

eISSN 2300-262X (online)



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

ARCHIVES of ACOUSTICS

QUARTERLY

Vol. 49, No. 4, 2024

WARSAW



ARCHIVES of ACOUSTICS

QUARTERLY, Vol. 49, No. 4, 2024

Research Papers

J. Sun, J. Zhu, J. Shao, <i>Fine-grained recognition of fidgety-type emotions using multi-scale one-dimensional residual Siamese network</i>	471
B. Avşar Aksu, D. Şahin Ceylan, G. Gültekin, <i>Mismatch negativity as a tool to detect the intensity just noticeable difference</i>	483
P. Kania, D. Kania, T. Łukaszewicz, <i>A real-time key-finding algorithm based on the signature of fifths</i>	491
M. Ahsan, D. Bismor, P. Fabiś, <i>Analysis of the vehicle engine misfires using frequency-domain approaches at various RPMs with ADXL1002 accelerometer</i>	507
C. Bartmański, A. Bramorska, <i>Feasibility of using wavelet analysis and machine learning method in technical diagnosis of car seats</i>	517
J. Li, L. An, Y. Cheng, H. Wang, <i>Sound quality prediction method of dual-phase Hy-Vo chain transmission system based on MFCC-CNN and fuzzy generation</i>	527
E.O. Do Nascimento, P.H.T. Zannin, <i>Combined evaluation of room acoustic descriptors in different structural configurations via ODEON simulations and artificial neural networks</i>	543
A. Nowicki, J. Tasinkiewicz, P. Karwat, I. Trots, N. Żolek, R. Tymkiewicz, <i>Ultrasound imaging of nonlinear media response using a pressure-dependent nonlinearity index</i>	557
Z. Wang, Z. Peng, F. Zhou, L. Tan, <i>Time-domain analysis of echoes from solid spheres and spherical shells with separated transmit-receive configurations</i>	565
S. Xu, Z. Peng, F. Zhou, X. Miu, H. Ke, <i>Prediction method and characteristics of static acoustic scattering for marine composite propellers</i>	575
J. Radosz, <i>Sound insulation of an acoustic barrier with layered structures of sonic crystals – comparative studies of physical and theoretical models</i>	591
D. Zhou, X. Yang, Y. Tang, Y. Miao, <i>A study of acoustic emission based RA-AF characteristics of polypropylene fiber-reinforced recycled aggregate concrete under uniaxial compression</i>	601
Y. Liu, X.-P. He, S.-P. He, <i>An algorithm for ultrasonic identification of ceramic materials and virtual prototype realization</i>	613
A. Perelomova, <i>Excitation of the secondary modes by the broad spectrum sound in a liquid with relaxation losses</i>	625

Editorial Board

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

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

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)

KOMPALA Janusz (Central Mining Institute, Poland)

LENIOWSKA Lucyna (University of Rzeszów, Poland)

PIECHOWICZ Janusz (AGH University of Krakow, Poland)

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

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

Advisory Editorial Board

Chairman: WESOŁOWSKI Zbigniew (Polish Academy of Sciences, Poland)

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

Fine-Grained Recognition of Fidgety-Type Emotions Using Multi-Scale One-Dimensional Residual Siamese Network

Jiu SUN*, Jinxin ZHU, Jun SHAO

*School of Information Technology, Yancheng Institute of Technology
Yancheng, Jiangsu, China**Corresponding Author e-mail: sunjiu@ycit.edu.cn*(received August 12, 2023; accepted April 4, 2024; published online July 15, 2024)*

Fidgety speech emotion has important research value, and many deep learning models have played a good role in feature modeling in recent years. In this paper, the problem of practical speech emotion is studied, and the improvement is made on fidgety-type emotion using a novel neural network model. First, we construct a large number of phonological features for modeling emotions. Second, the differences in fidgety speech between various groups of people were studied. Through the distribution of features, the individual features of fidgety emotion were studied. Third, we propose a fine-grained emotion classification method, which analyzes the subtle differences between emotional categories through Siamese neural networks. We propose to use multi-scale residual blocks within the network architecture, and alleviate the vanishing gradient problem. This allows the network to learn more meaningful representations of fidgety speech signal. Finally, the experimental results show that the proposed method can provide the versatility of modeling, and that fidgety emotion is well identified. It has great research value in practical applications.

Keywords: residual convolutional neural network; multi-scale neural network; fidgety speech emotion; fine-grained emotion classification; Siamese neural networks.



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/>).

Acronyms

1-D – one-dimensional,
AI – artificial intelligence,
CNN – convolutional neural network,
GMM – Gaussian mixture model,
LSTM – long short-term memory,
PCM – pulse code modulation,
RNN – recurrent neural network,
SEU – Southeast University,
SVM – support vector machine,
USB – Universal Serial Bus,
WAV – Waveform Audio File Format.

1. Introduction

Emotion recognition is a fundamental aspect of human communication and understanding. It plays a crucial role in various domains, including psychology, human-computer interaction, and social robotics.

Traditional approaches to emotion recognition have primarily focused on categorical classification (LATIF *et al.*, 2023; YAN *et al.*, 2013), but there is a growing need for more detailed analysis, especially in capturing subtle variations and specific types of emotions.

Various feature analysis and modeling algorithms have been applied to speech emotion recognition, including feature normalization, stochastic parameter optimization, neural networks and Gaussian mixture models (JIN *et al.*, 2009; 2014; HUANG *et al.*, 2009a; WANG, TASHEV, 2017; LIESKOVSKÁ *et al.*, 2021). CHEN and HUANG (2021), proposed to study hybrid features in speech emotion recognition. DUPUIS and PICHORA-FULLER (2014) recommended to study behavioural features in emotional speech. ATILA and ŞENGÜR (2021) proposed to use the novel convolutional neural network and long-short term memory network for emotion recognition. In their study, deep neural network structures were reviewed and studied. Large amount of data is required for deep learning.

PRASEETHA and VADIVEL (2018) also studied deep learning models. In their studies only basic emotions were investigated.

Other researchers studied practical problems in emotion recognition, including text and speaker independent emotion recognition, practical types of emotions, cognitive related states, and language specific emotion models (HUANG *et al.*, 2013a; 2013b; 2016; WU *et al.*, 2018; JIN *et al.*, 2011; XU *et al.*, 2014; ZOU *et al.*, 2011).

ZHOU *et al.* (2021) suggested to study a cough sound event using acoustic features. In their study, a Mel-spectrogram was used for feature analysis and a convolutional neural network was used for modeling. COVID-19 influenced cough sound recognition has wide potential applications. ATSAVASIRILERT *et al.* (2019) proposed to study the computational efficiency in speech emotion recognition. In their study, the light weight convolutional neural network was proposed, and the real world challenges in computing resources were given their work has important practical value. They further studied Mel-spectrograms and treated the speech signal processing problem as 2-D information processing. However, in their work, emotion types were limited.

Emotion recognition is an important field in understanding human behavior, with traditional machine learning models and deep neural networks being widely used for classification. However, limited research has been conducted on emotions with specific practical values, such as fidgety emotions, which have unique significance.

This research paper addresses the gap in fine-grained practical speech emotion recognition by providing a more detailed categorization of emotions. While the traditional approach considers six main emotional categories (sadness, joy, anger, disgust, surprise, and fear), this paper aims to explore emotions with special practical value, including fidgety emotions. By considering specific application scenarios, fine-grained subtypes, and composite types of emotions, this paper offers a comprehensive framework for emotion detection in practical applications.

Fidgety emotion represents a significant emotional category distinct from traditional emotion research, which primarily focuses on basic emotional categories. Fidgety is a complex emotion with practical value, playing a crucial role in the realms of learning and cognition. It holds particular significance in influencing cognitive abilities, behavioral control, and psychological stability. While conventional emotion recognition research extensively explores the six basic emotions, happiness, anger, surprise, sadness, fear, and disgust, there has been limited investigation into complex emotions.

Fidgety emotion, characterized by its complexity, is particularly triggered in repetitive and tedious cogni-

tive tasks, especially during prolonged periods of repetitive work. It remains a complex emotion with practical implications, significantly impacting cognitive abilities, behavioral control, and psychological stability within the processes of learning and cognition.

The paper explores the use of a Siamese neural network architecture, which excels in metric distance learning, for comparing and classifying fidgety-type emotions. We further propose to use a 1-D convolutional residual neural network, to improve the Siamese network structure. By constructing a large number of phonological features and analyzing group differences, the model captures individual characteristics and enables precise identification of emotional subcategories.

The empirical prowess of 1-D convolutional networks has been well-documented, asserting their supremacy in diverse time-serial feature extraction and modeling tasks. Numerous instances have showcased their state-of-the-art performance in extracting intricate patterns from temporal data streams, such as vibration signal processing, fault detection, and ECG signal processing (ABDELJABER *et al.*, 2017; AVCI *et al.*, 2018; 2019; KIRANYAZ *et al.*, 2019; XIONG *et al.*, 2017). However, the use of residual shortcut and multi-scale receptive fields in specific emotion recognition has not been studied yet.

The proposed approach takes into account the nuances and complexities of fidgety emotions, which have important practical implications. By providing a more detailed understanding of these emotions, the research contributes to the development of effective emotion recognition systems. Additionally, by considering the specific contexts and characteristics of fidgety emotions, the proposed framework is tailored to address their unique practical challenges. This research serves as a valuable contribution to the field of fine-grained practical speech emotion recognition, providing insights and techniques for improved detection and understanding of fidgety emotions.

The key contribution of this research lies in its practical application of fine-grained fidgety-type emotion recognition using the improved Siamese network structure. The proposed method demonstrates versatility in modeling emotions across different ages and genders, showcasing its potential for real-world applications. The experimental results validate the effectiveness of the approach, giving promising practical implications in emotion recognition.

The paper is structured as follows: Sec. 2 provides an overview of the database used for training and evaluation. Section 3 presents the methodology employed for fine-grained fidgety-type emotion recognition as a few-shot learner. Section 4 presents the experimental results obtained from applying the proposed methodology. Finally, Sec. 5 concludes the paper by summarizing the key findings and discussing the implications and future directions of the research.

2. Database

We have employed a local database from Southeast University (SEU) to validate our method of emotion recognition (HUANG *et al.*, 2009b; 2011; 2014; 2020), for fidgety-type emotions.

The recording software uses Adobe Audition. During recording, a monaural channel is used. The recorded speech signals are saved in the WAV (Waveform Audio Format) format encoded with PCM (pulse-code modulation). The recording hardware includes: one high-performance computer, one M-audio MobilePre USB sound card, one large-diaphragm condenser microphone, and one pair of monitoring headphones.

The recording process takes place in a quiet laboratory. After each recording, data verification and supplementation should be carried out. The recorded speech files should be manually checked promptly to eliminate any possible errors that may occur during the recording process. For example, inspecting and removing segments with signal overload, irregular noises (such as coughing), and long periods of silence caused by abnormal pauses. If the recording files have significant errors, supplementary recording may be necessary.

The collected data within this database encompasses speech-based emotions of a cognitive nature, encompassing emotions such as annoyance, fatigue, confidence, and joy.

For the purpose of this paper, a specific subset of utterances are chosen from the SEU database. To capture elicited emotional speech, negative emotions are induced through mathematical calculation tasks, involving the verbal reporting of calculated results and recording emotional speech, all conducted in Chinese. In the experimental dataset, 8 male and 8 female native Chinese-speaking participants volunteered, with careful selection to ensure gender balance, resulting in 3000 utterances for each gender category. The induction experiment avoided a standardized text, opting for the emotional speech collection in a natural state, in contrast to the scripted nature of a standardized text often used in acted speech recording. The recorded dataset comprises 6000 sentences, totaling 18 662 seconds, dis-

tributed across 2000 samples for fidgetiness, 2000 for happiness, and an additional 2000 for a neutral emotional state, forming a comprehensive subset of 6000 samples.

In addition to utilizing the SEU database for our research on speech emotion recognition, we have undertaken the task of manual annotation to achieve a fine-grained level of more detailed emotion types, as shown in Fig. 1. This meticulous process adds significant value to our research problem. By annotating the data ourselves, we ensure a comprehensive and nuanced understanding of the emotions expressed in the speech samples. This granular approach enables us to capture subtle variations and nuances within emotions, contributing to a more accurate and comprehensive analysis. The annotators are carefully selected with a background in psychology study and proper training of emotion utterance annotation. The annotation results are cross-confirmed. We adopted a multiple annotation approach with a voting strategy.

The five fidgety levels are divided into five categories based on the general discriminative ability of human annotators, using ratings of 1, 3, 5, 7, and 9. Different intensity levels are assigned based on the strength of emotions. This annotation is employed to distinguish fine-grained emotional intensities, facilitating supervised learning to differentiate between specific emotional nuances.

3. Methodology

3.1. Few-shot fine-grained fidgety-type emotion recognition

Fine-grained fidgety-type emotion recognition refers to the accurate detection and classification of subtle variations in emotions, particularly those expressed through fidgety behavior. Few-shot learning is a machine learning approach that can generalize from a small number of training examples, which is crucial for emotion recognition tasks where obtaining large labeled datasets is challenging.

In our paper, a few-shot learning framework is applied to fine-grained fidgety-type emotion recognition. The paper introduces the concept of a Siamese neural

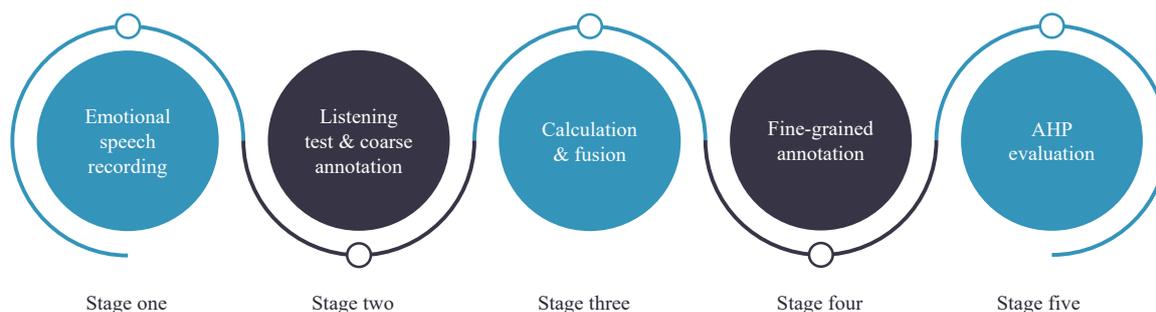


Fig. 1. Flow chart of the fine-grained annotation for emotional speech.

network, which is well-suited for metric distance learning. The Siamese network compares the input samples with templates and learns to measure the similarity or dissimilarity between them, as shown in Fig. 2.

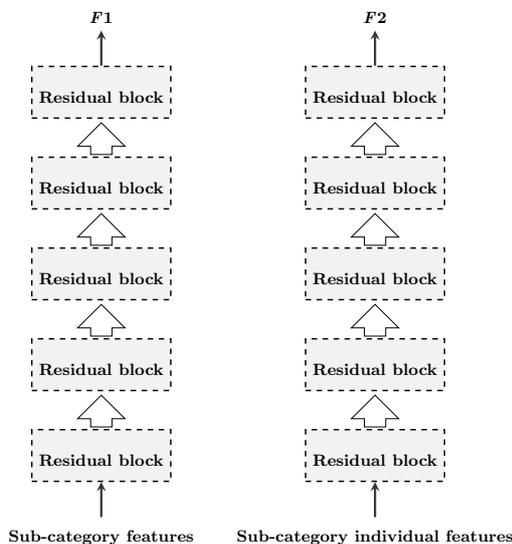


Fig. 2. Siamese network for fine-grained fidgety-type emotion recognition.

The novelty of the proposed solution lies in the utilization of the Siamese network as a few-shot learner. By leveraging this architecture, the model can effectively learn to recognize and classify fidgety-type emotions, even with limited training data. The Siamese network ability to learn meaningful representations of emotional features, combined with the few-shot learning approach, enhances the accuracy of fine-grained identification.

By constructing a large number of phonological features, analyzing group differences, and utilizing residual connections to address the vanishing gradient problem, the proposed method in the paper achieves a fine-grained emotion classification. This approach allows for the precise analysis of subtle differences between emotional categories. The experimental results demonstrate the versatility of the proposed method, highlighting its potential for practical applications in emotion recognition tasks involving fidgety speech.

First, we generate pairs of emotional samples, divided into positive and negative matches, and when we collect a small number of fidgety emotional types of specific speakers, we randomly select samples that are not sub-category and pair them to produce a negative training dataset. It is necessary to focus on the generation of negative sample pairs of similar sub-categories to improve fine-grained identification.

Within this system, we have incorporated a speaker recognition module that utilizes MFCC features with an i-vector approach. The i-vector approach is a commonly used technique in speaker recognition. It is a statistical modeling method that represents speaker char-

acteristics using a low-dimensional fixed-length vector called the i-vector. This module enables fine-grained matching of fidgety subcategories, specifically within the sample range of individual speakers. The goal is to enhance the accuracy of fine-grained identification.

Not all components of the output contribute equally to the comparison process of the Siamese network outputs. As a result, we have implemented a fully connected layer that takes the outputs of the two sub-networks and generates the final classification output.

3.2. Improved Siamese network based on multi-scale residual network

In the realm of fine-grained modeling and recognition, our proposed incorporation of “multi-scale” architectures with various receptive fields is a promising avenue. This approach allows for a more intricate understanding of intricate details within data. Alongside this, the fusion of few-shot learning principles with distance learning methodologies has proven to be a potent combination in the pursuit of enhancing recognition capabilities.

In a typical Siamese network, we have two identical subnetworks that process input examples independently and produce fixed-length embeddings. These embeddings are then compared to determine their similarity or dissimilarity. In an improved Siamese network, to enhance the network’s performance, we can incorporate residual connections within each subnetwork.

3.2.1. 1-D Convolution

1-D convolution is a fundamental operation in signal processing and data analysis, particularly for analyzing time-series signals. It involves combining two input signals to produce an output signal by sliding one signal (known as the kernel or filter) over the other, element by element, and computing the sum of element-wise products at each step. This operation is often used for various tasks such as feature extraction, filtering, and pattern recognition within time-series data.

The input speech signal is denoted as $x[n]$. The convolution kernel is denoted as $h[k]$.

Sliding operation: the convolution operation involves sliding the kernel over the input signal. At each step, the kernel is aligned with a portion of the input signal, and an element-wise multiplication is performed between the kernel and the overlapping portion of the input signal.

The convolution operation at a given time index n is calculated by sliding the kernel $h[k]$ over the speech signal $x[n]$ and performing the element-wise multiplication followed by summation, as shown in Eq. (1):

$$y[n] = \sum_{k=-\infty}^{\infty} x[n-k] \cdot h[k]. \quad (1)$$

In practice, the summation is limited to the valid range of k where both $x[n - k]$ and $h[k]$ are defined.

The resulting convolved signal $y[n]$ is obtained by performing the above convolution operation for each time index n , as shown in Eq. (2):

$$y[n] = \sum_{k=-\infty}^{\infty} x[n - k] \cdot h[k] \quad \text{for all } n. \quad (2)$$

One of the primary applications of 1-D convolution in time-series analysis is feature extraction and filtering. Using compact 1-D convolution we can highlight specific patterns and features within the fidgety speech signal.

3.2.2. Multi-scale residual convolution

Let us consider a specific layer, denoted as the layer L . The output of the layer L can be represented as $\mathbf{H}_L(\mathbf{x})$, where \mathbf{x} is the input to that layer. To introduce a residual connection, we define the residual function $\mathbf{R}_L(\mathbf{x})$, which captures the difference between the input and output of the layer L . The output of the layer L with the residual connection, denoted as $\mathbf{F}_L(\mathbf{x})$, is given by:

$$\mathbf{F}_L(\mathbf{x}) = \mathbf{H}_L(\mathbf{x}) + \mathbf{R}_L(\mathbf{x}), \quad (3)$$

where $\mathbf{F}_L(\mathbf{x})$ represents the desired output of the layer L . By adding the residual function $\mathbf{R}_L(\mathbf{x})$ to the input \mathbf{x} , we allow the network to learn the residual mapping.

The residual function $\mathbf{R}_L(\mathbf{x})$ can be defined as:

$$\mathbf{R}_L(\mathbf{x}) = \mathbf{W}_L \cdot \mathbf{x}, \quad (4)$$

where \mathbf{W}_L represents the weights of the residual connection, which are learned during the training process. Multiplying the input \mathbf{x} by \mathbf{W}_L allows the network to capture the residual information that needs to be added to the output.

With the addition of residual connections, the output of layer $L + 1$ can be expressed as:

$$\mathbf{H}_{L+1}(\mathbf{F}_L(\mathbf{x})) = \mathbf{H}_{L+1}(\mathbf{H}_L(\mathbf{x}) + \mathbf{R}_L(\mathbf{x})). \quad (5)$$

Convolutional kernels of different scales can extract features of varying precision, with smaller kernels capturing finer details. If a single layer uses only kernels of the same scale, it may overlook features of other precisions, resulting in incomplete information being represented by the extracted features. Consequently, we have designed three distinct resolutions for feature extraction, as illustrated in Fig. 3.

By incorporating residual connections in this manner, the gradient can flow directly from the output of a layer to its input, facilitating the flow of gradients during training. This alleviates the vanishing gradient problem and enables the network to learn more meaningful representations.

In the improved Siamese network, multiple residual connections can be added at different layers. By utilizing residual connections, the improved Siamese network can effectively learn complex patterns and relationships in the input data, leading to better similarity or distance measurements and improved performance in fine-grained emotion recognition. The overall framework is shown in Fig. 4.

Our innovative approach to the fine-grained fidgety emotion recognition challenge involves the utilization

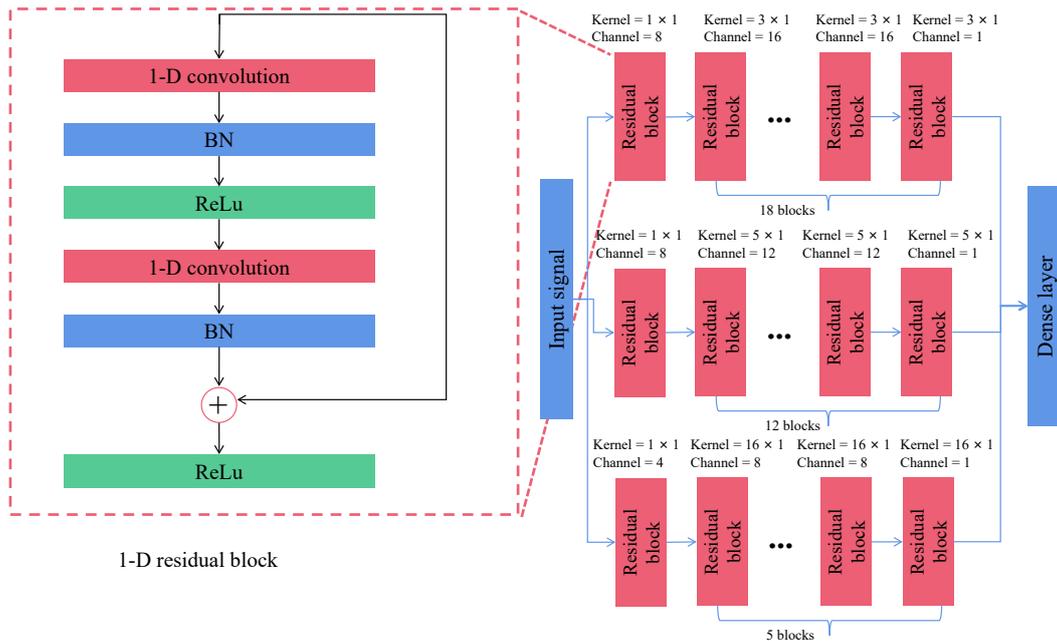


Fig. 3. Proposed multi-scale residual Siamese network structure.

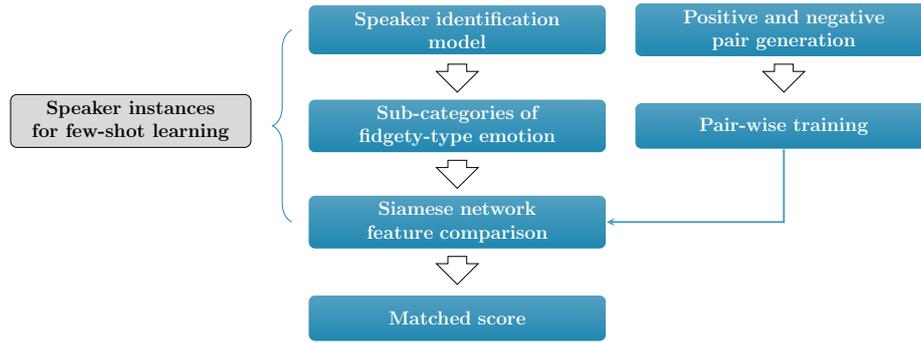


Fig. 4. Metric learning using speaker instances as few-shot learning.

of a 1-D convolutional residual neural network, strategically designed to augment the traditional Siamese network. The integration of residual blocks within this framework plays a pivotal role in enhancing convergence during the training process. By capitalizing on the inherent advantages of 1-D convolutions, particularly their proficiency in processing time-serial signals, our architecture demonstrates remarkable potential.

3.3. Training samples generation

Let S be a speech sample, C_{fid} be the main category of the fidgety emotion, and c_{fid}^j be the subcategory of the fidgety emotion within the main category. The method for generating positive and negative sample pairs is as follows:

$$S_1 \in C_{\text{fid}}, \quad (6)$$

$$S_2 \notin C_{\text{fid}}. \quad (7)$$

$\text{Neg_coarse} = \{S_1, S_2\}$ forms a negative sample pair:

$$S_1 \in c_{\text{fid}}^j, \quad (8)$$

$$S_2 \in c_{\text{fid}}^k, \quad (9)$$

$$j \neq k. \quad (10)$$

$\text{Neg_fine} = \{S_1, S_2\}$ forms a negative sample pair, representing samples that require fine-grained distinction. $\text{Neg_fine} : \text{Neg_coarse} > 3 : 1$ This ensures that the model has a higher resolution for fine-grained samples:

$$S_1 \in c_{\text{fid}}^j, \quad (11)$$

$$S_2 \in c_{\text{fid}}^j. \quad (12)$$

$\text{Pos} = \{S_1, S_2\}$ forms a positive sample pair, used to supervise the output results of the Siamese network. The distance between samples in the same fine-grained subclass should be relatively close.

4. Experimental results

4.1. Experimental data

In our experiments on speech emotion recognition, we have recognized the critical role of the emotion corpus. While basic emotion types have received considerable attention, the study of emotions with practical value remains insufficient. Particularly, the scarcity of negative practical emotions in existing databases poses a challenge. Therefore, we have made a deliberate decision to exclusively employ the SEU database for our research. Unlike other databases that predominantly focus on basic or positive emotions within ordinary settings, the SEU database offers a unique advantage by providing a comprehensive collection of practical emotions, including the elusive fidgety-type emotion. This strategic selection enables us to delve deeper into understanding and accurately recognizing the nuanced emotions encountered in real-world scenarios.

4.2. Models comparison

In this research study, we aim to investigate the effectiveness of the proposed multi-scale residual Siamese network for fine-grained fidgety-type emotion recognition. We compare it against four other classifiers: baseline Siamese network, LSTM, support vector machine (SVM), and Gaussian mixture model (GMM).

The baseline Siamese network is a deep neural network architecture that learns to measure similarity between input samples. It consists of two identical sub-networks that share weights, enabling it to compute a similarity metric between two inputs. The residual Siamese network builds upon this architecture by incorporating residual connections, which help alleviate the vanishing gradient problem and enable easier optimization.

Long short-term memory (LSTM) is a widely used recurrent neural network (RNN) architecture that has shown remarkable success in various sequence-based tasks, including natural language processing and speech recognition. Unlike traditional RNNs, LSTM

incorporates specialized memory cells that can capture and retain information over long periods. This unique characteristic enables LSTM to effectively learn and model complex temporal dependencies in sequences.

SVM is a supervised machine learning algorithm used for classification tasks. It aims to find an optimal hyperplane that maximally separates different classes in the feature space. SVMs are known for their ability to handle high-dimensional data and work well when there is a clear margin of separation between classes.

GMM is a probabilistic model that represents the distribution of data points as a mixture of Gaussian distributions. It can capture complex data patterns by estimating the parameters of Gaussian components. GMMs are versatile and can handle a wide range of data distributions, making them suitable for modeling fine-grained emotions.

4.3. Parameter settings

For the Siamese networks, we use a learning rate of 0.001, batch size of 32, and training for a fixed number of epochs (100). For training the LSTM model, the chosen parameter setting was a learning rate of 0.001, a batch size of 64, training for approximately 50 epochs. SVM parameter settings: $C = 1$, kernel = radial basis function (RBF), $\gamma = 0.1$. GMM parameter settings: number of Gaussian components = 12, mean and covariance initialization based on data, maximum number of iterations = 100.

“Epoch” refers to a single pass through the entire training dataset, and it is used to optimize the model’s parameters by adjusting them based on the accumulated error to improve overall performance during training. “Learning rate” in the context of machine learning is a hyperparameter that determines the size of the steps taken during the optimization process, influencing how quickly or slowly a model converges to the optimal set of parameters. “Batch size” refers to the number of training examples utilized in one iteration, influencing the efficiency of model training and the amount of computational resources required.

The training-to-validation-to-testing ratio is 6:1:3, totaling 6000 samples. Training dataset consists of 3600 utterances; validation dataset consists of 600 utterances; testing dataset consists of 1800 utterances.

In the experimental process of comparing models, we utilized different parameters to obtain the empirically optimal performance for each model. For example, we conducted a search for SVM parameters to set the optimal values. We compared different kernel functions, including RBF, linear, and polynomial, and the results indicated that RBF performed the best. We optimized the values of C and γ through the grid search. For GMM, we experimented with different values for the number of mixture components (4, 12, 16, 24) and employed a diagonal matrix initialization method to optimize the empirically best results for the model. In the case of LSTM, we compared different optimizers, with Adam yielding the best results. We conducted a search for different learning rates, selecting the empirically optimal learning rate based on $F1$ scores.

The purpose of comparing these models is to evaluate the efficacy of the proposed multi-scale residual Siamese network for fine-grained fidgety-type emotion recognition. By contrasting its performance with other established classifiers, such as the baseline Siamese network, LSTM, SVM, and GMM, we can determine whether the additional architectural enhancements of the residual Siamese network yield improved accuracy and robustness in recognizing fine-grained emotions characterized by fidgety behaviors.

4.4. Results

In our experiment, we adopt the confusion matrix as a crucial tool for evaluating and comparing different emotion recognition models. As shown in Tables 1–5, the confusion matrix provides a comprehensive summary of the models’ predictions, enabling us to analyze the true positives, true negatives, false positives, and false negatives in classifying emotions. By utilizing the confusion matrix, we can gain insights into the performance of each model in accurately recognizing and classifying different emotions. This evaluation allows us to compare the effectiveness of various models and make informed decisions regarding their suitability for emotion recognition tasks.

As shown in Fig. 5, we compared various popular machine learning models to gain insights into their performance and effectiveness in our study. By examining

Table 1. Confusion matrix for fine-grained fidgety-type emotion recognition using multi-scale residual Siamese network.

Actual emotion	Predicted emotion					
	Fidgety level 1	Fidgety level 2	Fidgety level 3	Fidgety level 4	Fidgety level 5	Neutral
Fidgety level 1	80.1	8.4	4.5	3.5	1.0	2.5
Fidgety level 2	7.5	81.9	3.5	0.8	1.2	5.1
Fidgety level 3	3.9	6.1	75.2	5.0	2.5	7.3
Fidgety level 4	2.5	5.5	6.0	77.8	4.0	4.2
Fidgety level 5	1.8	2.2	8.0	7.5	75.1	5.4
Neutral	1.5	2.1	2.1	3.9	3.1	87.3

Table 2. Confusion matrix for fine-grained fidgety-type emotion recognition using baseline Siamese network.

Actual emotion	Predicted emotion					
	Fidgety level 1	Fidgety level 2	Fidgety level 3	Fidgety level 4	Fidgety level 5	Neutral
Fidgety level 1	72.7	10.4	6.3	3.1	4.5	2.0
Fidgety level 2	4.7	77.1	6.0	5.5	5.4	1.3
Fidgety level 3	5.8	8.0	70.2	6.4	4.5	5.1
Fidgety level 4	3.4	3.7	0.3	74.8	11.3	6.5
Fidgety level 5	4.5	3.3	7.9	8.5	70.3	5.5
Neutral	1.1	3.4	2.3	4.4	8.4	80.4

Table 3. Confusion matrix for fine-grained fidgety-type emotion recognition using LSTM.

Actual emotion	Predicted emotion					
	Fidgety level 1	Fidgety level 2	Fidgety level 3	Fidgety level 4	Fidgety level 5	Neutral
Fidgety level 1	70.1	8.6	8.2	7.1	1.2	4.8
Fidgety level 2	8.6	75.2	2.4	4.4	7.4	2.0
Fidgety level 3	6.5	7.4	64.9	7.4	6.3	3.5
Fidgety level 4	5.7	7.4	5.3	70.3	6.5	4.8
Fidgety level 5	6.3	6.3	5.2	7.8	64.8	9.6
Neutral	7.8	6.4	3.7	3.3	3.7	75.1

Table 4. Confusion matrix for fine-grained fidgety-type emotion recognition using SVM.

Actual emotion	Predicted emotion					
	Fidgety level 1	Fidgety level 2	Fidgety level 3	Fidgety level 4	Fidgety level 5	Neutral
Fidgety level 1	77.2	4.7	6.3	4.5	3.7	3.6
Fidgety level 2	8.3	75.4	7.4	6.3	2.1	0.5
Fidgety level 3	3.2	10.9	66.8	8.4	8.4	2.3
Fidgety level 4	4.6	8.6	7.3	70.1	5.5	3.9
Fidgety level 5	3.3	1.4	5.6	7.5	71.4	10.8
Neutral	6.4	4.5	2.4	3.3	3.1	80.3

Table 5. Confusion matrix for fine-grained fidgety-type emotion recognition using GMM.

Actual emotion	Predicted emotion					
	Fidgety level 1	Fidgety level 2	Fidgety level 3	Fidgety level 4	Fidgety level 5	Neutral
Fidgety level 1	70.2	7.3	6.3	5.5	3.1	7.6
Fidgety level 2	7.4	77.9	6.2	3.9	2.4	2.2
Fidgety level 3	7.8	8.4	70.8	7.9	3.4	1.7
Fidgety level 4	1.7	8.9	8.4	69.1	6.9	5.0
Fidgety level 5	4.1	1.9	7.3	7.4	70.9	8.4
Neutral	9.8	2.3	5.8	4.5	2.4	75.2

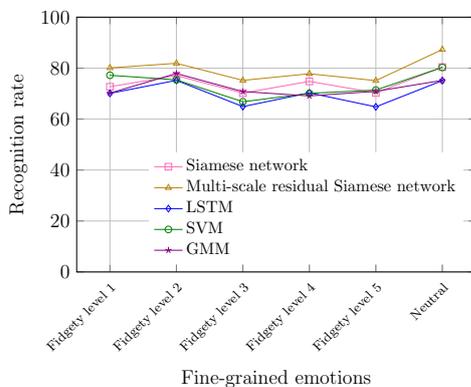


Fig. 5. Comparison among various modeling algorithms for averaged recognition rates.

and comparing the different curves generated by these models, we were able to assess their recognition rates and classification accuracy for the task at hand. This comparative analysis allowed us to evaluate the strengths and weaknesses of each model, identify areas of specialization, and uncover potential limitations.

From the experimental results, we can see that various modeling algorithms exhibit distinctive recognition rates for different fine-grained emotions.

Siamese network: the Siamese network exhibits moderate recognition rates across all fine-grained emotions, ranging from 70.2 % to 80.4 %. It achieves relatively higher rates for fidgety level 1 and fidgety level 2 compared to the other emotions.

Multi-scale residual Siamese network: the residual Siamese network demonstrates consistent performance, with recognition rates ranging from 75.1 % to 87.3 %. It achieves higher rates for fidgety level 1, fidgety level 2, and neutral emotions, indicating its effectiveness in recognizing these categories.

LSTM: the LSTM model showcases relatively lower recognition rates, ranging from 64.8 % to 75.2 %. It may require a much larger training database to capture the subtle distinctions between fine-grained emotions, resulting in slightly lower overall performance.

SVM: the SVM model demonstrates varied recognition rates, ranging from 66.8 % to 80.3 %. It performs relatively well for fidgety level 1 and fidgety level 2 emotions, but its performance drops for fidgety level 3 and fidgety level 4.

GMM: the GMM model achieves recognition rates ranging from 69.1 % to 77.9 %. It displays relatively lower rates compared to other models, particularly for fidgety level 1, fidgety level 3, and fidgety level 4 emotions.

Overall, the multi-scale 1-D residual Siamese network stands out with the highest recognition rates across various fine-grained emotions. The Siamese network and SVM models perform reasonably well, but their rates are slightly lower compared to the residual Siamese network. The LSTM and GMM models exhibit comparatively lower recognition rates, indicating the need for further improvement in capturing fine-grained emotional nuances.

4.5. Discussions

The advantages of the multi-scale residual Siamese network lie in its ability to enhance the model depth and, consequently, improve representation capability by introducing residual results. The use of the Siamese network structure enables fine-grained differentiation of emotion categories. However, its drawback is its reliance on a substantial amount of data for training, making it highly data-dependent.

The baseline Siamese network excels in distinguishing subtle differences between different emotions but lacks the introduction of residual structures, leaving room for improvement in representation capability.

LSTM's strength lies in its structure, which is conducive to modeling time series data. However, its computational complexity and convergence in modeling may not always achieve ideal results, especially under conditions of limited objective data.

SVM exhibits strong discriminative power under small-sample conditions, but it lacks the ability for automatic representation learning, making it challenging to fully exploit the value of training data.

GMM's advantage lies in its strong fitting capability and ability to model arbitrary feature distributions.

However, this is contingent upon having sufficient and diverse data coverage, resulting in a high dependence on data.

The performance of a model is influenced by the characteristics of different input data because the statistical machine learning approach is inherently dependent on data. To address this challenge, a strategy is to separate training, validation, and testing data. This allows for objective and reasonable testing on an unseen test set, effectively demonstrating the model's generalization ability.

The multi-scale residual Siamese network proposed by us exhibits high reliability and stability. This is ensured through the separation of our testing, validation, and training sets. Experimental results indicate that its recognition performance surpasses that of other traditional models.

To substantiate the efficacy of our proposed model, we conducted a comprehensive comparative analysis. Our proposed network was meticulously pitted against the traditional Siamese network, as well as other prominent machine learning algorithms. Through rigorous experimentation and meticulous evaluation, our results unveil the prowess of our approach, demonstrating its superior performance in the realm of fine-grained fidgety-type emotions modeling and recognition tasks. This novel fusion of multi-scale architectures, few-shot learning, and distance learning principles, bolstered by the advancements of 1-D convolutional residual neural networks, introduces a pioneering stride towards unraveling the complexities of intricate data domains.

5. Conclusions

This paper focuses on the practical application of fidgety speech emotion recognition. Our contributions are centered around the development of phonological features and the implementation of a meticulous emotion classification method that utilizes Siamese neural networks with residual connections.

To enhance the precision of emotion classification, we have introduced a meticulous approach employing Siamese neural networks. By integrating residual connections, we have effectively addressed the challenge of the vanishing gradient, enabling the network to acquire more meaningful representations of fidgety speech emotions.

Experimental results have demonstrated the efficacy and adaptability of our approach, as we have successfully achieved accurate identification of fidgety emotions. Our proposed approach exhibits significant potential for practical applications and lays the foundation for further advancements in this field.

In future endeavors, it would be valuable to explore the integration of contextual information, such as situational cues and temporal dynamics, in order to

enhance both the accuracy and contextual comprehension of fidgety speech emotion recognition.

Acknowledgments

The authors would like to thank Professor Li Zhao from the Southeast University for providing the raw speech emotional database used in this research.

References

1. ABDELJABER O., AVCI O., KIRANYAZ S., GABBOUJ M., INMAN D.J. (2017), Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks, *Journal of Sound and Vibration*, **388**: 154–170, doi: [10.1016/j.jsv.2016.10.043](https://doi.org/10.1016/j.jsv.2016.10.043).
2. ATILA O., ŞENGÜR A. (2021), Attention guided 3D CNN-LSTM model for accurate speech based emotion recognition, *Applied Acoustics*, **182**: 108260, doi: [10.1016/j.apacoust.2021.108260](https://doi.org/10.1016/j.apacoust.2021.108260).
3. ATSAVASIRILERT K. *et al.* (2019), A light-weight deep convolutional neural network for speech emotion recognition using mel-spectrograms, [in:] *Proceedings of 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (ISAI-NLP)*, pp. 1–4, doi: [10.1109/isai-nlp48611.2019.9045511](https://doi.org/10.1109/isai-nlp48611.2019.9045511).
4. AVCI O., ABDELJABER O., KIRANYAZ S., HUSSEIN M., INMAN D.J. (2018), Wireless and real-time structural damage detection: A novel decentralized method for wireless sensor networks, *Journal of Sound and Vibration*, **424**: 158–172, doi: [10.1016/j.jsv.2018.03.008](https://doi.org/10.1016/j.jsv.2018.03.008).
5. AVCI O., ABDELJABER O., KIRANYAZ S., INMAN D.J. (2019), Convolutional neural networks for real-time and wireless damage detection, *Dynamics of Civil Structures*, **2**: 129–136, doi: [10.1007/978-3-030-12115-0_17](https://doi.org/10.1007/978-3-030-12115-0_17).
6. CHEN Q., HUANG G. (2021), A novel dual attention-based BLSTM with hybrid features in speech emotion recognition, *Engineering Applications of Artificial Intelligence*, **102**: 104277, doi: [10.1016/j.engappai.2021.104277](https://doi.org/10.1016/j.engappai.2021.104277).
7. DUPUIS K., PICHORA-FULLER M.K. (2014), Recognition of emotional speech for younger and older talkers, *Ear & Hearing*, **35**(6): 695–707, doi: [10.1097/aud.000000000000082](https://doi.org/10.1097/aud.000000000000082).
8. HUANG C., CHEN G., YU H., BAO Y., ZHAO L. (2013a), Speech emotion recognition under white noise, *Archives of Acoustics*, **38**(4): 457–463, doi: [10.2478/aoa-2013-0054](https://doi.org/10.2478/aoa-2013-0054).
9. HUANG C., JIN Y., ZHAO Y., YU Y., ZHAO L. (2009a), Speech emotion recognition based on re-composition of two-class classifiers, [in:] *Proceedings of 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–3, doi: [10.1109/acii.2009.5349420](https://doi.org/10.1109/acii.2009.5349420).
10. HUANG C., JIN Y., ZHAO Y., YU Y., ZHAO L. (2009b), Recognition of practical emotion from elicited speech, [in:] *Proceedings of the First International Conference on Information Science and Engineering*, pp. 1–4, doi: [10.1109/icise.2009.875](https://doi.org/10.1109/icise.2009.875).
11. HUANG C., LIANG R., WANG Q., XI J., ZHAO C., ZHAO L. (2013b), Practical speech emotion recognition based on online learning: From acted data to elicited data, *Mathematical Problems in Engineering*, **2013**: 265819, doi: [10.1155/2013/265819](https://doi.org/10.1155/2013/265819).
12. HUANG C., SONG B., ZHAO L. (2016), Emotional speech feature normalization and recognition based on speaker-sensitive feature clustering, *International Journal of Speech Technology*, **19**(4): 805–816, doi: [10.1007/s10772-016-9371-3](https://doi.org/10.1007/s10772-016-9371-3).
13. HUANG C., ZHAO Y., JIN Y., YU Y., ZHAO L. (2011), A study on feature analysis and recognition of practical speech emotion, *Journal of Electronics & Information Technology*, **33**(1): 112–116, doi: [10.3724/sp.j.1146.2009.00886](https://doi.org/10.3724/sp.j.1146.2009.00886).
14. JIN Y., HUANG C., ZHAO L. (2011), A semi-supervised learning algorithm based on modified self-training SVM, *Journal of Computers*, **6**(7): 1438–1443, doi: [10.4304/jcp.6.7.1438-1443](https://doi.org/10.4304/jcp.6.7.1438-1443).
15. JIN Y., SONG P., ZHENG W., ZHAO L. (2014), A feature selection and feature fusion combination method for speaker-independent speech emotion recognition, [in:] *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4808–4812, doi: [10.1109/icassp.2014.6854515](https://doi.org/10.1109/icassp.2014.6854515).
16. JIN Y., ZHAO Y., HUANG C., ZHAO L. (2009), Study on the emotion recognition of whispered speech, [in:] *Proceedings of 2009 WRI Global Congress on Intelligent Systems*, pp. 242–246, doi: [10.1109/gcis.2009.175](https://doi.org/10.1109/gcis.2009.175).
17. KIRANYAZ S., GASTLI A., BEN-BRAHIM L., AL-EMADI N., GABBOUJ M. (2019), Real-time fault detection and identification for MMC using 1-D convolutional neural networks, *IEEE Transactions on Industrial Electronics*, **66**(11): 8760–8771, doi: [10.1109/tie.2018.2833045](https://doi.org/10.1109/tie.2018.2833045).
18. LATIF S., SHAHID A., QADIR J. (2023), Generative emotional AI for speech emotion recognition: The case for synthetic emotional speech augmentation, *Applied Acoustics*, **210**: 109425, doi: [10.1016/j.apacoust.2023.109425](https://doi.org/10.1016/j.apacoust.2023.109425).
19. LIESKOVSKÁ E., JAKUBEC M., JARINA R., CHMULÍK M. (2021), A review on speech emotion recognition using deep learning and attention mechanism, *Electronics*, **10**(10): 1163, doi: [10.3390/electronics10101163](https://doi.org/10.3390/electronics10101163).
20. PRASEETHA V.M., VADIVEL S. (2018), Deep learning models for speech emotion recognition, *Journal of Computer Science*, **14**(11): 1577–1587, doi: [10.3844/jcssp.2018.1577.1587](https://doi.org/10.3844/jcssp.2018.1577.1587).

21. WANG Z.Q., TASHEV I. (2017), Learning utterance-level representations for speech emotion and age/gender recognition using deep neural networks, [in:] *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5150–5154, doi: [10.1109/icassp.2017.7953138](https://doi.org/10.1109/icassp.2017.7953138).
22. WU C., HUANG C., CHEN H. (2018), Text-independent speech emotion recognition using frequency adaptive features, *Multimedia Tools and Applications*, **77**(18): 24353–24363, doi: [10.1007/s11042-018-5742-x](https://doi.org/10.1007/s11042-018-5742-x).
23. XIONG Z., STILES M., ZHAO J. (2017), Robust ECG signal classification for the detection of atrial fibrillation using novel neural networks, [in:] *Proceedings of 2017 Computing in Cardiology Conference (CinC)*, **44**, doi: [10.22489/cinc.2017.066-138](https://doi.org/10.22489/cinc.2017.066-138).
24. XU X., HUANG C., WU C., WANG Q., ZHAO L. (2014), Graph learning based speaker independent speech emotion recognition, *Advances in Electrical and Computer Engineering*, **14**(2): 17–22, doi: [10.4316/aee.2014.02003](https://doi.org/10.4316/aee.2014.02003).
25. YAN J., WANG X., GU W., MA L. (2013), Speech emotion recognition based on sparse representation, *Archives of Acoustics*, **38**(4): 465–470, doi: [10.2478/aoa-2013-0055](https://doi.org/10.2478/aoa-2013-0055).
26. ZHOU Q. *et al.* (2021), Cough recognition based on Mel-spectrogram and convolutional neural network, *Frontiers in Robotics and AI*, **8**: 1–7, doi: [10.3389/frobt.2021.580080](https://doi.org/10.3389/frobt.2021.580080).
27. ZOU C., HUANG C., HAN D., ZHAO L. (2011), Detecting practical speech emotion in a cognitive task, [in:] *Proceedings of 20th International Conference on Computer Communications and Networks (ICCCN)*, pp. 1–5, doi: [10.1109/icccn.2011.6005883](https://doi.org/10.1109/icccn.2011.6005883).

Research Paper

Mismatch Negativity as a Tool to Detect
the Intensity Just Noticeable DifferenceBusemnaz AVŞAR AKSU^{(1)*}, Didem ŞAHİN CEYLAN⁽²⁾, Gökçe GÜLTEKİN^{(2),(3)}⁽¹⁾ *Department of Neuroscience, Graduate School of Health Sciences, Üsküdar University
Istanbul, Turkey*⁽²⁾ *Department of Audiology, Faculty of Health Sciences, Üsküdar University
Istanbul, Turkey*⁽³⁾ *Department of Audiology, Language and Speech Disorders,
Institute of Graduate Studies, Istanbul University-Cerrahpaşa
Istanbul, Turkey**Corresponding Author e-mail: busemnaz.avsaraksu@uskudar.edu.tr

(received September 20, 2023; accepted June 11, 2024; published online October 1, 2024)

Mismatch negativity (MMN) essentially reflects auditory change detection. Although auditory change detection can potentially be assessed through behavioral auditory testing methods, the increased reliability of objective methods, such as MMN, makes them more valuable. The aim of this study was to detect and compare the intensity just noticeable difference using the MMN and a behavioral method. The level at which the intensity difference between the frequent stimulus and the infrequent stimulus was the lowest and the MMN wave elicited was accepted as the MMN threshold. A total of 60 subjects, 30 females (mean age 21.70, SD = 1.91 years) and 30 males (mean age 22.77, SD = 3.01), aged 20–30 years, were included in the study. In the whole sample, a significant difference was found between MMN thresholds obtained from the right ear side and MMN thresholds obtained from the left ear side, regardless of sex ($p < 0.05$). In the comparison of the values obtained using the behavioral method and MMN, no significant difference was found for either the right or the left side in both sexes ($p > 0.05$). The results showed that the values determined by the behavioral method and MMN on both the right and left ear sides were similar in both sexes.

Keywords: behavioral measurement; intensity just noticeable difference; auditory discrimination; mismatch negativity; loudness discrimination.



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

There are challenges in behavioral measurement methods, such as understanding the task, maintaining attention on the task, and changes in motivation. For this reason, research is needed to determine whether auditory evoked potentials can be an alternative to behavioral measurement methods. Studies conducted in this context have shown that methods using auditory evoked potentials provide objective and reliable findings when measuring central auditory processing abilities, such as auditory discrimination (HARRIS *et al.*, 2007). The mismatch negativity (MMN) is an objective and electrophysiological measurement that reflects the

neural encoding of the dissonance occurring when infrequent stimuli with different physical properties are presented against information stored in sensory memory, where the physical properties of frequent stimuli are stored (JOHNSON *et al.*, 2021; PAKARINEN *et al.*, 2007; SENDESEN *et al.*, 2022). Acoustic parameters that are changed for infrequent stimuli include frequency, intensity, location, and duration (RAO *et al.*, 2020).

Intensity is the most fundamental property of an auditory signal. It influences a multitude of functions, from sound source localization to the neural processing of the signal (RECANZONE, BECKERMAN, 2004). It has been reported that central auditory processing regions play a role in intensity discrimination indepen-

dent of performance (BELIN *et al.*, 1998; BENCH, 1969). There are studies suggesting that the coding of loudness in the central auditory system may be based on the firing rate of afferent neurons (MASLIN *et al.*, 2015). Therefore, it is valuable to measure intensity discrimination performance using auditory evoked late latency responses, which bring us closer to the central auditory system, and to examine the compatibility of these results with behavioral measures. In our review of the literature, we found that loudness discrimination performance has been investigated using objective methods (HARRIS *et al.*, 2007; HE *et al.*, 2012). However, in our study, we specifically investigated and compared the smallest level of intensity that we can discriminate using MMN and behavioral methods.

The MMN serves as an auditory processing index (RANA *et al.*, 2022). It has been proposed that MMN responses arise in the primary and secondary auditory cortices, as well as the frontal cortex (BONETTI *et al.*, 2018). Reports indicate sexual dimorphism in the auditory cortex (BERCHICCI *et al.*, 2021). Sexual dimorphism is known to impact event-related potential (ERP) results (BERCHICCI *et al.*, 2021; IKEZAWA *et al.*, 2008). Dimorphic effects on ERPs are typically analyzed through comparisons of latency and amplitude (NAGY *et al.*, 2003). No previous studies have compared male and female responses to MMN stimuli with minimal reductions (1 dB) in the difference between frequent and infrequent stimuli. Therefore, we conducted a study examining the smallest intensity difference (i.e., the difference in intensity level between frequent and infrequent stimuli) occurring during MMN and compared it between both sexes.

Previous studies investigated the intensity difference using MMN and compared it to the behavioral method, but the intensity change was not made in small steps as in our study (NÄÄTÄNEN *et al.*, 2004; O'REILLY, 2021; PAKARINEN *et al.*, 2007; RINNE *et al.*, 2006). In our study, we aim to identify the lowest possible intensity difference that can produce MMN. Therefore, we examine changes in intensity in smaller steps (1 dB) to detect thresholds.

The primary aim of this study is to investigate whether MMN is a valid tool for assessing intensity just noticeable difference (intensity JND). Our main hypothesis is that MMN can be used as an objective tool for intensity JND. In addition to the primary aim, it is of interest to determine whether there is a significant difference between men and women in the values of the smallest intensity change that can be detected both electrophysiologically and behaviorally.

2. Method

This study was approved by the Üsküdar University Ethics Committee (61351342/February 2021-33). Young adults with normal hearing, aged of 20–30 years,

were selected based on otoscopic examination, audiometric evaluations (MEHTA, OXENHAM, 2018; WILEY *et al.*, 1987), tympanometric examination (MISHRA *et al.*, 2021), acoustic reflex thresholds (WILEY *et al.*, 1987), and distortion product otoacoustic emission (HARRIS, 1990) among all volunteers. A total of 60 normal-hearing participants were included in our study (30 women – mean age 21.70, SD = 1.91, and 30 men – mean age 22.77, SD = 3.01), first enrolled in the MMN session, followed by the behavioral session.

The MMN was recorded with the Eclipse system by the Interacoustic device to detect the minimum amount of change in intensity. Stimuli were presented in a traditional oddball paradigm. The infrequent stimulus was set to differ only in intensity from the frequent stimulus, and all other parameters remained the same. The stimulus parameters are demonstrated in Table 1. A sampling number of 200 was used for each measurement. The wave with the most negative peak occurring 150 ms–250 ms after the onset of the stimulus was accepted as the MMN. MMN amplitudes were calculated by placing one reference cursor on the positive peak of the preceding wave and the other on the MMN trough.

Table 1. Stimulus parameters.

Frequency	1000 Hz tone burst
Time	70 ms (10–50–10)
Rate	0.7 Hz
Polarity	Rarefaction
Gain	±80 μ V
High pass filter	1.0 Hz (6 dB/oct)
Low pass filter	100 Hz
Frequent stimulus rate	80 %
Infrequent stimulus rate	20 %
Stop criteria/number of stimuli	100
Reject rate	<20 %
Impedance	2 k Ω –3 k Ω

In the study, the frequent stimulus was set at 60 dB nHL for a comfortable audible level. The infrequent stimulus was presented with 1 dB decrements, starting from 70 dB nHL (+10 dB), 69 dB nHL (+9 dB), ..., 66 dB nHL (+6 dB), 65 dB nHL (+5 dB), ..., 61 dB nHL(+1 dB), up to the final level at which the MMN wave could be elicited. In order to ensure the reliability of the waveform, two measurements were taken. Specifically, if the MMN waveform was elicited again in the second measurement, the wave was considered reliable. The level at which the intensity difference between the frequent stimulus and the infrequent stimulus was the lowest and the MMN wave was elicited again was accepted as the MMN threshold, representing the intensity JND that the participants could objectively detect.

The areas on the head where the electrodes were to be placed were cleaned and the reference electrodes

were placed on the right and left mastoids, the active electrode on the hairline (Fz), and the ground electrode on the middle region of the forehead (Fpz).

Stimuli were presented unilaterally with 3M E-A-R tone insert earphones. Participants were seated in a chair with their eyes open and were instructed not to pay attention to the stimulus. Instead, they were asked to count from 100 to 0 in steps of 3 (e.g., 100–97–94, ...). Since ocular monitorization was not available in the clinic, participants were asked not to move their eyes as much as possible to keep the reject rate low.

Behavioral measurements were conducted in the silent cabin where the audiological evaluation was made. Pairs of stimuli, each with a length of 500 ms and a frequency of 1000 Hz, were presented to the participants. They were asked to indicate the pairs as “different” or “the same” in terms of their intensity. The intensities of the stimuli were sent in such a way that the difference between them was high at the beginning, so that individuals could adapt to the experiment. The differences were then gradually decreased, as in the MMN protocol.

The pairs of stimuli were first sent as 60 dB HL–70 dB HL, 60 dB HL–69 dB HL, ..., 60 dB HL–66 dB HL, 60 dB HL–65 dB HL, ..., 60 dB HL–61 dB HL, with the first stimulus remaining constant and the second stimulus was decreased in 1 dB steps. The behavioral threshold, which represented the intensity JND that participants could detect, was defined as 1 dB above the level at which participants identified the pairs as the same.

Stimuli used in both electrophysiological and behavioral tests were presented monaurally and the two ear sides were tested separately.

2.1. Statistical analysis

IBM SPSS Statistics (v.25) was used in all statistical analyses. The Mann–Whitney *U*-test was used to compare numerical measurements based on sex, and the Wilcoxon signed-rank test was used for comparisons between the ear sides of the same individuals. The level of significance was set at $\alpha = 0.05$. Correlation analyses were conducted to examine the relationship between electrophysiological thresholds and behavioral threshold measures. Pearson correlation analysis was used for normally distributed data, and Spearman correlation analysis was used for non-normally distributed data. The significance value was accepted as $p < 0.05$. Correlation analyses were evaluated according to the classification proposed by EVANS (1996).

3. Results

According to the MMN test, the mean amplitude value obtained from the right ear side was 2.92 μ V (SD = 1.34), the mean latency value was 223.37 msn

(SD = 35.57), the mean amplitude value obtained from the left ear side was 2.95 μ V (SD = 1.37), and the mean latency value was 211.77 msn (SD = 32.49).

There was no significant difference between women (mean = 2.91 μ V, SD = 1.26) and men (mean = 2.93 μ V, SD = 1.43) in terms of MMN amplitude values obtained from the right side ($p = 0.988$). Likewise, there was no significant difference between women (mean = 220.46 msn, SD = 39.79) and men (mean = 226.26 msn, SD = 31.20) in terms of MMN latency values obtained from the right side ($p = 0.871$). Furthermore, there was no significant difference between women (mean = 2.91 μ V, SD = 1.26) and men (mean = 2.93 μ V, SD = 1.43) in terms of MMN amplitude values obtained from the left side ($p = 0.988$). Moreover, there was no significant difference between women (mean = 213.33 msn, SD = 36.83) and men (mean = 210.20 msn, SD = 28.03) in terms of MMN latency values obtained from the left side ($p = 0.790$).

When the MMN threshold values were compared between the right and left sides, the values obtained from the right side (mean = 5.41 dB, SD = 1.75) were significantly lower than those from the left side (mean = 5.88 dB, SD = 1.77) ($p < 0.01$).

When the behavioral threshold values were compared between the right and left sides, no significant difference was found between the values obtained from the right side (mean = 5.70 dB, SD = 2.05) and those from the left side (mean = 5.80 dB, SD = 2.04) ($p = 0.635$).

When the MMN threshold values were compared between the sexes, there was no significant difference between women (mean = 5.26 dB, SD = 1.85) and men (mean = 5.56 dB, SD = 1.67) on the right side ($p = 0.514$). Likewise, no significant difference was found on the left side between women (mean = 5.66 dB, SD = 1.72) and men (6.10 dB, SD = 1.82), ($p = 0.349$). In the comparison of the behavioral threshold values by sex, there was no significant difference on the right side between women (mean = 5.83 dB, SD = 2.29) and men (mean = 5.56 dB, SD = 1.81), ($p = 0.619$). Likewise, no significant difference was found on the left side between women (mean = 6.03 dB, SD = 2.39) and men (mean = 5.56 dB, SD = 1.61), ($p = 0.380$).

Furthermore, there was no significant difference between the MMN threshold (mean = 5.41 dB, SD = 1.75) and the behavioral threshold (mean = 5.70 dB, SD = 2.05) values obtained from the right side ($p = 0.359$). Likewise, there was no significant difference between the MMN threshold (mean = 5.88 dB, SD = 1.77) and the behavioral threshold (mean = 5.80 dB, SD = 2.04) values obtained from the left side ($p = 0.502$) (Fig. 1).

An exploratory analysis was performed due to the observation of a majority of perceptible values above 5 dB when the data were collected. The results obtained from the right side showed that the behav-

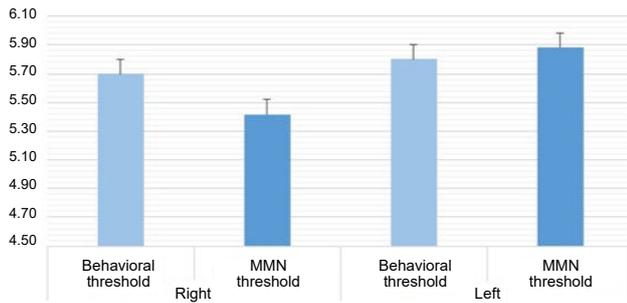


Fig. 1. Comparison of behavioral threshold and MMN threshold.

ioral thresholds (mean = 7.56 dB, SD = 1.50) were significantly higher than the MMN thresholds (mean = 5.25 dB, SD = 1.29) in women with a behavioral threshold above 5 dB on their right side ($p < 0.05$). Likewise, the behavioral thresholds (mean = 7.70 dB, SD = 1.56) were significantly higher than the MMN thresholds (mean = 6.00 dB, SD = 1.41) in men with a behavioral threshold above 5 dB on their right side ($p < 0.01$) (Table 2).

Table 2. Comparison of MMN and behavioral threshold values in male and female participants with behavioral threshold above 5 dB on the right side.

Sex	Threshold	n	Mean	SD	Z	P^1
Women	Behavioral threshold	16	7.56	1.50	-2.44	0.014*
	MMN threshold	16	5.25	1.29		
Men	Behavioral threshold	10	7.70	1.56	-3.15	0.002**
	MMN threshold	10	6.00	1.41		

¹ Wilcoxon test; * $p < 0.05$; ** $p < 0.01$.

Moreover, the results obtained from the left side showed that no significant difference was found between the behavioral thresholds (mean = 7.20 dB, SD = 1.93) and the MMN thresholds (mean = 7.00 dB, SD = 1.25) in women with a behavioral threshold above 5 dB on their left side ($p = 0.587$). However, the results obtained from the left side indicated that the behavioral thresholds (mean = 7.20 dB, SD = 1.93) were significantly higher than the MMN thresholds (mean = 5.70 dB, SD = 1.68) in men with a behavioral threshold above 5 dB on their right side ($p < 0.05$) (see Table 3).

Table 3. Comparison of MMN and behavioral threshold values in male and female participants with behavioral threshold above 5 dB on the left side.

Sex	Threshold	n	Mean	SD	Z	P^1
Women	Behavioral threshold	20	6.80	1.14	-0.54	0.587
	MMN threshold	20	7.00	1.25		
Men	Behavioral threshold	15	7.20	1.93	-2.35	0.018*
	MMN threshold	15	5.70	1.68		

¹ Wilcoxon test; * $p < 0.05$.

As a result of the correlation analysis, on the right side, no significant correlation was found between

MMN thresholds and behavioral thresholds in all participants ($p = 0.509$); however, on the left side, a significant but weak correlation was observed between MMN thresholds and behavioral thresholds in all participants ($r = 0.259$, $p = 0.044$).

4. Discussion

In our study, the mean behavioral threshold was 5.7 dB (SD = 2.05) for the right side and 5.8 dB (SD = 2.04) for the left side. The threshold values obtained for the right and the left sides were similar. When we look at the studies in the literature investigating intensity discrimination using behavioral measures, it is observed that the intensity JND is between 3 dB–5 dB (DORTA *et al.*, 2017; HE *et al.*, 1998). Participants whose level of intensity JND was greater than 5 dB were statistically evaluated separately. The proportion of participants with a behavioral threshold higher than 5 dB, regardless of right-left side and sex, was 58.30 %. When analyzed in terms of sex, women constituted 60 % of this group and men accounted for 41.20 %.

Moreover, in a study conducted with children, auditory discrimination was investigated with frequency variation and it was also examined whether there was a sex effect on the results. It was found that boys were better than girls at discriminating frequencies (ZALTZ *et al.*, 2014). In another study evaluating magnitude estimations, the sample consisted of 22 young adults and it was found that there was no sex difference in the results (WEDER *et al.*, 2020). Consistent with these findings, our results showed that intensity JND for both the right and left sides was similar for men and women. That is to say no sex difference was observed. In addition, the behavioral thresholds for the right side and left side showed similarity across the entire sample, regardless of sex.

Several studies investigated auditory discrimination with different variables such as frequency, duration, and intensity by both behavioral and electrophysiological methods. These studies have mostly shown that electrophysiological and behavioral methods are related in terms of the frequency, intensity, and duration parameters. HE *et al.* (2012) tested the auditory discrimination performance with electrophysiological and behavioral methods and compared the results. In their study, the acoustic change complex (ACC) is used as an electrophysiological method. For each of the frequency, intensity and duration variables, the minimum levels that can elicit the ACC were accepted as objective thresholds. As a result, the mean of the objective threshold for intensity discrimination was obtained as 2 dB and it was understood that it showed a significant correlation with behavioral thresholds.

Furthermore, in another study, which was conducted to determine whether ACC can be an objective indicator of intensity discrimination in children with

central auditory processing disorder, it was understood that ACC could be an objective tool for detecting the minimum amount of change in intensity as a result of comparing the behavioral discrimination thresholds (KUMAR *et al.*, 2020). In our study, the average objective and behavioral thresholds were obtained in the control group in the range of 2 dB–4 dB, as expected.

In another study, using the *N1–P2* response as the electrophysiological method, behavioral thresholds and *N1–P2* thresholds were compared, and it was shown that they were similar (HARRIS *et al.*, 2007). Electrophysiological tests, as demonstrated in exemplary studies, have proven to be viable alternatives to behavioral methods. In this context, our study, while comparing behavioral and electrophysiological thresholds in terms of intensity JND, showed that behavioral thresholds and MMN thresholds were similar for both the right and left sides.

Studies have found that behavioral hearing tests, which identify cortical areas stimulated by pure tones, show a broad distribution across the cortex (BELIN *et al.*, 1998; BIANCHI *et al.*, 2017; ZHANG *et al.*, 2006). However, the MMN response, which originates from the frontotemporal region, is more specific (FITZGERALD, TODD, 2020; WAGNER *et al.*, 2023). Our study compared behavioral thresholds and MMN thresholds, and found that participants with behavioral thresholds above 5 dB had significantly higher thresholds compared to MMN thresholds. This difference in performance was likely due to cognitive fatigue from the involvement of multiple cortical regions in the behavioral task. Interestingly, on the left side, the behavioral thresholds and MMN thresholds for women were similar (Table 3), which may be due to the small sample size rather than the expected increase in behavioral responses associated with a larger number of cortical regions.

On the left side, there was a weak correlation between MMN thresholds and behavioral thresholds. When we searched the literature, we could not find any grounds for discussion that could explain this result, so we hypothesize that the result was due to the inhomogeneous sample distribution (we used the Mann–Whitney *U*-test). It is anticipated that if the sample were more homogeneous, the significant correlation observed on the left side might also be evident on the right side. In such a case, the MMN could become a reliable tool for objectively detecting individual intensity JND.

BARRETT and FULFS (1998) reported that there was no significant difference in MMN latency values between sexes in healthy young adults when measured with a frequency of 1000 Hz. Our current study showed similar results in terms of latency as there was no significant difference in latency between men and women for both the right and left sides. In another study investigating the effect of sex on MMN in healthy young

adults, it was found that the MMN latencies were significantly longer in females compared to males (TOUFAN *et al.*, 2021). On the other hand, there is no study in which the MMN amplitude was compared by sex. Hence, further research with larger samples is needed to compare both amplitude and latency values of MMN with respect to sex and ear side.

The role of attention in the MMN test is a controversial topic in the literature. There are studies reporting that MMN formation requires active attention (SUSSMAN *et al.*, 2014; RAHNE *et al.*, 2014), while others report that it does not require active attention – such as when participants are watching a film or performing a mental task (IKEZAWA *et al.*, 2008; NÄÄTÄNEN, 1990; 1995; NÄÄTÄNEN, WINKLER, 1999). We developed our methodology based on references supporting the notion that “MMN can occur without active attention”. Furthermore, there is little information in the literature about ERPs elicited by passive listening. This information is crucial for better understanding on how the brain prepares and responds to sounds without an active task. As the result of our study confirms our hypothesis, MMN may have potential for clinical use for assessing JND intensity in populations that are unable to direct active attention to the task and are difficult to test behaviorally, such as children and the elderly.

The reliability of the traditional oddball paradigm is controversial, primarily because the infrequent stimulus differs from the frequent stimulus both physically and numerically (WIENS *et al.*, 2019). Alternative models to the traditional oddball paradigm have been created (RUHNAU *et al.*, 2012; SCHRÖGER, WOLFF, 1996). However, the problem with these models is that the MMN response is considerably reduced for various reasons (JACOBSEN, SCHRÖGER, 2003). The discussion on this subject is ongoing. Since the studies we used in our method were conducted with the traditional oddball paradigm, we preferred to use the traditional oddball paradigm in our study, in which 20 % of the infrequent stimuli and 80 % of the frequent stimuli were presented (SADIA *et al.*, 2013; SUSSMAN, 2007).

Although the use of MMN as an alternative to behavioral methods in auditory discrimination is controversial (SUSSMAN *et al.*, 2014), there was no significant difference between MMN thresholds and behavioral thresholds in our study. In other words, according to our study, MMN can be used as an alternative tool to behavioral methods.

5. Conclusions

In this study, we compared MMN and behavioral thresholds across the entire sample without discriminating the sample in terms of factors that may affect auditory discrimination performance, such as sex, intelligence level, and hemispheric dominance. However,

since a significant portion of the sample had behavioral discrimination thresholds above 5 dB, we performed a comparison of MMN and behavioral thresholds in this subgroup, accounting for sex discrimination. Future studies should analyze a larger sample and consider all the factors mentioned to provide a more comprehensive understanding.

6. Limitations

Since the equipment used for the behavioral study did not have the same setup as the MMN equipment, there was a discrepancy in stimulus duration; while 1000 Hz tone bursts with a duration of 70 ms were used in the MMN, the duration of 1000 Hz tone stimuli could be reduced to a maximum of 500 ms in the behavioral assessment. This variation in stimulus duration was considered a limitation because it introduced a difference in duration between the behavioral test and MMN stimuli.

Acknowledgments

We would like to thank the Department of Audiology at Üsküdar University for their support.

References

1. BARRETT K., FULFS J. (1998), Effect of gender on the mismatch negativity auditory evoked potential, *Journal of the American Academy of Audiology*, **9**(6): 444–451.
2. BELIN P. *et al.* (1998), The functional anatomy of sound intensity discrimination, *Journal of Neuroscience*, **18**(16): 6388–6394, doi: [10.1523/JNEUROSCI.18-16-06388.1998](https://doi.org/10.1523/JNEUROSCI.18-16-06388.1998).
3. BENCH J. (1969), Audio-frequency and audio-intensity discrimination in the human neonate, *International Audiology*, **8**(4): 615–625, doi: [10.3109/05384916909070234](https://doi.org/10.3109/05384916909070234).
4. BERCHICCI M., BIANCO V., DI RUSSO F. (2021), Electrophysiological sign of stronger auditory processing in females than males during passive listening, *Cognitive Neuroscience*, **12**(3–4): 106–111, doi: [10.1080/17588928.2020.1806224](https://doi.org/10.1080/17588928.2020.1806224).
5. BIANCHI F., HJORTKJÆR J., SANTURETTE S., ZATTORRE R.J., SIEBNER H.R., DAU T. (2017), Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians, *NeuroImage*, **163**: 398–412, doi: [10.1016/j.neuroimage.2017.07.057](https://doi.org/10.1016/j.neuroimage.2017.07.057).
6. BONETTI L. *et al.* (2018), Auditory sensory memory and working memory skills: Association between frontal MMN and performance scores, *Brain Research*, **1700**: 86–98, doi: [10.1016/j.brainres.2018.06.034](https://doi.org/10.1016/j.brainres.2018.06.034).
7. DORTA J., MARTÍN J.A., JORGE C. (2017), Intensity threshold: Beyond pure tones, *Estudios de Fonetica Experimental*, **26**: 133–163.
8. EVANS J.D. (1996), *Straightforward Statistics for the Behavioral Sciences*, Thomson Brooks/Cole Publishing Co.
9. FITZGERALD K., TODD J. (2020), Making sense of mismatch negativity, *Frontiers in Psychiatry*, **11**, doi: [10.3389/fpsy.2020.00468](https://doi.org/10.3389/fpsy.2020.00468).
10. HARRIS F.P. (1990), Distortion-product otoacoustic emissions in humans with high frequency sensorineural hearing loss, *Journal of Speech and Hearing Research*, **33**(3): 594–600, doi: [10.1044/jshr.3303.594](https://doi.org/10.1044/jshr.3303.594).
11. HARRIS K.C., MILLS J.H., DUBNO J.R. (2007), Electrophysiologic correlates of intensity discrimination in cortical evoked potentials of younger and older adults, *Hearing Research*, **228**(1–2): 58–68, doi: [10.1016/j.heares.2007.01.021](https://doi.org/10.1016/j.heares.2007.01.021).
12. HE N., DUBNO J.R., MILLS J.H. (1998), Frequency and intensity discrimination measured in a maximum-likelihood procedure from young and aged normal-hearing subjects, *The Journal of the Acoustical Society of America*, **103**(1): 553–565, doi: [10.1121/1.421127](https://doi.org/10.1121/1.421127).
13. HE S., GROSE J.H., BUCHMAN C.A. (2012), Auditory discrimination: The relationship between psychophysical and electrophysiological measures, *International Journal of Audiology*, **51**(10): 771–782, doi: [10.3109/14992027.2012.699198](https://doi.org/10.3109/14992027.2012.699198).
14. IKEZAWA S. *et al.* (2008), Gender differences in lateralization of mismatch negativity in dichotic listening tasks, *International Journal of Psychophysiology*, **68**(1): 41–50, doi: [10.1016/j.ijpsycho.2008.01.006](https://doi.org/10.1016/j.ijpsycho.2008.01.006).
15. JACOBSEN T., SCHRÖGER E. (2003), Measuring duration mismatch negativity, *Clinical Neurophysiology*, **114**(6): 1133–1143, doi: [10.1016/S1388-2457\(03\)00043-9](https://doi.org/10.1016/S1388-2457(03)00043-9).
16. JOHNSON N., SHIJU A.M., PARMAR A., PRABHU P. (2021), Evaluation of auditory stream segregation in musicians and nonmusicians, *International Archives of Otorhinolaryngology*, **25**(01): e77–e80, doi: [10.1055/s-0040-1709116](https://doi.org/10.1055/s-0040-1709116).
17. KUMAR P., SINGH N.K., SANJU H.K., KAVRAPPAPP G.M. (2020), Feasibility of objective assessment of difference limen for intensity using acoustic change complex in children with central auditory processing disorder, *International Journal of Pediatric Otorhinolaryngology*, **137**: 110189, doi: [10.1016/j.ijporl.2020.110189](https://doi.org/10.1016/j.ijporl.2020.110189).
18. MASLIN M.R.D., TAYLOR M., PLACK C.J., MUNRO K.J. (2015), Enhanced intensity discrimination in the intact ear of adults with unilateral deafness, *The Journal of the Acoustical Society of America*, **137**(6): EL408–EL414, doi: [10.1121/1.4914945](https://doi.org/10.1121/1.4914945).
19. MEHTA A.H., OXENHAM A.J. (2018), Fundamental-frequency discrimination based on temporal-envelope cues: Effects of bandwidth and interference, *The Journal of the Acoustical Society of America*, **144**(5): EL423–EL428, doi: [10.1121/1.5079569](https://doi.org/10.1121/1.5079569).
20. MISHRA S.K., RENKEN L., HERNANDEZ M., RODRIGO H. (2021), Auditory development of frequency discrimination at extended high frequencies, *Ear and Hearing*, **42**(3): 700–708, doi: [10.1097/aud.0000000000000972](https://doi.org/10.1097/aud.0000000000000972).

21. NÄÄTÄNEN R. (1990), The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function, *Behavioral and Brain Sciences*, **13**(2): 201–288, doi: [10.1017/S0140525X00078407](https://doi.org/10.1017/S0140525X00078407).
22. NÄÄTÄNEN R. (1995), The mismatch negativity: A powerful tool for cognitive neuroscience, *Ear and Hearing*, **16**(1): 6–18.
23. NÄÄTÄNEN R., PAKARINEN S., RINNE T., TAKEGATA R. (2004), The mismatch negativity (MMN): Towards the optimal paradigm, *Clinical Neurophysiology*, **115**(1): 140–144, doi: [10.1016/j.clinph.2003.04.001](https://doi.org/10.1016/j.clinph.2003.04.001).
24. NÄÄTÄNEN R., WINKLER I. (1999), The concept of auditory stimulus representation in cognitive neuroscience, *Psychological Bulletin*, **125**(6): 826–859, doi: [10.1037/0033-2909.125.6.826](https://doi.org/10.1037/0033-2909.125.6.826).
25. NAGY E., POTTS G.F., LOVELAND K.A. (2003), Sex-related ERP differences in deviance detection, *International Journal of Psychophysiology*, **48**(3): 285–292, doi: [10.1016/S0167-8760\(03\)00042-4](https://doi.org/10.1016/S0167-8760(03)00042-4).
26. O'REILLY J.A. (2021), Can intensity modulation of the auditory response explain intensity-decrement mismatch negativity?, *Neuroscience Letters*, **764**: 136199, doi: [10.1016/j.neulet.2021.136199](https://doi.org/10.1016/j.neulet.2021.136199).
27. PAKARINEN S., TAKEGATA R., RINNE T., HUOTILAINEN M., NÄÄTÄNEN R. (2007), Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP), *Clinical Neurophysiology*, **118**(1): 177–185, doi: [10.1016/j.clinph.2006.09.001](https://doi.org/10.1016/j.clinph.2006.09.001).
28. RAHNE T., PLONTKE S.K., WAGNER L. (2014), Mismatch negativity (MMN) objectively reflects timbre discrimination thresholds in normal-hearing listeners and cochlear implant users, *Brain Research*, **1586**: 143–151, doi: [10.1016/j.brainres.2014.08.045](https://doi.org/10.1016/j.brainres.2014.08.045).
29. RANA F.S., PAPE D., SERVICE E. (2022), The effect of increasing acoustic and linguistic complexity on auditory processing: An EEG study, [in:] *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pp. 4048–4052, doi: [10.21437/Interspeech.2022-10607](https://doi.org/10.21437/Interspeech.2022-10607).
30. RAO A., KOERNER T.K., MADSEN B., ZHANG Y. (2020), Investigating influences of medial olivocochlear efferent system on central auditory processing and listening in noise: A behavioral and event-related potential study, *Brain Sciences*, **10**(7): 1–17, doi: [10.3390/brainsci10070428](https://doi.org/10.3390/brainsci10070428).
31. RECANZONE G.H., BECKERMAN N.S. (2004), Effects of intensity and location on sound location discrimination in macaque monkeys, *Hearing Research*, **198**(1–2): 116–124, doi: [10.1016/j.heares.2004.07.017](https://doi.org/10.1016/j.heares.2004.07.017).
32. RINNE T., SÄRKKÄ A., DEGERMAN A., SCHRÖGER E., ALHO K. (2006), Two separate mechanisms underlie auditory change detection and involuntary control of attention, *Brain Research*, **1077**(1): 135–143, doi: [10.1016/j.brainres.2006.01.043](https://doi.org/10.1016/j.brainres.2006.01.043).
33. RUHNAU P., HERRMANN B., SCHRÖGER E. (2012), Finding the right control: The mismatch negativity under investigation, *Clinical Neurophysiology*, **123**(3): 507–512, doi: [10.1016/j.clinph.2011.07.035](https://doi.org/10.1016/j.clinph.2011.07.035).
34. SADIA G., RITTER W., SUSSMAN E. (2013), Category effects: Is top-down control alone sufficient to elicit the mismatch negativity (MMN) component?, *Biological Psychology*, **92**(2): 191–198, doi: [10.1016/j.biopsycho.2012.10.008](https://doi.org/10.1016/j.biopsycho.2012.10.008).
35. SCHRÖGER E., WOLFF C. (1996), Mismatch response of the human brain to changes in sound location, *NeuroReport*, **7**(18): 3005–3008, doi: [10.1097/00001756-199611250-00041](https://doi.org/10.1097/00001756-199611250-00041).
36. SENDESEN E., ERBIL N., TÜRKÜYLMAZ M.D. (2022), The mismatch negativity responses of individuals with tinnitus with normal extended high-frequency hearing – is it possible to use mismatch negativity in the evaluation of tinnitus?, *European Archives of Oto-Rhino-Laryngology*, **279**(7): 3425–3434, doi: [10.1007/s00405-021-07097-6](https://doi.org/10.1007/s00405-021-07097-6).
37. SUSSMAN E.S. (2007), A new view on the MMN and attention debate, *Journal of Psychophysiology*, **21**(3–4): 164–175, doi: [10.1027/0269-8803.21.34.164](https://doi.org/10.1027/0269-8803.21.34.164).
38. SUSSMAN E.S., CHEN S., SUSSMAN-FORT J., DINCES E. (2014), The five myths of MMN: Redefining how to use MMN in basic and clinical research, *Brain Topography*, **27**(4): 553–564, doi: [10.1007/s10548-013-0326-6b](https://doi.org/10.1007/s10548-013-0326-6b).
39. TOUFAN R., AGHAMOLAEI M., ASHAYERI H. (2021), Differential effects of gender on mismatch negativity to violations of simple and pattern acoustic regularities, *Brain and Behavior*, **11**(8): e2248, doi: [10.1002/brb3.2248](https://doi.org/10.1002/brb3.2248).
40. WAGNER L., LADEK A.S., PLONTKE S.K., RAHNE T. (2023), Electrically evoked mismatch negativity responses to loudness and pitch cues in cochlear implant users, *Scientific Reports*, **13**(1): 2413, doi: [10.1038/s41598-023-29422-1](https://doi.org/10.1038/s41598-023-29422-1).
41. WEDER S., SHOUSHARIAN M., OLIVARES V., ZHOU X., INNES-BROWN H., MCKAY C. (2020), Cortical fNIRS responses can be better explained by loudness percept than sound intensity, *Ear and Hearing*, **41**(5): 1187–1195, doi: [10.1097/AUD.0000000000000836](https://doi.org/10.1097/AUD.0000000000000836).
42. WIENS S., SZYCHOWSKA M., EKLUND R., VAN BERLEKOM E. (2019), Cascade and no-repetition rules are comparable controls for the auditory frequency mismatch negativity in oddball tasks, *Psychophysiology*, **56**(1): e13280, doi: [10.1111/psyp.13280](https://doi.org/10.1111/psyp.13280).
43. WILEY T.L., OVIATT D.L., BLOCK M.G. (1987), Acoustic-immittance measures in normal ears, *Journal of Speech and Hearing Research*, **30**(2): 161–170, doi: [10.1044/jshr.3002.161](https://doi.org/10.1044/jshr.3002.161).
44. ZALTZ Y., ROTH D.A.-E., GOVER H., LIRAN S., KISHON-RABIN L. (2014), The effect of gender on a frequency discrimination task in children, *Journal of Basic and Clinical Physiology and Pharmacology*, **25**(3): 293–299.
45. ZHANG Y.T., GENG Z.J., ZHANG Q., LI W., ZHANG J. (2006), Auditory cortical responses evoked by pure tones in healthy and sensorineural hearing loss subjects: Functional MRI and magnetoencephalography, *Chinese Medical Journal*, **119**(18): 1548–1554, doi: [10.1097/00029330-200609020-00008](https://doi.org/10.1097/00029330-200609020-00008).

