **12**(3): 1569, <https://doi.org/10.3390/app12031569>.
 79. ZIELIŃSKI S.K., ANTONIUK P., LEE H., JOHNSON D. (2022b), Automatic discrimination between front and back ensemble locations in HRTF-convolved binaural recordings of music, *EURASIP Journal on Audio, Speech, and Music Processing*, **2022**(1): 3, <https://doi.org/10.1186/s13636-021-00235-2>.
 80. ZIELIŃSKI S.K., LEE H., ANTONIUK P., DADAN O. (2020), A comparison of human against machine-classification of spatial audio scenes in binaural recordings of music, *Applied Sciences*, **10**(17): 5956, <https://doi.org/10.3390/app10175956>.

Research Paper

Method for Vocal Fold Paralysis Detection Based
on Perceptual and Acoustic Assessment

Rafał HALAMA, Krzysztof SZKLANNY*, Danijel KORŽINEK

Polish-Japanese Academy of Information Technology
Warsaw, Poland*Corresponding Author e-mail: kszkanny@pjawst.edu.pl

(received November 24, 2023; accepted November 30, 2024; published online December 12, 2024)

This study is aimed to evaluate a method for distinguishing between healthy and pathological voices. The evaluation was carried out using several acoustic parameters including COVAREP (collaborative voice analysis repository for speech technologies), the auditory-perceptual RBH (roughness, breathiness, hoarseness) scale, and AVQI (acoustic voice quality index). Finally, a classifier is trained using machine learning algorithms from the WEKA (Waikato Environment for Knowledge Analysis) platform.

The study group comprised 75 voice recordings of individuals affected by vocal fold paralysis. The control group consisted of 49 voice recordings of healthy individuals. The results indicate that the voice quality of the study group is significantly different than the voice quality of the control group. Acoustic parameters implemented in COVAREP and the RBH scale have proven to be reliable methods assessing voice quality. In addition, data classification achieved over 90 % accuracy for every classifier.

Keywords: voice quality; AVQI; COVAREP; RBH scale; vocal fold paralysis.



Copyright © 2024 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Voice is a key element in everyone's daily life as it is needed to communicate with other people. Three components are required for proper voice production: breathing, phonation, and articulation (MAJKOWSKA, 2004). For a human to produce a sound, simultaneous orchestration of several organs is required. The human breathing apparatus consists of lungs, diaphragm, trachea, and bronchi. It generates a driving force in the form of a stream of air exhaled from the lungs, which is needed to produce air turbulence and, therefore, sound. The phonation apparatus consists of the larynx with the vocal folds, vocal muscles, and the laryngeal nerve system. The airflow through the bronchi and trachea into the larynx causes vocal folds to vibrate, which are the sound source for voiced parts of speech. The articulation apparatus consists of the oral cavity, along with the tongue, the pharynx, and the nasal cavity. The oral cavity's role is to amplify and filter the sound produced in the larynx, thus transforming it into an articulated sound that is intelligible

as speech. When the uvula of the soft palate is properly positioned the sound wave is emitted through the nasal cavity and nostrils (TADEUSIEWICZ, 1988).

A healthy voice, also known as euphonic, is characterised by correct and clear articulation, good diction, and the smooth change of intensity and fundamental frequency depending on the content of the utterance. The air pressure of a person with such a voice is perfectly regulated. The close-ups of the vocal fold and the onset of exhalation occur at the same time. The opposite of a euphonic voice is a pathological voice. Voice pathology manifests itself in the form of aphonia and dysphonia. Dysphonia is characterised by hoarseness, abnormal timbre, loudness, and duration of the utterance (KOSZTYŁA-HOJNA *et al.*, 2014). Aphonia is defined as the inability to produce a voice. It may be caused by surgery, tumor, or psychological means (ROPER, 2014).

Vocal fold paralysis is caused by damage to the laryngeal nerves. The patient can suffer from unilateral or bilateral paralysis, the former of which is more common. We distinguish between central and periph-

eral vocal fold paralysis. Peripheral causes can be divided into traumatic and non-traumatic causes. Traumatic causes are mostly caused by surgery on the thyroid gland, either because of goiter or cancer. Other causes include communication injuries, heart, lungs, neck vessels or tracheal tumor surgeries, and intubation injuries. Non-traumatic causes include respiratory diseases such as tuberculosis, cancer, or enlarged lymph nodes. They also include viral infections such as shingles, influenza, esophageal, tracheal, and bronchial neoplasms, aortic aneurysm, myocardial hypertrophy, and mediastinal diseases. Patients with vocal fold paralysis have impaired defensive function of the larynx, which may cause choking on saliva or food. The voice of such a person is monotonous and dull. The fundamental frequency and timbre of the voice can change rapidly (CHEN *et al.*, 2007).

In the medical environment, the assessment of voice pathology is based on multiple different factors, including questionnaires for self-assessment, expert derived perceptual analysis (e.g., using the GRBAS scale (HIRANO, 1981)), acoustic analysis (e.g., jitter, shimmer, noise-to-harmonic ratio (BOERSMA, 2001)), aerodynamic analysis (e.g., maximum phonation time, mean airflow rate (SPEYER *et al.*, 2010)), and vocal range analysis (e.g., fundamental frequency and intensity range (COOPER, SORENSEN, 1981)). This assessment is repeated at several stages of administering medication or therapy, thus allowing for a correlation comparison of various methods of medical treatment – see Table 1 (JEONG *et al.*, 2022).

In recent years, especially spurred by the COVID-19 pandemic, much of the diagnosis and pre-screening have been performed in a purely remote setting (MONTALBARON *et al.*, 2023). As in this study, a common approach relies on using artificial intelligence (AI) in computer-aided diagnosis (VERIKAS *et al.*, 2006; CROWSON *et al.*, 2020). Early systems relied on simple Mel-frequency cepstrum coefficients and hidden Markov models (DIBAZAR *et al.*, 2006) – an approach common in speech recognition systems of the era. More modern solutions rely on deep learning and other novel machine learning techniques (COMPTON *et al.*, 2022; TIRRONEN *et al.*, 2023; SUVVARI, 2023). The results

outlined in the cited literature were relatively compared to those discussed in this paper. However, the comparison is difficult as the datasets of other authors are generally not available for a direct comparison.

This study aimed to prove that both acoustic and perceptual analysis are valuable tools for detecting changes in voice quality. Through a series of experiments using several classifiers, the data were successfully classified into voice recordings of people suffering from vocal fold paralysis and voice recordings of healthy individuals.

2. Speech database

The recordings were conducted in 1973–1996 in the Institute of Phoniatics at the Central Clinical Hospital, 1a Banacha St. in Warsaw. The Nagra IV S series professional tape recorder was used to record the speech in non-acoustically adapted room. (Wow and flutter (9.5 cm/s) $\pm 0.012\%$, according to DIN 45507 standard, 0.043% according to NAB standards. Signal-to-noise ratio (SNR) ASA *A*-weighted, reference 1 mW 125 dBm). The recordings contain 416 recordings of patients with various diseases affecting voice quality, such as after adenoidectomy, tubectomy, cordectomy with vocal fold paralysis, or dysphonia. Each patient underwent a phoniatic examination. A significant number of patients had their voices recorded repeatedly, which may allow us to compare the performance of our system on the same voice before and after rehabilitation.

Following speech signals were recorded: vowels / *i* / / *y* / / *a* / / *e* / / *o* / / *u* / read at equal intervals, simple announcing, and questioning sentences, and scientific text that the patient was not familiar with before the study began.

In addition, 10 sentences of text were recorded. All the recordings, which were conducted using a rarely employed cost-effective speed of 9.5 cm/s, are stored on analogue reel tapes in the Institute of Phoniatics's archives. The speed does not influence the quality of recorded speech.

The crucial process was to digitise the recordings. It was conducted at the Polish-Japanese Academy of

Table 1. Voice outcome measures as outlined by JEONG *et al.* (2022).

Category of outcome measurement	Definitions and examples
Visuo-perceptual	Subjective rating of laryngeal anatomy function, e.g., videostroboscopy, laryngoscopy, stroboscopy research tool
Auditory-perceptual	Subjective rating of the perceptual vocal quality, e.g., GRBAS (HIRANO, 1981), CAPE-V (NEMR <i>et al.</i> , 2012)
Acoustic	Computerized measurements of features of the speech sound signal, e.g., jitter, shimmer, noise-to-harmonic ratio, cepstral peak prominence
Aerodynamic	Measures of respiratory components of phonation, e.g., maximum phonation time, S/Z ratio, subglottal pressure
Voice-related quality-of-life measures	Patient rated assessment of the impact of dysphonia, e.g., vocal handicap index (WILSON <i>et al.</i> , 2004), V-RQOL (HOGIKYAN, 2004)

Information Technology using the Studer A812 reel-to-reel tape recorder (ROSLANOWSKI, 2008). The analogue signal from the recorder was sent to the computer via an E-MU 1616 audio interface. The connection was made using a symmetric cable with one end plugged into the CH1 connection of the recorder and the other end plugged into the audio interface's input. The signal was recorded in Sony Sound Forge at a sampling rate of 44.1 kHz and a 16-bit depth.

A database containing the patient's name, date of recording, disease description, ID, and file name recording, keywords, age, gender, and tape number was also created.

Examples of the transcript used are provided in Appendix A, together with its translation in Appendix B. A subset of the recordings, where patients phonated the sustained / : a/ vowel and uttered sentences in Polish: "Ten dzielny żołnierz był z nim razem. Ola lubi bezy", were included in the experiments. The voice recording was excluded if a vowel's phonation was not sustained for at least 1 second.

2.1. Pre-processing of recordings

The acoustic background and reverberation in the room used for recording exceeded appropriate levels, which affected the quality of the voice recordings. All of them had to be subjected to a noise reduction process. Firstly, the SNR was calculated for every voice recording. The SNR is a difference, measured in decibels, between the speech level and the background noise level.

Previous studies reported that recommended levels of SNR are above 42 dB, acceptable: above 30 dB, and unacceptable: below 30 dB (INGRISANO *et al.*, 1998; DELIYSKI *et al.*, 2005). To eliminate mains hum, we used a FIR high-pass filter to reduce all frequencies below 60 Hz (Fig. 1), which greatly improved the SNR levels of all recordings. Before the process, the SNR ranged from 17.9 dB to 40.9 dB, averaging 26.2 dB.

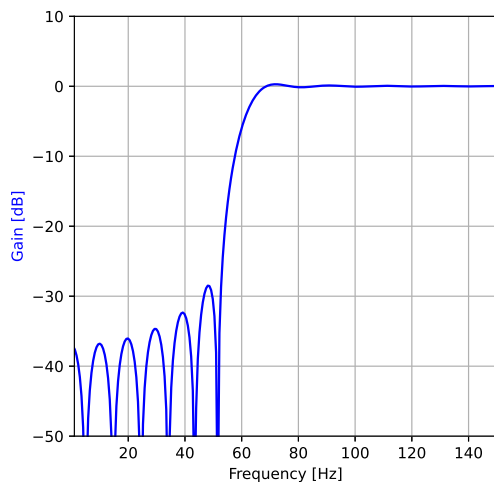


Fig. 1. FIR filter used to process the voice recordings.

After the process, the SNR ranged from 23.5 dB to 48.7 dB, averaging 36.1 dB. Only one recording was deemed unusable and was excluded from the study.

Voice recordings were sorted by the type of vocal disorder. Vocal cord paralysis, which is the goal of this study, was the only one that appeared more than a few times in the database. Only 75 recordings were used in further experiments. Forty-nine recordings which came from 17 healthy individuals were used as a control group.

2.2. Perceptual assessment of voice quality

All voice recordings used in this study were assessed by independent voice specialists using the RBH scale (NAWKA *et al.*, 1994). The scale is widely recognised as the easiest method of perceptual voice evaluation by institutions including the Committee on Phoniatrics of the European Laryngological Society (DEJONCKERE *et al.*, 2001). The RBH scale consists of three features: R – roughness; B – breathiness; H – hoarseness.

Every feature can receive a score from 0 to 3, which describes the severity of a vocal disorder: 0 – normal voice, 1 – a slight change, 2 – medium change, and 3 – high change.

The RBH scale, despite looking uncomplicated, is a reliable method of assessing voice quality, provided it is used by voice specialists such as phoniatrists or speech therapists (BEHRBOHM *et al.*, 2011).

Perceptual assessment of voice quality was carried out on both occasions by the same two independent voice specialists who had completed an RBH training program and had extensive experience in voice/speech signal assessment. On both occasions, the experts were blindfolded for the assessment duration. The two experts underwent an audiometry test, and the test results for both indicated normal hearing.

2.3. Control group recordings

Because the original dataset contained only voices with voice quality disorders, an additional set of recordings was created to capture the vocal properties of healthy individuals for control purposes. The recordings were made in the recording studio of the Polish-Japanese Academy of Information Technology. The microphone used in the recordings was a Rode NT-1A and it has the following parameters: frequency range 2 Hz–20 kHz, sensitivity 25 mV/Pa, equivalent noise level 5 dBA, maximum SPL – 137 dB SPL, polar pattern – cardioid. The signal was registered with a 48 kHz sampling rate and a 16-bit resolution (standard WAV PCM).

During the recording, the healthy individuals phonated the vowel /a : / three times with a sound pressure level of 60 dBA–80 dBA, 1 meter from the microphone, for a sustained period of at least 4 seconds.

Following that, the recorded individual was made to briefly strain his/her voice by reading out a few sentences, and then again to phonate the vowel / : a / four times.

The last four phonations of the vowel / : a / were used to calculate the acoustic parameters. All the participants phonated neutrally. Phonations with higher or lower values of the fundamental frequency of a speech signal, often denoted by F_0 , were not considered in the analyses.

A lot of consideration was taken to match the conditions of the original dataset while preparing the control samples. It is obviously impossible to recreate the conditions perfectly, but the chosen signal analysis methods were not affected by the differences in the acquisition and storage of the signal data. Given the overall low levels of background noise and good levels of SNR, both sets of recordings showed negligible levels of change in parameter values.

3. Acoustic voice evaluation

Acoustic methods for voice quality assessment are growing in popularity amongst clinicians focusing on voice research, because these methods benefit from being non-invasive and give the opportunity of utilising automation (MARYN *et al.*, 2009). They are an easy and reliable way of comparing voice dysphonia levels before surgeries and after them (MARYN *et al.*, 2009). Traditionally, sustained vowel phonation is used for testing instead of continuous speech (ASKENFELT, HAMMARBERG, 1986). In the case of vowels, features such as talking speed, pauses, the context of a sentence, accent, or type of language spoken are not relevant. On the other hand, this approach can sometimes be worse than continuous speech because sustained vowel phonation is not representative of everyday use of speech in a normal spontaneous setting (PARSA, JAMIESON, 2001). That is why the best results are obtained while using both methods.

One example of using acoustic analysis is an acoustic parameter, which evaluates the voice quality based on the parametrized sound signal. Collaborative voice analysis repository for speech technologies (COVAREP) is a free toolkit with many implementations of acoustic parameters (DEGOTTEX *et al.*, 2014) and it is available as an open source public repository online written in MATLAB. The following acoustic parameters which were used for our study were implemented in COVAREP: peak slope – PS (KANE, GOBL, 2011), normalised amplitude quotient – NAQ (ALKU *et al.*, 2002), parabolic spectral parameter – PSP (ALKU *et al.*, 1997), quasi-open quotient – QOQ (HACKI, 1989), cepstral peak prominence – CPP (HILLENBRAND, HOUE, 1996), H1H2 (HANSON, 1997), harmonic richness factor – HRF (CHILDERS, LEE, 1991), and maxima dispersion quo-

tient – MQD (KANE, GOBL, 2013). Voice recordings included only the sustained phonation of the / : a / vowel, which meant they could not be used in the experiments in which continuous speech was also needed.

3.1. Peak slope

The PS is calculated by observing the wavelet decomposition given the following formula for the mother wavelet:

$$g(t) = -\cos(2\pi f_n t) \cdot \exp\left(-\frac{t^2}{2\tau^2}\right), \quad (1)$$

where $f_n = \frac{f_s}{2}$, for f_s being the sampling frequency of 16 kHz and $\tau = \frac{1}{2f_n}$. This decomposition results in an octave band filter bank with centre frequencies at 8 kHz, 4 kHz, 2 kHz, 1 kHz, 500 Hz, and 250 Hz. From this filterbank, a local maximum is located for each band and a regression line is computed based on the amplitudes of the observed maxima (see Fig. 1 in (KANE, GOBL, 2011)).

This acoustic parameter differentiates between a modal, tense, or breathy voice. According to previous studies, the PS parameter has a certain advantage compared to other parameters (KANE, GOBL, 2011). It is completely independent, meaning that no other algorithm is used to compute its value. It is especially useful when the voice recording has an ambient noise that may disturb other algorithms and, consequently, affect the obtained values.

3.2. Normalised amplitude quotient

The NAQ is a time-based acoustic parameter used for speech signal analysis. Studies have suggested that the parameter effectively differentiates types of phonations and demonstrates resistance to the presence of noise in the speech signal (ALKU *et al.*, 2002).

It is computed for each glottal flow period using the following formula (ALKU *et al.*, 2002):

$$\frac{A_{ac}}{T_{av} \cdot d_{min}} = \frac{A_{max} - A_{min}}{T_{av} \cdot d_{min}}, \quad (2)$$

where A_{max} is the amplitude for each period of the signal, A_{min} is the lowest amplitude for each period of the signal, T_{av} is the average fundamental period length, d_{min} is the minimum derivative glottal flow, and A_{ac} is the maximal flow of amplitude.

3.3. Parabolic spectral parameter

The PSP is an acoustic parameter based on fitting a parabolic function to the low-frequency part of the calculated glottal flow spectrum. The parameter is a single numerical value that describes how the spectral decay of the resulting glottal flow behaves with respect to the theoretical limit corresponding to the

maximum decay. The PSP is commonly compared with other time-based acoustic parameters (ALKU *et al.*, 1997).

3.4. Maxima dispersion quotient

The MDQ is an acoustic parameter used to differentiate between modal, breathy, or tense voice. Previous studies show that the parameter is effective in assessing voice type based on sustained vowel phonation and continuous speech, which achieves better results than the NAQ parameter (KANE, GOBL, 2013).

For a tense voice, the maxima tend to appear around glottal closure instants (GCIs), which mark the moments of greatest excitation of vocal folds in the glottal airflow. Otherwise, if the voice is breathier, it has been observed that the maxima are scattered. The MDQ parameter recognises the scale of maxima scattering and thus effectively indicates the type of voice, and it obtains particularly good results during the analysis of continuous speech (KANE, GOBL, 2013).

3.5. Quasi-open quotient

The QOQ is an acoustic time domain parameter. It is calculated by measuring the distance between two points around and closest to the maximum of the glottal flow pulse, which are exactly 50 % of the maximum's amplitude value. This duration is also normalised with respect to the pitch period T_0 (Fig. 2).

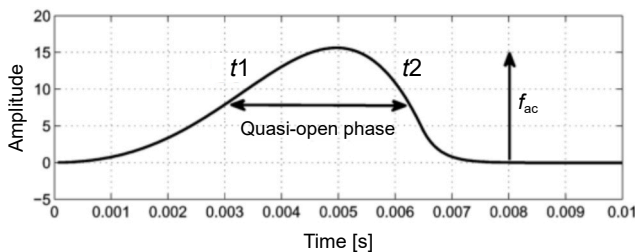


Fig. 2. Amplitude of a glottal flow impulse.

As confirmed in studies (KANE, GOBL, 2013), the QOQ parameter achieves weaker results than the MDQ and NAQ parameters. Only in the case of SNR ranging between 0 dB and 10 dB, this parameter works better.

3.6. Cepstral peak prominence

In 2018 CPP was recommended by the American Speech-Language-Hearing Association (ASHA) as a tool that allows to measure the degree of noise and other unwanted sounds in the voice signal as well as to detect the degree of dysphonia (PATEL *et al.*, 2018). CPP is defined by the distance between the top of the cepstrum and its regression line. As shown in the research of HILLENBRAND and HOUDE (1996),

the cepstral maxima are more visible in the cepstrum of a breathy voice than in the cepstrum of a modal voice which makes it possible to distinguish between these types of phonations using this parameter.

3.7. H1–H2

This acoustic parameter helps to distinguish between breathy and tense voices, which was confirmed in the studies by HANSON (1997), AIRAS and ALKU (2007). It is calculated by the difference between the amplitude of the first two vocal harmonies in the spectrum of the voice source. It is described in decibels [dB]. The H1–H2 parameter is less accurate than the MDQ and NAQ parameters (KANE, GOBL, 2013). Only when the SNR of the recording oscillates between 0 dB and 10 dB, this parameter achieves better results than its counterparts.

3.8. Harmonic richness factor

The HRF is described as the ratio of the sum of the harmonic amplitudes in the glottal flow to the component amplitude at the fundamental frequency. In previous studies (CHILDERS, LEE, 1991), the HRF parameter's scores were higher by 6.8 dB for a modal voice compared to a breathy voice, which effectively allows to distinguish between these types of phonations.

4. Acoustic voice quality index

The AVQI is a tool developed to measure overall voice quality using acoustic markers for clinical purposes. For the voice quality evaluation to be accurate and representative, the AVQI needs continuous speech and sustained vowel phonation, which lasts for a few seconds (MARYN, ROY, 2012).

The AVQI ranges between 0 and 10 and has a cut-off score between a healthy and pathological voice, which differs depending on the language, but generally, it is around 3 (Fig. 3). The more an AVQI score exceeds the cut-off threshold the higher the severity of voice dysphonia. The threshold for English and Australian equals 3.46 (REYNOLDS *et al.*, 2012), 2.70 for German (BARSTIES, MARYN, 2012), 3.07 for French (MARYN *et al.*, 2014), 2.95 for Dutch (MARYN *et al.*, 2010), 2.97 for Lithuanian (ULOZA *et al.*, 2017), 3.15 for Japanese (HOSOKAWA *et al.*, 2017), 2.02 for Korean (MARYN, WEENINK, 2015), and 3.09 for Finnish language (KANKARE *et al.*, 2020). Measurement errors must be considered while using the AVQI. The difference in results between the two recordings should be at least 0.54 (BARSTIES, MARYN, 2012) to mark that the voice quality has changed. To our knowledge, there is no data about AVQI parameters for the Polish language.

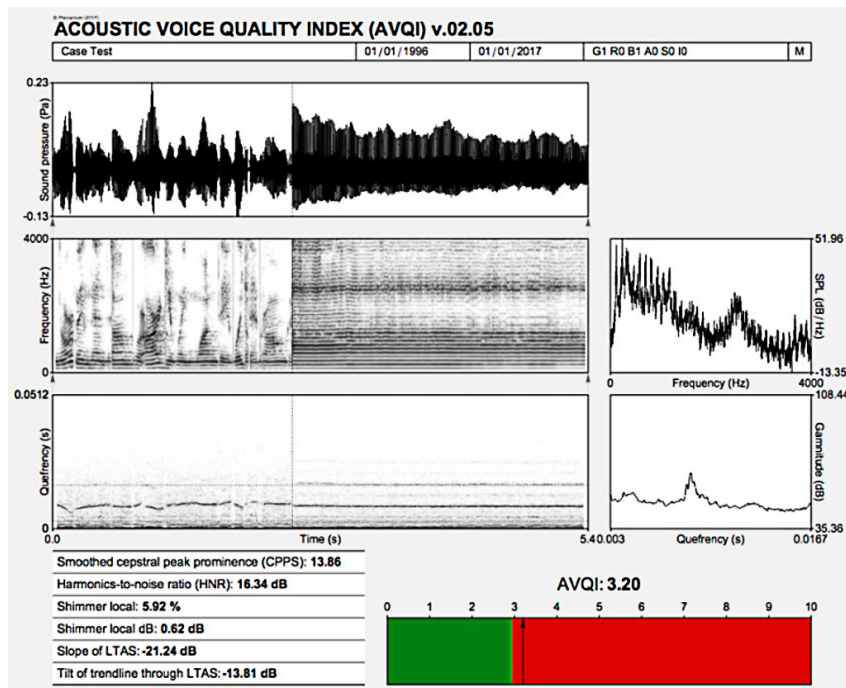


Fig. 3. Example of AVQI results.

5. Inter-rater reliability

Two independent experts who used the RBH scale to assess the voice quality of recordings in the database of non-healthy individuals were tested for inter-rater reliability because one of them had a sensitive hearing, which could heavily affect the results of experiments. Tests were conducted using MedCalc software and Real Statistics Resource Pack addon for Excel. To check the expert's agreement a single measure of the intraclass correlation coefficient (ICC) was used, which was previously used in other studies (MARYN *et al.*, 2014). The suggested limit between a good and a weak or an average agreement is 0.75 (PORTNEY, WATKINS, 2009). The obtained results for the R, B, and H parameters were as follows 0.56, 0.5, 0.46, which gave us an average of 0.51. In our experiment, we noticed a shift in the annotation of the recordings between the voice spe-

Table 2. Perceptual score distribution among experts. Scores of 0, 1, 2, and 3 are used for all parameters on the RBH scale, with reference to the different degrees of vocal disorder: 0 – a normal voice; 1 – a slight degree; 2 – a medium degree; 3 – a high degree.

Score	Expert	0	1	2	3
R	Expert 1	172	149	70	26
	Expert 2	38	172	148	59
B	Expert 1	100	195	87	35
	Expert 2	41	129	179	68
H	Expert 1	49	243	92	33
	Expert 2	1	110	149	157

cialists. The scores recorded by expert 2 proved to be more sensitive to changes in voice quality than those recorded by expert 1. The two experts underwent an audiometry test, and the test results for both indicated normal hearing. For further discussion on the inter-rater reliability of experts can be found in previous studies; Table 2 (SZKLANNY, WRZECIONO, 2019).

6. Acoustic analysis results

The AVQI score was tested to correlate with RBH scores for the same voice recordings. We used Spearman's rank-order correlation coefficient and RBH scores of experts were averaged. AVQI and the R feature had a weak correlation, while AVQI and two other features noted a higher-than-average level of correlation. Table 3 presents the results.

Table 3. Results of Spearman's Rank-Order Correlation coefficient for AVQI and RBH.

R and AVQI	B and AVQI	H and AVQI
0.371	0.655	0.594

Acoustic voice parameters obtained through COVAREP were tested for a correlation with the AVQI score for the same voice recording. With the use of Spearman's Rank-Order Correlation coefficient it was noted that the PS parameter from COVAREP had a significant correlation with the AVQI score amounting to 0.62 for a vowel, and 0.69 for a continuous speech. The parameter CPP, which is used for cal-

Table 4. Results of Spearman’s Rank-Order Correlation coefficient of AVQI with various acoustic parameters of non-healthy individuals.

Phonation type	NAQ	QOQ	H1H2	HRF	PSP	PS	MDQ	CPP
Vowel	−0.11	−0.4	0.03	−0.08	0.18	0.62	0.37	−0.84
Continuous speech	−0.35	−0.53	−0.16	−0.05	0.07	0.69	0.36	−0.77

Table 5. Results of Mann–Whitney U -test and Student t -test.

Parameter	Mean \pm SD for non-healthy individuals	Mean \pm SD for healthy individuals	Test results
CPP	12.41 \pm 0.66	11.47 \pm 0.47	$p < 0.0001$ $U = 481$
H1H2	12.97 \pm 7.69	5.36 \pm 3.82	$p < 0.0001$ $t = 6.76$
HRF	19.21 \pm 6.88	23.37 \pm 8.5	$p = 0.0014$ $U = 1214$
NAQ	0.174 \pm 0.05	0.11 \pm 0.02	$p < 0.0001$ $t = 8.647$
PSP	0.27 \pm 0.08	0.16 \pm 0.06	$p < 0.0001$ $t = 8.06$
QOQ	0.5 \pm 0.08	0.38 \pm 0.07	$p < 0.0001$ $U = 435$
MDQ	0.11 \pm 0.02	0.1 \pm 0.02	$p = 0.0001$ $t = 4.068$
PS	−0.42 \pm 0.05	−0.31 \pm 0.04	$p < 0.0001$ $U = 202$

culations in AVQI, had a significant negative correlation with the AVQI score amounting to -0.84 for a vowel, and -0.77 for a continuous speech. Similar results were observed in the study on the Finnish language, where the correlation between the CPP parameter and the AVQI score was equal to -0.35 (LAUKKANEN, RANTALA, 2022). Table 4 shows the tests results.

The Shapiro–Wilk test was used to check whether acoustic parameters had normal distribution or not. An F -test was run to check if the variance was equal. Two variants of the Student t -test were used for acoustic parameters with normal distribution: the Student t -test for an equal variance or the Student-test for unequal variance. As the distribution for other parameters was not normal, the Mann–Whitney U -test was used in their case. Table 5 shows that all acoustic voice parameters calculated for recordings of healthy individuals were statistically different from their counterparts calculated for individuals suffering from vocal fold paralysis.

The Shapiro–Wilk test was used to check whether RBH scores had a normal distribution or not. As their distribution was not normal, the Mann–Whitney U -test was used. Results showed that RBH scores for healthy individuals were statistically different from RBH scores for non-healthy individuals.

7. Classification

During the final experiment, we tried to differentiate a healthy voice from a voice affected by vocal cord paralysis using the classification based on acoustic parameters. All calculations were done in the WEKA software.

For the experiment, we used five classifiers, which were proven to be effective in previous studies on voice disorders (VERDE *et al.*, 2018): Naïve Bayes, support

vector machine (SVM), decision tree, logistic model tree, instance-based learning algorithm k -NN.

Naïve Bayes is a classifier based on Bayes’ theorem and the probability theory. The features of such a classifier are independent, so neither of them affects the other (FRIEDMAN *et al.*, 1997).

The SVM is a classifier defined by a hyperplane, that separates data belonging to different classes with the widest possible margin. This technique distinguishes between a healthy and pathological voice because it natively splits the data into up to two classes. The classification accuracy can be increased by changing the parameters and a function of the kernel (GODINO-LLORENTE *et al.*, 2005). This study used the polynomial function, which is one of the two most popular kernel functions used in the SVM (ALPAYDIN, 2004).

The decision tree is a technique used for classifying categorised data based on the training method represented by a decision tree. Decision trees are easy to interpret and can handle both continuous and categorical data. In this work, we used J48, based on the C4.5 algorithm (QUINLAN, 1999), which is the most popular tree-based classifier.

The logistic model tree is a technique that combines logistic regression, a probability-based machine learning algorithm, with a decision tree. In the WEKA software, it is implemented by the SimpleLogistic class (LANDWEHR *et al.*, 2005).

Instance-based learning algorithms are algorithms that use specific instances to obtain the results of a classifier. In this study, we used the k -NN algorithm (AHA *et al.*, 1991), which bases its results on the k -number of nearest neighbours in a new instance.

The dataset containing 75 voice recordings of non-healthy individuals and 49 voice recordings of healthy individuals with their acoustic parameters calculated

Table 6. Classification results.

Classifier	Parameters	Accuracy [%]	Sensitivity [%]	Specificity [%]	MAE
Naïve Bayes	NAQ, QOQ, H1H2, CPP, PSP, PS	95.16	98.59	90.57	0.059
SVM	NAQ, QOQ, H1H2, CPP, PSP, PS, HRF, MDQ	94.35	98.57	88.89	0.057
Decision tree	NAQ, QOQ, H1H2, CPP, PSP, PS	91.94	95.77	86.79	0.09
Logistic tree	NAQ, QOQ, H1H2, CPP, PSP, PS	94.35	97.22	90.38	0.12
<i>k</i> -NN	NAQ, QOQ, H1H2, CPP, PSP, PS, HRF	98.39	100	96.08	0.024

was prepared. Then, it was imported to the software WEKA and then underwent a classification process with the use of 10-fold cross-validation. We have calculated every classifier's accuracy, sensitivity, specificity, and mean absolute error (MAE). Accuracy describes the percentage of correctly classified data. Sensitivity describes the effectiveness of classifying positive cases. Specificity describes the effectiveness of the classification of negative cases. The MAE is a measure that determines how much on average the forecast period deviates from the real value.

Table 6 presents that the best results were received while using the *k*-NN classifier with a group of acoustic parameters (NAQ, QOQ, H1H2, CPP, PSP, PS, HRF). The decision tree (NAQ, QOQ, H1H2, CPP, PSP, PS) achieved the lowest accuracy: 91.94 %. The biggest MAE was received using the logistic model tree (NAQ, QOQ, H1H2, CPP, PSP, PS).

8. Discussion

Conducted experiments have shown that both the perceptual evaluation and the acoustic evaluation have the potential to distinguish a healthy voice from a pathological voice affected by vocal fold paralysis.

The biggest difficulty was encountered while processing the database of non-healthy individuals. This database contained voice recordings from 40–50 years ago, which were recorded on old analogue tapes. In addition, the standard of research has changed drastically over the last decades, so a significant part of the recordings could not be used for this study. Noise reduction due to unwanted background noise also turned out to be very time-consuming and the process should have been automated.

The perceptual assessment of experts who graded voice recordings of non-healthy individuals using the RBH scale was a significant problem. One expert's sensitive hearing led him to grade voice recordings differently from the other expert. Undoubtedly, this fact has influenced the results of some experiments.

An interesting finding was the negative correlation of the CPP parameter, which is one of the components needed to calculate the AVQI score. A similar correlation was found in the studies on the Finnish language (LAUKKANEN, RANTALA, 2022).

Every used classifier, whose accuracy was confirmed in the previous studies (VERDE *et al.*, 2018),

achieved over 90 % accuracy, which is a very high result for data classification. Such scores are reported in the literature to be on par with the level of human experts (SUVVARI, 2023).

A similar study was carried out in (SZKLANNY, 2019), which investigated the differences in the values of acoustic parameters between choral singers and individuals with a healthy voice. The values of acoustic parameters were compared with a group of men with a healthy voice. Significant differences were only observed for parameters H1H2 and HRF.

Other studies utilise deep learning approaches (COMPTON *et al.*, 2022) and transfer learning (TIRONEN *et al.*, 2023) providing a similarly high score at the cost of reduced interpretability of results.

9. Conclusion

The study shows that acoustic and perceptual analyses are valuable tools for detecting differences in voice quality. Using several classifiers, several experiments classified the data successfully into voice recordings of people suffering from vocal fold paralysis and voice recordings of healthy individuals.

Statistical tests have shown a medium-high correlation of the AVQI parameter with B and H features from the RBH perceptual scale. The acoustic parameter PS has shown a strong correlation with AVQI, while the CPP parameter has shown a strong, negative correlation with AVQI.

For further research, it would be advisable to expand the database with additional recordings of patients with vocal fold paralysis as well as healthy subjects, considering prolonged phonation of the vowel / : a/.

Appendix A.

Examples of recorded sentences in Polish

Ten dzielny żołnierz był z nim razem. Ola lubi bezy. Czy Ola lubi bezy? Idziemy do domu.

Czy idziemy do domu? Dzień dobry. Do widzenia. Warszawa miasto pokoju. Warszawa stolica Polski. Do widzenia Pani. Do zobaczenia Panu. Dziś jest ładna pogoda. Czy dziś jest ładna pogoda?

Przeszło sto lat minęło od pojawienia się na ulicach Warszawy pierwszego konnego tramwaju, łączącego

dworce na Pradze z Dworcem Wiedeńskim przy ulicy Marszałkowskiej. Jeszcze dwa razy Stolica przeżywała podobnie uroczyste momenty – w 1908 roku i 15 września 1945 roku. Wtedy w zniszczonej stolicy na lewym brzegu Wisły rozpoczął kursowanie pierwszy powojenny tramwaj. Odbudowa Stolicy i rozbudowa linii tramwajowych następowały równie szybko. Rejon otaczający Dworzec Centralny stanowi obecnie również wielki plac budowy, chociaż prowadzi się tu dopiero różne roboty przygotowawcze. Załogi wielu przedsiębiorstw inżynierskich przekładają urządzenia podziemne. Coraz bliżej jest termin zakończenia budowy objazdów tramwajowych w Alejach Jerozolimskich oraz w ulicach Marchlewskiego i Chałubińskiego. Na usunięcie czekają jeszcze słupy oświetleniowe, stojące na linii zastępczego torowiska. Długość objazdowych torów wynosi ponad dwa kilometry. Będą się one przecinały przy ulicy Chałubińskiego w miejscu gdzie rozebrano narożny budynek.

Appendix B.

Examples of recorded sentences translated to English

The brave soldier was with him. Ola likes meringue. Does Ola like meringue? We are going home. Are we going home? Good morning. Goodbye. Warsaw, the city of peace. Warsaw, the capital of Poland. Goodbye Mrs. Goodbye Mr. Today is nice weather. Is it nice weather today?