Research Paper

A Real-Time Key-Finding Algorithm Based on the Signature of Fifths

Paulina KANIA⁽¹⁾, Dariusz KANIA⁽²⁾, Tomasz ŁUKASZEWICZ^{(2)*}⁽¹⁾ Faculty of Physics, Adam Mickiewicz University
Poznań, Poland⁽²⁾ Faculty of Automatic Control, Electronics and Computer Science
Silesian University of Technology
Gliwice, Poland*Corresponding Author e-mail: tlukaszewicz@polsl.pl

(received April 6, 2024; accepted August 28, 2024; published online October 21, 2024)

The signature of fifths is a special kind of music representation technique enabling acquisition of musical knowledge. The technique is based on geometrical relationships existing between twelve polar vectors inscribed in the circle of fifths, which represent individual pitch-classes detected in a given composition. In this paper we introduce a real-time key-detection algorithm which utilizes the concept of the signature of fifths. We explain how to create the signature of fifths and how to derive its descriptors required by the algorithm, i.e., the main directed axis of the signature of fifths, the major/minor mode axis, the characteristic vector of the signature of fifths, the characteristic angle of the signature of fifths, and the angle of the major/minor mode. We performed a series of experiments to test the algorithm's effectiveness. The results were compared with those obtained using key-detection approaches based on key-profiles. All experiments were conducted using works composed by J.S. Bach, F. Chopin, and D. Shostakovich. The distinctive features of the presented algorithm, with respect to the considered key-detection approaches, are computational simplicity and stability of the decision-making process.

Keywords: music key-detection; tonality; music information retrieval; music classification.



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

Tonality, or key-detection, algorithms utilize various techniques. The foundations of tonality analysis date back to the time of Pythagoras, who defined numerical relationships between consonant and dissonant sounds. Major input to the tonality analysis is also ascribed to Leonard Euler, who in his tone network, commonly referred to as Tonnetz, “tied together” the tones that make up the major and minor chords. The fifths and thirds (major and minor) intervals present in Tonnetz constitute the founding elements of the harmonic relationships among major and minor scales. Similar interval relationships are reflected in the Longuet-Higgins tonal maps (LONGUET-HIGGINS, 1962a; 1962b). Harmonic networks also constitute the basis for various 3D spiral array models (SHEPARD, 1982; CHEW, 2000; 2007). Such models al-

low us to associate individual pitch-classes present in a given music piece with particular locations on the spiral, enabling chord and key recognition (MAUCH *et al.*, 2010; OSMALSKYJ *et al.*, 2012; SIGTIA *et al.*, 2015; CHUAN, CHEW, 2005; 2007).

More sophisticated models for representing the content of musical works have recently been proposed (HARTE *et al.*, 2006; BERNARDES *et al.*, 2016; HERREMANS, CHEW, 2019; TYMOCZKO, 2006; 2011). They have been used, for example, to detect harmonic changes (HARTE *et al.*, 2006; BOULANGER-LEWANDOWSKI *et al.*, 2013; CHEN, SU, 2018; JACOBY *et al.*, 2015; HORI *et al.*, 2017; NI *et al.*, 2013; PEISZER *et al.*, 2008; WU, LI, 2018), represent chords geometrically for visualization (TYMOCZKO, 2006; 2011; CANCINO CHACÓN *et al.*, 2014; SAPP, 2001), and assess changes in distribution of pitch-classes present in compositions from different epochs (YUST, 2019). Such

models have also been utilized in the algorithms for recognition of musical genres (ANGLADE *et al.*, 2010, PÉREZ-SANCHO *et al.*, 2010) and have been applied to the evaluation of the chord structure (BERNARDES *et al.*, 2016). Other applications of tonal models include generating structured music with constrained patterns, shaping the harmonic structure of musical pieces (ROIG *et al.*, 2014), and assessment and creation of tension in composition fragments (CHAPIN *et al.*, 2010; YANG *et al.*, 2021). Naturally, such models also find utility in computer-aided composition software (HUANG *et al.*, 2016; SABATHÉ *et al.*, 2017). In recent years, many tonal analysis solutions implementing artificial intelligence (FOSCARIN *et al.*, 2021; DAWSON, 2018; DENG, KWOK, 2017) or machine learning (KORZENIOWSKI, WIDMER, 2017; MASADA, BUNESCU, 2017; MCFEE, BELLO, 2017; ZHOU, LERCH, 2015) techniques have been proposed.

There are many ways to represent the content of a musical work. One of the most popular representations is Euler's Tonnetz, which illustrates chord relationships of the harmonic triad in 2D space. The spiral array model of CHEW (2000), which depicts chords in 3D space, is a more advanced approach. Other approaches introduce more dimensions, e.g., the tonal centroid 6D space of HARTE *et al.* (2006), or the solution provided by BERNARDES *et al.* (2016) which employs a space spanning as many as 12 dimensions. Increase in the dimensionality of the proposed models results from the constant quest for new ways to improve the accuracy of the analysis of musical works. However, improvement of computer analysis should not be sought only in the implementation of more and more complicated solutions. The signature of fifths is one example of a simple yet effective concept that can make a significant contribution to the development of algorithmic key-detection methods. Creation of the signature of fifths corresponding to a given musical piece (or its fragment) enables finding its key via analysis of the geometrical arrangement of the polar vectors comprising the signature. Details of this procedure are presented later in this paper.

The major-minor tonality is inherently associated with Western music. An important current research problem is algorithmic determination of a musical piece's key (BAUMANN, 2021; BERNARDES *et al.*, 2017; FOSCARIN *et al.*, 2021; KORZENIOWSKI, WIDMER, 2018; NÁPOLES LÓPEZ *et al.*, 2019; 2020; QUINN, WHITE, 2017; TOIVAINEN, KRUMHANSL, 2003). Input data used in key-finding algorithms is either in symbolic (e.g., MIDI, MusicXML) or audio (e.g., wav, mp3) format (BAUMANN, 2021; CHUAN, CHEW, 2005; GEBHARDT *et al.*, 2018; PEETERS, 2006; PAPADOPOULOS, PEETERS, 2012; RAPHAEL, STODDARD, 2004; WEISS, 2013; WU, LI, 2018). The computationally simplest key-detection approaches utilize the so-called key-profiles (HERREMANS, CHEW, 2019;

KORZENIOWSKI, WIDMER, 2017; ALBRECHT, SHANAHAN, 2013; GOMEZ, HERRERA, 2004; KANIA, KANIA, 2019; KANIA *et al.*, 2022; KRUMHANSL, KESSLER, 1982; KRUMHANSL, 1990; TEMPERLEY, 2004; TEMPERLEY, MARVIN, 2008). In the most basic scenario, determination of the key is based on searching for the maximum correlation coefficient of a given fragment of the analyzed composition with the 12 major and 12 minor key-profiles. It is still unclear which family of key-profiles and which fragment of a music piece are most appropriate for the analysis purposes. Sometimes the methods based on local-key estimation are also considered (NÁPOLES LÓPEZ *et al.*, 2020).

There are many families of key-profiles (HERREMANS, CHEW, 2019; KORZENIOWSKI, WIDMER, 2017; ALBRECHT, SHANAHAN, 2013; GOMEZ, HERRERA, 2004; KRUMHANSL, KESSLER, 1982; KRUMHANSL, 1990; TEMPERLEY, 2004; TEMPERLEY, MARVIN, 2008). They were created using a variety of methods, ranging from extensive experimental research based on cognitive psychology (KRUMHANSL, KESSLER, 1982; KRUMHANSL, 1990) to computationally intensive statistical/probabilistic analyses (AARDEN, 2003; BELLMANN, 2005; TEMPERLEY, 2004) anchored in the theory of music. In some cases, creation of new key-profiles resulted from experiences with well-established key-profiles. For example, analysis of the Krumhansl-Kessler profiles inspired Temperley to propose a new family of key-profiles (TEMPERLEY, 2004; TEMPERLEY, MARVIN, 2008). Temperley's proposal was backed with advanced models based on probabilistic reasoning. In some cases, key-profiles were created based on the analysis of numerous audio files (CHUAN, CHEW, 2014; GOMEZ, HERRERA, 2004; KORZENIOWSKI, WIDMER, 2018). Particularly interesting are the key-profiles developed with the use of artificial intelligence techniques (ALBRECHT, SHANAHAN, 2013). Experiments have proven their high key-detection efficacy (KANIA, KANIA, 2019).

Although the correlational approach to music key-detection based on key-profiles has low computational complexity, it is possible to create simpler solutions. Reducing the number of multiplication operations is usually a good way to speed up an algorithm. In this respect, the key-detection algorithm based on the signature of fifths, presented in (KANIA, KANIA, 2019), is much simpler than its strictly correlational alternatives. The simplification results from the fact that the correlation coefficients are calculated only for two relative key-profiles, not all 24 of them. In (KANIA *et al.*, 2021a), it was shown that the signature of fifths can also be utilized to determine the key signature of a given piece of music without calculating any correlation coefficient. The authors' proposed an algorithm that inspired the search for a simplified key-detection method and the study discussed in this paper. Computational simplification of the key-detection process is

particularly important when it comes to solutions implemented in hardware, e.g., for instruments presenting musical notation in a real-time manner.

The aim of this paper is to present an original algorithm for real-time key determination based solely on the signature of fifths. The novelty of the proposed approach lies in the assessment of the structure of the signature of fifths. Essentially, the method boils down to the analysis of the geometrical relationships existing between the so-called characteristic vector of the signature of fifths and the pair of directed axes – the main directed axis of the signature of fifths and the major/minor mode axis. The algorithm takes a symbolic description of the piece as input. Therefore, its application is justified only in the context of equal-temperament tuning, i.e., when all keys are equivalent, so there are no better or worse-sounding keys, as in the case of mean-tone or non-temperament tuning. It is assumed that the analyzed pieces are tonal works, which is not always true, especially in contemporary music since the mid-19th century. The essence of the algorithm lies in searching for the most populous set of notes forming a pentatonic scale consisting of steps I, II, III, V, and VI in the major mode and I, III, IV, V, and VII in the Aeolian minor mode (both consist of any five neighboring keys in the circle of fifths). Additionally, the durations of pitch classes are analyzed, but no additional information is considered, such as the significance of individual notes in chords, which impacts listeners' perception.

In the next section of the article, the basic concepts behind the proposed key-detection algorithm are presented. In Sec. 3, the algorithm is thoroughly described. The ideas presented in Secs. 2 and 3 are supported with simple examples. Section 4 presents the results of the carried out experiments, along with a discussion focused on identifying the strengths and weaknesses of the proposed algorithm. The article ends with a summary of the conducted study.

2. Theoretical background

A musical piece can typically be represented with tones belonging to twelve pitch-classes: C, C \sharp /D \flat , D, D \sharp /E \flat , E, F, F \sharp /G \flat , G, G \sharp /A \flat , A, A \sharp /B \flat , B. Let X be the set of durations of individual pitch-classes comprising a given fragment of a musical piece (1):

$$X = \{x_C, x_{C\sharp/D\flat}, x_D, x_{D\sharp/E\flat}, x_E, x_F, x_{F\sharp/G\flat}, x_G, x_{G\sharp/A\flat}, x_A, x_{A\sharp/B\flat}, x_B\}. \quad (1)$$

The vector \mathbf{K} , which represents the normalized aggregate durations of pitch-classes corresponding to the analyzed fragment of music, is given by:

$$\mathbf{K} = [k_A, k_D, k_G, k_C, k_F, k_{B\flat}, k_{E\flat}, k_{A\flat}, k_{D\flat}, k_{G\flat/F\sharp}, k_B, k_E], \quad (2)$$

where

$$k_i = \frac{x_i}{\max(x_A, x_D, x_G, x_C, x_F, x_{B\flat}, x_{E\flat}, x_{A\flat}, x_{D\flat}, x_{G\flat/F\sharp}, x_B, x_E)}, \quad (3)$$

and

$$i \in \{A, D, G, C, F, B\flat, E\flat, A\flat, D\flat, G\flat/F\sharp, B, E\}.$$

The values comprising the vector \mathbf{K} are ordered in accordance with the succession of the pitch-classes in the circle of fifths, beginning from the A tone and continuing counter-clockwise.

DEFINITION 1 (KANIA, KANIA, 2019):

The signature of fifths is a set of twelve polar vectors $\{\mathbf{k}_i : i = A, D, G, \dots, E\}$ whose coordinates (r_i, ϕ_i) fulfill the following conditions:

- the length of each vector is equal to the normalized corresponding component of a given pitch-class vector \mathbf{K} , i.e., $r_i = |\mathbf{k}_i| = k_i$;
- the direction of each vector is determined with the following relationship: $\phi_i = j \cdot 30^\circ$, where $j = 0|_{i=A}$, $j = 1|_{i=D}$, and so on.

Example 1 (KANIA, 2021):

Let us create the signature of fifths for the first bar of J.S. Bach's Prelude No. 1, BWV 846, whose musical notation was illustrated in Fig. 1.



Fig. 1. First bar of Bach's Prelude No. 1, BWV 846.

The vector \mathbf{K} corresponding to the tones presented in Fig. 1 can be expressed as:

$$\mathbf{K} = [0 \ 0 \ 0.2 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0.9], \quad (4)$$

whereas the signature of fifths associated with the above vector is shown in Fig. 2.

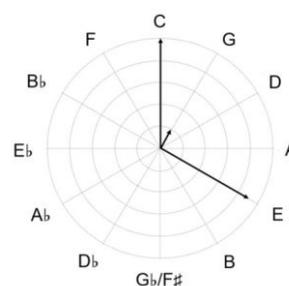


Fig. 2. Signature of fifths obtained for the fragment of the prelude shown in Fig. 1.

We can define a number of directed axes of the circle of fifths $Y \rightarrow Z$, which connect two opposite

pitch-classes. A given axis points from Y towards Z , where pair $(Y; Z) \in (C, F\sharp); (F, B); (B\flat, E); (E\flat, A); (A\flat, D); (D\flat, G); (F\sharp, C); (B, F); (E, B\flat); (A, E\flat); (D, A\flat); (G, D\flat)$. The value $[Y \rightarrow Z]$ is called the characteristic value of the directed axis $Y \rightarrow Z$. It is equal to $\mathbf{K}_R - \mathbf{K}_L$, where \mathbf{K}_R and \mathbf{K}_L are the sums of the lengths of the vectors comprising the signature of fifths, located on the right and left sides of the directed axis $Y \rightarrow Z$, respectively.

DEFINITION 2 (KANIA, KANIA, 2019):

The directed axis of the signature of fifths $Y \rightarrow Z$, for which $[Y \rightarrow Z]$ assumes the maximum value is called the main directed axis of the signature of fifths (MDASF).

DEFINITION 3:

The polar vector obtained as the sum of vectors comprising the signature of fifths is called the characteristic vector of the signature of fifths (CVSF).

The position of the characteristic vector of the signature of fifths can be described with the Cartesian coordinates (x, y) depicting its end or by providing polar coordinates (r_{SF}, ϕ_{SF}) .

DEFINITION 4:

The angle ϕ_{SF} is called the characteristic angle of the signature of fifths.

The MDASF, CVSF, and ϕ_{SF} corresponding to the signature of fifths presented in Fig. 2, are shown in Fig. 3. The outline of the plot illustrated in Fig. 3 was supplemented with the characteristic values associated with individual directed axes. The maximum value (2.1) indicates the direction of the MDASF, which is $B \rightarrow F$ in the considered case.

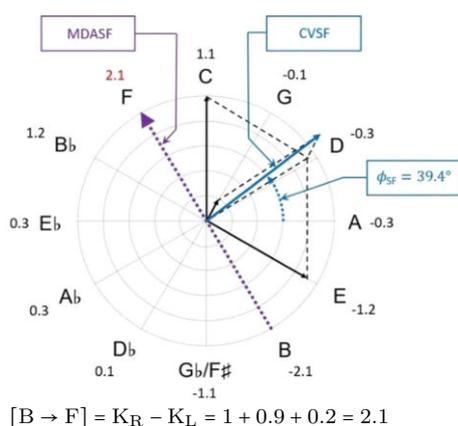


Fig. 3. Signature of fifths supplemented with the MDASF, CVSF, and ϕ_{SF} .

Let us assume that $Y \downarrow Z$ is the major/minor mode axis, which is perpendicular to the $Y \rightarrow Z$ axis. Its tip indicates the tone, which is 90° away in clockwise

direction from the tone indicated by the $Y \rightarrow Z$ axis. Let us then depict the inclination of the major/minor mode axis $Y \downarrow Z$ with respect to the abscissa as the angle ϕ_1 .

DEFINITION 5:

The angle between the CVSF and the major/minor mode axis is called the major/minor mode angle, depicted with the symbol ϕ_m . It is calculated as $\phi_m = \phi_{SF} - \phi_1$.

Figure 4 helps to clarify the way in which the major/minor mode angle corresponding to the signature of fifths presented in Fig. 2 was obtained.

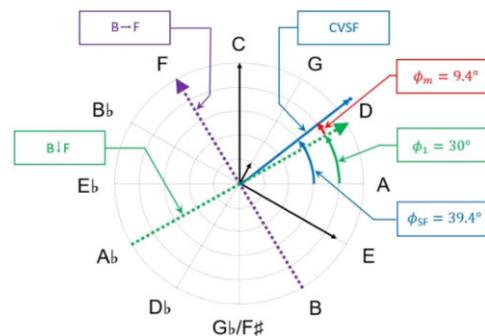


Fig. 4. Clarification of the way in which the angle of the major/minor mode ϕ_m is determined (for the signature of fifths presented in Fig. 2).

Determination of the key of a music composition becomes possible via computation of the major/minor mode angle ϕ_m , relating the direction of the major/minor mode axis and the direction of the characteristic vector of the signature of fifths.

3. Algorithm

In this section we present a real-time key-finding algorithm based on the signature of fifths. It is a simplified version of the algorithm shown by KANIA and KANIA (2019). It determines the major/minor mode of the analyzed piece of music via application of new descriptors of the signature of fifths, i.e., the major/minor mode axis, the characteristic angle of the signature of fifths, and the major/minor mode angle. The key-detecting procedure consists of several steps which lead to the calculation of the major/minor mode angle. In general, the value of this angle can be positive, negative, or zero. It is strictly associated with the key chosen from the obtained pair of two relative keys. A positive value indicates the major key whereas a negative value indicates the minor key. When the value of the major/minor mode angle is zero, i.e., the major/minor mode axis coincides with the characteristic vector of the signature of fifths, the analyzed fragment of the musical composition should be extended.

The proposed algorithm for the musical key-detection can be divided into the following steps:

- 1) Creation of the signature of fifths corresponding to the analyzed fragment of a music composition:
 - determination of the aggregate durations of individual pitch-classes within the considered fragment of music;
 - designation of the vector \mathbf{K} , representing the normalized aggregate durations of individual pitch-classes.
- 2) Determination of the MDASF:
 - calculation of the characteristic values corresponding to all possible directed axes;
 - selection of the directed axis with the maximum characteristic value, i.e., MDASF;
 - if MDASF cannot be determined:
 - a) increment the length of analyzed fragment;
 - b) jump to the point no. 1.
- 3) Determination of the two relative keys associated with the obtained MDASF (one of which is the key of the analyzed fragment of music). The relative keys corresponding to a given MDASF are pointed by the MDASF rotated clockwise by 30° .
- 4) Determination of the major/minor mode axis and the angle it creates with the axis OX (ϕ_1). The major/minor mode axis is perpendicular to the MDASF, and its tip points to the tone which is 90° away, clockwise, from the tone indicated by the MDASF.
- 5) Determination of the CVSF and the angle it creates with the OX -axis (ϕ_{SF}).
- 6) Calculation of the angle of the major/minor mode (ϕ_m):
 - if $\phi_m = 0$:
 - a) increment the length of the analyzed sample/fragment of music;
 - b) jump to the point no. 1;
 - if $\phi_m > 0$, the analyzed piece of music is in the major key from the pair of the two previously obtained relative keys;
 - if $\phi_m < 0$, the analyzed piece of music is in the minor key from the pair of the two previously obtained relative keys.

The essence of the developed algorithm is illustrated resorting to example 2.

Example 2 (KANIA, 2021):

Let us determine the keys of the first two preludes from the part I of J.S. Bach’s collection “The Well-Tempered Clavier” [in German: Das Wohltemperierte Klavier], the first bars of which are shown in Fig. 5.



Fig. 5. First bars of the preludes from the part I of Bach’s collection “The Well-Tempered Clavier”: a) Prelude No. 1, written in C major, BWV 846; b) Prelude No. 2, written in C minor, BWV847.

In the first step of the algorithm, after determining the aggregate durations of individual pitch-classes and calculating the vector \mathbf{K} , the signature of fifths is created (step no. 1). Knowing the lengths and directions of the vectors comprising the signature of fifths, it is possible to determine the MDASF (step no. 2). Figure 6 presents the signature of fifths and the MDASF for the analyzed excerpts of preludes.

Knowing the direction of the MDASF, we determined the pairs of the potential keys of the analyzed pieces of music as well as the directions of the major/minor mode axes (steps no. 3 and 4). The potential keys are shown in Figs. 6a–c (marked in red). They are pointed out by the MDASFs rotated by 30° clockwise (step no. 3). In each case, the major/minor mode axis is perpendicular to the MDASF and points towards the tone which is 90° away, clockwise, from the tone indicated by the MDASF (step no. 4). The directions of the major/minor mode axes and the angles they create with the OX -axis, marked as ϕ_1 , were shown in Fig. 6b. In the case of the Prelude No. 1, the angle ϕ_1 is equal to 30° , whereas for Prelude No. 2 it reaches 120° .

In the next phase of the procedure, the CVSF and the angle it creates with the OX -axis, denoted as ϕ_{SF} , are calculated (step no. 5). In the case of the Prelude No. 1, this angle is equal to 39.4° , whereas for the Prelude No. 2 it is 103.9° (Fig. 6c).

Knowing the angles ϕ_1 and ϕ_{SF} , we determined the value of the mode angle ϕ_m (step no. 6), which was then used to select one key from the previously obtained pair of relative keys. In the case of the Prelude No. 1, the value of angle ϕ_m is greater than zero ($\phi_m = 9.4^\circ$), hence the key is C major. For Prelude No. 2, ϕ_m is smaller than zero ($\phi_m = -16.1^\circ$), hence the key is C minor. In Fig. 6d, the keys detected using the proposed algorithm are marked in red. For readability, major keys are written in uppercase, whereas minor keys are written in lowercase. We apply this convention throughout figures and tables in the article.

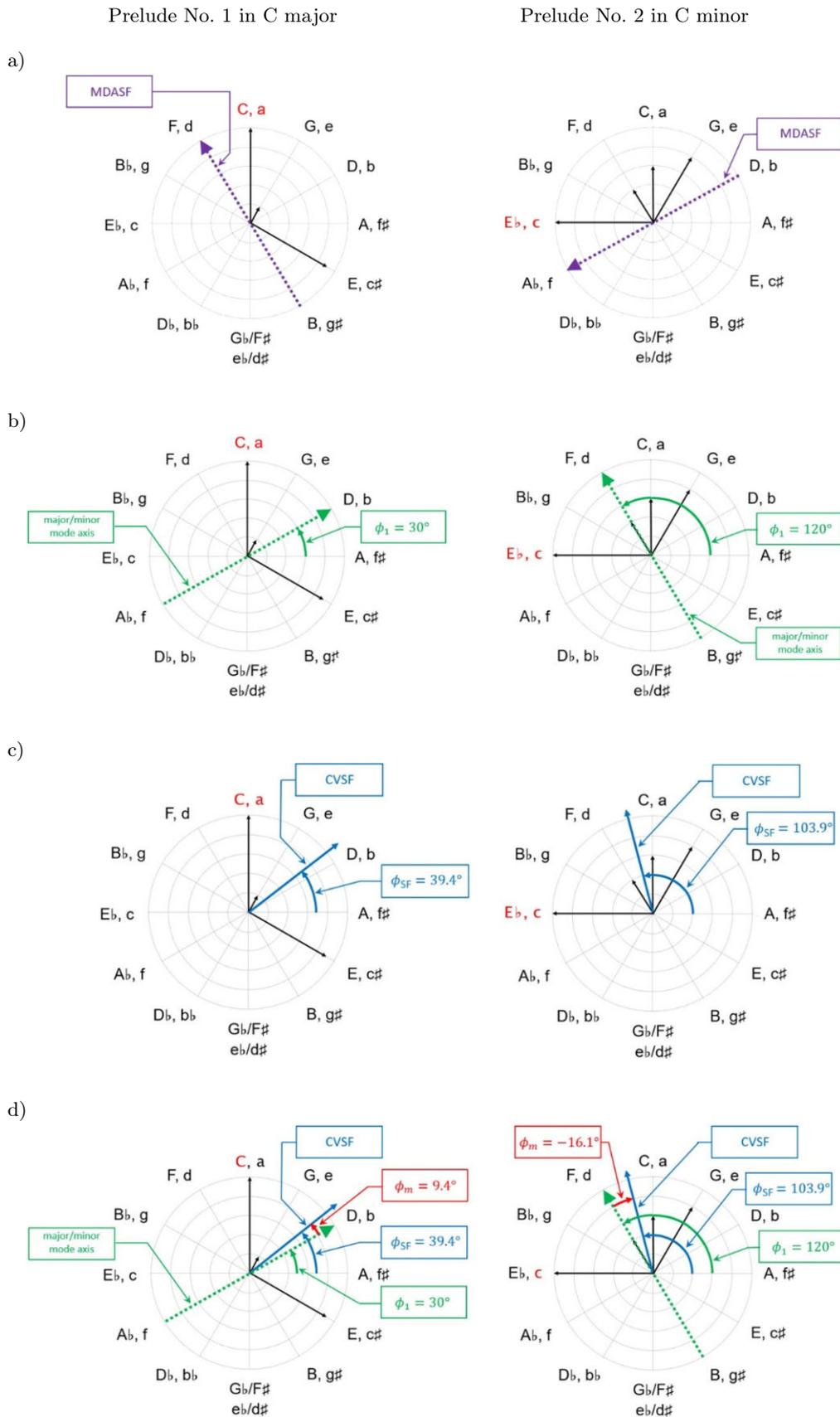


Fig. 6. Illustration of the successive steps of the proposed music key-detection algorithm based on Prelude No. 1 in C major and Prelude No. 2 in C minor, both from the part I of Bach’s collection “The Well-Tempered Clavier.”

4. Results and discussion

The aim of the conducted experiments was to compare the effectiveness of the developed algorithm with computationally simple key determination approaches based on correlation with well-known key-profiles. Four sets of preludes in 12 major and 12 minor keys were used in the experiments: the preludes by J.S. Bach from two collections of the “The Well-Tempered Clavier” – part I and part II (WTC I and WTC II), preludes by F. Chopin (Op. 28), and preludes by D. Shostakovich (Op. 87).

At first, we focused on determining the minimum number of notes needed for key-detection using the algorithm based on the signature of fifths. In each case the analysis was started with the minimum number of notes equal or greater than two. The analyzed fragment was extended by subsequent notes until it was possible to indicate the main directed axis of the signature of fifths (MDASF). All the constituent notes of any chords present were taken into account at once. For most of the analyzed works, the analysis process ended as soon as the main directed axis of the signature of fifths was determined for the first time. In a few cases, however, for which the major/minor mode axis coincided with the characteristic vector of the signature of fifths, the analyzed fragments were extended, and all the steps of the process reiterated.

The determined minimum numbers of notes needed to identify keys for different pieces of music are shown in Fig. 7. The vertical axis corresponds to the number of notes needed to detect the piece’s key, while the horizontal axis represents the numerical identifiers of the analyzed preludes. Correctly detected keys are marked with blue diamonds, whereas incorrectly detected ones are marked with orange triangles.

The analysis of the results indicates that the effectiveness of the developed algorithm, understood as the ratio of the number of correctly detected keys to the total number of analyzed works, was 85.4%. In the case of Bach’s preludes, it was 89.6%, whereas for Shostakovich’s and Chopin’s preludes the effectiveness reached 83.3% and 79.2%, respectively.

Figure 8 shows the effectiveness of the developed algorithm and the average number of notes needed to detect the key for each of the analyzed sets of works. Detection of the key was possible after analysis of 6.7 notes, on average (considering all collections). Excluding the significantly different result obtained for Shostakovich’s 14th prelude (207 notes), the average number of notes for key-detection was 4.6 notes. The fewest notes were needed for Bach’s works – 3.7 notes, on average. In the case of Chopin’s compositions, the key was found after 4.4 notes, on average, and in the case of Shostakovich’s works, 15 notes were required, on average. Again, after rejecting Shostakovich’s Prelude No. 14, the average num-

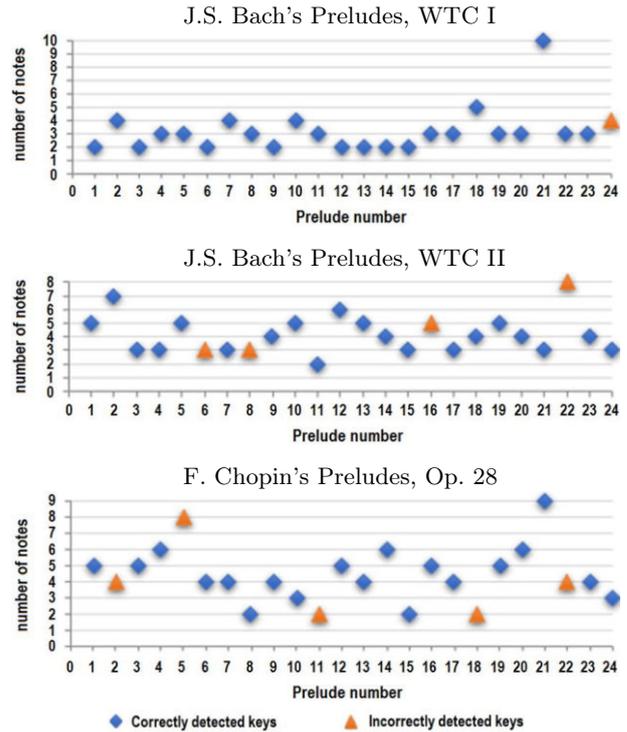


Fig. 7. Minimum number of notes required to detect the key using the proposed algorithm for individual preludes, where blue diamonds and orange triangles represent, respectively, the correctly and incorrectly detected keys (in the case of the Prelude No. 14 by Shostakovich the algorithm was able to correctly detect the key after analysis of 207 notes).

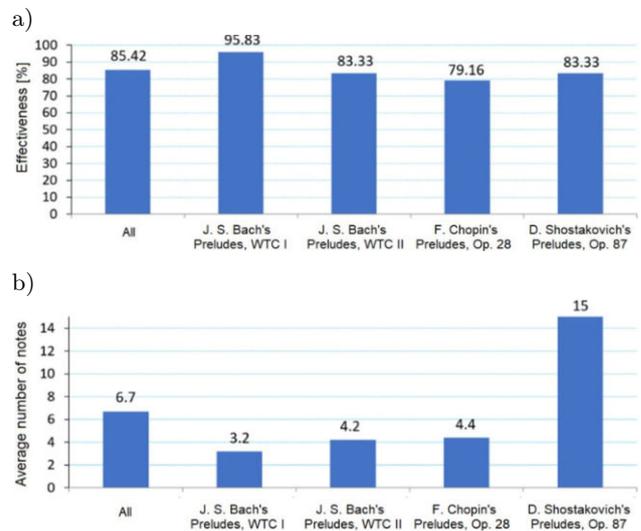


Fig. 8. Results of the music key-detection using the proposed algorithm based on the signature of fifths: a) effectiveness of the algorithm; b) average number of notes needed to detect the key.

ber of notes required to detect the key dropped to 7. It should be emphasized that the specificity of the musical notation often limits opportunities to use a small number of notes, as there are compositions which start with chords comprised of multiple notes. For exam-

ple, if a given piece of music starts with a five-note chord, it is not possible to perform the analysis for 2, 3 or 4 notes.

For further assessment, we compared the proposed key-detection algorithm with correlational approaches based on key-profiles. We chose to perform this comparison because key-profile methods in a certain way resemble the algorithm based on the signature of fifths. The way of determining the main directed axis of the signature of fifths can be associated with assigning the weight 1 to tones located on one side of the axis, and -1 to the tones located on its other side. This procedure resembles assignment of appropriate weights to particular tones, as in the considered key-profile approaches. As part of the assessment, we also accounted for the each method's computational complexity. The key-detection algorithm based on the signature of fifths is much simpler in this respect, as there is no need for calculating the correlation coefficients, hence no need for multiple (costly) multiplication operations. This is particularly important when it comes to hardware implementation of the key-detection process, e.g., in System on Chip (SoC) solutions.

In the conducted experiments, we used three sets of key-profiles: Krumhansl–Kessler (KRUMHANSL, KESSLER, 1982; KRUMHANSL, 1990), Temperley (TEMPERLEY, 2004),

and Albrecht–Shanahan (ALBRECHT, SHANAHAN, 2013). For each of the analyzed preludes the key was determined based on a short fragment taken from the beginning of a given composition. For the considered key-profile approaches, the key was detected after analysis of very short fragments of music, even ones comprised of just two notes. However, in many cases, extending the analyzed fragment of the piece resulted in the change of the previously detected key. For the algorithm based on the signature of fifths, the key was detected later, as typically more notes were needed to determine the MDASF. However, usually the determined key was stable and did not change with extension of the musical fragment.

Let us analyze in detail the results of the key-detection process for all the considered scenarios. We compared the minimum number of notes for which all key-detection approaches were able to indicate the key. The number of notes needed to determine the key was different for individual preludes but was always equal to the number of notes required by the algorithm based on the signature of fifths (due to the specificity of the performed comparison, as explained earlier in this article). Results for individual preludes are shown in Table 1. Correctly and incorrectly detected keys are illustrated for all considered key-detection approaches.

Table 1. Results obtained for different key-detection approaches, given the minimum number of notes for which all the considered methods were able to indicate the key.

Piece No.	Bach's Preludes, WTC I				Bach's Preludes, WTC II				Chopin's Preludes, Op. 28				Shostakovich's Preludes, Op. 87			
	SF	KK	T	AS	SF	KK	T	AS	SF	KK	T	AS	SF	KK	T	AS
1	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✓	×	✓	✓	×	×	×	×	✓	✓	✓	✓
3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓
4	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓	✓	✓	✓	✓
5	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	✓	✓	✓	✓
6	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	×	×	×	✓
7	✓	✓	✓	✓	✓	×	×	✓	✓	×	✓	✓	✓	✓	×	✓
8	✓	✓	✓	✓	×	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓
9	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓	✓	✓	✓	✓
11	✓	✓	✓	✓	✓	×	×	✓	×	×	×	×	✓	✓	✓	×
12	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓
13	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
14	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓
15	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	×	×	✓	✓	×
16	✓	✓	✓	✓	×	✓	✓	×	✓	×	✓	✓	✓	✓	✓	✓
17	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
18	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	×	✓	✓	✓	✓
19	✓	✓	✓	✓	✓	✓	×	✓	✓	×	×	✓	✓	✓	✓	✓
20	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	×	✓	✓	✓
21	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓	✓	✓	✓	✓
22	✓	✓	✓	✓	×	✓	✓	✓	×	×	×	×	✓	✓	✓	✓
23	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
24	×	×	✓	✓	✓	✓	✓	✓	✓	×	×	✓	✓	✓	✓	✓

SF – the method using the signature of fifths, KK – the correlation approach based on the Krumhansl–Kessler key-profiles, T – the correlation approach based on the Temperley key-profiles, AS – the correlation approach based on the Albrecht–Shanahan key-profiles, ✓ – correctly detected key, × – incorrectly detected key.

Table 2. Effectiveness of determining the key in particular groups of preludes and using different key-detection approaches.

Collection	SF [%]	KK [%]	T [%]	AS [%]
Bach's Preludes, WTC I	95.83	95.83	100	100
Bach's Preludes, WTC II	87.50	87.50	91.67	87.50
Chopin's Preludes, Op. 28	83.33	45.83	62.50	83.33
Shostakovich's Preludes, Op. 87	83.33	87.50	87.50	95.83
All Preludes	85.42	79.17	85.42	91.67

SF – the method using the signature of fifths, KK – the correlation approach based on the Krumhansl–Kessler key-profiles, T – the correlation approach based on the Temperley key-profiles, AS – the correlation approach based on the Albrecht–Shanahan key-profiles.

Table 2 presents a synthetic summary of the results. It lists the effectiveness of each of the considered key-detection approaches in different sets of preludes. The results of the effectiveness obtained for the set of all preludes are also given (last row).

Analyzing the results illustrated in Tables 1 and 2, one can get the impression that, in terms of the effectiveness, the algorithm using the signature of fifths does not differ significantly from the correlational approaches implementing key-profiles (the same or greater effectiveness was achieved with Albrecht–Shanahan key-profiles). Moreover, the algorithm utilizing the signature of fifths in the majority of cases required a greater number of notes to indicate a piece's key. In some cases, it was the only method that indicated the wrong key, e.g., for Prelude No. 22 from Bach's "The Well-Tempered Clavier" – part II – or Prelude No. 20 by Shostakovich. However, this algorithm does exhibit some unique and advantageous properties.

In order to show the distinctive features of the algorithm based on the signature of fifths, let us first inspect the Prelude No. 14, Op. 87, by Shostakovich, whose fragment is presented in Fig. 9.

At the beginning of this Prelude, in its left-hand part, one can notice many repeating tones of $B\flat$. It is also worth mentioning that up to the point marked with the index 5, where the note $G\flat$ appears in the right-hand part, there are only three tones present in the composition ($E\flat$, $D\flat$, $B\flat$). Determination of the MDASF (Fig. 10e) is not possible until reaching that point on the staff. Knowing the MDASF, one can determine the direction of the major/minor mode axis as well as the angle $\phi_m = -36.6^\circ$ (Fig. 10f). The negative value of the angle ϕ_m indicates the minor key mode – in the considered case it is $e\flat$ minor. The signatures of fifths corresponding to the increasingly longer fragments of the prelude (starting from its beginning to a given index) are presented in Fig. 10.

The proposed algorithm correctly identified the key after 207 notes. Table 3 presents the results of key-detection obtained for all the approaches considered, based on the analysis of fragments 0–1, 0–2, 0–3, 0–4, and 0–5.

Analyzing the results presented in Table 3, we can notice that the key-profile approaches need fewer notes to detect the key. Unfortunately, the initial indications are often incorrect and tend to vary as the

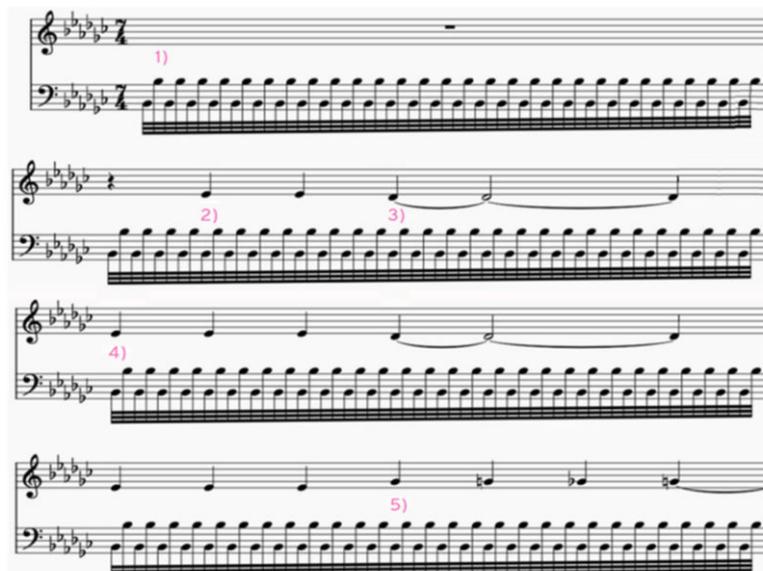


Fig. 9. Initial fragment of the Prelude No. 14, Op. 87, by Shostakovich.

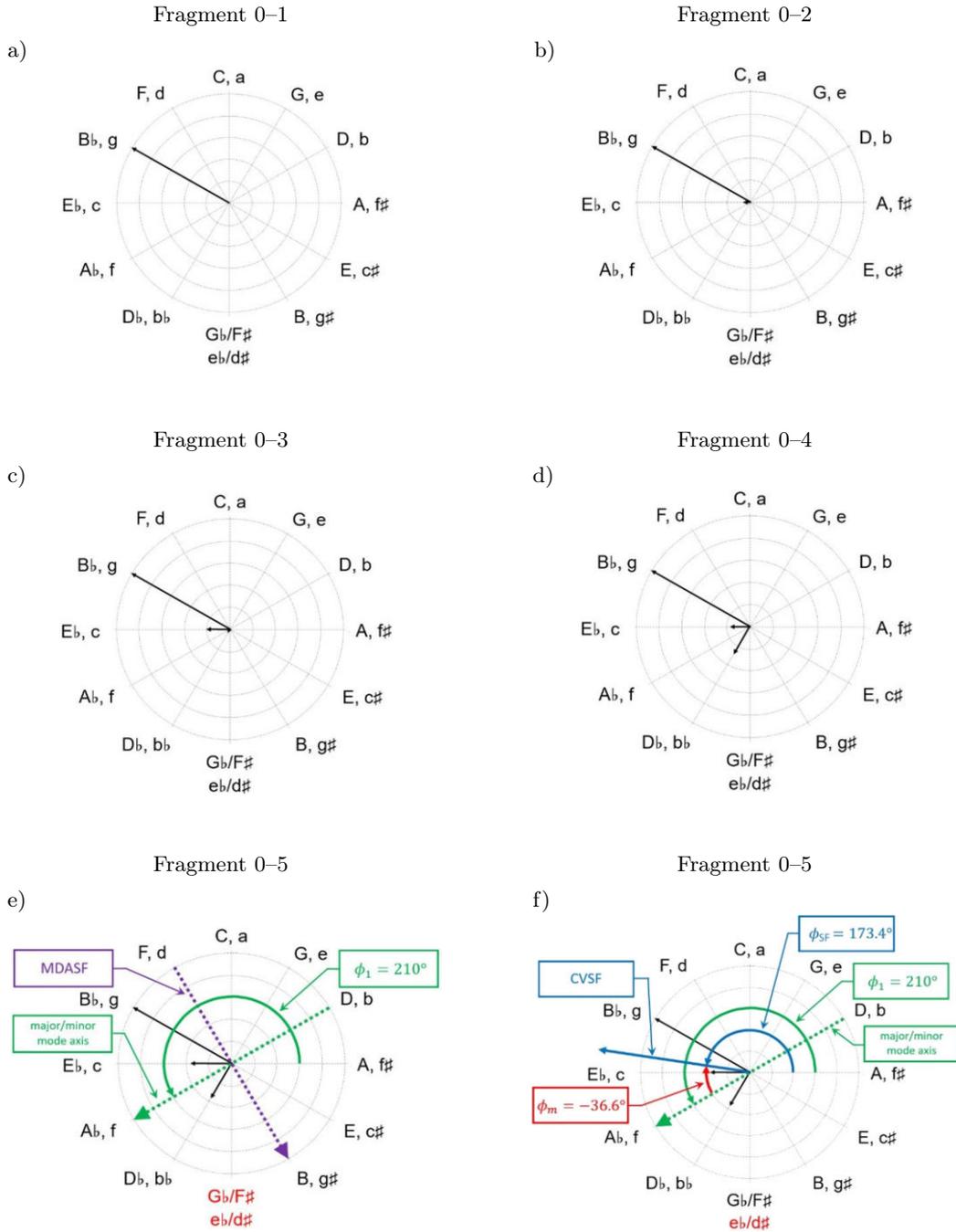


Fig. 10. Signatures of fifths corresponding to increasingly longer fragments of Prelude No. 14, Op. 87, by Shostakovich, shown in Fig. 9 (up to the point marked with index 5).

Table 3. Summary of the key-detection results obtained for the fragment of the prelude shown in Fig. 10, using the algorithm based on the signature of fifths (SF) as well as the considered approaches implementing the key-profiles of Krumhansl–Kessler (KK), Temperley (T), and Albrecht–Shanahan (AS).

Method	Analyzed fragment				
	0-1	0-2	0-3	0-4	0-5
SF	?	?	?	?	e \flat
KK	B \flat	B \flat	b \flat	b \flat	B \flat
T	b \flat	b \flat	b \flat	b \flat	b \flat
AS	B \flat	e \flat	b \flat	e \flat	e \flat

length of the analyzed music fragment increases. For the key-profiles of Krumhansl–Kessler and Temperley, the obtained keys were incorrect for all considered fragments – B♭ major (B♭) or B♭ minor (b♭). This result can be explained by the dominance of the sound B♭, which is the tonic of the indicated keys. The Albrecht–Shanahan key-profile approach detected various keys, among which was the correct one, i.e., E♭ minor (e♭). The algorithm using the signature of fifths needed more notes than the key-profile approaches. However, it indicated the key only when the MDASF was determined, and hence the detected key was usually correct.

Another distinctive feature of the key-detection algorithm proposed in this paper is the stability of the decision-making process, understood as low susceptibility to changes in the detected key as the length of the analyzed music fragment increases. This feature can easily be illustrated by the Prelude No. 21, Op. 28, by Chopin, the initial fragment of which is shown in Fig. 11.



Fig. 11. Initial fragment of the Prelude No. 21, Op. 28, by Chopin.

Table 4 presents the values of r_i , which represent the lengths of vectors making up the signatures of fifths calculated based on the aggregate durations of individual pitch-classes for a given number of notes (starting from the beginning of the prelude). The values from Table 4 were used to create the signatures of fifths shown in Fig. 12.

Table 4. Lengths of vectors representing the signatures of fifths corresponding to the Prelude No. 21, Op. 28, by Chopin, obtained for different numbers of initial notes.

Pitch-class	Number of notes							
	2	3	5	7	9	11	13	15
C						0.29	0.29	0.25
C♯/D♭								
D					0.17	0.14	0.29	0.38
D♯/E♭				0.2	0.17	0.14	0.14	0.13
E			0.25	0.2	0.17	0.14	0.14	0.13
F	1	1	1	1	1	1	1	1
F♯/G♭								
G			0.25	0.2	0.17	0.14	0.14	0.13
G♯/A♭								
A				0.2	0.17	0.14	0.14	0.13
A♯/B♭	1	0.33	0.25	0.2	0.33	0.29	0.43	0.38
B								

In Table 5, the results of the key-detection based on the signature of fifths were juxtaposed with the results

Table 5. Key-detection results obtained for the Prelude No. 21 by Chopin, Op. 28, using the algorithm based on the signature of fifths (SF) as well as the considered approaches implementing the key-profiles of Krumhansl–Kessler (KK), Temperley (T), and Albrecht–Shanahan (AS).

Method	Number of notes							
	2	3	5	7	9	11	13	15
SF	?	?	?	?	B♭	B♭	B♭	B♭
KK	F	F	F	F	F	F	d	d
T	f	f	f	F	F	F	B♭	B♭
AS	b♭	b♭	F	F	B♭	F	B♭	B♭

obtained for the correlational approaches utilizing key-profiles.

The results in Table 5 indicate that key-detection algorithm based on the signature of fifths offers greater stability. The key indicated after analysing the 9th note does not change as the length of the fragment increases further because the direction of MDASF does not change, whereas the sign of the angle ϕ_m experiences only insignificant variations (Figs. 12e–h). The keys found using the key-profile approaches changed at least once. This example clearly illustrates that the algorithm based on the signature of fifths requires more notes to determine the key, but once the key is detected the decision is usually correct and less prone to changes.

In summary, we can state the following:

- the proposed algorithm is computationally simple and easy to implement, as it does not require complex calculations;
- the key-detection algorithm based on the signature of fifths is competitive with the correlational approaches using key-profiles, especially if one wants to determine the key from a very short fragment of music;
- the algorithm using the signature of fifths usually needs a larger number of notes to determine the key than its key-profile alternatives, but once the key is detected the decision does not tend to change as the length of the analysed music fragment increases.

5. Conclusion

Development of multimedia systems is inextricably linked with methods enabling acquisition of musical knowledge. Currently, when almost all songs are only a few mouse clicks away, the problem for listeners is selection of music. Nowadays, many listeners use software applications which are able to suggest songs suitable for a given person. Such applications have become an integral part of the music industry.

Classification of music can be facilitated by various types of signal quantification and feature extraction techniques. Criteria of selection may include the

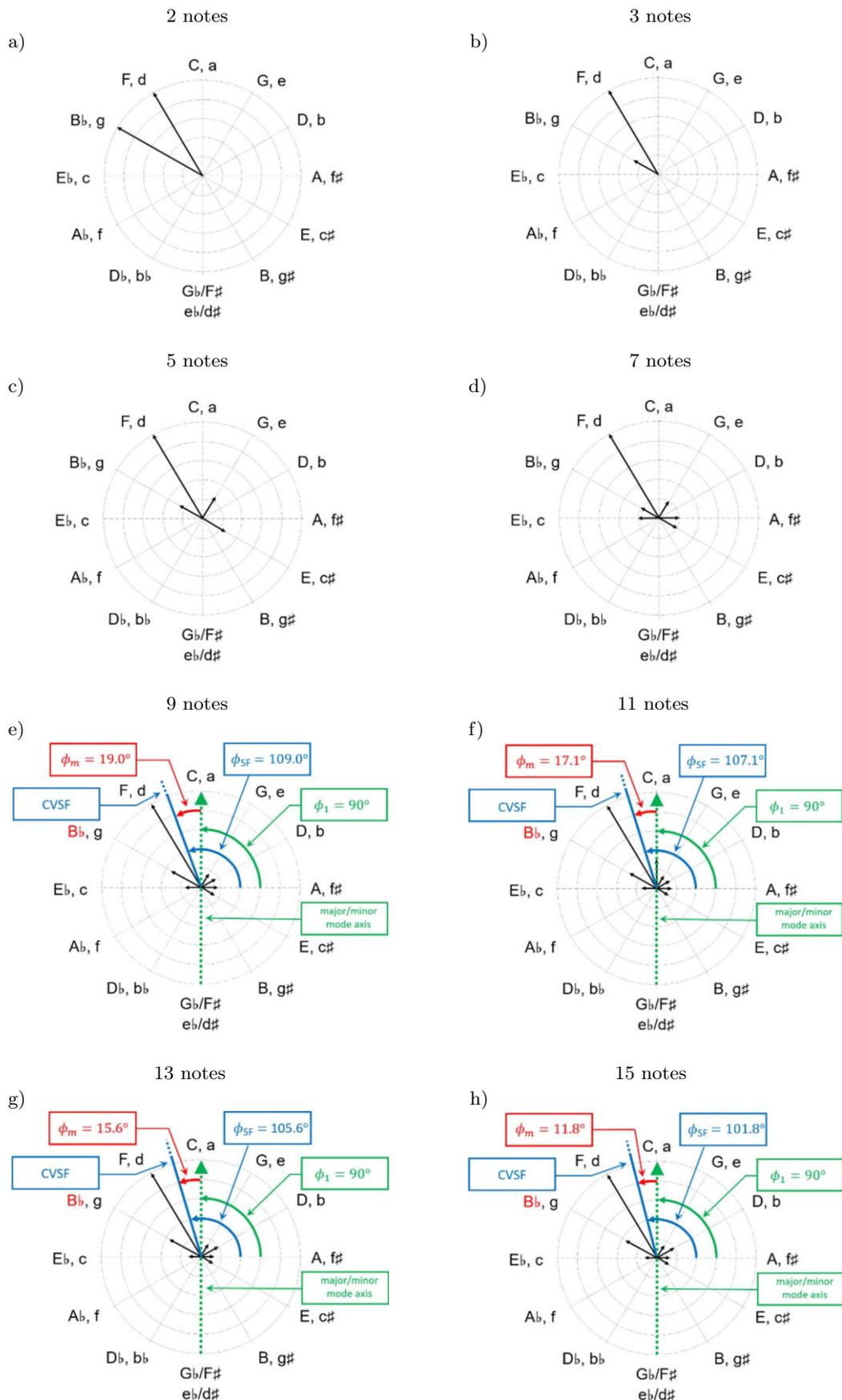


Fig. 12. Signatures of fifths obtained for different lengths of the analyzed fragments of the Prelude No. 21, Op. 28, by Chopin.

style, genre, character or key of a piece. The algorithm presented in this paper could be applied in such classification applications.

In this paper, we presented a novel algorithm enabling determination of the key of a musical piece based on the analysis of its signature of fifths. The algorithm is a simplified version of the method discussed by KANIA and KANIA (2019). The simplification boils down to the determination of the major/minor mode of the analyzed piece via application of new descriptors of the signature of fifths, i.e., the major/minor mode axis, the characteristic angle of the signature of fifths, and the major/minor mode angle. The effectiveness of the algorithm was tested on a collection of 96 preludes comprised of two sets of 24 preludes “The Well-Tempered Clavier”, BWV 846-869, by Bach, 24 preludes Op. 28 by Chopin, and 24 preludes Op. 87 by Shostakovich. Each set of preludes, individually, covered all possible keys.

The main advantage of the proposed key-detection algorithm is the stability of its decision-making process, i.e., low sensitivity to changes of the indicated key as the length of the analyzed fragment of music is increased. This feature clearly distinguishes the method from the tested correlational key-detection approaches based on key-profiles. Another advantage is its conceptual as well as computational simplicity. The latter advantage facilitates the method’s implementation in hardware, e.g., real-time presentation of musical notation on electronic displays. Calculations required to obtain the key signature with this method can be limited to execution of addition and comparison operations, which are convenient in terms of hardware implementation (only these two operations are needed to determine MDASF). Implementation of the proposed algorithm in a microprocessor system or SoC uses minimal resources, smaller than those required by the considered correlation-based approaches (using tonal profiles), in case of which many multiplication and division operations need to be performed.

The signature of fifths provides means for effective realization of the key-detection process. The effectiveness of the proposed algorithm, tested on the whole set of 96 preludes, was over 85 %. The correct key was detected after the analysis of 6.7 notes, on average.

The concept of the signature of fifths creates new opportunities in the area of music information retrieval. In addition to key determination, it has already been shown that the coefficients quantifying the variability of the signatures of fifths in time can be useful as feature coefficients in music classification processes (KANIA *et al.*, 2021b; ŁUKASZEWICZ, KANIA, 2022).

References

1. AARDEN B. (2003), *Dynamic melodic expectancy*, Ph.D. Thesis, Ohio State University.
2. ALBRECHT J., SHANAHAN D. (2013), The use of large corpora to train a new type of key-finding algorithm: An improved treatment of the minor mode, *Music Perception: An Interdisciplinary Journal*, **31**(1): 59–67, doi: [10.1525/mp.2013.31.1.59](https://doi.org/10.1525/mp.2013.31.1.59).
3. ANGLADE A., BENETOS E., MAUCH M., DIXON S. (2010), Improving music genre classification using automatically induced harmony rules, *Journal of New Music Research*, **39**(4): 349–361, doi: [10.1080/09298215.2010.525654](https://doi.org/10.1080/09298215.2010.525654).
4. BAUMANN S.A. (2021), Deeper convolutional neural networks and broad augmentation policies improve performance in musical key estimation, [in:] *Proceedings of the 22nd International Society for Music Information Retrieval Conference*, pp. 42–49.
5. BELLMANN H. (2005), About the determination of key of a musical excerpt, [in:] Kronland-Martinet R., Voinier T., Ystad S. [Eds.], *Computer Music Modeling and Retrieval, CMMR 2005, Lecture Notes in Computer Science*, **3902**: 76–91, doi: [10.1007/11751069_7](https://doi.org/10.1007/11751069_7).
6. BERNARDES G., COCHARRO D., CAETANO M., GUEDES C., DAVIES M.E.P. (2016), A multi-level tonal interval space for modelling pitch relatedness and musical dissonance, *Journal of New Music Research*, **45**(4): 281–294, doi: [10.1080/09298215.2016.1182192](https://doi.org/10.1080/09298215.2016.1182192).
7. BERNARDES G., DAVIES M., GUEDES C. (2017), Automatic musical key estimation with mode bias, [in:] *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 316–320, doi: [10.1109/ICASSP.2017.7952169](https://doi.org/10.1109/ICASSP.2017.7952169).
8. BOULANGER-LEWANDOWSKI N., BENGIO Y., VINCENT P. (2013), Audio chord recognition with recurrent neural networks, [in:] *Proceedings of the 14th International Conference on Music Information Retrieval (ISMIR)*, pp. 335–340.
9. CANCINO CHACÓN C.E., LATNER S., GRACHTEN M. (2014), Developing tonal perception through unsupervised learning, [in:] *The 15th International Society for Music Information Retrieval Conference*, pp. 195–200.
10. CHAPIN H., JANTZEN K., KELSO J.S., STEINBERG F., LARGE E. (2010), Dynamic emotional and neural responses to music depend on performance expression and listener experience, *PLOS ONE*, **5**(12): e13812, doi: [10.1371/journal.pone.0013812](https://doi.org/10.1371/journal.pone.0013812).
11. CHEN T-P., SU L. (2018), Functional harmony recognition of symbolic music data with multi-task recurrent neural networks, [in:] *Proceedings of the 19th ISMIR Conference*, pp. 90–97.
12. CHEW E. (2000), *Towards a mathematical model of tonality*, Ph.D. Thesis, Massachusetts Institute of Technology.
13. CHEW E. (2007), Out of the grid and into the spiral: Geometric interpretations of and comparisons with the spiral-array model, *Computing in Musicology*, **15**: 51–72.
14. CHUAN C.-H., CHEW E. (2005), Polyphonic audio key finding using the spiral array CEG algorithm, [in:] *2005 IEEE International Conference on Multimedia and Expo*, pp. 21–24, doi: [10.1109/ICME.2005.1521350](https://doi.org/10.1109/ICME.2005.1521350).

15. CHUAN C.-H., CHEW E. (2007), Audio key finding: Considerations in system design and case studies on Chopin's 24 Preludes, *EURASIP Journal on Advances in Signal Processing*, **2007**(1): 056561, doi: [10.1155/2007/56561](https://doi.org/10.1155/2007/56561).
16. CHUAN C.-H., CHEW E. (2014), The KUSC classical music dataset for audio key finding, *The International Journal of Multimedia & Its Applications*, **6**(4): 1–18, doi: [10.5121/ijma.2014.6401](https://doi.org/10.5121/ijma.2014.6401).
17. DAWSON M.R.W. (2018), *Connectionist Representations of Tonal Music. Discovering Musical Patterns by Interpreting Artificial Neural Networks*, AU Press, Athabasca, doi: [10.15215/aupress/9781771992206.01](https://doi.org/10.15215/aupress/9781771992206.01).
18. DENG J., KWOK Y.-K. (2017), Large vocabulary automatic chord estimation using deep neural nets: Design framework, system variations and limitations, doi: [10.48550/arXiv.1709.07153](https://doi.org/10.48550/arXiv.1709.07153).
19. FOSCARIN F., AUDEBERT N., FOURNIER-S'NIEHOTTA R. (2021), PKSpell: Data-driven pitch spelling and key signature estimation, [in:] *Proceedings of the 22nd International Society for Music Information Retrieval Conference*, pp. 197–204.
20. GEBHARDT R., LYKARTSIS A., STEIN M. (2018), A confidence measure for key labelling, *Proceedings of the 19th International Symposium on Music Information Retrieval (ISMIR)*, pp. 3–9.
21. GOMEZ E., HERRERA P. (2004), Estimating the tonality of polyphonic audio files: Cognitive versus machine learning modeling strategies, [in:] *Proceedings of the 5th International Conference on Music Information Retrieval*, pp. 92–95.
22. HARTE C., SANDLER M., GASSER M. (2006), Detecting harmonic change in musical audio, [in:] *Proceedings of Special 1st ACM Workshop on Audio and Music Computing Multimedia*, pp. 21–26, doi: [10.1145/1178723.1178727](https://doi.org/10.1145/1178723.1178727).
23. HERREMANS D., CHEW E. (2019), MorpheuS: Generating structured music with constrained patterns and tension, *IEEE Transactions on Affective Computing*, **10**(4): 520–523, doi: [10.1109/TAFFC.2017.2737984](https://doi.org/10.1109/TAFFC.2017.2737984).
24. HORI T., NAKAMURA K., SAGAYAMA S. (2017), Music chord recognition from audio data using bidirectional encoder-decoder LSTMs, [in:] *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1312–1315, doi: [10.1109/APSIPA.2017.8282235](https://doi.org/10.1109/APSIPA.2017.8282235).
25. HUANG C.-Z.A., DUVENAUD D., GAJOS K.Z. (2016), Chordripple: Recommending chords to help novice composers go beyond the ordinary, [in:] *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pp. 241–250, doi: [10.1145/2856767.2856792](https://doi.org/10.1145/2856767.2856792).
26. JACOBY N., TISHBY N., TYMOCZKO D. (2015), An information theoretic approach to chord categorization and functional harmony, *Journal of New Music Research*, **44**(3): 219–244, doi: [10.1080/09298215.2015.1036888](https://doi.org/10.1080/09298215.2015.1036888).
27. KANIA D., KANIA P. (2019), A key-finding algorithm based on music signature, *Archives of Acoustics*, **44**(3): 447–457, doi: [10.24425/aoa.2019.129260](https://doi.org/10.24425/aoa.2019.129260).
28. KANIA P. (2021), *The parametrization of signature of fifths aimed at music data mining* [in Polish: *Parametryzacja sygnatury kwintowej ukierunkowana na pozyskiwanie wiedzy muzycznej*], Msc. Thesis, Faculty of Physics, Adam Mickiewicz University in Poznań, Poland.
29. KANIA P., KANIA D., ŁUKASZEWICZ T. (2021a), A hardware-oriented algorithm for real-time music key signature recognition, *Applied Sciences*, **11**(18): 8753, doi: [10.3390/app11188753](https://doi.org/10.3390/app11188753).
30. KANIA D., KANIA P., ŁUKASZEWICZ T. (2021b), Trajectory of fifths in music data mining, *IEEE Access*, **9**: 8751–8761, doi: [10.1109/ACCESS.2021.3049266](https://doi.org/10.1109/ACCESS.2021.3049266).
31. KANIA M., ŁUKASZEWICZ T., KANIA D., MOŚCIŃSKA K., KULISZ J. (2022), A comparison of the music key detection approaches utilizing key-profiles with a new method based on the signature of fifths, *Applied Sciences*, **12**(21): 11261, doi: [10.3390/app122111261](https://doi.org/10.3390/app122111261).
32. KORZENIOWSKI F., WIDMER G. (2017), End-to-end musical key estimation using a convolutional neural network, [in:] *Proceedings of the 25th European Signal Processing Conference (EUSIPCO)*, pp. 966–970.
33. KORZENIOWSKI F., WIDMER G. (2018), Genre-agnostic key classification with convolutional neural networks, [in:] *19th International Society for Music Information Retrieval Conference*, pp. 264–270.
34. KRUMHANSL C.L. (1990), *Cognitive Foundations of Musical Pitch*, Oxford University Press, New York, doi: [10.1093/acprof:oso/9780195148367.001.0001](https://doi.org/10.1093/acprof:oso/9780195148367.001.0001).
35. KRUMHANSL C.L., KESSLER E.J. (1982), Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys, *Psychological Review*, **89**(4): 334–368, doi: [10.1037/0033-295X.89.4.334](https://doi.org/10.1037/0033-295X.89.4.334).
36. LONGUET-HIGGINS H.C. (1962a), Letter to a musical friend, *The Music Review*, **23**: 244–248.
37. LONGUET-HIGGINS H.C. (1962b), Second letter to a musical friend, *The Music Review*, **23**: 271–280.
38. ŁUKASZEWICZ T., KANIA D. (2022), A music classification approach based on the trajectory of fifths, *IEEE Access*, **10**: 73494–73502, doi: [10.1109/ACCESS.2022.3190016](https://doi.org/10.1109/ACCESS.2022.3190016).
39. MASADA K., BUNESCU R. (2017), Chord recognition in symbolic music using semi-Markov conditional random fields, [in:] *Proceedings of the 18th International Society for Music Information Retrieval Conference*, pp. 272–278.
40. MAUCH M., DIXON S. (2010), Approximate note transcription for the improved identification of difficult chords, [in:] *Proceedings of the 11th International Society for Music Information Retrieval Conference*, pp. 135–140.
41. MCFEE B., BELLO J.P. (2017), Structured training for large-vocabulary chord recognition, [in:] *Proceedings of the 18th International Conference on Music Information Retrieval (ISMIR)*, pp. 188–194.
42. NÁPOLES LÓPEZ N., ARTHUR C., FUJINAGA I. (2019), Key-finding based on a hidden markov model and key

- profiles, [in:] *Proceedings of the 6th International Conference on Digital Libraries for Musicology*, pp. 33–37, doi: [10.1145/3358664.3358675](https://doi.org/10.1145/3358664.3358675).
43. NÁPOLES LÓPEZ N., FEISTHAUER L., LEVE F., FUJINAGA I. (2020), On local keys, modulations, and tonicizations: A dataset and methodology for evaluating changes of key, [in:] *Proceedings of the 7th International Conference on Digital Libraries for Musicology*, pp. 18–26, doi: [10.1145/3424911.3425515](https://doi.org/10.1145/3424911.3425515).
44. NI Y., MCVICAR M., SANTOS-RODRÍGUEZ R., DE BIE T. (2013), Understanding effects of subjectivity in measuring chord estimation accuracy, *IEEE Transactions on Audio, Speech, and Language Processing*, **21**(12): 2607–2615, doi: [10.1109/TASL.2013.2280218](https://doi.org/10.1109/TASL.2013.2280218).
45. OSMALSKYJ J., EMBRECHTS J.-J., PIÉRARD S., VAN DROOGENBROECK M. (2012), Neural networks for musical chords recognition, *Journées d’Informatique Musicale*, pp. 39–46.
46. PAPAPOPOULOS H., PEETERS G. (2012), Local key estimation from an audio signal relying on harmonic and metrical structures, *IEEE Transactions on Audio, Speech, and Language Processing*, **20**(4): 1297–1312, doi: [10.1109/TASL.2011.2175385](https://doi.org/10.1109/TASL.2011.2175385).
47. PEETERS G. (2006), Musical key estimation of audio signal based on hmm modeling of chroma vectors, [in:] *Proceedings of the 9th International Conference on Digital Audio Effects*, pp. 127–131.
48. PEISZER E., LIDY T., RAUBER A. (2008), Automatic audio segmentation: Segment boundary and structure detection in popular music, [in:] *Proceedings of the International Workshop on Learning the Semantics of Audio Signals*, pp. 45–59.
49. PÉREZ-SANCHIO C., RIZO D., IÑESTA J.M., PONCE DE LEÓN P.J., KERSTEN S. (2010), Genre classification of music by tonal harmony, *Intelligent Data Analysis*, **14**(5): 533–545.
50. QUINN I., WHITE C.W. (2017), Corpus-derived key profiles are not transpositionally equivalent, *Music Perception*, **34**(5): 531–540, doi: [10.1525/mp.2017.34.5.531](https://doi.org/10.1525/mp.2017.34.5.531).
51. RAPHAEL Ch., STODDARD J. (2004), Functional harmonic analysis using probabilistic models, *Computer Music Journal*, **28**(3): 45–52.
52. ROIG C., TARDÓN L.J., BARBANCHO I., BARBANCHO A.M. (2014), Automatic melody composition based on a probabilistic model of music style and harmonic rules, *Knowledge-Based Systems*, **71**: 419–434, doi: [10.1016/j.knosys.2014.08.018](https://doi.org/10.1016/j.knosys.2014.08.018).
53. SABATHÉ R., COUTINHO E., SCHULLER B. (2017), Deep recurrent music writer: Memory-enhanced variational autoencoder-based musical score composition and an objective measure, [in:] *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 3467–3474, doi: [10.1109/IJCNN.2017.7966292](https://doi.org/10.1109/IJCNN.2017.7966292).
54. SAPP C.S. (2001), Harmonic visualizations of tonal music, *International Computer Music Conference (ICMC) 2001*, pp. 423–430.
55. SHEPARD R. (1982), Geometrical approximations to the structure of musical pitch, *Psychological Review*, **89**: 305–333, doi: [10.1037/0033-295X.89.4.305](https://doi.org/10.1037/0033-295X.89.4.305).
56. SIGTIA S., BOULANGER-LEWANDOWSKI N., DIXON S. (2015), Audio chord recognition with a hybrid recurrent neural network, [in:] *16th International Society for Music Information Retrieval Conference*, pp. 127–133.
57. TEMPERLEY D. (2004), Bayesian models of musical structure and cognition, *Musicae Scientiae*, **8**(2): 175–205, doi: [10.1177/102986490400800204](https://doi.org/10.1177/102986490400800204).
58. TEMPERLEY D., MARVIN E.W. (2008), Pitch-class distribution and key identification, *Music Perception*, **25**(3): 193–212, doi: [10.1525/mp.2008.25.3.193](https://doi.org/10.1525/mp.2008.25.3.193).
59. TOIVIAINEN P., KRUMHANSL C.L. (2003), Measuring and modeling real-time responses to music: the dynamics of tonality induction, *Perception*, **32**(6): 741–766, doi: [10.1068/p3312](https://doi.org/10.1068/p3312).
60. TYMOCZKO D. (2006), The geometry of musical chords, *Science*, **313**(5783): 72–74, doi: [10.1126/science.1126287](https://doi.org/10.1126/science.1126287).
61. TYMOCZKO D. (2011), *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*, Oxford University Press, New York.
62. WEISS C. (2013), Global key extraction from classical music audio recordings based on the final chord, [in:] *Proceedings of the Sound and Music Computing Conference*, pp. 742–747.
63. WU Y., LI W. (2018), Music chord recognition based on midi-trained deep feature and BLSTM-CRF hybrid decoding, [in:] *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 376–380, doi: [10.1109/ICASSP.2018.8461439](https://doi.org/10.1109/ICASSP.2018.8461439).
64. YANG S., REED C., CHEW E., BARTHET M. (2021), Examining emotion perception agreement in live music performance, *IEEE Transactions on Affective Computing*, **14**: 1442–1460, doi: [10.1109/TAFFC.2021.3093787](https://doi.org/10.1109/TAFFC.2021.3093787).
65. YUST J. (2019), Stylistic information in pitch-class distributions, *Journal of New Music Research*, **48**(3): 217–231, doi: [10.1080/09298215.2019.1606833](https://doi.org/10.1080/09298215.2019.1606833).
66. ZHOU X.H., LERCH A. (2015), Chord detection using deep learning, [in:] *Proceedings of SIMIR 2015*, pp. 52–58.

Research Paper

Analysis of the Vehicle Engine Misfires using Frequency-Domain Approaches at Various RPMs with ADXL1002 Accelerometer

Muhammad AHSAN^{(1)*}, Dariusz BISMOR⁽¹⁾, Paweł FABIŚ⁽²⁾⁽¹⁾ Department of Measurements and Control Systems, Silesian University of Technology
Gliwice, Poland⁽²⁾ Faculty of Transport and Aeronautical Engineering, Department of Road Transport
Silesian University of Technology
Katowice, Poland*Corresponding Author e-mail: ahsanmuhammad@aol.com/Muhammad.Ahsan@polsl.pl

(received December 31, 2023; accepted May 13, 2024; published online October 3, 2024)

Vehicle engine vibration signals acquired using MEMS sensors are crucial in the diagnosis of engine malfunctions, notably misfires due to unwanted signals and external noises in the recorded vibration dataset. In this study, the ADXL1002 accelerometer interfaced with the Beaglebone Black microcontroller is employed to capture vibration signals emitted by the vehicle engine across various operational states, including unloaded, loaded, and misfire conditions at 1500 RPMs, 2500 RPMs, and 3000 RPMs. In conjunction with the acquisition of this raw vibration data, frequency-domain signal processing techniques are employed to meticulously analyze and diagnose the distinct signatures of misfire occurrences across various engine speeds and loads. These techniques encompass the fast Fourier transform (FFT), envelope spectrum (ES), and empirical mode decomposition (EMD), each tailored to discern and characterize the nuanced vibration patterns associated with misfire events at different operational conditions.

Keywords: ADXL1002 accelerometer; MEMS sensors; vehicle engine vibration; misfire condition; fast Fourier transform; envelope spectrum; empirical mode decomposition.



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

Misfire, a persistent issue in vehicle engines, disrupts the harmonious sequence of combustion. Its prevalence is striking, causing substantial hurdles in engine performance and reliability, as highlighted in the research conducted by *HMIDA et al. (2021)*. This fault, defined by the inability of a cylinder to ignite its air-fuel mixture correctly, extends its impact across multiple domains. Recognizing and remedying misfires is of paramount importance in automotive engineering. *NAVEEN VENKATESH et al. (2022)* underscore the criticality of accurate detection and timely resolution of this issue. Essentially, mitigating misfires is not just about fixing a glitch; it is about preserving the efficiency and ecological integrity of vehicle engines.

Various contemporary techniques have emerged as effective means to pinpoint engine misfire faults, show-

ing the innovative strides in this field. For instance, one approach capitalizes on the ionic current (I.C.) presented at spark plug electrodes, leveraging them as sensors to detect misfiring cylinders. Studies referenced by *WANG et al. (2022)* and *KUMANO et al. (2020)* delve into this method, illustrating how monitoring changes in ionic current during engine operation offers insights into discerning misfiring events.

Another notable method involves monitoring the exhaust gas temperature at a reduced sampling rate. *TAMURA et al. (2011)* shed light on this technique, emphasizing its utility in detecting misfires within internal combustion engines. By scrutinizing irregular temperature fluctuations in the exhaust gas, this approach identifies deviations that serve as reliable indicators of misfiring occurrences.

Meanwhile, exploring the combustion characteristics and misfire mechanisms in passive pre-chamber

direct-injection gasoline engines presents a specialized avenue. ZHOU *et al.* (2023) contributed to this area, focusing on understanding the intricate combustion behavior and misfire processes unique to engines employing passive pre-chamber technology for gasoline direct injection. This method delves into the specificities of how these engines operate and experience misfire events, offering insights crucial for their optimization.

Moreover, a holistic approach to misfire detection in aircraft engines involves integrating both linear signal analysis for pattern recognition and non-linear methods. The study highlighted by SYTA *et al.* (2021) underscores this comprehensive strategy, demonstrating how combining diverse analytical techniques ensures a robust and multifaceted approach to identifying misfire incidents.

However, vibration signals offer an easy measurable method without affecting engine function (HMIDA *et al.*, 2021; SHARMA *et al.*, 2014). The examination of vehicle engine vibrations serves as a profound gateway to deciphering the inner workings of an internal combustion engine. These vibrations intricate in their composition, encapsulate a wealth of data pertaining dynamical performance of the engine (SHARMA *et al.*, 2014; DU *et al.*, 2021). The scrutiny of these vibration signals holds immeasurable significance as a diagnostic tool, particularly in isolating specific engine misfires (HMIDA *et al.*, 2021; SHARMA *et al.*, 2014). The majority of research has focused on utilizing various sensors like accelerometers, acoustic sensors, and knock sensors to measure vibration signals (FIRMINO *et al.*, 2021; TAO *et al.*, 2019), yielding satisfactory outcomes and broad implementation in monitoring machinery conditions like pumps, ball bearings, and gearboxes (AHSAN, BISMOR, 2022; AHSAN *et al.*, 2023). Nevertheless, these methods encounter challenges in assessing specific conditions of vehicle engines due to the nonstationary nature of the measured signals when conventional analysis techniques are applied.

A diverse array of sensors contributes to the comprehensive analysis of engine vibrations (BISMOR, 2019). Comparative studies in the literature have examined a piezoelectric acceleration sensor specifically designed for engine and transmission vibration measurement against commercially produced accelerometer signals (BISMOR, 2019; GÜL *et al.*, 2021). Existing literature consistently demonstrates that engine vibration intensity correlates with increased engine speed and load variations (YAŞAR *et al.*, 2019).

The renowned piezoelectric sensors boast remarkable sensitivity in detecting high-frequency vibrations, efficiently translating mechanical impulses into electrical signals. Conversely, MEMS sensors, particularly accelerometers, stand out for their adaptability and precision in registering vibrations across various axes (LI *et al.*, 2020; AHSAN, BISMOR, 2023). Despite higher noise levels compared to piezoelectric sensors, the

compact size, affordability, and expansive capacity of MEMS sensors to measure a broad spectrum of vibration frequencies make them particularly suited for the intricate demands of engine monitoring and diagnostic purposes. These attributes collectively render MEMS sensors highly adept for addressing the multifaceted needs of engine analysis and diagnostics.

MEMS sensors, despite their versatility, encounter two prominent challenges in the realm of engine vibration analysis (BISMOR, 2019). Firstly, the vibration signals obtained through MEMS sensors often exhibit higher levels of noise compared to their piezoelectric counterparts. This discrepancy in noise levels stems from several factors rooted in the construction and operating principles of MEMS sensors. The miniaturized size of MEMS sensors, while advantageous for their widespread applicability and cost-effectiveness, can also make them more susceptible to environmental interferences and internal noise generated within the sensor itself. Additionally, the fabrication process and material composition of MEMS sensors may introduce inherent noise that affects the accuracy of vibration signal acquisition. Literature and academic research underscore the significance of addressing these noise factors in MEMS sensors to enhance their performance and reliability in capturing vibration data accurately (ROSSI *et al.*, 2023). Secondly, MEMS sensors necessitate meticulous calibration to ensure synchronization between the reference frequency and the recorded frequency of vibration signals (AHSAN, BISMOR, 2023). Achieving this synchronization is critical for precise analysis and interpretation of the acquired data.

The research endeavors in this study encompass a comprehensive analysis of engine vibrations using the MEMS sensor technology, specifically focusing on the ADXL1002 accelerometer paired with the Beaglebone Black microcontroller. The ADXL1002 accelerometer interfaced with the BeagleBone Black aims to minimize costs compared to pricier piezoelectric sensors while ensuring efficient data acquisition and analysis. This cost-effective approach enhances access to reliable vibration data crucial for accurately detecting misfires at diverse engine operating speeds. A critical precursor to this investigation involves the calibration of the ADXL1002 accelerometer with the Beaglebone Black setup. Calibration procedures were meticulously conducted utilizing the vibrator and signal generator to ensure accuracy and reliability in capturing vibration data. Detailed insights into this calibration process are extensively documented in the conference paper (AHSAN, BISMOR, 2023), providing a foundation for the subsequent experimentation and analysis conducted in this research. The primary objective is to capture vibration signals emitted by the vehicle engine across various operational states, including unloaded, loaded, and misfire conditions at dif-

ferent RPMs. In conjunction with the acquisition of this raw vibration data, frequency-domain signal processing techniques are employed to meticulously analyze and diagnose the distinct signatures of misfire occurrences across varying engine speeds and loads. These techniques encompass the fast Fourier transform (FFT), envelope spectrum (ES), and empirical mode decomposition (EMD), each tailored to discern and characterize the nuanced vibration patterns associated with misfires at different operational conditions. This multifaceted approach aims to enhance the diagnostic capabilities for detecting and differentiating misfire events, contributing to a more robust understanding of engine performance under diverse circumstances.

2. Materials and methods

This section delineates the developed prototype of the ADXL1002 accelerometer interfaced with the Beaglebone Black and the meticulous calibration process undertaken to ensure its accuracy and reliability in capturing vibration data. Additionally, it details the methodology employed in recording vibration data from the vehicle engine across diverse operational conditions, encompassing normal, loaded, and misfire scenarios at various RPMs, and the acquired data is presented. This section also encompasses an in-depth description of the different signal processing techniques utilized for diagnosing misfire conditions, outlining the specific methodologies employed in the analysis of the acquired vibration data.

2.1. Calibration of ADXL1002 accelerometer

The ADXL1002 is a high-performance accelerometer that measures acceleration in a single in-plane axis. It has an analog output that is proportional to the supply voltage and a linear frequency response range from DC to 11 kHz. It also has a low noise density of $25 \mu\text{g}/\sqrt{\text{Hz}}$ and a high resonant frequency of 21 kHz, which make it suitable for vibration and shock sensing applications. The ADXL1002 has a self-test function, a sensitivity stability of 5% over temperature, and a low cross axis sensitivity of $\pm 1\%$. It operates from a single supply and has a low power consumption of 1.0 mA. It also has a standby mode that reduces power consumption and allows fast recovery. The ADXL1002 can withstand temperatures from -40°C to $+125^\circ\text{C}$ and comes in a small $5 \text{ mm} \times 5 \text{ mm} \times 1.80 \text{ mm}$ LFCSP package. It is RoHS compliant and meets the environmental standards for electronic devices (Analog Devices, n.d.).

The calibration protocol for the ADXL1002 accelerometer involves a systematic series of steps designed to ensure the accuracy and reliability of data collection (AHSAN, BISMOR, 2023). Initial setup includes interfacing the accelerometer with the Beagle-

Bone Black microcontroller, establishing robust communication protocols between the two systems, laying the groundwork for subsequent calibration stages.

The experimental setup for calibration, detailed in Fig. 1, incorporates essential components such as a signal generator, power amplifier, vibration exciter, measuring amplifier, the MEMS accelerometer (ADXL1002), and the BeagleBone Black microcontroller, all integral to the calibration procedure. In the experimental setup, sinusoidal signals spanning different frequencies were generated by the function generator, initiating the attached vibration exciter to induce vertical vibrations. An ADXL1002 accelerometer, interfaced with the BeagleBone Black, captured and recorded these induced vibrations.

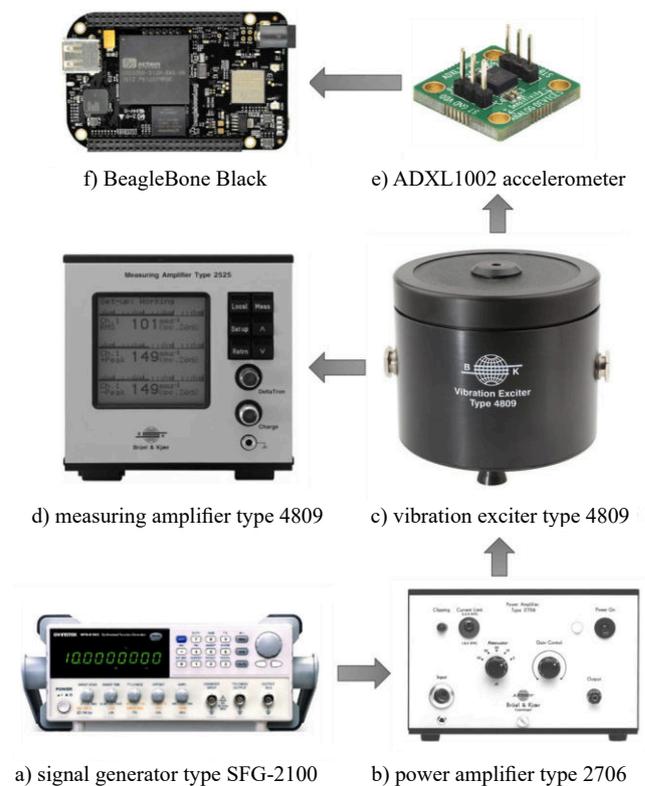


Fig. 1. Controlled environment setup for ADXL1002 accelerometer calibration: a) signal generator type SFG-2100; b) power amplifier type 2706; c) vibration exciter type 4809; d) measuring amplifier type 2525; e) ADXL1002 accelerometer; f) BeagleBone Black.

The FFT was utilized to analyze the time-domain vibration signals produced by the vibration exciter, computing the recorded frequencies and subsequently comparing them with the reference frequencies of the input signal to the vibration exciter. Remarkably, the plot depicted in Fig. 2 reveals a linear relationship between the reference and measured frequencies, affirming the efficiency of the ADXL1002 accelerometer and reliable detection of input frequencies.

Following the frequency comparison analysis, the sensitivity of the ADXL1002 accelerometer to recorded

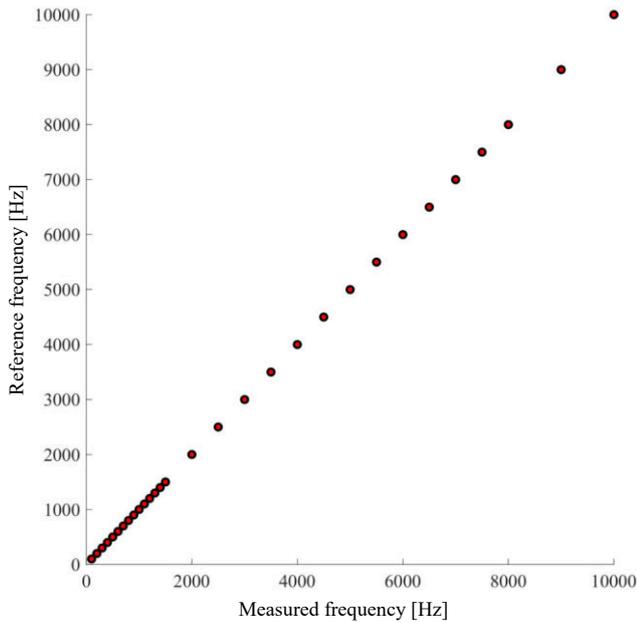


Fig. 2. Comparison of reference and computed frequencies using FFT analysis.

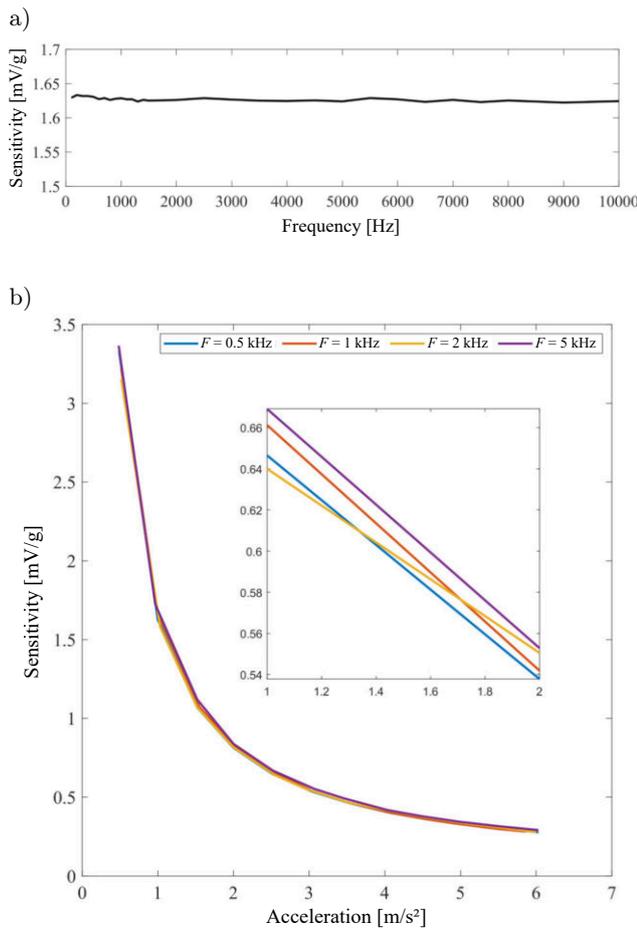


Fig. 3. Sensitivity analysis of ADXL1002 accelerometer: a) sensitivity plotted against frequencies showcasing consistent sensitivity across frequency ranges; b) sensitivity variation with gravitational forces, highlighting an inverse relationship between sensitivity and gravitational levels.

vibration signals was determined using the following equation:

$$S_i = \frac{\text{RMS}(v(f_i))}{g_i}, \quad (1)$$

where $v(f_i)$ represents the time-domain vibration signal at a specific frequency f_i with $i = 1, 2, 3, \dots, k$, g_i denotes the acceleration of the vibration signals, and $\text{RMS}(\cdot)$ denotes the root-mean-square of the signal. This sensitivity analysis was depicted in Fig. 3a, demonstrating consistent sensitivity levels across various frequency ranges. Additionally, Fig. 3b illustrates the inverse proportionality between sensitivity and gravitational force, indicating that higher gravitational levels resulted in reduced sensitivity.

Utilizing the developed prototype of the ADXL1002 accelerometer interfaced with the BeagleBone Black, the calibration process lays the foundation for advanced applications in misfire detection within vehicle engine vibration data.

2.2. Vibration dataset of vehicle engine using ADXL1002 accelerometer

Figure 4 visually depicts the placement of the accelerometer on the vehicle engine, illustrating its positioning for data capture purposes. The vehicle engine vibration dataset given in Table 1 offers a structured overview of engine conditions across various RPMs, loads, and misfire occurrences. Each entry in the table corresponds to a specific engine scenario, denoted by RPM (revolutions per minute), load intensity, and misfire status. The dataset captures instances across different RPMs, including 1500, 2500, and 3000, coupled with load conditions ranging from no load to half load and full load. Moreover, the dataset signifies whether a misfire was present or absent in each specific scenario. This dataset serves as a comprehensive representation of engine performance variations under different operational settings, allowing for in-depth analysis and exploration of how RPM, load, and misfire interrelate within the context of vehicle vibrations.

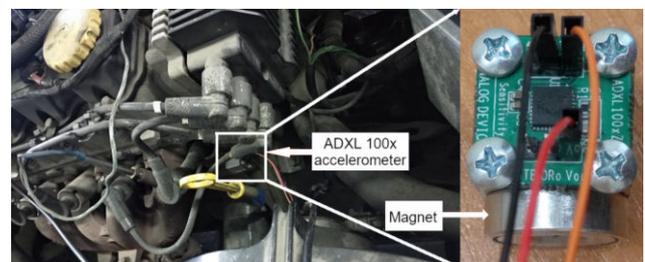


Fig. 4. Placement of ADXL1002 accelerometer on vehicle engine for vibration data collection.

To further elucidate the engine vibration dataset captured through the ADXL1002 accelerometer, a representation in the time domain is crucial for visual

Table 1. Vehicle engine vibration dataset.

RPM	Frequency [Hz]	Load	Misfire
1500	25	no load	no misfire
3000	50	no load	no misfire
1500	25	half load	no misfire
1500	25	full load	no misfire
2500	41.667	half load	no misfire
2500	41.667	full load	no misfire
3000	50	half load	no misfire
3000	50	full load	no misfire
1500	25	half load	misfire
1500	25	full load	misfire
2500	41.667	half load	misfire
2500	41.667	full load	misfire
3000	50	half load	misfire
3000	50	full load	misfire

comprehension. The selected vibration signals, depicted in Fig. 5 showcase the temporal characteristics of engine vibrations across distinct operational states. These signals, plotted against time, offer a direct insight into the fluctuations and patterns within the recorded vibrations during various engine conditions at 3000 RPMs, such as load levels, and misfire occurrences.

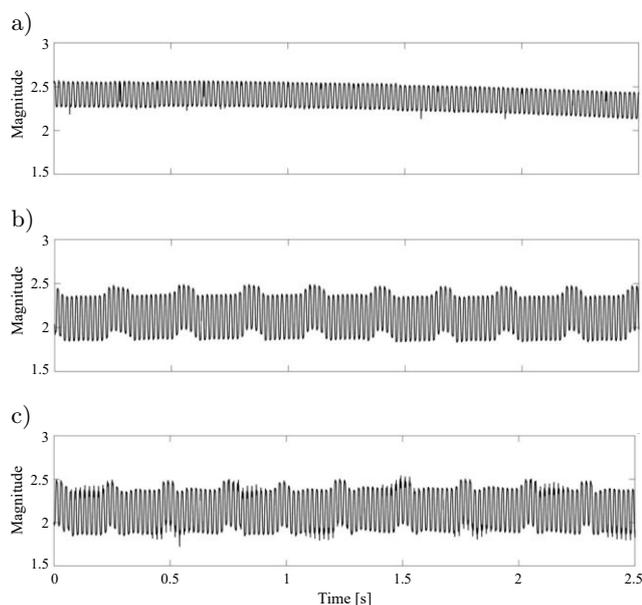


Fig. 5. Engine vibration signals at 3000 RPMs under varied load conditions and misfire scenarios: a) unloaded without misfire condition; b) loaded without misfire condition; c) loaded with misfire condition.

2.3. Frequency-domain approaches for misfire diagnoses

Misfires in a vehicle engine indeed have a tangible impact on the vibration signal's spectrum. These dis-

ruptions occur when the typical firing sequence of the engine cylinders is disturbed, resulting in irregularities in the vibration pattern. The introduction of additional frequency components, particularly around the harmonics or multiples of the engine's firing frequency, is a hallmark of misfires. Consequently, these irregular firing patterns lead to distinct spikes or alterations in the spectral content of the vibration signal. Detecting these deviations becomes pivotal for diagnostics and maintenance, necessitating advanced signal processing techniques capable of differentiating between load-induced alterations and the presence of misfires within the intricate vibration signals generated during engine operation.

2.3.1. FFT approach for misfire detection

Utilizing the FFT in the misfire and engine analysis involves decomposing vibration signals into their fundamental frequency components, unveiling spectral patterns inherent within the signals (LIN, YE, 2019).

The FFT algorithm computes the discrete Fourier transform (DFT) efficiently. Mathematically, the DFT of a discrete-time function $f(n)$ can be represented as follows (LIN, YE, 2019):

$$F(k) = \sum_{n=0}^{N-1} f(n)e^{-i2\pi kn/N}, \quad (2)$$

where $F(k)$ denotes the complex values within the frequency domain at a specific index k . This complex value represents the transformed signal's amplitude and phase at a particular frequency component. On the other hand, $f(n)$ signifies the discrete-time signal in the time domain at a distinct index n . The variable N represents the total count of samples constituting the time-domain signal. The term $e^{-i2\pi kn/N}$ is a complex exponential expression encapsulating the phase shift and frequency of individual components within the signal.

2.3.2. Envelope spectrum approach for misfire detection

The computation of the ES stands as an essential method to discern nuanced variations induced by both load and misfire events. This technique entails the extraction of the signal's envelope, thereby highlighting alterations in the overall vibration behavior attributed to changes in load or misfires. This approach furnishes accurate diagnostic insights into the operational state of the engine, pinpointing specific modifications in the vibration patterns influenced by varying loads or misfire occurrences (AHSAN, BISMOR, 2022).

The ES represents the amplitude of specific frequency components within a signal $f(n)$. Mathematically, the computation of the ES involves obtaining the

magnitude spectrum of the positive frequency components obtained from the signal's FFT. The equation for the ES for a signal $f(n)$ can be expressed as $ES(k)$:

$$ES(k) = \text{FFT}_{\text{positive}}(k), \quad (3)$$

where the symbol $ES(k)$ signifies the ES at a specific index k within the frequency domain representation. Meanwhile, $\text{FFT}_{\text{positive}}(k)$ represents the positive frequency components extracted from the FFT computation conducted on the signal. The operation $\|\cdot\|$ denotes the magnitude operation, commonly known as the absolute value operation. This mathematical operation retrieves the amplitude information from the positive frequency components derived through the FFT process, allowing for the extraction of the ES showcasing the amplitudes of distinct vibration frequencies within the signal.

In essence, the ES is derived by taking the absolute values of the positive frequency components obtained through the FFT process of the signal $f(n)$. This representation highlights the amplitudes of selected vibration frequencies within the signal.

2.3.3. EMD approach for misfire detection

The EMD serves as an invaluable tool in analyzing the intricacies of non-linear and non-stationary signals, offering a means to disentangle them into components of varying resolutions (LIU *et al.*, 2021). The process of EMD involves several steps. Initially, extrema, comprising local maxima and minima points, are identified within the signal $f(n)$. Subsequently, upper and lower envelopes are formed by connecting these extrema points. The mean signal, obtained by calculating the mean of the upper and lower envelopes, is then subtracted from the original signal $f(n)$ to derive the first intrinsic mode function (IMF). This process is iteratively applied to the obtained IMF, treating it as the new signal in each iteration, until specific convergence or stopping criteria are met. The iterative nature of EMD allows it to adapt to the characteristics of the input signal, resulting in the successive extraction of IMFs that collectively represent the signal's intrinsic oscillatory modes.

Misfires within the engine introduce additional, unfamiliar frequencies and harmonics into these vibration signals, intensifying the intricacy of the diagnostic process. To dissect and interpret these signals accurately, advanced signal processing techniques are crucial in distinguishing, analyzing, and understanding the new frequency components attributed to misfires. EMD, by its inherent nature, offers a potent means to handle these challenges. Its ability to effectively decompose signals into varying resolutions aligns with the demands posed by the intricate nature of vibration signals, especially in the presence of misfires. Thus, EMD emerges as a suitable technique to disentangle and in-

terpret these complex signals, facilitating a deeper understanding of the new frequency components introduced by misfires and aiding in precise engine diagnostics.

3. Results and discussion

Figure 6 illustrates the FFT representations of vibration signals at 1500 RPMs under varied load conditions and misfire scenarios: unloaded without misfire, loaded without misfire, and loaded with misfire. Notably, the FFT demonstrates harmonic frequency components within the loaded vibration signals, yet fails to reveal the 1500 RPM (25 Hz) frequency component in the presence of a misfire, as indicated in Fig. 6.

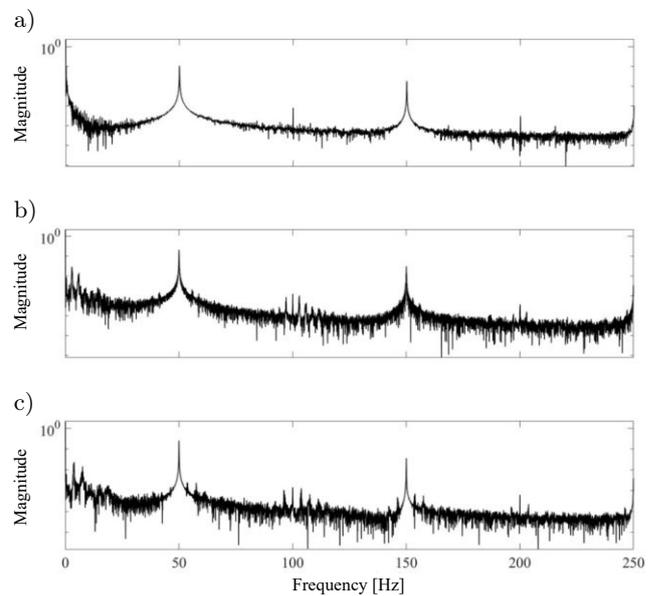


Fig. 6. FFT of raw vibration signals at 1500 RPMs under varied load conditions and misfire scenarios: a) unloaded without misfire condition; b) loaded without misfire condition; c) loaded with misfire condition.

Vibration signals recorded from vehicle engines via MEMS sensors, like the ADXL1002 accelerometer, often encompass unwanted frequencies that deteriorate the signal-to-noise ratio. During engine misfires, the power associated with the misfire frequency substantially diminishes, rendering it imperceptible within the FFT representation (Fig. 6).

However, misfires within the engine generate discernible periodic impacts in the time-domain vibration signals, as depicted in Fig. 5. ES analysis serves as a potent frequency-domain signal processing tool capable of highlighting misfire frequencies within vibration signals. Figure 7 showcases the ES representations of loaded signals without misfires, at 1500 RPMs and 2500 RPMs. These representations underscore the effectiveness of ES in discerning the absence of misfires

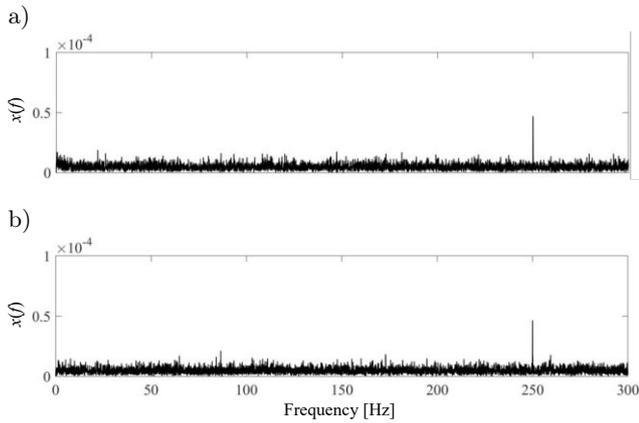


Fig. 7. Envelope spectrum of raw vibration signals at different RPMs under loaded and no misfire conditions: a) 1500 RPMs; b) 2500 RPMs.

within the engine, thereby exhibiting a clear spectral output.

Moreover, Fig. 8 illustrates the misfire frequency at 1500 RPMs (25 Hz) for both half load and full load conditions. Similarly, Figs. 9 and 10 showcase the ES representations at 2500 RPMs (41.667 Hz) and 3000 RPMs (50 Hz), correspondingly. Specifically, Fig. 9a displays ES for half load, while Fig. 9b present ES for full load. Similarly, Fig. 10a exhibits ES for half load, and Fig. 10b showcases ES for full load, illustrating the frequency components pertinent to their respective misfire frequencies.

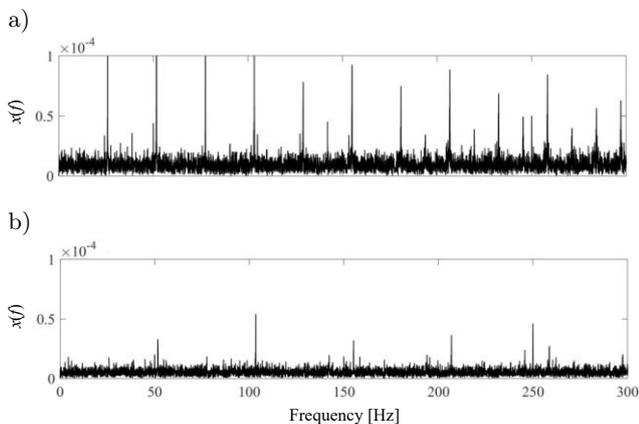


Fig. 8. Envelope spectrum of raw vibration signals at 1500 RPMs under loaded and misfire conditions: a) half load; b) full load.

The observations and discussions from the preceding analysis suggest that the ES offers enhanced proficiency in detecting misfires compared to the FFT applied to the raw vibration signals acquired using the ADXL1002 accelerometer. The diagnosis of misfires in vehicle engines utilizing the ADXL1002 accelerometer revolves around identifying periodic impulses. Upon misfire detection, requisite measures for periodic main-

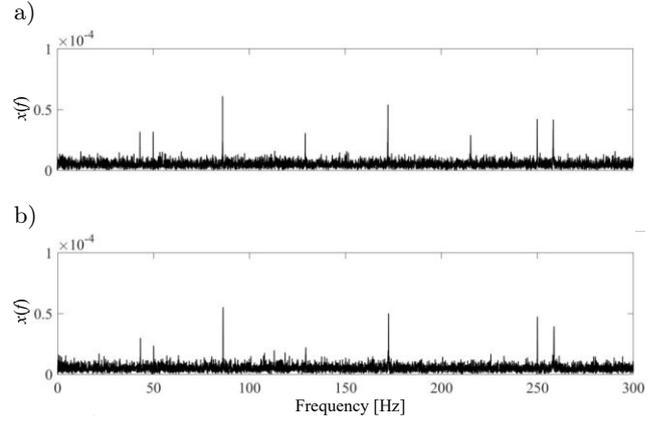


Fig. 9. Envelope spectrum of raw vibration signals at 2500 RPMs under loaded and misfire conditions: a) half load; b) full load.

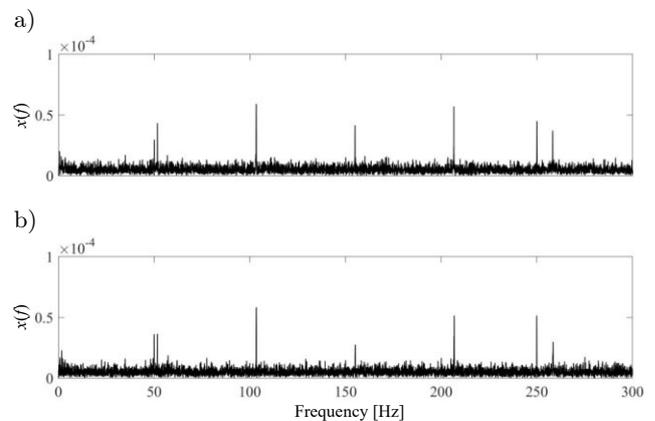


Fig. 10. Envelope spectrum of raw vibration signals at 3000 RPMs under loaded and misfire conditions: a) half load; b) full load.

tenance are undertaken. However, this task poses challenges due to the inherent characteristics of low signal-to-noise ratio and interference from unwanted external signals, as evidenced in the FFT results depicted in Fig. 6.

Figures 11 and 12 illustrate the IMF components pertaining to loaded conditions at 2500 RPMs and 3000 RPMs, respectively. Specifically, Fig. 11 displays the IMF components at 2500 RPMs for the loaded condition without any misfires, while Fig. 12 showcases the IMF components at 3000 RPMs under loaded conditions with misfires.

Figure 13 exhibits the FFT analysis conducted on the first IMF extracted from vibration signals observed at 1500 RPMs, 2500 RPMs, and 3000 RPMs, accounting for both half load and full load conditions. In Figs. 13a and 13b, the display illustrates the presence of the misfire frequency at 1500 RPMs (25 Hz), accompanied by discernible side harmonic frequency components, all stemming from the combined influence of misfires and consistent load conditions – either half load or full load.

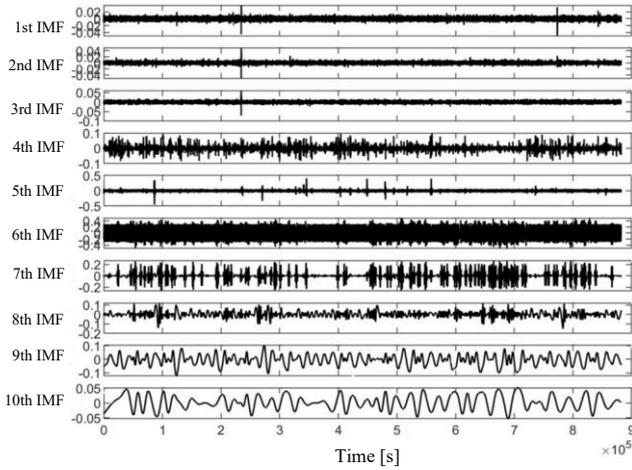


Fig. 11. EMD of vibration signal at 2500 RPMs under loaded and without misfire condition.

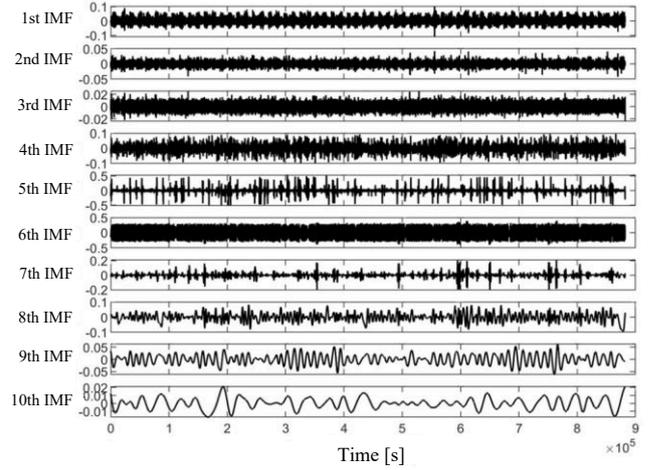


Fig. 12. EMD of vibration signal at 3000 RPMs under loaded and misfire condition.

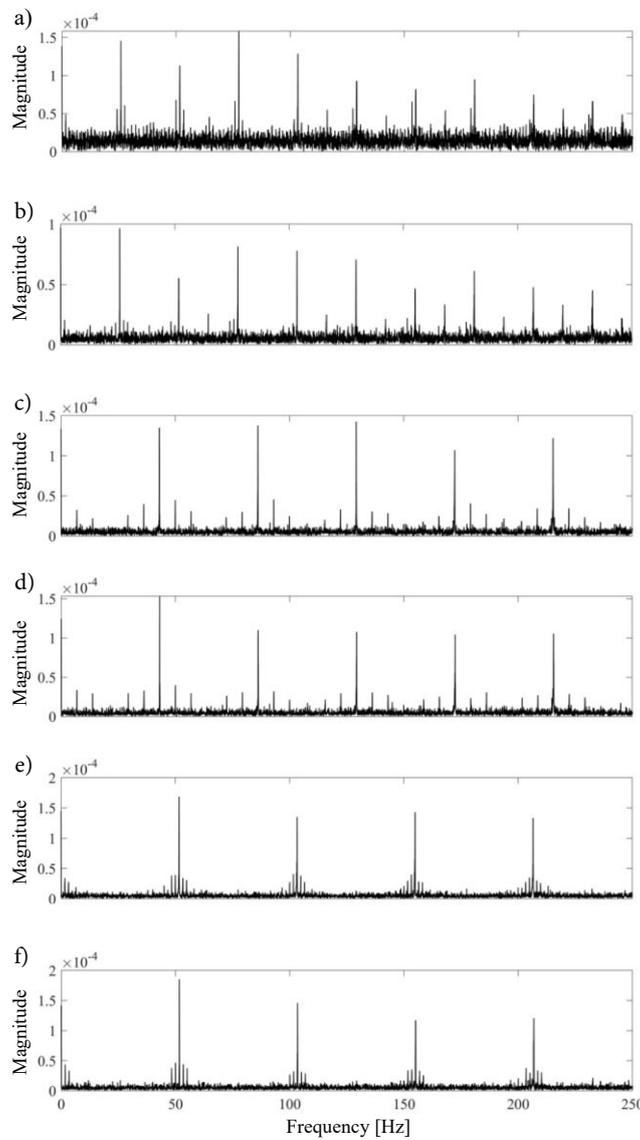


Fig. 13. FFT of the first IMF at different RPMs under loaded conditions and misfire scenarios: a) 1500 RPMs with half load; b) 1500 RPMs with full load; c) 2500 RPMs with half load; d) 2500 RPMs with full load; e) 3000 RPMs with half load; f) 3000 RPMs with full load.

Similarly, Figs. 13c and 13d delineate the 42 Hz misfire frequency, while Figs. 13e and 13f showcase the 50 Hz misfire frequency, specifically identifiable at distinct engine speeds. This analysis, applied to the first IMF, significantly reduces unwanted signals inherent in raw vibration data, offering a more refined representation compared to the observations depicted in Fig. 6.

4. Conclusions

This research highlights the vital role of MEMS sensors, notably the ADXL1002 accelerometer interfaced with the Beaglebone Black microcontroller, in diagnosing complex engine malfunctions. Specifically, it focuses on identifying misfires within intricate vibration datasets. The use of these sensors is crucial in efficiently detecting misfires and other engine irregularities.

The calibration process, meticulously detailed in a previous work (AHSAN, BISMOR, 2023), stands as a fundamental aspect of this study. This calibration ensured precision and reliability in capturing vibration data, serving as a solid foundation for subsequent experiments and analyses.

In this research, the primary objective was to record vibration signals across diverse operational scenarios, encompassing unloaded, loaded, and misfire conditions at varying RPMs. The vehicle engine was equipped with the ADXL1002 accelerometer, and vibration data were systematically recorded under distinct conditions: at 1500 RPMs, 2500 RPMs, and 3000 RPMs. The recorded scenarios included unloaded without misfire, loaded without misfire, and loaded with misfire conditions. Subsequently, the collected data was presented and analyzed in both the time-domain and the frequency-domain to visualize the effects of misfires and varying loads on vibration signals.

Additionally, this involved the application of advanced frequency-domain signal processing techniques such as FFT, ES, and EMD. These methods were carefully designed to distinguish distinctive patterns signifying engine misfires, specifically when considering diverse engine RPMs and loads. This strategic approach substantially amplified our capability to diagnose and understand potential engine issues. The efficiency of applying FFT directly to the raw vibration data was hindered by the presence of unwanted signals and external noises, making misfire detection challenging. To address this issue, EMD was employed to decompose the vibration signal into distinct frequencies. Subsequently, FFT was applied to the first IMF to pinpoint the misfire frequency variations at different RPMs. This method proved effective in diagnosing misfire frequencies within the vibration signals obtained through the use of the ADXL1002 accelerometer in the vehicle engine. Furthermore, it not only enabled the pre-

cise identification of misfires but also provided intricate insights into the precise vibrational characteristics associated with varying engine conditions.

References

1. AHSAN M., BISMOR D. (2022), Early-stage fault diagnosis for rotating element bearing using improved harmony search algorithm with different fitness functions, *IEEE Transactions on Instrumentation and Measurement*, **71**: 1–9, doi: [10.1109/TIM.2022.3192254](https://doi.org/10.1109/TIM.2022.3192254).
2. AHSAN M., BISMOR D. (2023), Calibration of a high sampling frequency MEMS-based vibration measurement system, [in:] Pawelczyk M., Bismor D., Ogonowski S., Kacprzyk J. [Eds.], *Advanced, Contemporary Control. PCC 2023. Lecture Notes in Networks and Systems*, Vol. 708, Springer, Cham, doi: [10.1007/978-3-031-35170-9_28](https://doi.org/10.1007/978-3-031-35170-9_28).
3. AHSAN M., BISMOR D., MANZOOR M.A. (2023), ARL-Wavelet-BPF optimization using PSO algorithm for bearing fault diagnosis, *Archives of Control Sciences*, **33**(3): 589–606, doi: [10.24425/acs.2023.146961](https://doi.org/10.24425/acs.2023.146961).
4. Analog Devices (n.d.), ADXL1002 Accelerometer Data-sheet. Analog Devices, Inc., <https://www.analog.com/en/products/adxl1002.html> (access: 2022).
5. BISMOR D. (2019), Analysis and comparison of vibration signals from internal combustion engine acquired using piezoelectric and MEMS accelerometers, *Vibration in Physical Systems*, **30**(1): 2019112.
6. DU C., JIANG F., DING K., LI F., YU F. (2021), Research on feature extraction method of engine misfire fault based on signal sparse decomposition, *Shock and Vibration*, doi: [10.1155/2021/6650932](https://doi.org/10.1155/2021/6650932).
7. FIRMINO J.L., NETO J.M., OLIVEIRA A.G., SILVA J.C., MISHINA K.V., RODRIGUES M.C. (2021), 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, doi: [10.1007/s40430-021-03052-y](https://doi.org/10.1007/s40430-021-03052-y).
8. GÜL M., KARAER M., DOĞAN A. (2021), Design of piezoelectric acceleration sensor for automobile applications, *International Journal of Automotive Science And Technology*, **5**(4): 398–403, doi: [10.30939/ijastech..1006197](https://doi.org/10.30939/ijastech..1006197).
9. HMIDA A., HAMMAMI A., CHAARI F., AMAR M.B., HADDAR M. (2021), Effects of misfire on the dynamic behavior of gasoline Engine Crankshafts, *Engineering Failure Analysis*, **121**: 1–19, doi: [10.1016/j.engfailanal.2020.105149](https://doi.org/10.1016/j.engfailanal.2020.105149).
10. KUMANO K., AKAGI Y., MATOHARA S., UCHISE Y., YAMASAKI Y. (2020), Using an ion-current sensor integrated in the ignition system to detect precursory phenomenon of pre-ignition in gasoline engines, *Applied Energy*, **275**: 115341, doi: [10.1016/j.apenergy.2020.115341](https://doi.org/10.1016/j.apenergy.2020.115341).
11. LI S., ZHANG Y., WANG L., XUE J., JIN J., YU D. (2020), A CEEMD method for diesel engine misfire fault diagnosis based on vibration signals, [in:] *2020*

- 39th Chinese Control Conference (CCC), pp. 6572–6577, doi: [10.23919/CCC50068.2020.9189312](https://doi.org/10.23919/CCC50068.2020.9189312).
12. LIN H.-C., YE Y.-C. (2019), Reviews of bearing vibration measurement using fast Fourier transform and enhanced fast Fourier transform algorithms, *Advances in Mechanical Engineering*, **11**(1): 1–12, doi: [10.1177/1687814018816751](https://doi.org/10.1177/1687814018816751).
 13. LIU X., SHI G., LIU W. (2021), An improved empirical mode decomposition method for vibration signal, *Wireless Communications and Mobile Computing*, pp. 1–8, doi: [10.1155/2021/5525270](https://doi.org/10.1155/2021/5525270).
 14. NAVEEN VENKATESH S. *et al.* (2022), Misfire detection in spark ignition engine using transfer learning, *Computational Intelligence and Neuroscience*, pp. 1–13, doi: [10.1155/2022/7606896](https://doi.org/10.1155/2022/7606896).
 15. ROSSI A., BOCCHETTA G., BOTTA F., SCORZA A. (2023), Accuracy characterization of a MEMS accelerometer for vibration monitoring in a rotating framework, *Applied Sciences*, **13**(8): 5070, doi: [10.3390/app13085070](https://doi.org/10.3390/app13085070).
 16. SHARMA A., SUGUMARAN V., BABU DEVA SENAPATI S. (2014), Misfire detection in an IC engine using vibration signal and decision tree algorithms, *Measurement*, **50**: 370–380, doi: [10.1016/j.measurement.2014.01.018](https://doi.org/10.1016/j.measurement.2014.01.018).
 17. SYTA A., CZARNIGOWSKI J., JAKLIŃSKI P. (2021), Detection of cylinder misfire in an aircraft engine using linear and non-linear signal analysis, *Measurement*, **174**: 108982, doi: [10.1016/j.measurement.2021.108982](https://doi.org/10.1016/j.measurement.2021.108982).
 18. TAMURA M., SAITO H., MURATA Y., KOKUBU K., MORIMOTO S. (2011), Misfire detection on internal combustion engines using exhaust gas temperature with low sampling rate, *Applied Thermal Engineering*, **31**(17–18): 4125–4131, doi: [10.1016/j.applthermaleng.2011.08.026](https://doi.org/10.1016/j.applthermaleng.2011.08.026).
 19. TAO J., QIN C., LI W., LIU C. (2019), Intelligent fault diagnosis of diesel engines via extreme gradient boosting and high-accuracy time-frequency information of vibration signals, *Sensors*, **19**(15): 3280, doi: [10.3390/s19153280](https://doi.org/10.3390/s19153280).
 20. WANG J. *et al.* (2022), Misfire and knock detection based on the ion current inside a passive pre-chamber of gasoline engine, *Fuel*, **311**: 122528, doi: [10.1016/j.fuel.2021.122528](https://doi.org/10.1016/j.fuel.2021.122528).
 21. YAŞAR A., KESKIN A., YILDIZHAN S., ULUDAMAR E. (2019), Emission and vibration analysis of diesel engine fuelled diesel fuel containing metallic based nanoparticles, *Fuel*, **239**: 1224–1230, doi: [10.1016/j.fuel.2018.11.113](https://doi.org/10.1016/j.fuel.2018.11.113).
 22. ZHOU H., MENG S., HAN Z. (2023), Combustion characteristics and misfire mechanism of a passive pre-chamber direct-injection gasoline engine, *Fuel*, **352**: 129067, doi: [10.1016/j.fuel.2023.129067](https://doi.org/10.1016/j.fuel.2023.129067).

Research Paper

Feasibility of Using Wavelet Analysis and Machine Learning Method
in Technical Diagnosis of Car Seats

Cezary BARTMAŃSKI, Alicja BRAMORSKA*

*Department of Acoustics, Electronics and IT Solutions Central Mining Institute
National Research Institute
Katowice, Poland**Corresponding Author e-mail: abramorska@gig.eu*(received October 24, 2023; accepted June 10, 2024; published online August 27, 2024)*

This paper presents the results of preliminary research aimed at developing a method for rapid, non-contact diagnostics of the electric drive of car seats. The method is based on the analysis of acoustic signals produced during the operation of the drive. Pattern recognition and machine learning processes were used in the diagnosis. A method of feature extraction (diagnostic symptoms) using wavelet decomposition of acoustic signals was developed. The discriminative properties of a set of diagnostic symptoms were tested using the “Classification Learner” application available in MATLAB. The obtained results confirmed the usefulness of the developed method for the technical diagnostics of car seats.

Keywords: acoustic diagnostics; wavelet decomposition; machine learning.



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

The subject of this study is the mechanical construction of a car seat with a driving device allowing to change the angle of the seat backrest. The device consists of an electric motor and a gearbox (Fig. 1) and is a compact module without the possibility of disassembling it, for example, for repair. In the event of failure or malfunction, the manufacturer replaces the entire module.

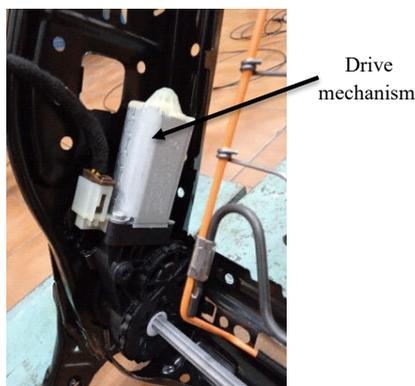


Fig. 1. Appearance of the drive mechanism of a car seat.

The main goal of the research is to develop a quick method for diagnosing the seat construction at the production stage. Due to the postulated speed of diagnosis and the small size of the drive device, it is decided to look for a solution based on the analysis of acoustic signals generated by the operating drive device. This choice is justified due to the generally well-known high content of relevant information about the state of the object in the acoustic signal generated during its operation (BASZTURA, 1996; LIN, 2001; GŁOWACZ, 2014; PAWLIK, 2019).

The starting point for the study was a small batch of seats (11 pieces) supplied by the manufacturer. Some of the seats were marked as good (meeting the manufacturer’s requirements) and the rest as defective. According to the manufacturer’s requirements, diagnostics of subsequent batches of seats should categorize them into two classes: good (technically efficient) and bad (defective) without identifying the specific defective component. The problem presented here falls under a topic referred to in literature as pattern recognition. It deals with the recognition of the affiliation of various objects to certain predefined classes. Objects within each class may differ more or less from each

other, and the number of objects within each class can be any.

The pattern recognition process is divided into two parts called feature extraction and classification. In the first part, characteristic features are extracted from the measured signal. In the second part, calculations are performed on the set of these features using the information contained in the so-called learning sequence. A learning sequence is a previously prepared set of features representing objects for which the correct classification is known. As a result of these calculations, a decision is made as to which class the recognized object belongs to (DUDA, HART, 1973; SUN *et al.*, 2004; XI *et al.*, 1997).

2. Test stand and measurement database

The work is performed on a test stand used in the laboratory for measuring the sound power level of mechanical equipment, which was adapted to the needs of the present work (Fig. 2).



Fig. 2. View of test stand.

The stand is located in a laboratory hall measuring $9.3 \text{ m} \times 7 \text{ m} \times 5.2 \text{ m}$. The hall has smooth walls and a floor covered with ceramic tiles to reflect acoustic waves. The room is partially filled with laboratory furniture. Following the guidelines in Annex A of (International Organization for Standardization, 2010), the mentioned test environment was classified as a rectangular industrial room with an average sound absorption coefficient of $a = 0.15$.

The seat structure is fixed to the floor by loading the seat base with two heavy steel fittings, ensuring its immobilization during testing (Fig. 2). In the initial stage of testing, microphones mounted on measurement stands were placed at the vertices and in the center of the walls of the virtual cuboid surrounding the test object – a total of 9 measurement points were established (Fig. 3). The dimensions of the cuboid were chosen so that the distance from the walls to the test

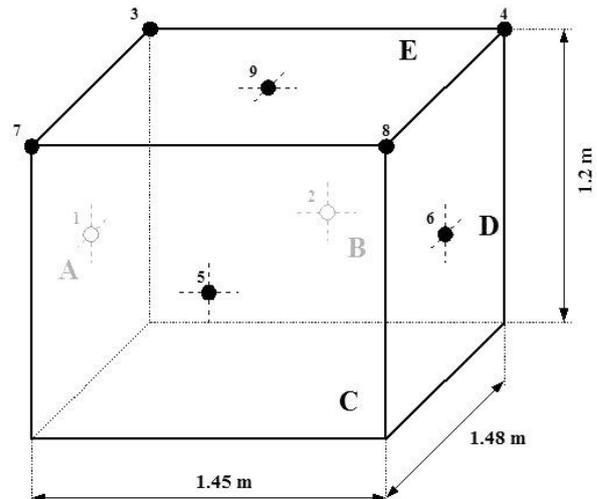


Fig. 3. Arrangement of microphones on test bench.

object was 0.5 m. The individual microphones were oriented so that their axes:

- were perpendicular to the measurement plane for measurement points no. 1, 2, 5, 6, and 9;
- indicated the point of intersection of the diagonals of the perpendicular for measurement points no. 3, 4, 7, and 8.

The small distance, compared to the size of the hall, minimizes the influence of the reflected wave on the recorded signal. Nevertheless, it should be noted that in the case of the presented research, it is not important to determine the exact value of the sound pressure level, but rather to determine the interrelationships of the different parts of the signal spectrum, as will be discussed later in the article.

The acoustic signal was recorded during the operation of the driving device for two directions of seat backrest movement: forward and backward. The time of backrest movement between extreme positions was approximately 20 s. The signals were recorded on a PULSE digital recorder, manufactured by Brüel & Kjær. The time courses of these signals and their measured features allow them to be classified, in accordance with the systematics used in the literature on signal analysis, as transient non-stationary signals (PIERSOL, 1989; SZABATIN, 2000). An example of the time course of the recorded signals is shown in Fig. 4.

Measurements of the acoustic background noise in the hall indicate that the S/N ratio of the recorded acoustic signals was more than 13 dB. This testifies to the negligible influence of the acoustic background noise on the signal under study.

It is necessary to ensure that the acoustic environment at the future location of the manufacturer's diagnostic station is comparable to the current one.

Initially, the signals were recorded in the acoustic range up to 20 kHz. Frequency analysis indicates that

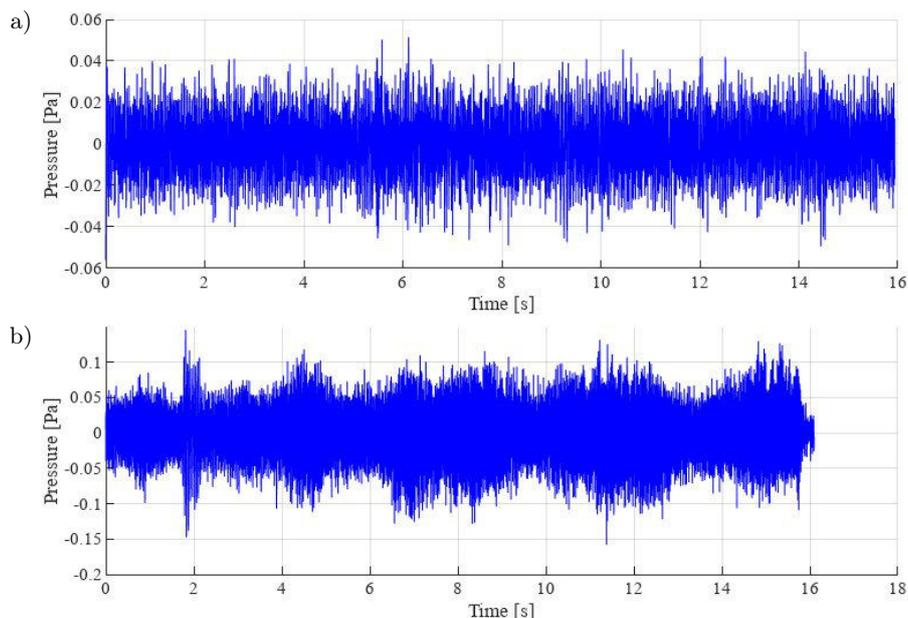


Fig. 4. Time course of the acoustic signal generated by the seat drive device: a) technically efficient seat; b) defective seat.

the spectra of these signals are dominated by components in the range of 20 Hz–60 Hz, and the maximum frequency practically does not exceed 200 Hz. Figure 5 shows an example of the spectra of the signals generated by the technically efficient and defective seats, respectively. Unfortunately, it was found that in each case, the differences of the spectra are not as clear as illustrated in the figures. Differences were also observed in the spectra of the signals generated for different directions of backrest movement for both the technically efficient and defective seats.

Measurements were made on all 11 seats – 5 seats marked as good and 6 seats marked as not meeting the

manufacturer's requirements. The recordings of measured signals, in the form of samples of instantaneous values, constituted the measurement database for the further part of the work.

The levels of signals recorded at particular points of the grid did not differ significantly; however, at measurement points 1, 2, and 5 (at the height of $h = 0.60$ m) this level was higher. Finally, for the planned future measurements, one point of microphone installation was selected, i.e., point 5, located near the drive mechanism of a seat, and for the recordings made at this point further signal processing and calculations were performed.

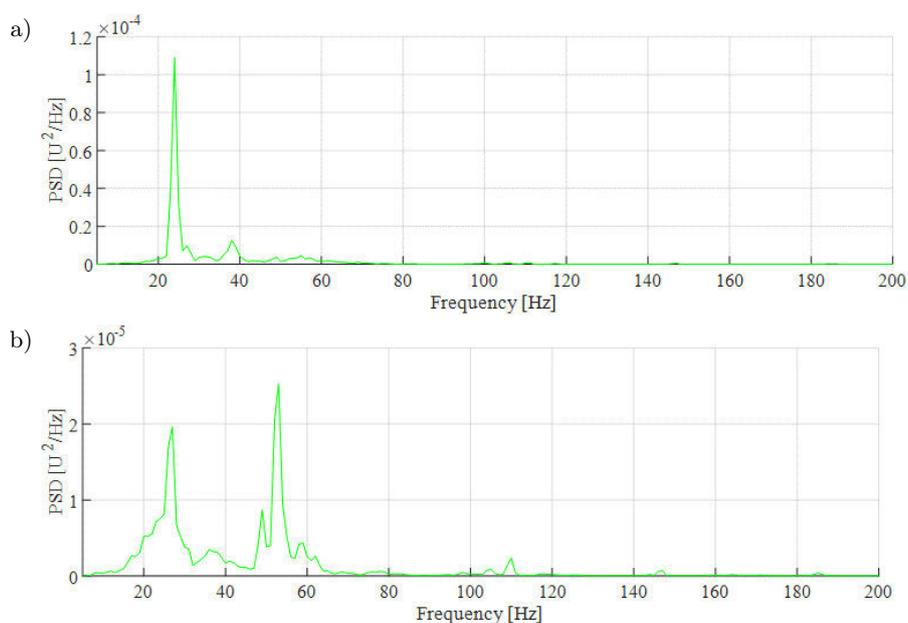


Fig. 5. Acoustic signal spectrum: a) technically efficient seat; b) defective seat.

3. Feature extraction

Vibroacoustic signals produced during the operation of mechanical equipment carry important information about the dynamic processes within them. Therefore, the analysis of vibroacoustic signals is one of the most important methods used in condition monitoring and technical diagnostics of equipment.

In practice, many methods of signal analysis, e.g., FFT, STFT, Wigner-Ville distribution are used in this area (TANG *et al.*, 2010). Among these, wavelet analysis is currently one of the most advanced tools of signal analysis, confirmed by numerous practical applications. It covers almost all aspects of technical diagnostics of mechanical equipment, including time-frequency analysis of signals, fault feature extraction, singularity detection for signals, denoising and extraction of weak signals, compression of signals, and system identification (BATKO *et al.*, 2005; HAN *et al.*, 2022; TANG *et al.*, 2010; PENG, CHU, 2004; LIN, QU, 2000; STASZEWSKI, 1998; HUANG, SOLORZANO, 1971; YANG *et al.*, 2022).

The review of the results of the spectral analysis of the acoustic signals produced by the working seat drive device indicates that there are differences in the spectra between the objects marked, according to the manufacturer's criteria, as good and bad. However, there are also differences between both good and bad objects.

The aim of processing recorded acoustic signals is to extract information about the individual characteristics of particular signals contained in their spectrum. In the discussed situation, it concerns the information allowing to make a decision about the technical condition of the tested object.

As the basic idea of creating the feature vector, we have established the development of a discrete representation of the acoustic signal into a functional series, followed by the separation of components of this expansion that carry significant energy of the signal. Coefficients of the selected components of this expansion will constitute the components of the feature vector (PENG, CHU, 2004).

In this paper, it was decided to use the method of wavelet analysis of signals to implement the feature extraction process. The basis of this analysis is the decomposition of the signal based on a set of orthogonal basis functions, called wavelets. The set of basis functions is generated by scaling and shifting the so-called mother wavelet in the time domain. The decomposition allows the signal to be represented as a superposition of wavelets. The coefficients of this superposition, called wavelet coefficients, are determined by the wavelet transform of the signal. The values of the wavelet coefficients measure the degree of correlation between the signal and the wavelet, making the proper choice of wavelet type crucial.

The values of the components of the feature vectors are a function of the type of wavelet selected for analysis. The purpose of selecting a particular wavelet is to obtain the strongest possible correlation between the signal with a small number of basis wavelets. As mentioned earlier, the values of wavelet coefficients are a measure of the degree of correlation between the wavelet and the signal.

The selection of the wavelet was done experimentally. A test was performed by decomposing the recorded signals using different wavelets and selecting the wavelet showing the highest degree of correlation with the signal. The parameter to be evaluated in this experiment was chosen as the maximum value of the modulus of the wavelet coefficients for each level of decomposition, according to the formula:

$$cDk_{\text{MAX}} = \max_{\{d_k\}} |d_k[i]|, \quad (1)$$

where k is the decomposition level index, i is the index of the element in the sequence of wavelet coefficients, $d_k[i]$ is the sequence of coefficients at the k -th decomposition level.

The wavelets from the Daubechie, Symlet, and Coiflet families were examined. In light of these studies, no specific wavelet type was found to be particularly advantageous for the feature extraction process. However, several wavelets performed favorably in this regard. For this reason, further studies were limited to two wavelets, i.e., db10 and sym7.

A key challenge in the effectiveness of the diagnostics is the correct selection of features and their number, which determines the structure of the feature vector. A review of the literature on the subject indicates that there are many ways to optimize the process of feature selection. It is worth mentioning the main approaches for solving this problem. These include: thresholding methods, using wavelet entropy to optimize parameters of the wavelet function, selecting the proper coefficients using statistical criteria, employing wavelet packet coefficients as features, and using principal component analysis (PCA) to reduce the size of the feature space extracted from wavelet coefficients (BIAŁASIEWICZ, 2004; PENG, CHU, 2004; SYED, MURALIDHARAN, 2022; QIU *et al.*, 2006; DING *et al.*, 2023).

For the sake of completeness, it is worth noting that the above-mentioned method of feature extraction based on wavelet coefficient selection does not exhaust the possibilities offered by wavelet analysis of signals. Other methods mentioned in the literature, which are not the subject of this paper, can be classified as: wavelet energy-based, singularity-based, and wavelet function-based methods, etc. (PENG, CHU, 2004).

In the classical wavelet analysis method, the signal is decomposed into two lower-resolution representations of the signal: a detailed representation

and a coarse representation (signal approximation). The coarse representation can also be represented as the sum of the detailed and coarse representations from the previous resolution level. Thus, the detailed representation at a given resolution level does not change after each subsequent decomposition step, while the sequence of detailed representations, which characterize the analyzed signal, increases by one element (Fig. 6).

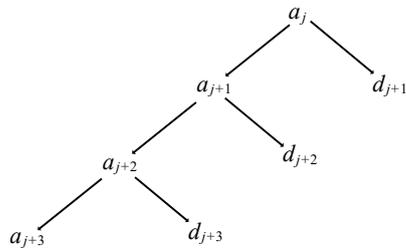


Fig. 6. Multiresolution signal decomposition scheme for three resolution levels.

The signal separation operation, which is a single level of wavelet decomposition from a signal of a given resolution level, is equivalent to its filtering by a set of digital quadrature low-pass and high-pass filters, and a subsequent downsampling operation (Fig. 7).

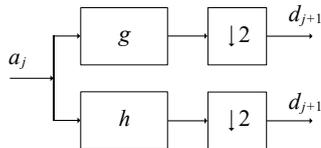


Fig. 7. Single level signal decomposition: *h* – low pass filter; *g* – high pass filter; ↓2 – decimation symbol.

The result of low-pass filtering is a sequence of samples that is the coarse approximation of the signal (a_i), while the result of high-pass filtering is a sequence of samples representing the details of the signal (d_i) at the immediately lower resolution level. Decimation, which involves removing every second sample from the resulting sample sequences at the output of the filters, prevents the introduction of redundant information into these sequences. The described iterative algorithm for determining the discrete wavelet transform (defined for discrete values of scale and shift parameters) is named, after its creator, the Mallat algorithm (MALLAT, 1989).

A decomposition of the signals up to level 8 resolution was performed. The components of the feature vector \mathbf{x} were computed from a sequence of wavelet coefficients of individual level decompositions. The individual components of the feature vector contain information about individual features of the signal, contained in a specific frequency range resulting from the division of the band into halves in subsequent stages of the decomposition of the analyzed signal (Fig. 8).

Considering the bandwidth of the measurement path for the acoustic signals, the decomposition components labeled D7 and D8 were excluded, as they fall outside the measurement bandwidth. Finally, the number of components in the feature vector, and thus the dimension of the feature space is $n_{wt} = 6$. The bandwidths of the decomposition components at the selected resolution levels, from which the individual components of the feature vectors were calculated, are shown in Table 1.

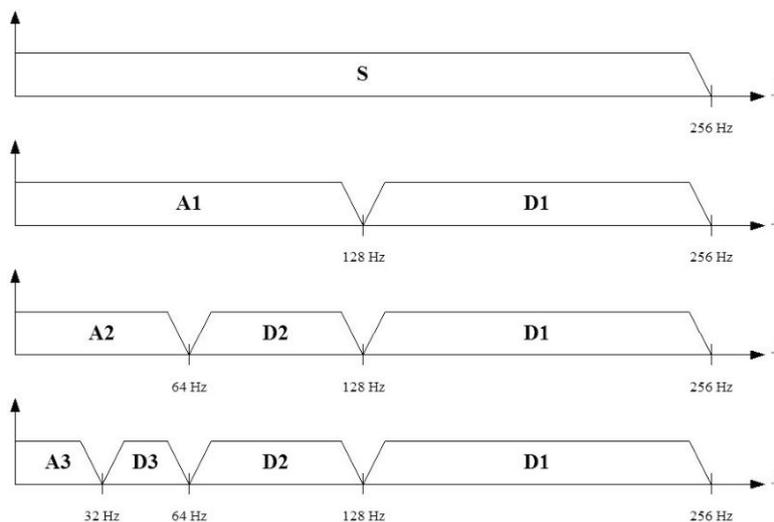


Fig. 8. Division of the signal band as a result of signal decomposition.

Table 1. Bandwidth of selected signal decomposition components.

Feature vector component	1	2	3	4	5	6
Component designation	D6	D5	D4	D3	D2	D1
Frequency range f [Hz]	4–8	8–16	16–32	32–64	64–128	128–256

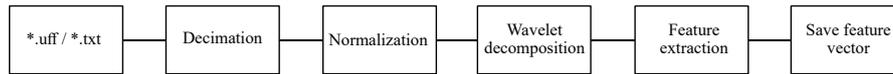


Fig. 9. Block diagram of the process of creating a feature vector.

The feature vector component value was defined as the root mean square (RMS) value of the reconstructed decomposition components based on the sequence of wavelet coefficients at a given decomposition level:

$$x_j = \sqrt{\frac{1}{N_k} \cdot \sum_{i=1}^{N_k} (d_k [i])^2}, \quad (2)$$

where j – the designation of the component of the feature vector, k – the decomposition level indicator, N_k – the number of wavelet coefficients for the k -th decomposition level, and $d_k [i]$ – the i -th wavelet coefficient of the k -th decomposition level.

Feature extraction is only one, but nevertheless an essential, component of the feature vector creation process. The entire process is shown in Fig. 9.

Files downloaded from the recorder (*.uff), containing samples of instantaneous values of the acoustic signal are processed into format suitable for MATLAB (*.txt). The sampling frequency during recording was $f_s = 65536$ Hz. This value was chosen during the preliminary research stage for problem recognition. Considering the found spectrum of the studied acoustic signals, this frequency can be significantly reduced, which allows for a decrease in the size of the registration files without losing the information contained in the spectrum. The recordings were resampled (decimated) to a sampling frequency of $f_{s_res} = 512$ Hz. After normalizing the signal energy, the recordings are subjected to wavelet decomposition as described above, and the feature vector values are calculated from the obtained sample sequences.

In classical wavelet analysis of signals, one should pay attention to high bandwidth of most decomposi-

tion components that may mask differences in the spectra of signals generated by objects in different technical conditions. Therefore, it was decided to investigate whether improving the frequency resolution of signal analysis by increasing the resolution would enhance the discriminative properties of the feature vectors. The decomposition of signals with the use of wavelet packets makes it possible to perform such tests.

Wavelet packets are a generalized method of signal decomposition using discrete wavelet transform. In this approach, the subsequent decomposition of the signal can be subjected to both a coarse representation and a detailed representation of the signal. This creates the possibility of analyzing different selected parts of the signal spectrum with higher resolution. A schematic of signal decomposition using wavelet packets, for example at three levels of resolution, is shown in Fig. 10.

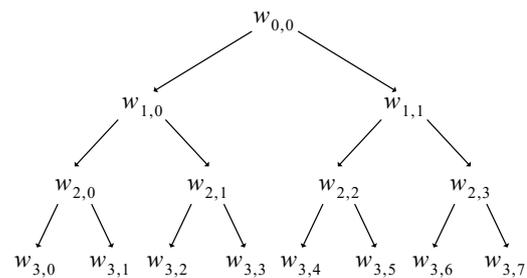


Fig. 10. Signal decomposition scheme using wavelet packets.

The division of the analyzed signal's frequency band corresponding to this decomposition scheme is shown in Fig. 11.

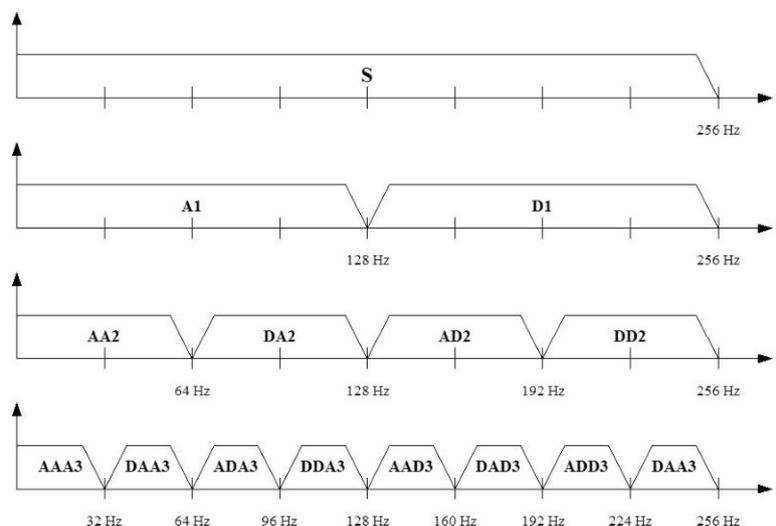


Fig. 11. Division of the signal band as a result of signal decomposition using wavelet packets.

Signal decompositions for the seventh resolution level were performed. The number of decomposition components is $2^7 = 128$ and the frequency bandwidth of individual decomposition components is $B_i = 2$ Hz. On the basis of the analysis of the registration spectra (Fig. 5), it should be expected that certain features of the acoustic signal, being a function of the dynamic properties of the seat mechanical system design, will manifest themselves in the frequency range of a few to several tens of Hz. Therefore, limiting the signal analysis to this frequency range allowed us to significantly reduce the number of decomposition components taken into account in further processing. This number was reduced to $n_{wp} = 26$.

The method for calculating the values of the individual components of the feature vector remained the same as previously described (Eq. (1)). Also, the signal processing operations, preceding the process of calculating the values of feature vector components, remained unchanged (Fig. 9).

Programming work, related to the implementation of the developed algorithms, was carried out using the MATLAB software platform. An example of the application's interface for signal decomposition using the classical wavelet analysis method is shown in Fig. 12.

The application was designed to support research related to the generation of a feature vector based on acoustic signal recordings produced during the operation of the seat drive device. It includes the last three stages of the feature vector generation process, shown in Fig. 9, i.e., wavelet decomposition, feature extraction, and save feature vector.

The application's input data are digital recordings of acoustic signals, provided as files containing samples of the signal's instantaneous values, after preprocessing (i.e., conversion to uff/txt format, decimation, normalization (Fig. 9)).

Pressing the “get and show” button expands the window to display a list of all registrations contained in the specified directory, indicated by the path. After selecting a registration, the time waveform and signal spectrum are displayed. Additional information, such as sampling frequency, number of registration samples, name of the downloaded file, parameters of the FFT algorithm is also displayed. The spectrum graph can be changed, to highlight important parts by changing some parameters of the FFT algorithm and activating these changes with the “change and show” button.

The main part of the algorithm, after selecting the type of wavelet, is initiated with the decomposition button. As a result, it displays the time waveforms of the reconstructed components for the assumed eight decomposition levels. They illustrate the energy of the signals representing each decomposition level. A graph and a table containing the RMS values of the wavelet coefficient sequences for the selected decomposition levels are also displayed. These values represent the components of the feature vector.

The result of the application is saved to disk in the directory specified by the path in the “folder for saving vector” after pressing the save vector button. Successful completion of this operation is indicated by the green color of the LED indicator.

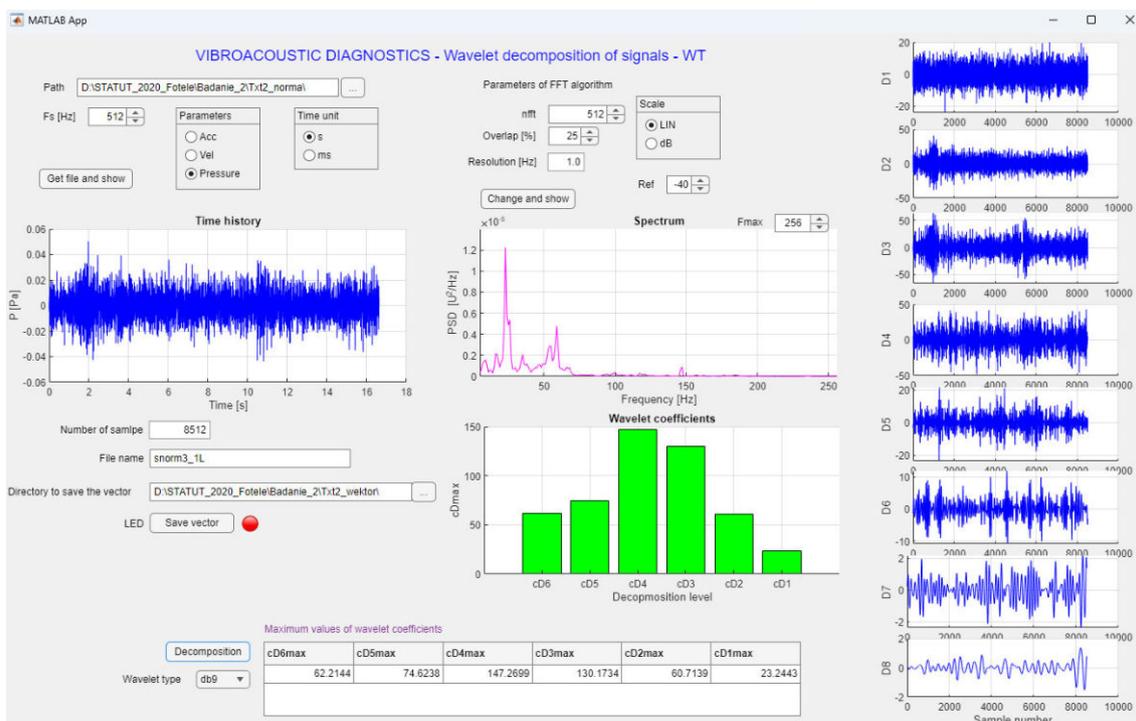


Fig. 12. Appearance of wavelet decomposition application interface.

4. Classification

The described algorithms were used to create sets of feature vectors, which serve as input data for the classification algorithms.

Each component of feature vector can be considered as a coordinate in a space called feature space. In this context, each diagnosed object corresponds to a point in that space called a picture, and a set of objects, belonging to one class corresponds to a certain area encompassing set of their pictures. From the point of view of diagnostics (the recognition process), a desirable situation occurs when, as a result of an appropriate choice of the structure of the feature space, images of objects of different categories occupy disconnected areas. In practice, these areas partially overlap (intermingle), causing diagnostic errors.

To test discriminative properties of the feature vectors we used the “Classification Learner” application available in MATLAB program. This application includes a set of classification algorithms allowing experimental selection of the optimal algorithm for a specific application. The classification results presented below were obtained for the K-nearest neighbor algorithm (for $K = 3$) and decision trees. It should be noted that the final choice of the classification algorithm should be made on a much larger dataset of diagnosed objects, and will most likely differ from the already indicated algorithms. Nevertheless, the obtained results suggest that the feature vectors generated using wavelet analysis of the signal exhibit discriminatory properties, enabling the effective diagnosis of the tested seats.

Due to the small number of tested objects, and the consequently small length of the learning sequence, the testing of particular algorithms included in the

mentioned application was performed according to the method known in the literature as leave-one-out (SOBCZAK, MALINA, 1985). From the set of feature vectors, one feature vector is selected and treated as the test vector. The rest of the set is treated as the learning set. This procedure is repeated for each vector in the set.

Figure 13 shows an example of a scatterplot in feature space for two selected components and the so-called confusion matrix for the learning sequence obtained by classical wavelet analysis.

What draws attention is the grouping of images (points) in the feature space corresponding to damaged objects (blue dots). For this learning sequence, two objects were incorrectly diagnosed, as indicated by the confusion matrix.

For feature vectors generated using wavelet packets, their discriminatory properties were tested with a feature space dimension of $n_{wp} = 26$, as already noted.

Figure 14 shows an example of the scatter of points in the feature space, for two selected components and the so-called confusion matrix, for the learning sequence, obtained by analysis using wavelet packets.

The feature space images corresponding to both good and bad objects occupy disjoint areas, although their close proximity may cause misdiagnosis. Qualitatively, the test results are somewhat better than the results of the testing a set of feature vectors obtained with the classical method. In this case, one object was misdiagnosed, as shown in the confusion matrix.

Due to the very small number of test objects and the consequently short length of the learning sequence, no far-reaching conclusions can be drawn from the above results, although they are promising.

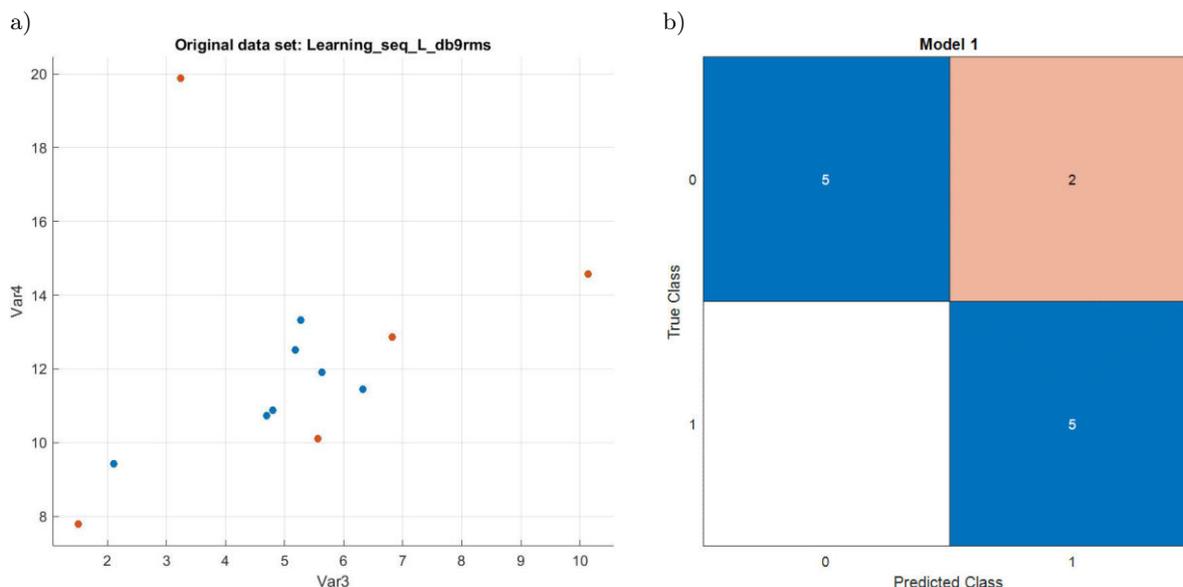


Fig. 13. Results of classical wavelet analysis: a) scatterplot of points in feature space for the two components of the learning sequence; b) confusion matrix.

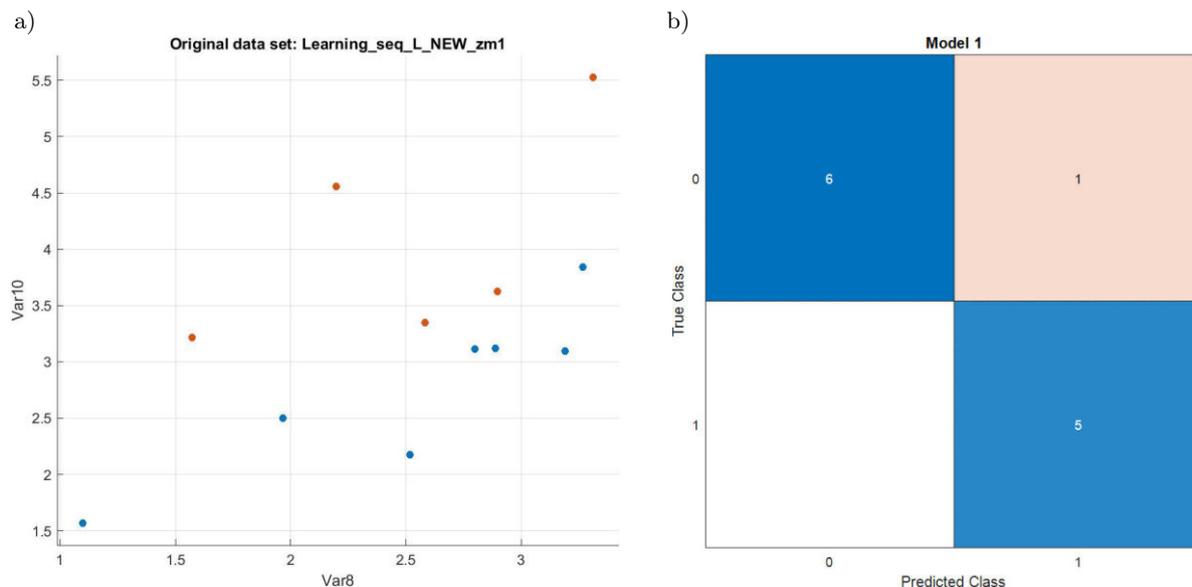


Fig. 14. Results of wavelet packed analysis: a) scatterplot of points in feature space for the two components of the learning sequence; b) confusion matrix.

The small number of research objects was caused by the Coronavirus pandemic and the resulting minimized possibilities of contacts with the seat manufacturer and the successive delayed acquisition of additional research objects. It is assumed that the possibility of accessing a large number of tested seats will soon be restored, allowing for the successive expansion of the measurement database, as was originally planned.

5. Conclusions

The obtained results indicate the potential of using wavelet analysis of acoustic signals generated during the operation of the diagnosed mechanical device in technical diagnostics. The feature vectors generated according to the presented algorithms have discriminative properties, allowing for diagnostics of the tested devices using machine learning methods.

As predicted, due to the increased frequency resolution of the diagnostic signal analysis, feature vectors generated using wavelet packets showed better discriminatory properties than those generated using classical wavelet analysis method.

Despite the fact that the obtained results are satisfactory, the effectiveness of the diagnostic method cannot be reliably assessed due to the small number of tested objects. This limitation was caused by objective difficulties.

The research is planned to be continued as originally intended. The machine learning method requires access to a large number of diagnosed objects and the creation, as a result of their research, of a database of diagnostic symptoms (features vectors). Based on these database resources, learning and testing sequences will be created. For optimal selection of

the classifier algorithm and to determine the effectiveness of the developed diagnostic method, it is necessary to use independent learning and testing sequences during testing.

Acknowledgments

This work was supported by the Ministry of Science and Higher Education of the Republic of Poland (Statutory activity no. 11131020-171).

References

1. BASZTURA C. (1996), *Acoustical Computer Diagnostic Systems* [in Polish: *Komputerowe Systemy Diagnostyki Akustycznej*], WNT, Warszawa.
2. BATKO W., DĄBROWSKI Z., ENGEL Z., KICIŃSKI J., WEYNA S. (2005), *Modern Methods of Vibroacoustics Processes Research* [in Polish: *Nowoczesne Metody Badania Procesów Wibroakustycznych*], Wydawnictwo Instytutu Technologii Eksploatacji – PIB, Radom.
3. BIAŁASIEWICZ J. (2004), *Wavelets and Approximations* [in Polish: *Falki i Aproksymacje*] WNT, Warszawa.
4. DING L., PENG J., ZHANG X., SONG L. (2023), Sleep snoring sound recognition based on wavelet packet transform, *Archives of Acoustics*, **48**(1): 3–12, doi: [10.24425/aoa.2022.142906](https://doi.org/10.24425/aoa.2022.142906).
5. DUDA R., HART P. (1973), *Pattern Classification and Scene Analysis*, John Wiley & Sons, Inc.
6. GŁOWACZ A. (2014), Diagnostics of synchronous motor based on analysis of acoustic signals with the use of line spectral frequencies and K-nearest neighbor classifier, *Archives of Acoustics*, **39**(2): 189–194, doi: [10.2478/aoa-2014-0022](https://doi.org/10.2478/aoa-2014-0022).

7. HAN X., XU J., SONG S., ZHOU J. (2022), Crack Fault diagnosis of vibration exciter rolling bearing based on genetic algorithm-optimized Morlet wavelet filter and empirical mode decomposition, *International Journal of Distributed Sensor Networks*, **18**(8), doi: [10.1177/15501329221114566](https://doi.org/10.1177/15501329221114566).
8. HUANG W.Y., SOLORZANO M.R. (1971), *Wavelet Pre-processing of Acoustic Signals*, Naval Ocean System Center.
9. International Organization for Standardization (2010), *Acoustics – Determination of sound power levels and sound energy levels of noise sources using sound pressure – Engineering methods for an essentially free field over a reflecting plane* (ISO Standard No. 3744:2010), <https://www.iso.org/standard/52055.html>.
10. LIN J. (2001), Feature extraction of machine sound using wavelet and its application in fault diagnosis, *NDT & E International*, **34**(1): 25–30, doi: [10.1016/S0963-8695\(00\)00025-6](https://doi.org/10.1016/S0963-8695(00)00025-6).
11. LIN J., QU L. (2000), Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis, *Journal of Sound and Vibration*, **234**(1): 135–148, doi: [10.1006/jsvi.2000.2864](https://doi.org/10.1006/jsvi.2000.2864).
12. MALLAT S.G. (1989), A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(7): 674–693, doi: [10.1109/34.192463](https://doi.org/10.1109/34.192463).
13. PAWLIK P. (2019), The use of the acoustic signal to diagnose machines operated under variable load, *Archives of Acoustics*, **45**(2): 263–270, doi: [10.24425/aoa.2020.133147](https://doi.org/10.24425/aoa.2020.133147).
14. PENG Z.K., CHU F.L. (2004), Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography, *Mechanical Systems and Signal Processing*, **18**(2): 199–221, doi: [10.1016/S0888-3270\(03\)00075-X](https://doi.org/10.1016/S0888-3270(03)00075-X).
15. PIERSOL J.S. (1989), *Random Data*, John Wiley & Sons.
16. QIU H., LEE J., LIN J., YU G. (2006), Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics, *Journal of Sound and Vibration*, **289**(4–5): 1066–1090, doi: [10.1016/j.jsv.2005.03.007](https://doi.org/10.1016/j.jsv.2005.03.007).
17. SOB CZAK W., MALINA W. (1985), *Information Selection and Reduction Methods* [in Polish: *Metody Selekcji i Redukcji Informacji*], WNT, Warszawa.
18. STASZEWSKI W.J. (1998), Wavelet based compression and feature selection for vibration analysis, *Journal of Sound and Vibration*, **211**(5): 735–760, doi: [10.1006/jsvi.1997.1380](https://doi.org/10.1006/jsvi.1997.1380).
19. SUN Q., CHEN P., ZHANG D., XI F. (2004), Pattern recognition for automatic machinery fault diagnosis, *Journal of Vibration and Acoustics*, **126**(2): 307–316, doi: [10.1115/1.1687391](https://doi.org/10.1115/1.1687391).
20. SYED S.H., MURALIDHARAN V. (2022), Feature extraction using discrete wavelet transform for fault classification of planetary gearbox – A comparative study, *Applied Acoustics*, **188**: 108572, doi: [10.1016/j.apacoust.2021.108572](https://doi.org/10.1016/j.apacoust.2021.108572).
21. SZABATIN J. (2000), *Basics of Signal Theory* [in Polish: *Podstawy Teorii Sygnałów*], WKŁ, Warszawa.
22. TANG B., LIU W., SONG T. (2010), Wind turbine fault diagnosis based on Morlet wavelet transformation and Wigner-Ville distribution, *Renewable Energy*, **35**(12): 2862–2866, doi: [10.1016/j.renene.2010.05.012](https://doi.org/10.1016/j.renene.2010.05.012).
23. XI F., SUN Q., KRISHNAPPA G. (1997), Bearing diagnostics based on pattern recognition of statistical parameters, [in:] *5th International Congress on Sound and Vibration*, Adelaide, South Australia.
24. YANG J., ZHOU C., LI X. (2022), Research on fault feature extraction method based on parameter optimized variational mode decomposition and robust independent component analysis, *Coatings*, **12**(3): 419, doi: [10.3390/coatings12030419](https://doi.org/10.3390/coatings12030419).

Research Paper

Sound Quality Prediction Method of Dual-Phase Hy-Vo Chain Transmission System Based on MFCC-CNN and Fuzzy Generation

Jiabao LI, Lichi AN, Yabing CHENG*, Haoxiang WANG

*School of Mechanical and Aerospace Engineering, Jilin University
Changchun, Jilin, China**Corresponding Author e-mail: chengyb@jlu.edu.cn*(received January 6, 2024; accepted July 1, 2024; published online October 21, 2024)*

The sound quality of transmission system noise significantly impacts user experience. This study aims to predict the sound quality of dual-phase Hy-Vo chain transmission system noise using a small sample size. Noise acquisition tests are conducted under various working conditions, followed by subjective evaluations using the equal interval direct one-dimensional method. Objective evaluations are performed using the Mel-frequency cepstral coefficient (MFCC). To understand the impact of the MFCC order and the frame number on prediction accuracy, MFCC feature maps of different specifications are analyzed. The dataset is expanded threefold using fuzzy generation with an appropriate membership degree. The convolutional neural network (CNN) is developed, utilizing MFCC feature maps as inputs and evaluation scores as outputs. Results indicate a positive correlation between the frame number and prediction accuracy, whereas higher MFCC orders introduce redundancy, reducing accuracy. The proposed CNN method outperforms three traditional machine learning approaches, demonstrating superior accuracy and resistance to overfitting.