References

1. AHA D.W., KIBLER D., ALBERT M.K. (1991), Instance-based learning algorithms, *Machine learning*, **6**: 37–66, <https://doi.org/10.1007/bf00153759>.
2. AIRAS M., ALKU P. (2007), Comparison of multiple voice source parameters in different phonation types, [in:] *Eighth Annual Conference of the International Speech Communication Association*, <https://doi.org/10.21437/interspeech.2007-28>.
3. ALKU P., BÄCKSTRÖM T., VILKMAN E. (2002), Normalized amplitude quotient for parametrization of the glottal flow, *The Journal of the Acoustical Society of America*, **112**(2): 701–710, <https://doi.org/10.1121/1.1490365>.
4. ALKU P., STRIK H., VILKMAN E. (1997), Parabolic spectral parameter – A new method for quantification of the glottal flow, *Speech Communication*, **22**(1): 67–79, [https://doi.org/10.1016/s0167-6393\(97\)00020-4](https://doi.org/10.1016/s0167-6393(97)00020-4).
5. ALPAYDIN E. (2004), *Introduction to Machine Learning*, MIT Press.
6. ASKENFELT A.G., HAMMARBERG B. (1986), Speech waveform perturbation analysis: A perceptual-acoustical comparison of seven measures, *Journal of Speech, Language, and Hearing Research*, **29**(1): 50–64, <https://doi.org/10.1044/jshr.2901.50>.
7. BARSTIES B., MARYN Y. (2012), Der acoustic voice quality index [in German: Ein Messverfahren zur allgemeinen Stimmqualität], *HNO*, **60**(8): 715–720, <https://doi.org/10.1007/s00106-012-2499-9>.
8. BEHRBOHM H., KASCHKE O., NAWKA T., SWIFT A.C. (2011), *Ear, Nose and Throat Diseases with Head and Neck Surgery* [in Polish: *Choroby ucha, nosa i gardła z chirurgią głowy i szyi*], 2nd ed., Edra Urban & Partner.
9. BOERSMA P. (2001), Praat, a system for doing phonetics by computer, *Glott International*, **5**(9/10): 341–345.
10. CHEN H.-C., JEN Y.-M., WANG C.-H., LEE J.-C., LIN Y.-S. (2007), Etiology of vocal cord paralysis, *ORL*, **69**(3): 167–171, <https://doi.org/10.1159/000099226>.
11. CHILDERS D.G., LEE C.K. (1991), Vocal quality factors: Analysis, synthesis, and perception, *The Journal of the Acoustical Society of America*, **90**(5): 2394–2410, <https://doi.org/10.1121/1.402044>.
12. COMPTON E.C. et al. (2022), Developing an Artificial Intelligence tool to predict vocal cord pathology in primary care settings, *The Laryngoscope*, **133**(8): 1531–1595, <https://doi.org/10.1002/lary.30432>.
13. COOPER W.E., SORESENSEN J.M. (1981), *Fundamental Frequency in Sentence Production*, Springer Science & Business Media.
14. CROWSON M.G. et al. (2020), A contemporary review of machine learning in otolaryngology–head and neck surgery, *The Laryngoscope*, **130**(1): 45–51, <https://doi.org/10.1002/lary.27850>.
15. DEGOTTEX G., KANE J., DRUGMAN T., RAITIO T., SCHERER S. (2014), COVAREP – A collaborative voice analysis repository for speech technologies, [in:] *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 960–964, <https://doi.org/10.1109/icassp.2014.6853739>.
16. DEJONCKERE P.H. et al. (2001), A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques, *European Archives of Oto-rhino-laryngology*, **258**: 77–82, <https://doi.org/10.1007/s004050000299>.
17. DELIYSKI D.D., SHAW H.S., EVANS M.K. (2005), Adverse effects of environmental noise on acoustic voice quality measurements, *Journal of Voice*, **19**(1): 15–28, <https://doi.org/10.1016/j.jvoice.2004.07.003>.
18. DIBAZAR A.A., BERGER T.W., NARAYANAN S.S. (2006), Pathological voice assessment, [in:] *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, **2006**: 1669–1673, <https://doi.org/10.1109/IEMBS.2006.259835>.
19. FRIEDMAN N., GEIGER D., GOLDSZMIDT M. (1997), Bayesian network classifiers, *Machine Learning*, **29**: 131–163, <https://doi.org/10.1023/A:1007465528199>.
20. GODINO-LLORENTE J.I., GÓMEZ-VILDA P., SÁENZ-LECHÓN N., BLANCO-VELASCO M., CRUZ-ROLDÁN F., FERRER-BALLESTER M.A. (2005), Support vector machines applied to the detection of voice disorders,

- [in:] *Nonlinear Analyses and Algorithms for Speech Processing. NOLISP 2005. Lecture Notes in Computer Science*, Faundez-Zanuy M., Janer L., Esposito A., Satue-Villar A., Roure J., Espinosa-Duro V. [Eds.], pp. 219–230, <https://doi.org/10.1007/11613107-19>.
21. HACKI T. (1989), Classification of glottal dysfunctions on the basis of electroglottography [in German: Klassifizierung von glottiscysfunktionen mit hilfe der elektroglottographie], *Folia phoniatica*, **41**(1): 43–48, <https://doi.org/10.1159/000265931>.
 22. HANSON H.M. (1997), Glottal characteristics of female speakers: Acoustic correlates, *The Journal of the Acoustical Society of America*, **101**(1): 466–481, <https://doi.org/10.1121/1.417991>.
 23. HILLENBRAND J., HOUDE R.A. (1996), Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech, *Journal of Speech, Language, and Hearing Research*, **39**(2): 311–321, <https://doi.org/10.1044/jshr.3902.311>.
 24. HIRANO M. (1981), *Clinical Examination of Voice*, Springer-Verlag, New York.
 25. HOGIKYAN N.D. (2004), The voice-related quality of life (V-RQOL) measure: History and ongoing utility of a validated voice outcomes instrument, *Perspectives on Voice and Voice Disorders*, **14**(1): 3–5, <https://doi.org/10.1044/vvd14.1.3>.
 26. HOSOKAWA K. *et al.* (2017), Validation of the acoustic voice quality index in the Japanese language, *Journal of Voice*, **31**(2): 260.e1–260.e9, <https://doi.org/10.1016/j.jvoice.2016.05.010>.
 27. INGRISANO D.R., PERRY C.K., JEPSON K.R. (1998), Environmental noise: A threat to automatic voice analysis, *American Journal of Speech-Language Pathology*, **7**(1): 91–96, doi: <https://doi.org/10.1044/1058-0360.0701.91>.
 28. JEONG G.-E. *et al.* (2022), Treatment efficacy of voice therapy following injection laryngoplasty for unilateral vocal fold paralysis, *Journal of Voice*, **36**(2): 242–248, <https://doi.org/10.1016/j.jvoice.2020.05.014>.
 29. KANE J., GOBL C. (2011), Identifying regions of non-modal phonation using features of the wavelet transform, [in:] *Twelfth Annual Conference of the International Speech Communication Association*, pp. 177–180, <https://doi.org/10.21437/interspeech.2011-76>.
 30. KANE J., GOBL C. (2013), Wavelet maxima dispersion for breathy to tense voice discrimination, [in:] *IEEE Transactions on Audio, Speech, and Language Processing*, **21**(6): 1170–1179, <https://doi.org/10.1109/tasl.2013.2245653>.
 31. KANKARE E. *et al.* (2020), The acoustic voice quality index version 02.02 in the Finnish-speaking population, *Logopedics Phoniatrics Vocology*, **45**(2): 49–56, <https://doi.org/10.1080/14015439.2018.1556332>.
 32. KOSZTYŁA-HOJNA B., MOSKAL D., KURYLISZYN-MOSKAL A., RUTKOWSKI R. (2014), Visual assessment of voice disorders in patients with occupational dysphonia, *Annals of Agricultural and Environmental Medicine*, **21**(4): 898–902, <https://doi.org/10.5604/12321966.1129955>.
 33. LANDWEHR N., HALL M., FRANK E. (2005), Logistic model trees, *Machine Learning*, **59**: 161–205, <https://doi.org/10.1007/s10994-005-0466-3>.
 34. LAUKKANEN A.-M., RANTALA L. (2022), Does the acoustic voice quality index (AVQI) correlate with perceived creak and strain in normophonic young adult Finnish females?, *Folia Phoniatica et Logopaedica*, **74**(1): 62–69, <https://doi.org/10.1159/000514796>.
 35. MAJKOWSKA M. (2004), Basic issues of voice emission and hygiene [in Polish: Podstawowe zagadnienia emisji i higieny głosu], [in:] *Prace Naukowe Akademii im. Jana Długosza w Częstochowie*, **5**: 93–101.
 36. MARYN Y., CORTHALS P., VAN CAUWENBERGE P., ROY N., DE BODT M. (2010), Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels, [in:] *Journal of Voice*, **24**(5): 540–555, <https://doi.org/10.1016/j.jvoice.2008.12.014>.
 37. MARYN Y., DE BODT M., BARSTIES B., ROY N. (2014), The value of the acoustic voice quality index as a measure of dysphonia severity in subjects speaking different languages, *European Archives of Oto-Rhino-Laryngology*, **271**: 1609–1619, <https://doi.org/10.1007/s00405-013-2730-7>.
 38. MARYN Y., ROY N. (2012), Sustained vowels and continuous speech in the auditory-perceptual evaluation of dysphonia severity, *Jornal da Sociedade Brasileira de Fonoaudiologia*, **24**: 107–112, <https://doi.org/10.1590/s2179-64912012000200003>.
 39. MARYN Y., ROY N., DE BODT M., VAN CAUWENBERGE P., CORTHALS P. (2009), Acoustic measurement of overall voice quality: A meta-analysis, *The Journal of the Acoustical Society of America*, **126**(5): 2619–2634, <https://doi.org/10.1121/1.3224706>.
 40. MARYN Y., WEENINK D. (2015), Objective dysphonia measures in the program Praat: smoothed cepstral peak prominence and acoustic voice quality index, *Journal of Voice*, **29**(1): 35–43, <https://doi.org/10.1016/j.jvoice.2014.06.015>.
 41. MONTALBARON M.B. *et al.* (2023), Presumptive diagnosis in tele-health laryngology: A multi-center observational study, *The Annals of Otology, Rhinology, and Laryngology*, **132**(12): 1511–1519, <https://doi.org/10.1177/000348942311165811>.
 42. NAWKA, T., ANDERS, L., WENDLER, J. (1994), The auditory assessment of hoarse voices according to the RBH system [in German], *Sprache, Stimme, Gehör*, **18**: 130–133.
 43. NEMR K. *et al.* (2012), GRBAS and Cape-V scales: High reliability and consensus when applied at different times, *Journal of Voice*, **26**(6): 812.e17–218.e22, <https://doi.org/10.1016/j.jvoice.2012.03.005>.
 44. PARSA V., JAMIESON D.G. (2001), Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech, *Journal of Speech, Language, and Hearing Research*, **44**(2): 327–339, [https://doi.org/10.1044/1092-4388\(2001/027\)](https://doi.org/10.1044/1092-4388(2001/027)).

45. PATEL R.R. *et al.* (2018), Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function, *American Journal of Speech-Language Pathology*, **27**(3): 887–905, <https://doi.org/10.1044/2018-ajslp-17-0009>.
46. PORTNEY L.G., WATKINS M.P. (2009), *Foundations of Clinical Research: Applications to Practice*, 3rd ed., Pearson/Prentice Hall Upper Saddle River, NJ.
47. QUINLAN J.R. (1999), *C4.5: Programs for Machine Learning*, Morgan Kaufman.
48. REYNOLDS V. *et al.* (2012), Objective assessment of pediatric voice disorders with the acoustic voice quality index, *Journal of Voice*, **26**(5): 672.e1–372.e7, <https://doi.org/10.1016/j.jvoice.2012.02.002>.
49. ROPER T.A. (2014), *Clinical Skills*, 2nd ed., Oxford University Press.
50. ROSŁANOWSKI A. (2008), *Phoniatic database* [in Polish: *Baza nagrań foniatrycznych*], B.Eng., Polish-Japanese Academy of Information Technology.
51. SPEYER R. *et al.* (2010), Maximum phonation time: Variability and reliability, *Journal of Voice*, **24**(3): 281–284, <https://doi.org/10.1016/j.jvoice.2008.10.004>.
52. SUVVARI T.K. (2023), The role of Artificial Intelligence in diagnosis and management of laryngeal disorders, *Ear, Nose & Throat Journal*, <https://doi.org/10.1177/01455613231175053>.
53. SZKLANNY K. (2019), Acoustic parameters in the evaluation of voice quality of choral singers. prototype of mobile application for voice quality evaluation, *Archives of Acoustics*, **44**(3): 439–446, <https://doi.org/10.24425/aoa.2019.129257>.
54. SZKLANNY K., WRZECIONO P. (2019), Relation of RBH auditory-perceptual scale to acoustic and electroglot-tographic voice analysis in children with vocal nodules, *IEEE Access*, **7**: 41647–41658, <https://doi.org/10.1109/ACCESS.2019.2907397>.
55. TADEUSIEWICZ R. (1988), *Speech Signal* [in Polish: *Sygnał mowy*], Wydawnictwa Komunikacji i Łączności, Warszawa.
56. TIRRONEN S., JAVANMARDI F., KODALI M., REDDY KADIRI S., ALKU P. (2023), Utilizing Wav2Vec in database-independent voice disorder detection, [in:] *ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, <https://doi.org/10.1109/ICASSP49357.2023.10094798>.
57. ULOZA V., PETRAUSKAS T., PADERVINSKIS E., ULOZAITĖ N., BARSTIES B., MARYN Y. (2017), Validation of the acoustic voice quality index in the Lithuanian language, *Journal of Voice*, **31**(2): 257.e1–257.e11, <https://doi.org/10.1016/j.jvoice.2016.06.002>.
58. VERDE L., DE PIETRO G., SANNINO G. (2018), Voice disorder identification by using machine learning techniques, *IEEE access*, **6**: 16246–16255, <https://doi.org/10.1109/access.2018.2816338>.
59. VERIKAS A., GELZINIS A., BACAUSKIENE M., ULOZA V. (2006), Towards a computer-aided diagnosis system for vocal cord diseases, *Artificial Intelligence in Medicine*, **36**(1): 71–84, <https://doi.org/10.1016/j.artmed.2004.11.001>.
60. WILSON J., WEBB A., CARDING P., STEEN I., MACKENZIE K., DEARY I. (2004), The voice symptom scale (VoiSS) and the vocal handicap index (VHI): A comparison of structure and content, *Clinical Otolaryngology & Allied Sciences*, **29**(2): 169–174, <https://doi.org/10.1111/j.0307-7772.2004.00775.x>.

Research Paper

Implementation of a Cost-Effective, Accurate Photoacoustic Imaging System Based on High-Power LED Illumination and FPGA-Based Circuitry

Maryam AHANGAR DARBAND^{ID}, Esmail NAJAFIAGHDAM^{*ID}*Department of Electrical Engineering, Sahand University of Technology
Tabriz, Iran*^{*}Corresponding Author e-mail: najafiaghdam@sut.ac.ir*(received November 29, 2023; accepted October 22, 2024; published online January 8, 2025)*

Imaging based on the photoacoustic (PA) phenomenon is a type of hybrid imaging approach that combines the advantages of pure optical and pure acoustic imaging, achieving good results. This method, which offers high resolution, suitable contrast, and non-ionizing radiation, is valuable for the early detection of various types of cancer. Recently, multiple studies have focused on improving different components of this imaging system. In this presentation, we implemented a simplest form of a PA imaging system for detecting blood vessels, given that angiogenesis is recognized as a common symptom of many cancers. For the first time, we implemented a high-power light-emitting diode (LED), to replace bulky and expensive lasers, and integrated circuit technologies such as field-programmable gate arrays (FPGAs) for a simple LED driver circuit and data acquisition (DAQ). Using an FPGA block, we successfully generated a 200-ns square pulse wave with a repetition frequency of 25 kHz, whose amplified form can drive a high-power LED at 1050 nm for appropriately stimulating the sample. By using ultrasonic sensors with a central frequency of 1 MHz and a DAQ system with 16-bit accuracy, along with a suitable algorithm for image reconstruction, we successfully detected blood vessels in a breast tissue mimic. With the use of the FPGA-based block, the image reconstruction algorithm was accelerated. Finally, the simultaneous and first-time use of LED and FPGA-based circuit technology for driving the LED, output information processing and image reconstruction were performed in PA imaging.

Keywords: photoacoustic imaging (PAI); light emitting diode (LED); pulsed light; breast tumor; field-programmable gate array (FPGA).



Copyright © 2024 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A wide variety of imaging methods are involved in the diagnostic process. They include X-rays, computed tomography (CT), and magnetic resonance imaging (MRI), which can help to pinpoint diagnoses and rule out other conditions that may be causing symptoms. However, our primary focus is on non-invasive, accessible, inexpensive, and highly accurate imaging modalities. Hybrid imaging methods that combine the advantages of multiple methods seem to be promising options. One such hybrid imaging modality is photoacoustic imaging (PAI), which detects optical absorption contrast acoustically via the PA effect (XU, WANG, 2006). In PAI, nanosecond and non-ionizing pulsed lasers with relatively high energy [mJ] are di-

rected at the sample. Then, a portion of this energy is absorbed and converted into heat, leading to thermoelastic expansion and, consequently, the propagation of wideband ultrasonic waves [MHz] (WANG, 2008).

The physics of the PA phenomenon has been thoroughly studied in our previous works (AHANGAR DARBAND *et al.*, 2023a; 2023b); here, we provide a brief overview. The general equation describing the PA phenomena can be expressed as follows:

$$\left[\nabla^2 - \frac{1}{v_s^2} \frac{\partial^2}{\partial t^2} \right] p(r, t) = -\frac{p_0}{v_s^2} \frac{\partial \delta(t)}{\partial t}, \quad (1)$$

where the acoustic wave is the $p(r, t)$ at position (r) and time (t), initiated by an initial source $p_0(r) = \Gamma(r)A_e(r)$, where $A_e(r)$ is the spatial electromagnetic (EM) absorption function, v_s is the speed of sound, and

$\Gamma(r) = v_s^2 \beta / C_p$ is the Grüneisen parameter, defined by the following parameters: $\beta = (1/V)(\partial V / \partial T)_p$, where V is the volume, T is the temperature, p is the ambient density, and C_p is the specific heat. To improve the optimal feature of the PA signal, several factors must be considered: stimulation with pulsed radiation at nanosecond [ns] duration, stress confinement, and thermal confinement (AHANGAR DARBAND *et al.*, 2023a; 2023b). During the past decade, extensive efforts have been made to improve the performance of various components of PAI (PALTAUF *et al.*, 2020). PAI has a range of applications in medicine, including tissue imaging, functional imaging, and molecular imaging (PONIKWICKI *et al.*, 2019; LINDE *et al.*, 2014). However, our focus in this report is on breast tissue imaging using the PA phenomenon for breast tumor diagnosis, as breast cancer is one of the most common cancers (American Cancer Society, 2019) and if it can be detected early in time, its treatment will be easier. Therefore, we intend to use the PAI system for the early detection of breast tumors.

The PAI system consists of four main components: 1) a stimulation light generation block and a channel for conversion of emitted light into a uniform, homogenized output (HARDER *et al.*, 2004); 2) a sample, which can be either real tissue or a tissue mimic; 3) ultrasonic sensors collecting the propagated ultrasonic wave; 4) a DAQ card combined with an image reconstruction algorithm to convert the captured data into images.

Despite the progress and improvements made in various aspects of PAI over recent decades, there remains room for further improvements in several components of this method. We have previously published the simulation results for different parts of the PAI system on different platforms (AHANGAR DARBAND *et al.*, 2023a), as well as studies on our proposed image reconstruction algorithm in separate research works (AHANGAR DARBAND *et al.*, 2023b). However, this review aims to show our practical implementation of a breast tissue PAI system that is designed to be as simple and cost-effective as possible (FATIMA *et al.*, 2019). One of the most important components of the PAI system is the sample stimulation module (TAM, 1986). Some optical sources commonly used as excitation modules in PAI include lasers, such as the Q-switched Nd:YAG laser (KHOSROSHAHI, MANDELIS, 2015), diode laser, optical parametric oscillator laser system, frequency-doubled YAG laser, and Ti:sapphire laser (WANG, 2017). However, there are some disadvantages to the laser stimulation system. They include high price, bulky size, complexity in both laser device structure and driving circuit, and a requirement to use a dedicated wrapped light homogenization system (XAVIERSELVAN, MALLIDI, 2020). It has previously been reported that high-power LED can serve as an alternative excitation source for PAI (HANSEN, 2011;

ZHU *et al.*, 2020; ALLEN, BEARD, 2016) given their low-cost (tens of dollars), ease of integration, and smaller size. Therefore, to leverage these advantages of high-power LED, we used a high-power LED in our PAI stimulation module. Based on the findings in (AGRAWAL *et al.*, 2021), although LEDs provide lower output energy than laser, their high pulse repetition rate offers the possibility to average more frames and thus improve the signal-to-noise ratio (SNR). LED-based PAI holds strong potential for point-of-care PA imaging, where an imaging depth of 2 cm–3 cm is sufficient (JO *et al.*, 2020). In addition to the benefits of using LED, the circuit presented in this article for LED driving is as simple as possible due to the use of a field programmable gate array (FPGA)-based module, further reducing system complexity.

The use of real tissue in studies of the implemented system would significantly help in the detailed analysis of the captured data. But, due to possible unwanted concerns, in this report, we used a tissue-mimic sample to test the device.

The induced PA signal was detected by ultrasonic transducers and then amplified and acquired by a DAQ card to modify signals for use in the MATLAB environment, where they were processed using an image reconstruction algorithm. Actually, the LED driving circuits, digital signal processing, recording, and reconstruction were all implemented using an FPGA-based hardware system. This approach allowed us to establish a PAI system in the simplest possible form (UPPUTURI, PRAMANIK, 2017; LIU *et al.*, 2023).

The image reconstruction algorithm is the most important component of a PAI system, as it determines sensitivity, speed, and resolution (WANG, 2017). In our previous work (AHANGAR DARBAND *et al.*, 2023ab), we reviewed different image reconstruction algorithms and presented our own algorithm. Since this article mainly focuses on the practical setup of the imaging system based on the PA phenomenon, we do not go into the details of the image reconstruction algorithm. Tailored to the geometry of the imaging system, the algorithm used in this study was based on a phase-controlled algorithm (ZHOU *et al.*, 2011).

2. Principles and construction

2.1. Proposed method

We describe a simple and cost-effective PAI system that can be used for in vitro mapping of breast tissue mimics. For a detailed description of the different components of the proposed PAI system see Fig. 1.

To propagate detectable PA waves, the light irradiated to the tissue should be short-pulsed or modulated with specific energy levels. Also, a proper wavelength must be selected to detect specific markers in breast tissue and tumor (WANG, 2017). Among the commer-

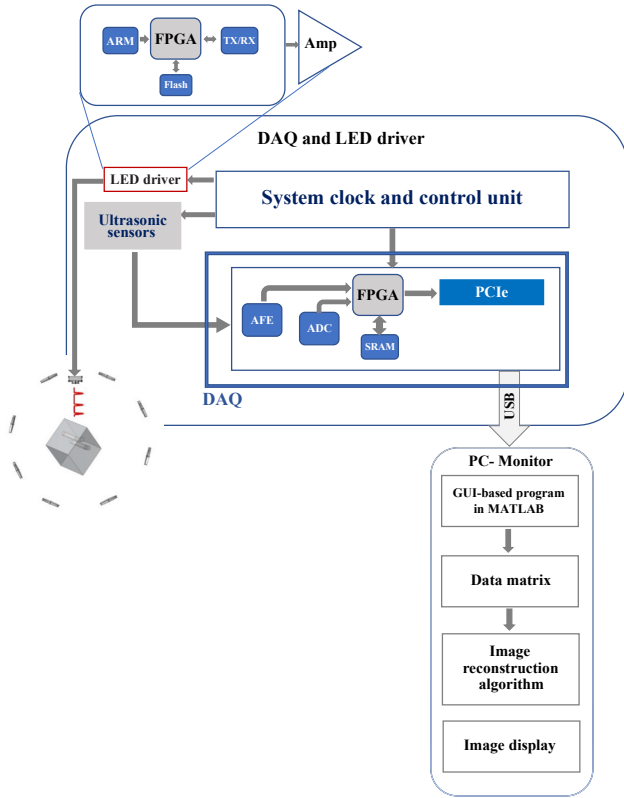


Fig. 1. Block diagram of the proposed implemented system.

cially available LEDs, we used an infrared high-power LED (LZ1-10R802, OSRAM). Easy-to-use evaluation circuit and cooling system enabled by FPGA-based technology, a low price (only a few tens of dollars), and small dimensions are the most important advantages of these LEDs. The LED driver circuit consists of an FPGA module and a wide-band two-stage amplifier. The FPGA-based LED driver circuit used in this work offers greater flexibility and speed capabilities compared to traditional MOSFET-based LED driver circuits (ZHU *et al.*, 2020; LIU *et al.*, 2023).

With the help of an FPGA-based module (Mojo Plus – FPGA Spartan 6) and appropriate programming of the FPGA, the required pulsed signal could be generated from the Mojo output pins. In our LED driving circuit, the Mojo module was programmed to generate a 200 ns square pulse (proper pulse to satisfy the stress confinement criteria (WANG, 2008)). This pulse features a rise-time of 1.9 ns, a fall-time of 3.7 ns, and a repetition rate of 25 kHz (Fig. 2). Afterward, the voltage amplitude of the pulse was amplified to the necessary level for optimal LED operation. Our two-stage wideband amplifier (designed and manufactured by the Microelectronics Laboratory of Sahand University) provides a wide dynamic range with high output voltage and high current. This amplifier was designed using a very high-speed, high-output current, high-voltage feedback amplifier integrated circuit (IC) (LM7171).

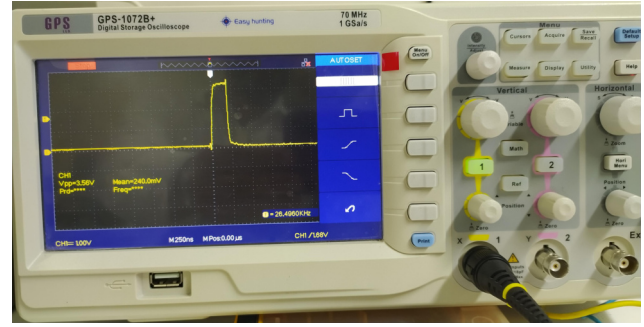


Fig. 2. LED driving pulse wave: a 200 ns square pulse with a repetition rate of 25 kHz.

It offers an adjustable gain between 22 and 150, with a maximum operating frequency of 12 MHz for the digital input. Additionally, the maximum output current is 120 mA. Finally, the driven LED radiates to an area of 1 mm × 1 mm with a power output of 1.2 W at a wavelength of 1050 nm, which is the appropriate absorption wavelength for PAI (WANG, 2017). Since the LED is placed in the water environment of the test chamber, its internal circuit, ensuring no obstruction to its radiation, is completely sealed within plastic containers. For complete tomography, the LED can rotate 270° around the sample at a distance of 5 mm from the tissue and can move along a straight line approximately 2 cm in length along the z-axis. A schematic representation in Fig. 3 illustrates the structure and geometry of this radiation block.

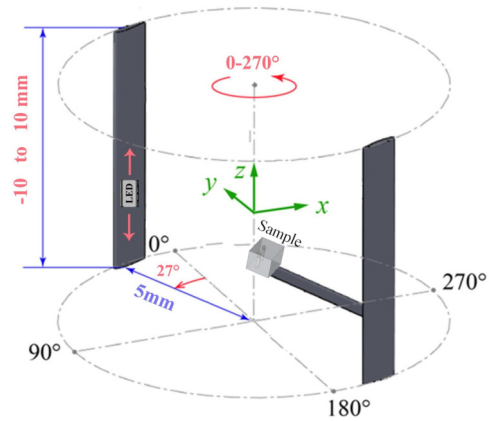


Fig. 3. Schematic structure and geometry of the radiation block.

Breast cancer is often associated with angiogenesis, and diagnosing abnormal blood vessels in the tissue seems to be the best shortcut for diagnosing breast tumors (American Cancer Society, 2019). Accordingly, a mimic breast tissue model with blood vessels was created in such a way that two low-density polyethylene (LDPE) tubes with an inner diameter of 1 mm were embedded in the middle and along the side edge of a chicken breast tissue (a square cube approximately 2 cm long) at a depth of 3 mm. These tubes were filled

with human blood that had been treated with heparin (0.1 mg/ml–0.2 mg/ml low-molecular-weight heparin (LMWH)) as an anticoagulant (Fig. 4). To determine imaging accuracy, a portion of the tube was emptied. Also, to accurately study the PAI system, we focused solely on the mimic blood vessel. The sample was fixed 5 mm in front of the LED radiation in an aqueous environment to optimize ultrasonic coupling between the sample and the ultrasonic transducer.

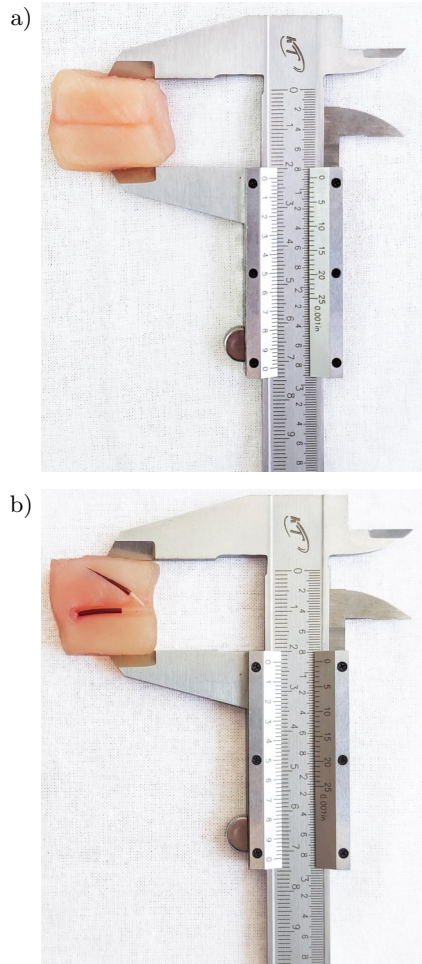


Fig. 4. Breast tissue-mimicking phantoms: a) chicken breast model with dimensions of 2 cm × 2 cm diameter; b) LDPE tubes embedded as vessels filled with blood inoculated with heparin as an anticoagulant.

The signals received by the ultrasonic sensors contain important information about the properties of the tissue being imaged. The process of receiving, processing, and transmitting the appropriate data to MATLAB to reconstruct the images was done by an analog front-end block. The PA signal propagated from the tissue were detected by eight ultrasonic sensors (PSC1.0M014083H2AD2-B0, Zhejiang Jiakang Electronics Co., Ltd.) with a central frequency of 1 MHz (1.0 MHz \pm 0.1 MHz) and a focal length of 66.74 mm (Fig. 5). These eight ultrasonic sensors were arranged at equal distances from each other to receive the PA



Fig. 5. Schematic of ultrasonic sensors with a central frequency of 1 MHz and a focal length of 66.74 mm.

signal around the circular area of the cross-section of the cylindrical chamber. The cylindrical chamber has a diameter of 8 cm and a height of 14 cm, with the sensors placed 6 cm above the bottom of the chamber (see Figs. 6–7).

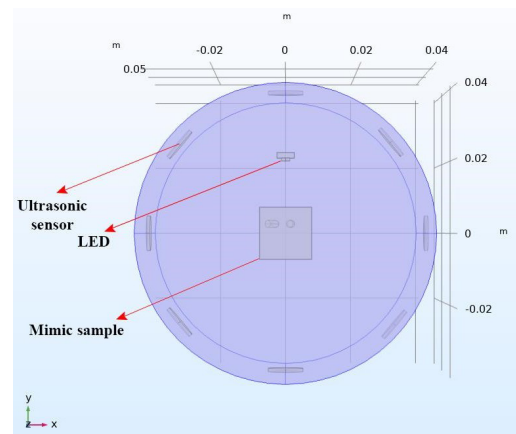


Fig. 6. Top view of the test sample chamber and the arrangement of the eighth sensors, sample, and LED simulated in COMSOL.

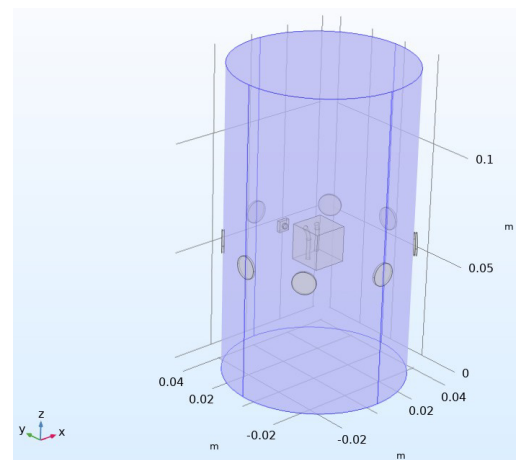


Fig. 7. Side view of the layout of the test chamber simulated in COMSOL. The mimic blood vessels inside the sample are visible in this view.

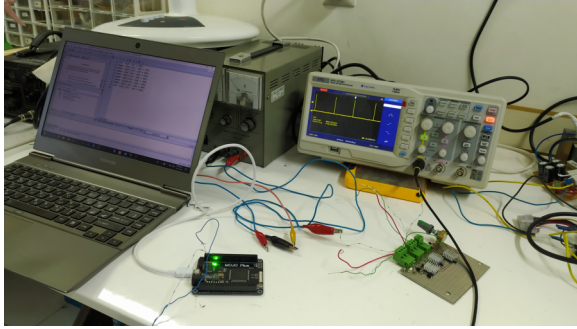


Fig. 8. View of the implemented system in the laboratory.

A DAQ (PC7KS02, Rizpardaz Electronics Company) was used to properly process the data on which Xilinx Spartan™ 7 FPGA was located. It comprises eight input channels to receive analog PA signals, which are connected to the rest of the circuit with four analog-to-digital (ADC) converters. The sampling frequency of these ADC is up to 250 million samples per second and has an accuracy of 14 bits to 16 bits. In our proposed PAI system, the sampling rate is equal to 50 million samples per second, which provides 16-bit data at the output of the ADC. The digital signal is saved inside the FPGA and then transmitted to the computer at a speed of 1 million bit/s through the network interface. In the MATLAB environment, a graphical user interface (GUI) has been developed to easily view and draw the received data. The LED driver circuit and GUI saving data in proper time are in complete synchronization.

Choosing the right algorithm based on the scanning geometry is a critical task that significantly affects image reconstruction accuracy and speed. Based on the structure of the presented here system, we used a spatial phase-controlled algorithm (ZHOU *et al.*, 2011) to reconstruct the images. We also used the FPGA-based platform to accelerate the image reconstruction speed (GAO *et al.*, 2022). To obtain enough information about the tissue structure and accurately detect blood vessels, stimulation was done from various directions. The light source was moved at different angles and along the z -axis. At first, the LED was placed 5 mm from the sample and its opposite point, and then it was rotated by ten 27° counterclockwise rotations from its initial position. This 270° rotation is repeated five times around the sample in a straight line, along the z -axis at 4 mm intervals ranging from -10 mm to 10 mm.

After processing and sampling, the data captured from the ultrasonic sensors were entered into MATLAB as 50 matrices with dimensions of 8×1019 . The feasibility of our PAI system was first validated by diagnosing vessels in tissue-mimicking phantoms. The following section presents the results of tissue tomography in the form of two-dimensional slices, which were the output of the image reconstruction algorithm.



3. Results

To demonstrate the potential of our proposed PAI system and its accuracy in detecting vessels in the mimic tissue, two scenarios were tested: 1) two vessels and 2) a single vessel embedded in the mimic tissue. In the first case, where two vessels were included in the mimic tissue, we used the sample shown in Fig. 4b. As previously described, 50 different positions were considered for LED placement. Figure 9 displays the results of image reconstruction by processing the information captured by the sensors in these modes. The total computation time for each image was 1 minute and 55.52 seconds. Figures 10 and 11 depict the output data from the sensors for the first scenario, where two mimic vessels were embedded in the sample, and the LED was positioned at point $Z = 0$, directly in front of the sample. Figure 10 shows the sensor data as an image using the full range of colormap colors to illustrate data intensity over time at each sensor position. Figure 11 presents each sensor's normalized data plotted against time for the same LED position. The procedure was then repeated for the second scenario in which only one mimic vessel was placed inside the sample (Fig. 12). The processed sensor data for the single

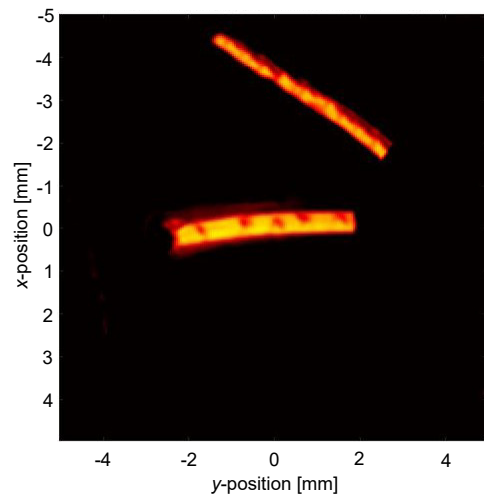


Fig. 9. Output from the image reconstruction algorithm for the case in which two mimic vessels were embedded in the sample.