Keywords: sound quality; dual-phase transmission; Hy-Vo chain; MFCC; fuzzy generation.



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

The silent chain transmission system is widely used in automobiles, motorcycles, and forklifts because of its low noise, high reliability, and high motion accuracy. As an advanced product of the silent chain, the Hy-Vo chain transmission system reduces the polygon effect because of the rocker pin. Based on the principle of bidirectional superposition, the dual-phase Hy-Vo chain transmission system can further reduce the polygon effect, vibration, and noise. In previous studies, researchers mainly focused on the design of the dual-phase Hy-Vo chain transmission system, with emphasis on the coupling effect between size parameters and the polygon effect (CHENG *et al.*, 2015; 2016a; 2016b; 2023). So far, the noise related research of the dual-phase Hy-Vo chain transmission system has not been involved.

A lot of studies have shown that noise can seriously harm people's mental and physical health (BASNER *et al.*, 2014; DRATVA *et al.*, 2012). Therefore, con-

sumers are also paying more attention to the use experience of low noise. Recently, there have been more and more researches on the sound quality in various fields (SONG, YANG, 2022; RUAN *et al.*, 2022; PARK *et al.*, 2020). In common sound quality prediction methods, acoustic parameters such as *A*-weighted sound pressure level (*A*-SPL), loudness, sharpness, roughness, fluctuation, and articulation index (AI) are used as inputs (WANG *et al.*, 2022; CHEN *et al.*, 2022). WANG *et al.* (2022) proposed a nonlinear sound quality modeling method that uses an extreme gradient boosting algorithm to predict the overall sound quality inside a pure electric car. CHEN *et al.* (2022) used the backpropagation neural network and support vector regression (SVR) to predict the sound quality of tractors, and used a genetic algorithm to optimize the parameters of the prediction models. To predict the sound quality using the convolutional neural network (CNN), the researchers introduced various feature maps as inputs (HUANG *et al.*, 2021; JIN *et al.*, 2021). HUANG *et al.* (2021) converted the objective parameter evaluation

into feature graphs and proposed a prediction method with an adaptive learning rate tree based on CNN. JIN *et al.* (2021) demonstrated that MFCC can distinguish noise of different sound qualities and used MFCC feature maps as inputs to predict the transmission sound quality. In the above studies on sound quality prediction, neural networks are widely used because of their strong ability to adjust to nonlinearity. However, when the number of samples is insufficient, the accuracy of a prediction model will be poor.

To predict the sound quality in the case of small samples, we have the following studies in this paper: firstly, we collected the noise of the dual-phase Hy-Vo chain transmission system under different working conditions. Random 5 s clips are taken from each noisy audio for subsequent processing. Based on the equal interval direct one-dimensional evaluation method, all noise samples are subjectively evaluated by the testers. Secondly, we calculate the MFCC for each sample. The standard MFCC only reflects the static characteristics of the noise, and the dynamic characteristics can be described by the difference of these static characteristics. To further study the influence of MFCC order and frame number on the prediction effect, we construct MFCC feature maps of different sizes as inputs of the prediction model. Thirdly, we propose a data enhancement method called fuzzy generation based on the fuzzy phenomenon in the subjective evaluation. By constructing the membership function of each noise sample, the appropriate membership degree is selected for sample generation. After the dataset is expanded, we build a CNN model for the sound quality prediction, and the prediction results show that the full-frame standard MFCC feature map has the best prediction effect when the membership degree is 0.9.

The more frames, the more complete the information contained in the MFCC, and the higher the prediction accuracy. However, higher order MFCC contains more redundant information, which will damage the prediction accuracy of the model. Finally, three common sound quality prediction methods are used in this paper, including the generalized regression neural network (GRNN), SVR, and ridge regression (RR). For each noise sample, we calculate six acoustic parameters (A -SPL, loudness, sharpness, roughness, fluctuation, and AI) as inputs. The comparative results show that the proposed new method has the lowest prediction error and strong resistance to overfitting. The flow chart of the sound quality research in this paper is shown in Fig. 1 and the structure of this paper is as follows: Sec. 2 involves the noise acquisition test and subjective evaluation of the noise sample. After the samples are preprocessed, we organize the testers to score the noise annoyance degree and test the correlation of the subjective evaluation results. In Sec. 3, the MFCC of all noise samples is calculated as an objective evaluation. After constructing the MFCC feature maps of different dimensions, the original dataset for the sound quality prediction is obtained by combining the subjective evaluation results. To train a more accurate prediction model, we use fuzzy generation to triple the size of the original dataset. In Sec. 4, we use MFCC feature maps of different specifications as input for the sound quality prediction and compare their prediction effects. After obtaining the optimal prediction model based on CNN, we compare it with the traditional sound quality prediction method. The results show that the prediction method proposed in this paper is more advantageous. Lastly, Sec. 5 presents the study's conclusion and summary.

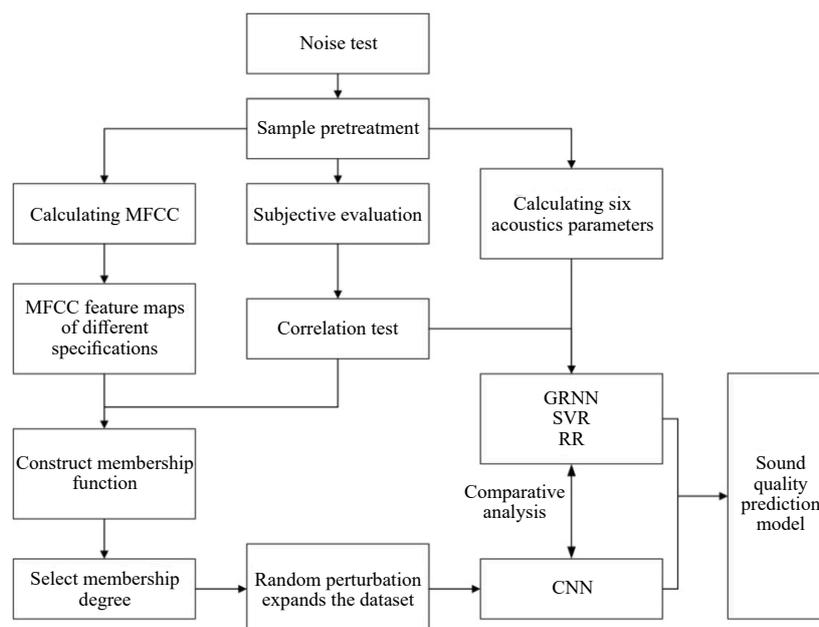


Fig. 1. Flow chart of sound quality prediction model construction.

2. Noise test and subjective evaluation

Different from single-phase transmission, the dual-phase sprocket teeth have phase difference. In our noise test, the drive sprocket tooth number is 35 with 5.14° phase difference, the driven sprocket tooth number is 37 with 4.86° phase difference, the pitch is 9.525 mm, the number of links is 84 and the chain form is 4×3 .

As shown in Fig. 2, the noise test is conducted in an indoor reverberation environment. We use measurement microphone (MINIDSP UMIK-1) to collect the noise and the measurement microphone is positioned at the same height as the center of the drive sprocket. There are two measurement points we selected, the first one is at the distance of 0.5 m from the center of the drive sprocket, the second one is at the distance of 1 m from that. The minimum speed of the test is 500 rpm and the maximum speed is 4000 rpm. The test loads are 500 N, 600 N, and 750 N. Starting from 500 rpm, noise data is collected under three loads for each 500 rpm increase. There are two collection points (0.5 m and 1 m from the center of the drive sprocket), eight speeds (500 rpm, 1000 rpm, 1500 rpm, 2000 rpm, 2500 rpm, 3000 rpm, 3500 rpm, 4000 rpm), and three loads (500 N, 600 N, 750 N), so $2 \times 8 \times 3 = 48$ original noise samples can be obtained. The sampling frequency is 48 000 Hz, and the noise data is recorded using Adobe Audition 2022 software. All noise acquisition times are longer than 30 s, we randomly intercept 5 s segment for subsequent data processing. Under the same working conditions, the time-domain waveform of the single-phase and dual-phase transmissions are shown in Fig. 3. The orange line on the left represents

the dual-phase transmission, and the blue line on the right represents the single-phase transmission.

As can be seen in Fig. 3, we can find that due to the principle of dual-phase superposition, the waveform of the dual-phase transmission is more uniform and denser at low speeds. When the speed is medium, the waveforms of the two transmissions are very similar. However, when running at high speed, the noise energy of the dual-phase transmission is obviously greater than that of the single-phase transmission. Therefore, the noise of the dual-phase transmission is different from that of other transmissions, and it is of great significance to study the sound quality prediction method of the dual-phase Hy-Vo chain transmission system.

After obtaining 48 noise samples, we organize twelve testers to conduct a subjective evaluation test. All of the testers are between 20 and 30 years old, and the ratio of men to women is 5:1. In addition, all testers have normal hearing and driving experience. As shown in Table 1, the subjective evaluation method is equal to the interval direct one-dimensional evaluation method (GUSKI, 1997). We rate the sound quality on a scale of discomfort, and there are five uncomfortable levels. A score of 0 is extremely uncomfortable level and a score of 10 is not uncomfortable level. Each of the remaining three uncomfortable levels has three scores, each score indicating the degree of discomfort in the same level. The subjective evaluation test is conducted in a quiet indoor environment, and the maximum SPL does not exceed 30 dB. The tester sits in a chair with headphones, and all the noise samples are played three times by Groove software. After listening, the tester gives the score and records it in a table.

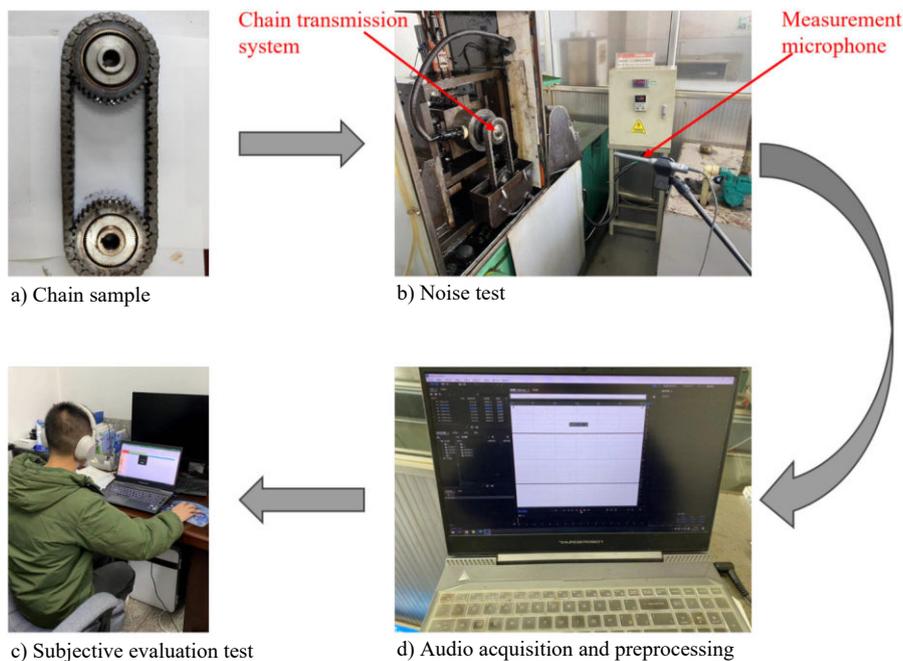


Fig. 2. Noise acquisition and data processing.

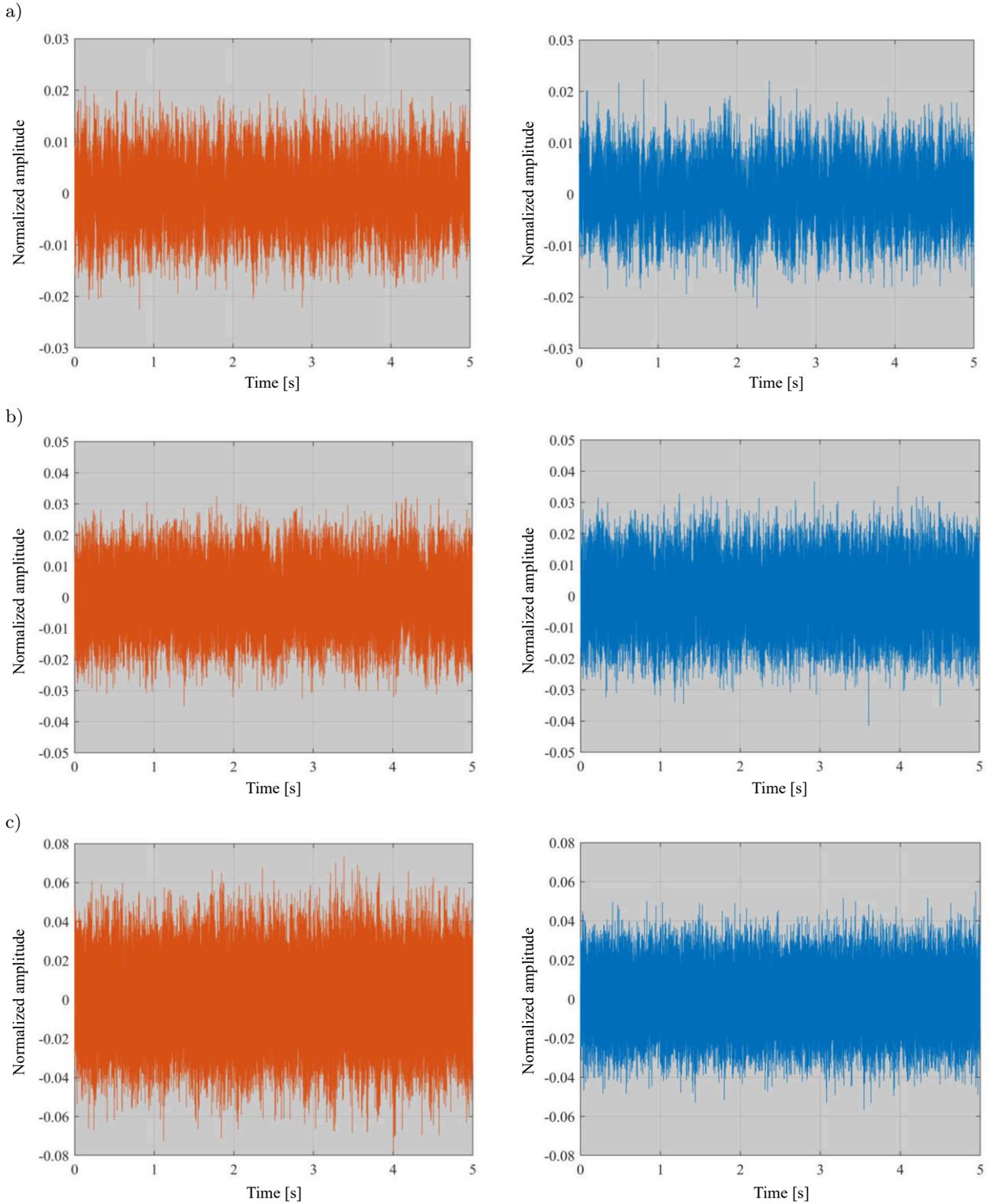


Fig. 3. Time-domain waveform comparison:
 a) 1000 rpm – 0.5 m – 1000 N; b) 2500 rpm – 0.5 m – 1000 N; c) 4000 rpm – 0.5 m – 1000 N.

Table 1. Subjective evaluation scoring table.

Uncomfortable level	Extremely uncomfortable	Very uncomfortable	Moderately uncomfortable	Little uncomfortable	Not uncomfortable
Scores	0	1–3	4–6	7–9	10

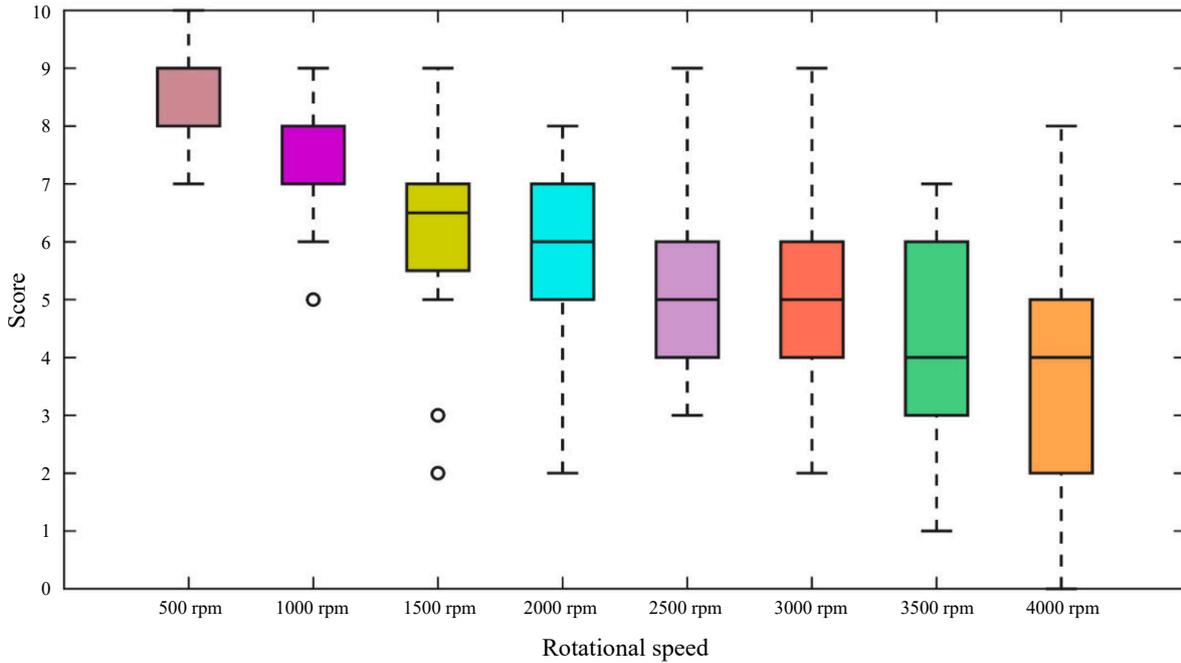


Fig. 4. Score boxplot for each speed.

All scores of each speed are presented in Fig. 4. Based on the horizontal line in the middle of the boxplot, in the range of 500 rpm–2500 rpm, it can be seen that the score decreases with the increase of the rotational speed. In this speed range, the sound quality of the chain transmission system becomes worse as the speed increases. The score of 3000 rpm remains unchanged compared to the score of 2500 rpm. However, the score continues to decline at 3500 rpm. As for the score of 4000 rpm, it is the same as the score of 3500 rpm. Therefore, in the case of medium and high speed, the sound quality of the chain transmission system shows a step-like decline trend. The length of box reflects the dispersion of scores. We can see that the scores are more dispersed at medium and high speeds, and there are even outliers at 1000 rpm and 1500 rpm. In the subjective evaluation test, we want all testers to have relatively consistent feelings about the same noise sample. The Spearman correlation analysis is performed on the scores of twelve testers and the results with poor correlation will be excluded. In the Spearman correlation analysis, the greater the coefficient R , the stronger the correlation. The equation of correlation coefficient (R) is:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (1)$$

where x_i and y_i represent the corresponding elements of the two variables, \bar{x} and \bar{y} represent the average value of the corresponding variables.

Based on Eq. (1), the R between the twelve testers are calculated, as illustrated in Fig. 5. The numbers from P1 to P12 represent the twelve testers, and it can be seen that P3–P6, P3–P11, and P6–P11 have a maximum correlation of 0.96. The correlation between P5–P6 and P6–P9 are both less than 0.7, indicating a weak correlation. According to Fig. 5, we calculate the average correlation coefficient (ACC) for each tester, as shown in Table 2.

In Table 2, all testers have an ACC of more than 0.7, indicating that the scores of each tester is reasonable. Generally speaking, the average score of the twelve testers is used as the final score of each noise sample, as shown in Table 3.

Table 2. ACC of each tester.

Tester	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
ACC	0.850	0.842	0.846	0.866	0.800	0.813	0.817	0.893	0.796	0.835	0.854	0.882

Table 3. Average score of each noise sample.

Sample	1	2	3	4	5	6	7	...	42	43	44	45	46	47	48
Score	8.75	8.58	8.33	9.08	8.75	8.67	7.08	...	5.25	3.00	2.33	1.92	5.50	4.67	4.17

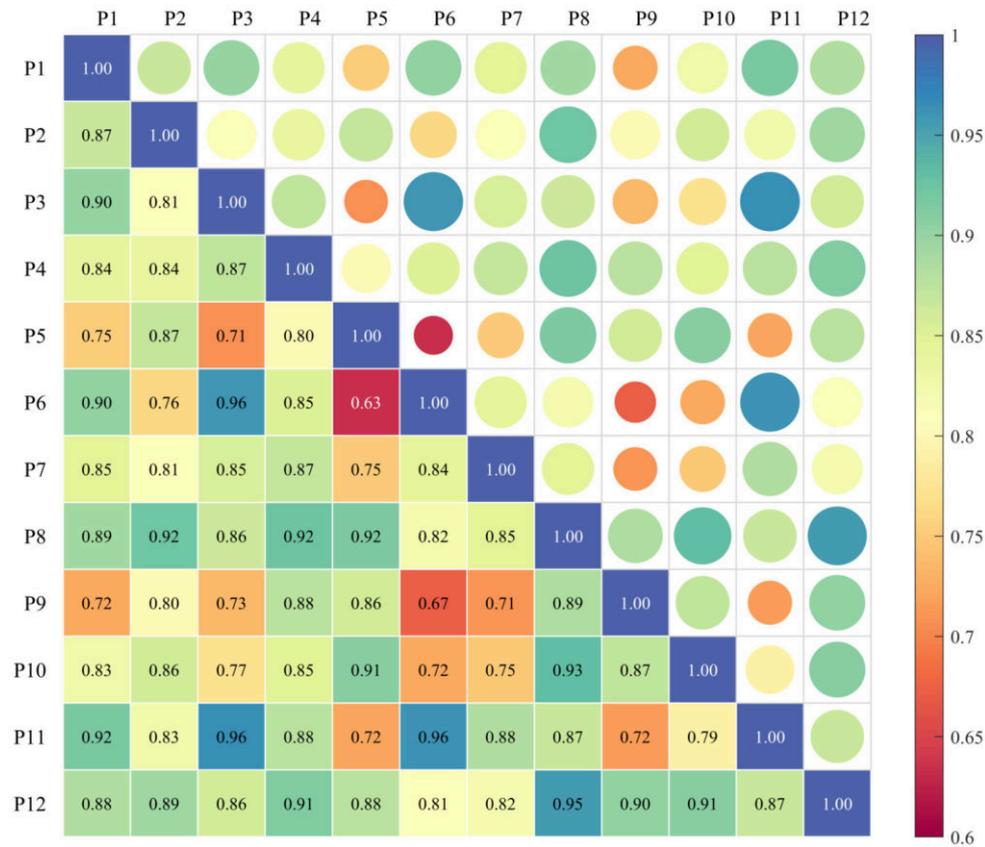


Fig. 5. Correlation heat map.

3. Objective evaluation and fuzzy generation

3.1. Construct Mel-frequency cepstral coefficient feature map

The MFCC is a feature extraction method commonly used in speech processing and audio analysis. It is based on the hearing characteristics of the human ear, by simulating the human ear's ability to perceive sounds of different frequencies, the sound signal is converted into a set of coefficients describing its characteristics. The advantage of MFCC is that they can effectively capture the main features of speech signals and have good adaptability for different speech processing tasks. However, they also have limitations, such as sensitivity to noise and possible degradation of performance in some complex environments. Therefore, in practical applications, MFCC is often used in combination with other types of feature and signal processing technologies (ABDUL, AL-TALABANI, 2022; MOONDRA, CHAHAL, 2023). The extraction process of MFCC is as follows:

- 1) Preprocessing: the sound signal is pre-weighted to increase the energy of the high frequency part:

$$y(t) = x(t) - \alpha x(t-1), \quad (2)$$

where $x(t)$ is the original signal, $y(t)$ is the pre-weighted signal, and α usually takes 0.95 or 0.97.

- 2) Framing: the segmentation of the sound signal into a series of short-time frames, each frame usually contains 20 ms–40 ms of data.
- 3) Windowing: the data of each frame is windowing processed, usually using hamming windows:

$$y(n) = x(n) \cdot \omega(n), \quad (3)$$

where $x(n)$ is the signal in a frame, $\omega(n)$ is the window function, and $y(n)$ is the signal after the window is added. The hamming window function is as follows:

$$\omega(n) = (1 - a) - a \cdot \cos(2\pi n/N) \quad 1 < n < N, \quad (4)$$

where N is the number of sampling points, and different values of a will produce different hamming windows, in general, $a = 0.46$.

- 4) The Fourier transform: a fast Fourier transform (FFT) is performed on each frame of data to convert it into a signal in the frequency domain:

$$Y(k) = \sum_0^{N-1} y(n) \cdot e^{-j \frac{2\pi}{N} kn}, \quad (5)$$

where $Y(k)$ is the k -th component in the frequency domain, and N is the number of FFT points.

5) Mel filtering: the frequency domain signal is passed through a set of Mel filter banks to simulate the human ear’s perception of different frequencies. Compared with the normal frequency mechanism, the Mel value is closer to the hearing mechanism of the human ear. It grows fast in the low frequency range, but it grows slowly in the high frequency range. Each frequency value corresponds to a Mel value, and the corresponding relationship is as follows:

$$m = 2595 \cdot \log_{10} \left(1 + \frac{f}{700} \right). \quad (6)$$

If we want to convert the Mel-frequency m to the frequency f , we can get it by sorting the above Eq. (6):

$$f = 700 \cdot (10^{m/2595} - 1). \quad (7)$$

The response $H_m(k)$ of each filter is usually defined as a triangular filter that is uniformly distributed on the Mel scale, and the output $S(m)$ is the signal energy that passes through the filter:

$$k = \frac{(1+N) \cdot f_m}{f_s}, \quad (8)$$

$$H_m(k) = \begin{cases} 0 & k < f(m-1), \\ \frac{2(k-f(m-1))}{a^*} & f(m-1) \leq k \leq f(m), \\ \frac{2(f(m+1)-k)}{a^*} & f(m) \leq k \leq f(m+1), \\ 0 & k \geq f(m+1), \end{cases} \quad (9)$$

$$S(m) = \sum_{k=0}^{K-1} |Y(k)|^2 \cdot H_m(k), \quad (10)$$

where

$$a^* = (f(m+1) - f(m-1))(f(m) - f(m-1)).$$

6) Log the output of the Mel filter bank to obtain the logarithmic energy spectrum:

$$L(m) = \log(S(m)), \quad (11)$$

where $L(m)$ is the logarithmic energy spectrum.

7) Discrete cosine transforms: perform a discrete cosine transform (DCT) on the logarithmic energy spectrum to obtain the MFCC coefficient:

$$C(n) = \sum_{m=0}^{M-1} L(m) \cdot \cos \left[\frac{\pi}{M} (m + 0.5)n \right], \quad n = 1, 2, \dots, L, \quad (12)$$

where $C(n)$ is the n -th cepstral coefficient, M is the number of Mel filters, and L refers to the MFCC coefficient order, usually 12–16.

From Eq. (3) to Eq. (12), we can get the standard MFCC, which only reflects the static properties of audio. The dynamic characteristics of audio can be described by the difference of these static characteristics, as follows:

$$\Delta C_t = \frac{\sum_{n=1}^N n(C_{t+n} - C_{t-n})}{2 \sum_{n=1}^N n^2}, \quad (13)$$

$$\Delta \Delta C_t = \frac{\sum_{n=1}^N n(\Delta C_{t+n} - \Delta C_{t-n})}{2 \sum_{n=1}^N n^2}. \quad (14)$$

Equations (13) and (14) represent the first- and second-order difference, respectively. In this paper, we take each frame as 32 ms, the noise sample is divided into K frames and the MFCC of L order is calculated. As shown in Fig. 6, we can obtain $K \times L$ feature maps of different orders. The standard full-frame MFCC feature map is 311×13 , 311×26 with only first-order differences, and 311×39 also with

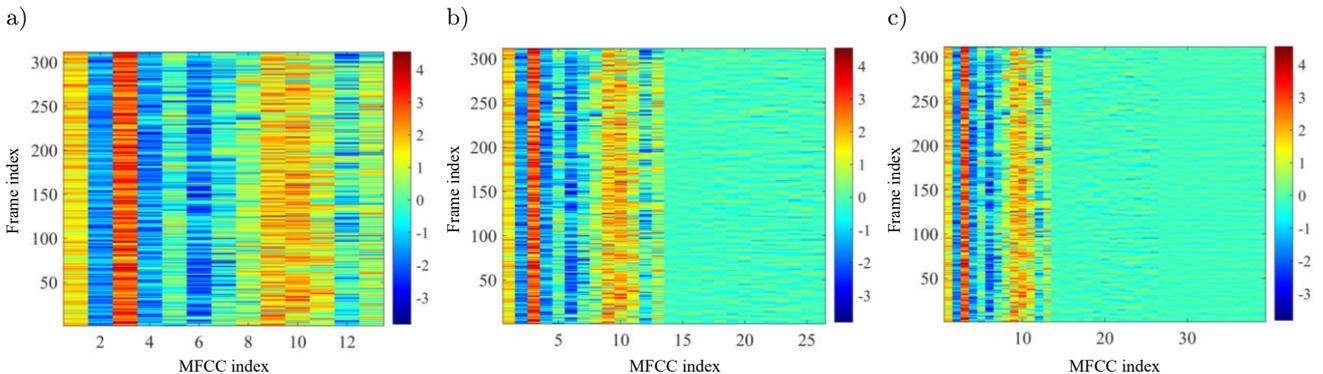


Fig. 6. MFCC feature map: a) 311×13 ; b) 311×26 ; c) 311×39 .

second-order differences. We also get the feature maps of two frame numbers, 208×13 and 104×13 , respectively. Finally, we can get 5 input features of different sizes.

3.2. Fuzzy generation

Fuzzy mathematics is a mathematical method to deal with uncertain information. Compared with traditional binary logic and precise mathematics, it pays more attention to the description and processing of fuzzy and uncertain phenomena in the real world. The core concept of fuzzy mathematics is a fuzzy set. Unlike traditional sets, where the elements either belong to or do not belong to the set, the degree to which an element in a fuzzy set belongs to the set is a numerical value between 0 and 1, called membership. This makes fuzzy sets more flexible in describing uncertainty and ambiguity in the real world. Membership functions are used to describe the degree to which an element belongs to a fuzzy set. The value of this function is between 0 and 1. The core strength of fuzzy mathematics is that it provides an effective way to deal with the uncertainty and ambiguity that are prevalent in the real world. By introducing fuzzy concepts, it allows for the more flexible and realistic problem solving and decision-making process (RUAN, LI, 2021; GÜNDOĞDU, KAHRAMAN, 2019; BUSTINCE *et al.*, 2016).

In the previous subjective evaluation, there is a fuzzy problem. Generally speaking, for the same noise sample, researchers only calculate the average score as the final subjective evaluation score. In fact, the scores of all testers are reasonable after the correlation test. Therefore, we believe that in the range of minimum and maximum scores, the average score as a label value is when the membership degree is 1, and the fuzzy mapping is constructed as follows:

$$\begin{aligned} F: V &\rightarrow [0, 1], \\ m &\mapsto F(m), \end{aligned} \quad (15)$$

where V is value field $[0, 10]$, F is the fuzzy interval of V , and $F(m)$ is the membership function.

For each noise sample, we can construct its fuzzy interval and membership function. In Table 4, the average score is the core of the fuzzy interval, the minimum score is the left boundary (LB), and the maximum score is the right boundary (RB). We construct the membership function on the fuzzy interval and select the appropriate membership degree to delimit the sample generation interval. Then the label value is randomly perturbed over the sample generation interval to expand the dataset. The membership function is defined as follows:

$$\frac{F(m_d) - 0}{d - r} = \frac{1 - 0}{k - r} \Rightarrow F(m_d) = \frac{1}{r - k}(r - d), \quad (16)$$

Table 4. Fuzzy intervals.

Sample	LB	Core	RB	Sample	LB	Core	RB
1	8	8.75	10	25	4	5.17	6
2	8	8.58	9	26	3	4.33	6
3	7	8.33	9	27	3	4.17	6
4	8	9.08	10	28	5	6.83	9
5	8	8.75	10	29	4	5.83	8
6	7	8.67	10	30	4	5.75	8
7	6	7.08	8	31	3	4.33	6
8	5	6.92	8	32	2	3.75	5
9	6	7.00	8	33	2	3.83	6
10	7	8.33	9	34	4	6.33	9
11	6	8.00	9	35	3	5.83	8
12	6	7.58	9	36	3	5.50	8
13	3	5.58	8	37	2	3.67	5
14	3	5.25	7	38	1	3.25	5
15	2	5.25	7	39	1	3.25	6
16	7	7.75	9	40	3	5.50	7
17	6	7.17	8	41	3	5.25	7
18	6	6.83	8	42	3	5.25	7
19	4	5.50	7	43	1	3.00	5
20	3	5.08	7	44	0	2.33	5
21	2	4.83	7	45	0	1.92	5
22	6	7.25	8	46	3	5.50	8
23	6	6.83	8	47	2	4.67	8
24	5	6.50	7	48	1	4.17	7

$$\frac{F(m_d) - 0}{d - l} = \frac{1 - 0}{k - l} \Rightarrow F(m_d) = \frac{1}{k - l}(d - l), \quad (17)$$

$$F(m_d) = \begin{cases} 0 & (0 \leq d < l), \\ \frac{1}{k - l}(d - l) & (l \leq d < k), \\ \frac{1}{r - k}(r - d) & (k \leq d \leq r), \\ 0 & (r < d \leq 10), \end{cases} \quad (18)$$

where k is the core point, l is the LB point, r is the RB point, d is a random generation point, and $F(m_d)$ is the membership of d .

As can be seen in Fig. 7, the farther away from the core point, the smaller the membership degree. For different samples, the span of their membership function is usually different. Under the same membership degree, the larger the span, the larger the sample generation interval. In the generation interval, the sample label values are randomly perturbed to expand the dataset. However, the larger the interval, the more noise the new sample points contain. In this paper, we choose four membership degrees of 0.3, 0.5, 0.7, and 0.9 for fuzzy generation. The dataset is expanded to three times its original size, including 144 samples.

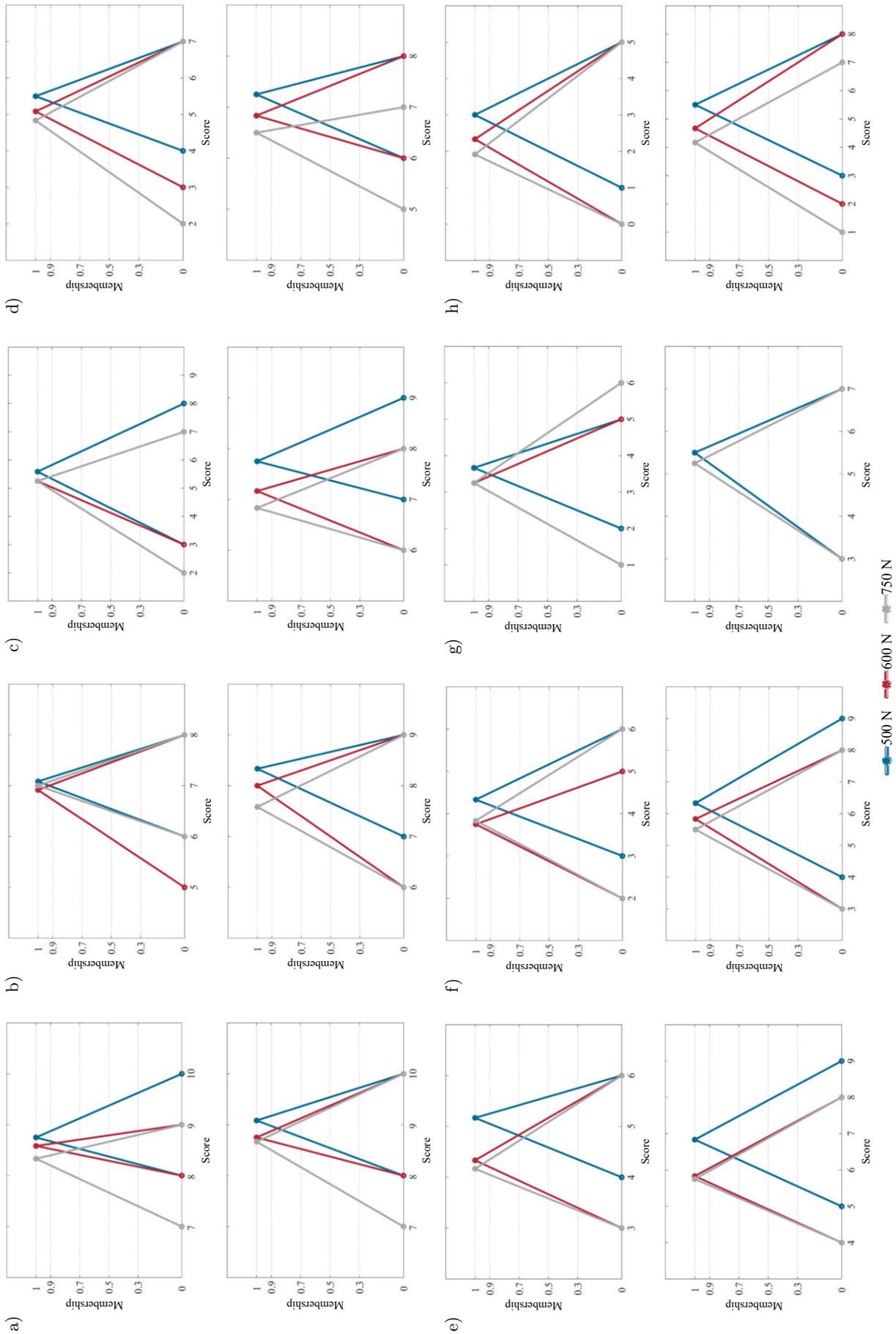


Fig. 7. All membership functions: a) 500 rpm; b) 1000 rpm; c) 1500 rpm; d) 2000 rpm; e) 2500 rpm; f) 3000 rpm; g) 3500 rpm; h) 4000 rpm. The upper graph in each subgraph represents the membership function for the samples of 0.5 m measurement point, and the lower graph in each subgraph represents the membership function for the samples of 1 m measurement point.

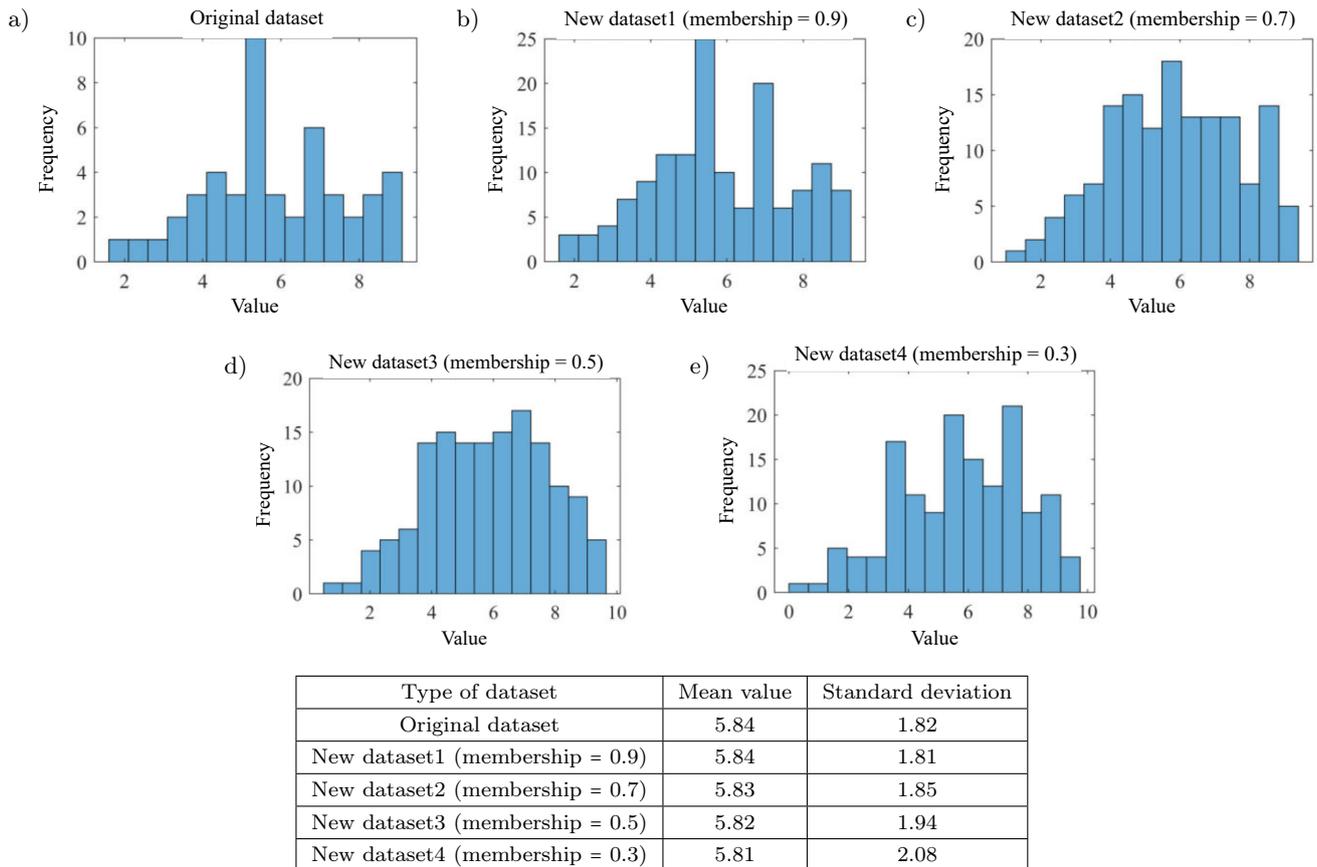


Fig. 8. Histogram comparison of original dataset and new datasets.

Figure 8 shows the histogram comparison between the original dataset and the expanded new dataset. The histogram of the original dataset shows a relatively symmetric unimodal distribution with a mean of 5.84 and a standard deviation of 1.82. The new dataset1 is very similar in shape to the original dataset, with the mean remaining at 5.84 and the standard deviation slightly reduced to 1.81. The new dataset2 has a slightly changed distribution shape, with a mean of 5.83 and a standard deviation of 1.85, slightly increasing the variability. The distribution shape of the new dataset3 has a significant change, with the mean of 5.82 and the standard deviation increasing to 1.94, indicating a further increase in variability. The new dataset4 has the most significant change in distribution shape, with a mean of 5.81 and a standard deviation of 2.08, indicating the greatest variability. As the membership value decreases, the standard deviation of the new dataset gradually increases, indicating that the perturbation introduces more variability. The mean remains essentially unchanged, indicating that the new dataset is still centered around the mean of the original data. By analyzing Fig. 8, we can conclude that higher membership values (such as 0.9 and 0.7) retain the main features of the original dataset and increase the number of datasets while maintaining low variability. Lower mem-

bership values (such as 0.5 and 0.3) introduce more variability and outliers, and may introduce more noise despite increasing the diversity of the dataset.

4. Modeling and prediction

4.1. Convolutional neural network

The CNN is a kind of deep learning model that has achieved great success in image recognition, video analysis, natural language processing and other fields. CNN is particularly suited for working with data with a grid structure, such as images and time series data. The core idea of CNN is to use convolutional layers to automatically learn features of spatial hierarchy from data. These features are gradually abstracted and combined through multiple convolution layers and subsampling layers (usually pooling layers) to accomplish complex tasks. The advantage of CNN is its ability to automatically learn and extract features without the need for manual feature engineering (BHATT *et al.*, 2021; GUMIRI *et al.*, 2023; MANDOUH *et al.*, 2023). CNN is generally used to solve the classification problem. To predict the sound quality, we set the output layer to have only one node, and do not use nonlinear activation function, so that the output is a linear

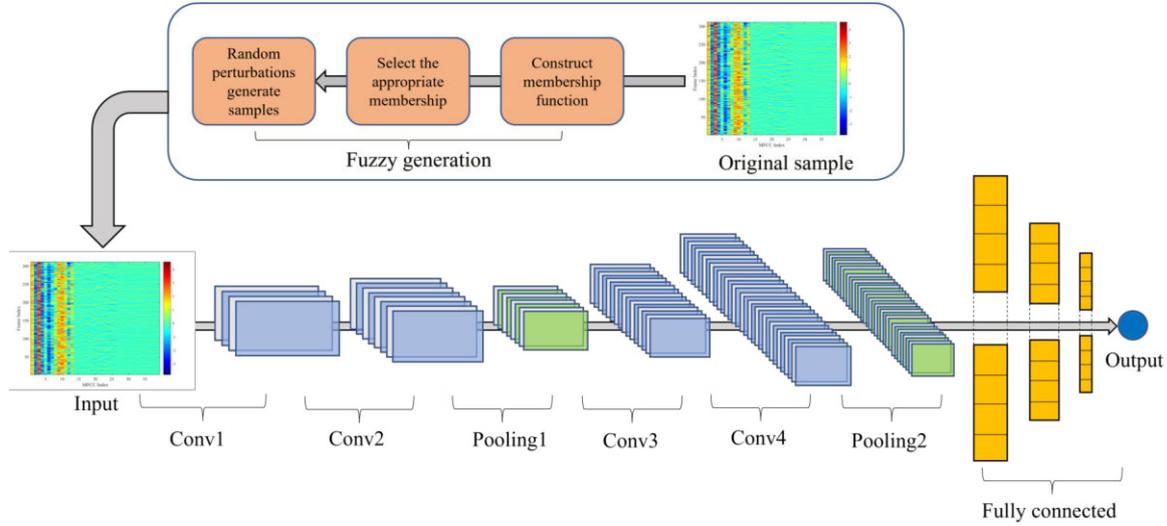


Fig. 9. Model structure.

transformation of the inputs, which can get a continuous value.

As shown in Fig. 9, the model structure consists of four convolution layers, two pooling layers, a flatten layer and three fully connected layers. In addition to the output layer, the activation function of the other layers is a rectified linear unit (ReLU). The pooling layer adopts the maximum pooling, the step size is 2 with 0 padding. The convolution layer has a step size of 1 without 0 padding. The numbers of neurons in the three fully connected layers are 1024, 128, and 1, respectively. Using dropout technology in the first fully connected layer, proceeds with the dropout rate set to 0.5. The last layer is the output layer, which outputs the evaluation score. Taking the input feature map $MFCC_{311 \times 13}$ as an example, the model structure parameters are shown in Table 5.

Table 5. Structural parameters.

Layer type	Channels/Units
Input 311×13	3
3×3 Conv1 ReLU, stride 1	6
3×3 Conv2 ReLU, stride 1	12
2×2 Maxpooling1 ReLU, stride 2	12
3×3 Conv3 ReLU, stride 1	24
3×3 Conv4 ReLU, stride 1	48
2×2 Maxpooling2 ReLU, stride 2	48
Flatten	3600
Fully connected (1)	1024
Dropout	1024
Fully connected (2)	128
Fully connected (3)	1

The MFCC feature map is taken as input, the evaluation score is taken as output, and the ratio of training set to test set is 5:1. Using the Adam optimizer, the initial learning rate is 0.001, a root mean squared

error (RMSE) is the loss function, and the epoch is set to 200. With 5 input feature maps and 4 membership degrees, the average of 5 training results is taken, and the model is trained $5 \times 4 \times 5 = 100$ times in total. In model training, we choose the R , the RMSE, and the mean absolute error (MAE) as evaluation indexes, and the calculation formula is as follows:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (19)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2}, \quad (20)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - y_i|, \quad (21)$$

where n is the number of samples, x_i is the predicted value of the sample, and y_i is the true value of the sample.

The prediction effects of MFCC feature maps with different orders are shown in Table 6, and Δ represents the increment compared to the results in the first row. Based on the training results of standard full-frame $MFCC_{311 \times 13}$ feature map, we can see that with the increase of the MFCC order, the three evaluation indexes are deteriorating. The high order MFCC contains too much useless information and damages the performance of the model. In addition, the influence of the MFCC frame number on the training results is shown in Table 7.

In Table 7, all three evaluation indexes get worse as the number of frames decreases. Compared with $MFCC_{311 \times 13}$, the frame number of $MFCC_{208 \times 13}$ decreases by 33.3 %, but RMSE and MAE increase by

Table 6. Prediction effects of different MFCC orders.

Feature maps	Training			Prediction		
	R (Δ)	RMSE (Δ)	MAE (Δ)	R (Δ)	RMSE (Δ)	MAE (Δ)
MFCC _{311×13}	0.979	0.394	0.314	0.971	0.474	0.371
MFCC _{311×26}	0.966 (-0.013)	0.505 (+0.111)	0.392 (+0.078)	0.953 (-0.018)	0.603 (+0.129)	0.467 (+0.096)
MFCC _{311×39}	0.923 (-0.056)	0.670 (+0.276)	0.539 (+0.225)	0.902 (-0.069)	0.785 (+0.311)	0.617 (+0.246)

Table 7. Prediction effects of different MFCC frame number.

Feature maps	Training			Prediction		
	R (Δ)	RMSE (Δ)	MAE (Δ)	R (Δ)	RMSE (Δ)	MAE (Δ)
MFCC _{311×13}	0.979	0.394	0.314	0.971	0.474	0.371
MFCC _{208×13}	0.977 (-0.002)	0.567 (+0.173)	0.481 (+0.167)	0.970 (-0.001)	0.663 (+0.189)	0.558 (+0.187)
MFCC _{104×13}	0.965 (-0.014)	1.248 (+0.854)	1.143 (+0.829)	0.958 (-0.013)	1.326 (+0.852)	1.211 (+0.840)

Table 8. Prediction effects of different membership degrees.

Membership degree	Training			Prediction		
	R (Δ)	RMSE (Δ)	MAE (Δ)	R (Δ)	RMSE (Δ)	MAE (Δ)
0.9	0.993	0.256	0.211	0.991	0.279	0.224
0.7	0.990 (-0.003)	0.319 (+0.063)	0.260 (+0.049)	0.985 (-0.006)	0.379 (+0.100)	0.305 (+0.081)
0.5	0.971 (-0.022)	0.447 (+0.191)	0.356 (+0.145)	0.966 (-0.025)	0.490 (+0.211)	0.400 (+0.176)
0.3	0.964 (-0.029)	0.552 (+0.296)	0.427 (+0.216)	0.941 (-0.050)	0.748 (+0.469)	0.554 (+0.330)

39.9 % and 50.4 %, respectively. For MFCC_{104×13}, the frame number continues to decline by 33.3 %, while RMSE and MAE increase sharply by 179.7 % and 226.4 %. Therefore, the prediction error is more sensitive to the frame number. Too few frames will lead to missing key information, and the accuracy of the model will be seriously degraded.

When the standard full-frame MFCC_{311×13} feature map is used as input, different membership degrees also affect the prediction results. As can be seen from Table 8, the prediction is best when the membership degree is 0.9. The membership degree gradually decreases, and the three evaluation indexes gradually deteriorate.

Based on the above comparative experiments, we can know that the model prediction is best when

the frame number is 311, the MFCC order is 13 and the membership degree is 0.9. Figure 10 shows the convergence curve under optimal conditions. In the first 21 iterations, the loss of the model decreases rapidly, but there are some fluctuations. In subsequent iterations, the model slowly converges. Finally, the RMSE of the training set is 0.142 and the RMSE of the test set is 0.153. The error is small enough to meet the scoring requirements of subjective evaluation, and the final prediction results are shown in Table 9.

Table 9. Final CNN prediction results.

Indexes	Training			Prediction		
	R	RMSE	MAE	R	RMSE	MAE
Results	0.997	0.142	0.110	0.996	0.153	0.127

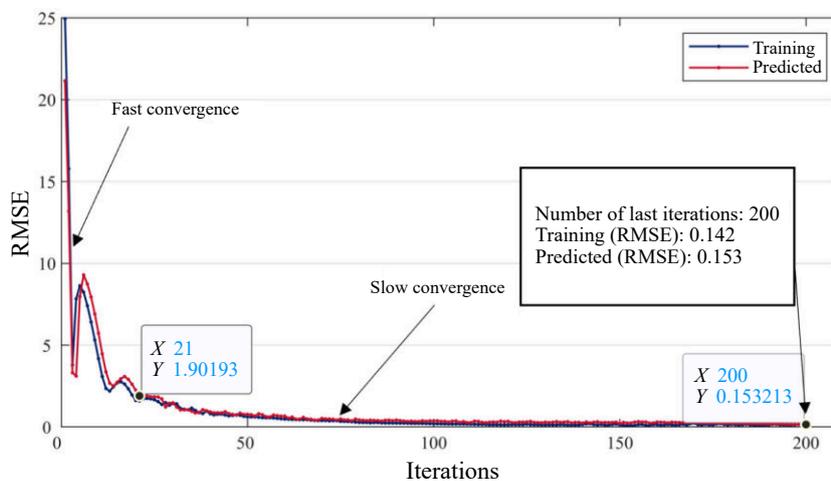


Fig. 10. Convergence curve.

To further verify the generalization ability of the proposed method in small samples, we used a five-fold cross-validation in the experiment. Five-fold cross-validation divides the dataset into five subsets, using one of the subsets as the validation set and the remaining four subsets as the training set, repeating five times to ensure that each subset is used as a single validation set. The final model performance is averaged by the results of five experiments. As shown in Table 10, training the model with the new dataset can significantly improve the model’s predictive performance and decreases the MAE value. This shows that the fuzzy generation method is effective under the condition of small samples and can enhance the generalization ability of the model. When the membership value is large (such as 0.9 and 0.7), the MAE value of the model is significantly reduced. At the same time, a low standard deviation is maintained, indicating that this degree of disturbance can effectively increase the data diversity without introducing too much noise. When membership values are small (such as 0.5 and 0.3), more noise is introduced into the dataset. Although the model performance is also improved, the effect is not as good as when the membership value is larger.

Table 10. Five-fold cross-validation results.

Type of the dataset	MAE	Standard deviation
Original dataset	3.241	1.034
New dataset1 (membership = 0.9)	0.736	0.121
New dataset2 (membership = 0.7)	0.885	0.178
New dataset3 (membership = 0.5)	1.078	0.141
New dataset4 (membership = 0.3)	1.448	0.149

4.2. Comparative analysis

To compare with traditional sound quality prediction methods, generalized regression neural network, SVR and RR models are used in this paper. We first use the Audio toolbox in MATLAB to calculate six acoustic parameters (A -SPL, loudness, sharpness, roughness, fluctuation, and AI) for all noise samples, as shown in Fig. 11. We take the six acoustic parameters as inputs, the evaluation scores as outputs, and the ratio of training set to test the set is also 5:1.

A generalized regression neural network (GRNN) is a type of neural network based on a radial basis func-

tion, mainly used to solve regression problems. The structure of GRNN is relatively simple, including input layer, pattern layer, summation layer and output layer. GRNN has applications in many fields, especially for scenarios that require fast and accurate regression predictions (ZHU *et al.*, 2022; YAO *et al.*, 2023). In the GRNN model, only one spread parameter σ needs to be optimized. By using the particle swarm optimization algorithm, the number of particles is 30, the maximum number of iterations is 20, and the optimal parameter $\sigma = 0.12$ is found on the interval [0.01 0.8].

The SVR is a regression method based on the principles of support vector machines. The core idea of SVR is to find a function that fits the training data as best as possible within a limited error range while maintaining the generalization ability of the model. For nonlinear data, the SVR uses kernel functions to map the data into a high-dimensional space, where linear regression is performed. Common kernel functions include linear kernel, polynomial kernel, radial basis function kernel, and so on (ZHAN *et al.*, 2022; SHI *et al.*, 2021). For the SVR model with radial basis function, we also use the particle swarm optimization algorithm to find the two optimal parameters (penalty parameter c and kernel parameter g). The number of particles is 30, the maximum number of iterations is 20, and the best $c = 34.83$, $g = 0.32$ are found on the interval [0.01 100].

The RR, also known as the Tikhonov regularization, is a linear regression method for dealing with multicollinearity problems. Multicollinearity refers to the fact that there is a high degree of correlation between predictor variables in a regression analysis. The RR solves this problem by introducing a regularization term, thereby improving the stability and predictive power of the model. The basic idea of RR is to add a regularization term to the loss function of ordinary least squares regression. Choosing proper regularization parameter λ is the key to applying RR (YASIN *et al.*, 2022; DAR *et al.*, 2023). In this paper, for the RR model, the 5-fold cross validation is used to find the optimal λ . The value range is $[10^{-6}, 10^{-5.76}, \dots, 10^6]$, and the best $\lambda = 1.33$ is found when the mean square error is minimum.

Table 11 shows the three evaluation indexes of three traditional sound quality prediction methods. Compared with the CNN model, we can see that the GRNN model performs slightly better in training than

Table 11. Comparison of prediction effect on different models.

Model	Training			Prediction		
	R (Δ)	RMSE (Δ)	MAE (Δ)	R (Δ)	RMSE (Δ)	MAE (Δ)
CNN	0.997	0.142	0.110	0.996	0.153	0.127
GRNN	0.998 (+0.001)	0.105 (-0.037)	0.055 (-0.055)	0.988 (-0.008)	0.239 (+0.086)	0.210 (+0.083)
SVR	0.991 (-0.006)	0.243 (+0.101)	0.288 (+0.178)	0.964 (-0.032)	0.407 (+0.254)	0.288 (+0.161)
RR	0.977 (-0.020)	0.360 (+0.218)	0.330 (+0.220)	0.966 (-0.030)	0.389 (+0.236)	0.336 (+0.209)

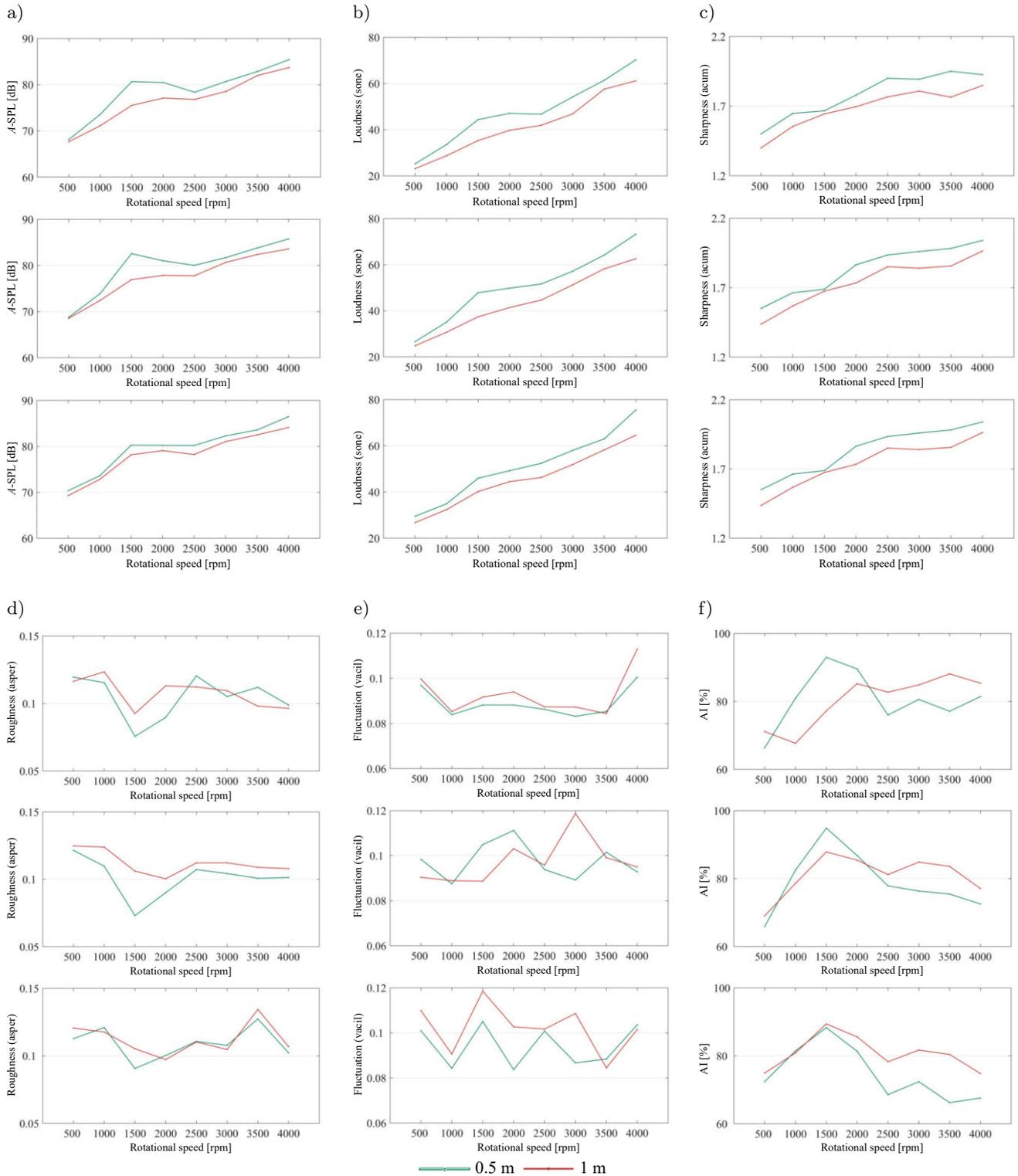


Fig. 11. Acoustic parameters: a) *A*-SPL; b) loudness; c) sharpness; d) roughness; e) fluctuation; f) AI. The top picture of each subgraph represents the parameters at 500 N, the middle picture of each subgraph represents the parameters at 600 N, and the bottom picture of each subgraph represents the parameters at 750 N.

the CNN. However, in the prediction, the *R* of the CNN model is the largest (0.996), while RMSE and MAE are the smallest (0.153 and 0.127, respectively). The effect of the CNN model on the test set is the least different

from that on the training set, and the overfitting degree of the three traditional methods is higher. Therefore, the new sound quality prediction method proposed in this paper is superior to other three methods.

5. Conclusion

The noise of the dual-phase Hy-Vo chain transmission system is different from that of the single-phase transmission. First of all, we have carried out the noise acquisition test of the dual-phase Hy-Vo chain transmission system. Then all the noise samples are subjectively evaluated, and the results are tested for correlation. The ACC of all testers is greater than 0.7, indicating that the subjective evaluation results are reasonable.

The MFCC feature maps of all noise samples are calculated as objective evaluation. By selecting different membership degrees for fuzzy generation, the original dataset is expanded by three times. The CNN model is constructed to predict the sound quality. The comparison results show that when the membership degree is 0.9, the prediction effect of standard full-frame MFCC feature map is the best.

Compared with the traditional sound quality prediction methods (GRNN, SVR, and RR), the CNN model has the best performance on the test set. The correlation coefficient is 0.996, the root mean square error is 0.153, and the MAE is 0.127. In addition, for the CNN model, the difference between the training effect and the prediction effect is small. Therefore, the new method proposed in this paper not only has the highest accuracy, but also has a strong ability to resist overfitting.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (no. 51775222) and the Science and Technology Development Project of Jilin Province in China (no. 20200401136GX).

References

1. ABDUL Z.K., AL-TALABANI A.K. (2022), Mel frequency cepstral coefficient and its applications: A review, *IEEE Access*, **10**: 122136–122158, doi: [10.1109/access.2022.3223444](https://doi.org/10.1109/access.2022.3223444).
2. BASNER M. *et al.* (2014), Auditory and non-auditory effects of noise on health, *Lancet*, **383**(9925): 1325–1332, doi: [10.1016/s0140-6736\(13\)61613-x](https://doi.org/10.1016/s0140-6736(13)61613-x).
3. BHATT D. *et al.* (2021), CNN variants for computer vision: History, architecture, application, challenges and future scope, *Electronics*, **10**(20): 2470, doi: [10.3390/electronics10202470](https://doi.org/10.3390/electronics10202470).
4. BUSTINCE H. *et al.* (2016), A historical account of types of fuzzy sets and their relationships, *IEEE Transactions on Fuzzy Systems*, **24**(1): 179–194, doi: [10.1109/tfuzz.2015.2451692](https://doi.org/10.1109/tfuzz.2015.2451692).
5. CHEN P., XU L., TANG Q., SHANG L., LIU W. (2022), Research on prediction model of tractor sound quality based on genetic algorithm, *Applied Acoustics*, **185**: 108411, doi: [10.1016/j.apacoust.2021.108411](https://doi.org/10.1016/j.apacoust.2021.108411).
6. CHENG Y., CHEN L., GE P., CHEN X., NIU J. (2023), Design and optimization of dual-phase chain transmission system based on single tooth chain plate, *Mechanics Based Design of Structures and Machines*, **51**(2): 899–913, doi: [10.1080/15397734.2020.1856683](https://doi.org/10.1080/15397734.2020.1856683).
7. CHENG Y., WANG Y., LI L., YIN S., AN L., WANG X. (2015), Design method of dual phase Hy-Vo silent chain transmission system, *Strojniški vestnik – Journal of Mechanical Engineering*, **61**(4): 237–244, doi: [10.5545/sv-jme.2014.2318](https://doi.org/10.5545/sv-jme.2014.2318).
8. CHENG Y., WANG Y., LI L., YIN S., AN L., WANG X. (2016a), Multi-variation characteristic of dual phase Hy-Vo silent chain transmission system, *Mechanism and Machine Theory*, **103**: 40–50, doi: [10.1016/j.mechmachtheory.2016.04.011](https://doi.org/10.1016/j.mechmachtheory.2016.04.011).
9. CHENG Y., YIN S., WANG X., AN L., LIU H. (2016b), Design and analysis of double-side meshing and dual-phase driving timing silent chain system, *Strojniški vestnik – Journal of Mechanical Engineering*, **62**(2): 127–136, doi: [10.5545/sv-jme.2015.2837](https://doi.org/10.5545/sv-jme.2015.2837).
10. DAR I.S., CHAND S., SHABBIR M., KIBRIA B.M.G. (2023), Condition-index based new ridge regression estimator for linear regression model with multicollinearity, *Kuwait Journal of Science*, **50**(2): 91–96, doi: [10.1016/j.kjs.2023.02.013](https://doi.org/10.1016/j.kjs.2023.02.013).
11. DRATVA J. *et al.* (2012), Transportation noise and blood pressure in a population-based sample of adults, *Environmental Health Perspectives*, **120**(1): 50–55, doi: [10.1289/ehp.1103448](https://doi.org/10.1289/ehp.1103448).
12. GOUMIRI S., BENBOUDJEMA D., PIECZYNSKI W. (2023), A new hybrid model of convolutional neural networks and hidden Markov chains for image classification, *Neural Computing and Applications*, **35**(24): 17987–18002, doi: [10.1007/s00521-023-08644-4](https://doi.org/10.1007/s00521-023-08644-4).
13. GUSKI R. (1997), Psychological methods for evaluating sound quality and assessing acoustic information, *Acta Acustica united with Acustica*, **83**(5): 765–774.
14. GÜNDOĞDU F.K., KAHRAMAN C. (2019), Spherical fuzzy sets and spherical fuzzy TOPSIS method, *Journal of Intelligent & Fuzzy Systems*, **36**(1): 337–352, doi: [10.3233/jifs-181401](https://doi.org/10.3233/jifs-181401).
15. HUANG H., WU J.H., LIM T.C., YANG M., DING W. (2021), Pure electric vehicle nonstationary interior sound quality prediction based on deep CNNs with an adaptable learning rate tree, *Mechanical Systems and Signal Processing*, **148**: 107170, doi: [10.1016/j.ymssp.2020.107170](https://doi.org/10.1016/j.ymssp.2020.107170).
16. JIN S., WANG X., DU L., HE D. (2021), Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN, *Applied Acoustics*, **172**: 107562, doi: [10.1016/j.apacoust.2020.107562](https://doi.org/10.1016/j.apacoust.2020.107562).
17. MANDOUH A.A., ALI M.E.N.O., MOHAMED M., TAHA L.G.E., MOHAMED S.A. (2023), A performance analysis of point CNN and mask R-CNN for building extraction from multispectral LiDAR data, *International Journal of Advanced Computer Science and Applications*, **14**(9), doi: [10.14569/ijacsa.2023.0140948](https://doi.org/10.14569/ijacsa.2023.0140948).

18. MOONDR A., CHAHAL P. (2023), Improved speaker recognition for degraded human voice using modified-MFCC and LPC with CNN, *International Journal of Advanced Computer Science and Applications*, **14**(4), doi: [10.14569/ijacsa.2023.0140416](https://doi.org/10.14569/ijacsa.2023.0140416).
19. PARK J.H., PARK H., KANG Y.J. (2020), A study on sound quality of vehicle engine sportiness using factor analysis, *Journal of Mechanical Science and Technology*, **34**(9): 3533–3543, doi: [10.1007/s12206-020-0805-0](https://doi.org/10.1007/s12206-020-0805-0).
20. RUAN K., LI Y. (2021), Fuzzy mathematics model of the industrial design of human adaptive sports equipment, *Journal of Intelligent and Fuzzy Systems*, **40**(4): 6103–6112, doi: [10.3233/jifs-189449](https://doi.org/10.3233/jifs-189449).
21. RUAN P., ZHENG X., QIU Y., ZHOU H. (2022), A bin-aural MFCC-CNN sound quality model of high-speed train, *Applied Sciences*, **12**(23): 12151, doi: [10.3390/app122312151](https://doi.org/10.3390/app122312151).
22. SHI M., LV L., GUO Z., SUN W., SONG X., LI H. (2021), High-low level support vector regression prediction approach (HL-SVR) for data modeling with input parameters of unequal sample sizes, *International Journal of Computational Methods*, **18**(08): 2150029, doi: [10.1142/s0219876221500298](https://doi.org/10.1142/s0219876221500298).
23. SONG X., YANG W. (2022), Research on the sound quality evaluation method based on artificial neural network, *Scientific Programming*, **2022**: 1–8, doi: [10.1155/2022/8686785](https://doi.org/10.1155/2022/8686785).
24. WANG Y., ZHANG S., MENG D., ZHANG L. (2022), Non-linear overall annoyance level modeling and interior sound quality prediction for pure electric vehicle with extreme gradient boosting algorithm, *Applied Acoustics*, **195**: 108857, doi: [10.1016/j.apacoust.2022.108857](https://doi.org/10.1016/j.apacoust.2022.108857).
25. YAO Q., WANG Y., YANG Y., YANG L. (2023), DOA estimation using GRNN for acoustic sensor arrays, *Multidimensional Systems and Signal Processing*, **34**(2): 575–594, doi: [10.1007/s11045-023-00877-9](https://doi.org/10.1007/s11045-023-00877-9).
26. YASIN A., AMIN M., QASIM M., MUSE A.H., MASTOR A.B.S. (2022), More on the ridge parameter estimators for the Gamma ridge regression model: Simulation and applications, *Mathematical Problems in Engineering*, **2022**: 1–18, doi: [10.1155/2022/6769421](https://doi.org/10.1155/2022/6769421).
27. ZHAN A., DU F., CHEN Z., YIN G., WANG M., ZHANG Y. (2022), A traffic flow forecasting method based on the GA-SVR, *Journal of High Speed Networks*, **28**(2): 97–106, doi: [10.3233/jhs-220682](https://doi.org/10.3233/jhs-220682).
28. ZHU Z., YIN H., LIANG Z. (2022), A prediction model for top-coal drawing capability in steep seams based on PCA-GRNN, *Geofluids*, **2022**: 1–9, doi: [10.1155/2022/3590764](https://doi.org/10.1155/2022/3590764).

Research Paper

Combined Evaluation of Room Acoustic Descriptors in Different Structural Configurations via ODEON Simulations and Artificial Neural Networks

Eriberto Oliveira DO NASCIMENTO*, Paulo Henrique Trombetta ZANNIN

Laboratory of Environmental and Industrial Acoustics and Acoustic Comfort
Federal University of Paraná – UFPR
Curitiba, PR, Brazil

*Corresponding Author e-mail: eriberto@ufpr.br

(received August 26, 2023; accepted June 18, 2024; published online September 26, 2024)

This study evaluated the combined sensitivity analysis of several room acoustic descriptors: reverberation time (T30), center time (Ts), early decay time (EDT), definition (D50), clarity (C50), useful-to-detrimental sound ratio (U50), and speech transmission index (STI); and also it assessed how these descriptors responded jointly to different acoustic-structural factors. The first-order factors were background noise (A), acoustic ceiling tile sound absorption coefficient (B), confinement (C), and occupancy (D), along with their interaction effects. A novel method is proposed for this joint evaluation of sensitivity factors. This method involves in situ measurements and an unreplicated 2^4 factorial design, which has been validated by ODEON software. The significance of input factors is determined using artificial neural networks (ANN) and the modified profile method (MPM), validated by multiple linear regression (MLR). Three significant correlation groups are identified at $p < 0.05$: group 1 (EDT, T30, Ts), group 2 (C50, D50), and group 3 (U50, STI). The ceiling material sound absorption (B) is found to affect reverberation (groups 1 and 2), while background noise (A) impacts STI and U50. A weak correlation is found between D50 and STI. These results are confirmed by the MLR and MPM methods.

Keywords: speech transmission index; reverberation time; artificial neural networks; room acoustics; ODEON simulation.



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

Classical approaches to acoustics are aimed at characterizing optimal classroom design for speech intelligibility based on reverberation time (RT) (LOCHNER, BURGER, 1960). In this context, other descriptors have been developed to consider exogenous effects such as background noise (BGN) to quantify the acoustic quality of rooms (KANG *et al.*, 2023). As a result, the speech transmission index (STI) was the leading objective descriptor designed to measure speech intelligibility (SI), and one of the driving forces for the development of STI was the recognition of the influence of BGN on intelligibility (HOUTGAST *et al.*, 1980). In this work, SI was assessed using the STI descriptor.

Many other descriptors have been proposed to quantify acoustic fields in classrooms; most of them are based on sound propagation and its energy ratios by

applying the diffuse field theory. These descriptors are: definition (D50), clarity (C50), center time (Ts), early decay time (EDT), and useful-to-detrimental sound ratio (U50). Additional descriptors, such as noise criteria (NC) curves and signal-to-noise ratio (SNR), are correlated with strictly subjective aspects of speech and sound perception (BRADLEY, 2011).

Many studies have sought to quantify the possible correlations between these parameters and the STI. For example, such studies included analyses of the relationship between STI and SNR (BRADLEY *et al.*, 1999), STI and RT (MIKULSKI, RADOSZ, 2011; RENNIES *et al.*, 2014), STI and energy descriptors (EDT, Ts, C50, D50) (SATO *et al.*, 2006; ANSAY, ZANNIN, 2016), and STI and U50 (BRADLEY *et al.*, 2003; SATO *et al.*, 2012; 2016; CHOI, 2017a; 2017b).

However, most of these studies have focused on quantifying the correlations between individual con-

struction design factors and STI, often examining only one aspect at a time. Furthermore, these studies have generally found that STI levels were low, often failing to meet the minimal threshold values of acoustic standards. Several studies have linked these low STI values to high BGN, which may be due to factors such as low sound absorption coefficients of ceiling panels, inadequate acoustic insulation, poor placement of sound-diffusion panels, or unfavorable building façade conditions (VISENTIN *et al.*, 2018; BISTAFA, BRADLEY, 2001; SALA, RANTALA, 2016; SECCHI *et al.*, 2017).

Therefore, the combined evaluation of acoustic descriptors is essential to determine how they are correlated and how room design and occupancy affect them. This paper seeks to contribute to the body of knowledge on this subject by quantitatively determining how the STI, T30 (reverberation time), EDT, C50, D50, Ts, and U50 respond jointly to the following factors: (A) background noise, (B) ceiling sound absorption; (C) confinement (if the room's windows and doors are open or closed), and (D) occupancy (if the room is or is not occupied by people, i.e., an audience), and by proposing a new method for this evaluation. An extensive comparison of results reported in the literature was also carried out.

Artificial intelligence in the form of artificial neural networks (ANNs) was used to ascertain the sensitivity of construction and occupancy conditions on acoustic parameters. This identification should lead to numerous benefits, contributing to improve the development of better classroom acoustic designs.

2. Materials and methods

2.1. In situ measurements

The speech transmission index measurements were taken in five classrooms by the indirect STI method, following the International Electrotechnical Commission 60268-16 standard (IEC, 2011) procedures. The BGN was measured for five minutes based on the equivalent continuous sound level (L_{eq}) in empty rooms with closed doors and windows to reduce external noise interference, using the B&K 2260 sound level meter. The NC value was then calculated as specified by American National Standard (ANSI, 2008).

The STI was measured using the following instruments: DIRAC room acoustics software (B&K type 7841), version 5.0, installed on a Sony VAIO notebook, Audio Interface ZE-0948 data acquisition board; Behringer FBQ800 equalizer; Lab. Gruppen LAB 300 amplifier; B&K 4227 mouth simulator, and a B&K 2260 real-time sound analyzer as the receiver microphone. According to IEC (2011) standard Sec. 7.2, the signal spectrum at the output of the B&K 4227 mouth simulator was equalized using the maximum length sequence (MLS) signal with a Pink + Blue filter.

The equalizer gains were then adjusted to the reference operational speech level of 60 dB, with a tolerable error of ± 1 dB, in the octave bands of 125 Hz and 8 kHz.

The descriptors T30, EDT, C50, D50, and Ts, were measured according to International Organization for Standardization [ISO] (2008), using the impulse response method. The e-sweep signal generated by DIRAC 5.0 software was used as an excitation signal. A B&K 4296 dodecahedral sound source was placed at a height of 1.5 m from the floor and more than 1.2 m from the walls. Five measurements were recorded with the receiver (B&K 2260 sound analyzer) in different positions in each classroom.

2.2. Classroom simulations

In this paper, five university classrooms were studied in the Federal University of Paraná, located in southern Brazil. Table 1 describes the maximum dimensions of the classrooms and their volumes. All the measured data, ODEON virtual models, sound absorbing materials, and classroom photos are available in the dataset reference provided by DO NASCIMENTO and ZANNIN (2023). ODEON software is widely used worldwide for predicting room acoustics parameters, for consulting, and for academic research. ODEON uses the image source method combined with a modified ray tracing algorithm for acoustic simulations. In this study, ODEON was used to simulate STI, T30, EDT, C50, D50, Ts, and U50 values based on various structural acoustic factors, such as A, B, C, and D.

Table 1. Dimensions of the measured classrooms.

Room	Width [m]	Length [m]	Height [m]	Volume [m ³]
1	12.40	13.40	2.75	358.11
2	10.00	8.70	2.95	236.06
3	11.14	7.57	3.91	355.16
4	11.50	5.53	4.20	289.27
5	13.99	13.99	5.84	1200.77

Classroom modeling and validation was performed using ODEON v. 11 software (RINDEL, 2012), which specifies that acoustic descriptors must be simulated in accordance with the guidelines (ISO, 2008; IEC, 2011) for in situ measurements. ISO (2008) establishes that T30, EDT, C50, D50, and Ts must be measured using an omnidirectional sound power level; hence, the Omni.SO-8 sound source was selected from the ODEON library. STI and U50 were simulated using the BB93_NORMAL_NATURAL.SO-8 sound source equalized at 60 dB. ODEON allowed for the weighting of male and female speech spectra using the virtual sound source, resulting in a total of three STI measures: STI_{male}, STI_{female}, and the non-gender filtered STI. A 20 cm \times 20 cm calculation meshgrid was placed at a height of 1.20 m from the floor, i.e., the average height of the ears of listeners sitting in typi-

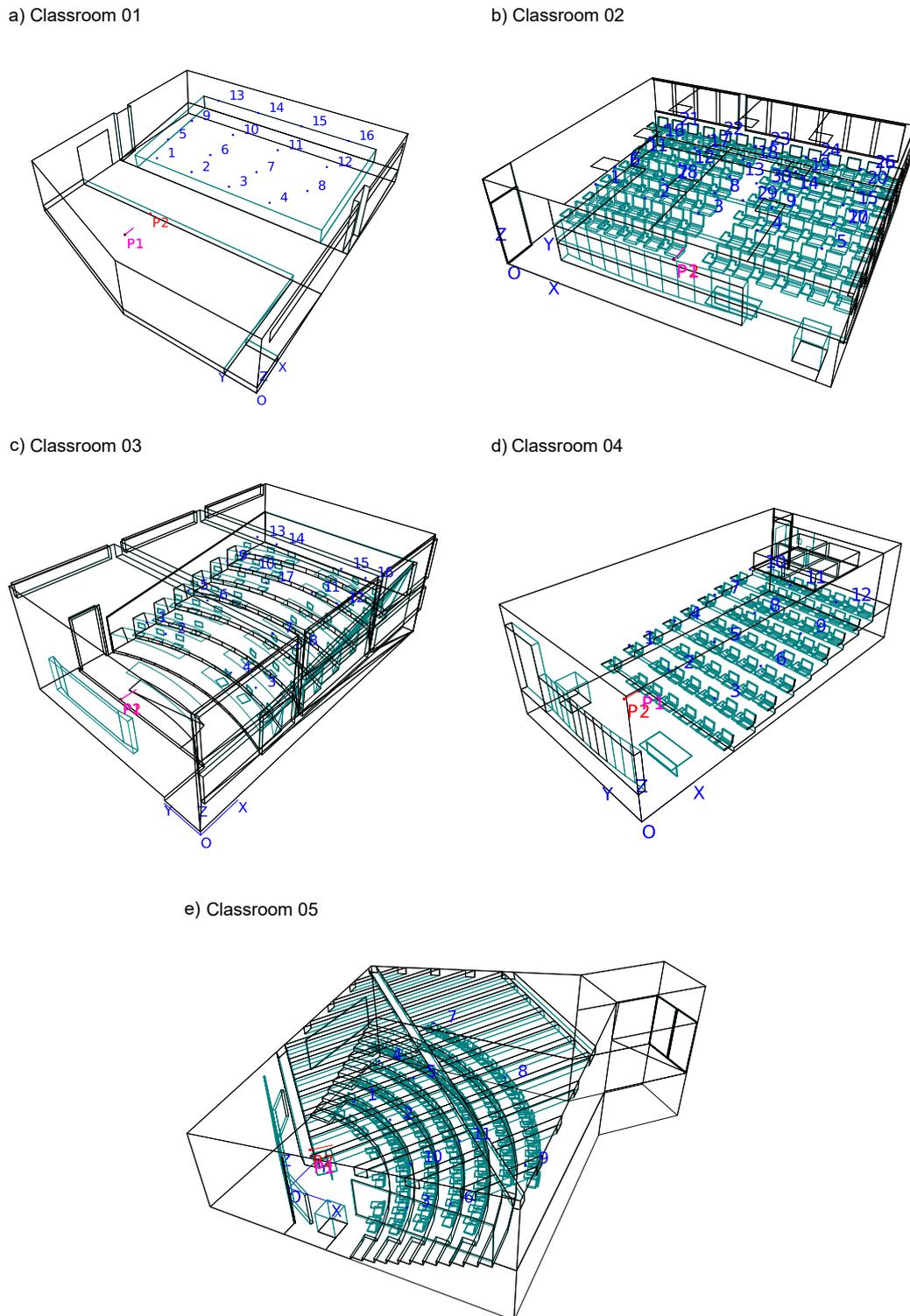


Fig. 1. Classrooms modeled using ODEON v. 11. The blue dots indicate the receiver positions, while the sound source is situated in the typical teacher position at the front of the class.

cal chairs. Figure 1 depicts the virtual models of the evaluated classrooms.

The receiver's placement in the virtual classrooms was chosen based on two criteria, to ensure full spatial coverage to accurately capture the room impulse

response (RIR) and to mimic the listener's perspective. The simulations were validated globally for each room using the root mean squared percentage error (RMSPE) – Eq. (1) – calculated between the measured and simulated values for descriptors T30 and STI in

their respective measurement positions in each classroom:

$$\text{RMSPE} = 100\% \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{\widehat{y}_i - y_i}{y_i} \right)^2}, \quad (1)$$

where n is the number of evaluated points, and \widehat{y}_i and y_i are the simulated (S) and measured (M) values, respectively, the mean percentage error (MPE) is $(100\%/n) \sum_{i=1}^n (\widehat{y}_i - y_i)/y_i$. Table 8 lists these errors for each classroom and for the T30 and STI descriptors.

To assess the goodness-of-fit of the validation in terms of simulated and measured values, the RMSPE was used as the deviation metric, because it is directly comparable to the just noticeable difference (JND) values, which were 5 % and 3 % for T30 and STI, respectively. Studies by BRADLEY (2011) and CHRISTENSEN *et al.* (2014) reported JND values more than five-fold higher than those of the current study. Therefore, a 10 % threshold for RMSPE was established as the decision criterion for validation, corresponding to 3 JND for STI and 2 JND for T30.

Each classroom was then calibrated by interactively fine-tuning the sound absorption coefficients of the various surface areas until the 10 % RMSPE threshold was reached. Lastly, the calibration curve for the acoustic model was built by line fitting of the measured versus simulated data, and the Pearson correlation coefficient was calculated. After experimentally validating a virtual room design created with ODEON, new conditions have been developed through the design of experiments (DOE), and a joint analysis was made to determine how the room acoustic parameters varied in these new conditions.

2.3. Creation of training dataset

MONTGOMERY (2012) described the DOE as a method to determine how modifying certain controllable factors (inputs) affects the system's response variables (outputs). The DOE expresses the relationships between the controllable factors (x) and responses (y), through a standard multilinear model, whose interactions are given in Eq. (2):

$$y = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{i < j} \beta_{ij} x_i x_j + \epsilon, \quad (2)$$

where β_0 is the model's linear intercept coefficient, β_j is the regressor coefficient of the main factors x_j , β_{ij} represents the coefficients of the regressors for the effects of the combinatorial interaction between the main factors $x_i x_j$, and k is the number of factors. A 2^k DOE was applied in this work, with $k = 4$ controllable factors and two levels, ranging from minimum (-1) to maximum ($+1$). The 2^k factorial design with a range of -1 to $+1$ was chosen for its efficiency in screening factors

and its representation of the standardized effect, which simplifies analysis and interpretation. Table 2 describes the factors and their levels.

Table 2. DOE levels and factors.

Level	Controllable factors – natural scale			
	Noise criteria	Ceiling absorption	Confinement	Occupancy
Lowest value	15	0.10	Open	Present
Highest value	40	0.90	Closed	Absent
	Controllable factors – encoded scale			
	x_1	x_2	x_3	x_4
	A	B	C	D
Lowest value	-1	-1	-1	-1
Highest value	$+1$	$+1$	$+1$	$+1$

Within the context of the DOE (see Table 2), the transition of variables from -1 to $+1$ in the encoded or natural space has significant implications for multiple linear regression (MLR) models. A positive shift from -1 to $+1$ signifies an ascent towards the higher end of the experimental range, indicating a positive influence on the evaluated response variable. Conversely, a negative shift denotes a descent towards the lower range of the response, suggesting a negative impact.

The simulations with ODEON software were used to emulate classrooms and their respective acoustic parameters as a function of four controllable factors (see Table 2). The combinatorial experimental conditions were based on the factors called experimental runs, which correspond to the design matrix lines x_{ij} listed in Table 3.

The effect of room occupancy, i.e., audience, was estimated by assigning sound absorption coefficients to the equivalent absorption surface area of the audience in ODEON classroom model, based on the literature (see Table 4).

Specifically, for a full 2^4 factorial design, the equivalent linear model with 3rd and 4th-order interactions is shown in Eq. (3):

$$\begin{aligned} y_{k,ij} = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 \\ & + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 \\ & + \beta_{34} x_3 x_4 + \beta_{123} x_1 x_2 x_3 + \beta_{124} x_1 x_2 x_4 \\ & + \beta_{134} x_1 x_3 x_4 + \beta_{234} x_1 x_2 x_3 + \beta_{1234} x_1 x_2 x_3 x_4, \quad (3) \end{aligned}$$

here, the room acoustics parameters are frequency-dependent and are written in octave bands to express the response of a given parameter. Therefore, a nested output ($y_{k,ij}$) was used in each classroom. Let k (= T30, EDT, C50, D50, Ts, STI) indicate the room acoustic parameters, i (= 1 to 16) represent the experimental runs, and j (= 63 Hz, 125 Hz, 250 Hz, 500 Hz,

Table 3. Design matrix used in simulations with ODEON software.

	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	ACD	BCD	ABCD
Run	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4	x_2x_3	x_2x_4	x_3x_4	$x_1x_2x_3$	$x_1x_2x_4$	$x_1x_3x_4$	$x_1x_2x_3$	$x_1x_2x_3x_4$
1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1
2	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1
3	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	1	-1
4	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1
5	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1
6	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1
7	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	1
8	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1
9	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1
10	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1
11	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1
12	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1
13	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1
14	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1
15	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	-1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 4. Sound absorption coefficient of materials used to emulate room occupancy.

Room	Occupancy – audience area	ODEON number	Reference	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
1, 2, 3, 5	Audience on heavily upholstered seats	11007	BERANEK, HIDAKA (1998)	0.72	0.80	0.86	0.89	0.90	0.90	0.90
4	Audience on wooden chairs, 2/m ²	11004	MEYER <i>et al.</i> (1964)	0.24	0.40	0.78	0.98	0.96	0.87	0.87

1 kHz, 2 kHz, 4 kHz, and 8 kHz) indicate the octave bands.

Consequently, to obtain the $y_{i,kj}$ output used to train the ANNs in each classroom, 51 responses were generated in each run, 48 corresponding to the descriptors T30_{63 Hz–8 kHz}, EDT_{63 Hz–8 kHz}, C50_{63 Hz–8 kHz}, D50_{63 Hz–8 kHz}, and Ts_{63 Hz–8 kHz} in the octave bands at 63 Hz, 125 Hz, 250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz, while the other three responses were STI_{male}, STI_{female}, and the STI in nominal values.

2.4. Artificial neural networks

The fundamental objective of ANNs is to produce generalizations through a training process (RUSSELL, NORVIG, 1996). Generalization refers to estimates for patterns not used in training. Training occurs when the weights of the connections of an ANN's neurons are adjusted to minimize a loss function. Hence, an ANN is a set of nested functions, with regard to, $Y_n = W_n X_{n-1}$, that have undergone nonlinear point-wise transformations, resulting in a generalized model, $X_n = \Phi(Y_n)$, where W_n is the weight matrix of layer n , and X_{n-1} represents the input patterns for a previous layer $n - 1$. The generalized multilayer model is expanded, as shown in Eq. (4):

$$Y_n = \Phi_n \left[W_n \Phi_{n-1} \left[W_{n-1} \Phi_{n-2} \left[\dots \Phi_1 \left[W_1 X \right] \right] \right] \right]. \quad (4)$$

The nonlinear transformations performed by Φ are called activation functions. As proposed by LECUN *et al.* (1998), the weight update rule depends on a loss function that, for a pattern (p) compares the ground truth value (D^p) with the neural networks' output model, given $M(X_n, W)^P$, using the residual error metric, as shown in Eq. (5):

$$E^P = \frac{1}{2} \left(D^p - M(X_n, W)^P \right)^2, \quad (5)$$

where E^P is the residual sum of squares about the $M(X_n, W)^P$, which is the ANN output value estimated as a function of the weight matrices (W), and the input patterns (X_n). Accordingly, neural network training can be interpreted as an optimization paradigm, in which the weights are adjusted as the loss function is minimized. Notably, in regression problems, a widely used loss function is the mean square error (MSE) shown in Eq. (6):

$$\text{MSE}_{\text{train}} = \frac{1}{P} \sum_{p=1}^P E^P, \quad (6)$$

where P stands for the total number of samples. Moreover, the weight update rule can be applied using various approaches, such as the gradient-based learning paradigm, e.g., the stochastic gradient descent, given in Eq. (7):

$$W(t) = W(t-1) - \eta \frac{\partial E}{\partial W}. \quad (7)$$

In this work, the error backpropagation learning algorithm was employed to update the weight matrices. Equations (8)–(10) portray the generic approach to backpropagation learning, which is interpreted as the automatic differentiation in the reverse mode (BAYDIN *et al.*, 2017):

$$\frac{\partial E^P}{\partial Y_n} = \Phi'(Y_n) \frac{\partial E^P}{\partial X_n}, \quad (8)$$

$$\frac{\partial E^P}{\partial W_n} = X_{n-1} \frac{\partial E^P}{\partial Y_n}, \quad (9)$$

$$\frac{\partial E^P}{\partial X_{n-1}} = W_n^T \frac{\partial E^P}{\partial Y_n}. \quad (10)$$

In this work, ANNs were applied as universal functional approximators to map the relationship between inputs and outputs described in Subsec. 2.2.

2.4.1. Input and target pre-processing

The ANN model’s input-output training pairs were based on the DOE method described in Table 3, which served not only to create an optimal experimental design but also as an augmentation method. According to the DOE runs, augmentation refers to the virtual response values for the acoustic parameters obtained through ODEON software.

Firstly, an ANN was trained individually for each acoustic parameter. Although the same input matrix was used in all the ANN models, only the output variable varied according to the evaluated acoustic parameters. As a result, a single training sample consisted of a 15-input feature vector and a scalar output. The input feature vector was derived from the DOE model inputs, A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, and ABCD, and the scalar target corresponded to the reduced single band acous-

tic parameter. The $Y_{16,8}$ octave band output matrix was dimensionally reduced to the $Y_{16,1}$ octave single band vector encoded in the first principal subspace, using principal component analysis – PCA (JOLLIFFE, CADIMA, 2016). Thus, the entire training set for one classroom was composed of an $X_{16,15}$ input matrix and a $Y_{16,1}$ output vector.

The PCA worked by deflating redundant information through multicollinearity reduction, using eigenvector information as a threshold measure. For example, for an $X_{m \times n}$ dataset, where m is the number of observations of n variables, the PCA diagonalizes X through its variance-covariance matrix, given as $S = (1/N - 1) X X^T$, where N is the number of observations, $n = 1, 2, \dots, N$. The eigendecomposition of S results in $S u_k = \lambda_k u_k$, thus determining the eigenvectors u_k and their eigenvalues λ_k . The sorted eigenvalue λ_k informs the cumulative explained variance in the original X dataset (JOLLIFFE, 2011).

Secondly, RUSSELL and NORVIG (1996) recommend applying the min-max scaling range -1 to 1 , in the training dataset. The input data set was then assigned to the design matrix, which was already in a scaled form; hence, only the target variables underwent scaling through Eq. (11):

$$y_{ij} = a + \frac{(d_{ij} - \min(d_{ij}))(b - a)}{\max(d_{ij}) - \min(d_{ij})}, \quad (11)$$

where y_{ij} is the j scaled target feature, i.e., j ($= T30, EDT, C50, D50, Ts, STI$), $i = 1, 2, \dots, 16$ is the i -th sample of the j feature, simulated d_{ij} stands for the non-scaled feature, and a and b correspond to -1 and $+1$, which represent the transformed range.

The hyperbolic tangent activation function, which has the same image range as min-max scaling, was employed here. Figure 2 shows the assignment of a training sample, with the input, $x_{1,15}$ (vector), and output

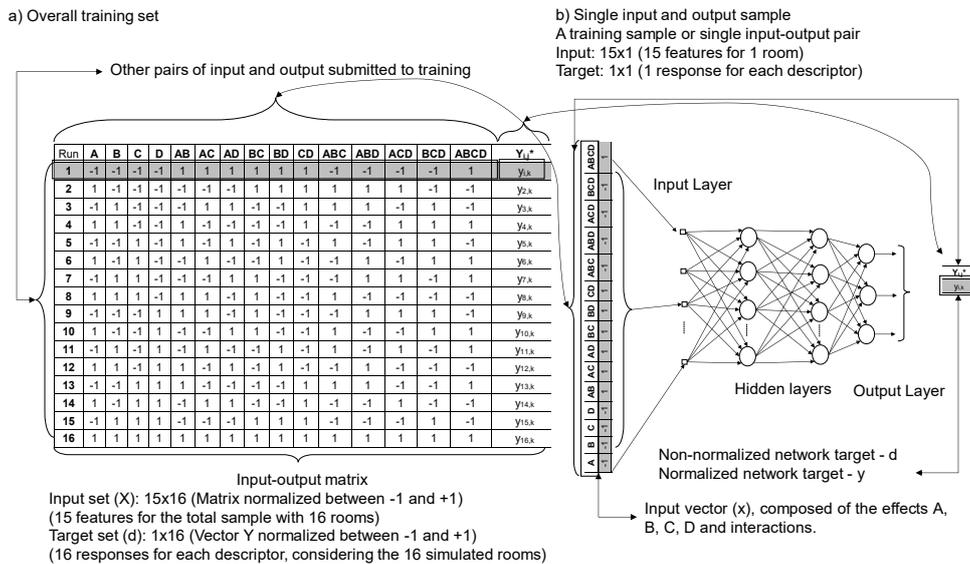


Fig. 2. Attribution of the input/output relationship in training the network.

$y_{1,1}$ (scalar) in the ANN and its relationship with the DOE for a classroom.

The algorithm described in Fig. 2 was used for each of the five classrooms where measurements were taken, and 16 virtual rooms were generated corresponding to the curves (runs). Thus, a total of 80 virtual classrooms were created using ODEON software. As mentioned earlier, the $X_{16,15}$ matrix was then adopted as the ANN input, while the target for subsequent ANN training was the $Y_{PCA_{16,6}}$ vector. Lastly, the ANN was trained with an input composed of 15 features, which were the main factors A, B, C, D and their respective 2nd order (AB, AC, AD, BC, CD), 3rd order (ABC, ABD, ACD, BCD) and 4th order (ABCD) interactions, and only one scalar as output.

Regarding the training dataset, a correlation matrix was evaluated in two ways: (a) by intra-correlations between the responses for the same descriptor, in the octave bands, and (b) the intercorrelations between the different descriptors T30, Ts, EDT, D50, U50, and STI in their octave bands.

2.4.2. Artificial neural network topological structures

The multilayer perceptron (MLP) architecture was utilized to train the ANN. The architecture of a network determines how artificial neurons are arranged and how signals flow between them (LECUN *et al.*, 2015). The term network topology indicates the number of neurons contained in these hidden layers. The ANN model was employed as a regression problem for multidimensional function approximation, leveraging the universal approximation theorem. This approach enabled the sensitivity analysis of the effects of controllable factors on the ANN models, as described in Subsec. 2.5. Table 5 outlines the design of the ANN applied in this study.

ANNs are subject to overfitting, which occurs when they approximate the training data too closely, causing them to lose their ability to generalize. Therefore, early stopping and holdout were employed to reduce the possibility of overfitting, according to the heuristics proposed by PIOTROWSKI and NAPIORKOWSKI (2013).

Having trained the ANN, the network inference phase began, after the synaptic weights had already been adjusted. Thus, the inference of a network consisted of applying its input values (X) and calculating its estimated output (y). The quality of ANN training was evaluated by comparing the target (d) and the estimated neural network (y), using Pearson's R^2 correlation coefficient and the MSE as performance metrics. Lastly, an ANN was trained for each of the outputs $T30_{PCA}$, EDT_{PCA} , $C50_{PCA}$, $D50_{PCA}$, Ts_{PCA} , and STI_{PCA} .

The optimized response from the 50 independent training sessions of the neural network, called the average equivalent network, was calculated for each topology. All the computational implementations were developed in MATLAB[®].

2.5. Artificial neural networks input variable sensitivity

In this work, the modified profile method (MPM) (GEVREY *et al.*, 2003; DO NASCIMENTO, OLIVEIRA, 2016) was applied to determine the relative significance of the input variables. The MPM calculates the significance of the ANN input variables, using the angular coefficient of the linear regression of the profile curve. Hence, the significance rating for the 15-input vector, A, B, C, D, AB, AC, AD, BC, BD, CD, ABC, ABD, ACD, BCD, and ABCD was calculated as the average of the effects in the five evaluated classrooms. Based on the non-replicated DOE, the significance of the MPM effects was estimated by transforming the angular coefficients into the z -score.

Multiple linear regression (MLR) was used to benchmark the MPM results. MLR was applied directly into the design matrix to compare the sensitivity of each input variable in the output. The sensitivity was taken as the regressor values shown in Eq. (2). The regressors were estimated using the least squares method, which yields the following equation, $\hat{\beta} = (X^T X)^{-1} y$, where X is the design matrix shown in Table 3, X^T is the transpose design matrix, $(\cdot)^{-1}$ is the inverse matrix operator, y is the output variable under analysis,

Table 5. Configurations of the neural network design.

Architecture	MLP
Number of inputs/number of outputs	15/1
Topology 1	MLP 15-5-5-1
Topology 2	MLP 15-10-10-1
Topology 3	MLP 15-15-15-1
Topology 4	MLP 15-20-20-1
Topology 5	MLP 15-30-30-1
Topology 6	MLP 15-35-35-1
Training algorithm	Error backpropagation, optimized by Levenberg–Marquardt
Activation functions in hidden layers	Hyperbolic tangent function
Activation function in the output layer	Linear function

y (= T30, EDT, C50, D50, Ts, STI), one output at a time. The estimated model is $\hat{y} = X\beta$, and the residual values are $e = y - \hat{y}$. The quality of the regression was evaluated using the sum of residual squares, given as $SS_r = y^T y - \hat{\beta}^T X^T y$ and through Pearson's squared linear correlation coefficient (R^2), usually known as the determination coefficient.

3. Results and discussion

Figure 3 shows the values of the measured descriptors. As can be seen, the hypothesis proposed by MONTGOMERY (2012), who recommended that the DOE should be carried out in the most differentiated conditions possible, was met since the descriptors showed high statistical variability.

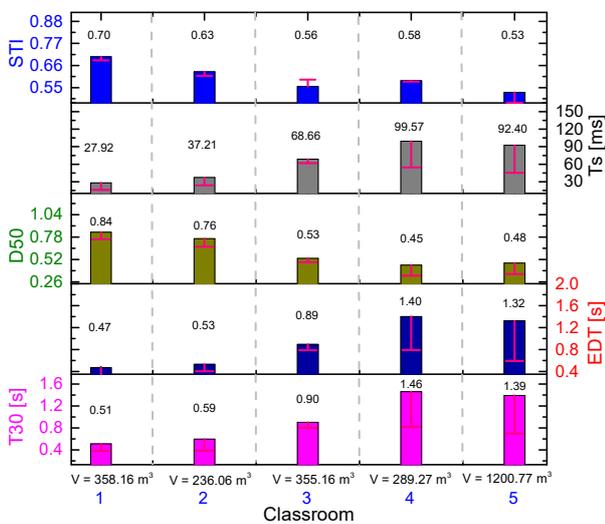


Fig. 3. Mean STI, T30, EDT, D50, and Ts values measured in the classrooms.

Table 6 lists the measured background noise levels and the respective NC of the classrooms. The NC was

used as an input for ODEON software, as indicated in Fig. 2.

Table 7 lists the average measured values of SNR in the octave bands. The STI values for each classroom were calculated according to the IEC (2011) standard Annex M: adjustments to measured STI and STIPA results for simulation of occupancy noise and different speech levels. The BGN level is listed in Table 6, and the input sound source signal level was equalized at L_{eq} 60 dB, as described in Sec. 2. The corresponding mean STI values measured in classrooms 1 up to 5 were 0.70 ± 0.06 , 0.63 ± 0.02 , 0.56 ± 0.02 , 0.58 ± 0.02 , and 0.53 ± 0.03 , respectively.

According to Subsec. 2.2, the classroom simulations were validated by comparing the most significant percent error values through the calibration curve and the R^2 value, as shown in Fig. 4.

Table 8 describes the generalization of the procedure highlighted in Fig. 4 to calculate validation errors pertaining to the other evaluated classrooms.

Using the validated virtual classroom models, new simulations were generated corresponding to the conditions illustrated in Fig. 2, which comprised the 15 combinations between factors A, B, C, and D. Figure 5 illustrates the distribution of the responses of descriptors T30, EDT, Ts, D50, U50, and STI as a function of each DOE run, using the results of the classroom two as an example.

More than 95 % of the variance was explained solely by the first principal component through the PCA method, as was also the case in the other classrooms. All the multi-band information contained in Fig. 5 was therefore condensed in Fig. 6.

Figure 7 shows the correlation matrix for the descriptors EDT, T30, Ts, D50, C50, U50, and STI obtained after applying the PCA.

As can be seen, Fig. 7 and Table 9 indicate the correlations and significance between the groups. Thus,

Table 6. Background noise level [dB] measured in the classrooms and noise criteria (NC) rating.

Room	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz	NC
1	34.5	30.0	23.2	21.4	21.1	19.6	17.0	23
2	48.3	38.6	32.2	32.0	29.9	23.0	19.5	35
3	48.0	41.0	40.2	38.9	30.0	25.3	24.9	30
4	48.9	42.6	34.6	28.9	25.1	21.9	21.4	32
5	50.4	44.9	36.6	31.9	29.3	22.8	18.0	34

Table 7. Mean measured SNR values.

Room	SNR, mean ± standard deviation						
	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
1	18.61 ± 4.62	25.61 ± 1.75	22.39 ± 0.68	19.22 ± 3.44	11.83 ± 2.43	8.11 ± 2.95	-3.61 ± 2.66
2	21.28 ± 2.67	20.48 ± 0.92	19.48 ± 0.59	17.44 ± 1.23	10.96 ± 1.59	6.04 ± 1.51	-9.32 ± 2.21
3	15.94 ± 3.73	20.19 ± 2.64	19.13 ± 3.01	18.31 ± 3.34	13.63 ± 3.03	10.13 ± 2.50	-1.44 ± 1.93
4	8.00 ± 3.28	16.17 ± 3.69	20.67 ± 3.26	23.17 ± 2.66	21.33 ± 2.77	20.83 ± 1.64	15.92 ± 1.62
5	14.09 ± 2.43	22.55 ± 3.08	21.91 ± 2.30	19.36 ± 1.80	11.09 ± 2.02	9.27 ± 1.85	-2.82 ± 1.83

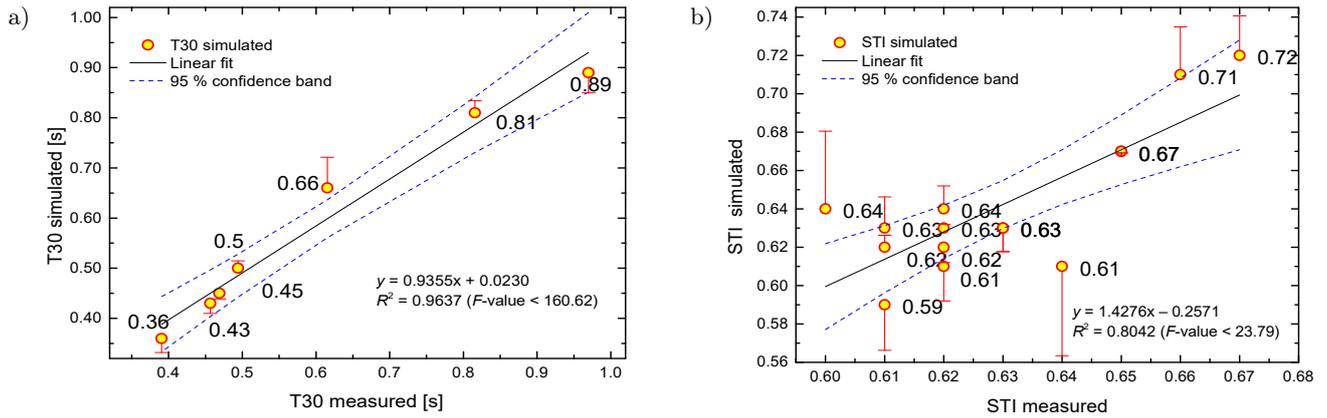


Fig. 4. Calibration curve of: a) T30; b) STI in classroom 2. The error bar is the residual value, i.e., the difference between the measured value and the corresponding value estimated by the linear fit.

Table 8. Forecast accuracy validation metrics for ODEON acoustic model via MPE, RMSPE and correlation.

Room	Estimation type	STI		T30 [s]								Mean		RMSPE, r^2
		Mean	RMSPE, r^2	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz				
1	Measured	0.70	7.14 %, 0.47	0.78	0.64	0.45	0.37	0.44	0.46	0.46	0.51	6.03 %; 0.96		
	Simulated	0.72		0.75	0.62	0.42	0.40	0.43	0.47	0.41	0.50			
	MPE	1.33 %	–	4.03 %	2.67 %	7.18 %	7.02 %	3.04 %	1.73 %	10.87 %	–	–		
2	Measured	0.63	3.99 %, 0.80	0.97	0.82	0.62	0.49	0.47	0.46	0.39	0.60	5.77 %; 0.96		
	Simulated	0.64		0.89	0.81	0.66	0.50	0.45	0.43	0.36	0.59			
	MPE	2.00 %	–	8.22 %	0.67 %	7.27 %	1.21 %	4.00 %	5.81 %	7.81 %	–	–		
3	Measured	0.56	5.03 %, 0.57	0.71	0.83	0.92	1.02	1.02	0.92	0.81	0.89	6.35 %; 0.91		
	Simulated	0.54		0.77	0.81	0.89	0.95	0.98	0.84	0.74	0.85			
	MPE	3.48 %	–	8.33 %	2.90 %	3.13 %	6.61 %	3.52 %	8.48 %	8.14 %	–	–		
4	Measured	0.58	3.88 %; 0.32	2.62	2.08	1.58	1.08	1.07	1.06	0.86	1.29	4.91 %; 0.99		
	Simulated	0.59		2.68	1.99	1.60	1.01	0.99	1.03	0.80	1.24			
	MPE	0.43 %	–	2.43 %	4.42 %	1.54 %	6.12 %	7.34 %	2.63 %	6.52 %	–	–		
5	Measured	0.53	6.63 %; 0.77	2.60	2.11	1.31	1.01	1.01	0.95	0.77	1.39	4.19 %; 0.99		
	Simulated	0.55		2.54	2.10	1.22	0.98	0.98	0.88	0.78	1.35			
	MPE	5.01 %	–	2.13 %	0.50 %	6.71 %	3.09 %	2.91 %	7.29 %	1.38 %	–	–		

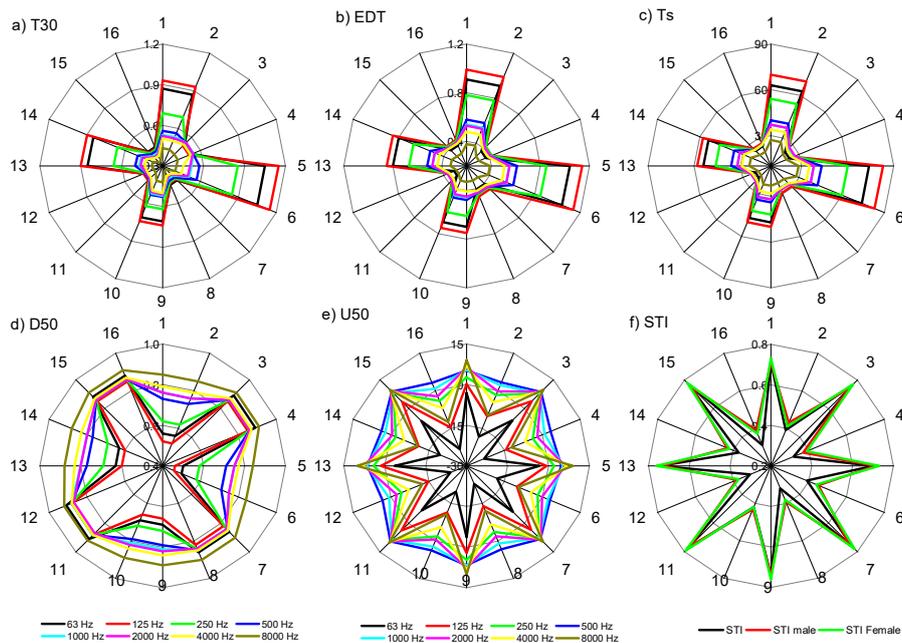


Fig. 5. Simulation of descriptors for each run for classroom 2.

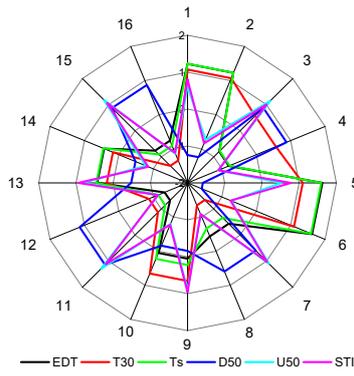


Fig. 6. Dimensional reduction of the acoustic descriptors.

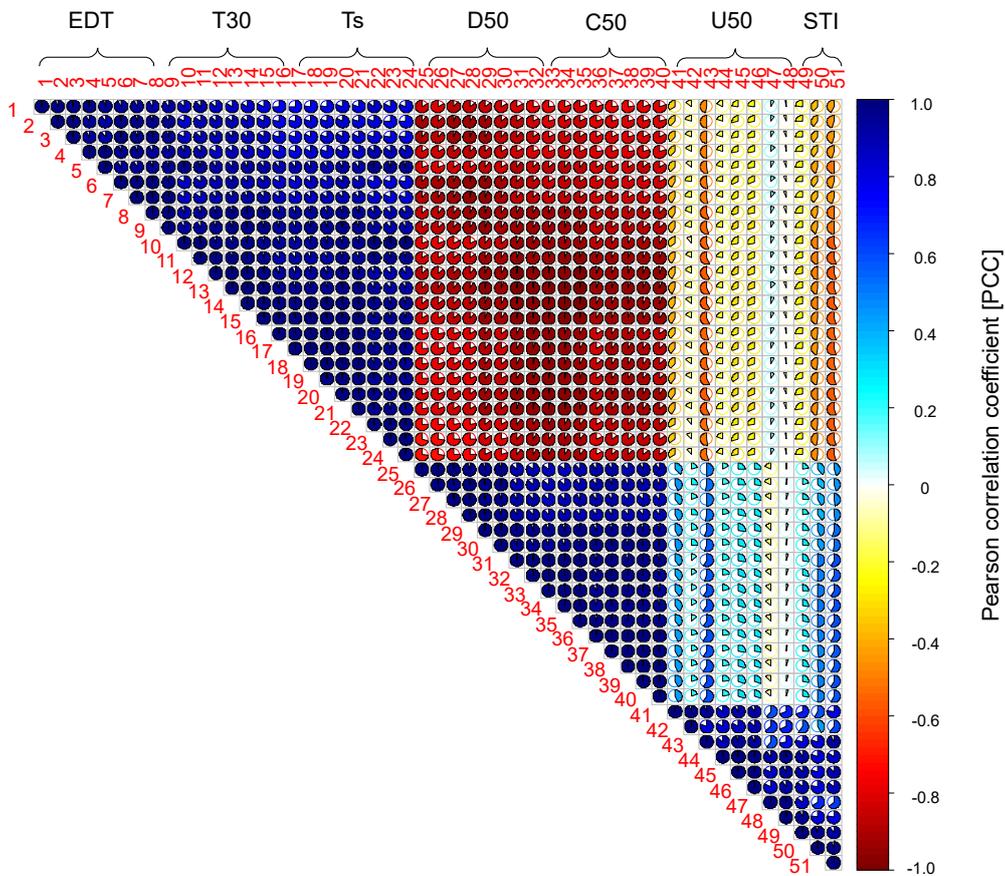


Fig. 7. Correlation matrix plot between all the acoustic descriptors.

Table 9. Correlations between the evaluated descriptors and their significance.

	EDT [s]	T30 [s]	Ts [ms]	D50	C50 [dB]	U50 [dB]	STI
EDT [s]	1						
T30 [s]	0.947*	1					
Ts [ms]	0.996*	0.948*	1				
D50	-0.963*	-0.902*	-0.965*	1			
C50 [dB]	-0.914*	-0.863*	-0.912*	0.977*	1		
U50 [dB]	-0.273*	-0.254*	-0.266*	0.267*	0.257*	1	
STI	-0.281*	-0.252*	-0.278*	0.269*	0.256*	0.975*	1

Note: a 2-tailed test of significance is used; *correlation is significant at the 0.05 level.

upon analyzing the correlation matrix, three groups were observed, identified by the dark blue samples, i.e., the 1st (EDT, T30, and Ts), the 2nd (C50 and D50), and the 3rd (U50 and STI).

In the benchmarking literature, TANG (2008) made a joint evaluation of the speech-related acoustic descriptors using regression models. The author reported the following correlations:

$$\begin{aligned} \text{STI (D50)} \ r^2 &= 0.893 (+); & \text{STI (C80)} \ r^2 &= 0.916 (+); \\ \text{STI (Ts)} \ r^2 &= 0.907 (-); & \text{STI (RT)} \ r^2 &= 0.903 (-). \end{aligned}$$

For comparison, the following correlations were extracted from Table 9:

$$\begin{aligned} \text{STI (D50)} \ r^2 &= 0.269 (+); & \text{STI (C50)} \ r^2 &= 0.256 (+); \\ \text{STI (Ts)} \ r^2 &= 0.278 (-); & \text{STI (RT)} \ r^2 &= 0.252 (-). \end{aligned}$$

Thus, it can be inferred that the effects followed the same vein.

Additionally, a literature review by MINELLI et al. (2022) found that the acoustic descriptors are significantly affected by the occupation condition of built environments (factor D), i.e., unoccupied or occupied. This finding is in line with ours, which indicated that the entire acoustic response was altered in various ways by a combined change in construction factors (A, B, C, and D).

Likewise, CROCE et al. (2023) and CHOI (2020) independently established relationships between STI,

U50, and RT and assessed the influence of construction factors. CHOI (2020) determined that, quantitatively, C50(125–4 kHz) is more strongly correlated with U50(125–4 kHz) $r^2 = 0.824$ than with STI (0.627). Hence, the results of linear regression indicated that C50, specifically, plays a more significant role in increasing r^2 values of both U50 and STI, outweighing the impact of the background noise component (factor A). This conclusion can be seen in the last two rows of Table 10, where the encoded effects on U50 and STI were the same (-)A, (+)B, (-)AB, and (+)D, and as proxy, the correlation STI(U50) was 0.975 (+), while the other correlations were lower: C50(U50) 0.257 (+), C50(STI) 0.269 (+).

The r^2 between U50(125–4 kHz) and C50(125–4 kHz) is 0.824, while that between STI and C50(125–4 kHz) is 0.627. This indicates that the correlation between STI and C50 is not much stronger, which resembles the result described in Table 9, i.e., C50(STI) 0.269 (+). In an extended discussion about the role of early reflections on C50 and their implications for speech intelligibility, PRODI et al. (2022) reported the same findings. This was highlighted in the current work by the MPM, which separated C50 and D50 from the STI and U50 group.

Graphically, to assess the significance rating, Fig. 8 illustrates the effects of A, B, C, and D, and their respective interactions in the classrooms on the evaluated descriptors, using the MPM.

In Fig. 8, note the spatial uniformity of the significance rating, which is why the discussion was gener-

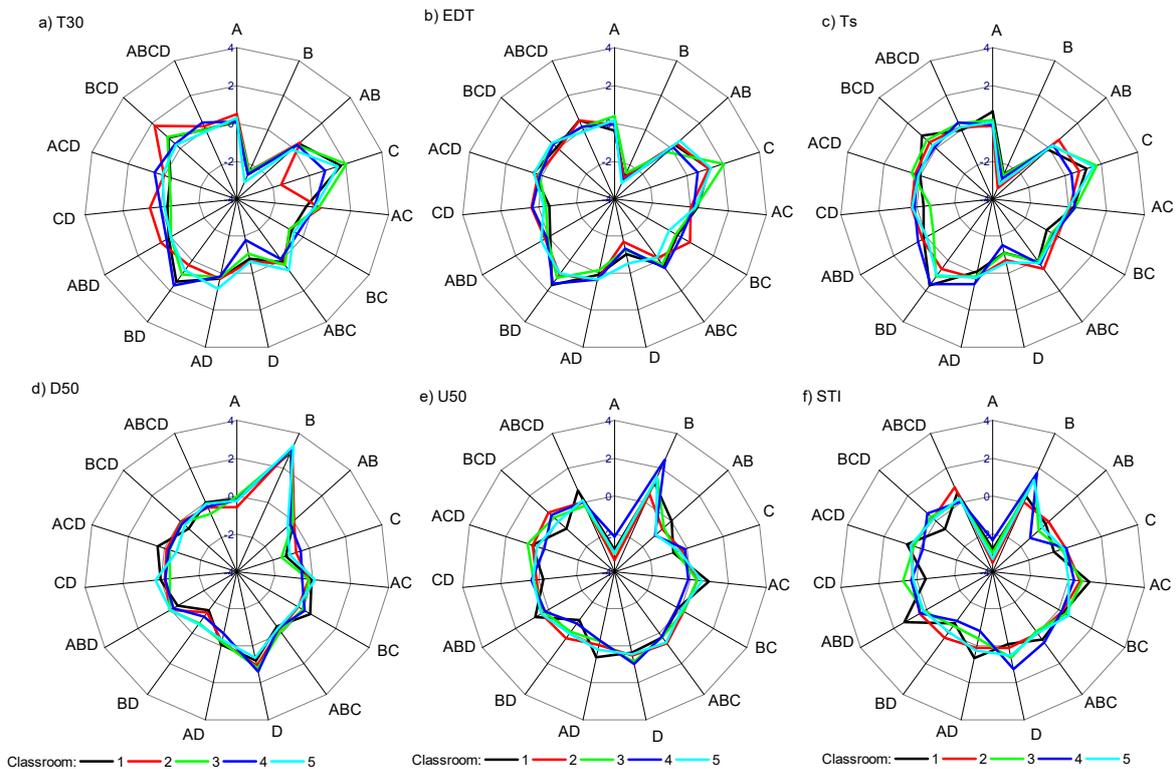


Fig. 8. Comparison of the significance rating in the classrooms.

Table 10. Comparison of the absolute effects on the descriptors: MPM versus MLR.

Group	Parameter	Rating and effects				
			1	2	3	4
1	EDT [s]	Coded effect	(-)B	(+)C	(+)BD	(-)D
		MPM magnitude	2.902	1.242	1.212	1.202
		MLR magnitude	2.963	1.348	1.133	1.222
	T30 [s]	Coded effect	(-)B	(+)BD	(-)D	(+)C
		MPM magnitude	2.728	1.210	1.076	0.834
		MLR magnitude	2.858	1.492	1.175	1.155
	Ts [ms]	Coded effect	(-)B	(+)BD	(-)D	(+)C
		MPM magnitude	2.944	1.236	1.146	1.134
		MLR magnitude	2.968	1.365	1.221	1.084
2	D50	Coded effect	(+)B	(-)C	(-)BD	(+)D
		MPM magnitude	3.096	1.096	1.074	1.044
		MLR magnitude	3.168	1.076	1.048	1.046
3	U50 [dB]	Coded effect	(-)A	(+)B	(-)AB	(+)D
		MPM magnitude	2.906	1.288	0.754	0.704
		MLR magnitude	3.347	0.910	0.598	0.582
	STI	Coded effect	(-)A	(+)B	(+)D	(-)AB
		MPM magnitude	3.072	0.996	0.632	0.582
		MLR magnitude	3.340	0.904	0.631	0.565

alized. Moreover, it was found that the groups in the effect analysis were the same as those in the correlation analysis. Table 10 quantifies the significance ratings of the impact shown in Fig. 8. The four main effects in each group are highlighted with a (+) sign, indicating an increase in the descriptor, while the (-) sign indicates a decrease in the descriptor as the effect varies from -1 to +1.

As can be seen in Table 10, the acoustic parameters were given the same rating by both MPM and MLR, matching the construction design factors. Likewise, it was found that the rating was identical for each acoustic parameter, confirming that the MPM showed excellent convergence when benchmarked with the MLR.

In the first group (EDT, T30, and Ts), the most important factors were: (-)B, (+)BD, (-)D, and (+)C. The increase in absorption (B) from -1 to +1, implied a decrease in EDT, T30, and Ts. The same finding was reported by BERANEK (2006).

The second group (C50, D50) responded, in terms of magnitude, as (+)B, (-)C, (-)BD, and (+)D. The increase in absorption (B), from -1 to +1 implied an increase in D50. Similarly, a positive correlation was found between D50 and both U50 and STI, and these results are promising, since D50 is strongly associated with STI (BRADLEY *et al.*, 2003; CROCE *et al.*, 2023; CHOI, 2020). It should be noted that the higher the D50, the greater the intelligibility, according to the relationship described by BRADLEY *et al.* (2003).

In the third group (U50 and STI), the first effect with the highest impact on intelligibility was background noise (A). In contrast, the second most significant effect was sound absorption (B). As a result, the increase in background noise from -1 to +1

caused a decrease in STI, while sound absorption alone did not exert a significant effect. Thus, the importance ratio of effect A on B was approximately three-fold. The factor A explained most of the variation in STI. A similar finding has been reported in several studies (BISTAFÀ, BRADLEY, 2000; RENNIES *et al.*, 2014; SATO *et al.*, 2012; LECCESE *et al.*, 2018).

It is essential to note that while this work provided a comprehensive understanding of the relationships between various acoustic descriptors and building design, there is room for improvement to enhance the consistency and reliability of results. Firstly, the study focused on a specific set of factors (A, B, C, and D) and their influence on acoustic parameters, potentially overlooking other variables that could contribute to the acoustic environment. Moreover, some limitations of this work must be recognized, including the potential for unaccounted variables or interactions that could influence the acoustic descriptors, as well as reliance on specific models and methods (MPM and MLR) that may not detect all the nuances of the acoustic environment. To improve the reliability of the results, further studies involving larger sample sizes or different types of classrooms would be advisable to validate the findings. Furthermore, incorporating more advanced modeling techniques or considering additional variables, such as room geometry or occupants' positions, could provide a more comprehensive picture of the relationships between acoustic descriptors and building design factors. Lastly, the study's reliance on correlational analyses means that other unmeasured variables could potentially lead to confusion about the observed relationships, underscoring the need for additional research to explore these relationships.

4. Conclusions

In conclusion, this work involved an in-depth comparison of acoustic parameters using both MPM and MLR methods, revealing a high degree of convergence between the two methods. This analysis, guided and benchmarked by a review of the literature, provides a comprehensive understanding of the complex relationships between various acoustic descriptors and building design.

The correlation analyses of U50, STI, and C50, as discussed in the context of (CHOI, 2020) and validated in the current study, elucidate the intricate interaction between these predictors of speech intelligibility. The observed predominance of C50 in influencing both U50 and STI values, as described in Table 10, supports CHOI's (2020) quantitative findings, and underscores the pivotal role of room acoustic parameters in enhancing speech intelligibility.

As for the significance of controllable factors, it was concluded that the sound absorption, the factor (B), interferes more strongly in reverberation-related descriptors (groups 1 and 2), while the background noise, the factor (A), strongly interferes with the STI. The factor (C) was the one that interfered the least with the acoustic descriptors in general. Moreover, it was shown that D50 responds to the acoustic conditions of the classroom in the same way as EDT, Ts, and T30 do. The only difference is that RT decreases and D50 increases in response to an increased classroom sound absorption. Surprisingly, the combined 3rd and 4th order interactions had negligible effects in the classrooms of this study.

Acknowledgments

This study was partly financed by the Brazilian research funding agency CAPES (Federal Agency for the Support and Improvement of Higher Education) under the Finance Code 001.

References

1. American National Standard (2008), *Criteria for evaluating room noise*, Standard ANSI S12.2:2008.
2. ANSAY S., ZANNIN P.H.T. (2016), Using the parameters of definition, D50, and reverberation time, RT, to investigate the acoustic quality of classrooms, *Canadian Acoustics/Acoustique Canadienne*, **44**(4): 6–11.
3. BAYDIN A.G., PEARLMUTTER B.A., RADUL A.A., SISKIND J.M. (2017), Automatic differentiation in machine learning: A survey, *Journal of Machine Learning Research*, **18**(1): 5595–5637.
4. BERANEK L.L. (2006), Analysis of Sabine and Eyring equations and their application to concert hall audience and chair absorption, *The Journal of the Acoustical Society of America*, **120**(3): 1399–1410, doi: [10.1121/1.2221392](https://doi.org/10.1121/1.2221392).
5. BERANEK L.L., TAKAYUKI H. (1998), Sound absorption in concert halls by seats, occupied and unoccupied, and by the hall's interior surfaces, *The Journal of the Acoustical Society of America*, **104**(6): 3169–3177, doi: [10.1121/1.423957](https://doi.org/10.1121/1.423957).
6. BISTAFA S.R., BRADLEY J.S. (2000), Reverberation time and maximum background-noise level for classrooms from a comparative study of speech intelligibility metrics, *The Journal of the Acoustical Society of America*, **107**(2): 861–875, doi: [10.1121/1.428268](https://doi.org/10.1121/1.428268).
7. BISTAFA S.R., BRADLEY J.S. (2001), Predicting speech metrics in a simulated classroom with varied sound absorption, *The Journal of the Acoustical Society of America*, **109**(4): 1474–1482, doi: [10.1121/1.1354199](https://doi.org/10.1121/1.1354199).
8. BRADLEY J.S. (2011), Review of objective room acoustics measures and future needs, *Applied Acoustics*, **72**(10): 713–720, doi: [10.1016/j.apacoust.2011.04.004](https://doi.org/10.1016/j.apacoust.2011.04.004).
9. 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, doi: [10.1121/1.427932](https://doi.org/10.1121/1.427932).
10. 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, doi: [10.1121/1.1570439](https://doi.org/10.1121/1.1570439).
11. CHOI Y.-J. (2017a), Comparison of two types of combined measures, STI and U50, for predicting speech intelligibility in classrooms, *Archives of Acoustics*, **42**(3): 527–532, doi: [10.1515/aoa-2017-0056](https://doi.org/10.1515/aoa-2017-0056).
12. CHOI Y.-J. (2017b), Predicting classroom acoustical parameters for occupied conditions from unoccupied data, *Applied Acoustics*, **127**: 89–94, doi: [10.1016/j.apacoust.2017.05.036](https://doi.org/10.1016/j.apacoust.2017.05.036).
13. CHOI Y.-J. (2020), Evaluation of acoustical conditions for speech communication in active university classrooms, *Applied Acoustics*, **159**: 107089, doi: [10.1016/j.apacoust.2019.107089](https://doi.org/10.1016/j.apacoust.2019.107089).
14. CHRISTENSEN C.L., KOUTSOURIS G., RINDEL J.H. (2014), Estimating absorption of materials to match room model against existing room using a genetic algorithm, [in:] *Forum Acusticum 2014*, pp. 7–12, doi: [10.13140/2.1.1588.8647](https://doi.org/10.13140/2.1.1588.8647).
15. CROCE P., LECCESE F., SALVADORI G., BERARDI U. (2023), Proposal of a simplified tool for early acoustics design stage of classrooms in compliance with speech intelligibility thresholds, *Energies*, **16**(2): 813, doi: [10.3390/en16020813](https://doi.org/10.3390/en16020813).
16. DO NASCIMENTO E.O., DE OLIVEIRA L.N. (2016), Sensitivity analysis of cutting force on milling process using factorial experimental planning and artificial neural networks, [in:] *IEEE Latin America Transactions*, **14**(12): 4811–4820, doi: [10.1109/TLA.2016.7817015](https://doi.org/10.1109/TLA.2016.7817015).
17. DO NASCIMENTO E.O., ZANNIN P. (2023), Deep learning applied to speech transmission index prediction: Simulations and measurements, *Harvard Dataverse*, doi: [10.7910/DVN/RZRUTT](https://doi.org/10.7910/DVN/RZRUTT).
18. GEVREY M., DIMOPOULOS I., LEK S. (2003), Review and comparison of methods to study the contribution

- of variables in artificial neural network models, *Ecological Modelling*, **160**(3): 249–264, doi: [10.1016/S0304-3800\(02\)00257-0](https://doi.org/10.1016/S0304-3800(02)00257-0).
19. HOUTGAST T., STEENEKEN H.J.M., PLOMP R. (1980), Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics, *Acta Acustica united with Acustica*, **46**(1): 60–72.
 20. International Electrotechnical Commission (2011), *Sound system equipment. Part 16: Objective rating of speech intelligibility by speech transmission index* (IEC Standard No. 60268-16).
 21. International Organization for Standardization (2008), *Acoustics – Measurement of room acoustic parameters – Part 2: Reverberation time in ordinary rooms* (ISO Standard No. 3382-2), <https://www.iso.org/standard/36201.html>.
 22. JOLLIFFE I. (2011), Principal component analysis, [in:] *International Encyclopedia of Statistical Science*, Lovric M. [Ed.], Springer, Berlin, Heidelberg, doi: [10.1007/978-3-642-04898-2_455](https://doi.org/10.1007/978-3-642-04898-2_455).
 23. JOLLIFFE I.T., CADIMA J. (2016), Principal component analysis: A review and recent developments, *Philosophical Transactions of the Royal Society A*, **374**(2065): 20150202, doi: [10.1098/rsta.2015.0202](https://doi.org/10.1098/rsta.2015.0202).
 24. KANG S., MAK C.M., OU D., ZHOU X. (2023), Effects of speech intelligibility and reverberation time on the serial recall task in Chinese open-plan offices: A laboratory study, *Applied Acoustics*, **208**: 109378, doi: [10.1016/j.apacoust.2023.109378](https://doi.org/10.1016/j.apacoust.2023.109378).
 25. LECCESE F., ROCCA M., SALVADORI G. (2018), Fast estimation of Speech Transmission Index using the Reverberation Time: Comparison between predictive equations for educational rooms of different sizes, *Applied Acoustics*, **140**: 143–149, doi: [10.1016/j.apacoust.2018.05.019](https://doi.org/10.1016/j.apacoust.2018.05.019).
 26. LECUN Y., BENGIO Y., HINTON G. (2015), Deep learning, *Nature*, **521**: 436–444, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
 27. LECUN Y., BOTTOU L., BENGIO Y., HAFNER P. (1998), Gradient-based learning applied to document recognition, [in:] *Proceedings of IEEE*, **86**(11): 2278–2324, doi: [10.1109/5.726791](https://doi.org/10.1109/5.726791).
 28. LOCHNER J., BURGER J. (1960), Optimum reverberation time for speech rooms based on hearing characteristics, *Acta Acustica*, **10**: 394–399.
 29. MEYER E., KUNSTMANN D., KUTTRUFF H. (1964), On the measurements of sound absorption of audiences [in German: Über einige Messungen zur Schallabsorption von Publikum], *Acta Acustica*, **14**(2): 119–124.
 30. MIKULSKI W., RADOSZ J. (2011), Acoustics of classrooms in primary schools – Results of the reverberation time and the speech transmission index assessments in selected buildings, *Archives of Acoustics*, **36**(4): 777–793, doi: [10.2478/V10168-011-0052-6](https://doi.org/10.2478/V10168-011-0052-6).
 31. MINELLI G., PUGLISI G.E., ASTOLFI A. (2022), Acoustic parameters for learning in classroom: A review, *Building and Environment*, **208**: 108582, doi: [10.1016/j.buildenv.2021.108582](https://doi.org/10.1016/j.buildenv.2021.108582).
 32. MONTGOMERY D.C. (2012), *Design and Analysis of Experiments*, 8th ed., John Wiley & Sons, Inc.
 33. PIOTROWSKI A.P., NAPIORKOWSKI J.J. (2013), A comparison of methods to avoid overfitting in neural networks training in the case of catchment runoff modelling, *Journal of Hydrology*, **476**: 97–111, doi: [10.1016/j.jhydrol.2012.10.019](https://doi.org/10.1016/j.jhydrol.2012.10.019).
 34. PRODI N., PELLEGATTI M., VISENTIN C. (2022), Effects of type of early reflection, clarity of speech, reverberation and diffuse noise on the spatial perception of a speech source and its intelligibility, *The Journal of the Acoustical Society of America*, **151**: 3522–3534, doi: [10.1121/10.0011403](https://doi.org/10.1121/10.0011403).
 35. RENNIES J., SCHEPKER H., HOLUBE I., KOLLMEIR B. (2014), Listening effort and speech intelligibility in listening situations affected by noise and reverberation, *The Journal of the Acoustical Society of America*, **136**: 2642–53, doi: [10.1121/1.4897398](https://doi.org/10.1121/1.4897398).
 36. RINDEL J.H. (2012), ODEON APPLICATION NOTE – ISO 3382-3 Open plan offices. ODEON – Room Acoustic Software.
 37. RUSSELL S., NORVIG P. (1996), Artificial intelligence – A modern approach, [in:] *The Knowledge Engineering Review*, doi: [10.1017/S0269888900007724](https://doi.org/10.1017/S0269888900007724).
 38. SALA E., RANTALA L. (2016), Acoustics and activity noise in school classrooms in Finland, *Applied Acoustics*, **114**: 252–259, doi: [10.1016/j.apacoust.2016.08.009](https://doi.org/10.1016/j.apacoust.2016.08.009).
 39. SATO H., MORIMOTO M., MIYAGAWA Y., SUZUKI Y. (2016), Relationship between speech intelligibility and objective measures in sound fields with a discrete long-path echo, *The Journal of the Acoustical Society of America*, **140**(4): 3193, doi: [10.1121/1.4970046](https://doi.org/10.1121/1.4970046).
 40. SATO H., MORIMOTO M., WADA M. (2012), Relationship between listening difficulty rating and objective measures in reverberant and noisy sound fields for young adults and elderly persons, *The Journal of the Acoustical Society of America*, **131**: 4596–4605, doi: [10.1121/1.4714790](https://doi.org/10.1121/1.4714790).
 41. SATO H., NISHIKAWA Y., SATO H., MORIMOTO M. (2006), The relation between speech transmission index, clarity, and reverberation time and listening difficulty in the impulse response database of AIJ, *The Journal of the Acoustical Society of America*, **120**(5): 3321, doi: [10.1121/1.4781197](https://doi.org/10.1121/1.4781197).
 42. SECCHI S. *et al.* (2017), Effect of outdoor noise and façade sound insulation on indoor acoustic environment of Italian schools, *Applied Acoustics*, **126**: 120–130, doi: [10.1016/j.apacoust.2017.05.023](https://doi.org/10.1016/j.apacoust.2017.05.023).
 43. TANG S.K. (2008), Speech-related acoustic parameters in classrooms and their relationships, *Applied Acoustics*, **69**(12): 1318–1331, doi: [10.1016/j.apacoust.2007.08.008](https://doi.org/10.1016/j.apacoust.2007.08.008).
 44. 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, doi: [10.1016/j.buildenv.2018.03.020](https://doi.org/10.1016/j.buildenv.2018.03.020).

Research Paper

Ultrasound Imaging of Nonlinear Media Response
Using a Pressure-Dependent Nonlinearity IndexAndrzej NOWICKI¹, Juriy TASINKIEWICZ, Piotr KARWAT¹, Ihor TROTS^{1*},
Norbert ŻOLEK¹, Ryszard TYMKIEWICZ¹*Department of Ultrasound
Institute of Fundamental Technological Research
Polish Academy of Sciences
Warsaw, Poland**Corresponding Author e-mail: igortr@ippt.pan.pl*(received July 17, 2024; accepted August 26, 2024; published online October 10, 2024)*

It has been shown that within the range of acoustic pressures used in ultrasound imaging, waveforms are distorted during propagation in tissue due to the physically nonlinear behavior of the tissue. This distortion leads to changes in the spectrum of the received ultrasound echoes, causing the transfer of signal energy from the fundamental frequency to higher harmonics. Interestingly, adipose tissue exhibits up to 50 % stronger nonlinear behavior compared to other soft tissues. The tissue nonlinearity parameter B/A is typically measured *ex vivo* using an ultrasound method in transmission mode, which requires extensive receiving systems. Currently, there is no improved ultrasound method for measuring the B/A nonlinearity parameter *in vivo*, which could be used in assessing the degree of fatty liver disease.

We propose a new, simple approach to estimating nonlinear tissue properties. The proposed method involves transmitting ultrasound waves at significantly different acoustic pressures, recording echoes only in the fundamental frequency band at various depths, and introducing a nonlinearity index (NLI) based on specific echo amplitude ratios.

The NLI at a given depth is calculated using the ratio of two dimensionless parameters. The first parameter is a predetermined constant obtained by dividing the total echo values from transmitting a signal at higher sound pressure by those from a signal at lower sound pressure, summed over a small tissue sample volume located near the transducer. The second parameter is calculated at a fixed distance from the transducer, determined by dividing the total echo values from transmitting a signal at higher sound pressure by those from a signal at lower pressure, summed over a small tissue volume of the tissue at that distance from the transducer. The reliability of the proposed measurements for assessing tissue nonlinearity has been substantiated through experimental confirmation of the existing correlations between the values of NLI and B/A in water, sunflower oil, and animal liver tissue samples with oil-enriched regions. The NLI was more than 15 % higher in sunflower oil than in water. The NLI in bovine liver sample below the area with injected oil (mimicking “steatosis”) was more than 35 % higher than in regions without oil. This method represents a promising modality for the nonlinear characterization of tissue regions *in vivo*, particularly for diagnosing fatty liver disease.

Keywords: ultrasound imaging; abdominal ultrasound; nonlinear propagation; tissue harmonic imaging; nonlinearity index.



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

Ultrasound imaging has become a cornerstone of modern medicine, offering a safe, portable, and cost-effective method to visualize internal organs and structures. Modern ultrasonography relies on linear and

nonlinear properties (such as tissue harmonic imaging) of ultrasound wave propagation in tissue. By analyzing these nonlinearities, physicians can gather additional information about tissues, potentially providing deeper insights into physiological processes and disease states.

A broad overview of the current state of research and challenges in estimating the B/A nonlinearity parameter is presented in (PANFILOVA *et al.*, 2021). So far, with a very wide range of B/A measurements, none of the methods has been successfully implemented in real-time ultrasound. One significant challenge lies in the limited bandwidth of linear arrays and convex type of ultrasonic transducers, which does not exceed 60 %–70 %. This narrow bandwidth significantly limits the reception of second harmonic echoes, consequently limiting the ability to estimate the B/A parameter in real time along the propagation path. The first attempts to reconstruct B/A profiles were made using the pump wave method (ICHIDA *et al.*, 1983; 1984). Another approach, using parametric array tomography was described in (GONG *et al.*, 2004; WANG *et al.*, 2003). The finite amplitude method, called FAM (GONG *et al.*, 2004; AKIYAMA, 2000; TOULEMONDE *et al.*, 2015) might be considered as an inspiration for our method. However, to the best of our knowledge, none of the B/A estimation/reconstruction techniques have yet been applied to real-time US imaging.

The phenomenon of nonlinear propagation of ultrasonic waves is used in tissue harmonic imaging (THI) because it increases the resolution of ultrasound images of examined organs (VARRAY *et al.*, 2010; VAN WIJK, THIJSEN, 2002). Several methods have been developed to measure this nonlinear echo, including amplitude modulation, pulse inversion, and second harmonic inversion (SIMPSON *et al.*, 1999). However, THI is far from being optimal in the sense that only half of the available transducer bandwidth is used for image formation – the lower half for transmission and the upper half during reception. The importance of reduced dynamic range and penetration encountered in THI was also pointed out (AVERKIOU *et al.*, 1997; AVERKIOU, 2001). Since images are formed with only the first harmonic components, which are usually at least 20 dB below the fundamental, the dynamic range is limited. Additionally, COILA and OELZE (2020) conducted a study on the influence of nonlinear propagation in tissue on the estimation of attenuation/absorption in specific tissue regions.

The linearized pressure-density equation of state has a form:

$$p = P - P_0 = (\rho - \rho_0) \left(\frac{\partial P}{\partial \rho} \right)_{\rho_0}, \quad (1)$$

where P is the total pressure, which is the sum of the equilibrium pressure P_0 and the acoustic pressure p ; ρ_0 and ρ are the equilibrium density and the small increase in density produced by the sound, respectively. Equation (1) is applicable when the speed of sound (c) is much greater than the local flow velocity (v), $c \gg v$, i.e., when the Mach number (M) is much less than one, $M = v/c \ll 1$.

In the context of ultrasonic pressures, typically ranging from hundreds of kilopascals to several megapascals, tissue volume experiences compression and stretching, leading to pressure-dependent changes in the speed of sound.

At sufficiently high pressure magnitudes, wave propagation ceases to be linear, necessitating the inclusion of additional nonlinear terms to the fluid equation of state. These terms account for both the tissue elasticity (A) in the linear regime of density changes and the higher-order elasticity coefficient (B) as the first correction taking into account nonlinear (quadratic) changes in density (HAMILTON, BLACKSTOCK, 2008).

Consequently, the pressure-density relation $p = f(\rho)$ can be approximated by a Taylor series expansion of the adiabatic equation of state:

$$\begin{aligned} p &= (\rho - \rho_0) \left(\frac{\partial P}{\partial \rho} \right)_{\rho_0} + \frac{(\rho - \rho_0)^2}{2} \left(\frac{\partial^2 P}{\partial \rho^2} \right)_{\rho_0} + \dots \\ &= A \left(\frac{\rho - \rho_0}{\rho_0} \right) + \frac{B}{2} \left(\frac{\rho - \rho_0}{\rho_0} \right)^2 + \dots \end{aligned} \quad (2)$$

After truncating Eq. (2) to the second-order term, the ratio of B/A and the nonlinearity coefficient of the medium $\beta = 1 + (1/2)B/A$ can be estimated, where $A = \rho_0 (\partial P / \partial \rho)_{\rho=\rho_0}$, and $B = \rho_0^2 (\partial^2 P / \partial \rho^2)_{\rho=\rho_0}$. The parameter A corresponds to the elasticity coefficient of increases and hence the speed of sound propagation increases. The nonlinear parameter B/A carries information about the distortion of the propagating wave, resulting in the transfer of some energy to the second and higher harmonics.

In the domain of linear acoustics, specifically in a lossless linear medium, the neglect of wave distortions and the emergence of higher harmonics result in $B/A = 0$. However, in real materials, B/A assumes finite values. Documented B/A values for various fluids and biological media were outlined in (DUCK, 2002): for water at 20 °C and 40 °C, $B/A = 4.96$ and 5.38, respectively; for 3.5 % saline at 20 °C, $B/A = 5.25$; for blood plasma at 30 °C, $B/A = 5.74$; for whole blood (26 °C) $B/A = 6.1$; for nonfat soft tissues $B/A = 6.3$ –8.0; and for fatty soft tissues $B/A = 9.6$ –11.3 (VARRAY *et al.*, 2010). DONG *et al.* (1999) provided the ultrasonic nonlinearity parameter B/A for nine versions of water-based, macroscopically uniform ultrasonically tissue-mimicking (TM) nonfat and fat materials. Notably, the B/A parameter is 1.5 to 2 times greater in adipose compared to other tissues, thereby partially elucidating the advantages of utilizing harmonic imaging in challenging cases involving excess body fat.

In what follows, we describe a preliminary investigation focused on evaluating the nonlinearity of the medium by comparing echoes from selected areas subjected to various scanning signals with differing amplitudes. In a linear medium, the magnitude of the echoes should be directly proportional to the amplitude of the

transmitted signals. Any deviation from the linear relationship between the transmitted signal's amplitude and the backscattered amplitude is contingent upon the physically nonlinear properties of the tissue area under scrutiny.

The rest of the paper is organized as follows: the next section presents a brief overview of the proposed method, the results obtained using proposed method are presented in Sec. 3, and the discussion is presented in Sec. 4.

2. Materials and methods

In the linear model of ultrasound propagation, the echoes from reflectors or biological tissue for different sound pressures measured at the same depth should vary proportionally to the transmitted sound pressures. However, in the case of nonlinear wave propagation, part of the acoustic energy is transferred to higher harmonics. This transfer increases with increasing acoustic pressure, resulting in a real decrease in echo amplitudes for the first harmonic.

We propose a new, straightforward approach to quantifying the nonlinear properties of tissue by analyzing the ratio of the energy of scattered echoes from a tissue/medium with different nonlinear ultrasound propagation characteristics using different acoustic pressures during transmission.

This new approach involves using several consecutive wave transmissions of identical waveform but with significantly different acoustic pressures (varying by several times), along with successive recordings of ultrasound images in the baseband of the head (only the first harmonic is recorded). Images of "linear" tissues will differ only in amplitude, which will be proportional to the amplitude of the transmitted wave. Therefore, the ratios of the echoes' amplitudes recorded for the sequence of low- and high-pressure transmission (after compensating for different transmit pressures) should be close to one. If there are areas in the imaging space with different B/A nonlinearity ratio, the amplitude ratios will differ from one and this value should increase with increasing nonlinearity coefficient of the imaged tissue.

The Verasonics Vantage 256 (Verasonics, USA) and us4R-lite ultrasound research system (us4us ltd., Poland) with a convex probe (ATL C4-2) were used for the measurements. The research was carried out in three stages: hydrophone measurements in water and sunflower oil, reflection measurements in tissue *in vitro*. First two steps were performed the using us4R-lite system, while the measurements in the liver sample were conducted with the Verasonics Vantage 256. In all three experiments, all transducer elements were activated simultaneously. For this purpose, short pulses (two sine cycles) at a nominal fre-

quency of 3.125 MHz and a sampling rate of 62.5 MHz were generated, with amplitudes corresponding to various sound pressures (see discussion below).

In the first step of the evaluation, three pulses with different driving voltages were successively transmitted into water. The corresponding peak-to-peak pressure amplitudes were measured with a needle hydrophone with a sensor diameter of 0.075 mm (Precision Acoustics, UK) at a distance of 0.5 cm from the face of the transmitting transducer. The measured amplitudes P_1 , P_2 , and P_3 were equal to 0.19 MPa, 0.39 MPa, and 1.55 MPa, respectively. Then, for each of these transmission pressures, the amplitudes of the first and higher harmonics were measured in both water and sunflower oil. The hydrophone was positioned at depths of 1 cm, 2 cm, 3 cm, 4 cm, 5 cm, and 6 cm from the transducer.

In the second step of our research, we used the same transmission pressures, to measure the echoes' amplitudes from a thread phantom made of thin (0.2 mm) nylon threads spaced 1 cm apart, immersed in both water and sunflower oil.

In the last step, we performed backscattered ultrasound measurements in a fresh beef liver sample. At a depth of 4 cm from the surface of the sample, 1 cm³ of sunflower oil was injected into the sample to mimic fatty tissue. Then, we determined the average amplitude of echoes in selected areas along the radiation axis of the examined tissue. We used regions of interest (ROIs) in the shape of a scan stripe with an arbitrarily selected width of 0.5 cm (10 image lines) and 25 RF samples along each line corresponding to 0.5 cm in depth (Fig. 1). The averaged echoes' amplitudes were calculated for each of the applied sound pressures P_1 – P_3 .

Finally, the echo amplitude values were normalized, i.e., divided by the transmitted pressures, and these values are denoted in this article as E_{P1} – E_{P3} . Theoretically, for purely linear propagation, such normalized backscattered echoes' amplitudes E should be the same across all applied pressures throughout the penetration depth. The pressures reported in the experiments are given as peak-to-peak values. The corresponding mechanical indices (MI) calculated for peak negative pressures at a nominal frequency of 3.125 MHz are given in Table 1. For all applied pressures, the MI remains below the FDA regulated limits for diagnostic ultrasound, $MI < 1.9$.

Table 1. MI values for pressures used in the experiments.

Peak-to-peak pressure [MPa]	Peak negative pressure [MPa]	MI
0.19	0.096	0.049
0.39	0.196	0.1
1.55	0.68	0.345

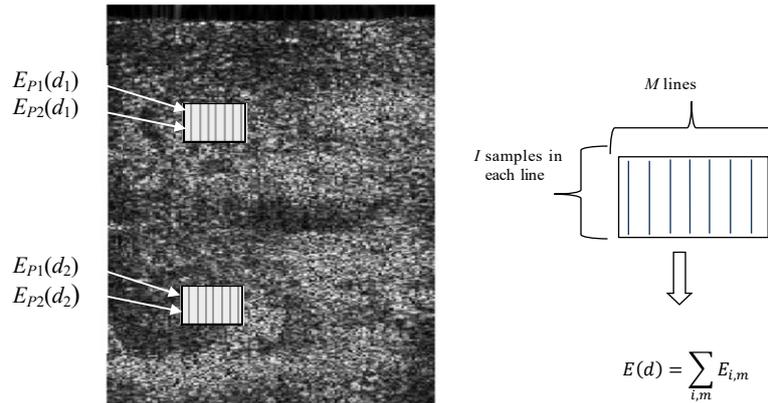


Fig. 1. Principle of calculating the average amplitude of echoes $E_{P_1}(d_1)$, $E_{P_2}(d_1)$ and $E_{P_1}(d_2)$, $E_{P_2}(d_2)$ at different depths d_1 and d_2 of the examined tissue and for two different pressures P_1 and P_2 of the transmitted ultrasonic wave. The echo amplitude $E(d)$ in each region is calculated as the sum of amplitudes of I samples in M lines.

We introduce the nonlinearity index (NLI) in the form of the quotient of two average echo amplitudes $E_{P_i}(d)$ and $E_{P_j}(d)$ measured at depth d for two different transmitted acoustic pressures P_i (lower pressure) and P_j (higher pressure):

$$\text{NLI}(d; P_i, P_j) = \text{NR} \frac{E_{P_i}(d)}{E_{P_j}(d)}, \quad (P_j > P_i). \quad (3)$$

The ratio of echo amplitudes $E_{P_i}(d)$ and $E_{P_j}(d)$ for both transmitted pressures is normalized by the factor NR, which is the ratio of the echo amplitudes E_{P_j} (for the higher pressure P_j) and E_{P_i} (for the lower pressure P_i) measured 0.5 mm below the transducer face. In our experiments, NR was equal to 9.3 for the pressures of 1.55 MPa and 0.19 MPa, and 4.4 for the pressures of 1.55 MPa and 0.19 MPa, respectively.

The instantaneous amplitude of the ultrasound echo signals was obtained using Hilbert transform-based envelope detection of the ultrasound radiofrequency (RF) signals (HAHN, 1996).

3. Results

As stated in the previous section, the measurements were carried out in three stages. First, we determined the intrinsic first harmonic levels in water and sunflower oil using a needle hydrophone placed at various depths from the face of the scanning head, ranging from 1 cm to 6 cm. Then, the amplitudes of backscattered echoes from thin threads immersed in both water and sunflower oil were measured. In the last stage of the research, the NLI was determined in areas of the beef liver sample with the “steatosis” area introduced by injecting 1 ml of sunflower oil.

3.1. Hydrophone axial field measurements in water and sunflower oil

The Fourier spectra (up to the third harmonic) of the acoustic pressures measured with a needle hydrophone in water and sunflower oil at depths of 10 mm and 60 mm are shown in Figs. 2a–b and Figs. 3a–b,

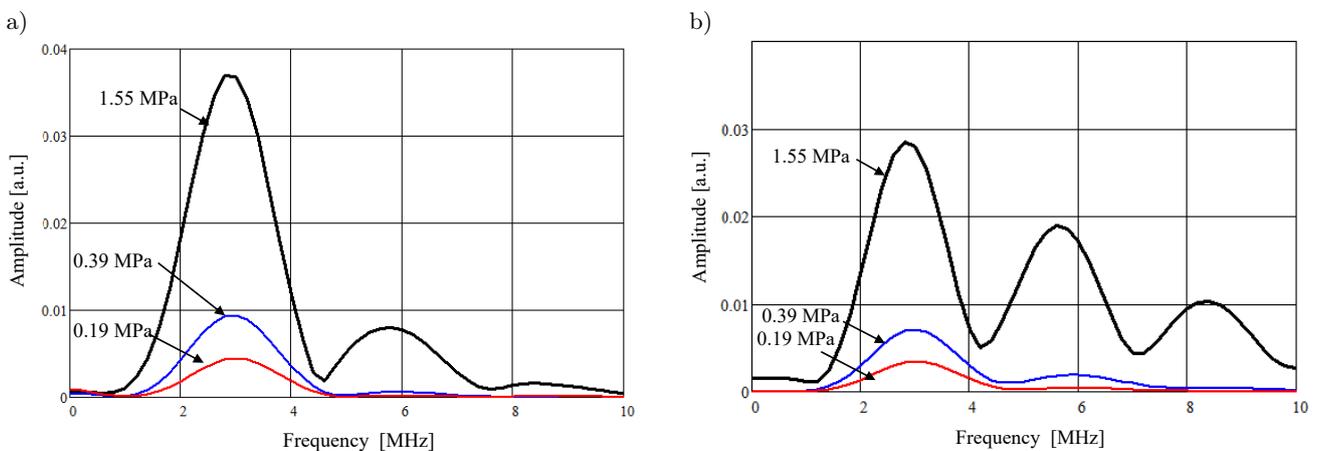


Fig. 2. Amplitudes of the first three harmonics 1 through 3 measured in water with a needle hydrophone (along the radiation axis) at depths of 1 cm (a) and 6 cm (b) from the transducer face. The transmitted pressures ranged from 0.19 MPa to 1.55 MPa.

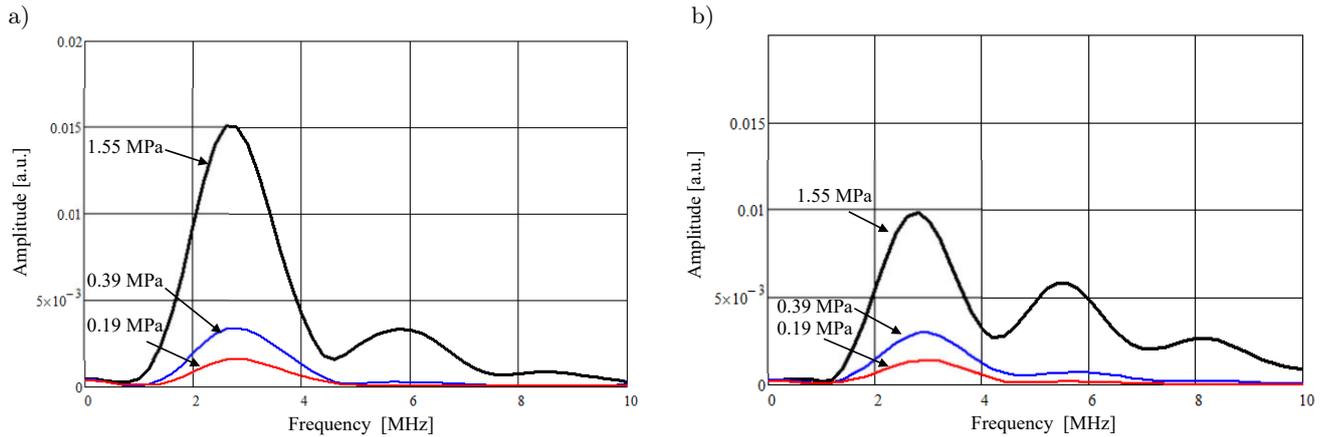


Fig. 3. Amplitudes of the first three harmonics measured in sunflower oil with a needle hydrophone (along the radiation axis) at depths of 1 cm (a) and 6 cm (b) from the transducer face. The transmitted pressures ranged from 0.19 MPa to 1.55 MPa.

respectively. The transmitted pressure amplitudes measured at a depth of 5 mm were recorded at values of 0.19 MPa, 0.39 MPa, and 1.55 MPa.

In water, at a depth of 1 cm, the amplitude of the first harmonic is 0.0382 (in arbitrary units – a.u.) for a pressure of 1.55 MPa, and at a depth of 6 cm, its value is 0.0285. For a pressure of 0.39 MPa, the amplitudes of the first harmonics are 0.0094 and 0.0071 at depths of 1 cm and 6 cm, respectively. For a pressure of 0.19 MPa, the amplitudes of the first harmonics are 0.0041 and 0.0037 at depths of 1 cm and 6 cm, respectively.

In sunflower oil, at a depth of 1 cm, the amplitude of the first harmonic for a pressure of 1.55 MPa is 0.015 (in a.u.), while at a depth of 6 cm its value is 0.0098. For a pressure of 390 kPa, the amplitudes of the first harmonics are 0.0034 and 0.0029 at depths of 1 cm and 6 cm, and for a pressure of 190 kPa, the amplitudes of the first harmonics are 0.0016 and 0.0014 at depths of 1 cm and 6 cm, respectively.

Figure 4 shows the dependence of NLIs defined by Eq. (3) on depth in both water and sunflower oil based on hydrophone data.

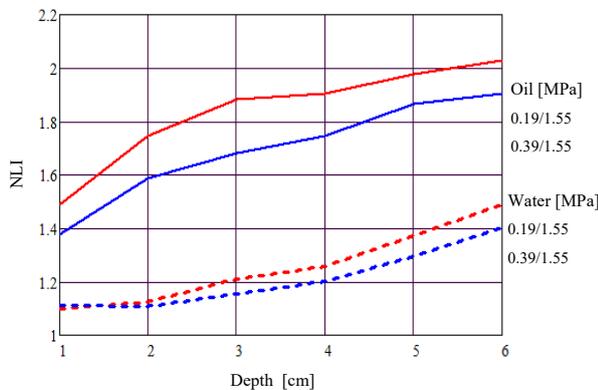


Fig. 4. Plots of NLIs in water and oil based on hydrophone data for pressures of 0.39 MPa and 0.19 MPa, using a reference pressure of 1.55 MPa (see Figs. 2 and 3).

The amplitude of the first harmonic for the reference pressure of 1.55 MPa decreases with increasing depth faster in sunflower oil than in water, particularly when comparing to the first harmonic at lower pressures of 0.39 MPa and 0.19 MPa. This is consistent with the fact that the nonlinearity coefficient B/A is larger for oil than for water (water ≈ 5 , sunflower oil > 8). As a result, starting from a depth of 2 cm, the NLI increases up to 2.02 in oil and 1.5 in water, at a depth of 6 cm for the applied pressures of 1.55 MPa and 0.19 MPa. This shows that higher values of the NLI correspond to higher values of B/A parameter, describing the nonlinear propagation characteristics of the medium. The corresponding NLI values for the reference pressures of 1.55 MPa and 0.39 MPa in oil and water were about 20 % lower, i.e., 1.9 and 1.4, respectively.

3.2. Thread phantom experiment

The NLI for echoes from the threads placed at depths of 1 cm, 2 cm, 3 cm, 4 cm, 5 cm, and 6 cm from the transducer head in both water and sunflower oil, using transmitted pressures of 190 kPa, 0.39 MPa, and a reference pressure of 1.55 Pa, is shown in Fig. 5.

The NLI is approximately 15 % greater in oil than in water for depths below 2 cm. However, the value of NLI in oil flattens out above 5 cm due to much greater attenuation – over 60 times greater than in water at 3 MHz (0.07 dB/cm for water and close to 4.5 dB/cm for oil). In practice, attenuation in water can be neglected.

3.3. In vitro measurements

A *B*-mode image of the fresh beef liver sample with three marked areas for NLI determination: area 1 – 2.5 cm from the surface; area 2 – 4.5 cm from the surface; area 3 – 6.5 cm from the surface, is shown in Fig. 6.

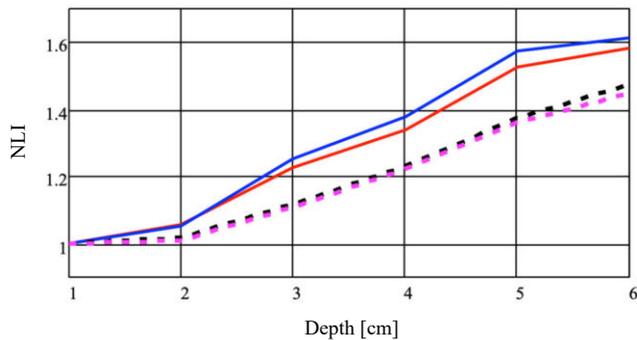


Fig. 5. NLIs for reflections from the nylon threads in oil and water. Threads are placed every 1 cm. The blue and red solid lines correspond to the NLI for sunflower oil at pressures of 0.19 MPa/1.55 MPa (solid blue line) and 0.39 MPa/1.55 MPa (solid red line). The blue and red dashed lines represent the NLI in water for pressures of 0.19 MPa/1.55 MPa (blue dashed line) and 0.39 MPa/1.55 MPa (red dashed line).