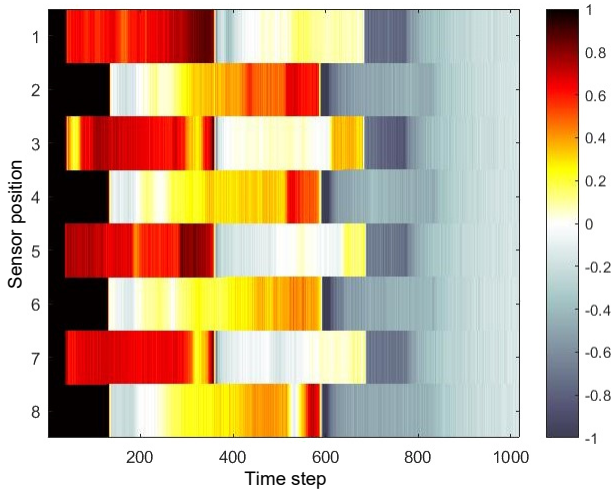


Fig. 10. Display of the sensor data as an image using the full range of colors in the colormap in the first case.

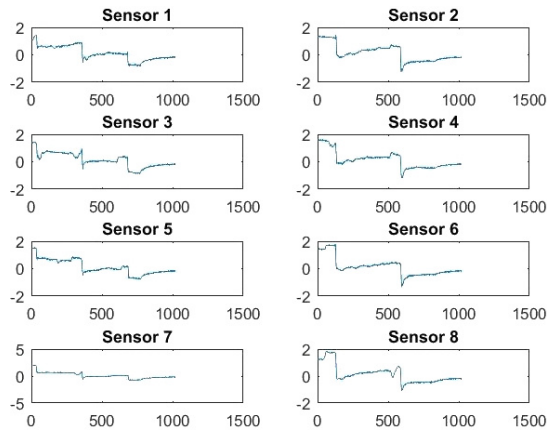


Fig. 11. Plotting of each sensor's normalized data graph against time [s] separately in the first state.

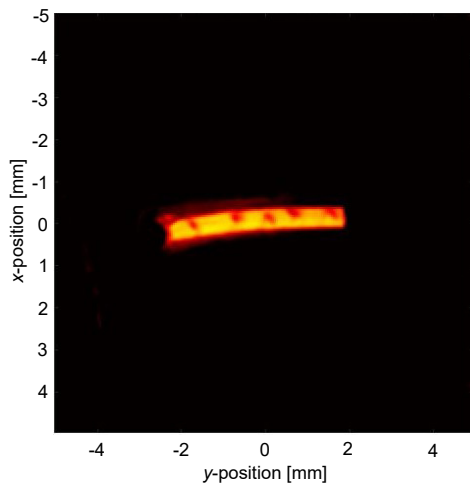


Fig. 12. Output of the image reconstruction algorithm output for the case in which one mimic vessel was embedded in the sample.

mimic vessel case is shown in Fig. 13, utilizing the full colormap range for clarity. In addition, the normalized data for each individual sensor in the second case is

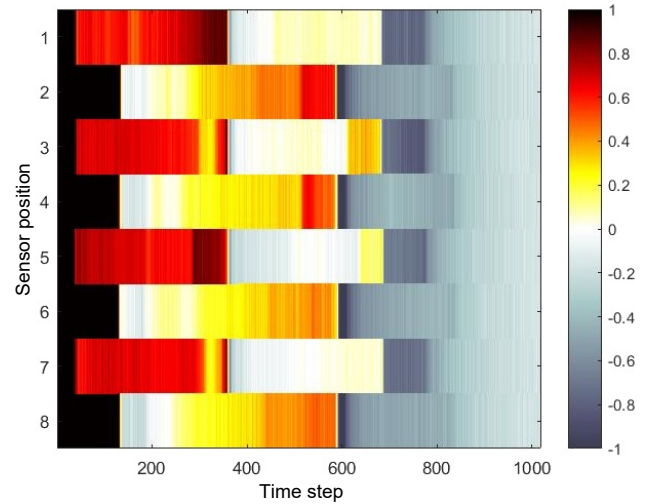


Fig. 13. Display of the sensor data as an image using the full range of colors in the colormap in the second case, testing a mimic vessel embedded in the sample.

shown in Fig. 14. This figure presents the sensor outputs over time, highlighting how the signal patterns differ with a single mimic vessel in the sample, compared to the two-vessel configuration in the first scenario (Fig. 11).

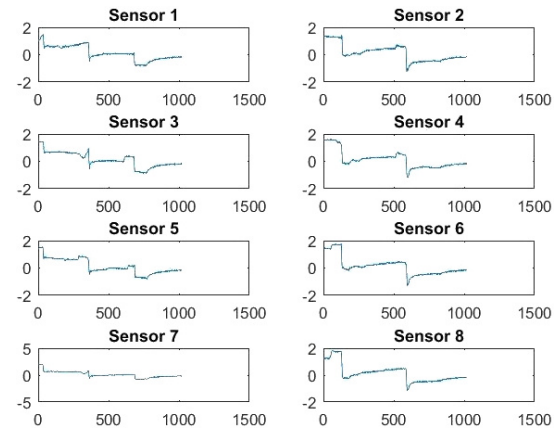


Fig. 14. Plotting of each sensor's normalized data graph against time [s] separately in the second case.

Based on the results presented, this report utilized a high-power LED to achieve good penetration depth. We generated a suitable pulse to drive the LED using FPGA technology, which successfully met the stress confinement requirements and generated efficient acoustic waves.

4. Conclusion

In summary, this article demonstrated the implementation of a miniature PAI system capable of detecting blood vessels at a depth of 3 mm using high-power LED radiation, a driving circuit, and data

acquisition (DAQ) based on FPGA technology. The simple, low-cost, and compact setup makes it possible to conveniently use it in non-invasive and label-free diagnoses. Furthermore, in future work, by using high-power LEDs in arrays (VAN HEUMEN *et al.*, 2023; JOSEPH *et al.*, 2020) or specialized light delivery system (KURIAKOSE *et al.*, 2020), we could enhance the accuracy and depth of the system's diagnostic capabilities and overcome many problems associated with using LEDs (ZHU *et al.*, 2020) in PAI systems.

Supplemental material

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

References

1. AGRAWAL S., KUNIYIL AJITH SINGH M., JOHNSTON-BAUGH K., HAN D.C., PAMEIJER C.R., KOTHAPALLI S.-R. (2021), Photoacoustic imaging of human vasculature using LED versus laser illumination: A comparison study on tissue phantoms and in vivo humans, *Sensors*, **21**(2): 424, <https://doi.org/10.3390/s21020424>.
2. AHANGAR DARBAND M., NAJAFI AGHDAM E., GHARIBI A. (2023a), Numerical simulation of breast cancer in the early diagnosis with actual dimension and characteristics using photoacoustic tomography, *Archives of Acoustics*, **48**(1): 25–38, <https://doi.org/10.24425/aoa.2023.144263>.
3. AHANGAR DARBAND M., QORBANI O., NAJAFI AGHDAM E. (2023b), Modified algebraic reconstruction technique based on circular scanning geometry to improve processing time in photoacoustic tomography, *Microwave and Optical Technology Letters*, **65**(8): 2456–2463, <https://doi.org/10.1002/mop.33714>.
4. ALLEN T.J., BEARD P.C. (2016), High power visible light emitting diodes as pulsed excitation sources for biomedical photoacoustics, *Biomedical Optics Express*, **7**(4): 1260–1270, <https://doi.org/10.1364/BOE.7.001260>.
5. American Cancer Society (2019), *Breast cancer facts & figures 2019–2020*, Atlanta: Cancer Society, Inc.
6. FATIMA A. *et al.* (2019), Review of cost reduction methods in photoacoustic computed tomography, *Photoacoustics*, **15**: 100137, <https://doi.org/10.1016/j.pacs.2019.100137>.
7. GAO Z., SHEN Y., JIANG D., LIU F., GAO F., GAO F. (2022), FPGA acceleration of image reconstruction for real-time photoacoustic tomography, ArXiv preprint, <https://doi.org/10.48550/arXiv.2204.14084>.
8. HANSEN R.S. (2011), Using high-power light emitting diodes for photoacoustic imaging, [in:] *Medical Imaging 2011: Ultrasonic Imaging, Tomography, and Therapy*, **7968**: 83–88, <https://doi.org/10.1117/12.876516>.
9. HARDER I., LANO M., LINDLEIN N., SCHWIDER J. (2004), Homogenization and beam shaping with microlens arrays, [in:] *Photon Management*, **5456**: 99–107, <https://doi.org/10.1117/12.549015>.
10. JO J., XU G., SCHIOPU E., CHAMBERLAND D., GANDIKOTA G., WANG X. (2020), Imaging of enthesitis by an LED-based photoacoustic system, *Journal of Biomedical Optics*, **25**(12): 126005, <https://doi.org/10.1117/1.JBO.25.12.126005>.
11. JOSEPH FRANCIS K., BOINK Y.E., DANTUMA M., AJITH SINGH M.K., MANOHAR S., STEENBERGEN W. (2020), Tomographic imaging with an ultrasound and LED-based photoacoustic system, *Biomedical Optics Express*, **11**(4): 2152–2165, <https://doi.org/10.1364/BOE.384548>.
12. KHOSROSHAHI M.E., MANDELIS A. (2015), Combined photoacoustic ultrasound and beam deflection signal monitoring of gold nanoparticle agglomerate concentrations in tissue phantoms using a pulsed Nd:YAG laser, *International Journal of Thermophysics*, **36**: 880–890, doi: <https://doi.org/10.1007/s10765-014-1773-3>.
13. KURIAKOSE M., NGUYEN C.D., KUNIYIL AJITH SINGH M., MALLIDI S. (2020), Optimizing irradiation geometry in LED-based photoacoustic imaging with 3D printed flexible and modular light delivery system, *Sensors*, **20**(13): 3789, <https://doi.org/10.3390/s20133789>.
14. LINDE B.B.J., SIKORSKA A., ŚLIWIŃSKI A., ŻWIRBLA W. (2014), Molecular association and relaxation phenomena in water solutions of organic liquids examined by photoacoustic and ultrasonic methods, *Archives of Acoustics*, **31**(4(S)): 143–152.
15. LIU X., KALVA S.K., LAFICI B., NOZDRIUKHIN D., DEÁN-BEN X.L., RAZANSKY D. (2023), Full-view LED-based optoacoustic tomography, *Photoacoustics*, **31**: 100521, <https://doi.org/10.1016/j.pacs.2023.100521>.
16. PALTAF G., NUSTER R., FRENZ M. (2020), Progress in biomedical photoacoustic imaging instrumentation toward clinical application, *Journal of Applied Physics*, **128**(18): 180907, <https://doi.org/10.1063/5.0028190>.
17. PONIKWICKI N. *et al.* (2019), Photoacoustic method as a tool for analysis of concentration-dependent thermal effusivity in a mixture of methyl alcohol and water, *Archives of Acoustics*, **44**(1): 153–160, <https://doi.org/10.24425/aoa.2019.126361>.
18. VAN HEUMEN S., RIKSEN J.J., SINGH M.K.A., VAN SOEST G., VASILIC D. (2023), LED-based photoacoustic imaging for preoperative visualization of lymphatic vessels in patients with secondary limb lymphedema, *Photoacoustics*, **29**: 100446, <https://doi.org/10.1016/j.pacs.2022.100446>.
19. TAM A.C. (1986), Applications of photoacoustic sensing techniques, *Reviews of Modern Physics*, **58**(2): 381, <https://doi.org/10.1103/RevModPhys.58.381>.
20. UPPUTURI, P.K., PRAMANIK M. (2017), Recent advances toward preclinical and clinical translation of photoacoustic tomography: A review, *Journal of Biomedical Optics*, **22**(4): 041006, <https://doi.org/10.1117/1.JBO.22.4.041006>.

21. WANG L.V. [Ed.] (2017), *Photoacoustic Imaging and Spectroscopy*, CRC Press.
22. WANG L.V. (2008), Prospects of photoacoustic tomography, *Medical Physics*, **35**(12): 5758–5767, <https://doi.org/10.1118/1.3013698>.
23. XAVIERSELVAN M., MALLIDI S. (2020), LED-based functional photoacoustics – Portable and affordable solution for preclinical cancer imaging [in:] *LED-Based Photoacoustic Imaging: From Bench to Bed-side*, pp. 303–319, https://doi.org/10.1007/978-981-15-3984-8_12.
24. XU M., WANG L.V. (2006), Photoacoustic imaging in biomedicine, *Review of Scientific Instruments*, **77**(4): 041101, <https://doi.org/10.1063/1.2195024>.
25. ZHU Y. *et al.* (2020), Towards clinical translation of LED-based photoacoustic imaging: A review, *Sensors*, **20**(9): 2484, <https://doi.org/10.3390/s20092484>.
26. ZHOU Q., JI X., XING D. (2011), Full-field 3D photoacoustic imaging based on plane transducer array and spatial phase-controlled algorithm, *Medical Physics*, **38**(3): 1561–1566, <https://doi.org/10.1118/1.3555036>.

Research Paper

Inference of Bubble Size Distribution in Sediments Based on Sounding by Chirp Signals

Xiaohong YANG^{(1),(2)}, Guangying ZHENG^{(1),(2),(3)*}, Fangyong WANG^{(1),(2),(3)},
Fangwei ZHU^{(1),(2)}, Linlang BAI^{(1),(2),(3)}

⁽¹⁾ *Science and Technology on Sonar Laboratory*
Hangzhou, China

⁽²⁾ *Hangzhou Applied Acoustics Research Institute*
Hangzhou, China

⁽³⁾ *Hanjiang National Laboratory*
Wuhan, China

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

(received February 15, 2024; accepted November 18, 2024; published online February 12, 2025)

A method is proposed to estimate the bubble void fraction and bubble size distribution in marine sediments based on measured sound speed and attenuation data in gas-bearing sediments. The new inversion approach employs an effective density fluid model, corrected for gas bubble pulsations, as the forward model and represents the unknown gas bubble size distribution using a finite sum of cubic B -splines. An in situ acoustic monitoring experiment was conducted at an intertidal site in the Yellow Sea to investigate gassy sediments and validate the method. The measured sound speed and attenuation show significant fluctuations due to bubble resonance, with resonance peaks shifting to higher frequencies as water depth and hydrostatic pressure increase. This method simultaneously estimates the bubble size distribution from sound speed and attenuation data.

Keywords: gassy sediment; bubble size distribution; sound speed; attenuation; cubic B -splines.



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

1. Introduction

Characterizing the amount of gas within marine sediments is crucial, as gas bubbles can significantly affect acoustic reflection and penetration (RICHARDSON *et al.*, 1998; ANDERSON *et al.*, 1998; CHEN *et al.*, 2023). FLEISCHER *et al.* (2001) reported the global distribution of gas-bearing sediments and noted that these sediments are predominantly found in the Northern Hemisphere, particularly in shallow areas near Europe and the United States. In China, there is also a noticeable presence of shallow gas near the seabed in the South China Sea, East China Sea, and Yellow Sea. Generally, sediments containing gas bubbles exhibit pronounced sensitivity to acoustic waves, characterized by high dispersion and attenuation (LEIGHTON, 2007; YARINA *et al.*, 2023; ZHANG *et al.*, 2023). These

acoustic properties of marine sediments are valuable for assessing the ecological status of the seabed. Consequently, the impact of bubbles on acoustic propagation is typically used to estimate the gas content and bubble size distribution within marine sediments (KARPOV *et al.*, 1996; LEIGHTON, ROBB, 2008). However, determining the bubble size distribution is generally more difficult than estimating the gas content of marine sediments.

In addressing the inverse problem of bubble size distribution (WILKENS, RICHARDSON 1998; BEST *et al.*, 2004; TÓTH *et al.*, 2015; EDRINGTON, CALLOWAY, 1984; SHANKAR *et al.*, 2005; 2006; FONSECA *et al.*, 2002), the forward model is typically based on Anderson and Hampton's (A&H) model (ANDERSON, HAMPTON, 1980a; 1980b), which remains the most widely used geoaoustic model for gas-bearing sediments.

The A&H model is designed for linear bubble pulsation; however, its expression for complex sound speed includes both positive and negative signs, leading to ambiguities in the inverse problem concerning the bubble characterization (MANTOUKA *et al.*, 2016).

The literature also highlights other effective inversion methods for estimating bubble size distributions in water and sediments. For example, COM-MANDER and McDONALD (1991) and DOGAN *et al.* (2015) utilized a linear B -spline to approximate an unknown bubble size distribution, transforming the integral equation into a system of linear equations involving the coefficients of linear B -splines. Nonetheless, in this inversion method, attenuation is derived from the scattering and extinction cross-section a forward model that may not be suitable for accurately modeling gas-bearing sediments.

Accordingly, we propose an inversion method in this paper to estimate bubble size distributions. Our model integrates an acoustic model tailored for gas-bearing sediments with B -spline expansions. Additionally, both sound speed and attenuation are considered simultaneously in the inverse problem. This new inversion method utilizes an effective density fluid model (ZHENG *et al.*, 2017), adapted to account for gas bubble pulsations as the forward model, and employs a finite sum of cubic B -splines to represent the unknown gas bubble size distribution. The inverse problem is reformulated as solving systems of equations that involve the coefficients of the cubic B -splines. This method has been validated using sound speed and attenuation data obtained from an in situ experiment conducted in the Yellow Sea.

2. Methodology

A corrected equivalent fluid density model, developed by ZHENG and HUANG (2016) and ZHENG *et al.* (2017), is utilized to predict sound speed and attenuation in gassy sediments. The model is expressed as

$$\nabla [K_{\text{eff}} \nabla \cdot \mathbf{u}_{\text{eff}}] = -\omega^2 \tilde{\rho}_{\text{eff}} \mathbf{u}_{\text{eff}}, \quad (1)$$

where \mathbf{u}_{eff} denotes the effective displacement. The effective modulus K_{eff} is expressed as follows:

$$K_{\text{eff}} = \left(\frac{(1-\beta)}{K_g} + \frac{\beta}{K_w} \right)^{-1}, \quad (2)$$

where K_g denotes the grain bulk modulus, K_w denotes the water bulk modulus, and β denotes the porosity. The corrected effective density is expressed as

$$\tilde{\rho}_{\text{eff}} = \rho_{\text{eff}} + \int_0^\infty \frac{4\pi\beta K_{\text{eff}} \rho_{\text{eff}} a f(a)}{\{\rho_w [\omega_0^2(a) - \omega^2 + 2ib(a)\omega]\}} da, \quad (3)$$

$$\rho_{\text{eff}} = \frac{(\rho \tilde{\rho} - \rho_w^2)}{(\tilde{\rho} + \rho - 2\rho_w)}, \quad (4)$$

$$\tilde{\rho} = \frac{\alpha \rho_w}{\beta} - \frac{iF\eta}{\kappa\omega}, \quad (5)$$

$$\rho = \beta \rho_w + (1-\beta) \rho_g, \quad (6)$$

where ω denotes the angular frequency, η denotes the water viscosity, ρ_w denotes the pore fluid density, ρ denotes the sediment density, the permeability satisfies $\kappa = \frac{(d^2 \beta^3)}{[180(1-\beta)^2]}$. The complex correction factor F is given by:

$$F(\varsigma) = \frac{\varsigma T(\varsigma)}{4 \left(1 - \frac{2T(\varsigma)}{\varsigma} \right)},$$

$$T(\varsigma) = \frac{(ber'(\varsigma) + ibei'(\varsigma))}{(ber(\varsigma) + ibei(\varsigma))}, \quad (7)$$

$$\varsigma = a \left(\frac{\omega \rho_w}{\eta} \right)^{1/2},$$

where the pore size satisfies $a = \sqrt{\frac{8\alpha\kappa}{\beta}}$. The second item in Eq. (3) is the correction term for bubble pulsation, where $f(a) da$ is the number of bubbles per unit volume with radii between a and $a + da$. The parameter a is the bubble radius, and we use R to generally denote the bubble radius, ω_0 denotes the bubble resonance frequency:

$$\omega_0^2 = \left[\text{Re} \varphi - \frac{2\sigma}{(R P_{\text{in},0})} \right] \frac{P_{\text{in},0}}{\rho_w R^2}, \quad (8)$$

b denotes the damping term:

$$b = \frac{2\eta}{(\rho_w R^2)} + \frac{\omega^2 R}{(2c)} + \frac{\text{Im}(P_{\text{in},0}\varphi)}{(2\omega \rho_w R^2)}, \quad (9)$$

where φ is the polytropic exponent of the gas, expressed as

$$\varphi = \frac{3\gamma_g}{\left\{ 1 - 3(\gamma_g - 1) i\chi \left[\left(\frac{i}{\chi} \right)^{1/2} \coth \left(\frac{i}{\chi} \right)^{1/2} - 1 \right] \right\}}, \quad (10)$$

where c denotes the fluid phase velocity, σ denotes the surface tension, $P_{\text{in},0} = P_\infty + \frac{2\sigma}{R}$, P_∞ denotes the equilibrium pressure, $\chi = \frac{D}{\omega R^2}$, γ_g denotes the ratio of specific heat, and D denotes the thermal diffusivity of gas.

The complex velocity of gassy sediment is denoted as $c_1 = \sqrt{\frac{K_{\text{eff}}}{\tilde{\rho}_{\text{eff}}}}$, and the phase velocity is

$$c_p = \frac{1}{\text{Re} \left(\frac{1}{c_1} \right)}. \quad (11)$$

The attenuation in decibels per meter is

$$\alpha^{(m)} = -\frac{20\omega \text{Im} \left(\frac{1}{c_1} \right)}{\ln 10}. \quad (12)$$

The absolute value of the effective density changes significantly due to bubble pulsation, leading to a decrease in the velocity of the porous medium. The imaginary part of the effective density accounts for an additional dissipation mechanism related to bubble pulsations. However, when the coefficient ratio defined in Eq. (13) is used to divide Eq. (1), Eq. (14) can be derived. Notably, the ratio is complex, with a modulus greater than 1 as long as the gas content is nonzero. As demonstrated in Eq. (14), bubble pulsation also modifies the sediment's effective modulus. In contrast, the sediment's effective density remains constant, resulting in a significant decrease in the sound speed of the medium:

$$\begin{aligned} \text{ratio} &= \frac{\tilde{\rho}_{\text{eff}}}{\rho_{\text{eff}}} \\ &= 1 + \int_0^\infty \frac{4\pi\beta K_{\text{eff}} a f(a)}{\{\rho_w [\omega_0^2(a) - \omega^2 + 2ib(a)\omega]\}} da, \end{aligned} \quad (13)$$

$$\nabla \left[\frac{K_{\text{eff}}}{\text{ratio}} \nabla \cdot \mathbf{u}_{\text{eff}} \right] = -\omega^2 \rho_{\text{eff}} \mathbf{u}_{\text{eff}}. \quad (14)$$

Notably, in comparison to existing acoustic theories of gas-bearing sediments, the proposed model offers two significant advantages:

- 1) it incorporates the dispersion mechanism resulting from the relative motion between the pore water and the solid frame;
- 2) it provides independent expressions for sound speed and attenuation, in contrast to the A&H model, which enhances the applicability of the proposed model to inverse problems.

The reciprocal of the complex velocity can be derived from Eqs. (11) and (12):

$$\frac{1}{c_1} = \frac{1}{c_p} - i \frac{\alpha^{(m)} \ln 10}{20\omega}. \quad (15)$$

Combining Eq. (15) with $c_1 = \sqrt{\frac{K_{\text{eff}}}{\tilde{\rho}_{\text{eff}}}}$, we obtain the following equation:

$$\frac{\tilde{\rho}_{\text{eff}}}{K_{\text{eff}}} = \left(\frac{1}{c_p} - i \frac{\alpha^{(m)} \ln 10}{20\omega} \right)^2. \quad (16)$$

$\frac{\tilde{\rho}_{\text{eff}}}{K_{\text{eff}}}$ can be derived from Eq.(3):

$$\begin{aligned} \frac{\tilde{\rho}_{\text{eff}}}{K_{\text{eff}}} &= \frac{\rho_{\text{eff}} + \int_0^\infty \frac{4\pi\beta K_{\text{eff}} \rho_{\text{eff}} a f(a)}{\{\rho_w [\omega_0^2(a) - \omega^2 + 2ib(a)\omega]\}} da}{K_{\text{eff}}} \\ &= \frac{\rho_{\text{eff}}}{K_{\text{eff}}} + \int_0^\infty \frac{4\pi\beta \rho_{\text{eff}} a f(a)}{\{\rho_w [\omega_0^2(a) - \omega^2 + 2ib(a)\omega]\}} da. \end{aligned} \quad (17)$$

A function $E(f)$ that varies with frequency is introduced to satisfy the relation in Eq. 17):

$$\begin{aligned} &\int_0^\infty \frac{4\pi\beta \rho_{\text{eff}} a f(a)}{\{\rho_w [\omega_0^2(a) - \omega^2 + 2ib(a)\omega]\}} da \\ &= \left(\frac{1}{c_p} - i \frac{\alpha^{(m)} \ln 10}{20\omega} \right)^2 - \frac{\rho_{\text{eff}}}{K_{\text{eff}}} = E(f). \end{aligned} \quad (18)$$

The inverse of the bubble size distribution $f(a)$ is used to solve the first kind of Fredholm integral equation:

$$\int_0^\infty \frac{4\pi\beta \rho_{\text{eff}} a f(a)}{\{\rho_w [\omega_0^2(a) - \omega^2 + 2ib(a)\omega]\}} da = E(f). \quad (19)$$

To solve the integral Eq. (19), we use a finite sum of cubic B -splines to denote $f(a)$:

$$f(a) = \sum_{j=0}^{n+2} C_j \Omega_3 \left(\frac{a - a_{j-1}}{h} \right), \quad a_0 \leq a \leq a_1, \quad (20)$$

where $a_j = a_0 + jh$ ($j = 0, 1, \dots, n$), $h = \frac{a_1 - a_0}{n}$, and C_j is the coefficient to be determined.

Substituting Eq. (20) into Eq. (19) yields a linear set of equations:

$$E(f_i) = \sum_{j=0}^{n+2} C_j K_{ij}, \quad (21)$$

where the elements of the matrix are

$$K_{ij} = \int_0^\infty \frac{\Omega_3 \left(\frac{a - a_{j-1}}{h} \right) 4\pi\beta \rho_{\text{eff}} a}{\{\rho_w [\omega_0^2(a) - \omega^2 + 2ib(a)\omega]\}} da. \quad (22)$$

In matrix notation, Eq. (21) may be written as follows:

$$\begin{pmatrix} K_{11} & K_{12} & K_{13} & \cdots & K_{1N} \\ K_{21} & K_{22} & K_{23} & \cdots & K_{2N} \\ K_{31} & K_{32} & K_{33} & \cdots & K_{3N} \\ \vdots & \vdots & \vdots & & \vdots \\ K_{N1} & K_{N2} & K_{N3} & \cdots & K_{NN} \end{pmatrix} \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_N \end{pmatrix} = \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ \vdots \\ E_N \end{pmatrix}. \quad (23)$$

The proposed inversion method has two primary advantages:

- 1) it transforms the nonlinear inverse problem by solving linear equation sets, thereby reducing computational demands;
- 2) it overcomes the ambiguities inherent in inversions based on the A&H model.

Consequently, this method is particularly well-suited for inverse problems that combine both sound speed and attenuation data, optimizing computational efficiency while accounting for both parameters.

To aid comprehension of the proposed method, we summarize it in the following steps:

- 1) select the control points for the cubic B -spline;
- 2) utilize Eq. (22) to calculate the matrix kernel K_{ij} based on frequency, measured physical parameters, and the selected control points;

- 3) apply Eq. (18) to calculate the function $E(f)$ based on the measured sound speed and attenuation coefficient;
- 4) calculate the coefficients of the cubic B -spline using the pseudo-inverse of matrix \mathbf{K} , and subsequently determine the bubble size distribution using Eq. (20).

Next, we present the simulation analysis of the inversion process for bubble size distribution utilizing the method developed in this section. A finite sum of cubic B -splines is employed to represent the unknown bubble size distribution, with the control points and their corresponding coefficients detailed in Table 1. Figure 1 illustrates the resulting bubble size distribution, while the physical parameters of the marine sediments are provided in Table 2. The sound speed and attenuation coefficient as functions of frequency are shown in Fig. 2. The results indicate that as the insonifying frequencies approach the resonance frequency of the bubble, the acoustic properties of gassy sediment exhibit significant dispersion, and the attenuation reaches its peak.

Table 1. Control points and their coefficients of cubic B -spline.

Control point [mm]	Coefficients ($\times 10^5$)
0	0
1	-6.8
2	33
3	-5.8
4	4.9
5	1.2
6	8.3
7	1.6
8	3.3
9	6.3
10	1.4
11	0

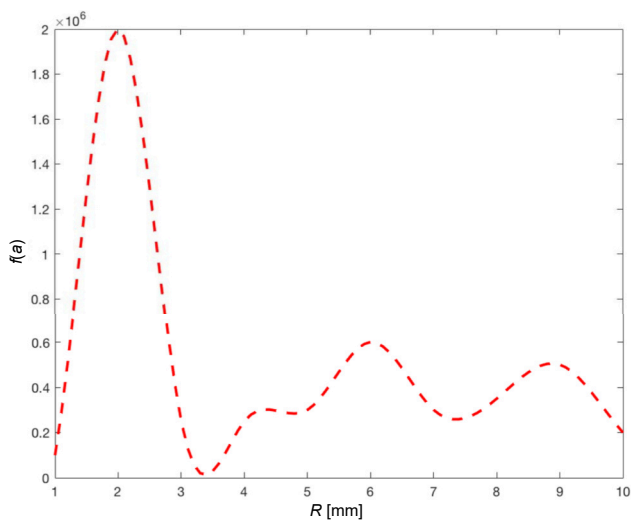


Fig. 1. Bubble size distribution.

Table 2. Input parameters.

	Parameters	Values
Sediment parameters	Grain density	2465 kg/m ³
	Grain diameter	0.781 mm
	Fluid bulk modulus	2.193 Pa $\times 10^9$ Pa
	Grain bulk modulus	3.6 Pa $\times 10^{10}$ Pa
	Fluid viscosity	1.002 Pa \cdot s $\times 10^{-3}$ Pa \cdot s
	Porosity	0.37
	Fluid density	998.2 kg/m ³
	Structure factor	1.25
Gas parameters	Gas density	1.1691 kg/m ³
	Gas velocity	340 m/s
	Equilibrium pressure	1.01 Pa $\times 10^5$ Pa
	Thermal diffusivity	2.4 m ² /s $\times 10^{-5}$ m ² /s
	Surface tension	72.75 N/m $\times 10^3$ N/m
	Ratio of specific heat	1.4

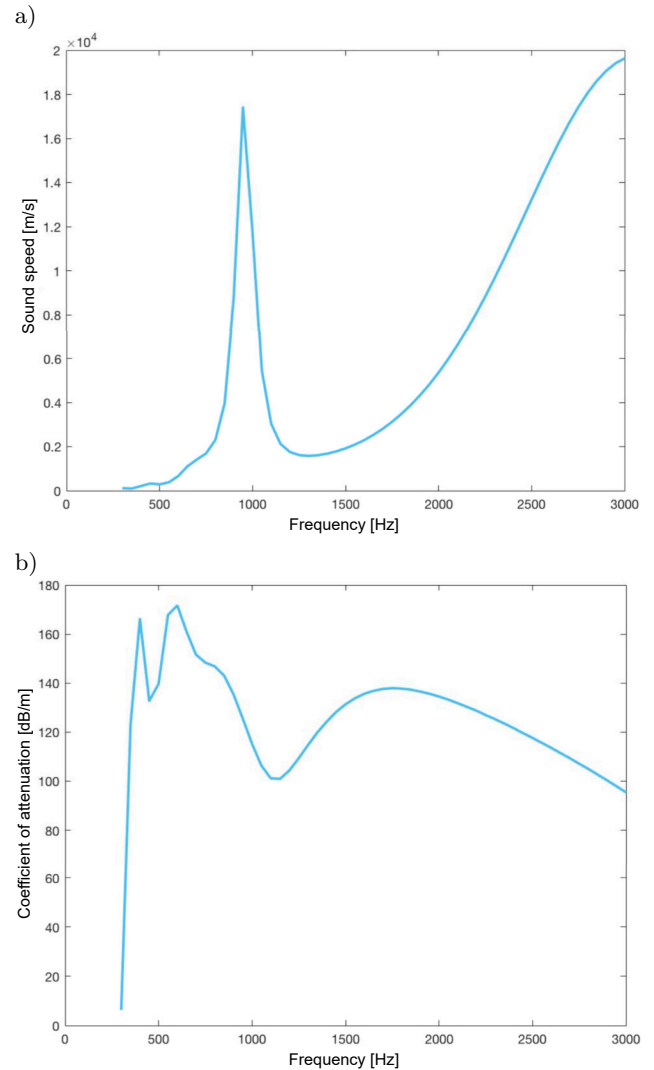


Fig. 2. Variation in: a) sound speed; b) coefficient of attenuation with frequency.

In the simulation of bubble size distribution, this study evaluates the effects of random errors in sound

speed and attenuation data to analyze the robustness of the proposed inversion method in the presence of data errors. When controlling for attenuation data, the inversion results for bubble size distribution across varying sound speed error ranges are depicted in Fig. 3. Conversely, when controlling for sound speed data, the inversion results for bubble size distribution across different attenuation error ranges are presented in Fig. 4. Additionally, Fig. 5 illustrates the inversion results for bubble size distribution considering data errors in both sound speed and attenuation. In these figures, the solid line represents the inversion results derived from multiple random errors, whereas the dashed line indicates the true value of the selected bubble size distribution. From Figs. 3 and 4, we observe that satisfactory inversion results can be achieved with a sound speed error range of 1×10^{-3} , while a larger local error in bubble size distribution occurs with an error range of 2×10^{-3} . Similarly, good inversion results can be at-

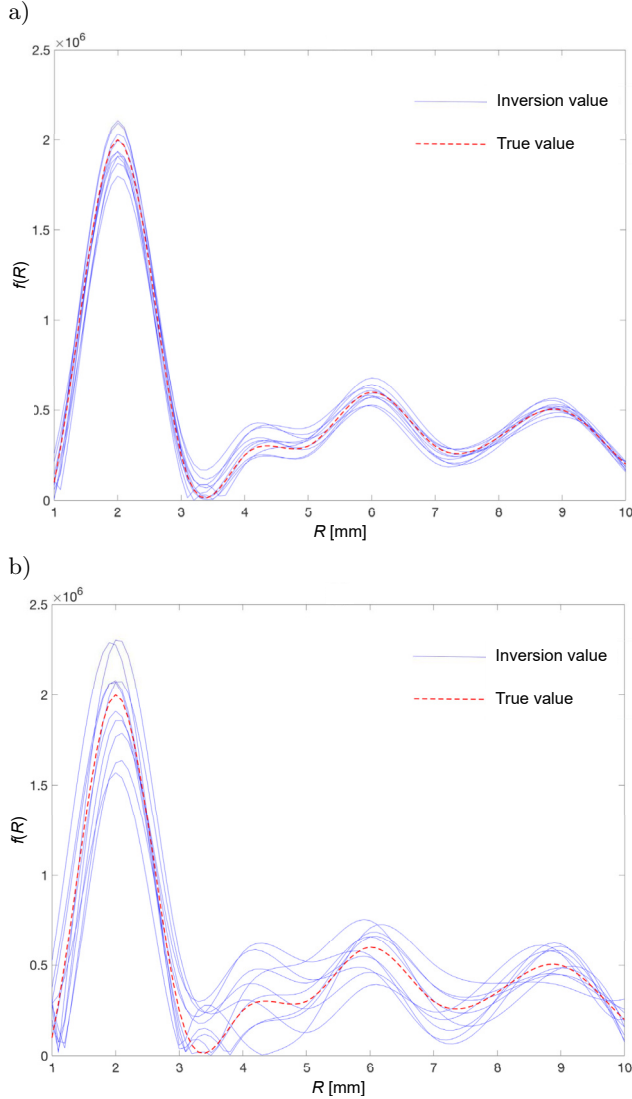


Fig. 3. Inversion results of bubble size distribution for sound speed error range of: a) 1×10^{-3} ; b) 2×10^{-3} .