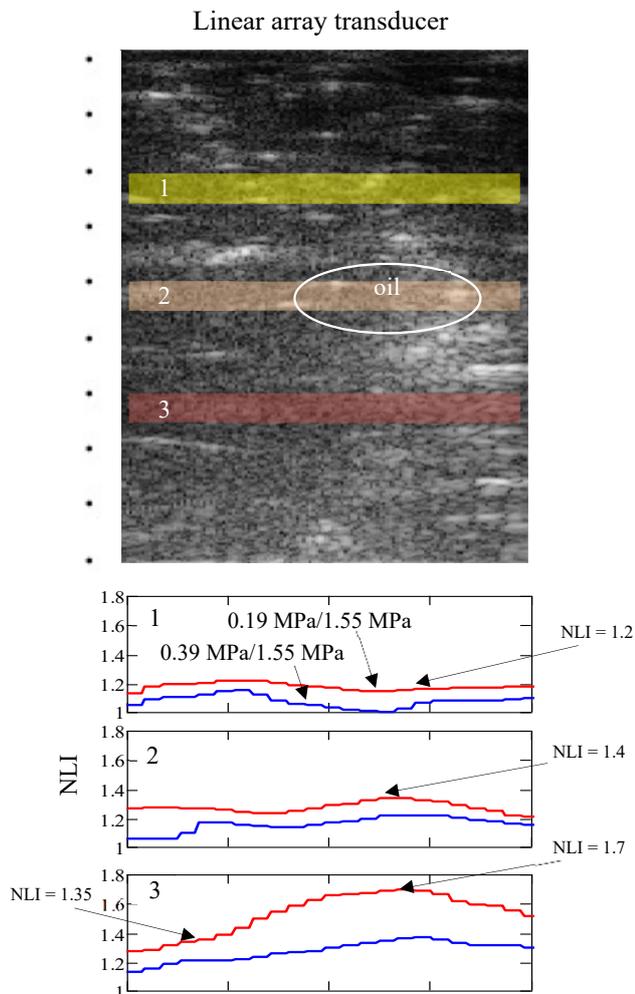


Fig. 6. *B*-mode image of a beef liver sample with three areas marked for NLI determination. The red lines correspond to pressures 0.19 MPa/1.5 MPa, while the blue lines correspond to pressures 0.39 MPa/1.5 MPa.

As shown in Fig 6, there is a significant increase in NLI to approximately 1.7 in area 3, at a depth of 5.5 cm, which is approximately 1.5 cm below the region where sunflower oil was injected. The results presented in Fig. 6 show that the NLI strongly depends on “steatosis” within the tissue. In the case of normal tissue (on the left side of the *B*-mode image of the sample) aside of the region with the injected oil, the NLI at a depth of about 6 cm does not exceed the value of 1.35, while it undergoes an increase of 20 % under the oil-injected region.

4. Discussion and conclusion

The nonlinear properties of the tissue cause the transfer of energy from the transmitted wave not only to the second harmonic component but also to the third, fourth, and higher harmonics. However, this energy transfer is ignored in THI. In the proposed technique for determining the nonlinear coefficient of the tested medium, the losses of the signal transferred to all higher harmonics are automatically taken into account – the amplitude of the recorded echoes at the first harmonic is smaller than the actual amplitude of the returning signals reflecting the energy loss to the second, third and higher frequency components. Moreover, since the proposed technique does not require the recording of the second and higher harmonics, the entire transducer bandwidth can be used for both transmission and reception.

In our preliminary experiments, we demonstrated that, both in water and sunflower oil, the relative decrease in echo amplitude with penetration depth is greater at higher initial pressures of the transmitted wave. Specifically, the quotient of the amplitude of the first harmonic for low and high transmit acoustic pressures increases faster in oil than in water. This result is consistent with what was expected, given that the nonlinearity coefficient B/A is over 1.5 times higher in oil than in water (water ≈ 5 , sunflower oil > 8).

Furthermore, the NLI at depths exceeding 2 cm in oil is approximately 15 % larger than in water. However, the NLI growth curve in oil begins to flatten out slightly beyond 5 cm (Fig. 5). This phenomenon is attributed to much greater attenuation of ultrasound in oil – over 60 times than in water. Specifically, at 3 MHz, the attenuation is 0.07 dB/cm for water and close to 4.5 dB/cm for oil. Practically, attenuation in water can be practically ignored.

The preliminary in vitro experiment using a beef liver sample, where a small region of the liver tissue was modified by injecting 1 ml of sunflower oil, confirmed the ability to assess the “steatosis” mimicking area by measuring the NLI.

The introduced NLI strongly depends on “steatosis” of the media where nonlinearity accumulation accelerates, and reaches maximum below this region. This is

confirmed in Fig. 6 where below the steatotic tissue (on the right side of the sample, the region with injected oil), the NLI at the depth of 6 cm reached 1.7 which is almost 35 % higher than the NLI value of 1.3 observed below normal tissue (on the left side of the sample).

The proposed method currently serves as a qualitative one and only allows the identification of areas in the examined tissue with smaller or larger nonlinear properties. At this stage, it does not allow for a quantitative correlation of the proposed NLI with the commonly established B/A nonlinearity parameter. However, the B/A ratio can only be determined in tissues prepared in vitro, while our method is suitable for direct application in vivo.

The nonlinearity parameters for healthy liver and fatty liver tissues are (according to the existing literature) 7 and over 10, respectively (ZHANG, 2001). We expect that the proposed NLI index could significantly help in the assessment of pathological changes in the examined tissue.

There is a number of key issues that we will address in our future research. First, we will extend the depth of NLI analysis (depth limitation is currently limited due to attenuation) and we will validate the proposed methodology using focused transmission techniques. We also plan to modify the presented approach for a more localized assessment of medium nonlinearity. Finally, our goal is to develop a methodology that will enable quantitative imaging of tissue nonlinear properties based on the introduced NLI.

References

- AKIYAMA I. (2000), Reflection mode measurement of nonlinearity parameter B/A , [in:] *AIP Conference Proceedings*, **524**: 321–324, doi: [10.1063/1.1309232](https://doi.org/10.1063/1.1309232).
- AVERKIOU M.A. (2001), Tissue harmonic ultrasonic imaging, *Comptes Rendus de l'Académie des Sciences*, **2**: 1139–1151.
- AVERKIOU M.A., ROUNDHILL D.R., POWERS J.E. (1997), A new imaging technique based on the nonlinear properties of tissues, [in:] *1997 IEEE Ultrasonics Symposium Proceedings. An International Symposium*, **2**: 1561–1566, doi: [10.1109/ULTSYM.1997.663294](https://doi.org/10.1109/ULTSYM.1997.663294).
- COILA A., OELZE M.L. (2020), Effects of acoustic nonlinearity on pulse-echo attenuation coefficient estimation from tissue-mimicking phantoms, *The Journal of the Acoustical Society of America*, **148**(2): 805–814, doi: [10.1121/10.0001690](https://doi.org/10.1121/10.0001690).
- DONG F., MADSEN E.L., MACDONALD M.C., ZAGZEBSKI J.A. (1999), Nonlinearity parameter for tissue-mimicking materials, *Ultrasound in Medicine & Biology*, **25**(5): 831–838, doi: [10.1016/s0301-5629\(99\)0016-2](https://doi.org/10.1016/s0301-5629(99)0016-2).
- DUCK F.A. (2002), Nonlinear acoustics in diagnostic ultrasound, *Ultrasound in Medicine & Biology*, **28**(1): 1–18, doi: [10.1016/S0301-5629\(01\)00463-X](https://doi.org/10.1016/S0301-5629(01)00463-X).
- GONG X., ZHANG D., LIU J., WANG H., YAN Y., XU X. (2004), Study of acoustic nonlinearity parameter imaging methods in reflection mode for biological tissues, *The Journal of the Acoustical Society of America*, **116**(3): 1819–1825, doi: [10.1121/1.1781709](https://doi.org/10.1121/1.1781709).
- HAHN S.L. (1996), *Hilbert Transforms in Signal Processing*, Artech House Publishers.
- HAMILTON M.F., BLACKSTOCK D.T. [Eds.] (2008), *Nonlinear Acoustics*, Acoustical Society of America.
- ICHIDA N., SATO T., LINZER M. (1983), Imaging the nonlinear ultrasonic parameter of a medium, *Ultrasonic Imaging*, **5**(4): 295–299.
- ICHIDA N., SATO T., MIWA H., MURAKAMI K. (1984), Real-time nonlinear parameter tomography using impulsive pumping waves, [in:] *IEEE Transactions on Sonics and Ultrasonics*, **31**(5): 635–641, doi: [10.1109/T-SU.1984.31548](https://doi.org/10.1109/T-SU.1984.31548).
- PANFILOVA A., VAN SLOUN R.J.G., WIJKSTRA H., SAPOZHNIKOV O.A., MISCHI M. (2021), A review on B/A measurement methods with a clinical perspective, *The Journal of the Acoustical Society of America*, **149**(4): 2200, doi: [10.1121/10.0003627](https://doi.org/10.1121/10.0003627).
- SIMPSON D.H., CHIEN T.C., BURNS P.N. (1999), Pulse inversion Doppler: A new method for detecting nonlinear echoes from microbubble contrast agents, [in:] *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, **46**(2): 372–382, doi: [10.1109/58.753026](https://doi.org/10.1109/58.753026).
- TOULEMONDE M., VARRAY F., BASSET O., CACHARD C. (2015), Nonlinearity parameter B/A of biological tissue ultrasound imaging in echo mode, [in:] *AIP Conference Proceedings*, **1685**(1): 040016, doi: [10.1063/1.4934411](https://doi.org/10.1063/1.4934411).
- VAN WIJK M.C., THIJSSSEN J.M. (2002), Performance testing of medical ultrasound equipment: Fundamental vs. harmonic mode, *Ultrasonics*, **40**(1–8): 585–591, doi: [10.1016/S0041-624X\(02\)00177-4](https://doi.org/10.1016/S0041-624X(02)00177-4).
- VARRAY F., CACHARD C., TORTOLI P., BASSET O. (2010), Nonlinear radio frequency image simulation for harmonic imaging: Creanuis, [in:] *2010 IEEE International Ultrasonics Symposium*, pp. 2179–2182, doi: [10.1109/ULTSYM.2010.5935538](https://doi.org/10.1109/ULTSYM.2010.5935538).
- WANG H., ZHU X., GONG X., ZHANG D. (2003), Computed tomography of the acoustic nonlinearity parameter B/A for biological tissues via difference frequency wave from a parametric array in reflection mode, *Chinese Science Bulletin*, **48**: 2427–2430, doi: [10.1360/03ww0065](https://doi.org/10.1360/03ww0065).
- ZHANG D., GONG X.F., CHEN X. (2001), Experimental imaging of the acoustic nonlinearity parameter B/A for biological tissues via a parametric array, *Ultrasound in Medicine & Biology*, **27**(10): 1359–1365, doi: [10.1016/S0301-5629\(01\)00432-X](https://doi.org/10.1016/S0301-5629(01)00432-X).

Research Paper

Time-Domain Analysis of Echoes from Solid Spheres and Spherical Shells with Separated Transmit-Receive ConfigurationsZhongkai WANG⁽¹⁾, Zilong PENG^{(1)*}, Fulin ZHOU⁽²⁾, Liwen TAN⁽¹⁾

⁽¹⁾ *School of Energy and Power
Jiangsu University of Science and Technology
Zhenjiang, China*

⁽²⁾ *State Key Laboratory of Ocean Engineering, Collaborative Innovation Center
for Advanced Ship and Deep-Sea Exploration
Shanghai Jiao Tong University
Shanghai, China*

*Corresponding Author e-mail: zlp_just@sina.com

(received February 18, 2024; accepted April 26, 2024; published online August 19, 2024)

The complexity of bistatic echo pulse sequences surpasses that of monostatic echo pulse sequences. Based on the scattering acoustic field of elastic spheres and spherical shells, a method is employed to calculate the time-domain echoes of solid spheres and spherical shells with transceiver separation under the condition of plane wave incidence. This is achieved by constructing the incident signal and performing a multiplication operation in the frequency domain with the target scattering acoustic field. Employing the contour integral method, we derive phase velocity and group velocity dispersion curves for circumferential waves propagating around these structures. Furthermore, under the assumption of plane wave incidence, we analyze the propagation paths of Rayleigh echoes for solid spheres and anti-symmetric Lamb waves for spherical shells. Estimation formulas for the arrival times of separated transmit-receive echoes are provided for both solid spheres and spherical shells. Our findings indicate that bistatic waves can be classified into clockwise and counterclockwise circulation patterns around the surfaces of these structures. Through a comparison with the time-angle spectrum of echoes, we demonstrate the accuracy of the proposed estimation formulas for echo arrival times. This study offers valuable insights for the identification of underwater targets.

Keywords: dispersion curves; time-domain echoes; bistatic configuration.



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

In the context of underwater detection and target identification, when acoustic waves impinge upon the surface of a target, they not only generate geometric reflection waves but also transmit through the target. Within the elastic body, various re-radiated elastic echoes are induced. These echoes carry characteristic information of the target, aiding active sonar systems in the detection and identification of targets (XIA *et al.*, 2016; TOO *et al.*, 2014). A considerable amount of research has been conducted globally on the recognition and formation mechanisms of elastic wave com-

ponents in submerged elastic shell targets, particularly in the domains of ultrasonics, underwater engineering for damage detection (crack detection), and the identification of submerged and seafloor targets (BEDNARZ, 2017; APOSTOLODIA *et al.*, 2007; KARGL *et al.*, 2012; QIAO *et al.*, 2016). However, there has been limited research on separated transmit-receive configurations. With the current trend in sonar detection moving towards multi-platform sonar joint detection, the investigation of target identification through the separated transmit-receive configuration, leveraging the scattering characteristics of elastic waves, holds significant importance.

Research on the acoustic scattering of spheres and spherical shells in free fields has been widely conducted both theoretically and experimentally worldwide (GAUNAURD, WERBY, 1987; 1991; AYRES *et al.*, 1987; MARSTON, SUN, 1992). GAUNAURD and ÜBERALL (1983; 1985) analyzed the circumferential waves of rigid spheres using singular expansion methods (SEM) and Watson transform techniques, and delved deeper into the scattering processes of solid spheres based on resonance scattering theory. ÜBERALL *et al.* (1982) established a direct connection between surface waves and complex frequency poles. FAN *et al.* (2012) applied the Sommerfeld–Watson Transformation (SWT) method to the problem of elastic spherical shells filled with water, employing contour integration to solve dispersion equations in the frequency and wavenumber domains, indicating the generation of a significant amount of fluid-added waves due to internal fluid loading. DING *et al.* (2023) proposed an acoustic encoding principle and method based on high-frequency time-domain echoes of stratified elastic spherical shells in water, studying the influence of shell thickness, material properties of each layer, and arrangement order on the characteristics of time-domain echoes. DIERCKS and HICKLING (1967) demonstrated through experiments with vacuum spheres that the target echo is related to the receiving point. ANDERSON (2012) utilized SPWV time-frequency analysis methods and ray theory analysis to show that in the presence of comprehensive effects caused by changes in circumferential wave paths and circumferential wave damping coefficients due to transmit-receive separation, time and frequency shifts occur in the echoes of free-field elastic spherical shells. THOMPSON (2023) investigated the application of time-frequency methods in detecting and identifying target echoes underwater.

FAWCETT (2015) proposed a method for solving the elastic scattering of near-spherical targets, studying the influence of shell thickness on polar angle transformation and showing changes in the mid-frequency enhancement region of near-spherical targets compared to shells with a constant radius. GUNDERSON *et al.* (2017) obtained forecast formulas for Rayleigh wave interference trajectories of solid spheres and solid cylinders by analyzing the path difference of circumferential waves along their surfaces, both clockwise and counterclockwise, which fit well with the angular frequency spectrum interference fringes of the targets. YU *et al.* (2014) proposed a subsonic antisymmetric Lamb wave separation method, which can identify subsonic antisymmetric Lamb waves even when the incident short pulse frequency is far from the maximum enhancement frequency. SU *et al.* (2017) proposed a signal processing method under low-frequency broadband long-pulse excitation, allowing the observation and analysis of weak elastic wave energy in strong specular reflection waves. LI and WU (2019) filtered and modulated tar-

get echoes, separating target elastic scattering components. The above studies did not analyze the mechanisms of bistatic echo paths, echo moments, and the information carried by echoes of solid spheres and shells.

This paper firstly establishes a theoretical analytical computational model for solid spheres and spherical shells. By multiplying the spectrum of the constructed sinusoidal pulse signal with the computed acoustic transfer functions of the solid sphere and the spherical shell, the multi-bounce echo spectra of the solid sphere and the spherical shell are obtained. Then, an inverse Fourier transform obtains the time-domain echo pulse sequence. The results reveal that both the bistatic echoes of solid spheres and spherical shells exhibit an X pattern, formed by waves circumnavigating the target in both clockwise and counterclockwise directions. By contrasting the reverse echoes of solid spheres and spherical shells, it is observed that spherical shell reverse echoes exhibit distinct wave packets, which can serve as a reference for the identification of solid spheres and spherical shells. Finally, formulas estimating the arrival times of Rayleigh waves in solid spheres and anti-symmetric waves in spherical shells are provided.

2. Theoretical research

2.1. The theoretical solution of elastic spheres

Harmonic plane sound waves with unit amplitude scatter from the elastic sphere as shown in Fig. 1. Assuming the incident direction is aligned with the z -axis, the backward scattering wave corresponds to $\theta = \pi$. Both the incident and scattered waves are independent of the azimuthal angle φ and symmetric about the z -axis.

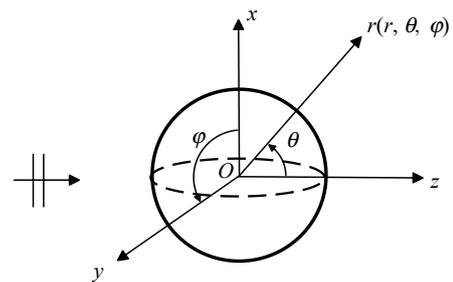


Fig. 1. Plane waves scattering from an elastic sphere.

The expression for the scattered acoustic field of the medium is given by (TANG *et al.*, 2018):

$$p_s = \sum_{n=0}^{\infty} i^n (2n+1) b_n h_n^{(1)}(kr) P_n(\cos \theta), \quad (1)$$

where θ is the receiving angle, \mathbf{r} is the receiving distance, and $b_n = -B_n/D_n$, with $D_n = 0$ being the characteristic equation of a solid sphere. The terms $h_n^{(1)}(kr)$,

$P_n(\cos\theta)$, and specific elements in the matrix are detailed in (WILLIAMS, MARSTON, 1985).

2.2. The theoretical solution of elastic spherical shells

Scattering of plane waves from a spatially fixed spherical shell is considered, with spherical coordinates r centered at the origin O , where the outer radius of the shell is a , the inner radius is b , the thickness is $h = a - b$, and the interior of the shell is a vacuum (Fig. 2). Assume a unit amplitude harmonic plane wave incident along the z -direction onto the spherical shell.

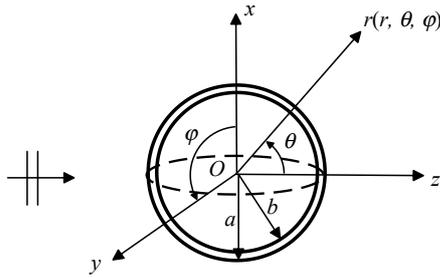


Fig. 2. Schematic diagram of elastic spherical shell scattering.

The boundary conditions of the spherical shell are:

$$\begin{cases} T_{rr} = -p, \\ u_r = \frac{1}{\rho_0 \omega^2} \frac{\partial p}{\partial r}, \\ T_{r\theta} = 0, \end{cases} \quad r = a, \quad (2)$$

$$\begin{cases} T_{rr} = 0, \\ T_{r\theta} = 0, \end{cases} \quad r = b, \quad (3)$$

where T_{rr} and $T_{r\theta}$ represent stress, and u_r represents displacement.

To solve for the scattering coefficient B_n from the boundary conditions, we use the expression $b_n = -B_n/D_n$, where $D_n = 0$ is the characteristic equation of the spherical shell, and the matrix elements B_n and D_n can be found in (GAUNAURD, WERBY, 1991). The scattered acoustic pressure in the surrounding water medium of the spherical shell can still be expressed using Eq. (1).

3. Algorithm verification

To ascertain the precision of the computational modeling of the scattering acoustic fields associated with elastic spheres and spherical shells, we conducted separate calculations for the scattering acoustic fields of a copper sphere with a radius of 0.5 m and a vacuum copper shell with equivalent dimensions and a thickness-to-diameter ratio of 0.05 (Fig. 3). The material properties are detailed in Table 1. The outcomes generated by the algorithm proposed in this study were juxtaposed with those derived from finite element software tailored for two-dimensional axisymmetric computations. The comparative findings are depicted in Fig. 4. It is evident from the results that the two computational approaches exhibit a high degree of agreement, thus affirming the accuracy of the proposed methodology.

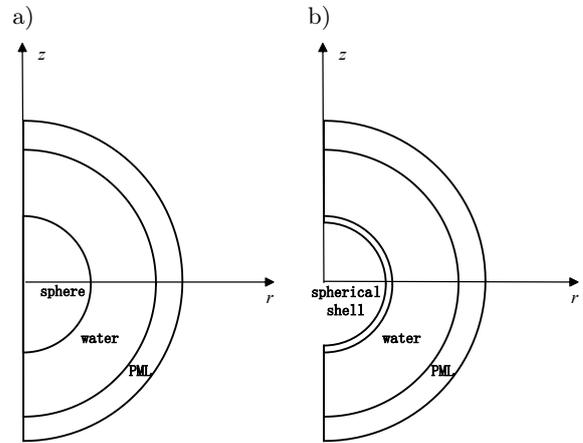


Fig. 3. Finite element software two-dimensional axisymmetric acoustic scattering model: a) sphere; b) spherical shell.

Table 1. Material parameters.

Material	Density [kg/m ³]	Longitudinal wave speed (c_l) [m/s]	Transverse wave speed (c_t) [m/s]
Water	1000	1500	–
Copper	8900	4759	2325

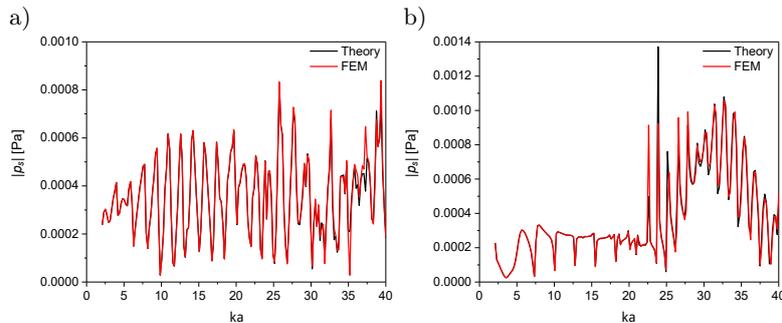


Fig. 4. Comparison of theoretical solution and finite element computational results (ka is dimensionless frequency): a) sphere; b) spherical shell.

4. Dispersion curve

4.1. Solid sphere dispersion curve

The object of study in this section is a copper sphere with a radius of 0.06 m. Utilizing both contour integration and Gaussian integration methods (LONG *et al.*, 1994), we computationally determine the phase velocity and group velocity dispersion curves for the elastic copper sphere with a radius of 0.06 m. The material parameters used for these calculations are outlined in Table 1. The resulting dispersion curves are visually represented in the subsequent figure (Fig. 5). It is evident that the principal types of circumferential waves in the elastic sphere include whispering gallery (W-G) waves, Rayleigh waves, and Franz waves. A higher imaginary component of the root signifies a more pronounced radiation capability; however, it also corresponds to increased attenuation during propagation (TANG *et al.*, 2018). Notably, the imaginary component of Franz waves surpasses that of Rayleigh waves, indicating that Franz waves experience heightened attenuation compared to Rayleigh waves (Fig. 6). Consequently, the observation of Franz waves in backward scattering scenarios poses considerable difficulty.

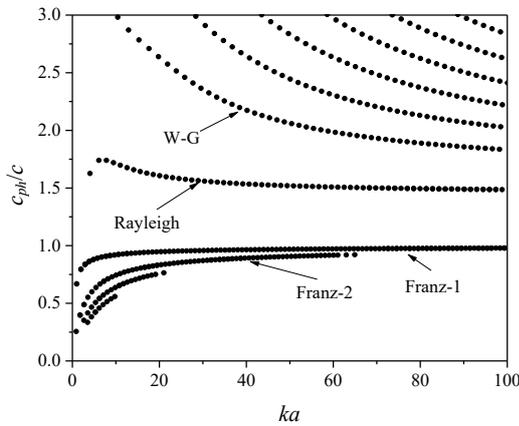


Fig. 5. Phase velocities.

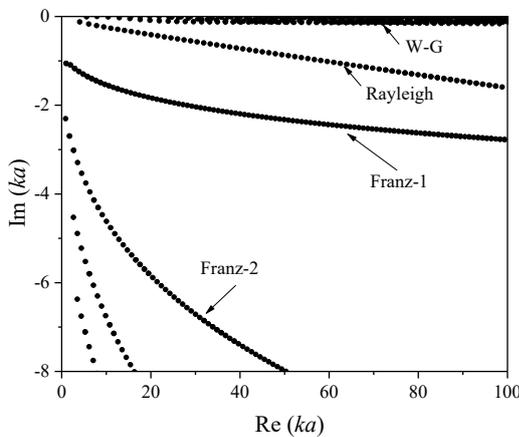


Fig. 6. Distribution plot of the real and imaginary parts.

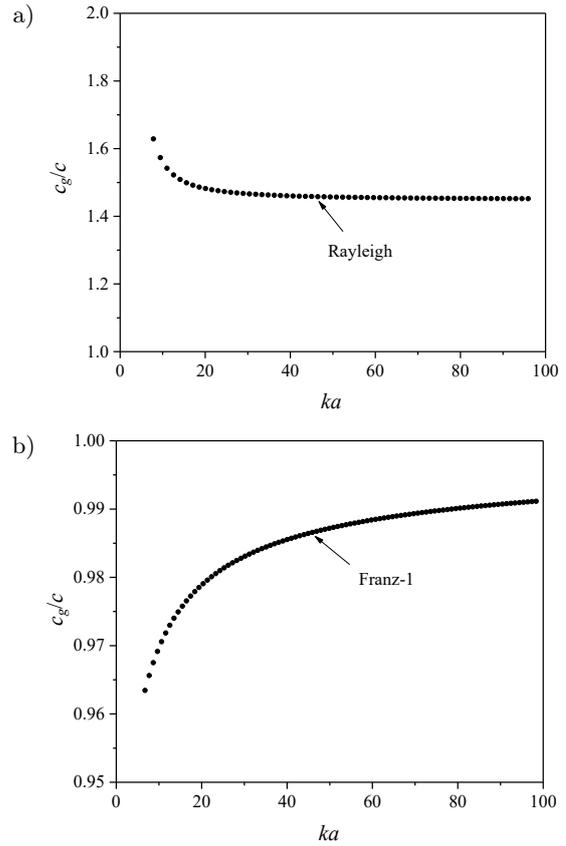


Fig. 7. Group velocities: a) Rayleigh; b) Franz-1.

4.2. Spherical shell dispersion curve

Similar to solving the dispersion curve of a solid sphere, all that is required is to modify the characteristic equation of the sphere to that of a spherical shell. The dispersion curve of the phase velocity for a vacuum copper spherical shell with a radius of 0.06 m and a thickness-to-radius ratio of 0.05 is depicted in Fig. 8. It can be observed that the a_{0+} and a_{0-} waves undergo a bifurcation near the resonance frequency.

From Fig. 9, it is evident that the imaginary component of the a_{0-} wave is nearly negligible at low fre-

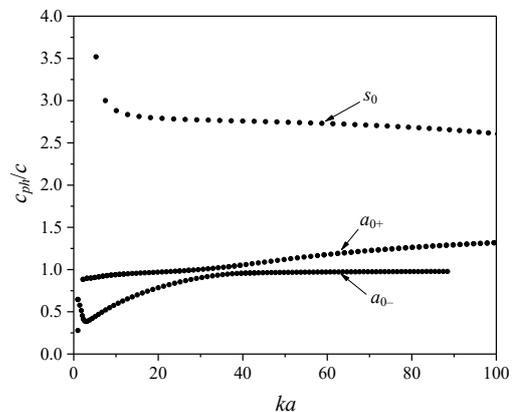


Fig. 8. Phase velocity dispersion curves.

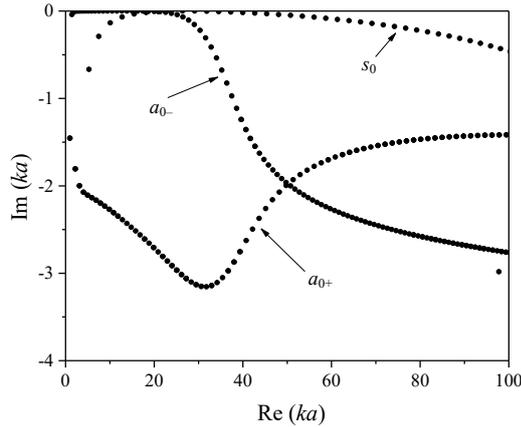


Fig. 9. Distribution plot of the real and imaginary parts.

quencies, suggesting a low radiation efficiency for this wave type within this frequency range. Consequently, the detection of the a_{0-} wave at low frequencies proves challenging. Moreover, a discernible trend is observed wherein the radiation efficiency of the a_{0-} wave increases from low to high frequencies, while conversely, the radiation efficiency of the a_{0+} wave decreases from high to low frequencies, with a notable inflection point near the resonance frequency. Notably, the imaginary component of the s_0 wave remains consistently minimal near the resonance frequency, indicating relatively weak radiation characteristics for this wave type. The calculated group velocity curve of the a_{0-} wave is shown in Fig. 10.

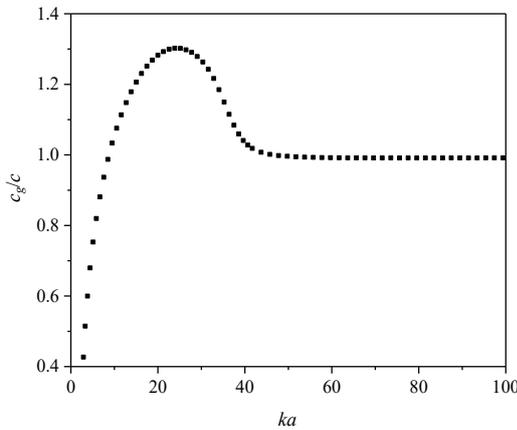


Fig. 10. Group velocities of the a_{0-} wave.

5. Time-domain analysis

The time domain echo calculation of the target mainly includes: constructing the sinusoidal signal as the incident wave, and using the fast Fourier transform to get the incident wave spectrum. The scattering sound pressure of the target is obtained as the transfer function by theoretical solution. The frequency spectrum of the incident signal and the transfer function are multiplied in the frequency domain, and the echo

pulse sequence of the target is obtained by inverse Fourier transform.

5.1. Sphere

A single-cycle 120 kHz ($ka \approx 30$) sine wave is used as the incident signal. At this point, the phase velocity is approximately 2325 m/s, and the group velocity is approximately 2194 m/s. According to the law of refraction:

$$\sin(\theta_c) = c/c_{ph}, \tag{4}$$

where c is the speed of sound in water, which is 1500 m/s. The critical angle θ_c is approximately 40° . This means that the incident wave enters the surface at the critical angle of 40° and radiates outward at the critical angle.

5.1.1. Monostatic configuration

For the backward wave, taking clockwise as an example, the path of the circumferential wave is shown in Fig. 11. The incident wave couples into the spherical surface at point A , decouples and propagates outward at point B . Afterward, a portion of the circumferential wave continues to circulate around the spherical surface for one revolution before decoupling from point B , and so on.

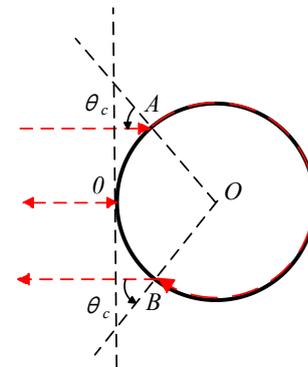


Fig. 11. Propagation path (clockwise).

In order to achieve a more succinct echo curve, a matched filtering technique is applied, utilizing the constructed incident signal as the reference signal. The resulting time-domain echo is matched filtered, yielding the echo curve as shown in Fig. 12. Using the mirror reflection echo at the top 0 as the time reference point, the propagation time between $R1$ and $R2$ is determined to be the time it takes for the wave to complete one full circulation around the spherical surface. This enables the calculation of the Rayleigh wave group velocity, which is approximately 2194 m/s. This value aligns with the Rayleigh wave velocity calculated in the previous section.

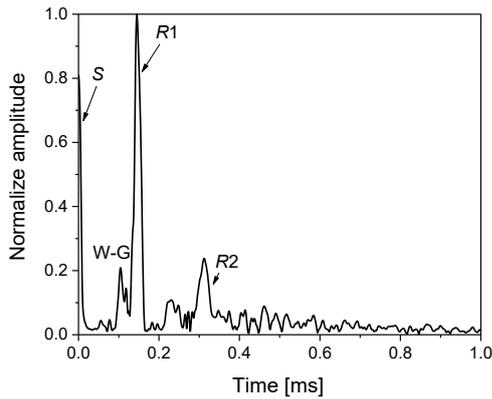


Fig. 12. Impulse responses.

5.1.2. Bistatic configuration

The time-domain echo of a separated transmit-receive elastic sphere is computed, with the foremost

bright line representing the specular echo. The position of this bright line changes with variations in the receiving angle. Upon further observation, it is noted that the time-domain echo pattern of the elastic sphere with a separated transmit-receive configuration exhibits an *X*-mode, which is attributed to the influence of the circumferential components of Rayleigh waves. Rayleigh waves propagate along the surface of the solid sphere in both clockwise (CW) and counterclockwise (CCW) directions. The sphere possesses a strictly symmetric structure, and the waves circumnavigate the sphere’s surface, eventually superimposing in the opposite direction of the shell ($\theta = 180^\circ$). Consequently, bright spots are observed at the 180° angle in Fig. 13b.

According to the propagation paths of the waves, they can be divided into clockwise circumferential propagation around the sphere and counterclockwise circumferential propagation around the sphere, as shown in Figs. 14 and 15.

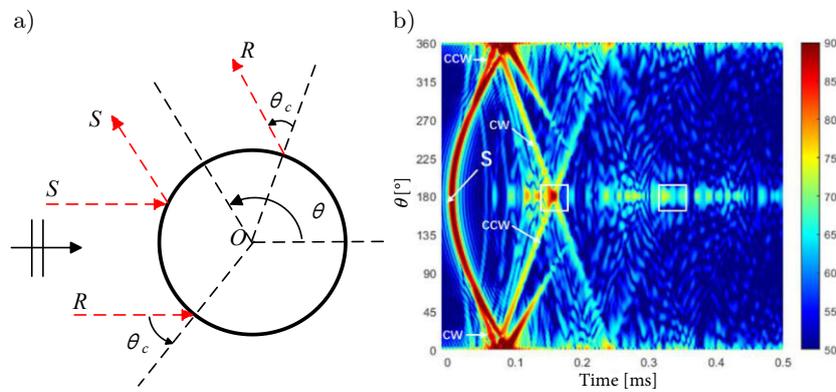


Fig. 13. a) Schematic for the acoustic scattering problem under consideration; b) the bistatic impulse response of the sphere.

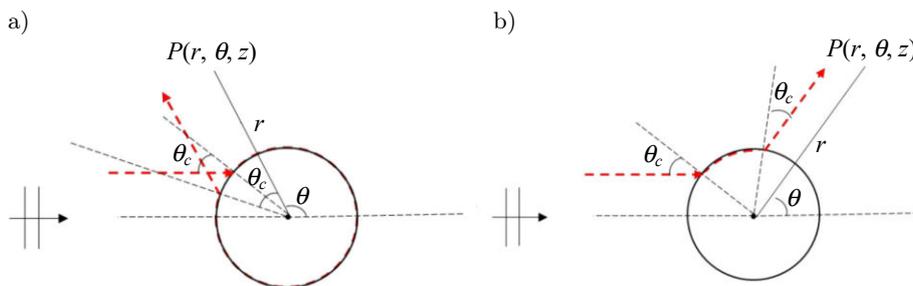


Fig. 14. Propagation path (CW): a) $\pi - \theta < 2\theta_c$; b) $\pi - \theta > 2\theta_c$.

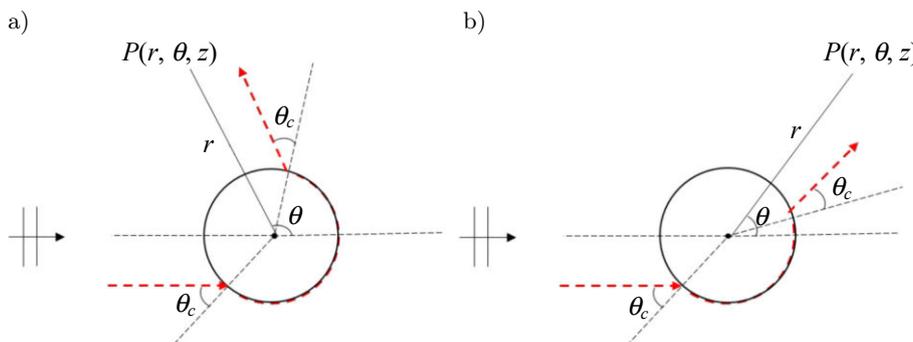


Fig. 15. Propagation path (CCW): a) path 1; b) path 2.

The formula for estimating the arrival time of the echo is given as follows, where $\theta = \pi$ corresponds to the moment of arrival of the reverse echo:

$$t_{CCW} = \frac{2a(1 - \cos(\theta_c))}{c} + \frac{a(\pi + \theta - 2\theta_c)}{c_R^g} + (n-1)\frac{2\pi a}{c_R^g}, \quad (5)$$

$$t_{CW} = \begin{cases} \frac{2a(1 - \cos(\theta_c))}{c} + \frac{a(\pi - \theta - 2\theta_c)}{c_R^g} + (n-1)\frac{2\pi a}{c_R^g}, & \pi - \theta > 2\theta_c, \\ \frac{2a(1 - \cos(\theta_c))}{c} + \frac{a(3\pi - \theta - 2\theta_c)}{c_R^g} + (n-1)\frac{2\pi a}{c_R^g}, & \text{else.} \end{cases} \quad (6)$$

The sphere is a strictly symmetrical structure, so the characteristics of echoes between angles $0^\circ \sim 180^\circ$ and angles $180^\circ \sim 360^\circ$ are consistent. This article uses the range of $0^\circ \sim 180^\circ$ as an example. Based on the formula mentioned earlier to determine the echo arrival time, it can be observed that the curves correspond well with the echo fringes. The relationship between the mirrored echo angle and the echo arrival time follows a sine function, causing the echo fringes to be curved. The first echo of Rayleigh waves has a slope of c_R^g/a , which is positively correlated with the velocity of Rayleigh waves, making the fringe a straight line. At angles below approximately 100° , clockwise echoes reach the receiving point earlier than geometric echoes. Besides the Rayleigh wave echoes (labeled as 1 in Fig. 16), other echoes are observed, possibly corresponding to a combination of W-G waves and other transmitted waves (WILLIAMS, MARSTON, 1985). It can also be observed that as the angle ap-

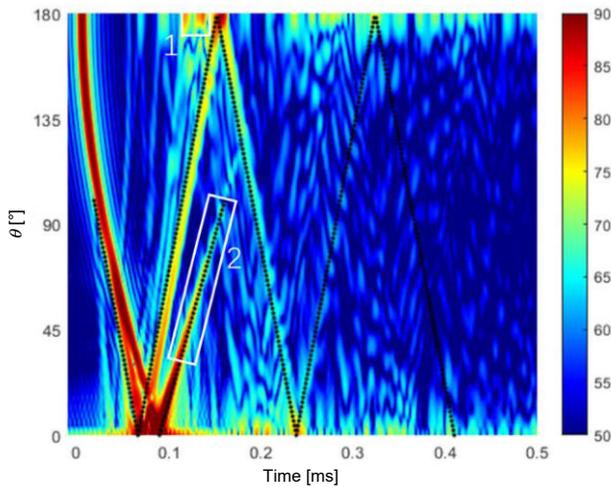


Fig. 16. Bistatic impulse response of the sphere.

proaches the forward direction, a segment of echoes is observed (labeled as 2 in Fig. 16). It is challenging to observe this segment of echoes near the reverse angle. By calculating the arrival time of echoes using the subsonic Franz-1 wave velocity (Fig. 7b), it is observed that this segment of echoes corresponds well with the echo fringes. Therefore, these echoes may be associated with subsonic waves.

5.2. Spherical shell

Calculating the form function of a vacuum copper spherical shell with a radius of 0.06 m and a thickness-to-diameter ratio of 0.05, it can be observed that the elastic shell exhibits a noticeable mid-frequency enhancement ($ka \approx 18 \sim 42$).

When using a single-cycle 120 kHz ($ka \approx 30$) sine signal as the incident signal, and referring to the dispersion curve calculated in Fig. 17, it can be observed that around the resonance frequency, the bending wave on the surface of the spherical shell in water corresponds to the a_{0-} wave. This indicates indicating that the incident acoustic wave enters tangentially ($\theta_c \approx \pi/2$) along the surface of the spherical shell and leaks into the water at an angle of θ_c .

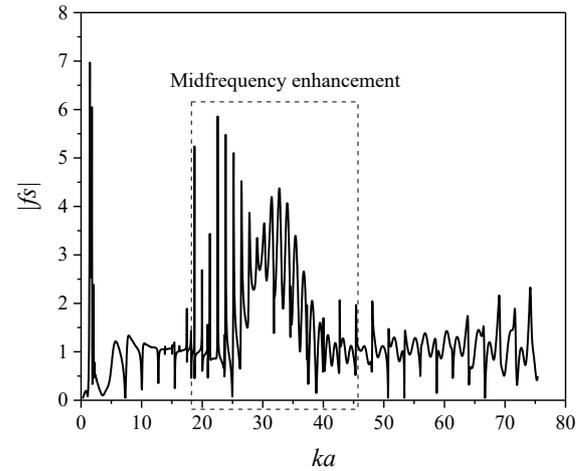


Fig. 17. Form function of a vacuum copper spherical shell.

5.2.1. Monostatic configuration

The incident wave enters tangentially onto the surface of the spherical shell, and the paths of clockwise and counterclockwise propagation of the circumferential waves are illustrated in Fig. 18. The incident wave couples into the sphere at points $A(A')$, decouples and propagates outward at points $B(B')$. Some of the circumferential waves continue to propagate along the surface of the shell for one complete revolution before decoupling and radiating outward from points $B(B')$. This process repeats, resulting in a sequence of echo pulses.

Due to the narrow pulse during signal emission, the propagation velocity needs to be determined using

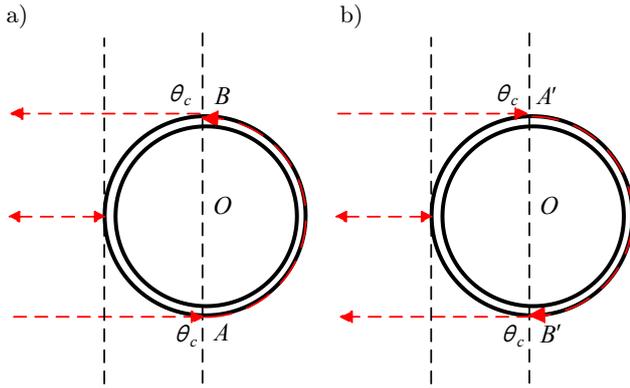


Fig. 18. Propagation path ($\theta_c \approx \pi/2$): a) CCW; b) CW.

the group velocity (TANG *et al.*, 2018). Calculations yield an approximate group velocity of about 1900 m/s for the a_{0-} wave. Using the estimated arrival time Eq. (9) based on the time when the echo arrives, with the echo reflected from the retroreflector as the time reference point, the calculated times are plotted against the time-domain echo curve. Three plotted curves from left to right (Fig. 19) represent the first, second, and third antisymmetric echoes. It can be observed that the calculated echo arrival times correspond well with the time-domain curves.

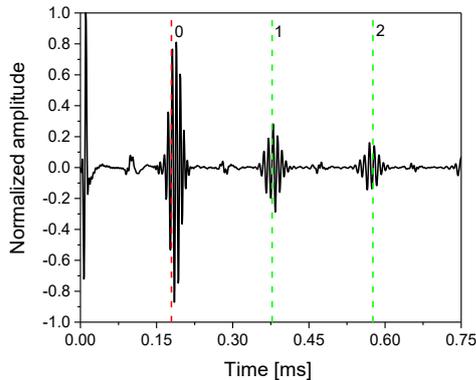


Fig. 19. Impulse responses.

5.2.2. Bistatic configuration

The time-domain echoes of the elastic spherical shell in the separate transmit-receive configuration were calculated. Upon observation, similarities were found with the characteristics of the bistatic echoes from the copper sphere. At the forefront, there is a bright line corresponding to the mirror image echo, and the position of this line changes with variations in the receiving angle. The time-domain echo pattern in the separate transmit-receive configuration for the elastic spherical shell resembles an X -mode, influenced by the circumferential components of the a_{0-} wave (ANDERSON, 2012). The a_{0-} wave propagates clockwise and counterclockwise along the surface of the spherical shell. Due to the strict symmetry of the shell

structure, the paths of waves propagating clockwise and counterclockwise along the surface are identical. Consequently, they intersect in the opposite direction of the shell ($\theta = 180^\circ$), resulting in a bright spot at 180° in the graph, which is the culmination of the superposition of circumferential waves (Fig. 20).

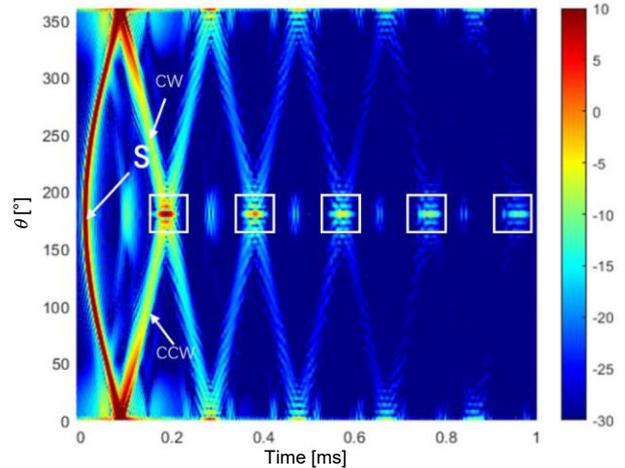


Fig. 20. Bistatic impulse response of the spherical shell.

According to the preceding section, the incident wave enters tangentially onto the surface of the spherical shell. When propagating counterclockwise, the incident acoustic wave couples from point A to point B and decouples to radiate outward. The angle corresponding to the circumferential wave path on the shell surface is θ_1 . When propagating clockwise, the incident acoustic wave couples from point A' to point B' and decouples to radiate outward. The angle corresponding to the circumferential wave path on the shell surface is θ_2 .

Observe the above diagram (Fig. 21) to obtain the angles corresponding to counterclockwise and clockwise circumferential waves:

$$\theta_1 = \theta, \quad (7)$$

$$\theta_2 = 2\pi - \theta. \quad (8)$$

Calculate the arrival times of counterclockwise and clockwise circumferential waves along the surface of the spherical shell based on the propagation paths:

$$t_{CCW} = \frac{2a}{c} + a \frac{\theta_1}{c_{a_{0-}}^g} + (n-1) \frac{2\pi a}{c_{a_{0-}}^g}, \quad (9)$$

$$t_{CW} = \frac{2a}{c} + a \frac{\theta_2}{c_{a_{0-}}^g} + (n-1) \frac{2\pi a}{c_{a_{0-}}^g}. \quad (10)$$

The time difference between the arrivals of counterclockwise and clockwise circumferential waves is given by Eq. (11). It can be observed that when the backward echo occurs $\theta = \pi$, the time difference between clockwise and counterclockwise waves is 0:

$$\Delta t = t_{CW} - t_{CCW} = a \frac{\theta_2 - \theta_1}{c_{a_{0-}}^g} = a \frac{2(\pi - \theta)}{c_{a_{0-}}^g}. \quad (11)$$

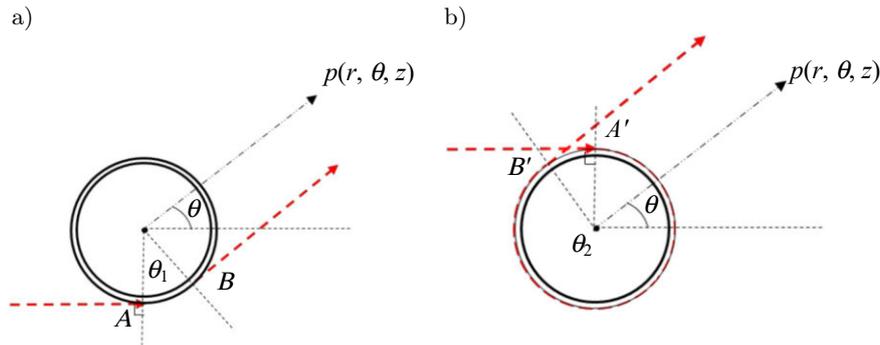


Fig. 21. Propagation path: a) CCW; b) CW.

When the circumferential wave reaches point B or B' , a portion decouples and radiates outward, while another portion continues to circumnavigate the spherical shell before decoupling and radiating outward. The remaining part continues to circumnavigate along the surface of the spherical shell. Therefore, the time difference between adjacent echoes of clockwise or counterclockwise echoes is the time it takes for the circumferential wave to complete one full rotation along the surface of the spherical shell:

$$\Delta t_0 = \frac{2\pi a}{c_{a_0}^g}. \quad (12)$$

According to the formula for estimating the echo timing, the bistatic echoes are calculated and plotted on the time-angle spectrum. Since the spherical shell is a strictly symmetric structure, this paper takes the example of $0^\circ \sim 180^\circ$. It can be observed that the calculated echo timings correspond well with the time-angle spectrum (Fig. 22).

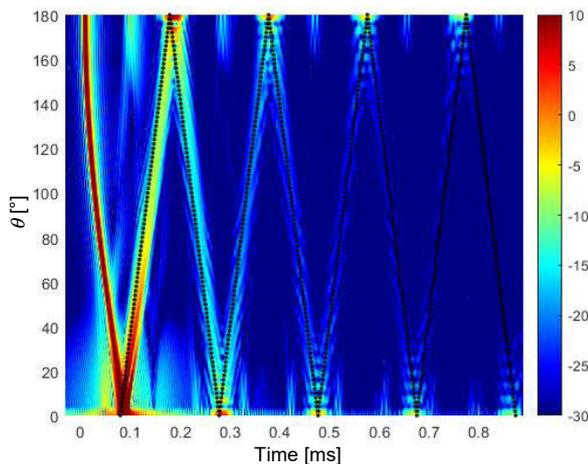


Fig. 22. Bistatic impulse response of the spherical shell.

6. Conclusions and discussion

This research computed the time-domain echoes of solid spheres and spherical shells under plane wave incidence by employing a method that constructs the

incident signal and performs multiplication operations in the frequency domain with the target scattering acoustic field. Based on the scattering acoustic field of elastic spheres and spherical shells, the following main conclusions are drawn:

- 1) Under the same planar wave incident signal, echoes from solid spheres are more complex compared to those from spherical shells. At specific frequencies, solid spheres predominantly exhibit Rayleigh and W-G waves, while spherical shells mainly manifest a_{0-} waves, with a noticeable mid-frequency enhancement phenomenon. In the time-domain echoes, this is evident as distinct wave packets, offering valuable insights for distinguishing between solid spheres and spherical shells underwater.
- 2) The study reveals that, for elastic spheres within the near-forward angle range, elastic waves arrive faster than specular echoes. This phenomenon is attributed to the azimuthal function of the circumferential wave arrival time on the spherical surface, which is dependent on the circumferential wave velocity and path.
- 3) By examining the relationship between echo arrival time and reception angle, it is observed that a small segment of the echo can be observed near the forward angle, whereas this segment of the echo is not observed near the backward angle. Computational analysis reveals that this segment of the echo corresponds to subsonic waves.
- 4) The echo arrival times of separated transmit-receive configurations for elastic spheres and spherical shells can be predicted. Estimation formulas for the bistatic echo arrival times of the two targets are provided, showing good agreement between the predicted echo arrival times and the time-domain echo curves.

Acknowledgments

This study was supported by the National Natural Science Youth Fund (52201397) and the Jiangsu Graduate Practical Innovation Program (SJCX23_2215).

References

1. ANDERSON S.D. (2012), *Space-time-frequency processing from the analysis of bistatic scattering for simple underwater targets*, Ph.D. Thesis, College of Engineering, George W. Woodruff School of Mechanical Engineering.
2. APOSTOLOUIDIA A., DOUKA E., HADJILEONTIADIS L.J., REKANOS I.T., TROCHIDIS A. (2007), Crack detection on beams by time-frequency analysis of transient flexural waves, *Archives of Acoustics*, **32**(4): 941–954.
3. AYRES V.M., GAUNAURD G.C., TSUI C.Y., WERBY M.F. (1987), The effects of Lamb waves on the sonar cross-sections of elastic spherical shells, *International Journal of Solids and Structures*, **23**(7): 937–946, doi: [10.1016/0020-7683\(87\)90088-6](https://doi.org/10.1016/0020-7683(87)90088-6).
4. BEDNARZ J. (2017), Operational modal analysis for crack detection in rotating blades, *Archives of Acoustics*, **42**(1): 105–112, doi: [10.1515/aoa-2017-0011](https://doi.org/10.1515/aoa-2017-0011).
5. DIERCKS K.J., HICKLING R. (1967), Echoes from hollow aluminum spheres in water, *The Journal of the Acoustical Society of America*, **41**(2): 380–393, doi: [10.1121/1.1910349](https://doi.org/10.1121/1.1910349).
6. DING D., CHEN C.X., KONG H.M., FAN J., PENG Z.L. (2023), Acoustic coding based on high frequency time domain echo of layered elastic spherical shells in water [in Chinese], *Applied Acoustics*, **42**(4): 781–791.
7. FAN W., FAN J., WANG X.N. (2012), Application of the SWT method to scattering from water-filled elastic spherical shells [in Chinese], *Journal of Ship Mechanics*, **16**(6): 705–715.
8. FAWCETT J.A. (2015), Computing the scattering from slightly deformed spherical shells, *IEEE Journal of Oceanic Engineering*, **41**(3): 682–688, doi: [10.1109/JOE.2015.2478995](https://doi.org/10.1109/JOE.2015.2478995).
9. GAUNAURD G., ÜBERALL H. (1985), Relation between creeping-wave acoustic transients and the complex-frequency poles of the singularity expansion method, *The Journal of the Acoustical Society of America*, **78**(1): 234–243, doi: [10.1121/1.392564](https://doi.org/10.1121/1.392564).
10. GAUNAURD G.C., ÜBERALL H. (1983), RST analysis of monostatic and bistatic acoustic echoes from an elastic sphere, *The Journal of the Acoustical Society of America*, **73**(1): 1–12, doi: [10.1121/1.388839](https://doi.org/10.1121/1.388839).
11. GAUNAURD G.C., WERBY M.F. (1987), Lamb and creeping waves around submerged spherical shells resonantly excited by sound scattering, *The Journal of the Acoustical Society of America*, **82**(6): 2021–2033, doi: [10.1121/1.395646](https://doi.org/10.1121/1.395646).
12. GAUNAURD G.C., WERBY M.F. (1991), Sound scattering by resonantly excited, fluid-loaded, elastic spherical shells, *The Journal of the Acoustical Society of America*, **90**(5): 2536–2550, doi: [10.1121/1.402059](https://doi.org/10.1121/1.402059).
13. GUNDERSON A.M., ESPAÑA A.L., MARSTON P.L. (2017), Spectral analysis of bistatic scattering from underwater elastic cylinders and spheres, *The Journal of the Acoustical Society of America*, **142**(1): 110–115, doi: [10.1121/1.4990690](https://doi.org/10.1121/1.4990690).
14. KARGL S.G., WILLIAMS K.L., THORSOS E.I. (2012), Synthetic aperture sonar imaging of simple finite targets, *IEEE Journal of Oceanic Engineering*, **37**(3): 516–532, doi: [10.1109/JOE.2012.2200815](https://doi.org/10.1109/JOE.2012.2200815).
15. LI X., WU Y. (2019), Feature extraction for acoustic scattering from a buried target, *Journal of Marine Science and Application*, **18**: 380–386, doi: [10.1007/s11804-019-00102-9](https://doi.org/10.1007/s11804-019-00102-9).
16. LONG Y.L., WEN X.L., XIE C.F. (1994), An implementation of a root finding algorithm for transcendental functions in a complex plane [in Chinese], *Journal on Numerical Methods and Computer Applications*, pp. 88–92.
17. MARSTON P.L., SUN N.H. (1992), Resonance and interference scattering near the coincidence frequency of a thin spherical shell: An approximate ray synthesis, *The Journal of the Acoustical Society of America*, **92**(6): 3315–3319, doi: [10.1121/1.404181](https://doi.org/10.1121/1.404181).
18. QIAO S., SHANG X., PAN E. (2016), Elastic guided waves in a coated spherical shell, *Nondestructive Testing and Evaluation*, **31**(2): 165–190, doi: [10.1080/10589759.2015.1079631](https://doi.org/10.1080/10589759.2015.1079631).
19. SU J., WANG F., DU S. (2017), An elastic wave enhancement method based on modified bright point model, [in:] *2017 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–4, doi: [10.1109/ICSPCC.2017.8242491](https://doi.org/10.1109/ICSPCC.2017.8242491).
20. TANG W.L., FAN J., MA Z.C. (2018), Elastic acoustic scattering mechanism of targets in water, *The Acoustic Scattering of Underwater Target* [in Chinese], pp. 79–84, Science Press, China.
21. THOMPSON M. (2023), *Time-frequency sonar detection of elastic wave reradiation*, Ph.D. Thesis, Electrical and Computer Engineering, Auburn University.
22. TOO G.P., LIN Y.W., KE Y.C. (2014), Echoes analysis from spherical elastic shells by using iterative time reversal mirror, [in:] *OCEANS 2014 – TAIPEI*, pp. 1–5, doi: [10.1109/OCEANS-TAIPEI.2014.6964463](https://doi.org/10.1109/OCEANS-TAIPEI.2014.6964463).
23. ÜBERALL H., GAUNAURD G.C., MURPHY J.D. (1982), Acoustic surface wave pulses and the ringing of resonances, *The Journal of the Acoustical Society of America*, **72**(3): 1014–1017, doi: [10.1121/1.388232](https://doi.org/10.1121/1.388232).
24. WILLIAMS K.L., MARSTON P.L. (1985), Backscattering from an elastic sphere: Sommerfeld–Watson transformation and experimental confirmation, *The Journal of the Acoustical Society of America*, **78**(3): 1093–1102, doi: [10.1121/1.393028](https://doi.org/10.1121/1.393028).
25. XIA Z., LI X., MENG X. (2016), High resolution time-delay estimation of underwater target geometric scattering, *Applied Acoustics*, **114**: 111–117, doi: [10.1016/j.apacoust.2016.07.016](https://doi.org/10.1016/j.apacoust.2016.07.016).
26. YU X., PENG L., YU G. (2014), Extracting the subsonic anti-symmetric lamb wave from a submerged thin spherical shell backscattering through iterative time reversal, *Journal of Ocean University of China*, **13**: 589–596, doi: [10.1007/s11802-014-2166-8](https://doi.org/10.1007/s11802-014-2166-8).

Research Paper

Prediction Method and Characteristics of Static Acoustic Scattering
for Marine Composite Propellers

Suchen XU⁽¹⁾, Zilong PENG^{(1)*}, Fulin ZHOU⁽²⁾,
Xuhong MIU⁽³⁾, Huicheng KE⁽¹⁾

⁽¹⁾ School of Energy and Power
Jiangsu University of Science and Technology
Zhenjiang, China

⁽²⁾ State Key Laboratory of Ocean Engineering
Collaborative Innovation Center for Advanced Ship and Deep-Sea Exploration
Shanghai Jiao Tong University
Shanghai, China

⁽³⁾ Unit 92578 of the People's Liberation Army
Beijing, China

*Corresponding Author e-mail: zlp_just@sina.com

(received May 10, 2024; accepted July 4, 2024; published online October 17, 2024)

This study introduces a hybrid approach to predict the acoustic scattering characteristics of composite propellers featuring variable thickness and complex curvature. The approach combines the Kirchhoff approximation (KA), which employs an intersection algorithm (IA) for determining the thickness of discrete surface elements, with the theory of orthotropic laminate transfer matrix (OLTM). The overall scattered sound field of the target is determined by solving the reflection coefficients of each surface element. To enhance computational efficiency, the scattered sound field of a complete composite propeller is ingeniously predicted by cloning mesh topology from a single propeller blade, taking advantage of the rotation symmetry characteristics of the composite propeller. The validity of this prediction method is confirmed through the finite element method (FEM) and static acoustic scattering characteristic experiments conducted on a lake. The predicted results for the target strength (TS) of the composite propeller closely align with the FEM. Additionally, the TS and time-domain echo characteristics of the steel propeller utilizing the KA exhibit strong agreement with the experimental findings. These research findings provide a significant reference value for predicting the acoustic scattering characteristics near the stern of underwater vehicles.

Keywords: acoustic scattering characteristics; composite propeller; Kirchhoff approximation; target strength.



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

When active sonar is utilized to detect the stern of an underwater vehicle, the stern sections, including the propeller, and the rudder, contribute the main echo (CHU, STANTON, 2010; KLAUSNER, AZIMI-SADJADI, 2014; TUCKER, AZIMI-SADJADI, 2011). Compared to traditional metal propellers, composites offer advantages such as higher strength, lower specific gravity, corrosion resistance, and good capability of stealth

(VARDHAN *et al.*, 2019; ISLAM *et al.*, 2022; UDDIN *et al.*, 2021; MOTLEY *et al.*, 2009). However, the acoustic scattering characteristics of the composite propeller are often overlooked in both domestically and internationally research. Modeling the acoustic response of static non motion propellers primarily focuses on their inherent acoustic characteristics in a non-working state, analyzing how material properties and geometric shapes influence acoustic scattering characteristics. In contrast, modeling the acoustic response of rotat-

ing propellers primarily addresses the hydrodynamic noise generated during operation, which is more complex than static propeller modeling. This involves accounting for the Doppler frequency shift effect induced by rotation and the acoustic scattering characteristics of the wake bubble cloud (CHENG *et al.*, 2023; SAFARI *et al.*, 2023). For a static composite propeller, it is crucial to consider variations in blade thickness and curvature.

Currently, prevalent numerical methods for predicting the acoustic scattering characteristics of underwater targets at low frequencies include the finite element method – FEM (JIANG *et al.*, 2023; ISAKSON, CHOTIROS, 2014; SABAT *et al.*, 2023; YANG *et al.*, 2024), and the boundary element method – BEM (LANGDON, CHANDLER-WILDE, 2006; VENĀS, KVAMSDAL, 2020; SEYBERT *et al.*, 1988; ZHANG *et al.*, 2020). However, these numerical methods necessitate significant computational resources, with increasing frequency. The Kirfchhoff approximation (KA) method is extensively employed for rapid prediction of the acoustic scattering characteristics of complex targets in a high-frequency range (KWON *et al.*, 2017). LAVIA *et al.* (2019) discretized the target into individual curved surface elements and computed the target's backscattering in high frequency through iterative integration using the Gauss-Legendre rule, as adopted in the KA. TANG *et al.* (1993) proposed a method called the planar element method, which has been broadly utilized in modeling acoustic echo of complex targets.

Numerous issues persist when employing the KA method for analyzing the acoustic scattering characteristics of the composite propeller. Each discrete surface element of a propeller blade possesses the distinct thickness and curvature (SAGAR *et al.*, 2013), with the primary variation pattern characterized by thinning towards the edges and thickening towards the leading edge, as well as transitioning from thicker at the blade root to thinner at the blade tip. The traditional KA method, used for assessing target strength (TS), only considers the reflection coefficients of uniformly thick targets. Variations in thickness and curvature across different parts of blades, as well as complex edge profiles, which inevitably lead to computational errors, have not yet been taken into account.

In this paper, a predictive method for the acoustic scattering characteristics of composite thin plates with variable thickness is proposed. By constructing discrete element thickness and material property information, and based on the theory of orthotropic laminate transfer matrix (OLTMT), the reflection coefficients of all elements are solved. This leads to the acquisition of the total scattered sound field for composite targets with variable thickness. By employing a mesh cloning topology approach, the prediction of acoustic scattering characteristic for the composite propeller with ro-

tational symmetry structures has been realized. The accuracy of this method in predicting the acoustic scattering characteristics of the static composite propeller has been validated through the FEM. Finally, this method is applied to the isotropic steel propeller, where the predicted results of TS and time-domain echo characteristics are in good agreement with experimental results.

2. Theoretical research

2.1. Physical method of geometrical acoustic scattering of underwater target

The KA formula is applied to the problem of acoustic scattering on nonrigid surfaces, as shown in Fig. 1. $Oxyz$ denotes a spatial Cartesian coordinate system, S represents the integral surface of the nonrigid target, Q_1 is the point of sound wave incidence, and Q_2 is the point of sound wave reception. The vectors \mathbf{r}_1 and \mathbf{r}_2 correspond to the positions of the incidence and reception points, respectively; α_1 and α_2 are the angles of incidence and reception, respectively, while \mathbf{n} represents the normal vector of the surface element.

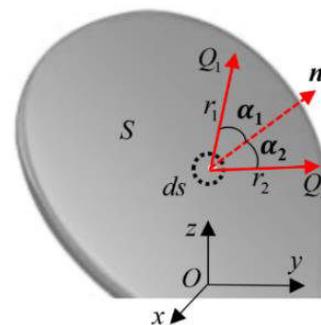


Fig. 1. Schematic diagram of the integral region.

Assume that the surface reflection coefficient is $V(\alpha)$ and the surface acoustic impedance is Z . In the illuminated region, the elastic surface conditions are:

$$\begin{cases} \phi_s = V(\alpha)\phi_i, \\ \frac{i\omega\rho_0(\phi_i + \phi_s)}{\partial(\phi_i + \phi_s)/\partial\mathbf{n}} = -Z, \end{cases} \quad (1)$$

where ϕ_i and ϕ_s are the potential functions of the incident sound wave and the scattered sound field, respectively, ω is the angular frequency of the incident wave, and ρ_0 is the density of the medium surrounding the target. From this, the potential function of the scattered sound wave is obtained as follows (FAN *et al.*, 2012):

$$\phi_s = -\frac{ikA}{4\pi} \int_S \frac{e^{ik(\mathbf{r}_1 + \mathbf{r}_2)}}{\mathbf{r}_1 \mathbf{r}_2} V(\alpha)(\cos \alpha_1 + \cos \alpha_2) dS, \quad (2)$$

where k is the sound wavenumber, A is the amplitude.

In monostatic configuration,

$$\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r},$$

$$\alpha_1 = \alpha_2 = \alpha,$$

it follows that:

$$\phi_s = -\frac{ikA}{2\pi} \int_S \frac{e^{2ikr}}{r^2} V(\alpha) \cos \alpha \, ds. \quad (3)$$

The TS under the far-field conditions is obtained as follows:

$$TS = 10 \lg \left(\frac{1}{\lambda^2} |I|^2 \right), \quad (4)$$

where λ is the sound wavelength,

$$I = \int_S e^{2ik\Delta r} \cos \alpha V(\alpha) \, ds,$$

and Δr is the path difference of a surface element relative to a reference point.

2.2. Theory of orthotropic laminate transfer matrix

The physical model for the propagation of plane waves in composite laminated plates is shown in Fig. 2. θ_i is the angle of incidence, θ_r is the angle of reflection, and θ_t is the angle of transmission, with the angular frequency denoted as ω . The x_3 refers to the direction of thickness, while x_1 refers to the principal direction. For a given single-layer medium, the x_3 coordinate of the transmission boundary is 0, and the x_3 coordinate of the incident boundary is d_k , representing the thickness of that layer. The total thickness D of the multi-layered solid medium is

$$D = \sum_{k=1}^n d_k.$$

Fluid media are present on both sides of the composite laminated plate. At the fluid-solid interface, the wave type conversion occurs: part of the wave is reflected into the external fluid, while the remainder converts into shear and longitudinal waves, continuing to propagate in the orthotropic medium and undergoing further reflection and transmission at the lower interface.

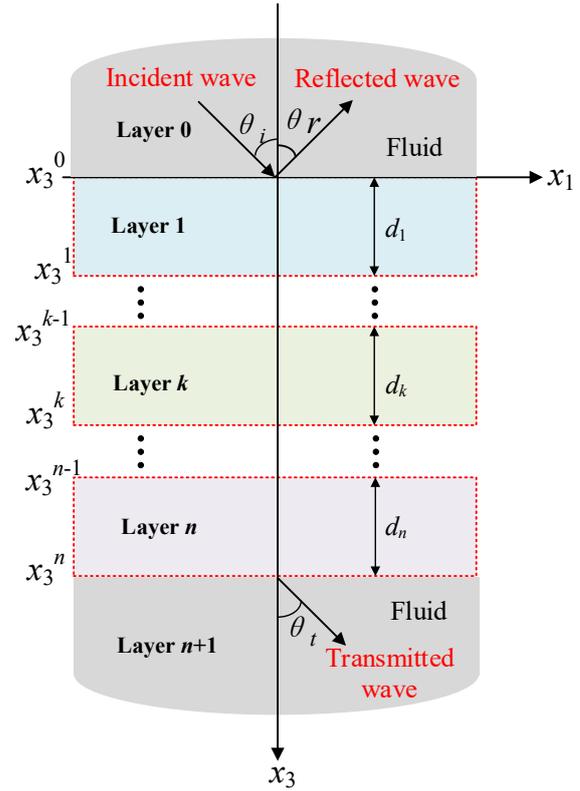


Fig. 2. Schematic diagram of interface between fluid and solid.

According to the theory of OLTM (Kuo et al., 2008; Lin et al., 2017), the relationship between stress, vibrational velocity, and amplitude in a single uniform medium is:

$$\begin{bmatrix} v_1 \\ v_3 \\ \sigma_{33} \\ \sigma_{13} \end{bmatrix} = [M] \begin{bmatrix} (R_1 + R'_1) \\ (R_1 - R'_1) \\ (R_2 + R'_2) \\ (R_2 - R'_2) \end{bmatrix}, \quad (5)$$

where v_1 and v_3 are the velocity of x_1 and x_3 direction, respectively; σ_{33} and σ_{13} , respectively, are the normal stress and shear stress on the solid surface; R_1 and R_2 are the amplitudes of the incident sound wave, while R'_1 and R'_2 are the amplitudes of the reflected sound wave.

The matrix M is:

$$[M] = \begin{bmatrix} \omega k_t \cos(k_{13}x_3) & -j\omega k_t \sin(k_{13}x_3) & j\omega k_{23} \sin(k_{23}x_3) & -j\omega k_{23} \cos(k_{23}x_3) \\ -j\omega_{13} \sin(k_{13}x_3) & \omega k_{13} \cos(k_{13}x_3) & \omega k_t \cos(k_{23}x_3) & -j\omega k_t \sin(k_{23}x_3) \\ -(Q_{13}k_t^2 + Q_{33}k_{13}^2) \cos(k_{13}x_3) & j(Q_{13}k_t^2 + Q_{33}k_{13}^2) \sin(k_{13}x_3) & j(Q_{33} - Q_{13})k_t k_{23} \sin(k_{23}x_3) & -(Q_{33} - Q_{13})k_t k_{23} \cos(k_{23}x_3) \\ 2jQ_{55}k_t k_{13} \sin(k_{13}x_3) & -2Q_{55}k_t k_{13} \cos(k_{13}x_3) & Q_{55}(k_{23}^2 - k_t^2) \cos(k_{23}x_3) & -jQ_{55}(k_{23}^2 - k_t^2) \sin(k_{23}x_3) \end{bmatrix}, \quad (6)$$

where Q_{ij} is the element of the stiffness matrix for orthotropic materials (LI, 2022); k_t is the wave number of the incident sound wave along the tangential direction of the plate; k_{13} and k_{23} are the compression wave and the shear wave in the direction of x_3 , respectively.

$$k_t = \frac{\omega}{c_0} \sin \theta_i, \quad (7)$$

$$k_{13} = \sqrt{\frac{\omega^2 - A_1 k_t^2}{A_3}} = \omega \sqrt{1 - \frac{A_1}{c_0^2} \sin^2 \theta_i / A_3}, \quad (8)$$

$$k_{23} = \sqrt{\frac{\omega^2 - a_1 k_t^2}{a_3}} = \omega \sqrt{1 - \frac{a_1}{c_0^2} \sin^2 \theta_i / a_3}, \quad (9)$$

where c_0 is the sound speed of fluid; $A_1 = (Q_{11} + Q_{13} + 2Q_{55})/2\rho$ and $A_3 = (Q_{33} + Q_{13} + 2Q_{55})/2\rho$; $a_1 = (Q_{11} - Q_{13})/2\rho$ and $a_3 = (Q_{33} - Q_{13})/2\rho$; ρ is the material density.

At the incident boundary, $x_3 = d$, while at the transmission boundary, $x_3 = 0$. Therefore, the mutual transfer relationships of vibrational velocity and stress at the upper and lower surfaces of each layer of the medium is:

$$\begin{bmatrix} v_1 \\ v_3 \\ \sigma_{33} \\ \sigma_{13} \end{bmatrix} \Big|_D = [M_d][M_0]^{-1} \begin{bmatrix} v_1 \\ v_3 \\ \sigma_{33} \\ \sigma_{13} \end{bmatrix} \Big|_0 = [TT_k] \begin{bmatrix} v_1 \\ v_3 \\ \sigma_{33} \\ \sigma_{13} \end{bmatrix} \Big|_0, \quad (10)$$

where $[TT_k]$ is the stress-strain transfer matrix of the layer no. k of the medium.

At the interfaces between different solid media, the stress and strain are continuous and consistent. Therefore, based on the consecutive multiplication of matrix, the overall transfer matrix $[T]$ of stress and vibrational velocity from the transmission boundary to the incident boundary of the multi-layered medium is:

$$\begin{bmatrix} v_1 \\ v_3 \\ \sigma_{33} \\ \sigma_{13} \end{bmatrix} \Big|_D = \prod_{k=1}^n [TT_k] \begin{bmatrix} v_1 \\ v_3 \\ \sigma_{33} \\ \sigma_{13} \end{bmatrix} \Big|_0 = [T] \begin{bmatrix} v_1 \\ v_3 \\ \sigma_{33} \\ \sigma_{13} \end{bmatrix} \Big|_0. \quad (11)$$

Applying the conditions of normal stress and normal displacement continuity on the upper and lower surfaces of orthotropic laminated plates (HU, 2017), the reflection coefficient R and the transmission coefficient D can be expressed as:

$$R = \frac{jk_{13}^{n+1}(a^*) - (b^*)\omega^2\rho_{n+1}}{jk_{13}^{n+1}(a^*) + (b^*)\omega^2\rho_{n+1}}, \quad (12)$$

$$D = \frac{\rho_1}{\rho_{n+1}} \frac{jk_{13}^{n+1}(1-R)}{jk_{13}^1 m_{22} + m_{23}\rho_1\omega^2}, \quad (13)$$

where

$$a^* = jm_{32}k_{13}^1 + m_{33}\rho_1\omega^2,$$

$$b^* = jm_{22}k_{13}^1 + m_{23}\rho_1\omega^2,$$

ρ_1 and ρ_{n+1} are the fluid medium density at the incident and transmission boundaries, respectively; k_{13}^1 and k_{13}^{n+1} are the wave numbers of the compressional wave at the incident and transmission boundaries, respectively; $m_{ik} = T_{ik} - T_{i1}T_{4k}/T_{41}$,

$$k_{13}^1 = \frac{\omega}{c_0} \cos \theta_i, \quad (14)$$

$$k_{13}^{n+1} = \frac{\omega}{c_0} \cos \theta_t. \quad (15)$$

2.3. Calculation method for static acoustic scattering of the composite propeller

For addressing the acoustic scattering issues of the composite propeller, a method is proposed that involves independent assessing and assigning values to the thickness of discrete surface elements during the KA process.

For the issue of element thickness discrimination, a simplified judgment method is proposed: the thickness determination of each independent element can be approximated as a problem of solving the spatial vector length at point M_2 of the rear interface plate penetrated by the normal located at the central point O_1 of the front plate. The spatial vectors from the origin to points O_1 and M_2 are represented as \mathbf{O}_1 and \mathbf{M}_2 , respectively, and the length of this vector is the thickness d of the plate, the method for determining the thickness of an element is shown in Fig. 3.

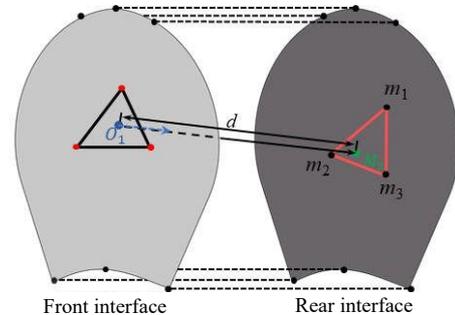


Fig. 3. Method for determining the thickness of surface elements.

Based on the intersection algorithm – IA (MÖLLER, TRUMBORE, 1997), $\mathbf{M}_2(uv)$ can be defined and represented as $\mathbf{M}_2(u, v) = \mathbf{O}_1 + d\mathbf{n}$, with the assumption that:

$$\mathbf{O}_1 + d\mathbf{n} = (1 - u - v)\mathbf{m}_1 + u\mathbf{m}_2 + v\mathbf{m}_3, \quad (16)$$

where u and v satisfy the conditions: $u \geq 0$, $v \geq 0$, and $u + v = 1$; \mathbf{m}_1 , \mathbf{m}_2 , \mathbf{m}_3 , respectively, represent the spatial position vectors of the vertices of the rear plate, the normalized vector in the direction of the normal of the front plate is \mathbf{n} . After organizing, it is obtained that:

$$\begin{bmatrix} -\mathbf{n} & \mathbf{m}_2 - \mathbf{m}_1 & \mathbf{m}_3 - \mathbf{m}_1 \end{bmatrix} \begin{bmatrix} d \\ u \\ v \end{bmatrix} = \mathbf{O}_1 - \mathbf{m}_1. \quad (17)$$

Move the vertex \mathbf{m}_1 of the rear interface element to the origin and transform it into a unit triangle within the Oyz plane, aligning the normal direction \mathbf{n} parallel to the x -axis. Define $\mathbf{E}_1 = \mathbf{m}_2 - \mathbf{m}_1$, $\mathbf{E}_2 = \mathbf{m}_3 - \mathbf{m}_1$, $\mathbf{T} = \mathbf{O}_1 - \mathbf{m}_1$, by applying Cramer's rule to Eq. (17), the solution is obtained as follows:

$$\begin{bmatrix} d \\ u \\ v \end{bmatrix} = \frac{1}{(\mathbf{n} \times \mathbf{E}_2) \cdot \mathbf{E}_1} \begin{bmatrix} (\mathbf{T} \times \mathbf{E}_1) \cdot \mathbf{E}_2 \\ (\mathbf{n} \times \mathbf{E}_2) \cdot \mathbf{T} \\ (\mathbf{T} \times \mathbf{E}_1) \cdot \mathbf{n} \end{bmatrix}. \quad (18)$$

The element thickness d can be determined by solving Eq. (18).

A mesh cloning topology method is proposed for rotationally symmetric structures to predict the acoustic scattering characteristics of multiple propeller blades. This is achieved by applying planar rotational operations to the mesh nodes of a single propeller blade within the same plane, thereby expanding it into a multi-blade mesh. The complete model mesh is then obtained through topological methods. This method is shown in Fig. 4. P_1 , P_2 , and P_3 represent the original mesh vertices, while P'_1 , P'_2 , and P'_3 represent the mesh vertices after topology, with \mathbf{a} representing the spatial vector from the origin to point P_1 , and \mathbf{b} representing the spatial vector from the origin to point P'_1 .

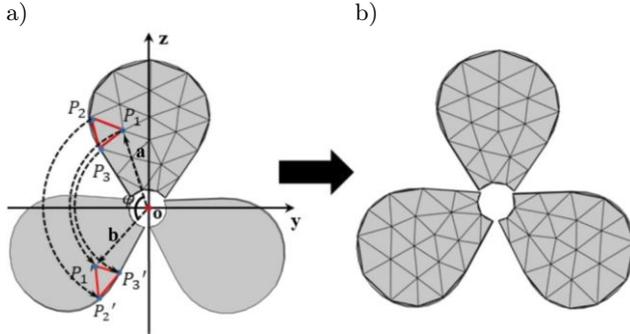


Fig. 4. Mesh cloning topology method: a) shows the topology process; b) shows the mesh after topology.

When cloning a single propeller blade to obtain three blades, the mesh rotates around the axis of 120° . Connect P_1 with the origin O of the coordinate axis to obtain the vector $\mathbf{a} = (y, z)$. The line connecting O and P_2 , which is obtained by rotating P_1 , forms the vector $\mathbf{b} = (y', z')$. By applying Eqs. (19)–(21), the rapid calculation of y' and z' can be realized, thereby obtaining the mesh node information for the three-blade propeller:

$$\cos \varphi = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|}, \quad \varphi = 120^\circ, \quad (19)$$

$$-\frac{1}{2} = \frac{yy' + zz'}{\sqrt{y^2 + z^2} \sqrt{y'^2 + z'^2}}, \quad (20)$$

$$y' = \frac{-y \pm \sqrt{3}z}{2}, \quad z' = \frac{\sqrt{3}y \pm z}{2}. \quad (21)$$

2.4. Research on time-domain echo characteristics of target

In the framework of linear acoustics, the problem of target scattering can be described using an acoustic transfer theory. The target can be regarded as a linear time-invariant network. In this network, the incident signal is the input, and the echo signal is the output. The time-domain transfer function of this linear transfer network is $h(\tau, \mathbf{r}_1, \mathbf{r}_2, \rho)$, where τ is the delay, \mathbf{r}_1 and \mathbf{r}_2 are the radius vectors of the incident and scattering points, respectively, and ρ is the radius vector of the target. The frequency domain transfer function $H(f, \mathbf{r}_1, \mathbf{r}_2, \rho)$ of the transfer network is denoted as the ratio of the scattering wave potential function to the incident wave potential function, where f is frequency. In the monostatic configuration, there is $\mathbf{r}_1 = \mathbf{r}_2$, and the transfer functions in the time-domain and frequency domain are $h(\tau, \mathbf{r}, \rho)$ and $H(f, \mathbf{r}, \rho)$, respectively.