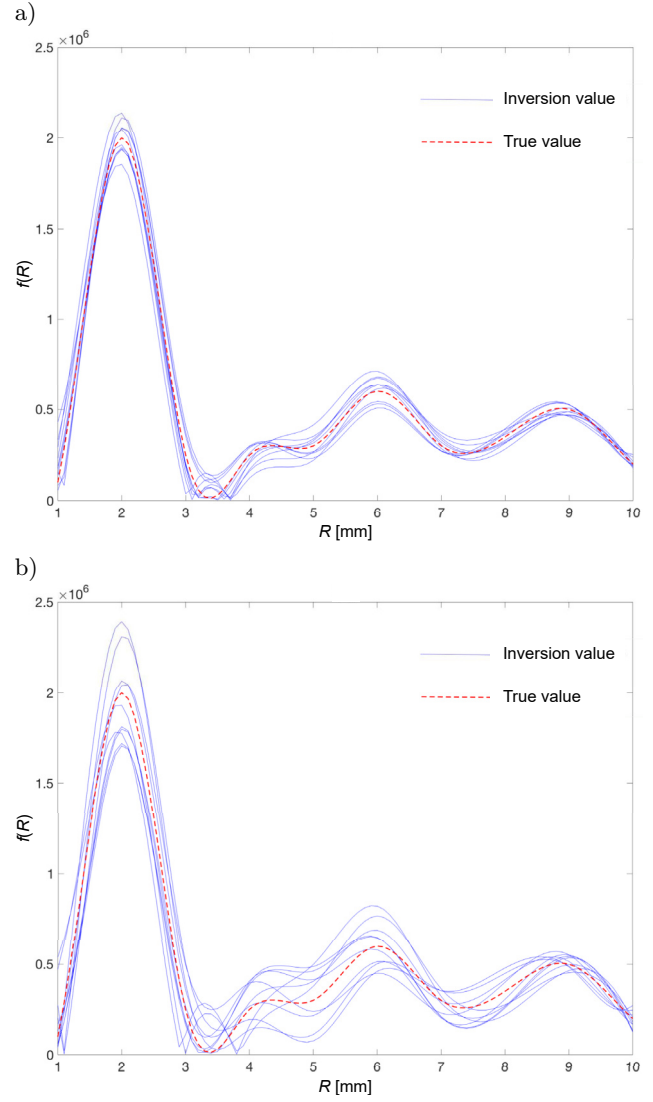


Fig. 4. Inversion results of bubble size distribution for attenuation coefficients error range of: a) 5×10^{-4} ; b) 1×10^{-3} .

tained with an attenuation coefficient error range of 5×10^{-4} , whereas a larger local error in bubble size distribution arises with an error range of 1×10^{-3} .

This analysis indicates that sound speed data exhibits a stronger resistance to interference compared to the attenuation data; therefore, the inversion of bubble size distribution is more sensitive to variations in attenuation data. Furthermore, when comparing the results in Fig. 5, it is evident that inaccuracies in attenuation data significantly influence the overall accuracy of the inversion of bubble size distribution when random errors are present in both sound speed and attenuation data. This is further corroborated by the objective function $E(f)$ in Eq. (18), where sound speed c_p appears in the denominator and the attenuation coefficient in the numerator. Consequently, the effect of the attenuation coefficient $\alpha^{(m)}$ on the objective function is more pronounced under the same perturbation range of sound speed.

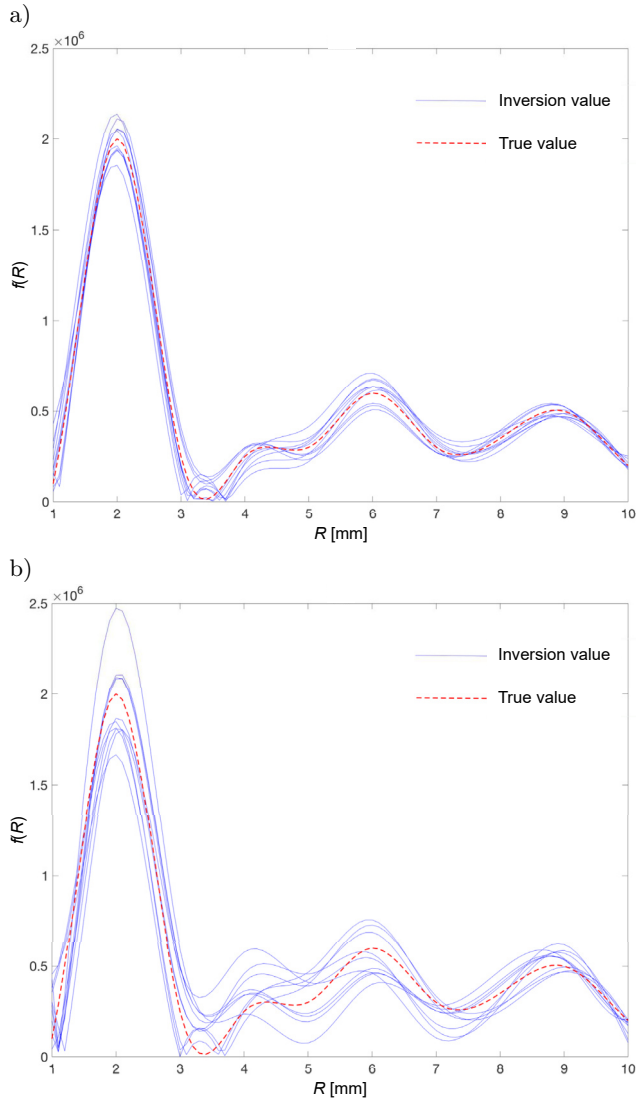


Fig. 5. Inversion results of bubble size distribution for data error range of: a) 5×10^{-4} ; b) 1×10^{-3} .

3. Experiments and verification

In this section, we present an analysis of the experimental results obtained by inverting the bubble size

distribution using the method developed in this article. The experimental site is located on the beach of a small island in the Yellow Sea, southwest of Beijingzi Town, Donggang City, Liaoning Province ($N39^{\circ}47'17.36''$, $E123^{\circ}49'0.52''$), as illustrated in Fig. 6. We selected an intertidal silt zone with a water depth of 3 m–4 m at high tide and a beach that emerges at low tide for the in situ acoustic experiment, due to substantial evidence indicating the presence of shallow gas.

The layout of the in-situ measurement experimental setup is depicted in Fig. 7. Two B&K8103 hydrophones (designated as H1 and H2) are positioned within the sediments at the same horizontal alignment, with a depth difference of 10 cm. These hydrophones are utilized to monitor the acoustic velocity and attenuation of gas-bearing sediments. To preserve the original structure of the sediment, we excavated 50 cm downward next to the designated burial location of the hydrophones and subsequently inserted the devices laterally into their predetermined positions.

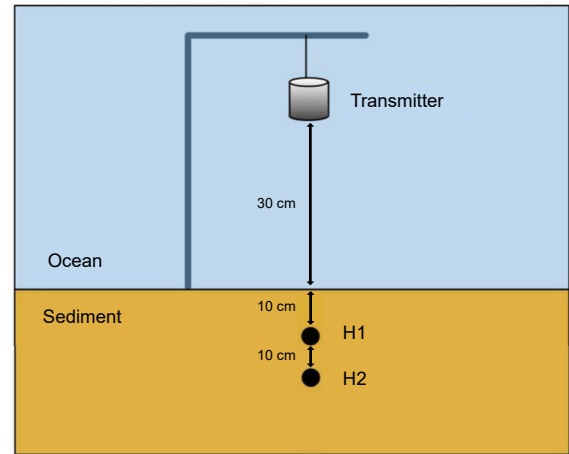


Fig. 7. Diagram of the experimental equipment layout.

The transmitter employed in the experiment is a cylindrical piezoelectric transducer, which operates within a frequency band ranging from 50 Hz to 20 kHz. It is suspended directly above the buried hydrophones

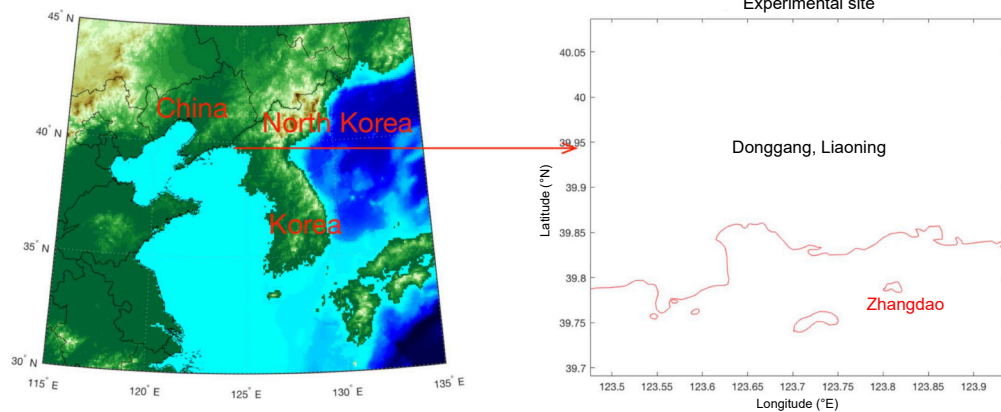


Fig. 6. Experimental site.

using an inverted L -shaped bracket, positioned 30 cm above the sediment surface, with the acoustic axis oriented vertically downward. The transmitted chirp signal spans a frequency range of 100 Hz to 15 kHz and utilizes Blackman window modulation, with a modulated pulse width of 8 ms and a pulse emitted every 1 s. The A/D sampling rate is set at 100 kHz, and the transmission data is recorded every 5 min for a duration of 200 min. Additionally, tidal height is monitored using a miniSVP.

Figure 8 illustrates the time series and frequency spectra recorded by H1 and H2 at two different tidal heights during the experiment. In Figs. 8a and 8b, the time series and frequency spectra captured by H1 and H2 at 21:00, when the water depth was 2.85 m, are presented. Conversely, Figs. 8c and 8d show the corresponding data recorded at 22:15, at a water depth of 3.41 m. Notably, the received signals vary significantly based on the tidal height and the depth of

the hydrophones. The acoustic signals at H1 and H2 can be distinctly identified, as the propagation path length of the acoustic signals received by H2 in the gas-bearing sediment is longer than that received by H1. Consequently, the signal attenuation recorded by H2 is greater than that of H1, leading to a lower amplitude for the H2 signal compared to H1.

The sound speed of gas-bearing sediments at each frequency point can be determined by analyzing the phase difference $\Delta\phi(f)$ between the signals received by the two hydrophones when the transmitted signals have a specified bandwidth. Additionally, the attenuation coefficient can be derived through a comparison of the power spectra of the signals captured by both hydrophones. The sound speed and attenuation coefficients are calculated as follows (YU *et al.*, 2015):

$$c_p(f) = c \left(1 + \frac{c\Delta\phi(f)}{\omega\delta x} \right), \quad (24)$$

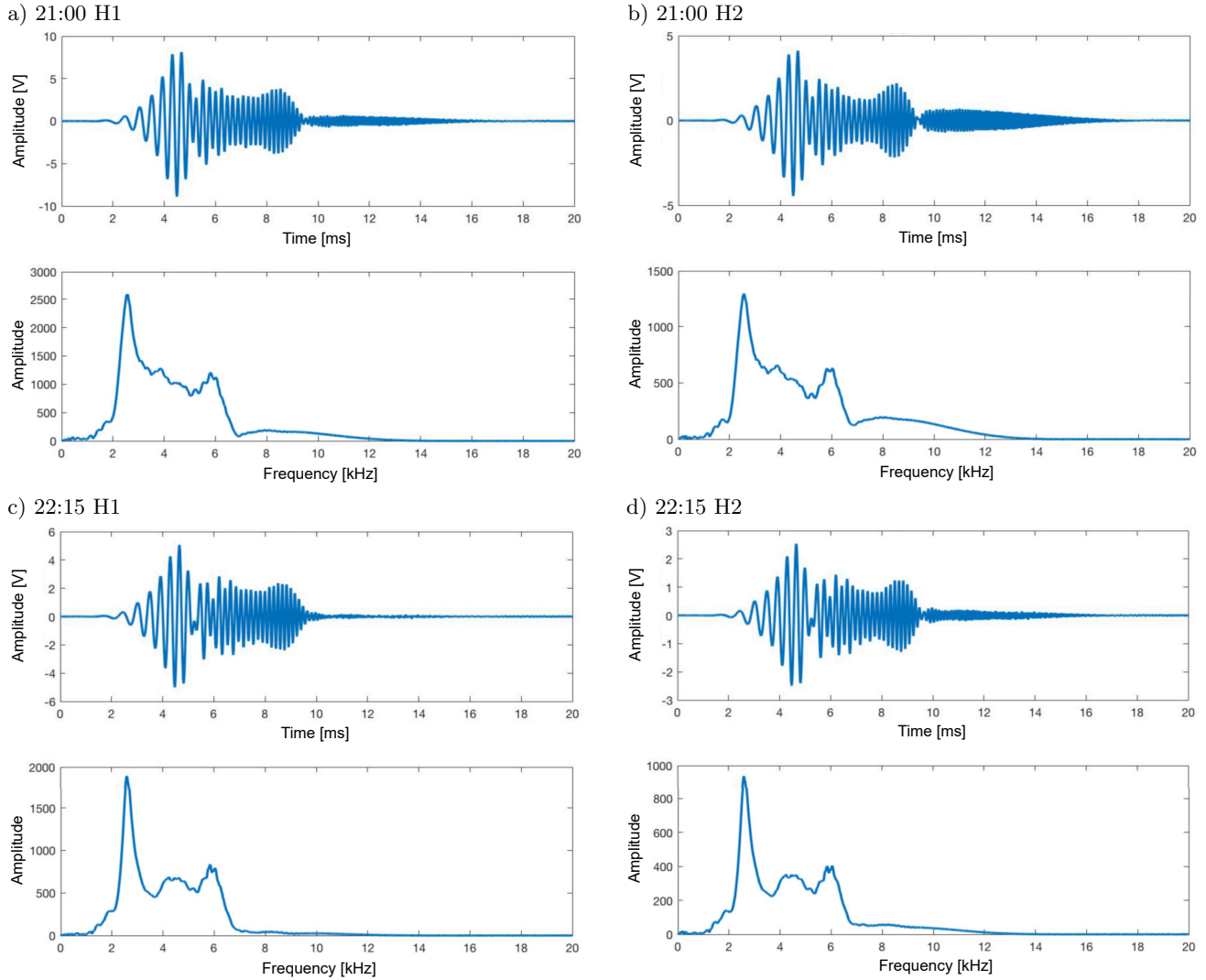


Fig. 8. Time series and frequency spectra recorded by hydrophones H1 and H2 for two different tidal heights during the experiment.

$$\alpha(f) = 10 \frac{1}{\delta x} \lg \left(\frac{A_1(f)}{A_2(f)} \right), \quad (25)$$

where δx is the distance between H1 and H2 (equal to 10 cm), and $A_1(f)$ and $A_2(f)$ are the power spectral density functions of the signals received by H1 and H2, respectively.

The analysis of sound speed and attenuation was conducted using a wide-band measurement method. Detailed derivations and implementation steps for this approach are provided in Appendix, which supports the experimental findings.

Figure 9 illustrates the relationship between sound speed and attenuation as a function of frequency (100 Hz–6000 Hz) over a half tidal cycle, with the solid line indicating water depth. The water depth varies from 0.65 m at 19:30 to 3.41 m at 22:15, before decreasing to 3.33 m at 22:45. In the upper graph of Fig. 9b, three prominent lines represent the three identified attenuation peaks. Notably, the attenuation

peak within the 4 kHz–6 kHz frequency band exhibits a slight shift towards higher frequencies, which can be attributed to the increase in tidal height and hydrostatic pressure in the sediments. However, no significant frequency shifts are observed for the other two attenuation peaks, likely due to minimal changes in hydrostatic pressure.

To better visualize the frequency shifts associated with hydrostatic pressure variations, the measured changes in sound speed and attenuation coefficient as a function of frequency at both initial and final times are presented in Figs. 9c and 9d. The resonance peaks labeled A–C in Fig. 9d correspond to the three bright lines in Fig. 9b. The frequency of the attenuation peak represented by the dotted line exceeds that of the peak indicated by the solid line, as evidenced by a comparison of the two attenuation curves. The attenuation peaks and their associated frequencies are detailed in Table 3, which also lists the approximate bubble radii for the peak frequencies. According to the frequency

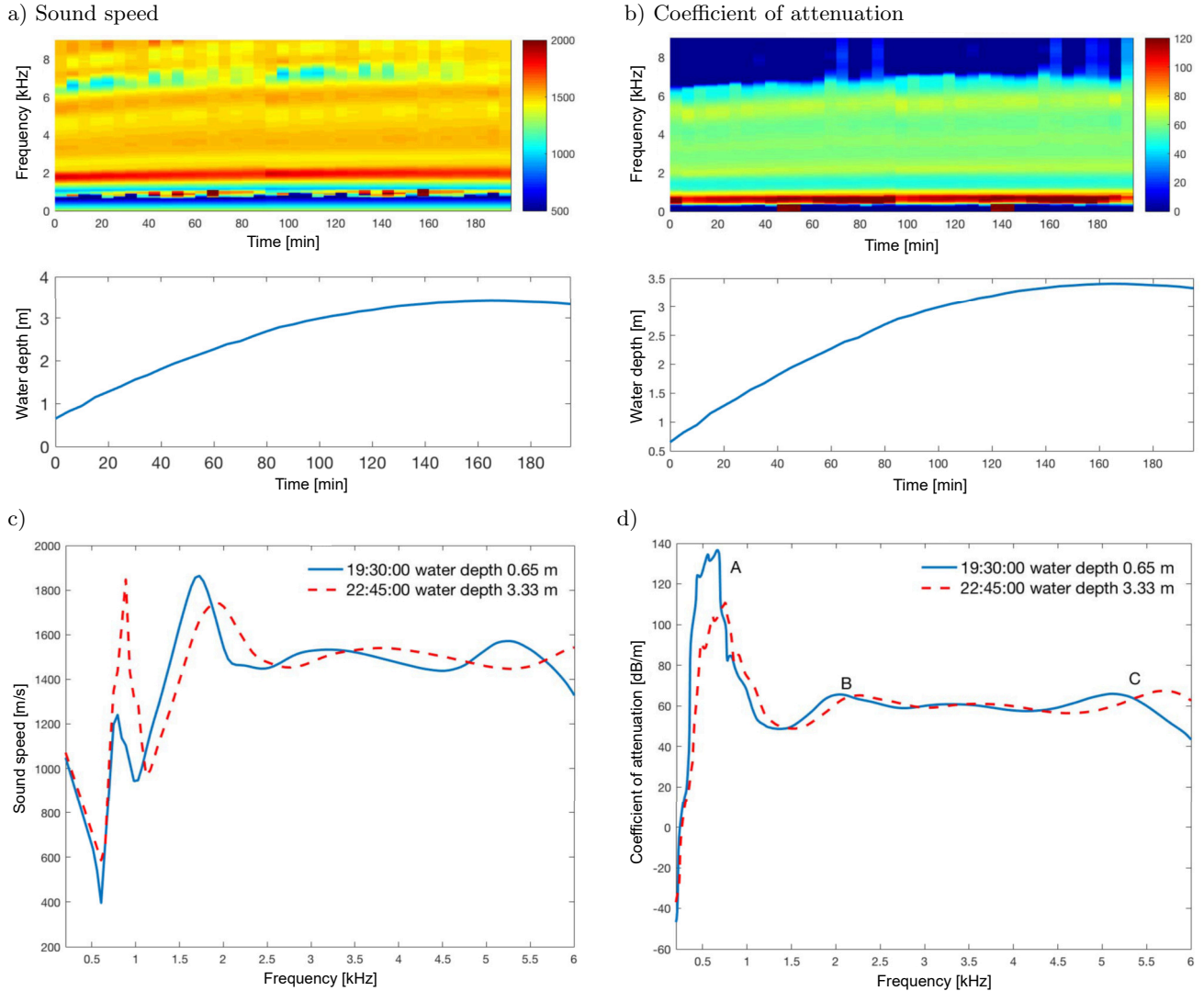


Fig. 9. Experimental results: frequency-dependent changes in the speed of sound and damping factor during changes in the height of the water column.

Table 3. Parameters of the attenuation peaks.

Attenuation peaks	19:30:00 Water depth 0.65 m		22:45:00 Water depth 3.33 m		Model predictions
	Resonance peak [Hz]	Attenuation coefficient [dB/m]	Resonance frequency [Hz]	Attenuation coefficient [dB/m]	Bubble radius [mm]
<i>A</i>	664	138	750	110	4.528
<i>B</i>	1853	69	2127	66	1.693
<i>C</i>	4923	68	5626	68	0.575

data in Table 3, the frequency shifts for peaks *A*, *B*, and *C* are 86 Hz, 274 Hz, and 703 Hz, respectively. This observation suggests that smaller bubbles result in larger frequency shifts due to increased hydrostatic pressure. Furthermore, the sound speed represented in Figs. 9a and 9c displays greater complexity, with variations that are more pronounced than the corresponding attenuation data. These fluctuations in sound speed can be attributed to changes in bubble behavior from inductive to capacitive near the bubble resonance frequency, leading to a phase jump at resonance and resulting in significant fluctuations in sound speed.

The frequency of attenuation peaks illustrated in Fig. 9 increases with water depth, confirming that the increase in hydrostatic pressure alters the resonance frequency of the bubble. According to Eq. (8), the resonance frequency of the bubble is proportional to the square root of hydrostatic pressure and inversely proportional to the bubble radius. This relationship suggests that smaller bubble radii will result in larger frequency shifts due to changes in hydrostatic pressure, an observation that aligns with the experimental data. However, the frequency shifts predicted by the model (78 Hz, 232 Hz, and 585 Hz) differ significantly from the measured data (86 Hz, 274 Hz, and 703 Hz). This discrepancy can be attributed to the fact that the measured sound speed and attenuation depend on a specific distribution of bubble sizes, making it inappropriate to interpret frequency shifts for bubbles of varying sizes.

Although the bubble radii correspond to peaks *A–C* (Table 3), bubble sizes are distributed across marine sediments. The bubble size distribution is derived from sound speed and attenuation data using the proposed inversion method, and the physical parameters of the measured sediment are detailed in Table 4. The best-fit bubble size distribution (ranging from 0.2 mm to 8 mm) is presented in Fig. 10c, with the highest gas content falling within the 6 mm–8 mm range. This range corresponds to the peak of the attenuation coefficient near the 500 Hz frequency and coincides with a sharp fluctuation in sound speed. The fitted sound speed models and attenuation curves are depicted in Figs. 10a and 10b, showing that the magnitudes of these sound speeds and attenuation coefficients

Table 4. Physical parameters of the measured sediment.

	Parameters	Values
Sediment parameters	Grain density	2478 kg/m ³
	Grain diameter	0.145 mm
	Fluid bulk modulus	2.193 Pa × 10 ⁹ Pa
	Grain bulk modulus	3.6 Pa × 10 ¹⁰ Pa
	Fluid viscosity	1.002 Pa · s × 10 ⁻³ Pa · s
	Porosity	0.45
	Fluid density	998.2 kg/m ³
	Structure factor	1.35
Gas parameters	Gas density	1.1691 kg/m ³
	Gas velocity	340 m/s
	Equilibrium pressure	1.01 Pa × 10 ⁵ Pa
	Thermal diffusivity	2.4 m ² /s × 10 ⁻⁵ m ² /s
	Surface tension	72.75 N/m × 10 ³ N/m
	Ratio of specific heat	1.4

are consistent with the measured attenuation data, thereby validating the proposed inversion method.

4. Conclusion

This study presented an inversion method for estimating bubble size distribution in gas-bearing sediments. The methodology integrates a corrected effective density fluid model with a cubic *B*-spline approach. The nonlinear inverse problem can be transformed by solving a set of equations involving the coefficients of cubic *B*-splines. Notably, this method allows for simultaneous estimation of bubble size distribution from measured sound speed and attenuation data. To validate the accuracy and robustness of this method, comparisons with other techniques for measuring bubble size distribution are necessary.

The method proposed in this paper integrates a suitable acoustic model – specifically, an effective density fluid model (ZHENG *et al.*, 2017), adapted to account for gas bubble pulsations – with *B*-spline expansions. This approach allows for the simultaneous consideration of sound speed and attenuation in addressing the inverse problem. Additionally, the proposed method offers greater applicability for real-time monitoring of shallow gas in marine sediments, compared to conventional inversion methods for bubble size distribution.

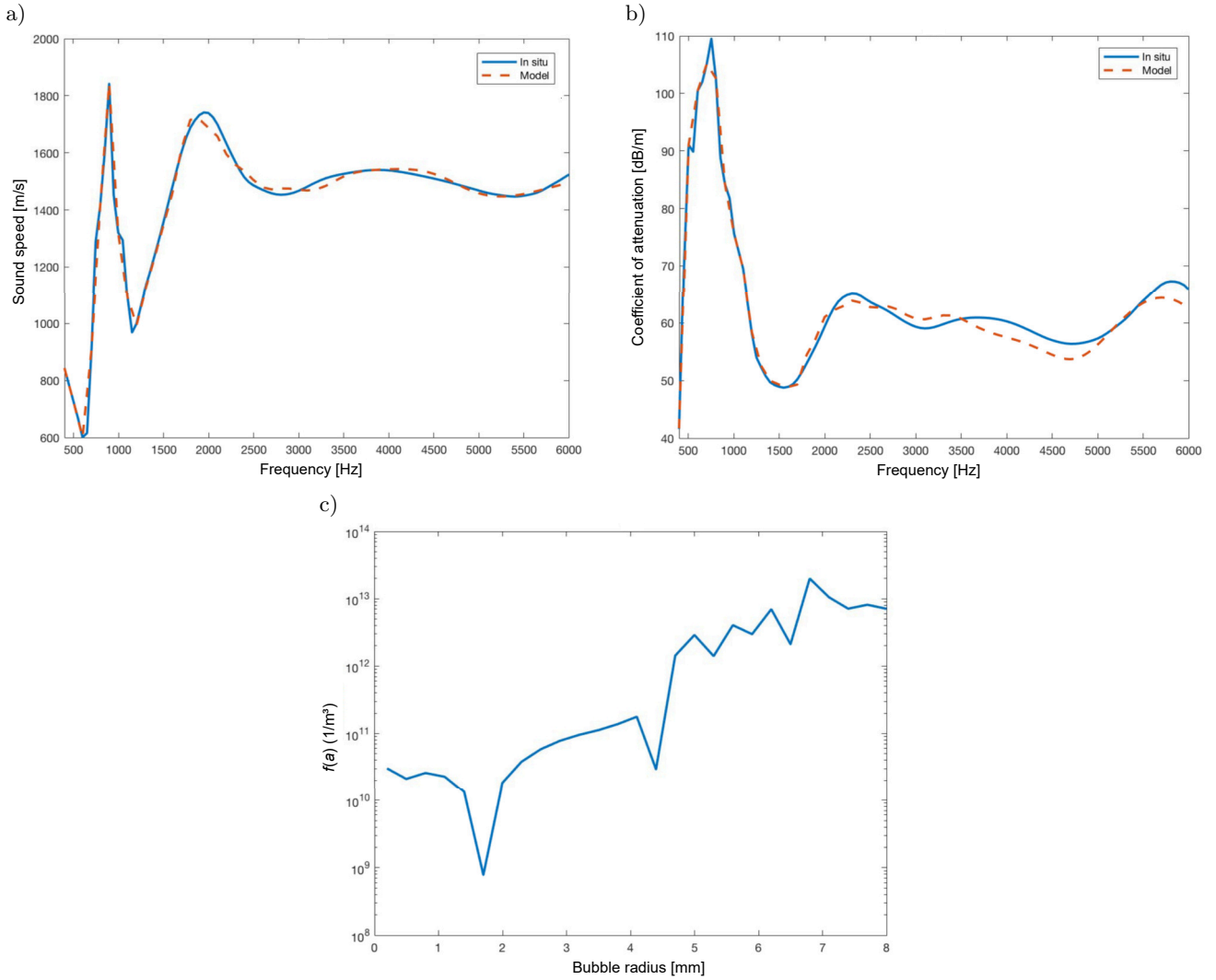


Fig. 10. Model fitting results: a) comparison of measured sound speed with model predictions; b) comparison of measured attenuation coefficient with model predictions; c) inversion results of bubble size distribution.

Appendix

In this appendix, we describe the wide-band method used for measuring sound speed and attenuation in marine sediments. This technique involves the use of a chirp signal modulated by a Blackman window to achieve accurate broadband measurements. The Blackman window offers several advantages, which are:

- 1) the compressed signal envelope is nearly free of sidelobes, unlike the normal signal, which retains smaller sidelobes. Ignoring these sidelobes can result in the loss of some information;
- 2) the signal's bandwidth is narrower, reducing distortion from the transmitter transducer, which has a limited bandwidth;
- 3) the reduced direct waveform shows less significant distortion at the band edges compared to the original signal. For these reasons, the Blackman window is used to modulate the amplitude of the

transmit signal. As long as the signal's bandwidth is wide enough and it has high time-delay resolution, the compressed signal can be separated in the time domain, minimizing amplitude and phase distortion. This ensures accurate broadband measurements of sound speed and attenuation.

When a sound wave passes through a sample with thickness $d_1 = x_2 - x_1$, let $p(x_1, \omega)$ be the sound pressure at x_1 . In the frequency domain, ignoring the time factor $e^{-j\omega t}$ and assuming a plane wave, the sound pressure received at x_2 can be written as

$$\begin{aligned}
 p(x_2, \omega) &= D_1 p(x_1, \omega) e^{jk(\omega)d_1} \\
 &= D_1 p(x_1, \omega) e^{j[\beta(\omega) + j\alpha(\omega)]d_1} \\
 &= D_1 p(x_1, \omega) e^{-\alpha(\omega)d_1} \exp(j\beta(\omega)d_1),
 \end{aligned} \tag{26}$$

where $k(\omega)$ is the complex wavenumber in the sample; its real part $\beta(\omega) = \omega/c_p(\omega)$ represents the phase velocity, whereas the imaginary part $\alpha(\omega)$ represents the

attenuation coefficient in [Np/m]; $c_p(\omega)$ is the compression wave phase velocity, and D_1 is the transmission coefficient at the water-sediment interface. If this sample with thickness d_1 is considered as a system, its transmission function can be expressed as

$$H_{s1}(j\omega) = D_1 e^{-\alpha(\omega)d_1} \exp(j\beta(\omega)d_1). \quad (27)$$

Assuming that the distance between the sound source and hydrophone is l and that the sound velocity dispersion and attenuation in the water column are neglected, the transfer function in the water column c_ω can be written as

$$H_{w1}(j\omega) = \exp(j\omega(l - d_1)/c_\omega). \quad (28)$$

Replacing the sample with a thickness of d_2 ($d_2 > d_1$) and keeping the same source-to-hydrophone distance, we have

$$\begin{aligned} H_{s2}(j\omega) &= D_2 e^{-\alpha(\omega)d_2} \exp(j\beta(\omega)d_2), \\ H_{w2}(j\omega) &= \exp(j\omega(l - d_2)/c_\omega). \end{aligned} \quad (29)$$

The ratio of the received signal spectrum is as follows:

$$\begin{aligned} H_r(j\omega) &= [H_{w2}(j\omega)H_{s2}(j\omega)] / [H_{w1}(j\omega)H_{s1}(j\omega)] \\ &= \frac{D_2}{D_1} e^{-\alpha(\omega)\Delta d} \exp\{j[\beta(\omega)\Delta d - \omega\Delta d/c_\omega]\}, \end{aligned} \quad (30)$$

where $\Delta d = d_2 - d_1$. Given $\Delta\phi = \beta(\omega)\Delta d - \omega\Delta d/c_\omega$, the sound speed and attenuation coefficient in the sample can be calculated as follows:

$$c_p = c_\omega \left(1 + \frac{c_\omega \Delta\phi}{\omega \Delta d}\right)^{-1}, \quad (31)$$

$$\alpha_p = -\frac{20 \lg e}{\Delta d} \ln \left[\frac{D_1}{D_2} |H_r(j\omega)| \right], \quad (32)$$

where α_p is the attenuation coefficient in [dB/m]. Thus, the sound speed is determined from the phase difference of the received signal, and the attenuation coefficient is calculated from the ratio of the amplitude spectra of the received signal.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (grant no. 12304501), the Science and Technology on Sonar Laboratory foundation (grant no. 2022-JCJQ-LB-031-02), and the Youth Elite Scientists Sponsorship Program by CAST (grant no. YESS20200330).

References

1. ANDERSON A.L., ABEGG F., HAWKINS J.A., DUNCAN M.E., LYONS A.P. (1998), Bubble populations and

acoustic interaction with the gassy floor of Eckernförde Bay, *Continental Shelf Research*, **18**(14–15): 1807–1838, [https://doi.org/10.1016/S0278-4343\(98\)00059-4](https://doi.org/10.1016/S0278-4343(98)00059-4).

2. ANDERSON A.L., HAMPTON L.D. (1980a), Acoustics of gas bearing sediments. I. Background, *The Journal of the Acoustical Society of America*, **67**(6): 1865–1889, <https://doi.org/10.1121/1.384453>.
3. ANDERSON A.L., HAMPTON L.D. (1980b), Acoustics of gas bearing sediments. II. Measurements and models, *The Journal of the Acoustical Society of America*, **67**(6): 1890–1903, <https://doi.org/10.1121/1.384454>.
4. BEST A.I., TUFFIN M.D.J., DIX J.K., BULL J.M. (2004), Tidal height and frequency dependence of acoustic velocity and attenuation in shallow gassy marine sediments, *Journal of Geophysical Research: Solid Earth*, **109**(B8): 589–600, <https://doi.org/10.1029/2003JB002748>.
5. CHEN J. *et al.* (2023), Elastic wave velocity of marine sediments with free gas: Insights from CT-acoustic observation and theoretical analysis, *Marine and Petroleum Geology*, **150**: 106169, <https://doi.org/10.1016/j.marpetgeo.2023.106169>.
6. COMMANDER K.W., McDONALD R.J. (1991), Finite-element solution of the inverse problem in bubble swarm acoustics, *The Journal of the Acoustical Society of America*, **89**(2): 592–597, <https://doi.org/10.1121/1.400671>.
7. DOGAN H., WHITE P. R., LEIGHTON T.G. (2015), Acoustic inversion for gas bubble distributions in marine sediments: Mercury results, *Seabed and Sediment Acoustics*, <https://doi.org/10.25144/16045>.
8. EDRINGTON T.S., CALLOWAY T.M. (1984), Sound speed and attenuation measurements in gassy sediments in the Gulf of Mexico, *Geophysics*, **49**(3): 297–299, <https://doi.org/10.1190/1.1441662>.
9. FLEISCHER P., ORSI T., RICHARDSON M., ANDERSON A. (2001), Distribution of free gas in marine sediments: A global overview, *Geo-Marine Letters*, **21**: 103–122, <https://doi.org/10.1007/s003670100072>.
10. FONSECA L., MAYER L., ORANGE D., DRISCOLL N. (2002), The high-frequency backscattering angular response of gassy sediments: Model/data comparison from the Eel River Margin, California, *The Journal of the Acoustical Society of America*, **111**(6): 2621–2631, <https://doi.org/10.1121/1.1471911>.
11. KARPOV S.V., KLUSEK Z., MATVEEV A.L., POTAPOV A.I., SUTIN A.M. (1996), Nonlinear interaction of acoustic waves in gas-saturated marine sediments, *Acoustical Physics*, **42**(4): 464–470.
12. LEIGHTON T.G. (2007), Theory for acoustic propagation in marine sediment containing gas bubbles which may pulsate in a non-stationary nonlinear manner, *Geophysics Research Letters*, **34**(17): 607, <https://doi.org/10.1029/2007GL030803>.
13. LEIGHTON T.G., ROBB G.B.N. (2008), Preliminary mapping of void fractions and sound speeds in gassy

- marine sediments from subbottom profiles, *The Journal of the Acoustical Society of America*, **124**(5): EL313–EL320, <https://doi.org/10.1121/1.2993744>.
14. MANTOUKA A., DOGAN H., WHITE P.R., LEIGHTON T.G. (2016), Modelling acoustic scattering, sound speed, and attenuation in gassy soft marine sediments, *The Journal of the Acoustical Society of America*, **140**(1): 276–282, <https://doi.org/10.1121/1.4954753>.
 15. RICHARDSON M.D., DAVIS A.M. (1998), Modeling methane-rich sediments of Eckernförde Bay, *Continental Shelf Research*, **18**(14–15): 1671–1688, [https://doi.org/10.1016/S0278-4343\(98\)00074-0](https://doi.org/10.1016/S0278-4343(98)00074-0).
 16. SHANKAR U., SINHA B., THAKUR N.K., KHANNA R. (2005), Amplitude-versus-offset modeling of the bottom simulating reflection associated with submarine gas hydrates, *Marine Geophysical Research*, **26**(1): 29–35, <https://doi.org/10.1007/s11001-005-2134-1>.
 17. SHANKAR U., THAKUR N.K., ASHALATHA B. (2006), Fluid flow related features as an indicator of potential gas hydrate zone: Western continental margin of India, *Marine Geophysical Research*, **27**(3): 217–224, <https://doi.org/10.1007/s11001-006-9001-6>.
 18. YU S., HUANG Y., LIU B., WANG F., ZHENG G. (2015), A wide-band method for sound speed and attenuation measurement in sediments, *Acta Acustica*, **40**(5): 682–694, <https://doi.org/10.15949/j.cnki.0371-0025.2015.05.009>.
 19. TÓTH Z., SPIESS V., KEIL H. (2015), Frequency-dependence in seismo-acoustic imaging of shallow free gas due to gas bubble resonance, *Journal of Geophysical Research-Solid Earth*, **120**(12): 8056–8072, <https://doi.org/10.1002/2015JB012523>.
 20. WILKENS R.H., RICHARDSON M.D. (1998), The influence of gas bubbles on sediment acoustic properties: In situ, laboratory, and theoretical results from Eckernförde Bay, Baltic sea, *Continental Shelf Research*, **18**(14): 1859–1892, [https://doi.org/10.1016/S0278-4343\(98\)00061-2](https://doi.org/10.1016/S0278-4343(98)00061-2).
 21. YARINA M., KATSNELSON B., GODIN O.A. (2023), Modal structure of the sound field in a shallow-water waveguide with a gassy sediment layer: Experiment and theory, *The Journal of the Acoustical Society of America*, **153**(3): A375–A375, <https://doi.org/10.1121/10.0019231>.
 22. ZHANG D., YANG J., WANG H., LI X. (2023), Prediction model of strength properties of marine gas-bearing sediments based on compressional wave velocity, *Applied Ocean Research*, **135**: 103562, <https://doi.org/10.1016/j.apor.2023.103562>.
 23. ZHENG G.Y., HUANG Y.W. (2016), Effect of linear bubble vibration on wave propagation in unsaturated porous medium containing air bubbles, *Acta Physica Sinica*, **65**(23): 234301, <https://doi.org/10.7498/aps.65.234301>.
 24. ZHENG G.Y., HUANG Y. W., HUA J., XU X., WANG F. (2017), A corrected effective density fluid model for gassy sediments, *The Journal of the Acoustical Society of America*, **141**(1): EL32–EL37, <https://doi.org/10.1121/1.4973616>.