When $x(t)$ represents the time-domain incident signal, $X(f)$ represents the frequency domain incident signal, and $Y(f)$ represents the frequency domain echo signal. The time-domain echo signal of the target can be expressed as:

$$y(t) = x(t) \otimes h(\tau, \mathbf{r}, \rho), \quad (22)$$

where \otimes is convolution operation, then the echo signal in the frequency domain is given by:

$$Y(f) = X(f) \cdot H(f, \mathbf{r}, \rho), \quad (23)$$

$$y(t) = F^{-1}[Y(f)]. \quad (24)$$

The time-domain echo signal can be obtained by taking the inverse Fourier transform of the frequency domain signal. This theory allows for the indirect extraction of the time-domain echo characteristics of underwater targets by calculating the frequency domain signals.

3. Simulation calculation and model

3.1. Composite propeller model

The computational model is a composite propeller of underwater vehicles, comprising a hub and blades. The propeller is made of carbon fiber, with its primary material parameters outlined in Table 1 (JING *et al.*, 2022). The basic dimensional parameters of the propeller are presented in Table 2.

In Table 1, E_1 is the longitudinal elastic modulus; E_2 and E_3 are the transverse elastic moduli; ν_{12} , ν_{23} , and ν_{13} are the Poisson ratios in the 12, 23, and 13 directions, respectively; G_{12} , G_{23} , and G_{13} are the shear moduli in the 12, 23, and 13 directions, respectively; ρ is the density.

Table 1. Material parameter.

Parameter	Numerical value
E_1 [GPa]	140
E_2 [GPa]	9
E_3 [GPa]	9
ν_{12}	0.32
ν_{23}	0.46
ν_{13}	0.32
G_{12} [GPa]	4.6
G_{23} [GPa]	3.08
G_{13} [GPa]	4.6
ρ [kg/m ³]	1620

Table 2. Basic parameters of propeller.

Parameter	Numerical value
Tip circle radius [mm]	261.5
Disk ratio	0.507
Number of blades	3
Propeller rake angle [°]	0
Hub diameter shape	linear type
Hub bore diameter [mm]	67
Hub outside diameter [mm]	81

Due to the unique design of the blade, which feature a larger thickness near the hub and along the midline of the blade surface, and a gradual decrease in thickness towards the blade edges and along the radial direction, the radial thickness variations of the propeller blade model are significant. Therefore, measurements of the sectional profile dimensions at different radial distances of a single blade are conducted. The construction of the blade model and the selected radial positions of the sections are shown in Fig. 5. In the figure, R represents the tip circle diameter.

Using traditional methods to measure propeller dimensions (ZHU, 2020), it is assumed that the generatrix is a curve located at a specific point along the longitudinal direction of the propeller blade. This curve

rotates around the propeller axis while moving axially, thus forming the three-dimensional surface of the propeller blade. The generatrix OH located in the middle of the blade surface is selected as the reference line, serving as the benchmark for measuring the shape and size of the propeller blade. The maximum thickness variation of the blade section is obtained from the blade section projection diagram. Table 3 illustrates the relative positional relationship of points on the blade camber line and the trailing edge line of section at different radii ranging from $0.2R$ to $0.95R$.

Table 3. Single blade section profile dimensions.

r/R	Distance from the generatrix to the leading edge [mm]	Sectional chord length [mm]	Sectional perimeter [mm]	Maximum sectional thickness [mm]
0.2	67.58	135.16	274.41	7.55
0.3	82.27	164.53	332.93	6.45
0.4	97.76	195.51	394.77	5.54
0.5	109.13	218.25	440.80	4.80
0.6	113.25	226.50	458.10	4.20
0.7	110.15	220.30	446.24	3.69
0.8	98.56	197.12	401.30	3.23
0.9	74.26	148.51	302.45	2.80
0.95	52.71	105.42	215.95	2.58
1.0	–	–	–	2.00

Measurements of certain parameters under a single blade’s top and upward view are conducted, with the specific parameters detailed in Table 4.

Table 4. Partial parameters of a single-blade propeller.

Parameter [mm]	Numerical value
a_{max}	206
c_1	10
c_2	2
d_1	88

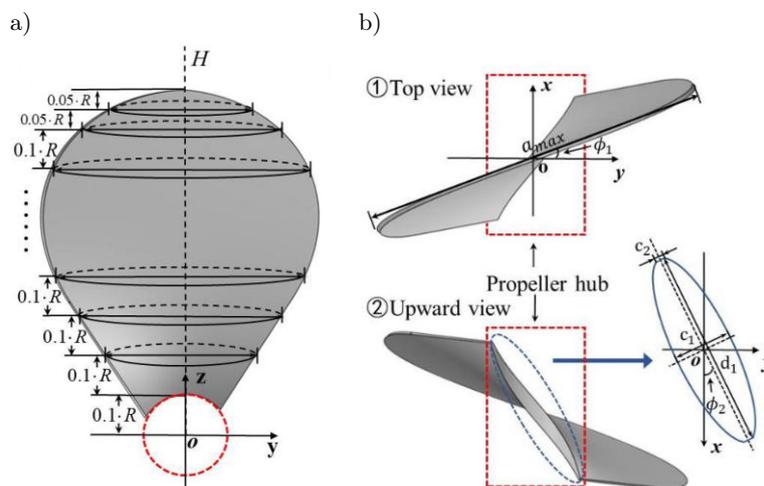


Fig. 5. Single blade structure: a) shows the front view; b) shows the top and upward views.

Measurements are taken of the outer edge of the blade root cross-section in Fig. 6b, combining circular curves and line segments to draw the lower sectional profile, ϕ_1 and ϕ_2 represent the rotational scales of the upper and lower sections, respectively, both set at 20° . By selecting cross-sectional profiles at multiple radial distances, the basic dimensions of the blade's shape are determined, completing the overall propeller blade construction.

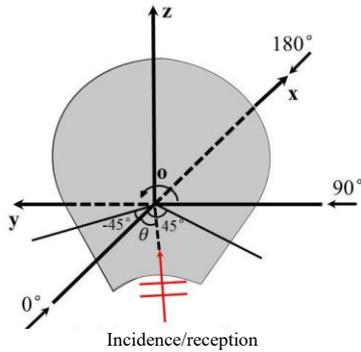


Fig. 6. Schematic diagram of model calculation angle.

3.2. Acoustic scattering characteristics of a single composite propeller blade

The acoustic scattering characteristics of a single composite propeller blade are simulated using finite element software, employing acoustic-solid coupling in the frequency domain for an underwater single composite propeller blade.

In simulations of the scattering acoustic field of a single composite propeller blade using the COMSOL FEM simulation software, the perfect matching layer (PML) technology is introduced to simulate the boundary conditions of an infinite free field. This technology serves as an absorbing boundary, preventing sound waves from reflecting at the interface between the conventional internal computation domain and the PML. This simulation solutions are compared with the KA method. Figure 6 presents the angle setting diagram, the schematic diagram for calculating acoustic scattering of a single composite propeller blade. Figure 7 illustrates a schematic diagram of the FEM calculation model.

The simulation model is established with water as the fluid medium and the incident sound wave

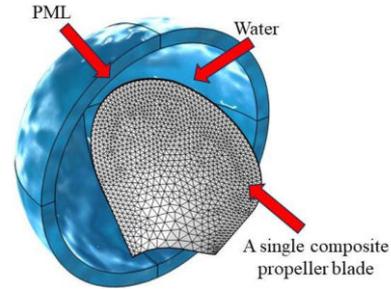


Fig. 7. FEM calculation model of a single composite propeller blade.

as a plane wave. The angle θ between the incident wave in the Oxy plane and the negative half-axis of the x -axis is considered. Two methods, the FEM and the KA, with a thickness variation interval ranging from 2 mm to 7.55 mm, are employed for comparative analysis of the TS. Figure 8 shows the directivity diagrams of TS at different frequencies, with θ ranging from 0° to 360° in 2° intervals (omnidirectional scattering). Figure 9 compares the directivity diagrams of TS at different frequencies, with an incident angle θ of -45° to 45° and a step size of 1° (forward scattering).

Figures 8 and 9 demonstrate that the KA method is largely consistent with the FEM results and offers a significant advantage in computation time. Table 5 provides the times required to compute the TS.

3.3. Acoustic scattering characteristics of the composite propeller

The acoustic scattering characteristics of the complete propeller model, derived from mesh cloning topology of a single propeller blade, are calculated. The TS results for the composite propeller are compared using both the KA and the FEM. The coordinate origin is centered at the propeller hub disk, with the incident wave direction at 0° perpendicular to the propeller hub disk. Calculation angles, θ , are set to range from 0° to 360° in 2° intervals. The comparison of the directivity diagrams of TS is shown in Fig. 10. The overall TS for the composite propeller, which includes multiple blades and the central hub, is obtained through the coherent superposition of the scattered sound fields, with the hub portion considered as a rigid target in the calculations. Table 6 presents the time required to compute the TS.

Table 5. Comparison of time required for TS calculation.

Omnidirectional scattering and forward scattering	FEM/KA	Number of mesh elements	Number of mesh nodes	Calculation time [s]
Omnidirectional scattering	FEM	192303	41910	2150.5
	KA	22396	11200	393.15
Forward scattering	FEM	192303	41910	2034.1
	KA	22396	11200	394.64

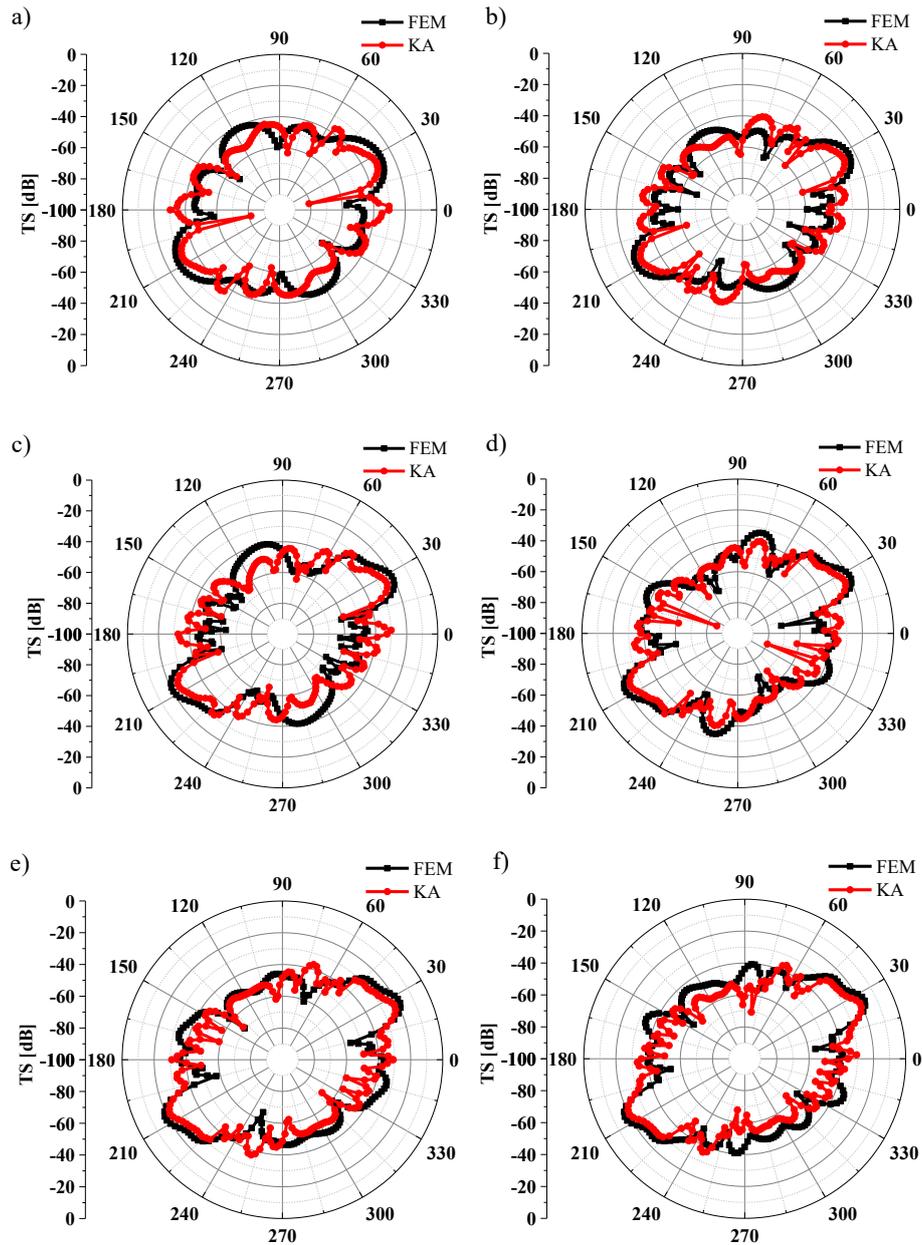


Fig. 8. Comparison of omnidirectional scattering TS for a single composite propeller blade at different frequencies: a) 12 kHz; b) 16 kHz; c) 21 kHz; d) 24 kHz; e) 26 kHz; f) 29 kHz.

Table 6. Comparison of time required for TS calculation.

FEM/KA	Number of mesh elements	Number of mesh nodes	Calculation time [s]
FEM	1503875	292033	16069.8
KA	80867	40645	877.7

As indicated in Fig. 10, the results from the KA are largely consistent with those of the FEM, especially in the forward and aft scattering regions (θ at 0° to 30° and 150° to 210°) at high frequencies. However, at certain oblique incident angles, discrepancies are observed due to the TS derived from the coherent superposition

of the hub and blades, which diverges from the unified calculations of a complete model.

The comparison with the FEM results reveals that for the composite propeller, a rotationally symmetric structure, the directivity diagrams of TS exhibit a rotationally symmetric distribution as well. Peaks oc-

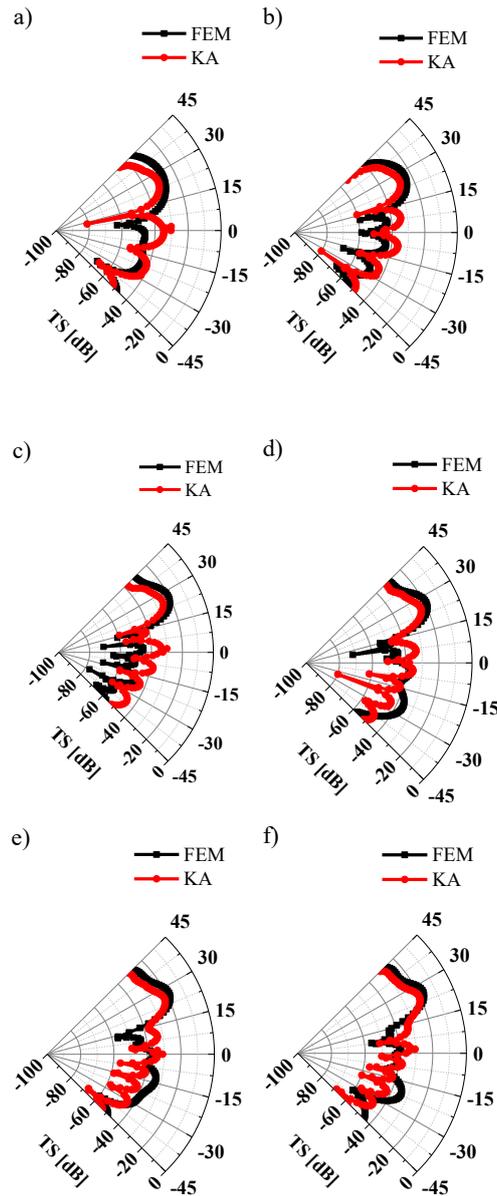


Fig. 9. Comparison of forward scattering TS for a single composite propeller blade at different frequencies: a) 12 kHz; b) 16 kHz; c) 21 kHz; d) 24 kHz; e) 26 kHz; f) 29 kHz.

cur in the 20° to 40° and 200° to 220° ranges, with strong TS regions near 30° and 210° . These are primarily due to the contributions from multiple composite propeller blades, where the unique inclined angle of the blades causes many normal vectors of surface elements to align almost parallel to the incident wave direction at these angles. As Eq. (12) indicates, an increase in the incident wave frequency enhances the reflection coefficient on the composite propeller blade surface, leading to more pronounced directivity within this angular range. Significant fluctuations in TS with multiple peaks are observed in the 0° to 30° , 75° to 105° , 150° to 210° , and 255° to 285° ranges. These are attributed to the contributions from the annular end face of the hub and some bright areas on the blades.

Echo interference from different parts of the composite propeller occurs, and the plane structure acoustic scattering characteristics of the propeller hub disk are significantly influenced by the angle.

3.4. Comparison between composite and steel propellers

The accuracy of the KA method was validated through comparison with the FEM. This was achieved by contrasting TS of the composite propellers with that of the steel propeller and analyzing the acoustical modulation effect of composite materials. Frequency computations were conducted within the range of 10 kHz to 30 kHz with a step size of 200 Hz, while an-

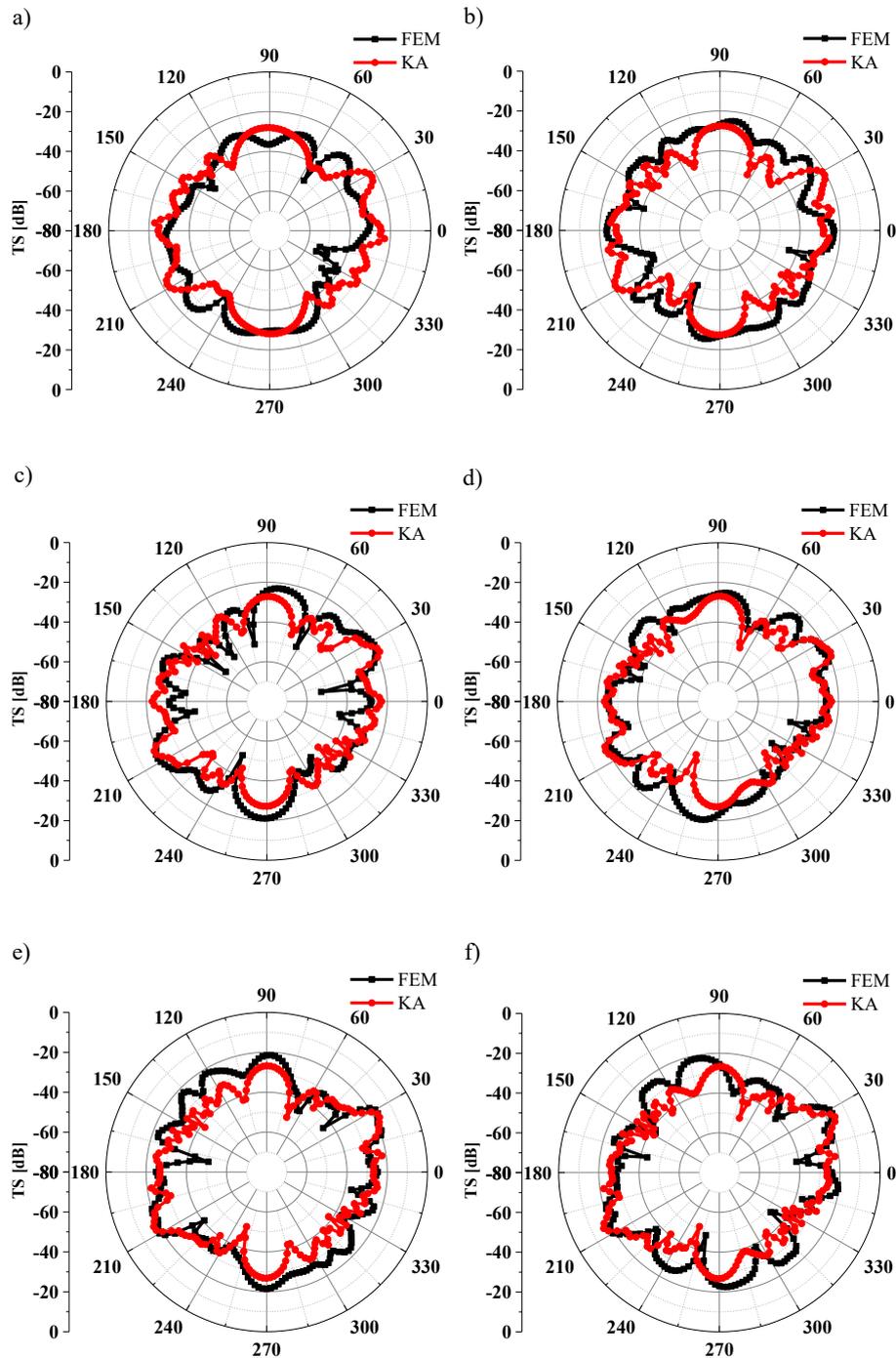


Fig. 10. Comparison of TS for composite propellers at different frequencies: a) 12 kHz; b) 17 kHz; c) 20 kHz; d) 21 kHz; e) 23 kHz; f) 24 kHz; g) 27 kHz; h) 29 kHz.

gle variations were assessed from 0° to 360° with a step size of 2° . In Fig. 11, the horizontal axis represents the incident angle of the sound wave, which the vertical axis denotes the incident frequency of the sound wave.

According to Figs. 11 and 12, compared to the steel propeller, the composite propeller exhibits a significant decrease in TS predominantly within the angular range of 15° to 45° and 195° to 225° . This reduction is

attributed to the parallel alignment between the normal vectors of discrete elements and the direction of sound waves within these angular intervals, which enhances the reflection coefficient of the composite materials, thereby facilitating effective acoustical modulation. However, with an increase in frequency, leading to the reduction in the wavelength of the sound waves, the extent of reduction in TS diminishes.

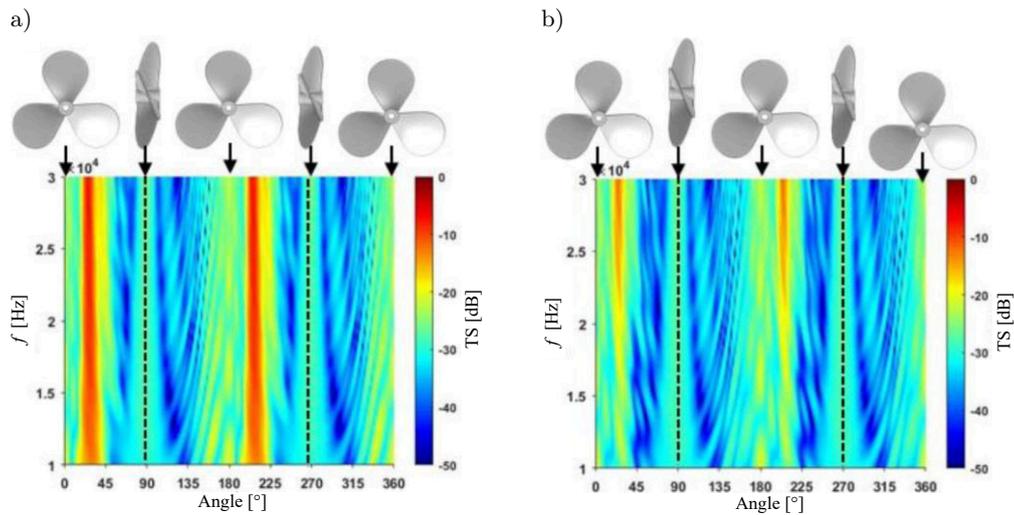


Fig. 11. Comparison of TS: a) shows TS of the steel propeller; b) shows TS of the composite propeller.

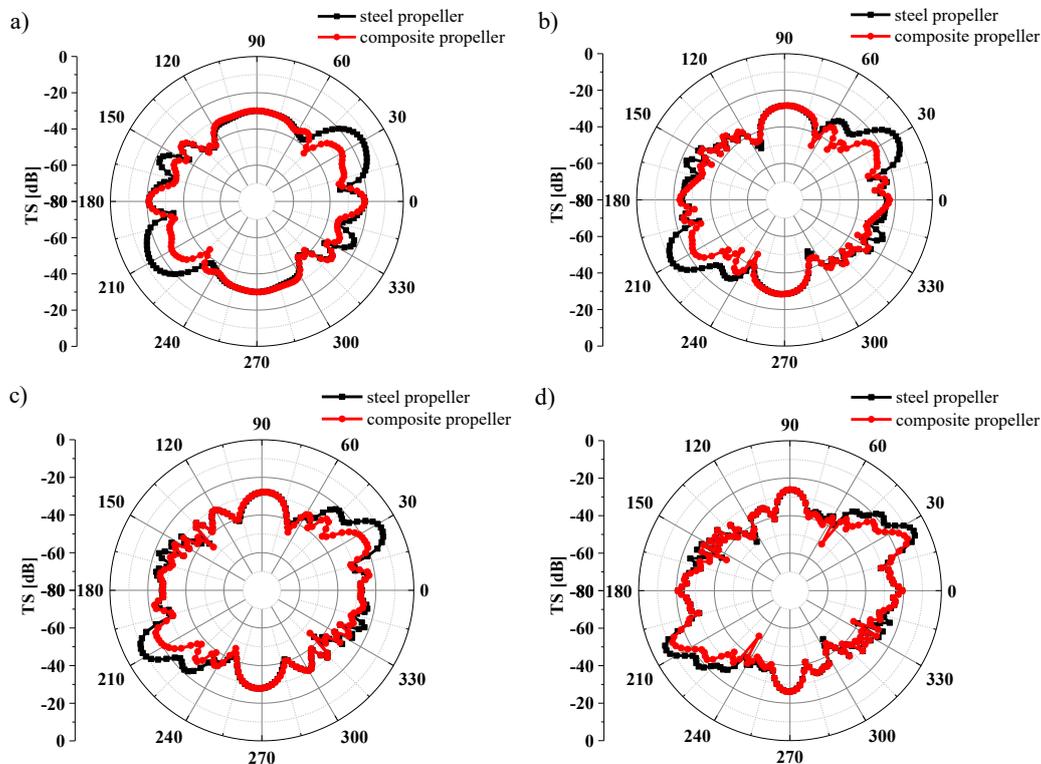


Fig. 12. Comparison of TS between steel propeller and composite propeller at different frequencies: a) 10 kHz; b) 15 kHz; c) 20 kHz; d) 30 kHz.

4. Comparison of acoustic scattering characteristics between the KA solutions and experimental results

4.1. Acoustic scattering characteristic experiment instrument and arrangement

Due to the demanding processing requirements and lengthy production cycles associated with composite material propellers, steel propellers, which are isotropic

and can be considered as special cases of orthotropic materials, offer a more universal option. Verification of the KA method suffices for these materials. Thus, a lake experiment on scaled models of steel propellers was conducted to verify this approach. By analyzing the echo characteristics of the steel propeller in relation to variations in incident angles and frequencies, this experiment provides essential data to support the accuracy of predictive methods for TS in the thin plate target with variable thickness.

The primary material parameters used include a density of 7800 kg/m^3 , Poisson’s ratio of 0.3, and Young’s modulus of $2.1 \times 10^{11} \text{ Pa}$. Figure 13 shows the steel propeller scale model.

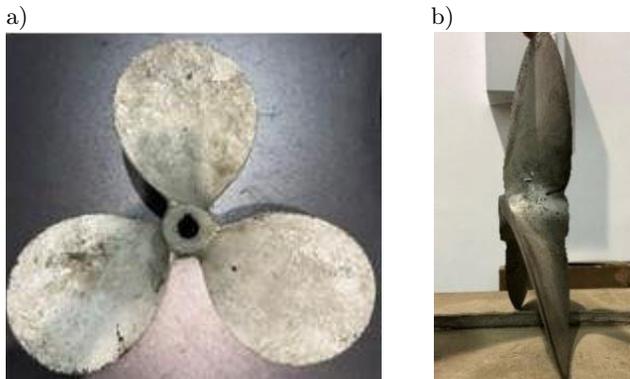


Fig. 13. Steel propeller experiment model: a) shows the front view; b) shows the side view.

Figure 14 is a schematic diagram of the experiment equipment arrangement. The transducer, hydrophone, and detection target are positioned at the same depth. This depth, measured from the water surface, is 7.38 m. The distance between the transducer

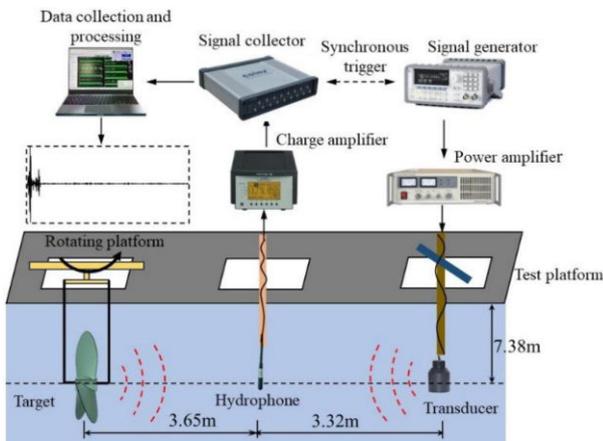


Fig. 14. Schematic diagram of the experiment equipment arrangement.

and the hydrophone is 3.65 m, while the distance between the hydrophone and the measurement target is 3.32 m. During the measurement process, the propeller rotates uniformly underwater in a suspended posture, with a rotation angle ranging from 0° to 360° .

4.2. Experimental data analysis of the time-domain echo characteristics of the steel propeller scale model

In the experiment, a scaled steel propeller model was utilized. Incident acoustic waves were directed at the model across a full 360° range. These signals were linear frequency modulated pulses, spanning a frequency band of 10 kHz to 20 kHz with a pulse width of 3 ms. The geometric center of the incidence setup was aligned with the hub disk center of the propeller, ensuring that the incident wave vector faced the center of the propeller hub disk directly at 0° incidence. The time-domain echo characteristics obtained from the experiment were then compared with the KA solutions as shown in the Fig. 15. The horizontal axis represents the attitude azimuth angle of the model, which the vertical axis denotes the the moment of the echo.

From Fig. 15, it is evident that strong echo bright spots are pronounced near the 30° and 210° . Both the experimental findings and the KA solutions reveal a similar distribution pattern of these bright spots, indicating a close correlation in their variation trends. This similarity suggests that the propeller blades’ side inclination angle plays a crucial role in the echo characteristics, especially when analyzing the time-domain acoustic scattering characteristics of the steel propeller. However, some discrepancies between the experimental data and the KA method predictions persist. These differences might stem from small size of the model, which could lead to weaker echo signals at the receiving point. Additionally, experimental limitations may have hindered accurate observation of the underwater model’s posture, potentially causing instability at certain angles and, consequently, inaccuracies in the results.

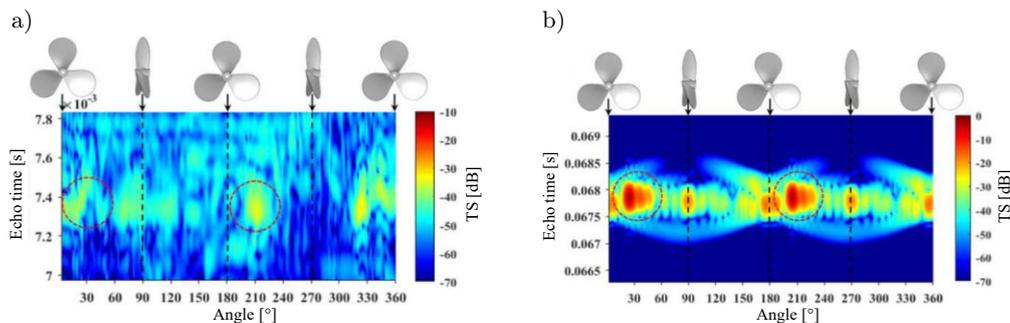


Fig. 15. Comparisons of the echo characteristics of the steel propeller scale model in the frequency band of 10 kHz–20 kHz between the experimental results and KA solutions: a) shows the experimental results; b) shows the KA solutions.

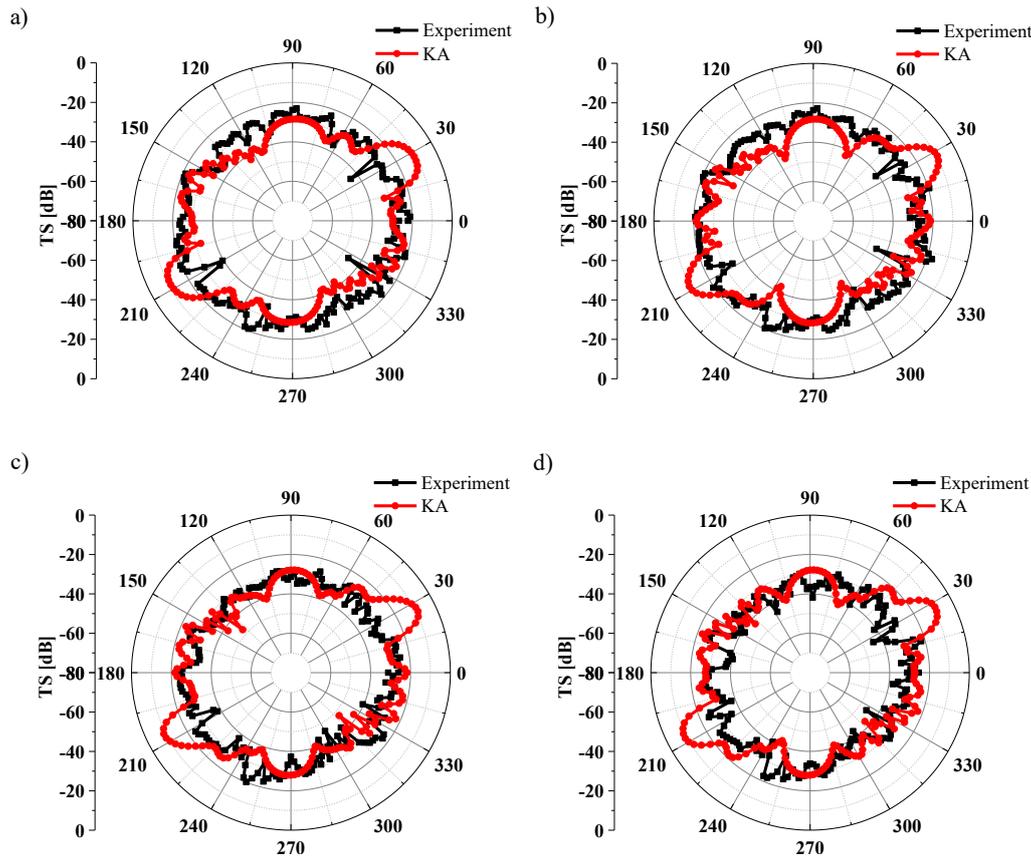


Fig. 16. The comparison results of the steel propeller TS at different frequencies:
a) 14 kHz; b) 16 kHz; c) 18 kHz; d) 20 kHz.

4.3. The steel propeller scale model monostatic TS experimental data analysis

The experimental data underwent the Fourier transform, and the outcomes of at different frequencies were then compared with those derived from the KA method.

As shown in Fig. 16, the experimental results show a trend that closely matches the KA solutions, with the numerical values being largely consistent. This indicates that the KA method has a reliable degree of accuracy in predicting the TS of propellers. However, it is observed that there are some differences in the peak values near the strong intensity points at 30° and 210° between experimental results for the propeller and the KA solutions. These discrepancies could potentially be attributed to issues related to fabrication and suspension of the model.

5. Conclusion

This study initially employs an IA to determine the thickness of each discrete surface element. Subsequently, based on material parameters, OLTM is utilized to calculate the reflection coefficient of each dis-

crete surface element. Finally, employing a coherent superposition approach, the scattered acoustic field of a single composite propeller blade is obtained. To streamline computations, a mesh topology cloning method is employed to predict the acoustic scattering characteristics of the entire composite propeller. Comparative analysis of the acoustic scattering characteristics of the composite propeller yielded the following key conclusions:

- 1) Compared to the traditional FEM, the KA method is shown to rapidly and accurately predict the high-frequency acoustic scattering characteristics of composite thin-plate models with spatially varying thickness.
- 2) For geometric models that exhibit rotational symmetry, complete multi-body models can be obtained by topologically cloning individual structures in the mesh. This method effectively predicts the acoustic scattering characteristics, with these typical structures displaying a consistent distribution pattern of rotational symmetry in their directivity diagrams of TS.
- 3) The propeller blade significantly affects the TS of the static composite propeller. An overall TS peak is observed when a considerable number of surface

elements on the propeller blade have an outward normal nearly parallel to the incident direction.

- 4) Compared to the acoustic scattering results from lake experiments, this study demonstrates that the applying the KA method to the isotropic steel propeller also ensures the high solution accuracy. This enhances the prediction accuracy of the stern direction acoustic scattering characteristics of underwater vehicles.

Acknowledgments

This work is sponsored and supported by the National Natural Science Fund of China (no. 52201397 and 52001211) and the Hunan Province Natural Science Fund (no. 2023JJ60544).

References

1. CHENG Z.K. *et al.* (2023), Observation of the rotational Doppler shift of a spinning object based on an acoustic vortex with a Fresnel-spiral zone plate, *Journal of Applied Physics*, **133**(11): 114502, doi: [10.1063/5.0141106](https://doi.org/10.1063/5.0141106).
2. CHU D., STANTON T.K. (2010), Statistics of echoes from a directional sonar beam insonifying finite numbers of single scatterers and patches of scatterers, *IEEE Journal of Oceanic Engineering*, **35**(2): 267–277, doi: [10.1109/JOE.2009.2037988](https://doi.org/10.1109/JOE.2009.2037988).
3. FAN J., TANG W.L., ZHUO L.K. (2012), Planar elements method for forecasting the echo characteristics from sonar targets [in Chinese], *Journal of Ship Mechanics*, **16**(1): 171–180, doi: [10.3969/j.issn.1007-7294.2012.01.020](https://doi.org/10.3969/j.issn.1007-7294.2012.01.020).
4. HU P. (2017), *Research on target strength of submarine sails made by sound-reflecting composites*, MSc. Thesis, China Ship Research and Development Academy.
5. ISAKSON M.J., CHOTIROS N.P. (2014), Finite element modeling of acoustic scattering from fluid and elastic rough interfaces, *IEEE Journal of Oceanic Engineering*, **40**(2): 475–484, doi: [10.1109/JOE.2014.2313060](https://doi.org/10.1109/JOE.2014.2313060).
6. ISLAM F., CALDWELL R., PHILLIPS A.W., ST JOHN N.A., PRUSTY B.G. (2022), A review of relevant impact behaviour for improved durability of marine composite propellers, *Composites Part C: Open Access*, **8**: 100251, doi: [10.1016/j.jcomc.2022.100251](https://doi.org/10.1016/j.jcomc.2022.100251).
7. JIANG B., YU J., LI W., CHAI Y., GUI Q. (2023), A coupled overlapping finite element method for analyzing underwater acoustic scattering problems, *Journal of Marine Science and Engineering*, **11**(9): 1676, doi: [10.3390/jmse11091676](https://doi.org/10.3390/jmse11091676).
8. JING J., HU Y.F., DING G.P. (2022), Simulation research on modal analysis of CFRP propeller, *Digital Manufacture Science*, **20**(1): 29–33.
9. KLAUSNER N., AZIMI-SADJADI MR. (2014), Non-Gaussian target detection in sonar imagery using the multivariate laplace distribution, *IEEE Journal of Oceanic Engineering*, **40**(2): 452–464, doi: [10.1109/JOE.2014.2328211](https://doi.org/10.1109/JOE.2014.2328211).
10. KUO Y.-M., LIN H.-J., WANG C.-N. (2008), Sound transmission across orthotropic laminates with a 3D model, *Applied Acoustics*, **69**(11): 951–959, doi: [10.1016/j.apacoust.2007.08.002](https://doi.org/10.1016/j.apacoust.2007.08.002).
11. KWON H.W., HONG S.Y., SONG J.H. (2017), A study for acoustic target strength characteristics of submarines using Kirchhoff approximation, *Marine Technology Society Journal*, **51**(4): 52–58, doi: [10.4031/MTSJ.51.4.5](https://doi.org/10.4031/MTSJ.51.4.5).
12. LANGDON S., CHANDLER-WILDE S.N. (2006), A wave-number independent boundary element method for an acoustic scattering problem, *SIAM Journal on Numerical Analysis*, **43**(6): 2450–2477, doi: [10.1137/S0036142903431936](https://doi.org/10.1137/S0036142903431936).
13. LAVIA E.F., GONZALEZ J.D., BLANC S. (2019), Modeling high-frequency backscattering from a mesh of curved surfaces using Kirchhoff approximation, *Journal of Theoretical and Computational Acoustics*, **27**(04): 1850057, doi: [10.1142/S2591728518500573](https://doi.org/10.1142/S2591728518500573).
14. LI D. (2022), *Analysis of Composite Laminates: Theories and Their Applications*, Science Press, China.
15. LIN H.-J., WANG C.-N., KUO Y.-M. (2017), Sound transmission loss across specially orthotropic laminates, *Applied Acoustics*, **68**(10): 1177–1191, doi: [10.1016/j.apacoust.2006.06.007](https://doi.org/10.1016/j.apacoust.2006.06.007).
16. MÖLLER T., TRUMBORE B. (1997), Fast, minimum storage ray-triangle intersection, *Journal of Graphics Tools*, **2**(1): 21–28, doi: [10.1145/1198555.1198746](https://doi.org/10.1145/1198555.1198746).
17. MOTLEY M.R., LIU Z., YOUNG Y.L. (2009), Utilizing fluid–structure interactions to improve energy efficiency of composite marine propellers in spatially varying wake, *Composite Structures*, **90**(3): 304–313, doi: [10.1016/j.compstruct.2009.03.011](https://doi.org/10.1016/j.compstruct.2009.03.011).
18. SABAT R. *et al.* (2023), Low frequency sound isolation by a metasurface of Helmholtz ping-pong ball resonators, *Journal of Applied Physics*, **134**(14): 144502, doi: [10.1063/5.0160267](https://doi.org/10.1063/5.0160267).
19. SAFFARI A., ZAHIRI S.H., GHANAD N.K. (2023), Using SVM classifier and micro-Doppler signature for automatic recognition of sonar targets, *Archives of Acoustics*, **48**(1): 49–61, doi: [10.24425/aoa.2022.142909](https://doi.org/10.24425/aoa.2022.142909).
20. SAGAR M.V., VENKAIAH M., SUNIL D. (2013), Static and dynamic analysis of composite propeller of ship using FEA, *International Journal of Engineering Research & Technology (IJERT)*, **2**(7): 2587–2594, doi: [10.17577/IJERTV2IS70418](https://doi.org/10.17577/IJERTV2IS70418).
21. SEYBERT A.F., WU T.W., WU X.F. (1988), Radiation and scattering of acoustic waves from elastic solids and

- shells using the boundary element method, *The Journal of the Acoustical Society of America*, **84**(5): 1906–1912, doi: [10.1121/1.397156](https://doi.org/10.1121/1.397156).
22. TANG W.L. (1993), Calculation of acoustic scattering of a nonrigid surface using physical acoustic method [in Chinese], *Acta Acustica*, **18**(1): 45–53, doi: [10.15949/j.cnki.0371-0025.1993.01.006](https://doi.org/10.15949/j.cnki.0371-0025.1993.01.006).
23. TUCKER J.D, AZIMI-SADJADI M.R. (2011), Coherence-based underwater target detection from multiple disparate sonar platforms, *IEEE Journal of Oceanic Engineering*, **36**(1): 37–51, doi: [10.1109/JOE.2010.2094230](https://doi.org/10.1109/JOE.2010.2094230).
24. UDDIN M.M., HOSSEN M.P., JAHAN M.M., ISLAM M.I. (2021), Structural analysis of composite propeller of ship using FEM, [in:] *AIP Conference Proceedings*, **2324**(1): 030001, doi: [10.1063/5.0037760](https://doi.org/10.1063/5.0037760).
25. VARDHAN D.H., RAMESH A., REDDY B.C.M. (2019), A review on materials used for marine propellers, *Materials Today: Proceedings*, **18**(7): 4482–4490, doi: [10.1016/j.matpr.2019.07.418](https://doi.org/10.1016/j.matpr.2019.07.418).
26. VENÅS J.V., KVAMSDAL T. (2020), Isogeometric boundary element method for acoustic scattering by a submarine, *Computer Methods in Applied Mechanics and Engineering*, **359**: 112670, doi: [10.1016/j.cma.2019.112670](https://doi.org/10.1016/j.cma.2019.112670).
27. YANG F., PENG Z., SONG H., TANG Y., MIAO X. (2024), A hybrid finite element method – Kirchhoff approximation method for modeling acoustic scattering from an underwater vehicle model with Alberich coatings with periodic internal cavities, *Archives of Acoustics*, **49**(2): 209–219, doi: [10.24425/aoa.2024.148777](https://doi.org/10.24425/aoa.2024.148777).
28. ZHANG X.G., ZHANG J.F., LYU S., CAO F., LI G.J., WANG N. (2020), Rapid method for large-scale target sound scattering calculation and experiment validation [in Chinese], *Journal of Ship Mechanics*, **24**(3): 409–418.
29. ZHU X.C. (2020), *Parametric modeling and CNC machining of ship propeller*, MSc. Thesis, Tianjin, Tian-gong University.

Research Paper

Sound Insulation of an Acoustic Barrier with Layered Structures of Sonic Crystals – Comparative Studies of Physical and Theoretical Models

Jan RADOSZ 

Central Institute for Labour Protection – National Research Institute
Warsaw, Poland; e-mail: jarad@ciop.pl

(received March 2, 2022; accepted June 11, 2024; published online September 26, 2024)

Recent years have seen a growing interest in the potential for the use of sonic crystals as noise barriers. The frequencies with the highest attenuation can be determined by assuming that an integer number of half wavelengths fits the distance between the scatterers. However, this approach limits the usefulness of sonic crystals as a viable noise barrier technology, as it necessitates a significant increase in the overall crystal size to cover a broader frequency range for noise reduction. Based on developed theoretical models, geometrical assumptions were made for the physical models of the acoustic barrier in terms of the materials used and the dimensions of structural elements. Three physical models were developed to verify the design intent. The method involved measuring the transmission loss (TL) and insertion loss (IL) of the sonic crystal structure and comparing these results with theoretical models. The aim of this work was to perform free-field measurements on a real-sized sample in order to verify the strengths and weaknesses of applying layered structures of sonic crystals based on calculations and measurements. The results of the conducted measurements showed satisfactory noise reduction by the developed physical models for key components of the analysed spectrum. It was also demonstrated that layered structures of sonic crystals can achieve greater noise reduction (up to 3.5 dB) and a wider frequency range of attenuation (up to the range of 2000 Hz–5000 Hz) compared to single-layer structures.

Keywords: noise; barrier; sonic crystal; diffuser.



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

Sonic crystals have received significant attention from the acoustical community over the past decade because of their unique acoustic properties. The ability of sonic crystals to function as stop-band filters in the audible frequency range, i.e., to attenuate waves within frequency bands known as band gaps, is the most attractive and extensively studied property of such materials. Further advantages of sonic crystal barrier in comparison to traditional sound barriers, include their ability to allow light to pass through and, uniquely, their non-obstruction of the free flow of air. The existence of band gaps in sonic crystals was demonstrated in early works (MARTINEZ-SALA *et al.*, 1995; RUBIO *et al.*, 1999; SÁNCHEZ-PÉREZ *et al.*, 1998). However, barriers made from these “conventional” sonic crystals suffer from the major disadvantage of providing attenuation only within a relatively

narrow frequency band and are therefore unsuitable as barriers for broad-band noise attenuation. To enhance the sound insulation properties of sonic crystals, researchers have recently focused on systems in which both Bragg scattering and local resonant phenomena are present (ELFORD *et al.*, 2011; FUSTER-GARCIA *et al.*, 2007; GOFFAUX, SÁNCHEZ-DEHESA, 2003; HIRSEKORN *et al.*, 2004; HO *et al.*, 2003; LIU *et al.*, 2000; ROMERO-GARCIA *et al.*, 2013; HU *et al.*, 2005; CASTIÑEIRA-IBAÑEZ *et al.*, 2012). These investigations showed that periodic arrays of scatterers composed of a small number of elements are capable of achieving sound attenuation values large enough to compete with other acoustic barriers.

To measure the effective screening effect of a barrier, MORANDI *et al.* (2016) proposed using transient sound signals and a suitable windowing technique, which is now standardised in EN 1793-6 (CEN, 2012). The paper aims to conduct laboratory testing of scat-

terers to experimentally determine their acoustic performance. According to two studies (MORANDI *et al.*, 2015; 2016), the availability of standardised values allows a direct comparison of the sound insulation and reflection properties of the sonic crystal noise barriers related to other classical. However, the results of measurements presented in the paper indicate that the number of measurement points according to EN 1793-6 (CEN, 2012) may not be sufficient.

In previous research (RADOSZ, 2019), the author explored the potential of sonic crystals as noise barriers, highlighting their ability to create band gaps that attenuate sound waves. The study focused on a multiple-resonance band gap system to enhance sound attenuation properties, using six concentric *C*-shaped resonators. However, the research identified limitations in the frequency range of attenuation, primarily due to the specific lattice parameters and packing fraction of the sonic crystals used. Further research is needed to evaluate layered structures of sonic crystals in terms of increasing sound insulation with a possible synergy effect of layers.

2. Materials and methods

The study is organised as follows. According to data obtained from measurements of a compressor unit (Fig. 1), single unit cells are studied by means of finite

element (FE) Bloch-type analyses in order to investigate the band structure of the unit cells to identify those capable of generating band gaps.

2.1. Analyses of band structures

Numerical analyses were performed to design a basic sonic crystal single-layer structure exhibiting a complete band gap within the frequency range in which the compressor noise spectrum shows a prominent peak (see Fig. 1). The lattice constant is set in accordance to Bragg scattering theory, looking for a band gap at approximately 2900 Hz and 4500 Hz corresponding to the FFT spectrum of the compressor unit (see Fig. 1). The resulting lattice constants for layers are 38 mm (A-1) and 60 mm (A-2), given $c_{\text{air}} = 343$ m/s, the speed of sound in air at 20 °C. PVC pipes with radii of 28 mm (A-1) and 44 mm (A-2) were considered as inclusion placed at the centre of the unit cell. The properties of the materials used in the calculations are listed in Table 1.

Table 1. Properties of material and medium used in the FE analyses.

Material/ medium	Density ρ [kg/m ³]	Longitudinal wave c_L [m · s ⁻¹]	Shear wave speed c_S [m · s ⁻¹]
Air	1.25	343	–
PVC	1400	2142	874

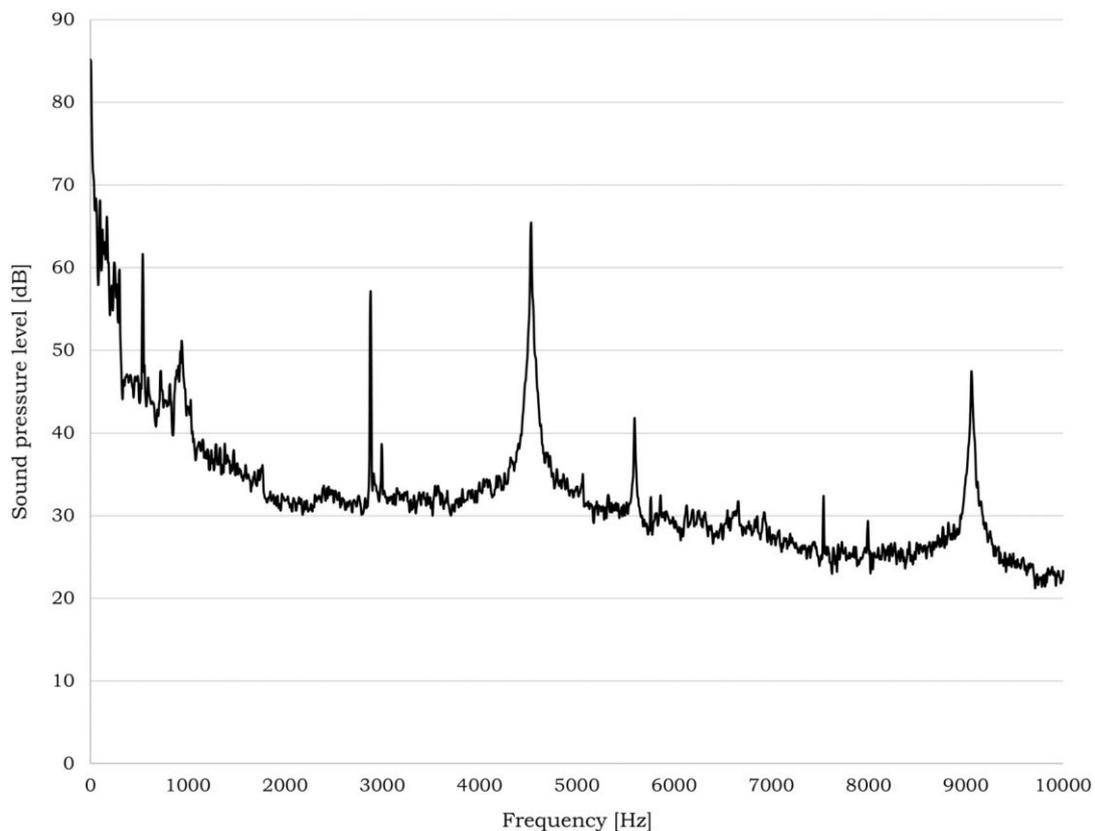


Fig. 1. Fast Fourier transform (FFT) spectrum at the measuring point at the inlets of an industrial compressor unit.

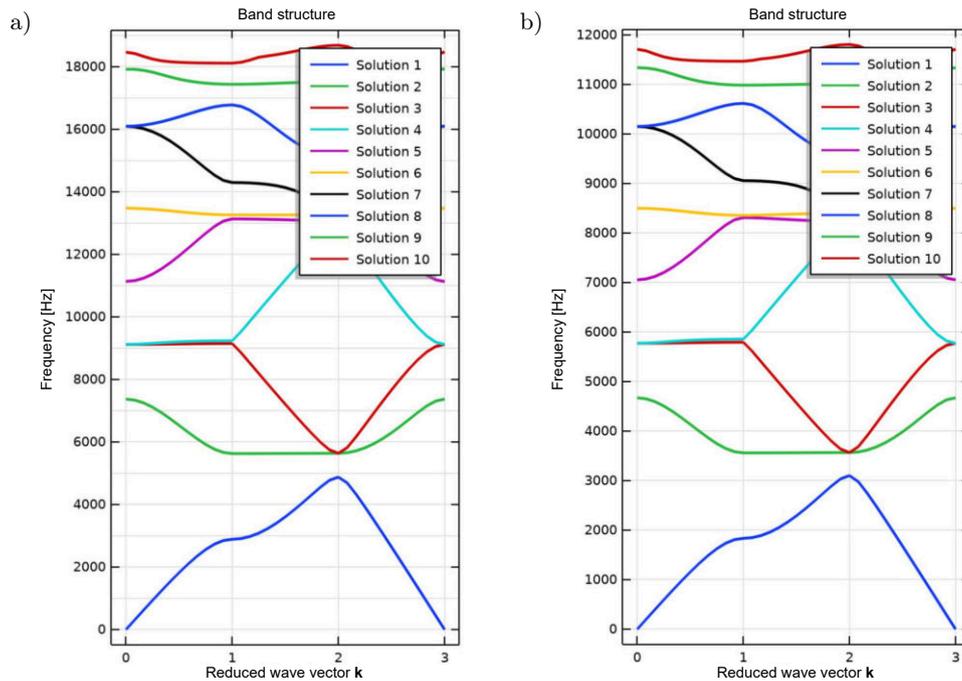


Fig. 2. Band structures for air-PVC pipe unit cell in the first irreducible Brillouin zone (for ten eigenfrequencies solutions): a) model A-1; b) model A-2.

The band structure was computed along the three high-symmetry directions of the first irreducible Brillouin zone ΓX , XM , and $M\Gamma$ using the plane wave expansion (PWE) method with the use of MATLAB software. Figure 2 presents the band structures in terms of the reduced wave vector $\mathbf{k} = \left[\frac{k_x a}{\pi}, \frac{k_y a}{\pi} \right]$, where \mathbf{k}_x and \mathbf{k}_y are the wave vectors in the x and y directions, respectively.

2.2. Design assumptions of physical models of the acoustic barrier

It was assumed that the three physical models proposed, based on selected theoretical models, are de-

signed using commonly available materials with standardised dimensions (pipes, profiles, etc.). The physical model is made from PVC pipes (diffusers) with a density of $\rho = 1400 \text{ kg/m}^{-3}$. The foundations of the physical models are made from MDF boards. The dimensions of the physical models are shown in Figs. 3–5.

2.3. Measurement method

The measurements of transmission loss (TL) were carried out using a laboratory stand, the scheme of which is shown in Fig. 6. According to the adopted test method, the stand was located in an anechoic chamber in order to ensure acoustic conditions similar to those

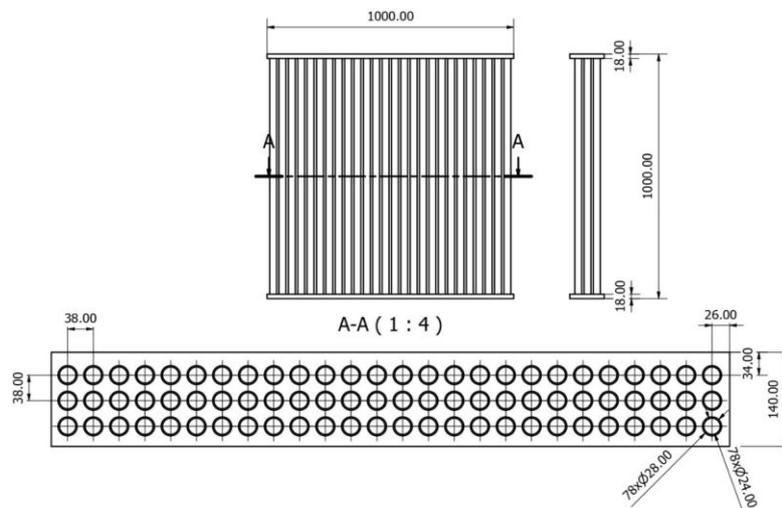


Fig. 3. Dimensions of the A-1 physical model of the acoustic barrier.

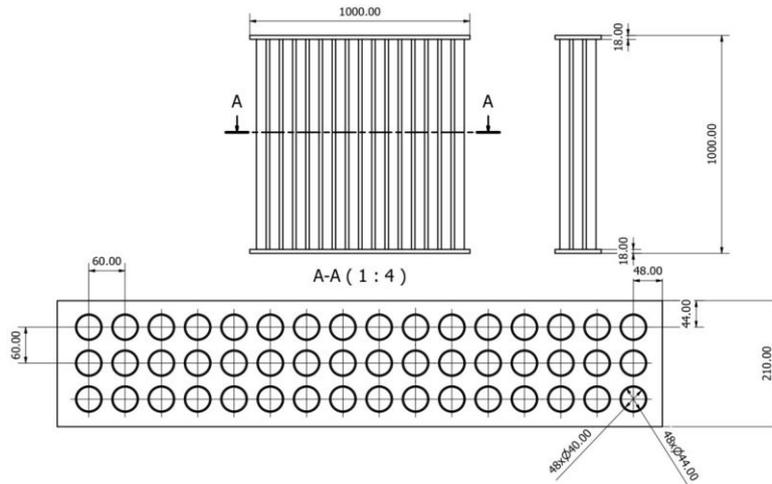


Fig. 4. Dimensions of the A-2 physical model of the acoustic barrier.

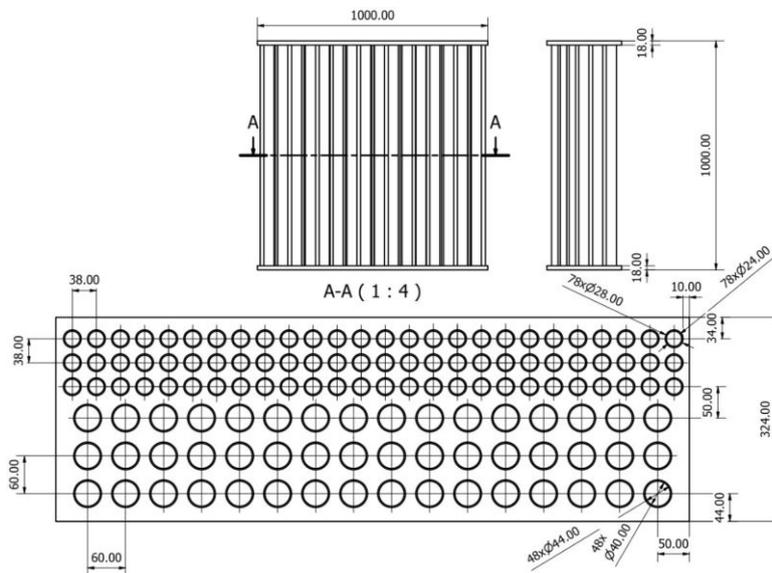


Fig. 5. Dimensions of the A-3 physical model of the acoustic barrier.

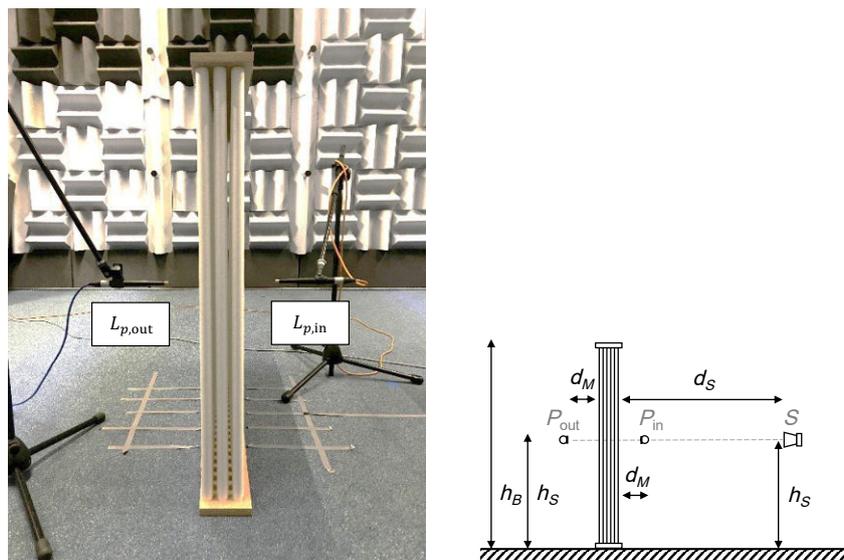


Fig. 6. Scheme of the laboratory stand for measuring TL: S – sound source; P_{in} , P_{out} – measurements points; $h_S = 0.5$ m; $h_B = 1$ m; $d_M = 0.1$ m; $d_S = 1$ m.

of a free field, which were used in the theoretical model tests.

The tested physical models of the acoustic barrier were situated on the floor structure. The positioning of the sound source and microphones in relation to the physical model corresponded to the setup adopted in the computational model.

The sound source (Bose S1 PRO) was placed at a distance of 100 cm from the model surface, with its main radiation axis perpendicular to the surface and passing through the centre of the model.

Microphones (DPA 4007) were placed successively at five points, spaced 10 cm apart, on the horizontal centreline of the module, at a distance of 10 cm from the model's surface.

The values of sound pressure levels for 1/3 octave bands were calculated from the impulse responses obtained with the use of maximum length sequence (MLS) signal (DIRAC software). Then, the values of sound TL were determined according to the following equation:

$$TL = \overline{L_{p,in}} - \overline{L_{p,out}} \text{ [dB]},$$

where $\overline{L_{p,in}}$ is the mean sound pressure level from five measurement points in front of the barrier, and $\overline{L_{p,out}}$ is the mean sound pressure level from five measurement points behind the barrier.

The values of insertion loss (IL) were determined based on the measured values of the sound pressure levels in the space behind the screen $\overline{L_{p,1}}$ (mean value from five measurement points) and in the space without the acoustic barrier $\overline{L_{p,2}}$ (mean value from five measurement points), assuming the same distances as specified in the EN 1793-6 (CEN, 2012) (Fig. 7).

Figure 8 shows the highest values of the repeatability standard deviation for three measurements performed at each measurement point using the impulse response with the MLS signal. These values did not exceed 0.022 dB, which indicates a high repeatability of the measurement method used.

2.4. Measurement results

Figures 9–11 present a comparison of TL (calculated as the average from five measurement points) between theoretical models and physical models of the acoustic barrier. The measurement results showed that the occurrence of the band gap was consistent with the results of theoretical calculations for the key spectrum bands resulting from Bragg's law.

For the case of the A-1 physical model (measurements), the highest attenuation was obtained in the third octave band with a centre frequency of 4000 Hz. The measured TL was 19.7 dB, which differs by 0.7 dB

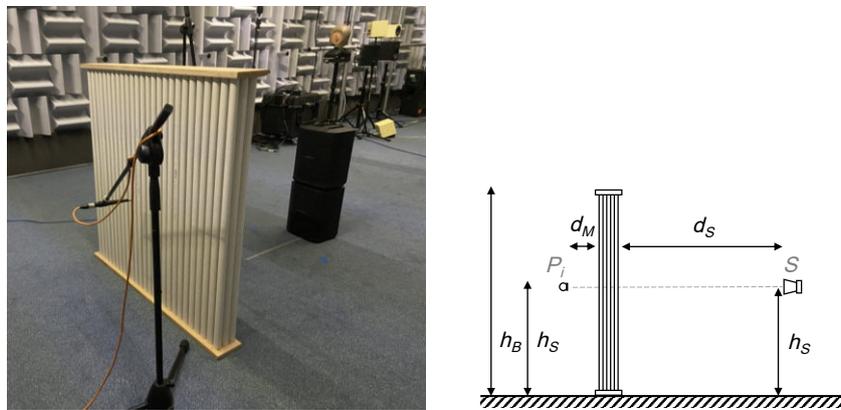


Fig. 7. Scheme of the laboratory stand for measuring IL: S – sound source; P_i – measurements points; $h_S = 0.5$ m; $h_B = 1$ m; $d_M = 0.25$ m; $d_S = 1$ m.

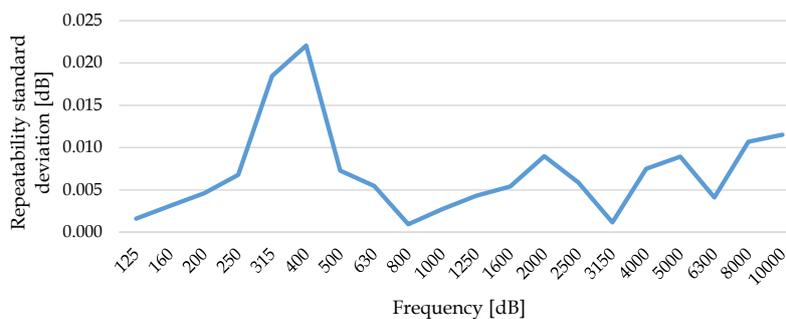


Fig. 8. Repeatability standard deviation of the measurements using the impulse response with the MLS signal (highest value from five measurement points).

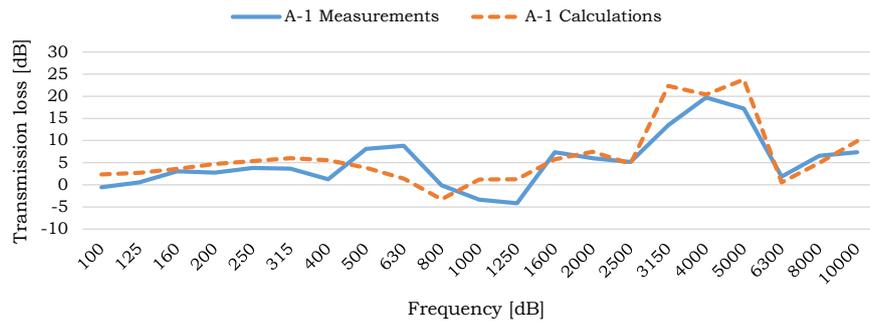


Fig. 9. Comparison of TL results for A-1 model.

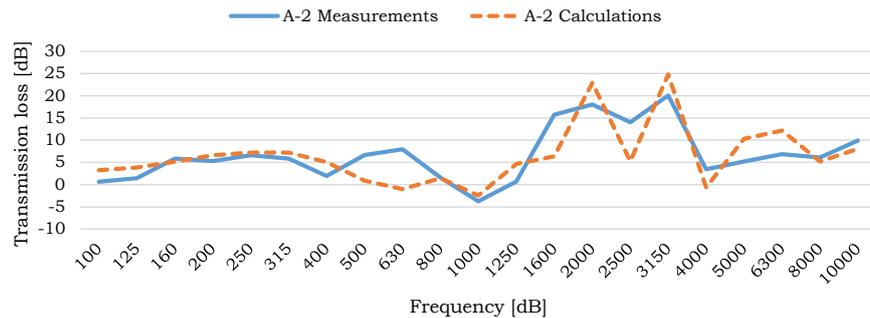


Fig. 10. Comparison of TL results for A-2 model.

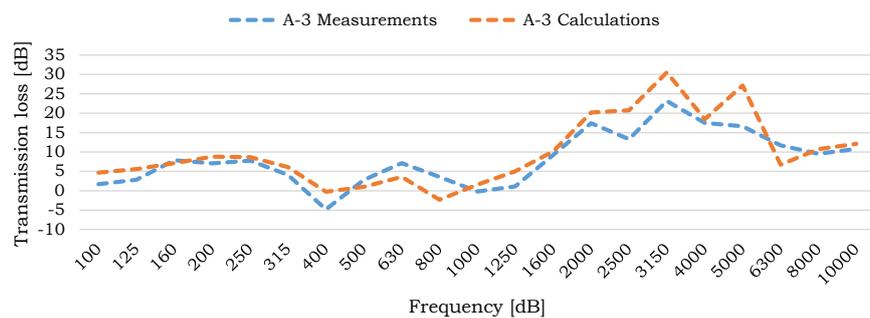


Fig. 11. Comparison of TL results for A-3 model.

from the theoretical model (calculations) for that band (Fig. 9).

For the A-2 physical model, the highest attenuation was achieved in the $\frac{1}{3}$ octave bands with centre frequencies of 2000 Hz and 3150 Hz. The measured TLs were 18.0 dB and 20.1 dB, respectively (Fig. 10). The differences from the theoretical model for these bands were 4.8 dB and 4.7 dB.

In the case of the A-3 physical model, the highest attenuation was achieved in the $\frac{1}{3}$ octave bands with centre frequencies ranging from 2000 Hz to 5000 Hz. The measured TLs varied from 13.3 dB to 23.2 dB (Fig. 11). The differences with the theoretical model in this range were between 1.0 dB and 10.6 dB.

Figures 12–14 present a comparison of IL (calculated as the average from five measurement points) between theoretical models and physical models of the acoustic barrier. The results of the measurements showed that the occurrence of the band gap was con-

sistent with the results of theoretical calculations for the key spectrum bands predicted by Bragg's law.

In the case of the A-1 model, the highest attenuation was observed in the $\frac{1}{3}$ octave bands with centre frequencies ranging from 4000 Hz to 5000 Hz. The measured ILs were 14.7 dB and 15.5 dB, respectively (Fig. 12). The differences from the theoretical model for the $\frac{1}{3}$ octave bands with centre frequencies of 4000 Hz and 5000 Hz were 4.1 dB and 0.6 dB, respectively.

In the case of the A-2 physical model, the highest attenuation was achieved in the $\frac{1}{3}$ octave bands with centre frequencies ranging from 2000 Hz to 3150 Hz. The measured ILs were from 13.0 dB to 16.4 dB (Fig. 13). The differences from the theoretical model in this range were from 2.3 dB to 3.4 dB.

In the case of the A-3 physical model, the highest attenuation was achieved in the $\frac{1}{3}$ octave bands with centre frequencies ranging from 2000 Hz to 5000 Hz.

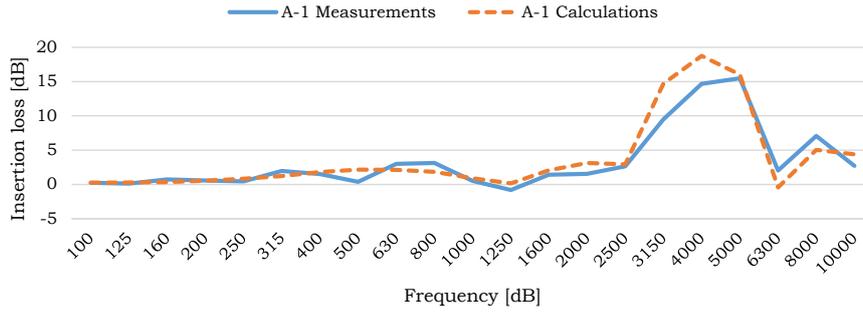


Fig. 12. Comparison of IL results for A-1 model.

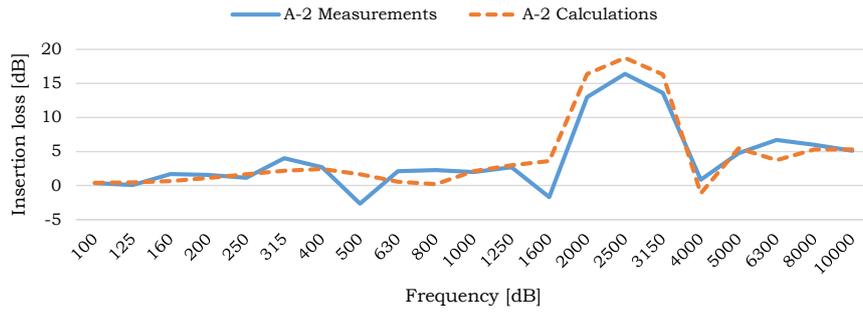


Fig. 13. Comparison of IL results for A-2 model.

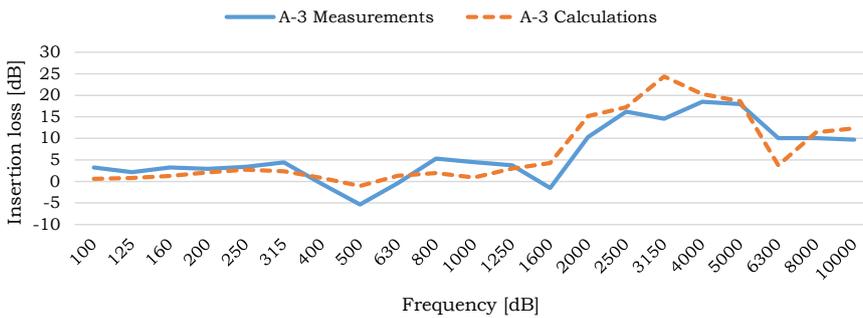


Fig. 14. Comparison of IL results for A-3 model.

The measured ILs were ranging from 10.3 dB to 18.5 dB (Fig. 14). The differences from the theoretical model in this range ranged from 0.7 dB to 9.8 dB.

Figure 15 presents a comparison of the measurement results for TL and IL across the three physical

models of the acoustic barrier (A-1–A-3). In both cases, it was shown that the layered physical model (A-3) provides greater noise attenuation and a wider frequency range of attenuation compared to the single-layer structures (A-1 and A-2).

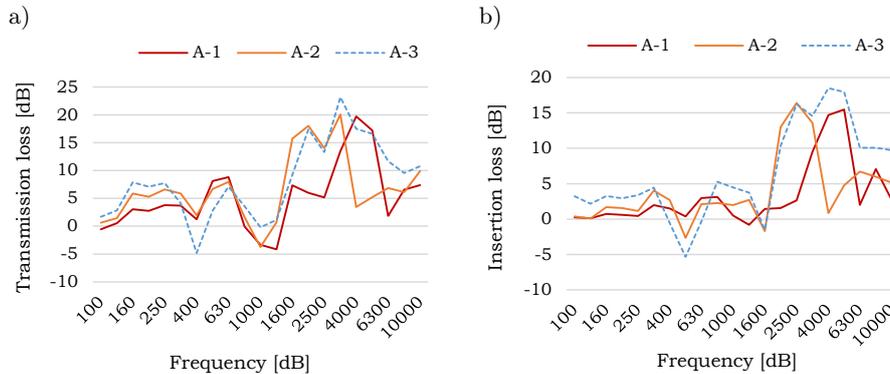


Fig. 15. Comparison of TL (a) and IL (b) measurement results for three physical models.

3. Conclusions

Compared to traditional partitions, barriers composed of sonic crystal structures do not exhibit continuous attenuation characteristics. The developed proposals of physical models, including layered structures, provide great opportunities for noise reduction in frequency bands where industrial noise sources emit substantial acoustic energy and significantly increase the frequency range of sound attenuation.

Three physical models were developed based on theoretical models. A laboratory test stand was prepared in a test room with conditions similar to a free field. The effectiveness of noise attenuation for sonic crystal structures was assessed using impulse response measurements for samples measuring 1 m × 1 m.

The results of the conducted measurements showed that the noise attenuation provided by the developed physical models was satisfactory for the key components of the analysed spectrum. It was also demonstrated that the layered structure allows to increase the noise attenuation (up to 3.5 dB) and increase the frequency range of attenuation (up to the range of 2000 Hz–5000 Hz) compared to single-layer structures. The measured of TL for the multilayer model was 23.3 dB, while the IL was 18.5 dB. The method used produced repeatable measurement results, with a standard deviation of repeatability not exceeding 0.02 dB.

The design assumptions were verified by comparing the measurement results of the physical models with those from theoretical models. The comparison of the results showed consistency in the occurrence of band gaps for key spectral bands predicted by Bragg's law, for both the TL and IL. Despite this agreement and the satisfactory effectiveness of noise suppression, differences were observed between the values in TL and IL between the physical and theoretical models. These differences result from the fact that theoretical models are 2D models and do not take into account the finite height of the barrier and the influence of the ground. The study results help estimate the influence of factors such as sound reflections or diffraction, which are not considered in 2D theoretical models.

Acknowledgments

This paper is based on research conducted as part of the fifth stage of the National Programme “Improvement of Safety and Working Conditions” partially supported from 2020 to 2022 by the Polish National Centre for Research and Development (project no.III.PB.01 titled “Development of an Industrial Acoustic Barrier for Attenuation of Narrowband Noise Components Using Layered Structures of Phonic Crystals”).

The Central Institute for Labour Protection – National Research Institute in Poland is the Programme's main coordinator.

References

- CASTIÑEIRA-IBÁÑEZ S., RUBIO C., ROMERO-GARCÍA V., SÁNCHEZ-PÉREZ J.V., GARCÍA-RAFFI L.M. (2012), Design, manufacture and characterization of an acoustic barrier made of multi-phenomena cylindrical scatterers arranged in a fractal-based geometry, *Archives of Acoustics*, **37**(4): 455–462, doi: [10.2478/v10168-012-0057-9](https://doi.org/10.2478/v10168-012-0057-9).
- CEN (2012), *Road traffic noise reducing devices – Test method for determining the acoustic performance – Part 6: Intrinsic characteristics – In situ values of airborne sound insulation under direct sound field conditions*, EN 1793-6:2012.
- ELFORD D.P., CHALMERS L., KUSMARTSEV F.V., SWALLOWE G.M. (2011), Matryoshka locally resonant sonic crystal, *The Journal of the Acoustical Society of America*, **130**(5): 2746–2755, doi: [10.1121/1.3643818](https://doi.org/10.1121/1.3643818).
- FUSTER-GARCIA E., ROMERO-GARCÍA V., SÁNCHEZ-PÉREZ J.V., GARCÍA-RAFFI L.M. (2007), Targeted band gap creation using mixed sonic crystal arrays including resonators and rigid scatterers, *Applied Physics Letters*, **90**(24): 244104, doi: [10.1063/1.2748853](https://doi.org/10.1063/1.2748853).
- GOFFAUX C., SÁNCHEZ-DEHESA J. (2003), Two-dimensional phononic crystals studied using a variational method: Application to lattices of locally resonant materials, *Physical Review B*, **67**(14): 144301, doi: [10.1103/PhysRevB.67.144301](https://doi.org/10.1103/PhysRevB.67.144301).
- HIRSEKORN M., DELSANTO P.P., BATRA N.K., MATIC P. (2004), Modelling and simulation of acoustic wave propagation in locally resonant sonic materials, *Ultrasonics*, **42**(1–9): 231–235, doi: [10.1016/j.ultras.2004.01.014](https://doi.org/10.1016/j.ultras.2004.01.014).
- HO K.M., CHUN K.C., YANG Z., ZHANG X.X., SHENG P. (2003), Broadband locally resonant sonic shields, *Applied Physics Letters*, **83**(26): 5566–5568, doi: [10.1063/1.1637152](https://doi.org/10.1063/1.1637152).
- HU X., CHAN C.T., ZI J. (2005), Two-dimensional sonic crystals with Helmholtz resonators, *Physical Review E*, **71**(5): 055601(R), doi: [10.1103/PhysRevE.71.055601](https://doi.org/10.1103/PhysRevE.71.055601).
- LIU Z. *et al.* (2000), Locally resonant sonic materials, *Science*, **289**(5485): 1734–1736, doi: [10.1126/science.289.5485.1734](https://doi.org/10.1126/science.289.5485.1734).
- MARTÍNEZ-SALA R., SANCHO J., SÁNCHEZ J.V., GÓMEZ V., LLINARES J., MESEGUER F. (1995), Sound attenuation by sculpture, *Nature*, **378**(6554): 241, doi: [10.1038/378241a0](https://doi.org/10.1038/378241a0).
- MORANDI F., MINIACI M., GUIDORZI P., MARZANI A., GARAI M. (2015), Acoustic measurements on a sonic crystals barrier, *Energy Procedia*, **78**: 134–139, doi: [10.1016/j.egypro.2015.11.128](https://doi.org/10.1016/j.egypro.2015.11.128).
- MORANDI F., MINIACI M., MARZANI A., BARBARRESI L., GARAI M. (2016), Standardised acoustic char-

- acterisation of sonic crystals noise barriers: Sound insulation and reflection properties, *Applied Acoustics*, **114**: 294–306, doi: [10.1016/j.apacoust.2016.07.028](https://doi.org/10.1016/j.apacoust.2016.07.028).
13. RADOSZ J. (2019), Acoustic performance of noise barrier based on sonic crystals with resonant elements, *Applied Acoustics*, **155**: 492–499, doi: [10.1016/j.apacoust.2019.06.003](https://doi.org/10.1016/j.apacoust.2019.06.003).
 14. ROMERO-GARCÍA V., KRYNKIN A., GARCIA-RAFFI L.M., UMNova O., SÁNCHEZ-PÉREZ J.V. (2013), Multi-resonant scatterers in sonic crystals: Locally multi-resonant acoustic metamaterial, *Journal of Sound and Vibration*, **332**(1): 184–198, doi: [10.1016/j.jsv.2012.08.003](https://doi.org/10.1016/j.jsv.2012.08.003).
 15. RUBIO C. *et al.* (1999), Existence of full gaps and deaf bands in two-dimensional sonic crystals, *Journal of Lightwave Technology*, **17**(11): 2202–2207, doi: [10.1109/50.803012](https://doi.org/10.1109/50.803012).
 16. SÁNCHEZ-PÉREZ J.V. *et al.* (1998), Sound attenuation by a two-dimensional array of rigid cylinders, *Physical Review Letters*, **80**: 5325, doi: [10.1103/PhysRevLett.80.5325](https://doi.org/10.1103/PhysRevLett.80.5325).