Research Paper

An Under-Sampled Line Array Element Signal Reconstruction Method
Based on Compressed Sensing Theory

Tongjing SUN*, Mengwei ZHOU, Lei CHEN

*Department of Automation, Hangzhou Dianzi University
Hangzhou, China**Corresponding Author e-mail: stj@hdu.edu.cn*(received August 9, 2024; accepted October 29, 2024; published online February 4, 2025)*

The half-wavelength spacing arrangement of underwater uniform linear arrays has been widely used for better anti-interference performance and higher signal gain. However, practical challenges of small element spacing, numerous elements, high hardware costs, large data storage requirements, high processing complexity, and mutual coupling effects between elements, have hindered its widespread use. This paper proposes an under-sampled array signal reconstruction method based on the compressed sensing (CS) theory in the element domain. This method is not limited by the array configuration and constructs a deterministic measurement matrix that satisfies the restricted isometry property (RIP). Based on the array configuration, to ensure reconstruction performance. The method uses a two-dimensional orthogonal matching pursuit (OMP) method for time-space joint reconstruction of under-sampled spatial signals. Our simulation and practical test data processing results demonstrate that this method can achieve high-precision reconstruction of under-sampled array element domain signals at low under-sampling rates and can reconstruct full array signals with minimal error. Even under low signal-to-noise ratio (SNR) conditions, offering a practical and efficient solution to the challenges of underwater acoustic array signal processing.

Keywords: underwater acoustic array; compressed sensing; under-sampled array; signal reconstruction; deterministic measurement matrix.



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

1. Introduction

Array-based reception methods are usually used to resist interference and improve gain, including uniform linear arrays (ULA), uniform circular arrays (UCA), L-shaped arrays, and planar arrays (BALANIS, 2016; SILVER, 2019; ZHANG *et al.*, 2013). The ULA is the most common, featuring uniformly spaced array elements. Studies have shown that a ULA performs best when the spacing between array elements is half the wavelength. However, with the advent of large arrays such as towed line arrays, a larger array aperture is required to cover more spatial data. Using half-wavelength spacing necessitates an increasing number of array elements. For example, the number of elements can reach thousands for a ULA operating at 28 kHz with an array length of tens of meters. This leads to greater data storage requirements and increased processing complexity, exceeding active sonar systems' hardware and software processing capabilities, and thus affecting performance.

To address this issue, researchers have explored sparse arrays. Sparse arrays sample a subset of elements from a ULA, allowing the spacing between elements to exceed the half-wavelength limit, thereby reducing the number of elements while still achieving the desired performance. Nested arrays (PAL, VAIDYANATHAN, 2010) and coprime arrays (VAIDYANATHAN, PAL, 2011) are typical examples of sparse arrays. A coprime array is formed by interleaving two subarrays with coprime numbers of elements. In contrast, a nested array is created by nesting multiple levels of subarrays, with the spacing of each level determined by the number of elements in the previous level. Various improved sparse arrays have been proposed based on the concepts of nested and coprime arrays (HE *et al.*, 2022; MOHSEN *et al.*, 2023; YANG *et al.*, 2023). Sparse arrays are widely used in array signal processing because they increase the degrees of freedom. One approach is to use the difference coarray of a sparse array to construct an equivalent virtual ULA and then obtain the covariance matrix of the virtual ULA by vectorizing

the covariance matrix (LEI *et al.*, 2015; LI, ZHANG, 2020; KAZARINOV, 2022). However, this virtual domain method has limitations regarding the array configuration. If the difference coarray of a sparse array has holes, it increases the processing complexity.

The advent of compressed sensing (CS) theory has provided an effective means for array signal processing (CANDÈS, WAKIN, 2008; ENDER, 2010). The core idea is to take advantage of the signal sparsity to reduce the amount of sampled data, which has been extensively applied in processing under-sampled signals in the time domain (LI, YANG, 2014; JURDANA *et al.*, 2023). Since signal sources are sparse in the spatial domain, naturally satisfying the sparsity requirement of CS, it has also been applied to signal reconstruction in the spatial domain. MIRZA *et al.* (2020) have proposed a CS technique based on a sparse array for direction of arrival (DOA) estimation, addressing grid mismatch issues in spatial CS, thereby enhancing the robustness of CS DOA techniques. KIKUCHI *et al.* (2022) applied CS theory to process ULA, effectively reducing the number of elements in antenna arrays. The measurement matrices used in these studies are random. Although the random measurement matrices satisfied the restricted isometry property (RIP) and yielded satisfactory results in reconstruction accuracy, it is impossible to determine the configurations of the sparse arrays obtained by sampling, thus hindering engineering implementation. For deterministic measurement matrices, SALAMA (2020), LAKSHMI *et al.* (2021), and CHEN *et al.* (2020) used the difference co-array of nested arrays to construct an equivalent ULA and vectorized the covariance matrix of the sparse array to reconstruct the ULA's received signal. However, these methods only reconstruct the covariance matrix of the ULA's received signal and cannot reconstruct the received signal in the element domain.

This paper applies CS theory to the reconstruction of element-domain signals. By constructing a sensing matrix and using a two-dimensional orthogonal matching pursuit (OMP) method, the time-domain signals are projected onto the element domain to achieve the reconstruction of under-sampled array signals. This approach imposes fewer restrictions on the array configurations of sparse arrays for signal reconstruction. Furthermore, reconstructing signals in the element domain allows sampling only a portion of the array elements to obtain the entire array's received data, effectively reducing the data storage requirements for large arrays.

2. Compressed sensing theory

For sparse signals, CS theory samples signals at a rate much lower than the Nyquist sampling theorem to obtain discrete samples of the original signals. These samples are then used to reconstruct the origi-

nal signals through reconstruction algorithms. If a signal can be sparsely represented, a measurement matrix unrelated to the transformation basis can be designed to observe it. The observed values can then be used to achieve exact or approximate signal reconstruction by solving optimization problems. The process mainly includes two parts: CS observation and signal reconstruction.

2.1. Compressed sensing observation part

Consider an N -dimensional discrete-time domain signal \mathbf{X} and an $N \times N$ -dimensional sparse representation matrix Ψ , consisting of $N \times N$ -dimensional basis vectors. If the signal \mathbf{X} can be represented as

$$\mathbf{X} = \sum_{i=1}^N \psi_i \alpha_i = \Psi \alpha, \quad (1)$$

where α is a sparse vector containing only K ($K \ll N$) non-zero values, this implies that \mathbf{X} can be sparsely represented. Then, a measurement matrix $\Phi \in R^{M \times N}$ ($M \ll N$) that satisfies certain conditions is used to “sense” the signal, resulting in an M -dimensional observation signal of \mathbf{X} :

$$\mathbf{Y} = \Phi \mathbf{X}. \quad (2)$$

The process of CS observation is illustrated in Fig. 1.

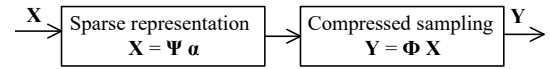


Fig. 1. Process of CS observation.

2.2. Signal reconstruction sections

After obtaining the linear observation vector \mathbf{Y} of the signal \mathbf{X} with respect to the measurement matrix Φ , the next step is to determine how to recover \mathbf{X} from \mathbf{Y} . Since directly solving the underdetermined Eq. (2) is infeasible, we use the sparse representation of \mathbf{X} in Eq. (1):

$$\mathbf{Y} = \Phi \mathbf{X} = \Phi \Psi \alpha = \Theta \alpha, \quad (3)$$

where $\Theta = \Phi \Psi$ is a $M \times N$ -dimensional matrix called the sensing matrix. We can think of \mathbf{Y} as the projection of α onto the sensing matrix Θ . Thus, the problem now becomes recovering α from \mathbf{Y} . Although Eq. (3) is also an underdetermined equation, the sparsity of α significantly reduces the number of unknowns, making signal reconstruction feasible.

CANDÈS and WAKIN (2008) proved that, under the condition that the signal α is sparse, if the sensing matrix Θ satisfies the condition that any $2K$ columns are linearly independent, the solution can be obtained using the following equation:

$$\begin{cases} \hat{\alpha} = \arg \min \|\alpha\|_0, \\ \text{subject to } \Theta \alpha = \mathbf{Y}. \end{cases} \quad (4)$$

Equation (4) is an NP-hard non-convex optimization problem, making it very challenging to solve. Numerous optimization algorithms have been proposed to address this issue (ZHAO, NEHORAI, 2014; WANG *et al.*, 2022). After recovering α through the reconstruction algorithm, the signal \mathbf{X} can be reconstructed according to Eq. (1).

The process of signal reconstruction can be illustrated in Fig. 2.

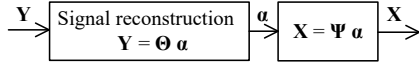


Fig. 2. Process of CS reconstruction signal.

From the aforementioned analysis, it can be concluded that the primary research focus of CS encompasses the following three aspects:

- 1) sparse representation: designing a sparse representation matrix to represent the original signal \mathbf{X} as a sparse vector α of the same length;
- 2) compressed sampling: using an $M \times N$ measurement matrix, CS observes the high-dimensional original signal \mathbf{X} to obtain the low-dimensional observed signal \mathbf{Y} ;
- 3) signal reconstruction: recovering the original signal \mathbf{X} from the observed signal \mathbf{Y} by solving Eq. (4).

3. Under-sampled array signal reconstruction method in the element domain

The application of CS in the time domain mainly deals with one-dimensional signals. However, array reception signals are typically two-dimensional, encompassing both the element domain (spatial domain)

and the time domain. Focusing on the three critical technologies of sparse representation, compressed sampling, and signal reconstruction, this paper uses the inherent sparsity of spatial arrays, constructing a sparse matrix from steering vectors of various angles for sparse representation in the spatial domain. Based on the configuration of the under-sampled array, a measurement matrix satisfying the RIP condition is constructed using a unit diagonal sampling method. It extends the OMP method to the two-dimensional space-time joint domain for signal reconstruction in the element domain. The implementation process is shown in Fig. 3.

3.1. Sparse representation of array signals

CS requires the original signal to be sparsely representable. When applied to the element domain, the target is sparse in the spatial domain, naturally satisfying the sparsity condition.

Suppose a ULA with N hydrophones spaced by d receives K signals with identical central frequency f_0 and wavelength λ . We first consider the case of a single snapshot, where the time-domain signal received by the array can be described as

$$\mathbf{X} = \mathbf{A} \mathbf{s} + \mathbf{N}, \quad (5)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T$, $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_N]^T$, \mathbf{x}_N , \mathbf{n}_N , respectively, represent the signal and additive noise received by the N -th array element. Additionally, $\mathbf{s} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_K]^T$, where \mathbf{s}_K represents the K -th incident signal on the array. The matrix \mathbf{A} is the $N \times K$ -dimensional array manifold matrix:

$$\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_k)]^T, \quad (6)$$

where $\mathbf{a}(\theta_k)$ is the steering vector of the array in the direction θ_k :

$$\mathbf{a}(\theta_k) = \left[1, \exp\left(-\frac{j2\pi d \sin(\theta_k)}{\lambda}\right), \exp\left(-j2\pi \cdot \frac{2d \sin(\theta_k)}{\lambda}\right), \dots, \exp\left(-j2\pi \cdot \frac{2d \sin(\theta_k)}{\lambda}\right) \right]^T. \quad (7)$$

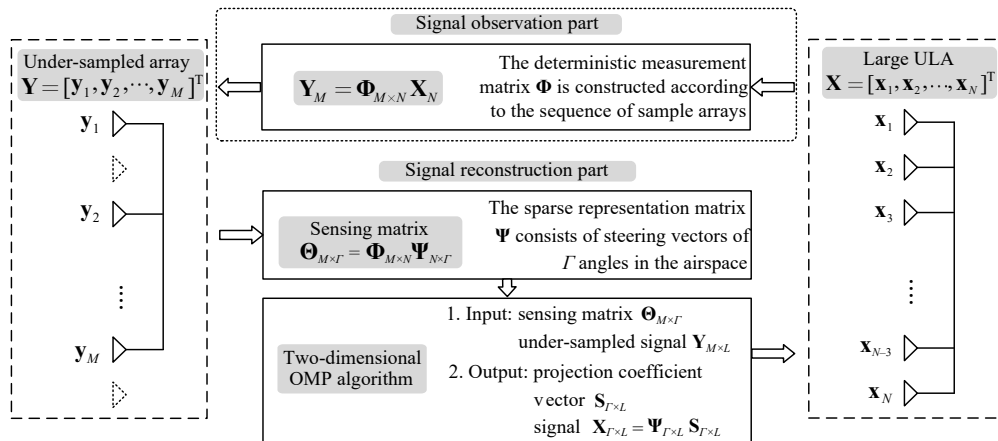


Fig. 3. Flow chart of signal reconstruction method of under-sampled array.

If the spatial domain from -90° to 90° is divided into Γ grids, and the incident angles of the K signal sources fall on these grids, we obtain Γ spatial angles. The array steering vectors at these Γ angles are used to form an extended array manifold matrix Ψ . Thus, Eq. (5) can be further expressed as:

$$\mathbf{X} = \Psi \mathbf{s} + \mathbf{N}, \quad (8)$$

where \mathbf{s} is a Γ -dimensional projection coefficient vector, and each element of \mathbf{s} corresponds to a grid. Since there are only K grids with incident signals among the Γ grids, \mathbf{s} is a K -sparse vector, having a form similar to $[0, 0, \dots, \mathbf{s}_1, 0, \dots, 0, \dots, \mathbf{s}_K, 0, \dots, 0]^T$, where non-zero values occur only at the grids with incident signals.

Equation (8) shows that the array received signal \mathbf{X} is sparsely represented as a sparse vector \mathbf{s} through the extended array manifold matrix Ψ . The extended array manifold matrix Ψ serves as the sparse representation matrix, constructed through the following steps:

- 1) divide the spatial domain from -90° to 90° into Γ grids of equal angles, resulting in $\{\theta_1, \theta_2, \dots, \theta_\Gamma\}$;
- 2) obtain the steering vectors of the array at these Γ angles: $\{\mathbf{a}(\theta_k)\}_{k=1}^\Gamma$;
- 3) form the sparse representation matrix:
 $\Psi = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_\Gamma)]^T$.

3.2. Construction of the measurement matrix based on under-sampled array configuration

In CS, the under-sampling of large ULAs is achieved through a measurement matrix. The critical difference between element-domain CS and time-domain CS is that the measurement matrix of element-domain CS does not require a linear combination of all element signals for compressive sampling. From a hardware perspective, linear combination of element signals still necessitates sampling each element. However, the under-sampled signals we obtain only contain the received signals from a subset of elements.

In the element domain, the received signal of an N -element ULA can be expressed as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T$, where \mathbf{x}_i ($i = 1, 2, \dots, N$) is the signal received by

the i -th element. The measurement matrix Φ consists of $M \times N$ -dimensional sampling basis vectors ϕ_i ($i = 1, 2, \dots, M$), each of which samples the original array signal \mathbf{X} once, obtaining one element signal. In total, the M sampling basis sample M element signals, forming the under-sampled array signal $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M]^T$. To ensure that each sampling basis samples only one element, each N -dimensional sampling basis vector can be a sparse vector containing only one non-zero value. Moreover, to avoid redundant sampling, the positions of the non-zero values in the M sampling basis should be different.

Figure 4 illustrates the process of element signal under-sampling. By sorting the M sampling basis vectors according to the positions of their non-zero values, the measurement matrix has a structure similar to that of Eq. (9). It can be viewed as M rows extracted from an identity diagonal matrix, where the columns are linearly independent, ensuring that the resulting measurement matrix satisfies the RIP:

$$\Phi = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & & & \ddots & & & \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}. \quad (9)$$

Thus, the steps to construct the measurement matrix are as follows:

- 1) determine the positions of the sampled elements in the under-sampled array: $D = [d_1, d_2, \dots, d_M]$;
- 2) construct an N -dimensional identity diagonal matrix \mathbf{E} ;
- 3) extract the i -th rows from \mathbf{E} to form the measurement matrix Φ .

3.3. Signal reconstruction based on two-dimensional OMP

Subsections 3.1 and 3.2 discussed the mathematical model for the single snapshot case. Now, we consider

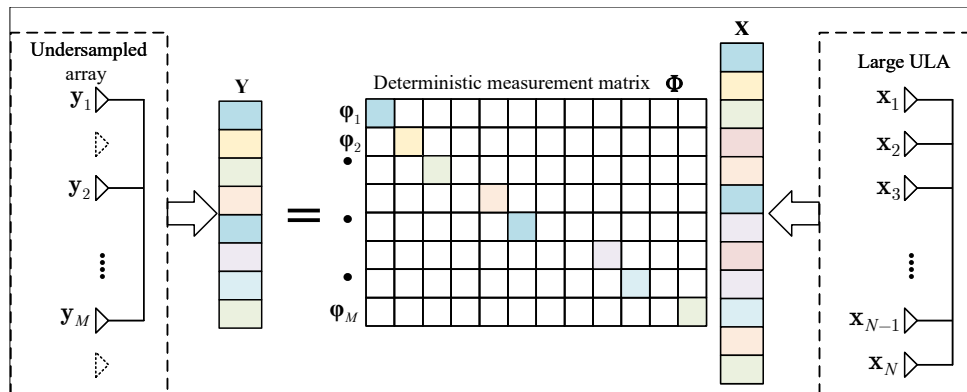


Fig. 4. Signal compression sampling in the element domain.

the scenario with L snapshots. The problem of reconstructing the original $N \times L$ -dimensional array signal \mathbf{X} from the $M \times L$ -dimensional array signal \mathbf{Y} can be described as

$$\begin{cases} \widehat{\mathbf{S}} = \arg \min \|\mathbf{S}\|_0, \\ \mathbf{X} = \mathbf{\Psi}\widehat{\mathbf{S}}, \\ \text{subject to } \mathbf{\Theta}\mathbf{S} = \mathbf{Y}, \end{cases} \quad (10)$$

where $\widehat{\mathbf{S}}$ is an $M \times L$ matrix containing L projection coefficient vectors; $\mathbf{\Psi}$ is an $N \times \Gamma$ -dimensional sparse representation matrix, and $\mathbf{\Theta}$ is an $M \times \Gamma$ sensing matrix obtained by $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Psi}$, with $\mathbf{\Phi}$ being the measurement matrix as detailed in Subsec. 3.2.

The reconstruction of the one-dimensional projection coefficient vector involves solving the problem in Eq. (4). In CS, there are numerous optimization algorithms available to solve Eq. (4). The OMP algorithm is one such reconstruction method, which uses an iterative approach to obtain the solution (TROPP, GILBERT, 2007). However, traditional OMP cannot handle two-dimensional signals as presented in Eq. (10). This paper utilizes a two-dimensional OMP algorithm. The under-sampled signal \mathbf{Y} is first divided into L vectors by columns, and then each vector is sequentially solved:

$$\begin{cases} \widehat{\mathbf{s}}_i = \arg \min \|\mathbf{s}_i\|_0, \\ \text{subject to } \mathbf{\Theta}\mathbf{s} = \mathbf{Y}_i, \end{cases} \quad (11)$$

where \mathbf{Y}_i denotes the i -th column of \mathbf{Y} , and $\widehat{\mathbf{s}}_i$ represents the projection coefficient vector reconstructed from it. Finally, the L projection coefficient vectors form the projection coefficient matrix $\widehat{\mathbf{S}}$.

The two-dimensional OMP algorithm process is as follows:

- 1) initialize the projection coefficient matrix $\mathbf{S}_{\Gamma \times L}$ and set the iteration count $i = 1$. repeat steps (2) to (4) L times until $i > L$, then proceed to step (5);
- 2) initialize the projection coefficient vector α_Γ , residual $\mathbf{r}_0 = \mathbf{Y}_i$, index set $\Lambda_0 = \emptyset$, and inner loop iteration count $n = 1$. Repeat steps (a) to (e) until the stopping criterion is met:

- a) find the atom column in $\mathbf{\Theta}$ most correlated with the residual and its index:

$$\lambda_n = \arg \max_{j \notin \Lambda_{n-1}} \|\langle \theta_j, \mathbf{r}_{n-1} \rangle\|,$$

where θ_j is the j -th column of $\mathbf{\Theta}$;

- b) update the index set: $\Lambda_n = \Lambda_{n-1} \cup \lambda_n$;
- c) solve the projection coefficient vector using least squares:

$$\mathbf{s}_n(t \in \Lambda_n) = \arg \min_x \|\mathbf{\Theta}_{\Lambda_n} x - \mathbf{Y}_i\|_2,$$

$$\mathbf{s}_n(t \notin \Lambda_n) = 0;$$

- d) update the residual: $\mathbf{r}_n = \mathbf{r}_{n-1} - \mathbf{\Theta}\mathbf{s}_n$;

- e) $n = n + 1$;

- 3) output the projection coefficient vector \mathbf{s} as the i -th row of $\mathbf{S}_{\Gamma \times L}$;

- 4) $i = i + 1$;

- 5) recover the signal: $\mathbf{X}_{N \times L} = \mathbf{\Psi}_{N \times \Gamma} \mathbf{S}_{\Gamma \times L}$.

Using the two-dimensional OMP algorithm, the entire array signal \mathbf{X} is reconstructed from the under-sampled array signal \mathbf{Y} .

4. Performance verification based on simulated and measured data

This section compares the reconstruction error under different under-sampling rates, array configurations, and signal-to-noise ratio (SNR) using simulated and measured data. For an N -element ULA, M elements are sampled. When $M < N$, the array is under-sampled, and the ratio M/N is the under-sampling rate. The reconstruction error is defined as

$$\text{Error} = \frac{\|\widehat{\mathbf{X}}_{N \times L} - \mathbf{X}_{N \times L}\|_2}{\|\mathbf{X}_{N \times L}\|_2}, \quad (12)$$

where $\widehat{\mathbf{X}}_{N \times L}$ is the reconstructed signal, and $\mathbf{X}_{N \times L}$ is the original signal.

4.1. Performance verification using simulated data

The simulation involves the transmission of linear frequency-modulated signals by active sonar with a center frequency of 28 kHz and a bandwidth of 16 kHz. A 32-element ULA receives the echo signal. The full array signal received by the ULA is the original signal \mathbf{X} . The under-sampled signal \mathbf{Y} is obtained using the constructed measurement matrix as described Eq. (2). The measurement matrix can be either deterministic, as shown is Subsec. 3.2, or random. The underwater sound speed is set to $c = 1500$ m/s, and the element spacing is half the wavelength of the echo signal.

4.1.1. Signal waveform comparison

In this section, the waveform of the original signal from unsampled elements is compared with the reconstructed signal at an under-sampling rate of 50 % and SNR = 5 dB. The positions of the 16 sampled elements are {1, 2, 3, 5, 10, 11, 15, 16, 17, 19, 22, 23, 24, 25, 31, 32}. The original and reconstructed signals from the 18th and 28th elements among the remaining 16 unsampled elements are compared, as shown in Fig. 5.

4.1.2. Reconstruction error of different under-sampled arrays

In this section, at an under-sampling rate of 50 % and SNR = 5 dB, 100 sets of under-sampled structures

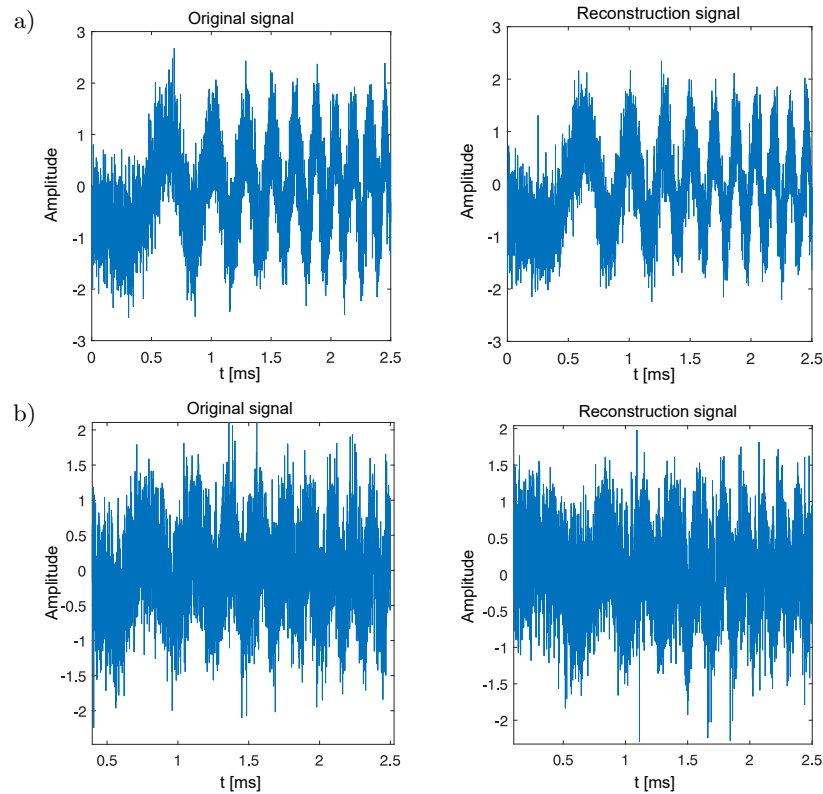


Fig. 5. Comparison of signal waveforms before and after reconstruction: a) the 18th element; b) the 28th element.

are independently tested using a random measurement matrix. Each set of structures undergoes five repeated experiments, and the average reconstruction error of the five experiments is taken as the reconstruction error for that set. The results are shown in Fig. 6, with reconstruction errors mainly ranging from 0.1 to 0.25, and some under-sampled structures exhibiting large errors.

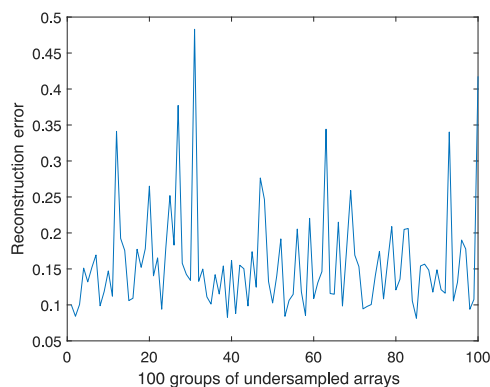


Fig. 6. Reconstruction errors of different under-sampled arrays.

4.1.3. Reconstruction error at different under-sampling rates

Five sets of under-sampled arrays are selected, and the under-sampling rates are gradually reduced from 87.5 % to 12.5 % by removing one redundant element

from the under-sampled array each time, while keeping other conditions unchanged. The reconstruction error at different under-sampling rates is then compared. Each set of under-sampled arrays undergoes five repeated experiments to avoid randomness, and the average error is computed. The results are shown in Table 1. From Table 1, it can be seen that when the under-sampling rate reaches 31.25 % or higher, the reconstruction error is generally below 0.2, indicating good reconstruction effect.

Table 1. Reconstruction errors of five under-sampled arrays at different under-sampling rates.

Under-sampling rate	25 %	31.25 %	37.5 %	50 %	75 %	87.5 %
1	0.702	0.151	0.14	0.094	0.071	0.063
2	0.677	0.124	0.109	0.102	0.069	0.062
3	0.895	0.124	0.1	0.081	0.055	0.059
4	0.47	0.23	0.195	0.165	0.078	0.068
5	0.528	0.265	0.251	0.185	0.095	0.061

4.1.4. Reconstruction error under different SNR

Ten groups of under-sampled arrays are selected to construct deterministic measurement matrices. Each group undergoes an independent experiment at an under-sampling rate of 50 %. The reconstruction performance under different SNRs is then analyzed. Each group of under-sampled arrays is subjected to five re-

peated experiments to avoid randomness. The results are shown in Table 2. It can be observed that high SNRs yield lower reconstruction errors. For signals without noise, the optimal reconstruction error can reach 0.009, which is almost negligible.

Table 2. Reconstruction errors of five under-sampled arrays under different SNR.

SNR [dB]	1	2	3	4	5
Noise is 0	0.066	0.011	0.094	0.009	0.009
5	0.079	0.115	0.155	0.111	0.095
7	0.067	0.095	0.133	0.064	0.080
10	0.05	0.074	0.114	0.048	0.067
15	0.055	0.047	0.097	0.035	0.051

4.2. Performance verification using measured data

The measured data is obtained from a lake test at the Xin'anjiang test site, where the underwater sound speed is approximately 1450 m/s. A linear frequency modulated signal with a frequency range of 20 kHz–36 kHz is transmitted with a pulse width of 2 ms. The test setup is shown in Fig. 7. A 32-element ULA receives the underwater echo signal, sampled at 1 MHz. The hydrophone array and target are 10 m underwater, and the transmitter is 9.5 m underwater. The target is a 0.6 m diameter spherical model.

Due to the complexity of the underwater environment, and to more clearly observe the target, we apply a matched filter to both the original signal \mathbf{X} and the reconstructed signal $\hat{\mathbf{X}}$, resulting in \mathbf{X}' and $\hat{\mathbf{X}}'$, respectively. Then, the reconstruction error is calculated using Eq. (13):

$$\text{Error} = \frac{\|\hat{\mathbf{X}}'_{N \times L} - \mathbf{X}'_{N \times L}\|_2}{\|\mathbf{X}'_{N \times L}\|_2}, \quad (13)$$

where $\hat{\mathbf{X}}'_{N \times L}$ is the matched filter signal of the reconstructed signal $\hat{\mathbf{X}}_{N \times L}$, and $\mathbf{X}'_{N \times L}$ is the matched filter signal of the original signal $\mathbf{X}_{N \times L}$.

4.2.1. Signal waveform comparison

In this section, the 32-element ULA is processed at a 75 % under-sampling rate. The positions of the sampled elements are selected as $\{1, 2, 3, 4, 5, 7, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 26, 27, 28, 29, 30, 32\}$. Among the remaining eight unsampled elements, the signals before and after reconstruction at the 23th element are compared along with the results of matched filter. The results are shown in Fig. 8.

4.2.2. Reconstruction error of different under-sampled arrays

This section processes the measured data at a 75 % under-sampling rate. One hundred groups of under-sampled arrays are sampled using a random measurement matrix. The results are shown in Fig. 9, where the reconstruction error fluctuates between 0.08 and 0.28.

4.2.3. Reconstruction errors of 10 under-sampled arrays

Five groups of under-sampled arrays are selected. Starting with an under-sampling rate of 87.5 %, redundant elements are gradually removed to reduce the under-sampling rate to 37.5 % while keeping other conditions unchanged. The reconstruction error at different under-sampling rates is compared. The results are shown in Table 3. As shown in Table 3, the proposed algorithm achieves optimal performance with measured data, with reconstruction errors below 0.1 when the under-sampling rate is above 50 %.

Table 3. Reconstruction errors of five groups of under-sampled arrays at different under-sampling rates.

Under-sampling rate	37.5 %	50 %	62.5 %	75 %	87.5 %
1	0.164	0.103	0.071	0.064	0.083
2	0.174	0.117	0.096	0.092	0.107
3	0.166	0.099	0.086	0.076	0.095
4	0.157	0.116	0.085	0.079	0.099
5	0.166	0.125	0.097	0.096	0.106

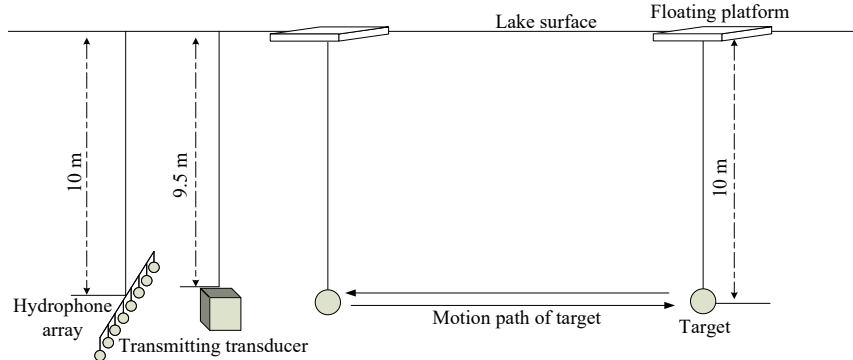


Fig. 7. Experimental setup on the lake.

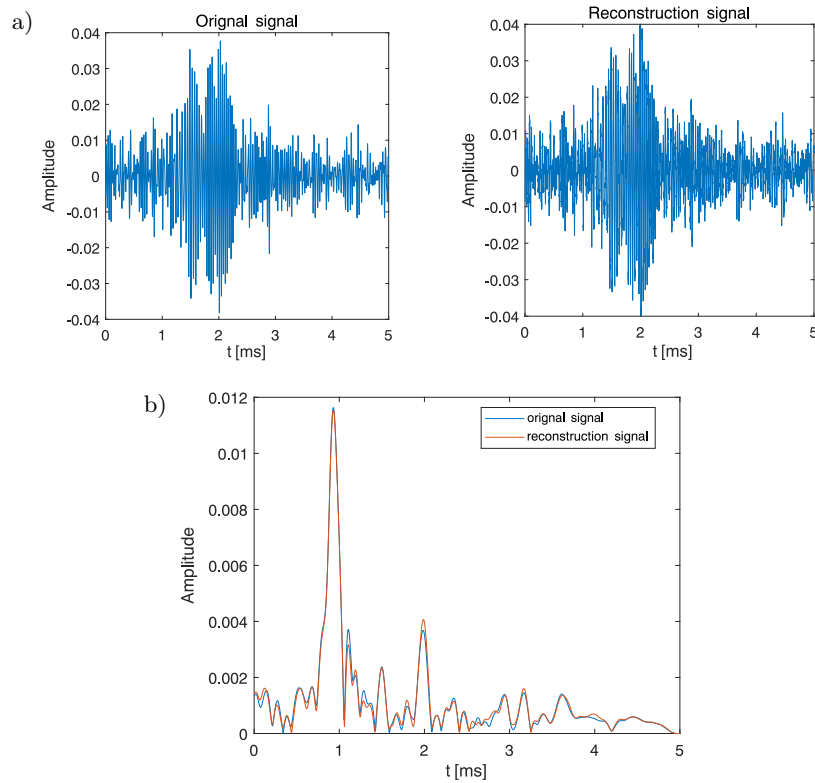


Fig. 8. Waveform comparison before and after reconstruction of measured data: a) signal comparison before and after reconstruction; b) comparison of matched filter results before and after reconstruction.

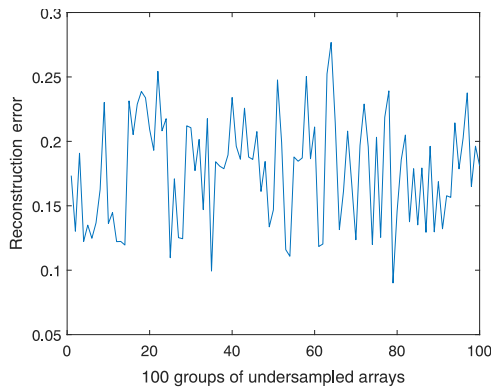


Fig. 9. Reconstruction error under different under-sampled arrays.

5. Conclusion

This paper addressed the under-sampling problem in large ULAs by applying CS theory to element-domain signal processing. The array signals were sparsely represented by exploiting the sparsity of signal sources in the spatial domain. Time-domain signals were projected onto the element domain through sparse representation. Then, reconstruction algorithms were used in the element domain to recover the full array signal from the under-sampled array signals. Compared to the method of reconstructing the original array covariance matrix, the element-domain signal reconstruction method directly processes the signal and

has broader applicability. Using CS for signal reconstruction allows recovering full array data from any under-sampled array, enabling data reception from redundant elements in large arrays without the need to sample them. The performance of this method is verified through the processing of both simulated and measured data, demonstrating that it can reconstruct element-domain signals with small errors even at low SNRs and varying under-sampling rates.

Acknowledgments

This article uses experimental data collected at the Dalian Test and Control Institute (China), and we gratefully acknowledge our colleagues for their experimental expertise. This work was supported by the Joint National Natural Science Foundation of China (no. U22A2044). It was also supported by the Key Laboratory Fund from Underwater Test and Control Technology (no. 2023-JCJQ-LB-030) and Underwater Acoustic Countermeasure Technology (no. JCKY2024207CH01).


Declaration of competing interest

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

References

1. BALANIS C.A. (2016), *Antenna Theory: Analysis and Design*, 4th ed., John Wiley & Sons, New Jersey.
2. CANDÈS E.J., WAKIN M.B. (2008), An introduction to compressive sampling, *IEEE Signal Processing Magazine*, **25**(2): 21–30, <https://doi.org/10.1109/MSP.2007.914731>.
3. CHEN L.H., MA X.C., LI X., SONG Q.Y. (2020), Sparse array beamforming method combined with compressed sensing model [in Chinese], *Journal of Signal Processing*, **36**(4): 475–485, <https://doi.org/10.16798/j.issn.1003-0530.2020.04.001>.
4. ENDER J.H.G. (2010), On compressive sensing applied to radar, *Signal Processing*, **90**(5): 1402–1414, <https://doi.org/10.1016/j.sigpro.2009.11.009>.
5. HE J., TANG M., SHU T., YU W.X. (2022), Linear co-prime sensor location arrays: Mutual coupling effect and angle estimation [in Chinese], *Journal of Electronics & Information Technology*, **44**(8): 2852–2858, <https://doi.org/10.11999/JEIT210489>.
6. JURDANA V., LOPAC N., VRANKIC M. (2023), Sparse time-frequency distribution reconstruction using the adaptive compressed sensed area optimized with the multi-objective approach, *Sensors*, **23**(8): 4148, <https://doi.org/10.3390/s23084148>.
7. KAZARINOV A.S. (2022), DOA estimation with sparse virtual arrays, [in:] *2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, pp. 1359–1362, <https://doi.org/10.1109/ElConRus54750.2022.9755489>.
8. KIKUCHI H., HOBARA Y., USHIO T. (2022), Compressive sensing to reduce the number of elements in a linear antenna array with a phased array weather radar, *IEEE Transactions on Geoscience and Remote Sensing*, **60**: 1–10, <https://doi.org/10.1109/TGRS.2022.3152998>.
9. LAKSHMI S., KUMAR N.S., HARI C.V., SRINATH S.A. (2021), Localization of underwater acoustic target using compressive sensing on nested array, [in:] *2021 Fourth International Conference on Microelectronics, Signals & Systems (ICMSS)*, pp. 1–5, <https://doi.org/10.1109/ICMSS53060.2021.9673646>.
10. LEI Y., WEI Y. S., LIU W. (2015), A novel adaptive beamforming technique for large-scale arrays, [in:] *2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 269–273, <https://doi.org/10.1109/ISSPIT.2015.7394341>.
11. LI P., YANG Y.X. (2014), Compressed sensing based acoustic data compression and reconstruction technology [in Chinese], *Technical Acoustics*, **33**(1): 14–20.
12. LI S., ZHANG X.P. (2020), A new approach to construct virtual array with increased degrees of freedom for moving sparse arrays, *IEEE Signal Processing Letters*, **27**: 805–809, <https://doi.org/10.1109/LSP.2020.2993956>.
13. MIRZA H.A., RAJA M.A.Z., CHAUDHARY N.I., QURESHI I.M., MALIK A.N. (2020), A robust multi sample compressive sensing technique for DOA estimation using sparse antenna array, *IEEE Access*, **8**: 140848–140861, <https://doi.org/10.1109/ACCESS.2020.3011597>.
14. MOHSEN N., HAWBANI A., WANG X.F., AGRAWAL M. (2023), Optimized sparse nested arrays for DoA estimation of non-circular signals, *Signal Processing*, **204**: 108819, <https://doi.org/10.1016/j.sigpro.2022.108819>.
15. PAL P., VAIDYANATHAN P.P. (2010), Nested arrays: A novel approach to array processing with enhanced degrees of freedom, *IEEE Transactions on Signal Processing*, **58**(8): 4167–4181, <https://doi.org/10.1109/TSP.2010.2049264>.
16. SALAMA A.A., AHMAD M.O., SWAMY M.N.S. (2020), Compressed sensing DOA estimation in the presence of unknown noise, *Progress in Electromagnetics Research C*, **102**: 47–62, <https://doi.org/10.2528/PIERC20031204>.
17. SILVER H.W. [Ed.] (2019), *The ARRL Antenna Book*, 24th ed., The American Radio Relay League, Newington.
18. TROPP J.A., GILBERT A.C. (2007), Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Transactions on Information Theory*, **53**(12): 4655–4666, <https://doi.org/10.1109/TIT.2007.909108>.
19. VAIDYANATHAN P.P., PAL P. (2011), Sparse sensing with co-prime samplers and arrays, [in:] *IEEE Transactions on Signal Processing*, **59**(2): 573–586, <https://doi.org/10.1109/TSP.2010.2089682>.
20. WANG J. et al. (2022), Improved adaptive beamforming algorithm based on compressed sensing [in Chinese], *China Academic Journal Electronic Publishing House*, **38**(5): 18–21+45, <https://doi.org/10.16328/j.htdz8511.2022.05.012>.
21. YANG Z.X., SHEN Q., LIU W., ELDAR Y.C., CUI W. (2023), High-order cumulants based sparse array design via fractal geometries – Part I: Structures and DOFs, [in:] *IEEE Transactions on Signal Processing*, **71**: 327–342, <https://doi.org/10.1109/TSP.2023.3244672>.
22. ZHANG X.F. et al. (2013), *Array Signal Processing and MATLAB Implementation* [in Chinese], 2nd ed., Publishing House of Electronics Industry, Beijing.
23. ZHAO T., NEHORAI A. (2014), Sparse direction of arrival estimation using co-prime arrays with off-grid targets, [in:] *IEEE Signal Processing Letters*, **21**(1): 26–29, <https://doi.org/10.1109/LSP.2013.2289740>.

Review Paper

A Review of the Sonication-Assisted Exfoliation Methods for MoX_2 (X: S, Se, Te) Using Water and EthanolSihan WANG⁽¹⁾, Yanshu YU⁽²⁾, Jianling MENG^{(1)*}*College of Mathematics and Physics, Beijing University of Chemical Technology*
Beijing, China; e-mails: wsh20010@163.com; 674991037@qq.com*Corresponding Author e-mail: mengjianling@buct.edu.cn

(received October 10, 2024; accepted December 18, 2024; published online March 7, 2025)

Two-dimensional transition metal dichalcogenides (MoX_2 , where $X = \text{S, Se, Te}$), have been the research hotspot over the past decade. The sonication-assisted liquid-phase exfoliation method is suitable for the mass production of MoX_2 in practical applications. Water and ethanol, rather than organic solvents, are increasingly chosen for liquid-phase exfoliation method due to their non-toxic, environmentally friendly properties. However, a systematic review of the method for MoX_2 preparation using water and ethanol is lacking. In this paper, recently published work on the sonication-assisted exfoliation method for MoX_2 preparation using water and ethanol is summarized. Three key parameters are focused on: solvents selection, sonication power, and sonication time. Finally, the application of MoX_2 flakes and the future outlook of the sonication-assisted liquid-phase exfoliation method using water and ethanol are presented. The review aims to provide guidance on exfoliating MoX_2 using the sonication-assisted exfoliation method with water and ethanol.

Keywords: two-dimensional transition metal dichalcogenides; sonication-assisted liquid-phase exfoliation; water; ethanol.



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

1. Introduction

Since graphene was discovered by [NOVOSELOV *et al.* \(2004\)](#), two-dimensional (2D) materials have become a hotspot in materials research. So far, various 2D materials have been studied, including the insulator boron nitride (BN), semiconductor transitional metal dichalcogenides (TMDs), magnetic materials such as CrX_3 (I, Br), Fe_3GeTe_2 , and the topological insulator Bi_2Se_3 , among others. TMDs MX_2 (where $M = \text{Mo, W}$, and $X = \text{S, Se}$) shows potential in many applications due to its excellent physical properties. For example, MoS_2 -based field-effect transistors (FETs) exhibit a $\sim 10^8$ on/off ratio with mobility $\sim 200 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ ([RADISAVLJEVIC *et al.*, 2011](#)). Monolayer MoS_2 also demonstrates a strong photoluminescence effect due to the indirect-to-direct bandgap transition (from 1.9 eV to 2.2 eV) from bulk to monolayer ([MAK *et al.*, 2010](#)). Actually, monolayer MoS_2 exists in three distinct phases: the semiconductive 2H phase, the

metallic 1T phase, and the 1T' phase. The 2H- MoS_2 phase holds potential development for applications in valleytronics. On one hand, 2H- MoS_2 displays a novel valley degree of freedom due to broken inversion symmetry. On the other hand, its valley and spin degrees of freedom are coupled due to spin-orbit splitting ([XIAO *et al.*, 2012](#)). MoSe_2 is more conductive with respect to MoS_2 due to selenium (Se) atoms being more conductive than molybdenum (Mo) atoms. The bandgaps of monolayer MoSe_2 and MoTe_2 are 1.55 eV, 1.1 eV, respectively, extending the spectral range of TMDs from the visible to near-infrared region ([WU *et al.*, 2020](#)).

The mass production of TMDs is required for their practical applications. Currently, there are two categories of TMDs synthesis methods: bottom-up and top-down approaches. Chemical vapor deposition (CVD) and metal-organic chemical vapor deposition (MOCVD), both of which are bottom-up approaches, can synthesize wafer-scale monolayer MoX_2 films. Although continuous efforts to produce large-scale wafers

show promise for practical applications (YU *et al.*, 2017; HU *et al.*, 2023; XU *et al.*, 2021), the cost remains high for industrial application at this stage. The mechanical exfoliation approach, a top-down approach, can produce high-quality MoX_2 , which is suitable for advanced fundamental research. However, its efficiency is relatively low. Another top-down approach is the liquid-phase exfoliation (LPE). LPE methods include shear exfoliation, ultrasonication exfoliation, and microfluidization exfoliation (SETHULEKSHMI *et al.*, 2024). Sonication-assisted liquid exfoliation is the most common technique for MoX_2 synthesis. For example, COLEMAN *et al.* (2011) demonstrated the feasibility of ultrasonic-assisted liquid-phase exfoliation as early as 2011. Their findings were further supported and expanded upon in subsequent studies (KHAN *et al.*, 2011; 2012; O'NEILL *et al.*, 2011; BARWICH *et al.*, 2013; COLEMAN *et al.*, 2013; HANLON *et al.*, 2015; GHOLAMVAND *et al.*, 2016; BACKES *et al.*, 2017; HARVEY *et al.*, 2017; SYNNAUSCHKE *et al.*, 2019; GRIFFIN *et al.*, 2020). Though the sonication-assisted liquid exfoliation process can cause problems such as high defect rate, low stability and impaired electronic properties of the nanosheets, its advantages are: (1) simplicity, universality, and low cost, making it suitable for mass production (AKEREDOLU *et al.*, 2024); (2) mild operating conditions (room temperature and pressure) (AGGARWAL *et al.*, 2024), and the properties of the nanosheets being controllable by adjusting process parameters (SETHULEKSHMI *et al.*, 2024).

The common solvents used in sonication-assisted liquid exfoliation for MoX_2 synthesis are organic polymer, typically N-methyl-2-pyrrolidone (NMP) (O'NEILL *et al.*, 2012). However, the polymer is toxic and hard to remove due to its generally high boiling point. Similarly, although alternative surfactant can exfoliate MoS_2 by expanding the layers, the surfactant molecules are usually difficult to recycle (MA *et al.*, 2018; POZZATI *et al.*, 2024). Recently, significant efforts have been devoted to utilizing green solvents that achieve comparable concentrations and sizes of TMDs dispersion as NMP and surfactant-based solvents. It has been demonstrated that phyto-extracted green solvents facilitate the production of few layer MoS_2

which enhances the photo-conversion efficiency of dye-sensitized solar cells and exhibits an excellent redox activity with high specific capacitance (KUMAR *et al.*, 2023). Polarclean, Iris and Cyrene have been reported as the most promising green solvents for the production of graphene, MoS_2 and WS_2 . In particular, Polarclean has been highlighted due to its low defect density (OCCHIUZZI *et al.*, 2023). RAFI *et al.* (2024) produced bilayered and trilayered MoS_2 nanosheets by employing isopropyl alcohol and deionized water in a 7:3 ratio as a cosolvent. Green solvents biomaterials are beyond the scope of this review, with relevant work summarized in other reviews (SETHULEKSHMI *et al.*, 2024). Among organic solvents, ethanol is considered more environmentally favorable based on environmental impact, health, and safety (EHS) statements (CAPELLO *et al.*, 2007; SHELDON *et al.*, 2019). Hence, our review focuses on water and ethanol solvents.

To our knowledge, no review has been reported on the sonication-assisted liquid exfoliation of MoX_2 employing water and/or alcohol, specifically in terms of factors of process, although reviews on water-mediated exfoliation of MoS_2 have been reported (AGGARWAL *et al.*, 2024).

Our review summarizes the sonication-assisted exfoliation formulation using water and/or ethanol from the following three aspects: solvents selection, sonication power, and sonication time, aiming to provide a guidance on exfoliating TMDs using water and/or ethanol via the sonication-assisted exfoliation method.

2. Sonication-assisted exfoliation recipe

A typical sonication-assisted exfoliation process is as follows (Fig. 1). Firstly, MoX_2 powder is mixed with appropriate solvents. Then, the mixture is ultrasonicated in ultrasonic instrument. Various techniques are used to prevent excessive temperature rise during sonication. For instance, intermittent ultrasound, for example, 40 seconds ultrasonic time followed by 20-second break time, are utilized. Additionally, an ice bath or water-cooling temperature control system is used to maintain a constant temperature. The resulting dispersion subsequently is centrifuged, and the

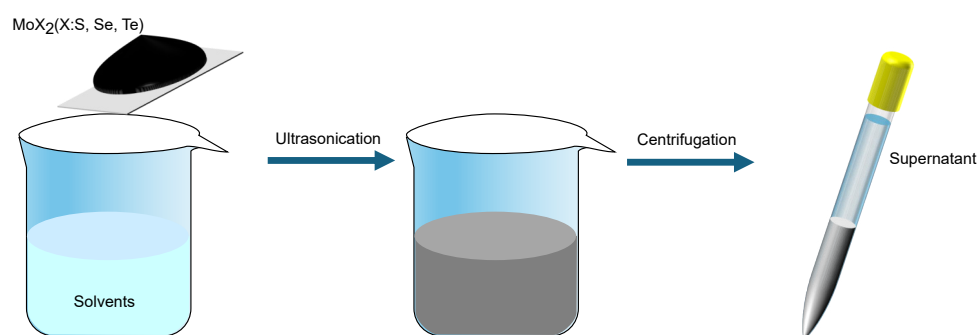


Fig. 1. Schematic diagram of the sonication-assisted exfoliation process.

supernatant is extracted. The speed and number of centrifugation steps help to roughly separate the MoX₂ flakes by size.

The mechanism underlying the sonication-assisted exfoliation process has been reported (GUPTA *et al.*, 2016; COLEMAN *et al.*, 2011; SMITH *et al.*, 2011; NICOLSI *et al.*, 2013; LI *et al.*, 2020). The materials discussed in detail are graphite, and the LPE solvents used are various organic solvents, such as IPA/H₂O mixture, sodium dodecylbenzenesulfonate (SDBS)/H₂O mixture and NMP (LI *et al.*, 2020). Actually, LPE involves two simultaneous structural modifications: exfoliation (reduction in thickness), and fragmentation (reduction in lateral dimension). The research explains exfoliation and fragmentation processes in detail. It was found that fragmentation and exfoliation take place during LPE in three distinct stages, with the kink-band-induced peeling process being one key stages (shown in Fig. 2). In the first stage, graphite flake rupture along existing defects, and kink bands are formed due to surface acoustic waves. The second stage involves the kink bands leading to increase in chemical activity, which promotes fragmentation and exfoliation, leading to the peeling off thin graphite stripes. Then, the last stage involves the peeled graphite strips being exfoliated into thin flakes, with a minimum of ~30 layers. Although the research did not discuss exfoliation of MoX₂ by LPE using water and/or ethanol, the mechanism is also applicable to the exfoliation of MoX₂ by LPE using organic solvents, as both materials possess analogous 2D layered structures.

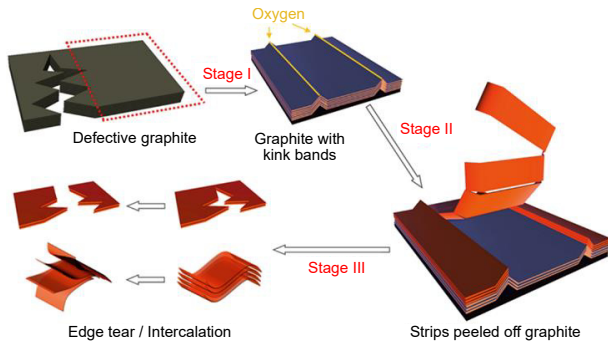


Fig. 2. Schematic diagram of the sonication-assisted LPE mechanism of graphite (reprinted with permission from (LI *et al.*, 2020)).

In general, the abovementioned procedures all have influence on the concentration and size of the exfoliated MoX₂ flakes. Here, we focus on three main influencing factors, including solvents selection, sonication power and sonication time, while other aspects are beyond the scope of this review.

2.1. Solvents selection

There are several theories for screening solvents, such as Hansen solubility parameters (HSP), Young's equation, and Shen's method for probing and matching surface tension components (MA *et al.*, 2020). The HSP theory is commonly used, with the HSP distance R_a employed to evaluate the level of dissolution process between solvents and solutes, as described by the following equation:

$$R_a = [4(\delta_{D,\text{solv}} - \delta_{D,\text{solu}})^2 + (\delta_{P,\text{solv}} - \delta_{P,\text{solu}})^2 + (\delta_{H,\text{solv}} - \delta_{H,\text{solu}})^2], \quad (1)$$

where δ_D , δ_P , δ_H represent dispersive, polar, and hydrogen-bonding solubility parameters of a solvent and solutes, respectively (ZHOU *et al.*, 2011). The reference HSP parameters of MoX₂, H₂O, and ethanol are shown in Table 1.

In general, pure water is a poor solvent for MoS₂ exfoliation. However, MA *et al.* (2018; 2020) demonstrated the feasibility of exfoliating MoS₂ using water. The authors concluded that the stability of MoS₂ in an aqueous solution is due to the fragmentation of the MoS₂ flakes induced by sonication. Compared to graphite, MoS₂ is easier to fragment. The obtained MoS₂ nanosheets have sizes ranging from 100 nm to 400 nm with a few layers (5–6 layers) or multilayers (15–20 layers) in thickness. Mesoporous sheets were also observed (shown in Fig. 3a). LI *et al.* (2015) reported that MoS₂ can be exfoliated in pure water due to defects and enlarged interlayer spacing induced by the fabrication process. ZHAO *et al.* (2016) exfoliated commercial MoS₂ in water using a specially designed sonication instrument with a stirring function. The slipping exfoliation was achieved by the tilted rotation of MoS₂ sheets during stirring. FORSBERG *et al.* (2016) exfoliated MoS₂ in water using a two-step method. First, an orbital sander was used for mechanical ex-

Table 1. HSP value for MoX₂, H₂O, and ethanol.

	δ_D (0.5 MPa)	δ_P (0.5 MPa)	δ_H (0.5 MPa)
MoS ₂ (ZHOU <i>et al.</i> , 2011)	17–19	6–12	4.5–8.5
MoSe ₂ (MAO <i>et al.</i> , 2018)	15.3–18.4	9–18	3.3–11.3
MoTe ₂ (CUNNINGHAM <i>et al.</i> , 2012)	17.8	8	6.5
H ₂ O (ZHOU <i>et al.</i> , 2011)	15.8	8.8	19.4
Ethanol (ZHOU <i>et al.</i> , 2011)	18.1	17.1	16.9

*The HSP parameters are obtained from (ZHOU *et al.*, 2011). Republished with permission from Angewandte Chemie International Edition, permission conveyed through Copyright Clearance Center, Inc.

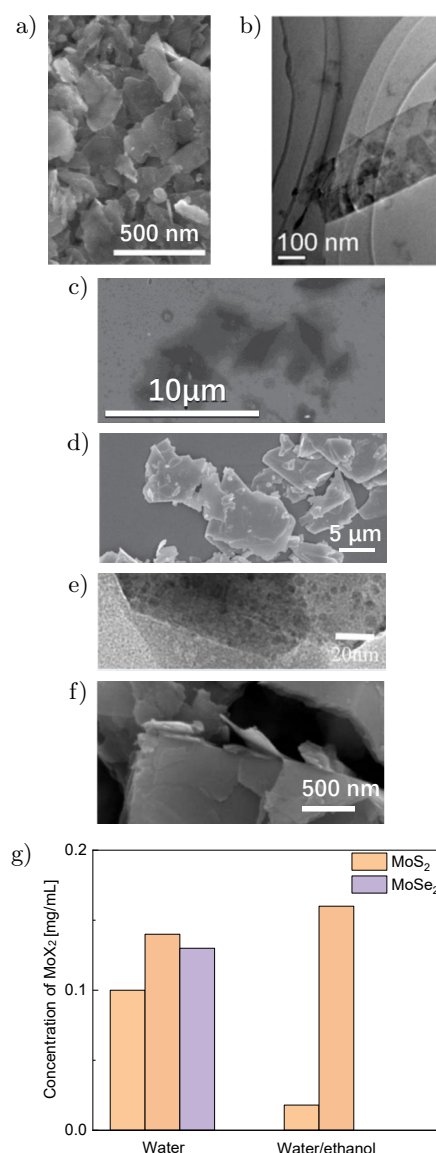


Fig. 3. Scanning electron microscope (SEM) images of exfoliated MoS₂ flake (a–f) except for (e), which is the transmission electron microscope (TEM) image of exfoliated MoS₂ with various vol% ethanol/water solvents. The concentration of MoX₂ using water and water/ethanol solvents is shown in (g).

(a) Reprinted from (MA *et al.*, 2018) with permission from Elsevier; (b) republished with permission from *Angewandte Chemie International Edition* from ZHOU *et al.* (2011), permission conveyed through Copyright Clearance Center, Inc; (c) republished with permission from *IEEE Transactions on Nanotechnology* from YUAN *et al.* (2021), permission conveyed through Copyright Clearance Center, Inc; (d) reprinted from WANG *et al.* (2013) with permission from Elsevier; (e) republished with permission from *Journal of Materials Science: Materials in Electronics* from YANG *et al.* (2017), permission conveyed through Copyright Clearance Center, Inc; (f) adapted with permission from TAGHAVI and AFZALZADEH (2021), Creative Commons License CC BY-SA 4.0.

foliation of MoS₂. Then, the obtained MoS₂ powder was exfoliated in water by sonication. MoS₂ was also exfoliated in water via sonication under an Ar/H₂ at-

mosphere (GUTIÉRREZ, HENGLEIN, 1989). LIU *et al.* (2018a) found that bulk MoSe₂ can be directly exfoliated in hot water at 50 °C, achieving intense exfoliation kinetics while maintaining high quality. Based on simulation at atomic and molecular scales, it was proposed that the stable dispersion of MoSe₂ nanosheets in water is achieved owing to the presence of platelet surface charges originating from edge functionalization and intrinsic polarity. A large number of atomically thin MoSe₂ layers are produced by 100 W sonication for 24 h and 8000 rpm centrifugation for 40 min. The lateral dimensions of the obtained MoSe₂ nanosheet range from 50 nm to 500 nm. A large proportion (>70 %) of these layers are less than 2.0 nm thick, and >40 % of them are thinner than 1.0 nm, corresponding to monolayers. Other studies have reported that atomically thin MoSe₂ platelets can be exfoliated from bulk MoSe₂ by 20 W sonication for 60 h and dispersed in pure water by centrifugation for 30 min under temperature control (KIM *et al.*, 2015). The exfoliated flakes have dimensions of 200 nm to 300 nm with 2–3 layers. The concentration is shown in Fig. 3g, and lateral size and number of layers of exfoliated MoX₂ using water as the solvent are shown in Fig. 4, with data coming from the above-mentioned studies.

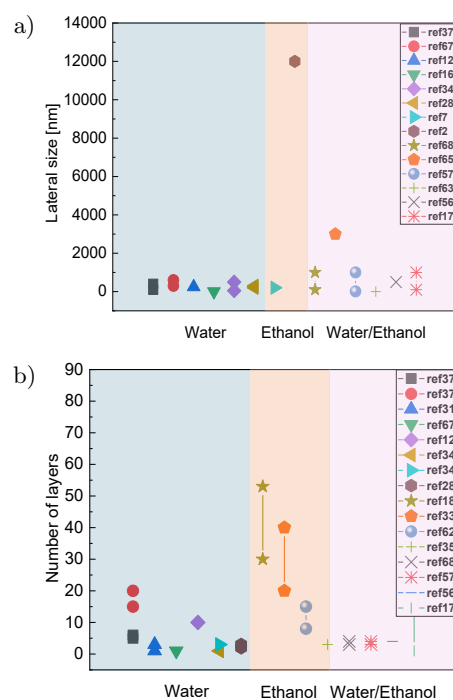


Fig. 4. MoX₂ using water, ethanol and water/ethanol solvents: a) lateral size; b) number of layers.

Anhydrous ethanol is used as the initial solvent for exfoliating MoSe₂ and MoTe₂. MoSe₂ nanosheets were obtained in anhydrous ethanol through the ultrasonic-assisted LPE method and were subsequently used as a gas sensor as the ethanol solvent evaporates (CHEN *et al.*, 2019). Absolute alcohol has also been reported

to be used in the preparation of MoTe₂ nanoflakes (HAN *et al.*, 2023; LIANG *et al.*, 2020). Exfoliation was obtained by sonication for 20 h followed by centrifugation. At the end of the process, the layer thickness of the stripped nanosheet ranged from 30 to 53 atomic layers (HAN *et al.*, 2023). Similarly, YAN *et al.* (2018) mixed bulk MoTe₂ powders with anhydrous alcohol and sonicated powders for 12 h. After centrifugation and additional processing, the number of layers of the resulting MoTe₂ ranged from 8 to 15. AHMAD *et al.* (2021) reported MoTe₂ nanosheets with a lateral size of about 12 μm by sonicating the mixture of MoTe₂ powder and absolute ethanol for 16 h. MoTe₂ nanosheets were also prepared with an ethanol-assisted ultrasound-assisted liquid-phase exfoliation (UALPE) method at 20 °C (LIU *et al.*, 2018b). The concentration is shown in Fig. 3g, and lateral size and number of layers of exfoliated MoX₂ using ethanol solvent are shown in Fig. 4.

Besides single-component solvents, the HSP theory can also be used for solvent mixtures. The HSP parameters of a mixture are a linear combination of the corresponding parameters of each component, as follows:

$$\delta_{\text{blend}} = \sum \phi_{n,\text{comp}} \delta_{n,\text{comp}}, \quad (2)$$

where δ_{blend} , $\phi_{n,\text{comp}}$, $\delta_{n,\text{comp}}$ represent the HSP parameters of the blend, the volume fraction of each component, and the HSP parameters of each component, respectively (ZHOU *et al.*, 2011). By choosing water and/or alcohol with the appropriate composition, a high dispersion concentration of MoS₂ can be achieved. Experimental results and theoretical predictions are consistent in showing that a 45 vol% ethanol/water provides the highest dispersion concentration, with a value of $0.018 \pm 0.003 \text{ mg/mL}$, which is approximately 13 times higher than that in pure ethanol and 68 times higher than that in pure water (ZHOU *et al.*, 2011). The size of the sheets varies from 100 nm to several micrometers and their thickness is 3–4 layers (shown in Fig. 3b). Other works also report using a 45 vol% alcohol/water mixture as the solvent for MoS₂ flake preparation. YUAN *et al.* (2021) prepared MoS₂ nanosheets by liquid phase exfoliation (LPE) for a formic acid gas sensor using a 45 vol% alcohol/water mixture as the solvent. The size of the nanosheets obtained is about 3 μm (shown in Fig. 3c). WANG *et al.* (2013) also prepared MoS₂ nanosheets by dispersing MoS₂ powder in a 45 vol% ethanol/water mixture, with a thickness of 3–4 layers and the size of nanosheets ranging from tens of nanometers to several micrometers (shown in Fig. 3d). HUANG *et al.* (2024) prepared few-layer MoS₂ flakes with an average thickness of 7 nm by sonication in 45 vol% ethanol/water mixture at 240 W for 90 min. In addition to the 45 vol% ethanol/water mixture, other proportions of ethanol/water have also been reported for liquid-phase exfoliation of MoS₂ flakes.

For example, MoS₂ quantum dots were obtained by dispersing defected MoS₂ nanosheets into a 25 vol% ethanol/water solution. Due to the inherent defects in MoS₂, the average lateral size of acquired MoS₂ quantum dots is 3.6 nm – shown in Fig. 3e (YANG *et al.*, 2017). A 23 vol% ethanol/deionized water solution was also reported to be utilized to exfoliate MoS₂, yielding flakes with an average of 4 layers and a lateral size of 500 nm – shown in Fig. 3f (TAGHAVI *et al.*, 2021). Furthermore, a 50 vol% ethanol/water solvent mixture has been reported to exfoliate MoS₂, resulting in nanosheets with lateral sizes of several micrometers and damaged surface edges (PRABUKUMAR *et al.*, 2018; JIN *et al.*, 2020). Compared to the same proportion of NMP/water, this exfoliation efficiency is poor (PRABUKUMAR *et al.*, 2018). HALIM *et al.* (2013) used Young's equation to determine the liquid-solid interfacial energy and predicted that the optimal cosolvent of alcohol-water mixtures should have a surface tension between 30 mJ/m² and 35 mJ/m². The concentration lateral size and number of layers of exfoliated MoX₂ using water/ethanol solvent are shown in Figs. 3g and 4, respectively.

2.2. Sonication power

In addition to the solvent type, which affects the quality of the final MoX₂ production in ultrasound-assisted liquid exfoliation, the ultrasonic power also plays a crucial role. Sonication power is an important parameter influencing the exfoliation process. The size of exfoliated MoS₂ flakes increases as the sonication power increases from 38.5 W, 47 W, to 65.5 W with bath sonication. At 84 W, the MoS₂ flakes begin to agglomerate (TAGHAVI, AFZALZADEH, 2021). This phenomenon can be explained by the collapse of the high-energy bubbles, which increases the size and number of bubbles. As a result, the shock waves produced by sonication are reduced while the bubble implosions increase. Unlike bath sonication, probe sonication uses an ultrasound probe to transmit vibrations. HAU *et al.* (2021) synthesized MoS₂ for 8 h via probe sonication at 420 W using a water/ethanol with a volume ratio of 2:1.

Some divergence exists between sonication-assisted LPE using water and/or ethanol and organic solvents. For instance, using a mixture of chloroform and acetonitrile in a 65:35 ratio as solvents for LPE, the average size of MoS₂ nanosheets decreases as the ultrasonic power increases from 350 W, 450 W to 550 W. Meanwhile, the concentration of produced MoS₂ increases correspondingly. This phenomenon is explained by the cavitation effect and micro-jet effect induced by ultrasound. The cavitation effect is the primary force for exfoliating layered MoS₂, involving the process of the formation, growth, and implosive collapse of bubbles. Simultaneously, the micro-jet effect induced by the collapse of bubbles is the force that fragments the

MoS₂ sheets (ZHANG *et al.*, 2014). The effect of ultrasonic power on exfoliation has also been studied using NMP as a solvent (QIAO *et al.*, 2014). The ultrasonic power was controlled from 100 W, 200 W, 250 W, 285 W, 320 W, 350 W to 400 W. The concentration of nanosheets increased as the sonication power increased, and then decreased after 320 W. Meanwhile, the size of the nanosheets initially decreased and then increased after 320 W. This behavior is associated with the ultrasonic cavitation effect. At low input power, the covalently bonded S-Mo-S sheet are broken into small flakes due to inertial cavitation. However, at high input power, the breaking intensity decreases due to fewer large bubbles being generated, a phenomenon known as ultrasonic cavitation shielding effect.

2.3. Sonication time

Sonication time is another important parameter. TAGHAVI and AFZALZADEH (2021) systematically studied the effect of sonication time on the exfoliation of MoS₂ using a mixture of 77 % deionized (DI) water and 23 vol% ethanol by volume. They found that the size of MoS₂ flakes increases as the effective sonication time increases from 15 min to 60 min, but then decreases with prolonged sonication time. This is due to the agglomeration process. A similar effect has been observed in the LPE of MoS₂ using NMP (O'NEILL *et al.*, 2012). The dimensions of MoS₂ flakes increase after 23 h of sonication, reaching a maximum after 60 h, and then decrease after 60 h of sonication. MITTAL *et al.* (2023) reported that the number of exfoliated MoSe₂ layers in DI water and ethanol decreases as the sonication time increases from 10 min to 60 min. LIU *et al.* (2018a) studied the effect of sonication time on exfoliation of MoSe₂ using water at 50 °C. For comparison, the authors sonicated bulk MoSe₂ for 8 h and 24 h, and found that the layers after 8 h of sonication could not withstand higher centrifugal speed. However, bulk MoSe₂ could be broken into high-quality layers at a longer ultrasound time (24 h) due to more sufficient exfoliation.

XU *et al.* (2024) first compared the effects of different ethanol contents on the dispersibility of MoTe₂, and then analyzed the relationship between sonication time and the thickness of nanosheets at intervals of 0.5 h, 1.5 h, 2.5 h, 3.5 h, 4.5 h, and 5.5 h. The results showed that the average thickness of the nanosheets decreased as the sonication time increased.

3. Application, perspective and conclusions

Exfoliated TMDs using water and/or ethanol solvents enable a wide range of applications, such as electrochemical application as supercapacitor electrodes, photoelectrochemical applications for photocurrent response material (KAJBAFVALA *et al.*, 2018), mechani-

cal reinforcement in polymers (O'NEILL *et al.*, 2012), electrocatalysts for hydrogen evolution reactions, hazardous gas sensor, batteries, surface coatings, and more. Due to its high carrier mobility, strong spin-orbit coupling, and extensive light absorption, MoSe₂ is considered as one of the most promising materials for optoelectronics in TMDs, making it suitable for flexible, lightweight optoelectronic devices (PATEL *et al.*, 2019). MoSe₂ exfoliated by alcohol solvents can also be applied to gas sensors by taking advantage of the volatilization of alcohol (ZHOU *et al.*, 2011). MoTe₂ can be transformed into many types of lasers following specific processing. Additionally, sonication-assisted LPE using water and/or ethanol solvents is an environment-friendly, low-cost, and easy-to-operate method for scaling up mass production of TMD flakes, making it suitable for industrial practical applications (CIESIELSKI, SAMORÌ, 2014). Non-toxic, environmentally friendly solvents and dispersants can extend the range of 2D TMD inks (LEE *et al.*, 2020). From a technological perspective, sonication-assisted LPE using water and ethanol not only facilitates upscaled production of TMDs flakes comparable to organic solvents, but it is also an economical and practical solution, as water and ethanol do not require additional post-processing for environmental compliance.

In this review, the preparation of MoX₂ flakes using the sonication-assisted LPE method with water and/or ethanol was summarized. Although many parameters influence this method, the review focused on three main parameters: solvent selection, sonication power, and sonication time. Solvent selection refers to the ratio of water and/or alcohol used. Related studies were summarized, revealing that a 45 vol% alcohol/water mixture is the optimal solvent for MoS₂, as explained by HSP theory. The effects of sonication power exhibit some inconsistencies, and even some divergences exist between LPE using water and/or ethanol solvents versus organic solvents. This variation may be attributed to differences in sonication equipment used by various research groups. Regarding sonication time, the size of MoS₂ flakes initially increases, and then decreases as sonication time increases. This phenomenon is analogous to that observed when using organic solvents. To further analyze the mechanism behind the LPE method using water and/or ethanol solvents, the advanced LPE mechanism, which includes three stages, is summarized from the literature. Finally, the wide applications of exfoliated MoX₂ flakes and the future outlook for the LPE method using water and ethanol solvents were discussed.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grant no. 12104281. All data generated or analyzed during this study are

included in this published article. The authors declare that they have no relevant financial or non-financial interests to disclose.

References

1. AGGARWAL R., SAINI D., MITRA R., SONKAR S.K., SONKER A.K., WESTMAN G. (2024), From bulk molybdenum disulfide (MoS_2) to suspensions of exfoliated MoS_2 in an aqueous medium and their applications, *Langmuir*, **40**(19): 9855–9872, <https://doi.org/10.1021/acs.langmuir.3c03116>.
2. AHMAD H. *et al.* (2021), Generation of four-wave mixing in molybdenum ditelluride (MoTe_2)-deposited side-polished fibre, *Journal of Modern Optics*, **68**: 425–432, <https://doi.org/10.1080/09500340.2021.1908636>.
3. AKEREDOLU B.J. *et al.* (2024), Improved liquid phase exfoliation technique for the fabrication of MoS_2 /graphene heterostructure-based photodetector, *Heliyon*, **10**(3): e24964, <https://doi.org/10.1016/j.heliyon.2024.e24964>.
4. BACKES C. *et al.* (2017), Guidelines for exfoliation, characterization and processing of layered materials produced by liquid exfoliation, *Chemistry of Materials*, **29**: 243–255, <https://doi.org/10.1021/acs.chemmater.6b03335>.
5. BARWICH S., KHAN U., COLEMAN J.N. (2013), A technique to pretreat graphite which allows the rapid dispersion of defect-free graphene in solvents at high concentration, *The Journal of Physical Chemistry C*, **117**(37): 19212–19218, <https://doi.org/10.1021/jp4047006>.
6. CAPELLO C., FISCHER U., HUNGERBÜHLER K. (2007), What is a green solvent? A comprehensive framework for the environmental assessment of solvents, *Green Chemistry*, **9**: 927–934, <https://doi.org/10.1039/B617536H>.
7. CHEN X. *et al.* (2019), Two-dimensional MoSe_2 nanosheets via liquid-phase exfoliation for high-performance room temperature NO_2 gas sensors, *Nanotechnology*, **30**: 445503, <https://doi.org/10.1088/1361-6528/ab35ec>.
8. CIESIELSKI A., SAMORÌ P. (2014), Graphene via sonication assisted liquid-phase exfoliation, *Chemical Society Reviews*, **43**(1): 381–398, <https://doi.org/10.1039/C3CS60217F>.
9. COLEMAN J.N. (2013), Liquid exfoliation of defect-free graphene, *Accounts of Chemical Research*, **46**(1): 14–22, <https://doi.org/10.1021/ar300009f>.
10. COLEMAN J.N. *et al.* (2011), Two-dimensional nanosheets produced by liquid exfoliation of layered materials, *Science*, **331**(6017): 568–571, <https://doi.org/10.1126/science.1194975>.
11. CUNNINGHAM G. *et al.* (2012), Solvent Exfoliation of transition metal dichalcogenides: Dispersibility of exfoliated nanosheets varies only weakly between compounds, *ACS Nano*, **6**(4): 3468–3480, <https://doi.org/10.1021/nm300503e>.
12. FORSBERG V. *et al.* (2016), Exfoliated MoS_2 in water without additives, *PLOS ONE*, **11**(4): e0154522, <https://doi.org/10.1371/journal.pone.0154522>.
13. GHOLAMVAND Z. *et al.* (2016), Comparison of liquid exfoliated transition metal dichalcogenides reveals MoSe_2 to be the most effective hydrogen evolution catalyst, *Nanoscale*, **8**: 5737–5749, <https://doi.org/10.1039/C5NR08553E>.
14. GRIFFIN A. *et al.* (2020), Effect of surfactant choice and concentration on the dimensions and yield of liquid-phase-exfoliated nanosheets, *Chemistry of Materials*, **32**(7): 2852–2862, <https://doi.org/10.1021/acs.chemmater.9b04684>.
15. GUPTA A., ARUNACHALAM V., VASUDEVAN S. (2016), Liquid-phase exfoliation of MoS_2 nanosheets: The critical role of trace water, *The Journal of Physical Chemistry Letters*, **7**(23): 4884–4890, <https://doi.org/10.1021/acs.jpclett.6b02405>.
16. GUTIÉRREZ M., HENGLEIN A. (1989), Preparation of colloidal semiconductor solutions of MoS_2 and WSe_2 via sonication, *Ultrasonics*, **27**(5): 259–261, [https://doi.org/10.1016/0041-624X\(89\)90066-8](https://doi.org/10.1016/0041-624X(89)90066-8).
17. HALIM U. *et al.* (2013), A rational design of cosolvent exfoliation of layered materials by directly probing liquid-solid interaction, *Nature Communications*, **4**: 2213, <https://doi.org/10.1038/ncomms3213>.
18. HAN C. *et al.* (2023), Theoretical and experimental investigations on sub-nanosecond KTP-OPO pumped by a hybrid Q-switched laser with AOM and MoTe_2 saturable absorber, *Optics & Laser Technology*, **167**: 109760, <https://doi.org/10.1016/j.optlastec.2023.109760>.
19. HANLON D. *et al.* (2015), Liquid exfoliation of solvent-stabilized few-layer black phosphorus for applications beyond electronics, *Nature Communications*, **6**: 8563, <https://doi.org/10.1038/ncomms9563>.
20. HARVEY A. *et al.* (2017), Exploring the versatility of liquid phase exfoliation: producing 2D nanosheets from talcum powder, cat litter and beach sand, *2D Materials*, **4**: 025054, <https://doi.org/10.1088/2053-1583/aa641a>.
21. HAU H.H. *et al.* (2021), Enhanced NO_2 gas-sensing performance at room temperature using exfoliated MoS_2 nanosheets, *Sensors and Actuators A: Physical*, **332**(Part 1): 113137, <https://doi.org/10.1016/j.sna.2021.113137>.
22. HU J., ZHOU F., WANG J., CUI F., QUAN W., ZHANG Y. (2023), Chemical vapor deposition syntheses of wafer-scale 2D transition metal dichalcogenide films toward next-generation integrated circuits related applications, *Advanced Functional Materials*, **33**(40): 2303520, <https://doi.org/10.1002/adfm.202303520>.
23. HUANG C., ZHOU W., GUAN W., YE N. (2024), Molybdenum disulfide nanosheet induced reactive oxygen species for high-efficiency luminol chemiluminescence, *Analytica Chimica Acta*, **1295**: 342324, <https://doi.org/10.1016/j.aca.2024.342324>.

24. JIN X.-F. *et al.* (2020), Inkjet-printed MoS₂/PVP hybrid nanocomposite for enhanced humidity sensing, *Sensors and Actuators A: Physical*, **316**: 112388, <https://doi.org/10.1016/j.sna.2020.112388>.
25. KAJBAFVALA M., FARBOD M. (2018), Effective size selection of MoS₂ nanosheets by a novel liquid cascade centrifugation: Influences of the flakes dimensions on electrochemical and photoelectrochemical applications, *Journal of Colloid and Interface Science*, **527**: 159–171, <https://doi.org/10.1016/j.jcis.2018.05.026>.
26. KHAN U., O'NEILL A., PORWAL H., MAY P., NAWAZ K., COLEMAN J.N. (2012), Size selection of dispersed, exfoliated graphene flakes by controlled centrifugation, *Carbon*, **50**(2): 470–475, <https://doi.org/10.1016/j.carbon.2011.09.001>.
27. KHAN U., PORWAL H., O'NEILL A., NAWAZ K., MAY P., COLEMAN J.N. (2011), Solvent-exfoliated graphene at extremely high concentration, *Langmuir*, **27**(15): 9077–9082, <https://doi.org/10.1021/la201797h>.
28. KIM J. *et al.* (2015), Direct exfoliation and dispersion of two-dimensional materials in pure water via temperature control, *Nature Communications*, **6**: 8294, <https://doi.org/10.1038/ncomms9294>.
29. KUMAR B.A., ELANGOVAN, T., Raju, K., RAMALINGAM G., SAMBASIVAM S., ALAM M.M. (2023), Green solvent exfoliation of few layers 2D-MoS₂ nanosheets for efficient energy harvesting and storage application, *Journal of Energy Storage*, **65**: 107336, <https://doi.org/10.1016/j.est.2023.107336>.
30. LEE H. *et al.* (2020), Zwitterion-assisted transition metal dichalcogenide nanosheets for scalable and biocompatible inkjet printing, *Nano Research*, **13**: 2726–2734, <https://doi.org/10.1007/s12274-020-2916-4>.
31. LI X., WANG W., ZHANG L., JIANG D., ZHENG Y. (2015), Water-exfoliated MoS₂ catalyst with enhanced photoelectrochemical activities, *Catalysis Communications*, **70**: 53–57, <https://doi.org/10.1016/j.catcom.2015.07.024>.
32. LI Z. *et al.* (2020), Mechanisms of liquid-phase exfoliation for the production of graphene, *ACS Nano*, **14**: 10976–10985, <https://doi.org/10.1021/acsnano.0c03916>.
33. LIANG Y. *et al.* (2020), Nano-seconds pulsed Er:Lu₂O₃ laser using molybdenum ditelluride saturable absorber, *Optics & Laser Technology*, **121**: 105791, <https://doi.org/10.1016/j.optlastec.2019.105791>.
34. LIU Y.T., ZHU X.D., XIE X.M. (2018a), Direct exfoliation of high-quality, atomically thin MoSe₂ layers in water, *Advanced Sustainable Systems*, **2**(1): 1700107, <https://doi.org/10.1002/adsu.201700107>.
35. LIU X.J. *et al.* (2018b), Highly Active, durable ultrathin MoTe₂ layers for the electroreduction of CO₂ to CH₄, *Small*, **14**(16): 1704049, <https://doi.org/10.1002/smll.201704049>.
36. MA H. *et al.* (2020), Investigating the exfoliation behavior of MoS₂ and graphite in water: A comparative study, *Applied Surface Science*, **512**: 145588, <https://doi.org/10.1016/j.apsusc.2020.145588>.
37. MA H., SHEN Z., BEN S. (2018), Understanding the exfoliation and dispersion of MoS₂ nanosheets in pure water, *Journal of Colloid and Interface Science*, **517**: 204–212, <https://doi.org/10.1016/j.jcis.2017.11.013>.
38. MAK K.F., LEE C., HONE J., SHAN J., HEINZ T.F. (2010), Atomically thin MoS₂: A new direct-gap semiconductor, *Physical Review Letters*, **105**: 136805, <https://doi.org/10.1103/PhysRevLett.105.136805>.
39. MAO B., GUO D., QIN J., MENG T., WANG X., CAO M. (2018), Solubility-parameter-guided solvent selection to initiate Ostwald ripening for interior space-tunable structures with architecture-dependent electrochemical performance, *Angewandte Chemie International Edition*, **57**(2): 446–450, <https://doi.org/10.1002/anie.201710378>.
40. MITTAL H., RAZA M., KHANUJA M. (2023), Liquid phase exfoliation of MoSe₂: Effect of solvent on morphology, edge confinement, bandgap and number of layers study, *MethodsX*, **11**: 102409, <https://doi.org/10.1016/j.mex.2023.102409>.
41. NICOLASI V., CHHOWALLA M., KANATZIDIS M.G., STRANO M.S., COLEMAN J.N. (2013), Liquid exfoliation of layered materials, *Science*, **340**(6139): 1226419, <https://doi.org/10.1126/science.1226419>.
42. NOVOSELOV K.S. *et al.* (2004), Electric field effect in atomically thin carbon films, *Science*, **306**(5696): 666–669, <https://doi.org/10.1126/science.1102896>.
43. OCCHIUZZI J., POLITANO G.G., D'OLIMPIO G., POLITANO A. (2023), The quest for green solvents for the sustainable production of nanosheets of two-dimensional (2D) materials, a key issue in the roadmap for the ecology transition in the flatland, *Molecules*, **28**(3): 1484, <https://doi.org/10.3390/molecules28031484>.
44. O'NEILL A., KHAN U., COLEMAN J.N. (2012), Preparation of high concentration dispersions of exfoliated MoS₂ with increased flake size, *Chemistry of Materials*, **24**(12): 2414–2421, <https://doi.org/10.1021/cm301515z>.
45. O'NEILL A., KHAN U., NIRMALRAJ P.N., BOLAND J., COLEMAN J.N. (2011), Graphene dispersion and exfoliation in low boiling point solvents, *The Journal of Physical Chemistry C*, **115**(13): 5422–5428, <https://doi.org/10.1021/jp110942e>.
46. PATEL A.B. *et al.* (2019), Electrophoretically deposited MoSe₂/WSe₂ heterojunction from ultrasonically exfoliated nanocrystals for enhanced electrochemical photoresponse, *ACS Applied Materials & Interfaces*, **11**(4): 4093–4102, <https://doi.org/10.1021/acsami.8b18177>.
47. POZZATI M. *et al.* (2024), Systematic investigation on the surfactant-assisted liquid-phase exfoliation of MoS₂ and WS₂ in water for sustainable 2D material inks, *physica status solidi (RRL) – Rapid Research Letters*, **18**(9): 2400039, <https://doi.org/10.1002/pssr.202400039>.
48. PRABUKUMAR C., SADIQ M.M.J., BHAT D.K., BHAT K.U. (2018), Effect of solvent on the morphology of MoS₂ nanosheets prepared by ultrasonication-assisted exfoliation, *AIP Conference Proceedings*, **1943**: 020084, <https://doi.org/10.1063/1.5029660>.

49. QIAO W. *et al.* (2014), Effects of ultrasonic cavitation intensity on the efficient liquid-exfoliation of MoS₂ nanosheets, *RSC Advances*, **4**(92): 50981–50987, <https://doi.org/10.1039/C4RA09001B>.
50. RADISAVLJEVIC B., RADENOVIC A., BRIVIO J., GIACOMETTI V., KIS A. (2011), Single-layer MoS₂ transistors, *Nature Nanotechnology*, **6**: 147–150, <https://doi.org/10.1038/nnano.2010.279>.
51. RAFI R., RAHULAN K.M., FLOWER N.A.L., ABITH M., CHIDAMBARAM S.G.T., RAJENDRAN A.S. (2024), Optical limiting performance of MoS₂ nanosheets exfoliated via liquid-phase sonication: Implications for laser shielding, *ACS Applied Nano Materials*, **7**(10): 11097–11106, <https://doi.org/10.1021/acsanm.3c06096>.
52. SETHULEKSHMI A.S., JACOB F.P., JOSEPH K., APREM A.S., SISUPAL S.B., SARITHA A. (2024), Biomaterials assisted 2D materials exfoliation: Reinforcing agents for polymer matrices, *European Polymer Journal*, **210**: 112943, <https://doi.org/10.1016/j.eurpolymj.2024.112943>.
53. SHELDON R.A. (2019), The greening of solvents: Towards sustainable organic synthesis, *Current Opinion in Green and Sustainable Chemistry*, **18**: 13–19, <https://doi.org/10.1016/j.cogsc.2018.11.006>.
54. SMITH R.J. *et al.* (2011), Large-scale exfoliation of inorganic layered compounds in aqueous surfactant solutions, *Advanced Materials*, **23**(34): 3944–3948, <https://doi.org/10.1002/adma.201102584>.
55. SYNNAITSCHKE K. *et al.* (2019), Length- and thickness-dependent optical response of liquid-exfoliated transition metal dichalcogenides, *Chemistry of Materials*, **31**(24): 10049–10062, <https://doi.org/10.1021/acs.chemmater.9b02905>.
56. TAGHAVI N.S., AFZALZADEH R. (2021), The effect of sonication parameters on the thickness of the produced MoS₂ nano-flakes, *Archives of Acoustics*, **46**: 31–40, <https://doi.org/10.24425/aoa.2021.136558>.
57. WANG G.-X., BAO W.-J., WANG J., LU Q.-Q., XIA X.-H. (2013), Immobilization and catalytic activity of horseradish peroxidase on molybdenum disulfide nanosheets modified electrode, *Electrochemistry Communications*, **35**: 146–148, <https://doi.org/10.1016/j.elecom.2013.08.021>.
58. WU X., WANG Y.-h., LI P.-l., XIONG Z.-z. (2020), Research status of MoSe₂ and its composites: A review, *Superlattices and Microstructures*, **139**: 106388, <https://doi.org/10.1016/j.spmi.2020.106388>.
59. XIAO D., LIU G.-B., FENG W., XU X., YAO W. (2012), Coupled spin and valley physics in monolayers of MoS₂ and other group-VI dichalcogenides, *Physical Review Letters*, **108**: 196802, <https://doi.org/10.1103/PhysRevLett.108.196802>.
60. XU X. *et al.* (2021), Seeded 2D epitaxy of large-area single-crystal films of the van der Waals semiconductor 2H MoTe₂, *Science*, **372**(6538): 195–200, <https://doi.org/10.1126/science.abf5825>.
61. XU X. *et al.* (2024), High-Performance flexible broadband photoelectrochemical photodetector based on molybdenum telluride, *Small*, **20**(27): 2308590, <https://doi.org/10.1002/smll.202308590>.
62. YAN Z.Y. *et al.* (2018), MoTe₂ saturable absorber for passively Q-switched Ho:Pr:LiLuF₄ laser at 3 μm, *Optics & Laser Technology*, **100**: 261–264, <https://doi.org/10.1016/j.optlastec.2017.10.012>.
63. YANG Q., HE Y., FAN Y., LI F., CHEN X. (2017), Exfoliation of the defect-rich MoS₂ nanosheets to obtain nanodots modified MoS₂ thin nanosheets for electrocatalytic hydrogen evolution, *Journal of Materials Science: Materials in Electronics*, **28**: 7413–7418, <https://doi.org/10.1007/s10854-017-6430-8>.
64. YU H. *et al.* (2017), Wafer-scale growth and transfer of highly-oriented monolayer MoS₂ continuous films, *ACS Nano*, **11**: 12001–12007, <https://doi.org/10.1021/acsnano.7b03819>.
65. YUAN Z., YANG C., GAO H., QIN W., MENG F. (2021), High response formic acid gas sensor based on MoS₂ nanosheets, *IEEE Transactions on Nanotechnology*, **20**: 177–184, <https://doi.org/10.1109/TNANO.2021.3059622>.
66. ZHANG S.-L., CHOI H.-H., YUE H.-Y., YANG W.-C. (2014), Controlled exfoliation of molybdenum disulfide for developing thin film humidity sensor, *Current Applied Physics*, **14**(3): 264–268, <https://doi.org/10.1016/j.cap.2013.11.031>.
67. ZHAO G., WU Y., SHAO Y., HAO X. (2016), Large-quantity and continuous preparation of two-dimensional nanosheets, *Nanoscale*, **8**(10): 5407–5411, <https://doi.org/10.1039/C5NR07950K>.
68. ZHOU K.-G., MAO N.-N., WANG H.-X., PENG Y., ZHANG H.-L. (2011), A mixed-solvent strategy for efficient exfoliation of inorganic graphene analogues, *Angewandte Chemie International Edition*, **50**(46): 10839–10842, <https://doi.org/10.1002/anie.201105364>.

Technical Note

Experimental and Numerical Investigations of Acoustic Variations in a Classroom Environment

Amit Kumar Singh CHAUHAN^{(1)*}, Ajitanshu VEDRTNAM⁽²⁾, Suryappa Jayappa PAWAR⁽³⁾

⁽¹⁾ *Department of Mechanical Engineering, Motilal Nehru National Institute of Technology Allahabad
Prayagraj, India*

⁽²⁾ *Department of Mechanical Engineering, Invertis University
Bareilly, Uttar Pradesh, India*

⁽³⁾ *Department of Applied Mechanics, Motilal Nehru National Institute of Technology Allahabad
Prayagraj, India*

*Corresponding Author e-mail: amitchauhan.ac@gmail.com

(received October 30, 2024; accepted February 4, 2025; published online March 6, 2025)

The acoustic behaviour of a classroom is vital for an effective teaching-learning process. The present work aims to experimentally determine the acoustic performance of a typical classroom. The full-scale experiment was conducted at the Seminar Hall, the Department of Applied Mechanics, MNNIT Allahabad, Prayagraj, using a method with limited resource requirements. The Seminar Hall was divided into four planes by threads, and the sound pressure level (SPL) was measured at 30 coordinates in each plane for the specified sound source location. Data were collected from three different sound source locations. The study revealed that the sound source location and frequency significantly influence the sound pressure levels in the classroom, impacting its acoustic performance. The broader implications of interior materials, such as wall material and the position of elements like the teaching board, door, and podium, are highlighted as critical considerations for future classroom acoustic optimization. Furthermore, a numerical model was developed to predict the variation in the SPL with change in the sound source locations and frequencies. The collected data validated with the finite element (FE) model. The verification experiments for the modeling results were performed for each plane. The results of the FE model and experiments were found consistent across all four planes of the seminar hall and the various sound source locations.

Keywords: acoustic measurements; finite element method; room acoustics; sound pressure level; sound source location.



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

1. Introduction

Comfortable acoustic conditions are essential in the workplace, as is intensive verbal communication for improved efficiency. Classrooms are often noisy and reverberant, making learning difficult (MEALINGS, 2023a). Specific classrooms are used for students to convey better acoustics and comfort (RABELO *et al.*, 2014). The acoustic parameters such as the sound pressure level (SPL) and the speech transmission index directly impact the audience's intelligence present in the classroom. Noise decreases the information sent from the source in the classroom (MEALINGS, 2023b; PENG *et al.*, 2016; RABELO *et al.*, 2014). The research com-

munity has conducted various studies to achieve the acoustic comfort of the classroom. PENG *et al.* (2016) investigated the background noise level and speech SPL for the Chinese word recognition test and found that high SPL could not guarantee good Chinese word recognition score for children present in the classroom because of its dependency on the background noise level. VISENTIN *et al.* (2018) used speech intelligibility, response time, and rating scales to analyze the effect of acoustic changes in the room. ZHANG *et al.* (2019) used two classrooms and conducted listening tests at different SPLs. The interaction effect of the sound types and the SPL was found to have practical significance for different noises. GRAMEZ and BOUBENIDER (2017)

measured the ambient noise and interior sound insulation for a conference room compared with the guidelines available in the literature. Poor room acoustics was found due to the low insulation and high reverberation time. MEALINGS *et al.* (2024) measured the acoustic performance of 166 rooms and found that reverberation time and noise level (SPL) are the two significant factors that impact the room's performance.

It is reported that the superior signal-to-noise ratio is significant in addition to reverberation time (BRADLEY, 1986; BRADLEY *et al.*, 1999; 2003; YANG, BRADLEY, 2009). BUDZYŃSKI (1986) mentioned that early reflections coming from sidewalls are responsible for increasing auditory distance localization. Installing sound-insulating material may help, but speech transmission quality could be better and more cost-effective. Increasing sound-absorbing material leads to a lower signal-to-noise ratio and a decreased speech intelligibility, specifically for distant listeners. Interestingly, the acoustic ceiling tiles used for the sound insulation absorb consonant sounds higher than the vowel sound, as vowels have lower frequencies (NÁBĚLEK *et al.*, 1989; NIJS, RYCHTÁRIKOVÁ, 2011). The optimum configurations of absorptive treatment for improved acoustical conditions using computer-based and numerical models were reported in (BISTAFA, BRADLEY, 2000; MIR, ABDU, 2005; REICH, BRADLEY, 1998; SMIRNOWA, OSSOWSKI, 2005). The authors reported the FE model, which effectively predicted the acoustic behaviour of a room in their previous work. The presented model was validated for a rectangular room made of laminated glass (VEDRTNAM, PAWAR, 2018). Many standards are reported in the literature, which provide reference values for the different parameters that may influence acoustic comfort (World Health Organisation, 1999; NEWMAN, SABINE, 1965). The studies on designing and measuring the acoustic properties of interiors, especially for small rooms and primarily SPL (VORLÄNDER, 1998; WEYNA, 1996) problems in estimating the acoustic behaviour of interiors, the effect of source directivity (VIGEANT *et al.*, 2006), and acoustical designing of classrooms (BRADLEY, 1986; BRADLEY *et al.*, 1999; 2003; GRAMEZ, BOUBENIDER, 2017; JERLEHAG *et al.*, 2018; PENG *et al.*, 2016; RABELO *et al.*, 2014; VISENTIN *et al.*, 2018; YANG, BRADLEY, 2009; ZHANG *et al.*, 2019) are already available.

Numerous studies have explored the influence of room geometry, materials, and sound source locations on classroom acoustics. For example, VISENTIN (2023) study explores how background noise, including student interactions, impacts task performance and listening comprehension in classrooms. The research highlights the critical role of signal-to-noise ratio and emphasizes designing acoustic environments that account for real-world noise levels beyond typical reverberation time measurements. HONGISTO *et al.* (2023) compared

two classrooms, one acoustically refurbished with enhanced sound-absorbing materials and reduced reverberation times. The study demonstrated significant reductions in noise annoyance and improved speech intelligibility, particularly during activity-based lessons. This reinforces the importance of targeted interventions in classroom design. VAN REENEN and MANLEY (2023) focused on the implementation of classroom acoustic standards globally. It discusses the effectiveness of mandatory standards accompanied by detailed design guidance in achieving optimal learning environments and identifies cost and accessibility as barriers to adoption.

Several standards for the acoustical property measurements, i.e., ISO 10534-2, ASTM E2611-09, ASTM E1050-98, JIS A1409, ISO 354-2003, ASTM C423, ISO 140-3, SAE J1400, ISO 140-4, and ASTM E90 are also available. The architect's job nowadays should essentially involve meeting the measurable standards set for designing acoustically comfortable living rooms, classrooms, workshops, laboratories, concerning halls, lecture halls, fictional rooms, dining halls, drawing rooms, factories, sports halls, mechanical rooms, hotels, restaurants and every enclosed space of human intervention including sound and noise. The minor changes in frequency, room dimensions, materials, goods, and interiors affect the SPL in the rooms.

Numerous studies have explored the acoustic performance of classrooms, focusing primarily on reverberation time (RT), speech intelligibility indices (STI), clarity (C50), noise reduction coefficients. However, the influence of spatial variability in SPL across different loudspeaker locations in a classroom using controlled frequency tones, such as 4000 Hz (a frequency crucial for speech clarity), has been underexplored. Also, many of these studies rely heavily on generalized assumptions and computational simulations, often needing to integrate detailed experimental validation. In this work, a method to determine SPL variation due to sound source (SS) location, directivity, and objects in the room is proposed. A numerical model is also proposed for predicting SPL variation as a function of SS location, frequency, and object.

2. Materials and methods

Figures 1a and 1b show the photograph and schematic diagram of the seminar hall. The dimensions of the seminar hall were 9.25 m × 7.23 m × 3.14 m. This seminar hall had tiles on the floor, concrete walls, a door, a teacher's desk, a podium made of wood, and a teaching board made of Balsa wood. The dimensions of the board, door, podium, and teacher's desk were 3.6 m × 1.2 m, 1.20 m × 2.05 m, 0.62 m × 0.62 m × 1.20 m, and 3.68 m × 0.62 m × 0.76 m, respectively. An air-conditioner was also mounted on the wall. The speaker was placed in three different positions.

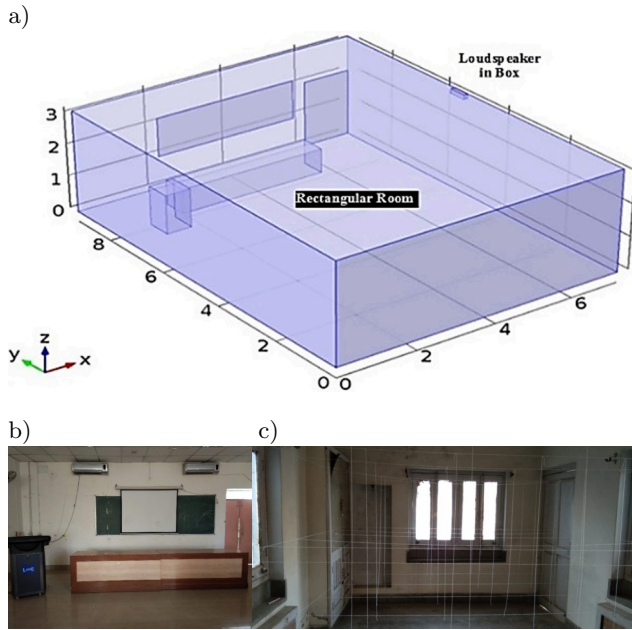


Fig. 1. Schematic diagram (a) and photograph (b) of seminar hall, photograph of the room used for verification experiment (c).

Four different planes (Fig. 2) in the seminar hall were created using the mesh of treads for accuracy and repeatability of a particular location while recording the SPL. The SPL was recorded at 30 points (six along the x -axis and five along the y -axis) in each plane. The coordinates were marked on the threads for the accuracy of the location while noting the SPL. The sound signal was produced using a directivity-controlled SS mounted in a cubic cabinet. The omnidirectional microphones were used. A filling of bonded acetate fibre significantly increased the effective volume of a sealed-box loudspeaker. An amplifier was used to enhance the amplitude of an electrical signal produced by the source. The amplifier was connected in between a sound-generating laptop and the 2-in electrodynamic loudspeaker. The horizontal and vertical input loudspeaker coverage are 50° and 30° , re-

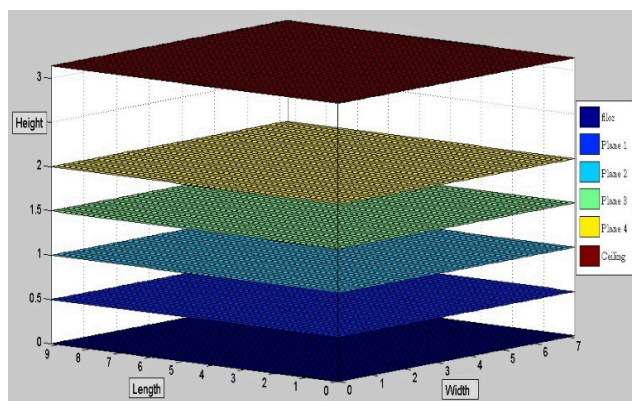


Fig. 2. Position of different planes selected for the work.

spectively. The directivity index of the loudspeaker is 18.9 dB at 2000 Hz.

The speaker sensitivity rating is 85 dB – 1 W – 1 m, i.e., 85 dB sound is produced at 1 m away from the speaker if 1 W input is given. The loudspeaker with a 50 mm driver was mounted on the front, attempting to block the sound backward, utilizing sound-insulating materials. The Indi 6182 Multifunctional Sound Level Meters were used to measure the SPL at different locations in the room. The SPL was measured by the sound level meters in L_{eq} (equivalent continuous sound level) mode. The Laser Distance Meter (Leica DISTOTM X310, Swiss technology by Leica Geosystem) was used for the distance measurement. The pure tone of 4000 Hz (sine wave) was generated following the authors' procedure in their earlier work (VEDRTNAM, PAWAR, 2018).

The typical frequencies under consideration for room acoustics are 125 Hz–4000 Hz, octave bands. Thus, the SPL was measured at 1000 Hz, 2000 Hz, and 3000 Hz at the selected coordinates of different plains for comparison purposes. To systematically analyze the variation of SPL at different frequencies, separate controlled experiments were conducted using pure sine wave signals at 1000 Hz, 2000 Hz, 3000 Hz, and 4000 Hz. The SPL measurements reported for each frequency correspond to independent experimental runs rather than being derived from a single 4000 Hz excitation. This approach ensures accurate assessment of frequency-dependent acoustic behaviour in the classroom environment. Further, the experimentation was repeated in a different room to verify the effect of frequency change on the SPL (Fig. 1c). The FE model was constituted using the acoustics module, pressure acoustics, and frequency domain of COMSOL 5.4. The actual dimensions of the seminar hall and other objects were considered for the geometry model (Fig. 1b). The SS geometry was taken from experimentation for simulation. The meshing was performed using a physics-controlled mesh with the extra fine element size. The full mesh comprises 103 811 domain elements, 6146 boundary elements, and 390 edge elements. The parametric sweep of coordinates for the speaker (similar to the experiment) was performed to compute the speaker's results for three locations. The standard material properties were utilized for the different materials present in the seminar hall (VEDRTNAM, PAWAR, 2018). The SPL of four virtual planes (Fig. 2) at similar locations to experiments were obtained from the FE model.

While acoustic performance is typically assessed using multiple parameters, including RT, STI, and C50, this study focuses specifically on SPL variations. The SPL is a critical factor in classroom acoustics as it directly influences speech intelligibility and sound distribution. By analyzing the SPL across different source locations and frequencies, this study provides valuable

insights into the spatial acoustic behaviour of the classroom. Future work will extend this analysis to incorporate additional acoustic metrics for a more comprehensive assessment. The SPL was measured up to the height of 2 m from the floor since the maximum range of height of humans for listening belongs to this region. The controlled harmonic tone 4000 Hz sine wave frequency was selected as a test signal in the mid to high-frequency range. It plays a significant role in understanding consonants due to its critical importance in speech intelligibility. It provides preciseness and repeatability for evaluating the frequency-dependent SPL distributions without neglecting the confounding effects of other variables, such as mixed-frequency content or background noise.

The SPL at 70 dB refers to the pressure value of 0.063 Pa and intensity of $1 \text{ W/m}^2 \times 10^{-5} \text{ W/m}^2$, and at 80 dB, the SPL refers to the pressure value of 0.2 Pa and intensity of $1 \text{ W/m}^2 \times 10^{-3} \text{ W/m}^2$ (SMIRNOWA, OSSOWSKI, 2005). Sound intensity as a “sound energy quantity” can be related to sound power (acoustic power) as $I \approx p^2$ (for progressive plane waves) (VEDRT-NAM, PAWAR, 2018).

The SPL was measured at 30 coordinates in every plane, and the results were plotted using MATLAB. Table 1 shows the locations of the loudspeakers used in the experiments. These positions were selected to represent different typical loudspeaker placements in a classroom environment. The loudspeakers were placed at varying distances and orientations from key

room features (e.g., the teacher’s desk, podium, and walls) to assess how the sound source location influences the SPL distribution. These positions were not based on any pre-existing loudspeakers in the room but were experimentally chosen to cover a variety of configurations that might be encountered in real-world classroom setups. Thereafter, the results for all three fixed positions of the loudspeaker (Table 1) in each plane are discussed.

Table 1. Location of the loudspeaker and their coordinates.

Loudspeaker location	x	y	z
First fixed position	4	0	3.14
Second fixed position	1.5	3.5	0.76
Third fixed position	5.5	3.5	0

3. Results and discussions

3.1. Measurement of SPL in seminar hall at first fixed position of the loudspeaker – $(x, y, z) = (4 \text{ m}, 0 \text{ m}, 3.14 \text{ m})$ – in different planes

Figure 3 shows the variation of the SPL in plane 1. It is found that the effect of source directivity plays a significant role in the SPL distribution curve (Fig. 3a). The higher SPL values (red colour) were on an axis parallel to the source as plenty of direct sounds reached that axis. The low SPL was measured below the speaker. The minimum SPL was measured behind the podium because sound waves could not reach

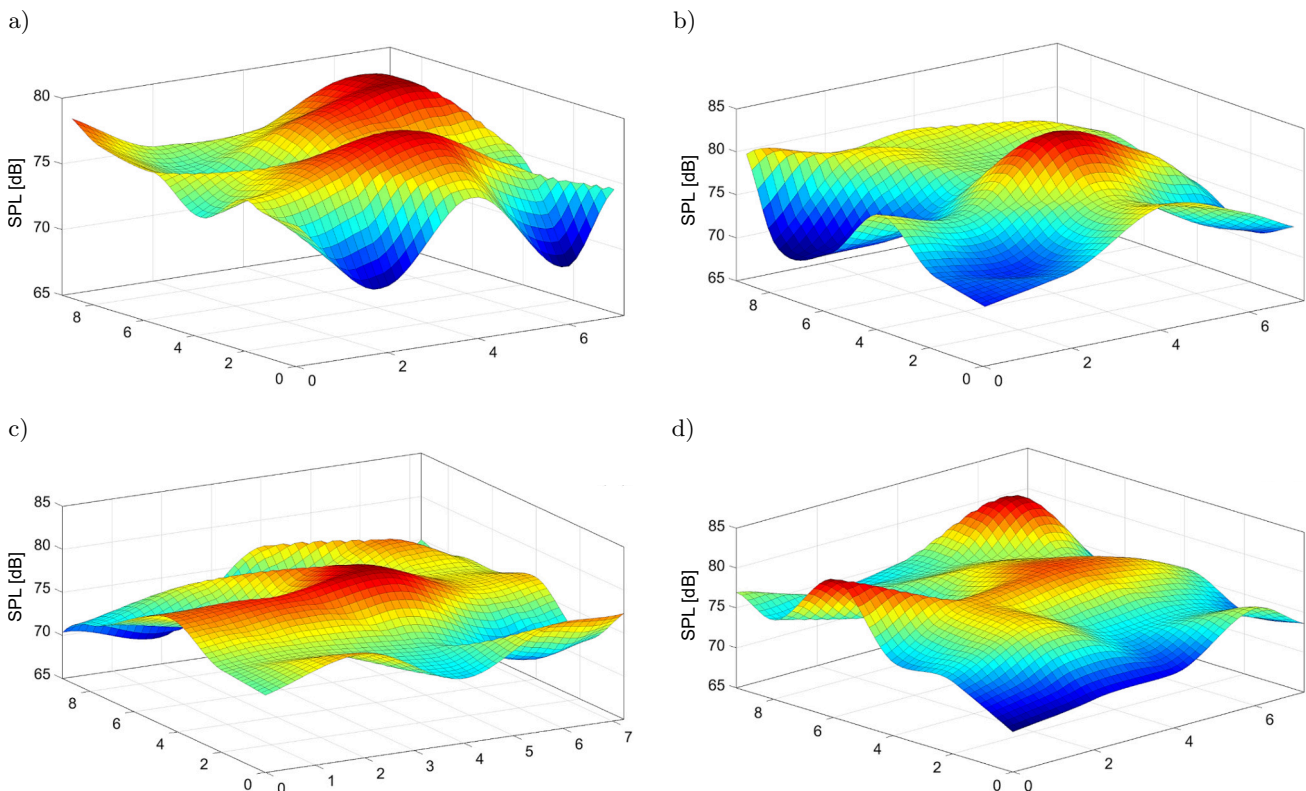


Fig. 3. Variation in SPL at first fixed position of the source in: (a) plane 1; (b) plane 2; (c) plane 3; (d) plane 4.

there directly. The lower reflection and lack of direct sound waves have resulted in the lowest SPL behind the podium. The sound waves coming toward the podium first struck it, then absorbed and partially reflected. The lowest values of the SPL (blue colour) were found beside the teacher's desk because of the lack of reach of direct sound waves.

The desk influences sound wave distribution by reflecting and diffusing the sound waves, with minimal contribution from material absorption. Hence, the SPL values were little higher in front of the teacher's desk. At the front wall, the SPL was measured lower near the air conditioner's presence. Generally, air conditioners are designed with sound-absorbing materials to dampen the sound. The front panels of the air conditioners act as barriers and help reflect and absorb sound waves. However, the SPL suddenly rose at the corners of the front wall because of constructive interference due to the intersection of two walls.

Figure 3b shows the variation of the SPL in plane 2. A similar trend was observed in plane 2.

The lowest SPL value was found on the wall, exactly below the speaker. The SPL was found most stable near the source directivity field (yellow colour). The area near the door (at the origin) had a lower SPL primarily due to the positioning and interaction of the sound waves with the wooden door, rather than significant absorption by the material itself. Additionally, due to the formation of destructive interference, the SPL values were low. The trends observed in Fig. 3c

and 3d were almost similar, with minimum variations because of the absence of obstructions in their planes. The comparison of the SPL for all four planes is shown in Fig. 4.

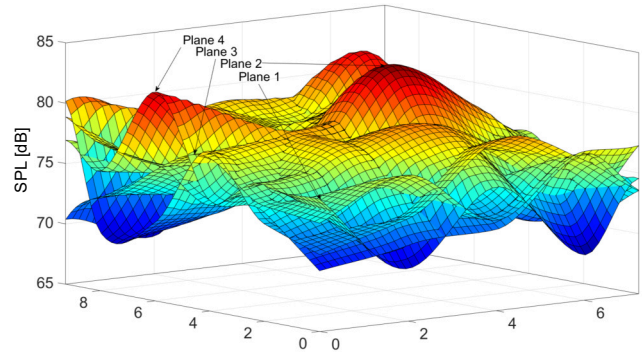
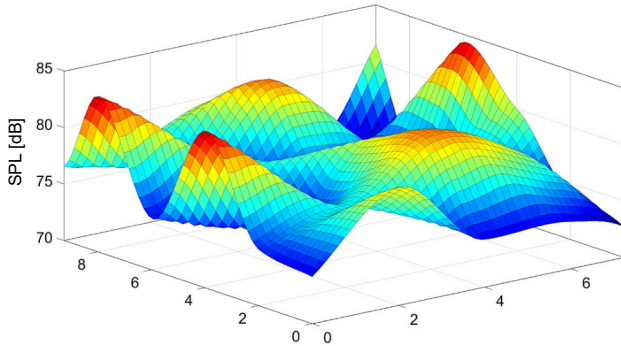


Fig. 4. Comparison of SPL at first fixed position of the source in all planes.

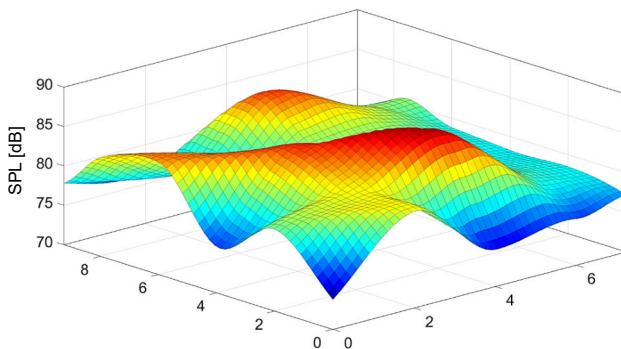
3.2. Measurement of SPL in seminar hall at second fixed position of the loudspeaker – $(x, y, z) = (1.5 \text{ m}, 3.5 \text{ m}, 0.76 \text{ m})$ – in different planes

In the second case, the loudspeaker was placed 0.76 m above the floor, facing the larger space in the opposite direction as the board. The SPL was measured and plotted in a similar manner to the previous. Figure 5a shows the variation in the SPL in plane 1 and the effect of source directivity on the SPL distribution,

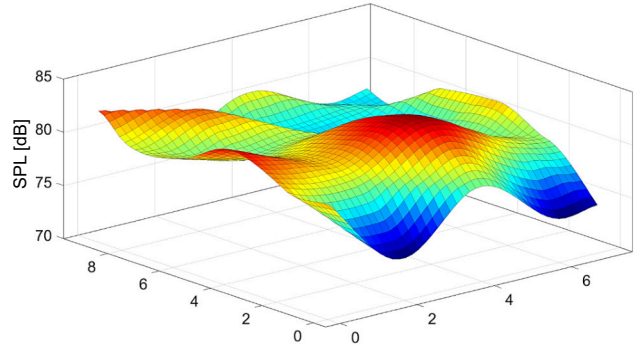
a)



c)



b)



d)

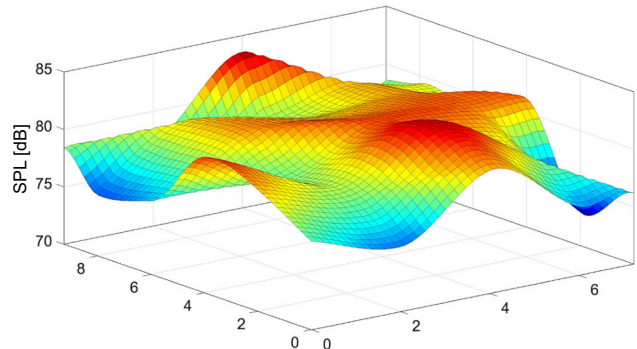


Fig. 5. Variation in SPL at the second fixed position of the source in: (a) plane 1; (b) plane 2; (c) plane 3; (d) plane 4.

which remains highly variable. Higher SPL values (red colour, near 6 m, 7.23 m, 0.5 m) were measured in the far-field region of the lower plane. This far-field region has a significant amount of space without interfering with room interiors, so a lot of direct sound reaches it. The average SPL values were measured on the same wall where the speaker was mounted, because the side walls were closer in this case.

Figure 5b shows the variation of the SPL in plane 2. The highest SPL values (red colour, near 2 m, 4 m, 1 m) were measured near the speaker field. The absorption of sound was maximum at the front wall location (9.25 m, 2 m, 1 m) and near the air conditioners (9.25 m, 6 m, 1 m), but the source's directivity to the receiving place was also maximum. As a result, the SPL in these areas is approximately average. Comparing Figs. 5a and 5b reveal that both curves have higher and lower values at the same locations and follow a nearly identical pattern while only varying in SPL intensity. Figure 5c shows the variation of the SPL in plane 3. The SPL was measured lower near the origin coordinates (0 m, 0 m, 1.5 m) because of the presence of a door, as sound absorption was maximum at that location due to the presence of wood material. The higher and lower points in Figs. 5a–c are almost identical. Figure 5d shows the variation of the SPL in plane 4. Since the sound distribution is more uniform in the presence of more free space, and there is less interruption of interiors, this plane had the fewest variations in the SPL distribution curve compared to all other planes.

The highest SPL value (near 2 m, 4 m, 2 m) is found in the near field region and on-axis to the source. The SPL values at the speaker's backside, as well as the corners of walls near the podium (0 m, 7.23 m, 2 m) were lower due to source directivity and the presence of absorbing materials. The comparison of the SPL for all four planes is shown in Fig. 6. The comparison shows that the plane 4 has the most stable SPL values because of the higher source directivity and least absorptivity. The corners of the room also helped in maintaining the SPL values at the far end by forming constructive interferences.

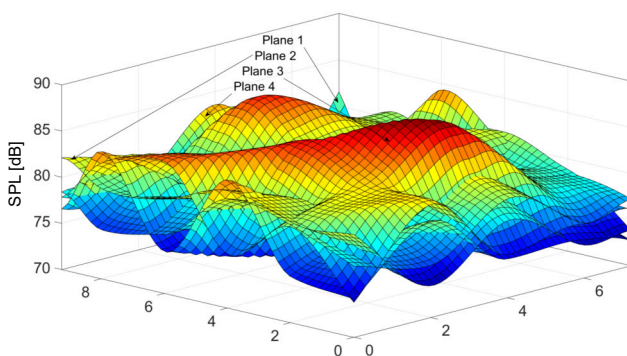


Fig. 6. Comparison of SPL at second fixed position of source in all planes.

3.3. Measurement of SPL in seminar hall at the third fixed position of the loudspeaker – $(x, y, z) = (5.5 \text{ m}, 3.5 \text{ m}, 0 \text{ m})$ – in different planes

In the third case, the loudspeaker was positioned at ground level (near the center of the room), away from the origin, and facing the teacher's desk and board.

Figure 7a shows the SPL variation and the source directivity effect in plane 1. This plane had the most significant variation in SPL values due to speaker location, less free space, and maximum interruption from interiors. In Figs. 7a and 7b, the highest SPL (red colour, near 4 m, 4 m, 0.5 m, and 4 m, 4 m, 1 m, respectively) were measured near the field region, speaker location, and on-axis to the source. In Fig. 7b, the SPL drops abruptly between the podium and the teacher's desk (2 m, 6 m, 1 m) due to the maximum amount of sound-absorbing material surrounding this area. Figures 7c and 7d show the variation of the SPL in planes 3 and 4, respectively. SPL distributions were relatively uniform due to the significant free space and minimal interruption of interiors. The area from the front to the speaker location was measured as the high SPL. Figure 7d shows the variation of the SPL in plane 4, which has a similar distribution to plane 3 with some apparent changes.

Figure 8 shows a comparison of the SPL across all four planes. The SPL behaviour was found most stable compared to the other two loudspeaker locations. The area near the speaker showed the maximum SPL in all four planes, whereas the SPL was found lower at the corner backside of the SS location.

After analyzing all speaker locations, it was found that the first plane had the most variations when compared to the other planes. The most apparent reason is the presence of objects in the room on this plane, such as air conditioners, the teacher's desk, and the podium. Because the material absorption coefficients of these interiors (beyond the scope of this study) can vary, the reverberant field may influence the value of SPL at different coordinates. The third speaker location, in the third and fourth planes, was constantly compared to the other two speaker locations because the speaker was placed in the center of the room, at ground level. As a result, the sound distribution was more uniform than the other speakers' locations.

Figure 9 shows the surface plot of the SPL obtained after solving the FE model using COMSOL for the different planes. However, as it was ambiguous to demonstrate the experimental results with the simulation results using this plot, a few verification experiments were also performed, and line graphs were plotted. The line graphs were plotted along the line parallel to the Y-axis at $X = 4 \text{ m}$ in four different planes as described previously, and results were compared to those obtained from the experiment. For comparison, 20 SPL readings from the investigation were collected

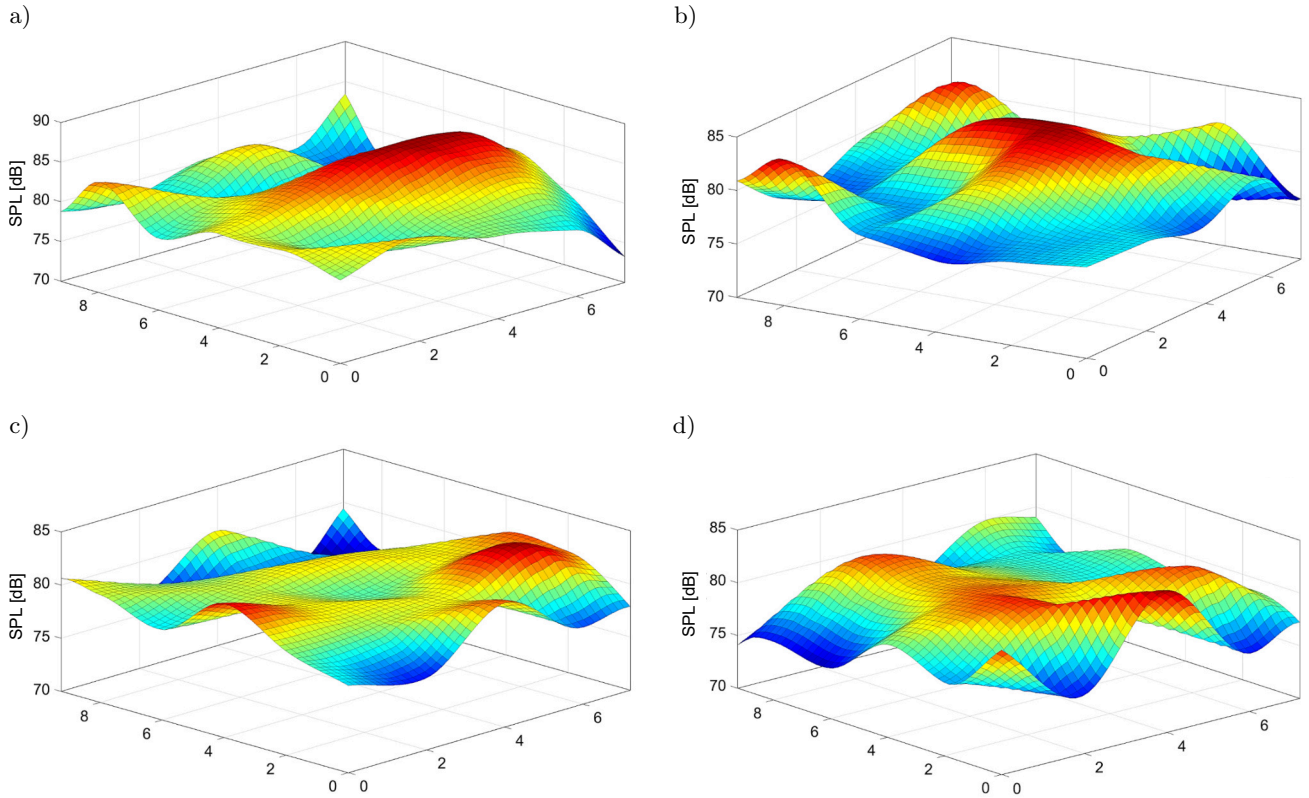


Fig. 7. Variation in SPL at the third fixed position of the source in: (a) plane 1; (b) plane 2; (c) plane 3; (a) plane 4.

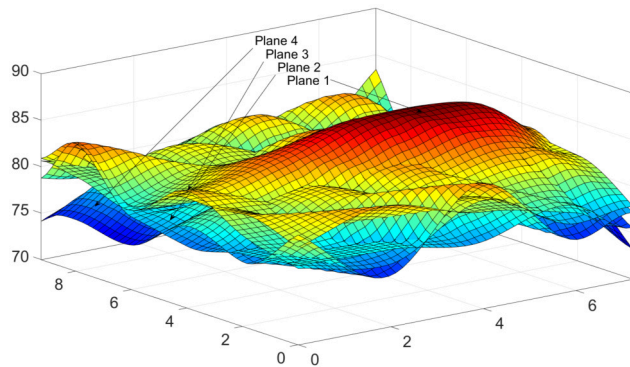


Fig. 8. Comparison of SPL at the third fixed position of the source in all planes.

for the first and third locations of the SS in four different planes, and the results were compared against the simulation results.

Figures 10a and 10d show line graphs that compare the experimental and numerical results. The line graphs in Fig. 10a represent the straight lines taken on plane 1. The graph showed that the variation in the SPL from modeling was uniform when compared to experimental results due to modeling data computed at continuous points on the line. After reaching a steady state, the sound level meter's equivalent continuous sound level mode provided the SPL without fluctuations. The SPL instability is visible in the simulation's steady state. The simulation fluctuations

$Sx = 4$, $Sy = 0$, $Sz = 3.14$, $freq(1) = 4000$ Hz.
Surface: Sound pressure level [dB].

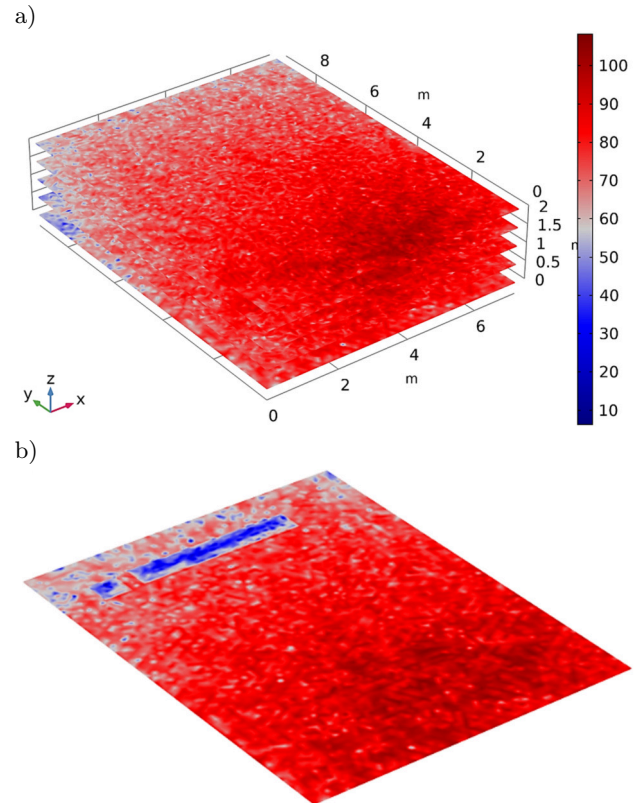


Fig. 9. (a) Surface plots of SPL (sample modeling results) and (b) surface plot of plane at $y = 0$.

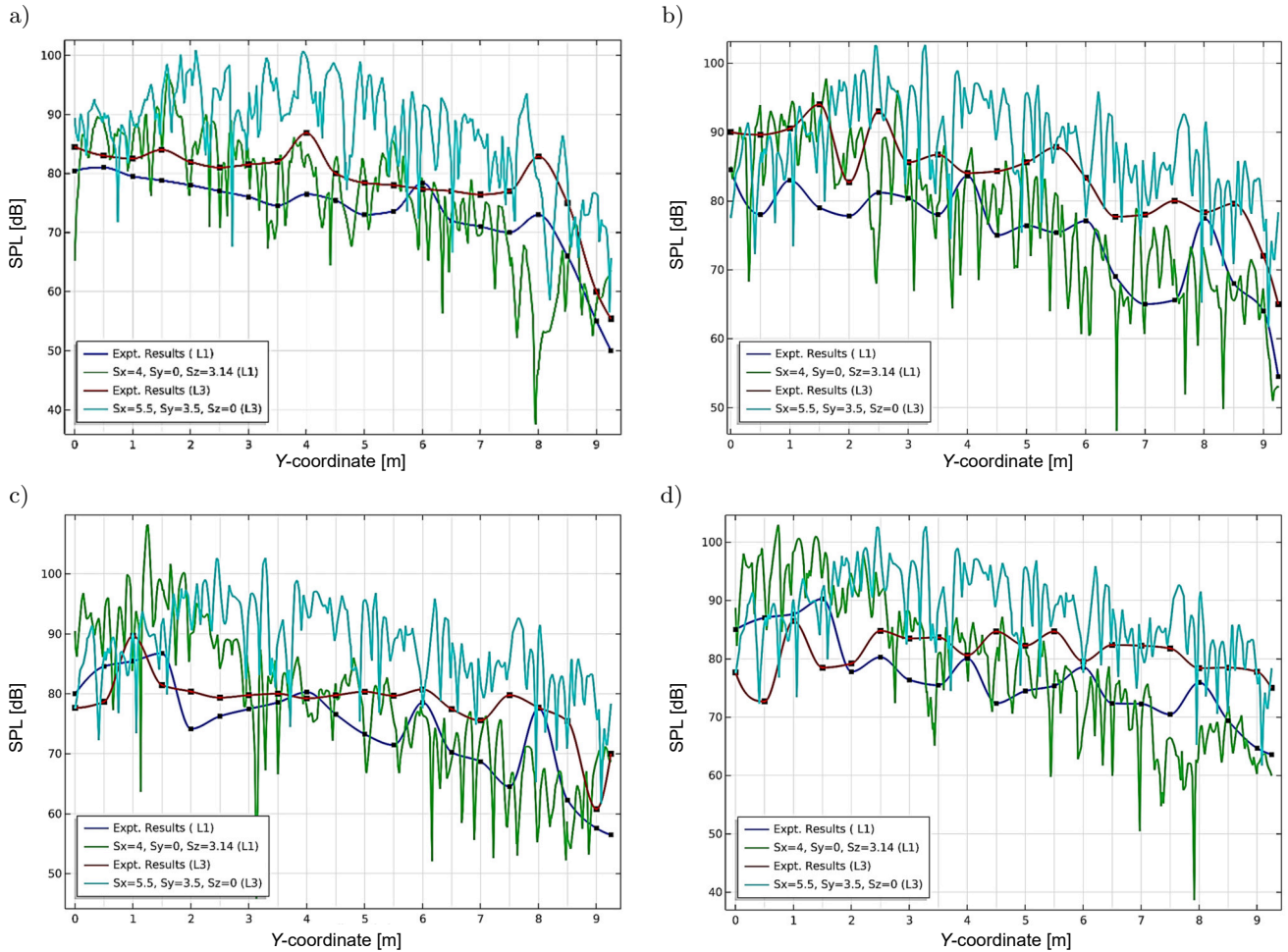


Fig. 10. Comparison of experimental and simulation results for: (a) plane 1; (b) plane 2; (c) plane 3; (d) plane 4.

show the values for each point in the room and do not vary with time.

Further, additional experiments are conducted to investigate the capability of a numerical model for predicting the SPL variation of any rectangular space for different frequency ranges with different objects and interiors if the velocity of sound and the absorption coefficient of the material are known. The additional experiments are conducted in the seminar hall and a different room. The SPL was noted for four randomly selected points.

Table 2 compares experimental and simulation results at different frequencies for the seminar hall. The SPL was reduced with the increment of frequency for the tested frequency values during the experiment. The numerical model captured this effect well, and the SPL was dropped in simulation results compared to the experiments. However, a slight variation in the SPL could be noticed; the SPL in simulation results is 3 %–5 % higher than the experimental results, possibly due to losses and unavoidable noise due to atmospheric factors present during the experiment. The trend of the SPL variation with frequency change was similar for experimentation and simulation.

Table 3 compares experimental and simulation results for a normal room at different frequencies. A similar observation was reported for the room and the seminar hall. The prediction of the SPL from the numerical model was in line with the experimentally evaluated SPL values for all randomly selected locations in the room for tested frequencies.

4. Conclusions

This study provides a comprehensive assessment of SPL distributions in a classroom environment, both experimentally and through FEM simulations. The findings demonstrate how the SPL varies with sound source location and frequency, providing critical insights for optimizing classroom acoustics. The results highlight the importance of considering spatial variability in the SPL for improving speech intelligibility, particularly in classrooms with complex geometries. This work also offers a replicable methodology for assessing classroom acoustics that can be extended to other indoor spaces, such as lecture halls and meeting rooms. It is concluded from the experiments that

Table 2. Comparison of experimental and simulation results at different frequencies for the seminar hall.

Simulation no.	Location [m] (1st fixed position of the speaker)			Frequencies [Hz] used for experimentation results of SPL in seminar hall				Frequencies [Hz] used for simulation results of SPL in seminar hall			
	x	y	z	1000	2000	3000	4000	1000	2000	3000	4000
1.	2	4	0.5	90.1	87	85	79.44	92.8	90.6	89.5	83.1
2.	4	6	0.5	85.1	82.9	82.1	75.36	87.5	86.2	86.4	81.2
3.	6	2	0.5	83.7	81.8	80.9	73.84	84.2	84.3	84.7	79.2
4.	8	4	0.5	87	84.7	83.1	76.52	90.7	85.8	87.5	80.4
5.	2	4	1.0	88.2	86.1	84.1	77.8	90.2	91.5	87.5	83.1
6.	4	6	1.0	90.2	87	85.1	79.42	92.7	87.4	86.5	83
7.	6	2	1.0	82.1	80.2	79.3	72.56	87.1	82.1	81.2	75.2
8.	8	4	1.0	88	85.9	84	77.46	91.6	87.6	85.8	81.1
9.	2	4	1.5	83.8	82	81	74.14	88.6	87.5	83.1	80
10.	4	6	1.5	86.9	84.7	83	76.18	91.2	88.5	85.6	80.2
11.	6	2	1.5	83.1	81.9	80.6	73.5	87.5	85.7	84	78.1
12.	8	4	1.5	83	81.2	80.5	73.36	86.9	85	84.2	80.3
13.	2	4	2.0	84.7	82.3	81.4	74.76	87.7	84.2	83.1	79
14.	4	6	2.0	89.6	83.5	84.6	78.72	92.1	86.3	87.9	81.1
15.	6	2	2.0	86.4	84.4	82.6	75.12	88.2	87	83.5	79.2
16.	8	4	2.0	86.7	84.5	82.9	75.92	90	88.2	86	81.5

Table 3. Comparison of experimental and simulation results at different frequencies for a normal room.

Simulation no.	Location [m] (1st fixed position of the speaker)			Frequencies [Hz] used for experimentation results of SPL in normal room				Frequencies [Hz] used for simulation results of SPL in normal room			
	x	y	z	1000	2000	3000	4000	1000	2000	3000	4000
1.	2	4	0.5	93.2	91.2	88.1	84.7	94.2	92	89.1	86.1
2.	4	6	0.5	88.3	86.1	85.8	80.7	89.2	86.7	86.4	83.1
3.	6	2	0.5	86.8	85.4	84.2	78.1	88.4	83.2	82.1	82
4.	8	4	0.5	90.1	87	86.5	81.2	93.5	89.5	85	84.6
5.	2	4	1.0	91.3	89.6	87.2	82.4	95.1	90.5	87	83.2
6.	4	6	1.0	93.5	90.1	88.5	84.2	94.3	88.4	86.1	81.2
7.	6	2	1.0	85.1	83.2	82.7	77	86.7	85.1	80.2	74
8.	8	4	1.0	91.5	88.7	87.1	82.1	93.9	89.7	87.9	85.1
9.	2	4	1.5	87.2	86.1	83.9	79.5	89.5	86.5	85	82.5
10.	4	6	1.5	90.3	88	86.1	81.4	92.5	88.1	88	83.2
11.	6	2	1.5	86.5	85.1	83.9	79	88	86.9	84	83.1
12.	8	4	1.5	86.7	84.6	84.2	78.6	89.1	84.7	82.1	79
13.	2	4	2.0	87.9	85.2	84.6	80	90.1	84.1	81	80.1
14.	4	6	2.0	93	86.2	88.2	83.2	93.5	88.9	88	85
15.	6	2	2.0	90.1	87.5	85	80.3	92.1	88.1	84.5	81.9
16.	8	4	2.0	89.8	87.1	85.4	81	88.5	86.1	82	81.1

source directivity is a significant factor as an on-axis to the source. The SPL was comparatively found as a continuous varying curve, but SPL values varied considerably for other axes also. At the corners, the variations in the SPL were found maximum due to the higher absorption coefficient variation. As the material absorption coefficient varies at the corner because of the connection of two walls, the sound wave will get interrupted, and a discrepancy occurs. At the corners, the variation in the SPL was significant due to

the source's directivity and construction or destruction of interference of waves. The SS location was also found as a significant factor in variation of the SPL behaviour. The SPL dropped for the tested sound frequency range with the increment in frequency. Changing the material in the interiors and surfaces of the room may alter the room's acoustic performance.

The FE model has predicted the SPL effectively and can be employed for the various concert halls, theatres, sports halls, and fictional rooms for the tested

frequency range. The computation time has significantly increased for higher frequency ranges. These structures' acoustic performance can be analyzed after evaluating the speed of sound and absorption coefficient of different materials used in interior parts of the room. The application of the FEM in this study provides unique insights into the spatial variation of the SPL at a specific frequency, revealing non-uniformities that may not be captured by simpler models. This study also demonstrates the utility of the FEM in providing detailed spatial and frequency-specific insights into classroom acoustics, which are critical for designing learning environments optimized for speech intelligibility. While harmonic tones serve as a controlled experimental approach, future work should incorporate broader spectra and real-world sound sources to extend these findings. Further investigations incorporating other acoustic parameters, such as RT, STI, C50, etc., may also be considered for a more holistic evaluation. The selection of these frequencies (1000 Hz, 2000 Hz, and 3000 Hz) was based on previous studies emphasizing the importance of mid-to-high frequency bands in determining speech clarity in typical classroom settings. However, including lower frequencies (250 Hz, 500 Hz, and 750 Hz) would provide a more comprehensive understanding of speech intelligibility and can be considered as future work.

The results of this study can help in the design of classrooms and other educational spaces by optimizing sound source placement, material choices, and overall room geometry to enhance speech clarity and reduce acoustic discomfort. By providing both experimental and numerical insights, this study bridges the gap between theory and practical application, offering a more effective approach for achieving acoustically comfortable learning environments. Additionally, the hybrid methodology introduced here can be applied to a wide range of indoor spaces that require acoustic optimization. Future challenges that could be incorporated into the current FE model include modeling of source and boundary properties as well as frequency assessments.

Acknowledgments

The authors acknowledge the support of TEQIP-II for using their measuring instruments and AMD, MNIT for using the seminar hall.

Declarations

All authors have guidance on competing interests, and none of the authors have any competing interests in the manuscript. All authors have read and approve this version of the article, and due care has been taken to ensure the integrity of the work.

References

1. BISTAFA S.R., BRADLEY J.S. (2000), Predicting reverberation times in a simulated classroom, *The Journal of the Acoustical Society of America*, **108**(4): 1721–1731, <https://doi.org/10.1121/1.1310191>.
2. BRADLEY J.S. (1986), Speech intelligibility studies in classrooms, *The Journal of the Acoustical Society of America*, **80**(3): 846–854, <https://doi.org/10.1121/1.393908>.
3. BRADLEY J.S., REICH R.D., NORCROSS S.G. (1999), On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility, *The Journal of the Acoustical Society of America*, **106**(4): 1820–1828, <https://doi.org/10.1121/1.427932>.
4. 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>.
5. BUDZYŃSKI G. (1986), Theory of the reflective localization of sound sources, *Archives of Acoustics*, **11**(1): 13–24.
6. GRAMEZ A., BOUBENIDER F. (2017), Acoustic comfort evaluation for a conference room: A case study, *Applied Acoustics*, **118**: 39–49, <https://doi.org/10.1016/j.apacoust.2016.11.014>.
7. HONGISTO V., LINDBERG M., LAHTI A., VEERMANS M., ALAKOIVU R., RADUN J. (2023), How acoustic refurbishment of a classroom affected pupils and sound levels – A natural experiment, *Proceedings of Forum Acusticum*, <https://doi.org/10.61782/fa.2023.0506>.
8. JERLEHAG C., LEE P.J., PARK S.H., JONES T., CARROLL N. (2018), Acoustic environments of patient room in a typical geriatric ward, *Applied Acoustics*, **133**: 186–193, <https://doi.org/10.1016/j.apacoust.2017.12.022>.
9. MEALINGS K. (2023a), A scoping review of the effect of classroom acoustic conditions on primary school children's numeracy performance and listening comprehension, *Acoustics Australia*, **51**(1): 129–158, <https://doi.org/10.1007/S40857-022-00284-3/FIGURES/12>.
10. MEALINGS K. (2023b), The effect of classroom acoustic treatment on listening, learning, and well-being: A scoping review, *Acoustics Australia*, **51**(2): 279–291, <https://doi.org/10.1007/S40857-023-00291-Y/FIGURES/3>.
11. MEALINGS K., MILES K., MATTHEWS N., BUCHHOLZ J.M. (2024), Towards an acoustically accessible campus: A case study of the acoustic conditions of an Australian University, *Acoustics Australia*, pp. 1–6, <https://doi.org/10.1007/s40857-024-00323-1>.
12. MIR S.H., ABDOL A.A. (2005), Investigation of sound-absorbing material configuration of a smart classroom utilizing computer modeling, *Building Acoustics*, **12**(3): 175–188, <https://doi.org/10.1260/135101005774353032>.
13. NÁBĚLEK A.K., LETOWSKI T.R., TUCKER F.M. (1989), Reverberant overlap- and self-masking in consonant

- identification, *The Journal of the Acoustical Society of America*, **86**(4): 1259–1265, <https://doi.org/10.1121/1.398740>.
14. NEWMAN R.B., SABINE W.C. (1965), Collected papers on acoustics, *Journal of Architectural Education* (1947–1974), **20**(2): 22, <https://doi.org/10.2307/1424115>.
 15. NIJS L., RYCHTÁRIKOVÁ M. (2011), Calculating the optimum reverberation time and absorption coefficient for good speech intelligibility in classroom design using U50, *Acta Acustica United with Acustica*, **97**(1): 93–102, <https://doi.org/10.3813/AAA.918390>.
 16. PENG J., WANG C., JIANG P., LAU S.K. (2016), Investigation of Chinese word recognition scores of children in primary school classroom with different speech sound pressure levels, *Applied Acoustics*, **110**: 235–240, <https://doi.org/10.1016/j.apacoust.2016.03.026>.
 17. RABELO A.T.V., SANTOS J.N., OLIVEIRA R.C., MAGALHÃES M. de C. (2014), Effect of classroom acoustics on the speech intelligibility of students, *Codas*, **26**(5): 360–366, <https://doi.org/10.1590/2317-1782/20142014026>.
 18. REICH R., BRADLEY J. (1998), Optimizing classroom acoustics using computer model studies, *Canadian Acoustics – Acoustique Canadienne*, **26**(4): 15–21, <https://jcaa.caa-aca.ca/index.php/jcaa/article/view/1179>.
 19. SMIRNOWA J., OSSOWSKI A. (2005), A method for optimising absorptive configurations in classrooms, *Acta Acustica United with Acustica*, **91**(1): 103–109.
 20. VAN REENEN C., MANLEY D. (2023), Classroom acoustics: Mainstreaming and application of standards, *Proceedings of Meetings on Acoustics*, **51**(1), <https://doi.org/10.1121/2.0001745>.
 21. VEDRTNAM A., PAWAR S.J. (2018), Experimental and simulation studies on acoustical characterisation of laminated safety glass, *Glass Technology: European Journal of Glass Science and Technology Part A*, **59**(2): 58–70, <https://doi.org/10.13036/17533546.59.2.008>.
 22. VIGEANT M.C., WANG L.M., RINDEL J.H. (2006), Room acoustics computer modeling: Study of the effect of source directivity on auralizations, *AEI 2006: Building Integration Solutions – Proceedings of the 2006 Architectural Engineering National Conference*, **2006**: 22, [https://doi.org/10.1061/40798\(190\)22](https://doi.org/10.1061/40798(190)22).
 23. VISENTIN C. (2023), Background noise in classrooms: How it affects performance, EIAS2023 presentation, <https://www.acousticbulletin.com/background-noise-in-classrooms-how-it-affects-performance-eias-2023/>.
 24. VISENTIN C., PRODI N., CAPPELLETTI F., TORRESIN S., GASPARELLA A. (2018), Using listening effort assessment in the acoustical design of rooms for speech, *Building and Environment*, **136**: 38–53, <https://doi.org/10.1016/j.buildenv.2018.03.020>.
 25. VORLÄNDER M. (1998), Objective characterization of sound fields in small rooms, [in:] *Audio Engineering Society Conference: 15th International Conference: Audio, Acoustics & Small Spaces*, **219**, <https://doi.org/999910121283>.
 26. WEYNA S. (1996), An image of the energetic acoustic field in reduced parallelepiped room models, *Acustica*, **82**(1): 72–81.
 27. World Health Organisation (1999), Guidelines for Community Noise, Berglund B., Lindvall T., Schwela D.H. [Eds.].
 28. YANG W., BRADLEY J. S. (2009), Effects of room acoustics on the intelligibility of speech in classrooms for young children, *The Journal of the Acoustical Society of America*, **125**(2): 922–933, <https://doi.org/10.1121/1.3058900>.
 29. ZHANG D., TENPIERIK M., BLUYSSSEN P.M. (2019), Interaction effect of background sound type and sound pressure level on children of primary schools in the Netherlands, *Applied Acoustics*, **154**: 161–169, <https://doi.org/10.1016/j.apacoust.2019.05.007>.