Research Paper

A Study of Acoustic Emission Based RA-AF Characteristics of Polypropylene Fiber-Reinforced Recycled Aggregate Concrete Under Uniaxial Compression

Daowen ZHOU^{(1),(2)}, Xin YANG^{(1),(2),(4)*}, Yu TANG^{(1),(2)}, Yutao MIAO⁽³⁾

⁽¹⁾ *School of Civil Engineering, Fujian University of Technology*
Fuzhou, China

⁽²⁾ *Key Laboratory of Underground Engineering, Fujian University of Technology*
Fuzhou, China

⁽³⁾ *Chengtong Branch, China Railway Second Group Co., Ltd.*
Chengdu, China

⁽⁴⁾ *School of Civil Engineering, Fuzhou University*
Fuzhou, China

*Corresponding Author e-mail: yangxin546@163.com

(received November 30, 2023; accepted April 10, 2024; published online August 6, 2024)

In order to research the acoustic emission characteristics of polypropylene fiber-reinforced recycled aggregate concrete under uniaxial load, 20 groups of test specimens with a coarse aggregate substitution rate of 25 % and 50 % are designed and fabricated to conduct the acoustic emission test under uniaxial compression, and the evolution laws of the acoustic emission b -value, the cracking modes and the acoustic emission RA-AF moving averages with time are studied. The laws of influence of the coarse aggregate substitution rate and coarse-fine polypropylene fiber on the acoustic emission b -value of RAC are discussed. The K -means clustering method is adopted for two-dimensional clustering analysis of the shear cracking and tensile cracking, and then the SVM is used to obtain the boundary between the two types of clusters. The time distribution laws of shear cracking and tensile cracking of the polypropylene fiber-reinforced recycled aggregate concrete are analyzed. The changes in the moving averages of RA and AF of RAC test specimens with time are studied, and the research indicates that as the RA value decreases, the shear cracking gradually reduces and the tensile cracking gradually increases and dominates.

Keywords: polypropylene fiber-reinforced recycled aggregate concrete; acoustic emission characteristics; b -value; RA-AF; moving average.



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

The current topic is how to improve the performance of recycled aggregate concrete (RAC). There are mainly two optimization schemes: first, when making recycled aggregate, the cement mortar on the surface is treated with chemical reagents to make the internal voids in aggregate smaller, thus increasing the RAC performance; second, fiber materials are added to improve the tensile and compressive strength and toughness of RAC. Steel fibers have been widely

used and popularized due to their very evident anti-crack effect since their application to concrete. However, steel fibers also have some shortcomings. When steel fibers were applied in a humid and corrosive environment, the service life of concrete would be significantly reduced due to its susceptibility to corrosion. Polypropylene fiber is a type of reinforced and toughened material, with good corrosion resistance, light weight, easy dispersion, little damage to mixing machines, low price, and other advantages, which better overcomes the shortcomings of steel fiber.

In recent years, many scholars have studied the acoustic emission characteristics of RAC, including RAC modification (adding silica fume – BAI *et al.*, 2022), replacement of coarse and fine aggregates (copper slag as fine aggregate – PREM *et al.*, 2018), replacement of coarse and fine aggregates with plastic wastes (BELMOKADDEM *et al.*, 2020), recycled sand as fine aggregate (FARDOUN *et al.*, 2022), and crumb rubber as fine aggregate (ASSAGGAF *et al.*, 2022), and aggregate proportion (GUO *et al.*, 2017). WATANABE *et al.* (2007) tested the microcracks in concrete under compression applying the acoustic emission technique. KENCANAWATI *et al.* (2013) introduced an acoustic emission parameter analysis model for analyzing the concrete cracking behavior.

The acoustic emission (AE) technique can be used to study damage evolution and identify failure modes (AKI, 1965; UTSU, 1965; LOCKNER *et al.*, 1991; UNANDER, 1993; WEISS, 1997; GREENHOUGH, MAIN, 2008; KWIA TEK *et al.*, 2014; LIU *et al.*, 2020a; 2020b; CHEN *et al.*, 2022). In addition, many achievements had been made in the research of acoustic emission characteristics of fiber-reinforced concrete, including compression performance (MENNA *et al.*, 2022), bending test (ADAMCZAK-BUGNO *et al.*, 2022a; 2022b; MANDAL *et al.*, 2022), shear behavior (GOYAL *et al.*, 2022), uniaxial tension (DE SMEDT *et al.*, 2022), and cyclic load (XARGAY *et al.*, 2021). GOYAL *et al.* (2021) established the empirical relationship between the damage index and the acoustic emission parameter by the genetic algorithm. ESSASSI *et al.* (2021) classified the acoustic emission signals by the *K*-means algorithm and found three different classes of cracking events: fiber cracking, matrix cracking, and fiber debonding. LAUFF *et al.* (2021) added short fibers to fresh concrete and processed the material with a 3D printer to orient the fibers, thus obtaining a material with high uniaxial strength properties. JUNG *et al.* (2021) proposed

a new parameter, namely, the composite *b*-value, to analyze the distribution of the acoustic emission amplitude at the crack origin. PREM *et al.* (2021) reported that acoustic emission parameters were closely related to different failure mechanisms (shear, shear flexure, and flexure).

To sum up, the polypropylene fiber-reinforced RAC test was conducted under uniaxial compression in this work to obtain the acoustic emission information in the fracturing process and analyze the evolution law of the AE based *b*-value and the AE based RA-AF (RA is defined as the ratio of rise time to amplitude; AF is defined as the ratio of counts to duration). Based on *K*-means clustering and support vector machine, the cracking modes of polypropylene fiber-reinforced recycled aggregate concrete were studied. On this basis, a whole-process analysis was made based on the time history of the acoustic emission characteristic, which was of great theoretical and practical significance.

2. Experimental details

2.1. Specimen preparation

The polypropylene fiber used for the test is produced by Hebei Xinqixiang Polypropylene Fiber Technology Co., Ltd., with its physical and mechanical parameters shown in Table 1.

The cement used is ordinary Portland cement of grade 42.5, the fine aggregate is medium sand in zone II with a particle size of 0.15 mm–4.75 mm and the natural coarse aggregate is the gravel with a particle size of 5 mm–20 mm, among which the 5 mm–10 mm and 10 mm–20 mm aggregates are in a 2:3 ratio. The recycled coarse aggregates are produced with waste concrete with a base material strength grade of C30 by way of mechanical crushing and screening, as shown in Fig. 1.

Table 1. Physical and mechanical indicators of polypropylene fiber.

Fiber no.	Diameter [mm]	Length [mm]	Tensile strength [MPa]	Fracture strength [MPa]	Elongation at break [%]	Initial modulus [GPa]	Density [g/cm ³]	Recommended dosage [kg/m ³]
FF1	0.036	12	≥ 300	360	30	3.5	0.91	0.9
FF2	0.036	19	≥ 450	450	17	4.8	0.91	0.9
CF1	0.9	30	≥ 550	500	24	6.6	0.91	6.0
CF2	0.9	50	≥ 550	500	24	6.6	0.91	6.0



Fig. 1. Coarse and fine aggregates of RAC: a) fine aggregates; b) natural coarse aggregates; c) recycled coarse aggregates.

Table 2. Test plan for C30 polypropylene fiber-reinforced RAC.

Specimen no.	Mixing amount of coarse aggregate [kg/m ³]		Coarse aggregate substitution rate [%]	Fiber length [mm]	Fiber dosage [kg/m ³]
	Natural	Recycled			
R-25-0	840.63	280.21	25	None	0
R-25-1	840.63	280.21		12	0.9
R-25-2	840.63	280.21		19	0.9
R-25-3	840.63	280.21		30	6
R-25-4	840.63	280.21		50	6
R-25-5	840.63	280.21		12 + 30	0.9 + 5.1
R-25-6	840.63	280.21		12 + 50	0.9 + 5.1
R-25-7	840.63	280.21		19 + 30	0.9 + 5.1
R-25-8	840.63	280.21		19 + 50	0.9 + 5.1
R-25-9	840.63	280.21	12 + 19 + 50	0.45 + 0.45 + 5.1	
R-50-0	560.42	560.42	50	None	0
R-50-1	560.42	560.42		12	0.9
R-50-2	560.42	560.42		19	0.9
R-50-3	560.42	560.42		30	6
R-50-4	560.42	560.42		50	6
R-50-5	560.42	560.42		12 + 30	0.9 + 5.1
R-50-6	560.42	560.42		12 + 50	0.9 + 5.1
R-50-7	560.42	560.42		19 + 30	0.9 + 5.1
R-50-8	560.42	560.42		19 + 50	0.9 + 5.1
R-50-9	560.42	560.42	12 + 19 + 50	0.45 + 0.45 + 5.1	

In this test, the concrete with a strength grade of C30 is used, and the number of test groups is 20, with 3 test specimens for each group. It is calculated that the mix proportion of C30 polypropylene fiber-reinforced RAC is: cement 358 kg/m³, medium sand 706.15 kg/m³, coarse aggregate 1,120.85 kg/m³, and water 215 kg/m³. Each group of concrete uses the same mix proportion, and only the coarse aggregate substitution rate and fiber dosage are different. Among them, the recycled coarse aggregate substitution rate for R-25 and R-50 is 25 % and 50 %, respectively. As a reference test specimen, R-25-0 should be made of plain concrete, without adding any fiber to it. R-25-1, R-25-2, R-25-3, and R-25-4 are four test specimens, each mixed with a different kind of polypropylene fiber. The fiber dosage is 0.9 kg/m³ for R-25-1 and R-25-2 test specimens into which one kind of fine polypropylene fiber is added, respectively, and 6.0 kg/m³ for R-25-3 and R-25-4 into which one kind of coarse fiber is added, respectively, according to the instruction manual of

the polypropylene fiber manufacturer. R-25-5, R-25-6, R-25-7, R-25-8, and R-25-9 are the test specimens into which both fine and coarse fibers are added. To ensure the test comparability, the total dosage of coarse and fine fibers is controlled to 6.0 kg/m³ as shown in Table 2.

2.2. Experimental setup

With reference to the related provisions in (GB/T 50081-2019, 2019), a 150 mm × 150 mm × 150 mm standard test cube is designed for the test. After being molded, all the test specimens are left at room temperature for 1D, numbered for removal of their forms, and cured in the standard curing room for 28 days before the cube compression test is carried out. The machine used for the cube compression test of polypropylene fiber-reinforced RAC is the HCT306B microcomputer-controlled electro-hydraulic servo press, and the test specimen is loaded at a speed of 0.5 MPa/s until the test specimen is ruptured.

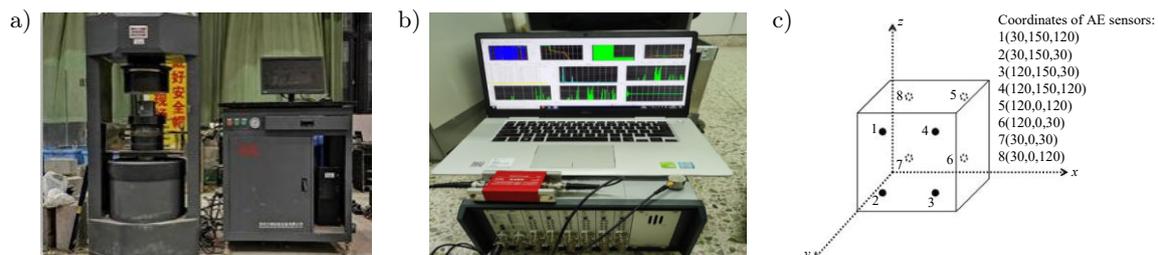


Fig. 2. Test apparatus: a) HCT306B press; b) AMSY-6 acoustic emission instrument; c) acoustics emission sensor arrangement (applied load acts parallel to the z -axis on the plane above the specimen).

The acoustic emission acquisition system is an AMSY-6 acoustic emission instrument. During the test, 8 acoustic emission sensors are fixed on the surface of polypropylene fiber-reinforced RAC (this work selects the channel with the third highest number of acoustic emission events for analysis). To ensure that the acoustic emission sensor and RAC can be fully coupled, a layer of vacuum adhesive is applied to the contact position between the acoustic emission sensor and RAC. To eliminate the effect of environmental noise upon the test, the threshold value and the sampling frequency of the acoustic emission instrument are set as 40 dB and 5 MHz, respectively.

2.3. Characteristics of AE based b -value

GUTENBERG and RICHTER (1944) proposed the famous expression of statistical relation between earthquake magnitude and frequency:

$$\log_{10} N = a - bM, \quad (1)$$

where M is the magnitude, N is the earthquake frequency of magnitude in ΔM , a and b are constants, and the b -value is a function of the relative magnitude distribution. In this work, the b -value is calculated by replacing the magnitude M with the acoustic emission amplitude, therefore the b -value is a function of the crack growth scale, whose dynamic change characteristics are of direct physical significance. The least square method is adopted in this work to calculate the b -value. A hundred pieces of acoustic emission amplitude data are taken as the sampling window each time, the sliding window is 50, and the magnitude interval ΔM is set to 0.2 dB. In the data processing process, the occurrence of the last acoustic emission amplitude from a hundred piece sampling window is regarded as the scale of the b -value.

Figure 3 illustrates the temporal variation of the AE based b -value of the RAC with a coarse aggregate substitution rate of 25 % and 50 % (the first specimen from each group was selected for the acoustic emission analysis in this work). The law of influence of the coarse aggregate substitution rate on the b -value is analyzed based on Fig. 3a. For no. 0 specimens not mixed with polypropylene fibers, the AE based b -value of R-25-0 specimen shows a continuous sharp decline in the initial phase, indicating a continuous increase of major event cracking; later, the b -value fluctuates up and down in a small range until the post-peak phase, indicating the stable growth of microcracks. The b -value of R-50-0 test specimen begins to increase slightly after 20 seconds, indicating an increase of minor event cracking; later, the b -value fluctuates in a small range; when the specimen is on the verge of failure, the b -value begins to decrease significantly until the post-peak phase, indicating a continuous increase of the major event cracking of the specimen.

It can be seen from the analysis of the changes of the AE based b -value in Fig. 3a that:

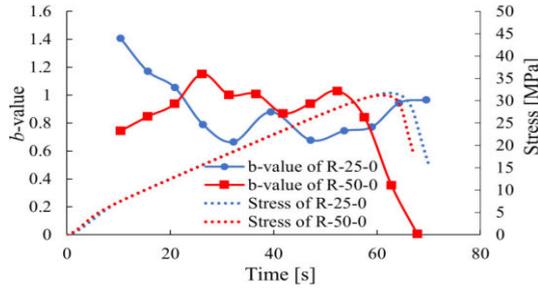
- 1) For both R-25-0 and R-50-0 RAC into which no polypropylene fibers are added, the b -value exhibits a continuous sharp decrease for R-50-0, indicating that a specimen (R-25-0) with a small content of recycled aggregate is probably more resistant to microcracking processes than R-50-0.
- 2) The b -value of R-50-0 test specimen decreases significantly in the post-peak phase, indicating a continuous increase of major event cracking in the post-peak phase. The b -value of R-25-0 shows a small increase in the post-peak phase. Therefore, on the whole, the post-peak strength of R-25-0 test specimen with a coarse aggregate substitution rate of 25 % is better than that of R-50-0 test specimen with a coarse aggregate substitution rate of 50 %.
- 3) The initial b -value of R-50-0 test specimen is smaller than that of R-25-0 test specimen probably because the more initial defects of R-50-0 test specimen lead to more major event cracking in the initial phase, making the initial b -value bigger.

From the analysis of the law of influence of adding one kind of fine polypropylene fiber on the b -value in Fig. 3b–c, it can be observed that, for R-25-1, R-25-2, R-50-1, and R-50-2 test specimens into which only one kind of fine fiber is added, respectively, their b -value do not decrease sharply until they are close to failing in the pre-peak phase, and that their b -value mainly increase slightly and slowly or fluctuate up and down in a small range. This indicates that with the addition of polypropylene fine fibers, the major event cracking is inhibited and the minor vent cracking dominates. When the test specimens are about to fail, the b -value decrease significantly, indicating that major event cracking increases gradually.

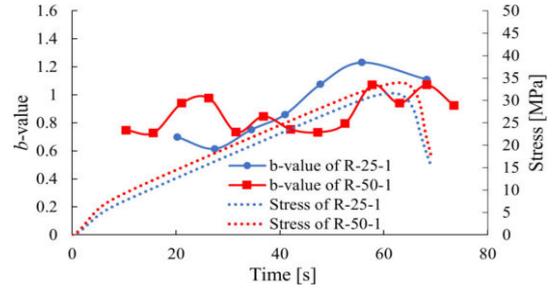
From the analysis of the law of influence of adding one kind of coarse polypropylene fiber on the b -value in Fig. 3b–c, it can be observed that, for R-25-3, R-50-3, R-25-4, and R-50-4 test specimens into which only one kind of coarse polypropylene fiber is added (fiber diameter: 0.9 mm; fiber length: 30 mm and 50 mm), their b -values are generally stable and begin to decrease significantly only when they are about to fail.

From the analysis of the law of influence of adding more than one kind of coarse polypropylene fiber on the b -value in Fig. 3f–j, it can be observed that, for R-25-6, R-50-6, and R-50-7 test specimens into which both coarse and fine fibers are added, their b -values are generally stable and begin to decrease significantly only when they are about to fail. The b -value of other test specimens into which more than one kind of fiber is added are characterized by ups and downs, demonstrating the intense evolution of microcracks.

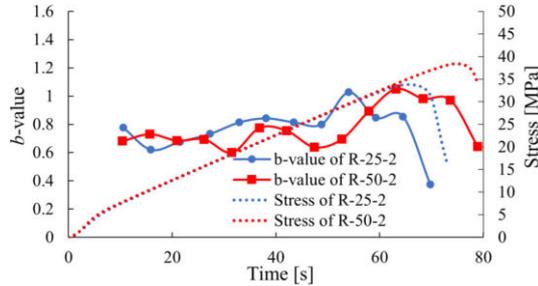
a) No. 0 specimens without polypropylene fiber (mean value of b -value is 0.9155 and 0.884 for R-25-0 and R-50-0).



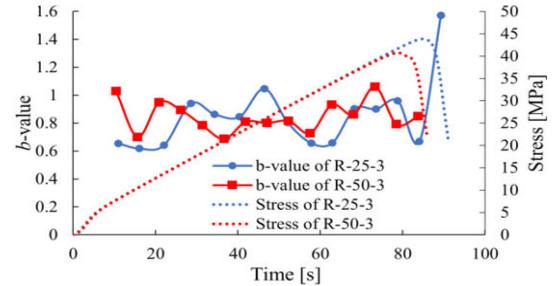
b) No. 1 specimens with fine fiber added (mean value of b -value is 0.9061 and 0.8655 for R-25-1 and R-50-1).



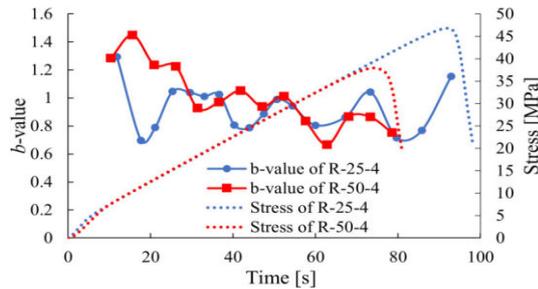
c) No. 2 specimens with fine fiber added (mean value of b -value is 0.7651 and 0.77 for R-25-2 and R-50-2).



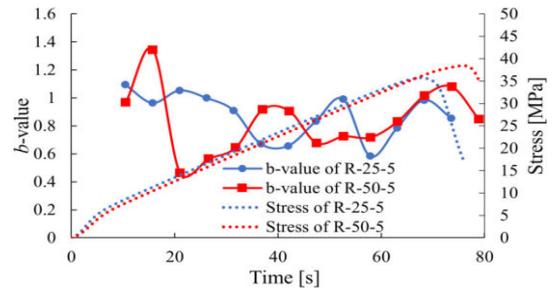
d) No. 3 specimens with coarse fiber added (mean value of b -value is 0.849 and 0.8457 for R-25-3 and R-50-3).



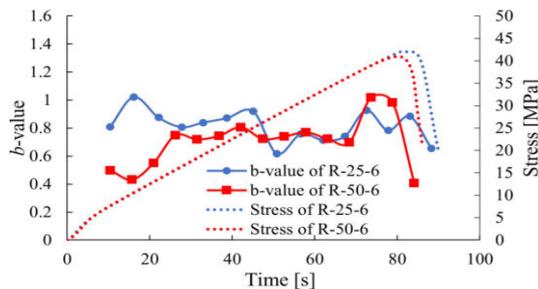
e) No. 4 specimens with coarse fiber added (mean value of b -value is 0.9252 and 1.005 for R-25-4 and R-50-4).



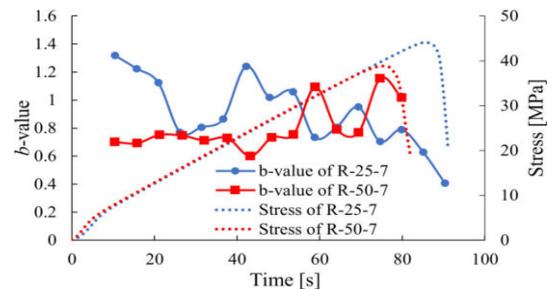
f) No. 5 specimens with coarse and fine fibers (mean value of b -value is 0.8757 and 0.8353 for R-25-5 and R-50-5).



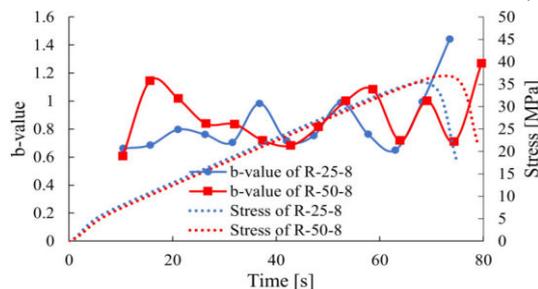
g) No. 6 specimens with coarse and fine fibers (mean value of b -value is 0.8144 and 0.7036 for R-25-6 and R-50-6).



h) No. 7 specimens with coarse and fine fibers (mean value of b -value is 0.9027 and 0.8031 for R-25-7 and R-50-7).



i) No. 8 specimens with coarse and fine fibers (mean value of b -value is 0.8392 and 0.8883 for R-25-8 and R-50-8).



j) No. 9 specimens with coarse and fine fibers (mean value of b -value is 0.852 and 0.7833 for R-25-9 and R-50-9).

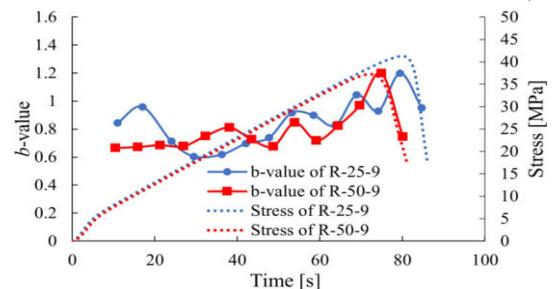


Fig. 3. Temporal variation of AE based b -value of RAC.

3. Analysis of polypropylene fiber-reinforced concrete cracking modes based on RA-AF method

Concrete cracking modes include tensile and shear modes of cracking. The correlation between RA and AF in acoustic emission parameters is related to the cracking modes of concrete materials and can be used to effectively identify the cracking modes of concrete materials.

Generally speaking, RA and AF values are different in terms of shear damage and tensile damage. In a tensile damage mode, the related AF value will be higher due to the release of energy. On the contrary, due to the long rise time and duration, the RA value related to the shear damage mode will be higher (ZHANG, 2020). The RA and AF calculation principles are indicated in Eqs. (2) and (3), respectively:

$$RA = \frac{\text{Rise time}}{\text{Amplitude}}, \quad (2)$$

$$AF = \frac{\text{Counts}}{\text{Duration}}, \quad (3)$$

where rise time is the time interval during which the acoustic emission signal is raised to its maximum amplitude through a threshold, whose unit is μs ; amplitude is the highest amplitude value of the acoustic emission waveform signal, whose unit is mV; counts is the number of times the single acoustic emission signal exceeds the threshold value, whose unit is dimensionless; duration is the time from the first time the acoustic emission signal exceeds the threshold value to the last time it drops to the threshold value, whose unit is μs .

Determining the RA-AF boundary is important to the differentiation between material cracking modes. The boundary is generally a straight line defined artificially. The acoustic emission event distributed to the upper left of the straight line is considered tensile cracking, while that distributed to the lower right of the straight line is considered shear cracking, as shown in Fig. 4. However, such a method can be easily affected by such factors as the sensor model and characteristic parameter selection. Consequently, an unsupervised machine learning technique – K -means cluster-

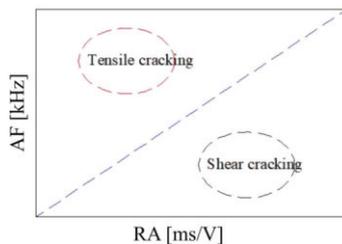


Fig. 4. Classification of typical cracking modes based on RA-AF relationship.

ing method is introduced in this work to differentiate between RA and AF.

3.1. K -means clustering

K -means is a method of continuous iterative clustering. Its operating principle is: assuming that the sample data is n variables X_1 to X_n , these n variables are divided into k categories, and m_i is the mean value of these variables (MACQUEEN, 1967). The distance formula adopts the Euclidean distance, and the specific steps are as follows:

- 1) randomly and uniformly select K observation samples as the initial cluster center m_1 to m_k ;
- 2) separate each sample data from its nearest cluster center;
- 3) update the mean vector of the cluster center according to the cluster center to which each sample belongs;
- 4) repeat steps 2 and 3. When the set number of iterations is reached or the mean vector of the cluster center is no longer updated, the model is built and the clustering algorithm results are derived.

3.2. Theory of support vector machines

The support vector machines (SVM) can find a suitable interface for the two types of data, so that the two types of data can be separated most completely. The interface is defined as the optimal hyperplane (VAPNIK, 1999). Forming a new vector \mathbf{y} ($\mathbf{y} = (X, L)$) with data X and its label L can form an n -dimensional vector space ($\mathbf{Y} = (y_1, \dots, y_j, \dots, y_n)$) for n data. In the vector space \mathbf{Y} , the hyperplane can be described by the following equation:

$$\omega^T X + d = 0, \quad (4)$$

where $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_f)$ is the hyperplane normal vector, which is used to describe the direction of the hyperplane; d is the hyperplane displacement term, which is used to describe the distance of the hyperplane from the origin.

The vectors closest to the hyperplane in the two types of vectors are called support vectors, and the sum of the distances between two heterogeneous support vectors and the hyperplane is:

$$\gamma = \frac{2}{\|\boldsymbol{\omega}\|}. \quad (5)$$

Find the optimal hyperplane, that is, calculate ω and d to make t the maximum value. If linear inseparability is considered, the objective function to be solved is shown in Eq. (5). At this time, some points are allowed not to meet the constraint condition (6):

$$f = \min_{\omega, d} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m L_{0/1}(y_i(\omega^T X_i + d) - 1), \quad (6)$$

$$y_i(\omega^T X_i + d) \geq 1, \quad i = 1, 2, \dots, n, \quad (7)$$

where $L_{0/1}$ is 0–1 loss function, C is a constant greater than 0. When C is taken as a valid value, some data are allowed not to meet the constraint conditions. When C is taken as infinity, all data will be forced to meet the conditions.

3.3. Cracking modes analysis

In this work, the K -means clustering method is used for the two-dimensional clustering analysis of RA-AF values, so as to differentiate between shear and tensile cracking. Then, the recognition and classification function of the SVM is used to obtain the boundary between the two types of clusters, i.e., shear cracking and tensile cracking. In order to make the data meet the constraint conditions as much as possible, the constant C in Eq. (6) is set to 1.000 in this work. See Figs. 5–6 (note: the boundary of the compaction phase is determined by the stress-strain curve,

which is the turning point of the stress-strain curve from an upward concave shape to an approximate straight line) and Table 3 for the calculation results. As can be seen from Fig. 5, the two types of cracking can be well differentiated by applying the K -means clustering method; the shear cracking is characterized by a high RA value, while the tensile cracking is characterized by a low RA value; the slopes of the linear RA-AF value boundaries of the two types of cracking obtained by the SVM are both close to 0. Most of the cracking of the polypropylene fiber-reinforced RAC is tensile cracking. As can be observed from Fig. 6 and Table 3, shear cracking is mainly distributed in the elastic-plastic phase and some test specimens have a few shear cracking in the compaction phase and the post-peak phase; there are a lot of tensile cracking in the three phases. In addition, in the latter half of the elastic-plastic phase, the RA value of the shear cracking reduces gradually; the shear cracking disappears within 5 s–18 s before the test specimen fails, and only tensile cracking exists in this time range. To sum up, the tensile cracking is in the majority in terms of the number of cracking modes. From the perspective of time distribution of cracking,

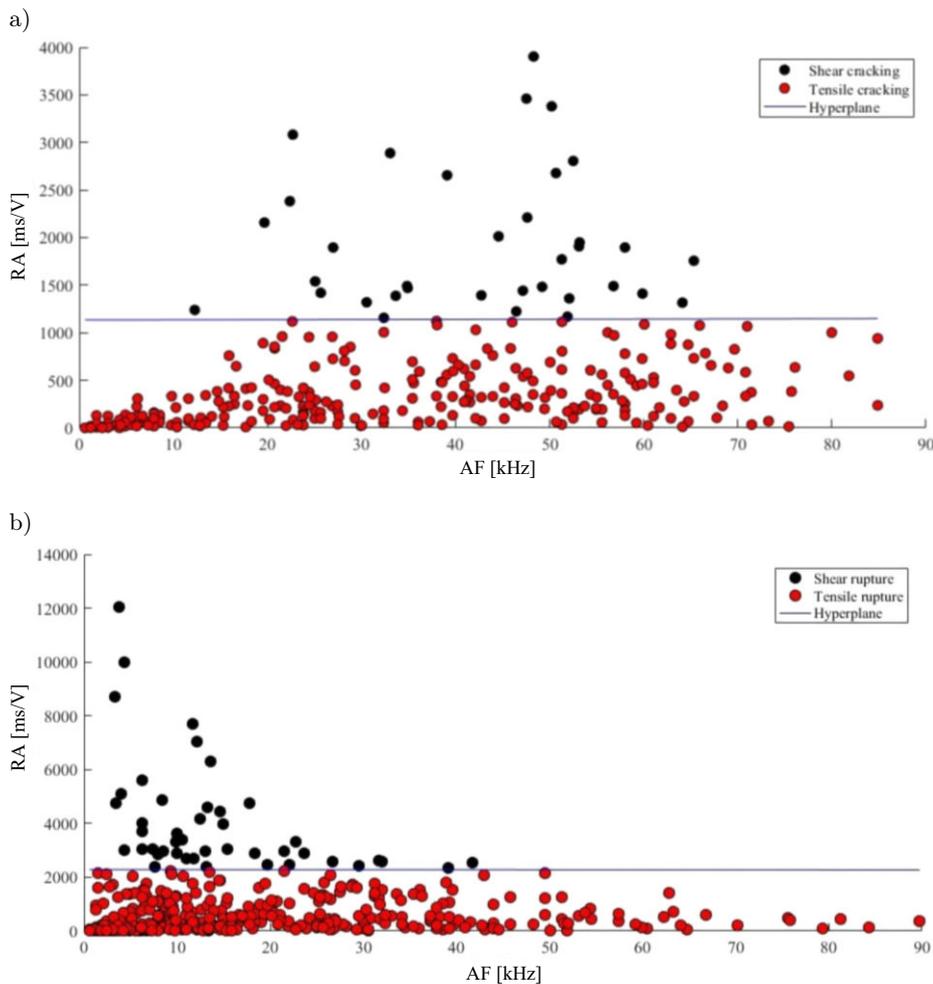


Fig. 5. Hyperplane calculation results of shear and tensile cracking: a) R-25-0; b) R-25-2.

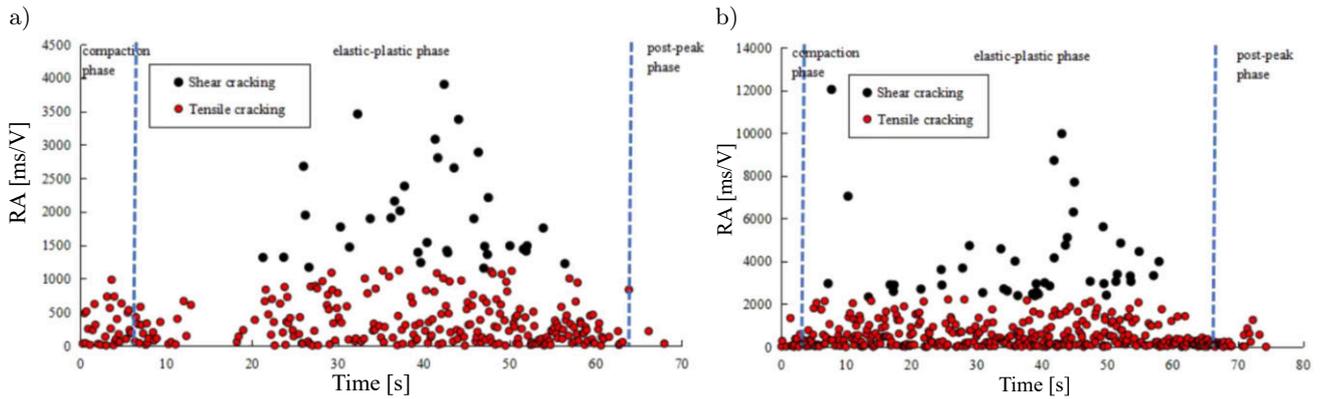


Fig. 6. Time distribution of shear and tensile cracking:
a) R-25-0; b) R-25-2.

Table 3. RA-AF value boundaries corresponding to different cracking modes of polypropylene fiber-reinforced RAC and shear cracking time distribution.

Specimen no.	Shear cracking and tensile cracking dividing line for RA-AF value $RA = k \cdot AF + e$		Time distribution of shear cracking [s]	Failure time of specimens [s]
	k [ms/V/kHz]	e [ms/V]		
R-25-0	0.1579	1134.5	21.3–56.4	62
R-25-1	0.0677	1433.8	2.2–52.5	62
R-25-2	-0.1487	2278.9	7.2–57.8	66
R-25-3	0.0885	1808.8	3.5–75	85
R-25-4	0.3166	2436.6	2.3–73.7	92
R-25-5	-0.368	1599.6	10.6–60.7	68
R-25-6	-0.0299	1918.2	2.4–72.9	81
R-25-7	0.1508	1679.8	14.6–78.8	86
R-25-8	-0.2016	1278.4	0.6–56.4	68
R-25-9	0.0592	2411.2	7.3–66	80
R-50-0	-0.4341	894.2	9.6–56.1, 63.1	61
R-50-1	0.8742	1375.4	3.2–49.4	63
R-50-2	0.6816	1284.7	7.9–65	75
R-50-3	-0.4768	1431.8	6.3–65.4	79
R-50-4	0.4683	677.0	3–66.6, 76.3	74
R-50-5	-0.3026	1170.5	8.5–68.1	76
R-50-6	0.0171	1522.7	2.8–65.3	80
R-50-7	0.1057	2084.3	5.3–66.9	75
R-50-8	0.0342	1941.9	7.6–60.7	73
R-50-9	0.3992	1553.7	3.4–59.7	72

the shear cracking is mainly concentrated in the first half of the elastic-plastic phase, and the tensile cracking is in the majority within 5 s–18 s before the test specimen fails. Therefore, the cracking mechanism of the polypropylene fiber-reinforced RAC under uniaxial compression is dominated by tensile cracking.

4. Whole-process analysis under uniaxial compression based on time history of acoustic emission characteristic parameters

As can be known from Eqs. (2) and (3) and the research results of Sec. 3, the RA value increases/

decreases mainly depending on the increase/decrease of the shear cracking, while the AF value increases/decreases mainly depending on the intensity of acoustic emission activities. Figure 7 illustrates the changes in RA and AF moving averages of RAC test specimens with time. In order to reduce scattering and clearly reveal the trend, each point on the curve is the moving average of the last 100 acoustic emission data points and the time of the last data point of every 100 acoustic emission data points is taken as the scales of RA and AF.

Research findings:

1) As can be seen from Fig. 7, the RA moving averages of most test specimens exhibit an overall upward trend in the first half phase and an overall downward

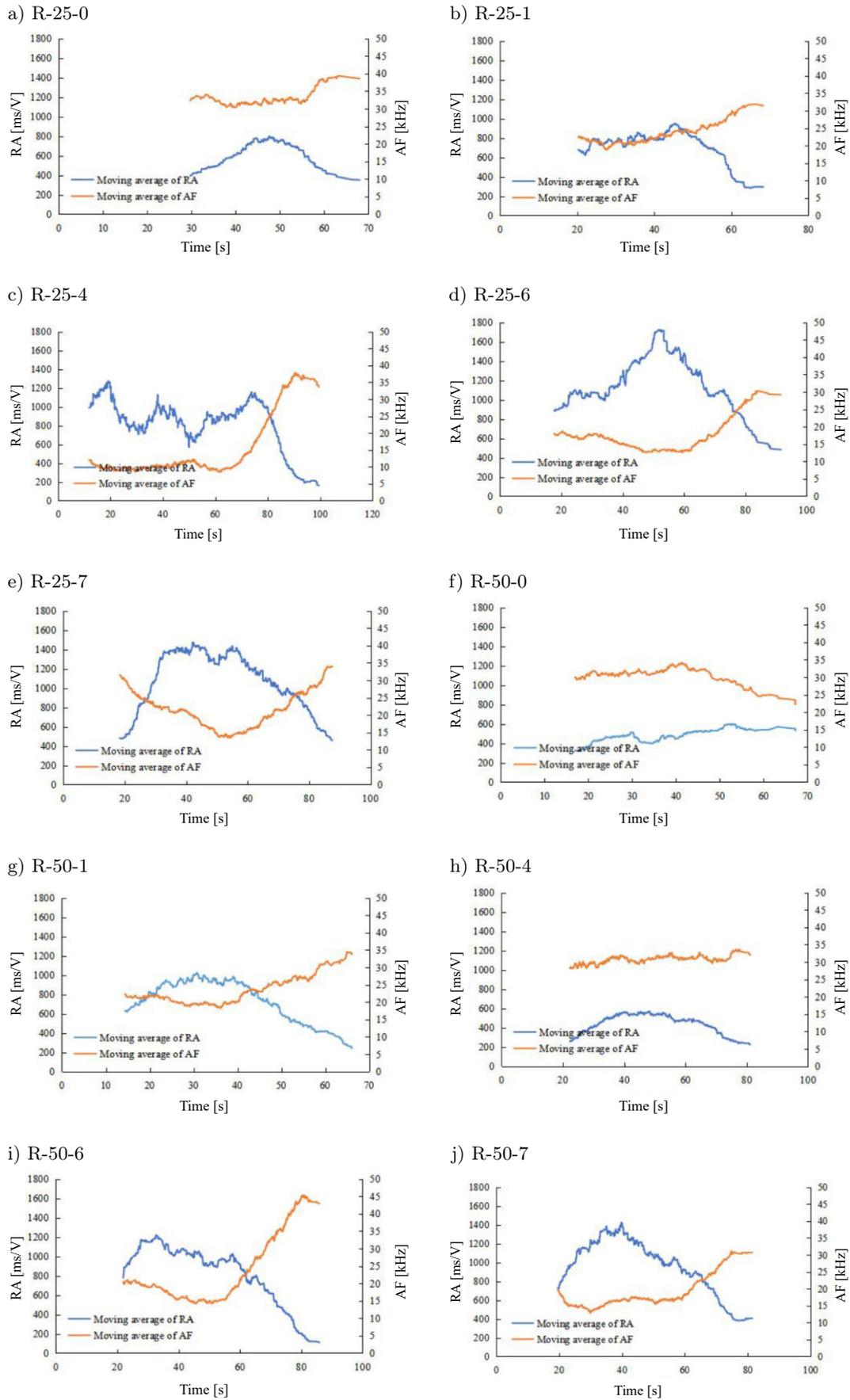


Fig. 7. Moving averages of acoustic emission RA and AF of RAC.

trend in the latter half phase. As the RA value decreases, the shear cracking gradually reduces and the tensile cracking gradually increases and dominates.

2) As can be observed from Fig. 7 and Table 4, the minimum values of the AF moving averages of most test specimens are negatively correlated with the RAC strength, namely the lower the minimum value of the AF moving average is, the higher the RAC strength will be. Among the RAC with a coarse aggregate substitution rate of 25 %, the minimum value of the AF moving average is 30.44 for R-25-0 test specimen into which no fiber is added and in the 8.53–18.87 range for the RAC into which coarse and fine fibers are added. Among the RAC with a coarse aggregate substitution rate of 50 %, the minimum value of the AF moving average is 22.31 for R-50-0 test specimen into which no fiber is added and is in the 12.91–28.13 range for the RAC into which coarse and fine fibers are added; Therefore, the minimum value of the AF moving average can be used to evaluate the reinforcement effect of polypropylene fibers on RAC. In concrete with a coarse aggregate substitution rate of 25 %, the minimum value of the AF moving average of R-25-4 test specimen is the minimum, showing the best reinforcement effect. In concrete with a coarse aggregate substitution rate of 50 %, the minimum values of the AF moving averages of R-50-6 and R-50-7 test specimens are the minimum, showing better reinforcement effects than other specimens.

3) There is a certain correlation between the AF value and the AE based b -value. When the AF value increases continuously, the corresponding b -value also increases continuously, indicating an increase in minor event cracking, as shown in Fig. 8a, R-25-1 test specimen in the 20 s–55 s range. When the AF value decreases continuously, there are two scenarios. One is that AF has a high initial value (more than 25 kHz) and decreases continuously. In this scenario, the AE based b -value also decreases continuously, indicating an increase in major event cracking, as shown in Fig. 8b,

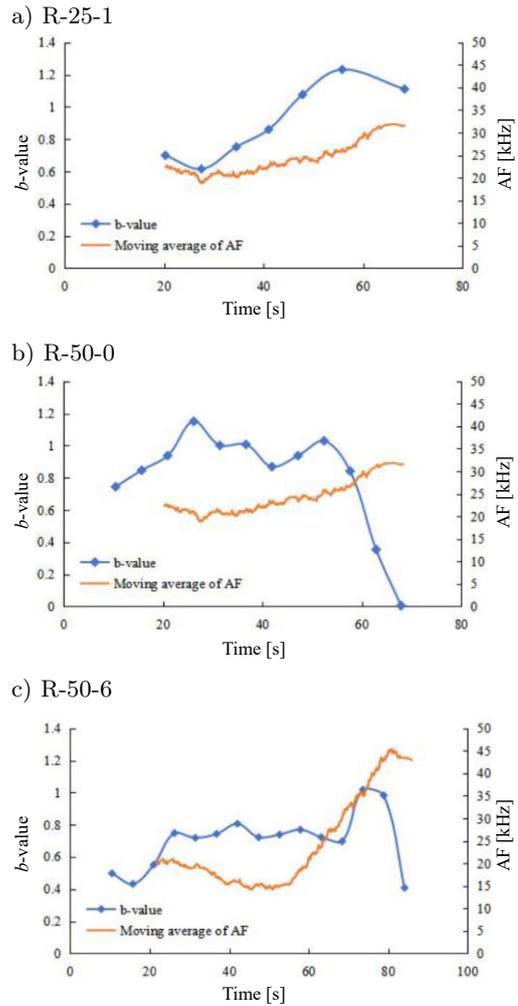


Fig. 8. Relationship between acoustic emission AF moving average of RAC and b -value.

R-50-0 test specimen in the 36 s–67 s range. The other is that AF has a low initial value (less than 25 kHz) and decreases continuously. In this scenario, the AE based b -value does not show a continuous decrease, indicating that the acoustic emission activity is attenuated, as shown in Fig. 8c, R-50-6 test specimen in the 22 s–48 s range.

5. Conclusions

In this work, the acoustic emission characteristics of polypropylene fiber-reinforced RAC have been researched under uniaxial compression, and the following main conclusions have been drawn:

1) For both R-25-0 and R-50-0 RAC into which no polypropylene fibers are added, the b -value exhibits a continuous sharp decrease, indicating that the existence of internal defects in RAC leads to the continuous increase of major event cracking. The initial b -value of R-50-0 test specimen is bigger because more initial defects lead to more major event cracking in the initial phase than R-25-0 test specimen. It is found in the re-

Table 4. Relationship between minimum value of acoustic emission AF moving average and peak strength.

Specimen no.	Minimum values of the AF moving averages [kHz]	Peak stress [MPa]
R-25-0	30.44	31.384
R-25-1	18.87	32.385
R-25-4	8.53	49.1
R-25-6	12.62	41.109
R-25-7	13.49	43.658
R-50-0	22.31	29.427
R-50-1	18.28	34.787
R-50-4	28.13	38.469
R-50-6	14.21	41.169
R-50-7	12.91	39.646

search that, for R-25-1, R-25-2, R-50-1, and R-50-2 test specimens into which one kind of fine fiber is added respectively, their b -values do not decrease sharply until the specimens are about to damage. This indicates that with the addition of polypropylene fine fibers, the major event cracking are inhibited. For R-25-4 and R-50-4 test specimens into which only one kind of coarse fiber is added (fiber diameter: 0.9 mm; fiber length: 50 mm), their b -values exhibit a short transition, indicating the intense evolution of microcracks, and then stay stable generally, indicating the stable growth of microcracks, which is mainly due to the bridging effect of coarse fibers.

2) The K -means clustering method has been adopted for two-dimensional clustering analysis of the shear cracking and tensile cracking of the polypropylene fiber-reinforced RAC. The shear cracking is characterized by a high RA value, while the tensile cracking is characterized by a low RA value. The tensile cracking is in the majority in terms of the number of cracking modes. From the perspective of time distribution of cracking, the shear cracking is mainly concentrated in the first half of the elastic-plastic phase; the tensile cracking exists in a large quantity in the compaction, elastic-plastic and post-peak phases; the shear cracking disappears and the tensile cracking is in the majority within 5 s–18 s before the test specimen fails. Therefore, the cracking mechanism of the polypropylene fiber-reinforced RAC under uniaxial compression is dominated by tensile cracking. The SVM has been used to give the hyperplane equations of shear cracking and tensile cracking of polypropylene fiber-reinforced RAC, and the slopes of the linear boundaries of the hyperplane equations are close to 0.

3) The changes in RA and AF moving averages of RAC test specimens with time have been researched. The research shows that, as the RA value decreases, the shear cracking gradually reduces and the tensile cracking gradually increases and dominates. The minimum values of the AF moving averages of most test specimens are negatively correlated with the RAC strength, namely the lower the minimum value of the AF moving average is, the higher the RAC strength will be. There is a certain correlation between the AF value and the AE based b -value. When the AF value increases continuously, the corresponding b -value also increases continuously, indicating an increase in minor event cracking. When the AF value decreases continuously, there are two scenarios. One is that AF has a high initial value (more than 25 kHz) and decreases continuously. In this scenario, the AE based b -value also decreases continuously, indicating an increase in major event cracking; The other is that AF has a low initial value (less than 25 kHz) and decreases continuously. In this scenario, the AE based b -value does not show a continuous decrease, indicating that the acoustic emission activity is attenuated.

Acknowledgments

This study was supported by the Fujian Provincial Natural Science Foundation Projects (grant no. 2022J01930), the authors gratefully acknowledge this support.

References

- ADAMCZAK-BUGNO A., LIPIEC S., VAVRUS M., KOTES P. (2022a), Non-destructive methods and numerical analysis used for monitoring and analysis of fibre concrete deformations, *Materials*, **15**(20): 7269, doi: [10.3390/ma15207268](https://doi.org/10.3390/ma15207268).
- ADAMCZAK-BUGNO A., ŚWIT G., KRAMPIKOWSKA A., PROVERBIO E. (2022b), Analysis of the significance of changes in the number and energy parameters of acoustic emission signals on the assessment of the strength of fibre-cement boards, *Materials*, **15**(16): 5757, doi: [10.3390/ma15165757](https://doi.org/10.3390/ma15165757).
- AKI K. (1965), Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits, *Bulletin of the Earthquake Research Institute* (Tokyo), **43**: 237–239.
- ASSAGGAF R., MASLEHUDDIN M., AL-OSTA M.A., AL-DULAIJAN S.U., AHMAD S. (2022), Properties and sustainability of treated crumb rubber concrete, *Journal of Building Engineering*, **51**: 104250, doi: [10.1016/j.jobee.2022.104250](https://doi.org/10.1016/j.jobee.2022.104250).
- BAI W.F., SHEN J.X., GUAN J.F., WANG J.Y., YUAN C.Y. (2022), Study on compressive mechanical properties of recycled aggregate concrete with silica fume at different strain rates, *Materials Today Communications*, **31**: 103444, doi: [10.1016/j.mtcomm.2022.103444](https://doi.org/10.1016/j.mtcomm.2022.103444).
- BELMOKADDEM M., MAHI A., SENHADJI Y., PEKMEZCI B.Y. (2020), Mechanical and physical properties and morphology of concrete containing plastic waste as aggregate, *Construction and Building Materials*, **257**: 119559, doi: [10.1016/j.conbuildmat.2020.119559](https://doi.org/10.1016/j.conbuildmat.2020.119559).
- CHEN D.L. *et al.* (2022), Effect of attenuation on amplitude distribution and b value in rock acoustic emission tests, *Geophysical Journal International*, **229**(2): 933–947, doi: [10.1093/gji/ggab480](https://doi.org/10.1093/gji/ggab480).
- DE SMEDT M., VANDECRUYS E., VRIJDAGHS R., VERSTRYNGE E., VANDEWALLE L. (2022), Acoustic emission-based damage analysis of steel fibre reinforced concrete in uniaxial tension tests, *Construction and Building Materials*, **321**: 126254, doi: [10.1016/j.conbuildmat.2021.126254](https://doi.org/10.1016/j.conbuildmat.2021.126254).
- ESSASSI K., REBIERE J.L., EL MAHI A., AMINE BEN SOUF M., BOUGUECHA A., HADDAR M. (2021), Health monitoring of sandwich composites with auxetic core subjected to indentation tests using acoustic emission, *Structural Health Monitoring*, **21**(5): 2264–2275, doi: [10.1177/14759217211053991](https://doi.org/10.1177/14759217211053991).
- FARDOUN H., SALIBA J., SAIYOURI N. (2022), Evolution of acoustic emission activity throughout fine recycled aggregate earth concrete under compressive

- tests, *Theoretical and Applied Fracture Mechanics*, **119**: 103365, doi: [10.1016/j.tafmec.2022.103365](https://doi.org/10.1016/j.tafmec.2022.103365).
11. GB/T 50081-2019 (2019), *Code for test methods of physical and mechanical properties of concrete* [in Chinese], China Architecture & Building Press, Beijing.
 12. GOYAL P., SHARMA S., KWATRA N. (2021), Evaluation of damage in GFRP repaired steel fiber reinforced concrete beams using acoustic emission technique, *Structural Concrete*, **23**(2): 907–922, doi: [10.1002/suco.202100408](https://doi.org/10.1002/suco.202100408).
 13. GOYAL P., SHARMA S., KWATRA N. (2022), Acoustic emission monitoring of steel fiber reinforced beams under simultaneous corrosion and sustained loading, *European Journal of Environmental and Civil Engineering*, **27**(4): 1535–1560, doi: [10.1080/19648189.2022.2087743](https://doi.org/10.1080/19648189.2022.2087743).
 14. GREENHOUGH J., MAIN I.G. (2008), A Poisson model for earthquake frequency uncertainties in seismic hazard analysis, *Geophysical Research Letters*, **35**(19): L19313, doi: [10.1029/2008GL035353](https://doi.org/10.1029/2008GL035353).
 15. GUO M.H., ALAM S.Y., BENDIMERAD A.Z., GRONDIN F., ROZIERE E., LOUKILI A. (2017), Fracture process zone characteristics and identification of the micro-fracture phases in recycled concrete, *Engineering Fracture Mechanics*, **181**: 101–115, doi: [10.1016/j.engfracmech.2017.07.004](https://doi.org/10.1016/j.engfracmech.2017.07.004).
 16. GUTENBERG B., RICHTER C.F. (1944), Frequency of earthquakes in California, *Bulletin of the Seismological Society of America*, **34**(4): 185–188, doi: [10.1785/BSSA0340040185](https://doi.org/10.1785/BSSA0340040185).
 17. JUNG D., YU W.R., AHN H., NA W.J. (2021), New b -value parameter for quantitatively monitoring the structural health of carbon fiber-reinforced composites, *Mechanical Systems and Signal Processing*, **165**: 108328, doi: [10.1016/j.ymsp.2021.108328](https://doi.org/10.1016/j.ymsp.2021.108328).
 18. KENCANAWATI N.N., IIZASA S., SHIGEISHI M. (2013), Fracture process and reliability of concrete made from high grade recycled aggregate using acoustic emission technique under compression, *Materials and Structures*, **46**(9): 1441–1448, doi: [10.1617/s11527-012-9986-z](https://doi.org/10.1617/s11527-012-9986-z).
 19. KWIATEK G., GOEBEL T.H.W., DRESEN G. (2014), Seismic moment tensor and b value variations over successive seismic cycles in laboratory stick-slip experiments, *Geophysical Research Letters*, **41**(16): 5838–5846, doi: [10.1002/2014GL060159](https://doi.org/10.1002/2014GL060159).
 20. LAUFF P. *et al.* (2021), Evaluation of the behavior of carbon short fiber reinforced concrete (CSFRC) based on a multi-sensory experimental investigation and a numerical multiscale approach, *Materials*, **14**(22): 7005, doi: [10.3390/ma14227005](https://doi.org/10.3390/ma14227005).
 21. LIU X.L., HAN M.S., HE W., LI X.B., CHEN D.L. (2020a), A new b -value estimation method in rock acoustic emission testing, *Journal of Geophysical Research-Solid Earth*, **125**(12): e2020JB019658, doi: [10.1029/2020JB019658](https://doi.org/10.1029/2020JB019658).
 22. LIU X.L., LIU Z., LI X.B., GONG F.Q., DU K. (2020b), Experimental study on the effect of strain rate on rock acoustic emission characteristics, *International Journal of Rock Mechanics and Mining Sciences*, **133**(9): 104420, doi: [10.1016/j.ijrmms.2020.104420](https://doi.org/10.1016/j.ijrmms.2020.104420).
 23. LOCKNER D.A., BYERLEE J.D., KUKSENKO V., PONOMAREV A., SIDORIN A. (1991), Quasi-static fault growth and shear fracture energy in granite, *Nature*, **350**: 39–42, doi: [10.1038/350039a0](https://doi.org/10.1038/350039a0).
 24. MACQUEEN J. (1967), Some methods for classification and analysis of multivariate observation, [in:] *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.
 25. MANDAL D.D. *et al.* (2022), Acoustic emission monitoring of progressive damage of reinforced concrete T-beams under four-point bending, *Materials*, **15**(10): 3486, doi: [10.3390/ma15103486](https://doi.org/10.3390/ma15103486).
 26. MENNA D.W., GENIKOMSOU A.S., GREEN M.F. (2022), Compressive and cyclic flexural response of double-hooked-end steel fiber reinforced concrete, *Frontiers of Structural and Civil Engineering*, **16**: 1104–1126, doi: [10.1007/s11709-022-0845-x](https://doi.org/10.1007/s11709-022-0845-x).
 27. PREM P.R., VERMA M., AMBILY P.S. (2018), Sustainable cleaner production of concrete with high volume copper slag, *Journal of Cleaner Production*, **193**: 43–58, doi: [10.1016/j.jclepro.2018.04.245](https://doi.org/10.1016/j.jclepro.2018.04.245).
 28. PREM P.R., VERMA M., AMBILY P.S. (2021), Damage characterization of reinforced concrete beams under different failure modes using acoustic emission, *Structures*, **30**: 174–187, doi: [10.1016/j.istruc.2021.01.007](https://doi.org/10.1016/j.istruc.2021.01.007).
 29. UNANDER T.E. (1993), The effect of attenuation on b -values in acoustic emission measurements – A theoretical investigation, *International Journal of Rock Mechanics and Mining Science & Geomechanics Abstracts*, **30**(7): 947–950, doi: [10.1016/0148-9062\(93\)90050-N](https://doi.org/10.1016/0148-9062(93)90050-N).
 30. UTSU T. (1965), A method for determining the value of b in a formula $\log n = a - bm$ showing the magnitude-frequency relation for earthquakes, *Geophysical Bulletin of Hokkaido University*, **13**: 99–103.
 31. VAPNIK V.N. (1999), An overview of statistical learning theory, [in:] *IEEE Transactions on Neural Networks*, **10**(5): 988–999, doi: [10.1109/72.788640](https://doi.org/10.1109/72.788640).
 32. WATANABE T., NISHIBATA S., HASHIMOTO C., OHTSU M. (2007), Compressive failure in concrete of recycled aggregate by acoustic emission, *Construction and Building Materials*, **21**(3): 470–476, doi: [10.1016/j.conbuildmat.2006.04.002](https://doi.org/10.1016/j.conbuildmat.2006.04.002).
 33. WEISS J. (1997), The role of attenuation on acoustic emission amplitude distributions and b -values, *Bulletin of the Seismological Society of America*, **87**(5): 1362–1367, doi: [10.1785/BSSA0870051362](https://doi.org/10.1785/BSSA0870051362).
 34. XARGAY H., RIPANI M., FOLINO P., NUNEZ N., CAGGIANO A. (2021), Acoustic emission and damage evolution in steel fiber-reinforced concrete beams under cyclic loading, *Construction and Building Materials*, **274**: 121831, doi: [10.1016/j.conbuildmat.2020.121831](https://doi.org/10.1016/j.conbuildmat.2020.121831).
 35. ZHANG L. (2020), *Study on Damage Evaluation Mechanism of Recycled Concrete in Cold Area* [in Chinese], Inner Mongolia University.

Research Paper

An Algorithm for Ultrasonic Identification of Ceramic Materials and Virtual Prototype Realization

Yu LIU⁽¹⁾, XiPing HE^{(1)*}, ShengPing HE⁽²⁾

⁽¹⁾ Shaanxi Key Laboratory of Ultrasonic, School of Physics and Information Technology
Shaanxi Normal University
Xi'an, China

⁽²⁾ High-Tech. Institute
Sichuan, Luzhou, China

*Corresponding Author e-mail: Hexiping@snnu.edu.cn

(received October 24, 2020; accepted April 2, 2020; published online July 5, 2024)

To prevent important items from being replaced by a forgery, an ultrasonic fingerprint identification algorithm is proposed and an identification program is developed. A virtual prototype for the ultrasonic identification of ceramics is developed based on an ultrasonic detection card. This virtual prototype allows for the simultaneous transmission and acquisition of signals. Numerous experimental tests were conducted using this virtual prototype. The results demonstrate that the virtual prototype achieves accurate identification of ceramics. This virtual prototype lays a good foundation for the development of intelligent, automated, integrated, and miniaturized ultrasonic identification systems.

Keywords: ceramic material; ultrasonic testing; ultrasonic scattering; ultrasonic fingerprint; virtual prototype.



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

In recent years, with the continuous progress of technology, the number of counterfeit ceramics has been increasing, and anti-counterfeiting technology is facing a great challenge. Resonance acoustic spectroscopy has been used at the Los Alamos National Laboratory, USA, to safely monitor nuclear containers of UF₆ and to inspect whether nuclear containers or chemical weapons have been tampered with (OLINGER *et al.*, 1993). Traditional ceramic identification mainly relies on organoleptic assessment, such as touching, observing, and smelling. Many counterfeit ceramics are difficult to identify by traditional methods due to the limitations of human sensory organs. Therefore, the issue of how to identify ceramics has been a concern of scholars.

There are many reports on identification methods for ceramics, including the use of X-ray diffraction, chemical element labeling, and X-ray computed tomography image reconstruction. In these methods, X-rays are transmitted through the object to be detected and

imaged on photographic film or recorded on a digital sensor. The objects can be identified according to the information presented in the image. However, there are two shortcomings to these methods: the impact of radiation on human health and radiation pollution (DEJOIE *et al.*, 2015; SCIAU *et al.*, 2011; PADELETTI *et al.*, 2010; FIGUEIREDO *et al.*, 2010).

In the current study, ultrasonic detection technology is used to identify ceramics. This method not only overcomes the limitations of traditional methods but also is not harmful to human health or the environment. All one needs to do is fix an ultrasound probe onto the surface of the object to be identified. The operating steps are more convenient and safer compared to traditional methods. In addition, the technology has many advantages, such as low cost, excellent discrimination, and high accuracy (SHI *et al.*, 2015).

The scattering of ultrasonic waves in contact with grains and interfaces in polycrystalline materials results in attenuation and dispersive velocities (BADIDI *et al.*, 2003). Random ultrasonic backscattering results from the random orientation distribu-

tion, shape, size, and interface of the grains. Moreover, the ultrasonic velocity is explicitly related to grain size. PALANICHAMY *et al.* (1995) estimated the average grain size in austenitic stainless steel using ultrasonic velocity measurements and obtained more accurate results compared to attenuation measurements. The attenuation of ultrasonic waves is closely related to the distribution of grains in polycrystalline materials. Even if the average grain size is the same for two different-shaped grains, their internal ultrasonic attenuation will be significantly different (SMITH, 1982).

An analytical equation to explain the relationship between backscattering and microtexture size and orientation has been proposed. The numerical result of this equation is consistent with the result measured using orientation image microscopy. In addition, research on the ultrasonic backscatter coefficient has clarified the complex interrelationship between polycrystalline grains and ultrasonic waves (SARPÜN *et al.*, 2005). The microstructure of polycrystalline materials determines the ultrasonic velocity, attenuation, and backscatter power of ultrasonic waves propagating inside the material (LAUX *et al.*, 2002; HIRAO *et al.*, 1987; ÖZKAN *et al.*, 2013; VIJAYALAKSHMI *et al.*, 2011; MURTHY *et al.*, 2008). Conversely, ultrasound can also effectively characterize the difference in the internal microstructure of polycrystalline materials. Each finished material in industrial production has unique microstructural characteristics, just like human fingerprints. Using ultrasonic fingerprint technology to prevent valuables from being replaced has been a new application in the field of ultrasonic testing in recent years. When ultrasonic waves propagate in an object, the scattering signal is highly correlated with its internal structure (LI *et al.*, 2014; BUENOS *et al.*, 2014).

In this study, scattering signals of three ceramic disks of identical material and appearance are extracted using an ultrasonic probe with a frequency of 5 MHz, and the ultrasonic fingerprints of the signals are extracted to identify the ceramic specimens.

2. Identification algorithm

Ultrasonic reflection signals and scattering signals are acquired from the ceramics that need to be protected. The features of the signals are processed to calculate the “target ultrasonic fingerprint”. When ceramics with the same appearance are mixed, the ultrasonic signals are acquired from each specimen, respectively. After processing the signals acquired by the virtual prototype, features are extracted from the signal of each specimen, and the ultrasonic fingerprints to be identified are calculated based on these features. The identification is completed by comparing the “fingerprints to be identified” with the “target ultrasonic fingerprint”. This process involves both time-domain identification and frequency-domain identifica-

tion. Whether the ceramic to be identified is the target ceramic is determined according to the identification results of the ultrasonic fingerprints.

First, to reduce the effect of errors and improve anti-interference capability, the acquired time-domain signals are normalized using:

$$A_n = \frac{X_n - X_{\min}}{X_{\max} - X_{\min}}, \quad (1)$$

where X_{\min} is the minimum amplitude of the signal, X_{\max} is the maximum amplitude, and X_n and A_n are, respectively, the amplitude of each sampling point before and after normalization (where $n = 1, 2, 3, \dots, N$).

Second, a parameter P_i (where $i = 1, 2, 3, \dots, 20$) is defined as follows:

$$P_i = \frac{A_n}{\sqrt{\sum_{n=1}^N A_n^2}}. \quad (2)$$

Twenty sets of signals were acquired as target signals in each experiment to reduce operational and systematic errors. Therefore, there are 20 sets of P_i . Subsequently, the arithmetic mean P_v was calculated based on P_i . The feature of each target time-domain signal F_i was calculated according to:

$$F_i = \sum_{n=1}^N \left| \ln \frac{P_i}{P_v} \right|. \quad (3)$$

The target signal features in the frequency domain are similar to those in the time domain, with the difference being that A_n is obtained using a fast Fourier transform after Eq. (1). P_i is calculated by using the frequency spectrum of A_n in Eq. (2), and the features in the frequency domain can be calculated by using Eq. (3). Finally, there is one feature in the time domain and one feature in the frequency domain, obtained by averaging the features of the 20 sets of signals. These two features are, respectively, the ultrasonic fingerprints in the time and frequency domains.

The steps for extracting the ultrasonic fingerprints of the signals to be identified are similar to those used for the target signals. It is worth noting that P_v of the signals to be identified is still the value while calculating the target signal features. This means that the fingerprint to be identified is calculated based on the target fingerprint. The fingerprint to be identified indicates the dissimilarity compared to the target fingerprint. The ultrasonic fingerprints in the time and frequency domains need to be compared with the target signal ultrasonic fingerprints, respectively, to improve the accuracy of the identification. Although the target ultrasonic fingerprints and the ultrasonic fingerprints to be identified have been obtained from the identical ceramic, they will differ somewhat under actual conditions because of inevitable errors such as operating error and system noise. Therefore, it is necessary to

calculate a threshold. The variation of signal amplitude caused by error is a random variable with independent distribution. In statistical analysis, all signals that may be collected are statistical populations, while the signals collected from the populations are statistical samples. According to the central limit theorem, the distribution of the sample mean will approximate a Gaussian distribution, regardless of the population distribution. Therefore, it is feasible to judge the outliers by using the Pauta criterion.

In statistics, three times the value of the standard deviation is considered as a reasonable margin of error and measurements beyond the reasonable margin of error are identified as outliers. Therefore, the standard deviation can be calculated by using:

$$\sigma = \sqrt{\frac{\sum_{i=1}^k (\bar{F} - F_k)^2}{k}}, \quad (4)$$

where \bar{F} is the mean value of the target signal ultrasonic fingerprints and k is the number of target signals ($k = 20$ in this study). The threshold Q is then given as:

$$Q = \bar{F} + 3\sigma. \quad (5)$$

Therefore, according to the Pauta criterion, the probability of a deviation greater than 3σ is ~ 0.0026 , which is a rare probability event.

The result of the identification is obtained by assessing whether the ultrasonic fingerprints to be identified are within the range of the threshold.

3. Experimental instruments

In this study, four ceramic plates of the same shape and material were used as specimens, as shown in Fig. 1. Each of the plates had a diameter of 175 mm and a bottom thickness of 5 mm. The velocity of ultrasound in the plates was 6250 m/s. The velocity of ultrasound propagation in the 5 mm thick plate was



Fig. 1. Right-angle positioner and specimens.

obtained by measuring the time difference between the first and second ultrasonic echoes. A dark blue plastic sheet cut at right angles was pasted on the surface of the ceramic plate. The position of the probe was determined when the right angle sides of the plastic slice were tangent to circular probe. The plastic sheet plays an important role in accurately fixing the probe position.

Figure 2 shows the ultrasonic flaw detection card (CTS-04PC) customized from Shantou Ultrasonic Electronics Company (China). This model of ultrasonic flaw detection card is available for Peripheral Component Interconnect (PCI) slots and ultrasonic probes. The function of the card depends on the operating status of the registers. The computer controls the base address and offset of the registers to control the working mode of the ultrasonic flaw detection card. Therefore, the probe is connected to the ultrasonic flaw detection card, which is then assembled in the PCI slot of the computer. Together, all these components form a virtual prototype, as shown in Fig. 3.



Fig. 2. Ultrasonic flaw detection card.



Fig. 3. Virtual prototype.

The internal structure of the virtual prototype is shown in Fig. 4. The operational program for the ultrasonic flaw detection card was written in C++. The sampling frequency of the system was 100 MHz, and the excitation waveform consisted of a positive half-wave, a negative half-wave, and a radio-frequency wave. The excitation voltage could be regulated from 0 V to 300 V, and the gain range could be adjusted from 0 dB to 110 dB. The pulse repetition rate was 10 kHz. Damping, ultrasonic velocity, display range of the waveform, and other parameters can all be adjusted according to actual conditions.

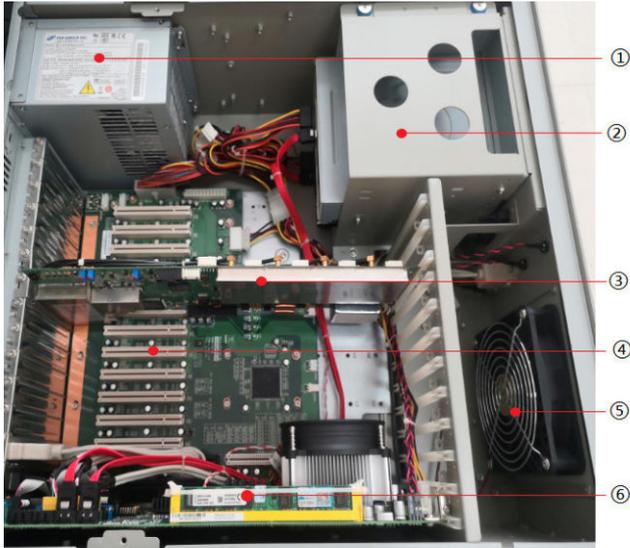


Fig. 4. Internal structure of the virtual prototype: 1) a PC power supply unit; 2) an optical disk driver; 3) an ultrasonic flaw detection card; 4) PCI slots; 5) a cooling fan; 6) a computer motherboard.

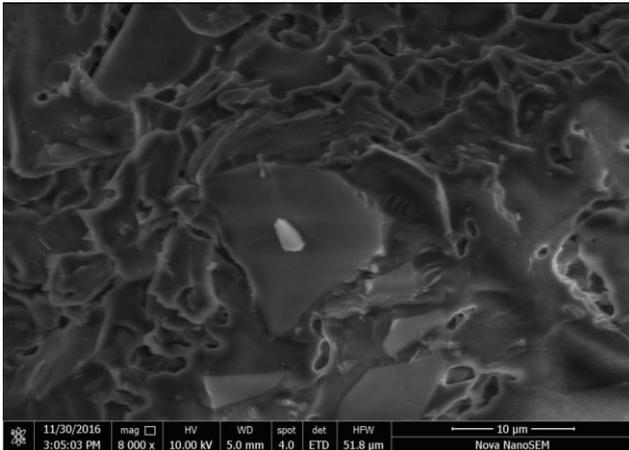
4. Identification of ceramics and results

4.1. Microstructure of ceramics

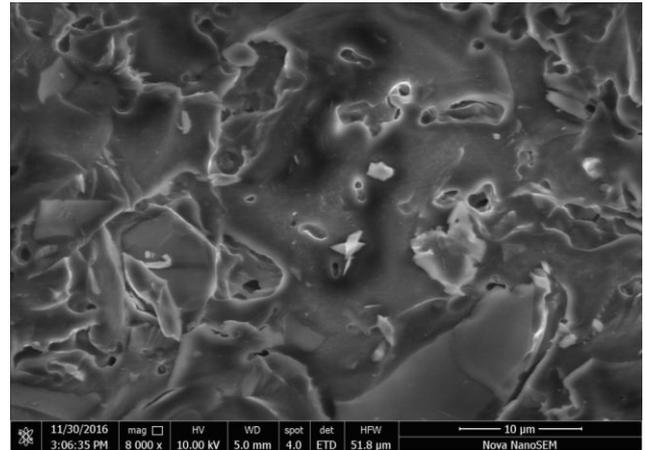
Cross-sectional scanning electron microscope (SEM) images of three ceramic specimens studied during the experiment are shown in Figs. 5a-c. The sizes, shapes, and interfaces of the ceramic microtextures are all quite different. According to the SEM results, the granularity of the ceramics used in the experiments ranges from $0.05\ \mu\text{m}$ to $0.3\ \mu\text{m}$. These grains are arranged in a disorganized manner, and the boundaries between the particles are not clear. The SEM images show that the internal microstructure of different portions of the same ceramic sample also varies greatly.

The interaction between ultrasonic waves and the ceramic material is closely related to the wavelength of ultrasound and the size of the ceramic particles. The wavelength λ of the 5 MHz ultrasonic waves used in the experiment is 1.2 mm inside the ceramic sample, and the particle sizes of ceramics are in the range

a)



b)



c)

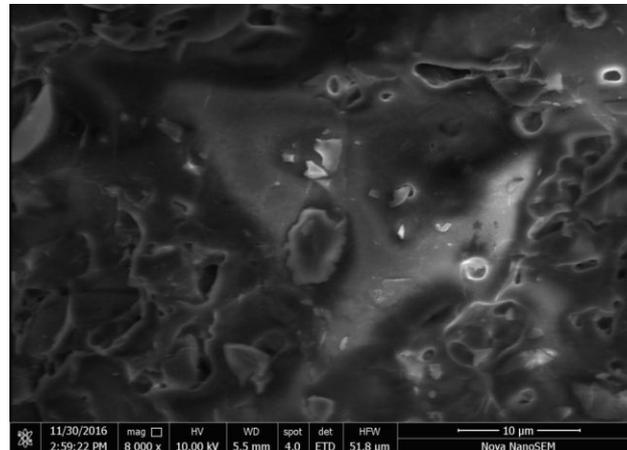


Fig. 5. SEM images of ceramic cross-sections.

of $0.5\ \mu\text{m}\sim 3\ \mu\text{m}$. The wavelength is much larger than the particle sizes, satisfying the condition for Rayleigh scattering. The scattering coefficient is proportional to the fourth power of the frequency, meaning higher frequency is conducive to improving the intensity of scattering signals and extracting more features from the specimens (YANG *et al.*, 2012). Because of the irregular structure inside ceramics, the scattered waves generated by identical parts of different ceramics vary, and consequently, the characteristics of the scattered signals received by the probe also differ, which is the basis for identifying the ceramic samples.

4.2. Identification steps

Firstly, signals were acquired from the samples to be protected using an ultrasonic probe and saved by the virtual prototype. Secondly, ultrasonic signals from all the samples were extracted using the same ultrasonic probe and saved as additional data. Thirdly, features of the target ultrasound fingerprints were calculated in both time and frequency domains. In this way, these feature thresholds were then obtained and saved as the “target ultrasonic fingerprint”. Fourthly, features of the ultrasonic fingerprints to be identified were calculated in both time and frequency domains and saved as the “ultrasonic fingerprint to be identified”. Finally, results were obtained according to the relationship between the features of “ultrasonic fingerprint to be identified” and the “target ultrasonic fingerprint”.

If the features of the “ultrasonic fingerprint to be identified” fall below the thresholds in both time and frequency domains, the result is “same item”. If the features are above the thresholds in both time and frequency domains, the result is “different item”. However, if the features are below the threshold in one domain and above it in another, the result is “uncertain”.

Increasing the number of signal acquisition makes the distribution of signals more stable and improves the accuracy of identification. Target signals should be acquired in 20 sets, while signals to be identified should be acquired in more than 5 sets based on previous experience. The result with the highest frequency of occurrences is the final result. If the result is “uncertain”, the signal to be identified needs to be re-extracted until a clear result is obtained.

4.3. Identification of ceramic specimens

The first step involves configuring the parameters of the virtual prototype by, for example, setting the ultrasonic velocity according to the specimen material. The frequency of the excitation signal is 5 MHz. The sampling time should be adjusted to ensure the number of the ultrasonic echoes is more than 5. In this study, a sampling time of $20\ \mu\text{s}$ is sufficient for the samples. It should be noticed that varying sampling

time will lead to different signal acquisitions. For consistent identification of a specific item, the sampling time must remain consistent in all signal acquisitions.

The second step includes adjusting the gain based on the amplitude feature of the specimen. Ultrasonic wave signals in various materials are different. Adjusting the gain settings according to the specific material features is beneficial to improve identification accuracy. The upper limit of the amplitude of the virtual prototype is 250 V, and it is important to avoid an amplitude exceeding this limit during gain adjustment.

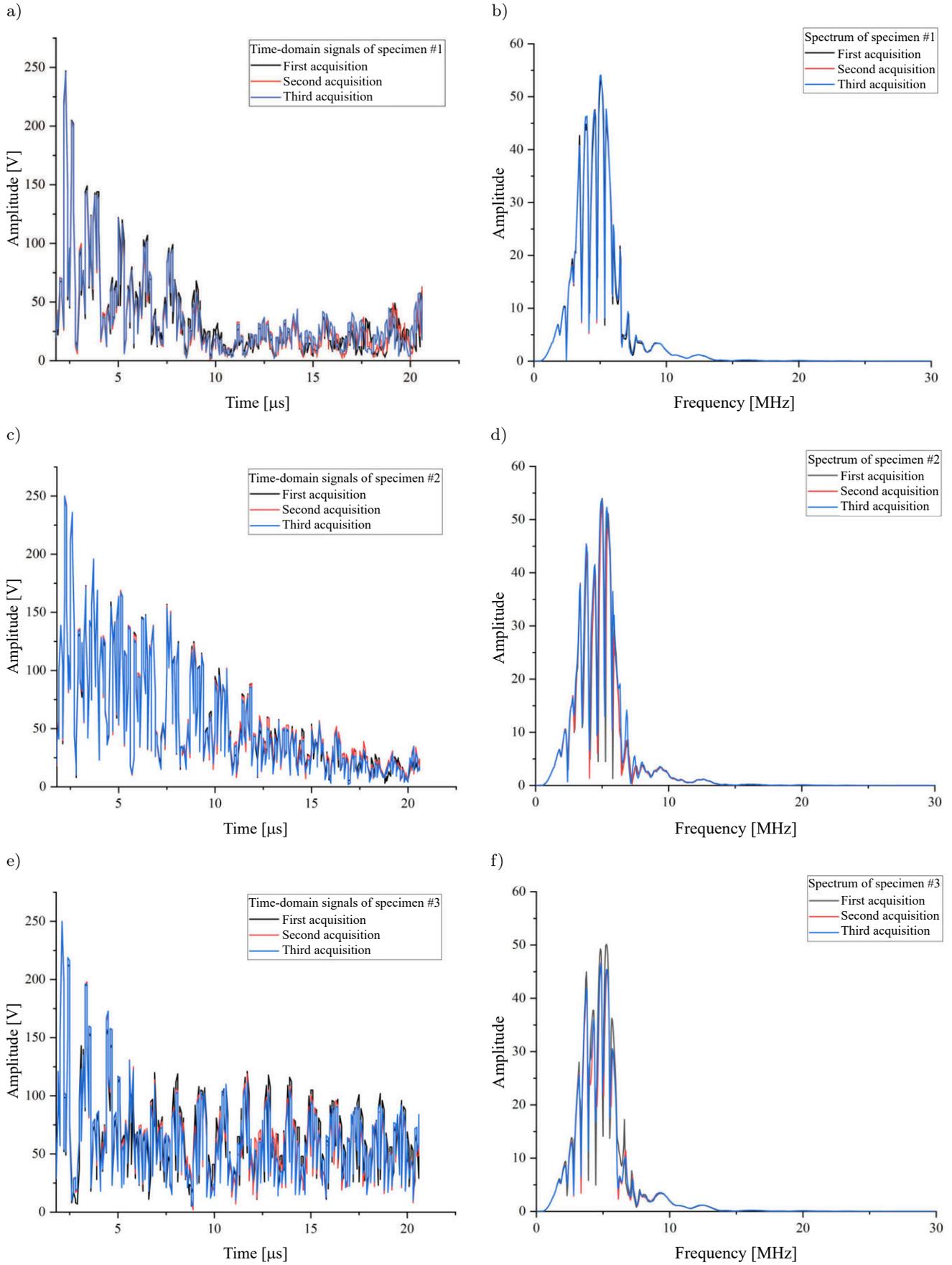
The third step involves signal acquisition. Making sure a probe is placed in the same position each time is important because signals vary in different probe’s positions, even when they are from the same specimen.

Eventually, water is used as the coupling agent. It is crucial to keep the bottom of the specimen dry because ultrasonic waves pass through the coupling agent to reach the bottom of the specimen, where they reflect off the surface on which the specimen is placed.

The experimental specimens were four ceramic disks made from the same material and having identical appearance, as shown in Fig. 1. The specimens were labeled as 1–4. Specimen #1 was designated as the target specimen. Twenty sets of ultrasonic signals were acquired for the target specimen (#1), while seven sets were acquired for every specimen to be identified (#1, #2, #3, and #4). These signals were processed to obtain ultrasonic fingerprints. The time-domain signals of the ceramic specimens are shown in Fig. 6. A comparison of the amplitudes of these time-domain signals clearly shows that the ultrasonic signals of these specimens varied significantly. Although the waveforms of specimens #3 and #4 appear similar, the details differed enormously, which could be identified by ultrasonic fingerprints.

The ultrasonic fingerprints of the target specimen (#1) are listed in Table 1. The ultrasonic fingerprints to be identified of specimens #1 and #2 are listed in Tables 2 and 3, respectively. For brevity, partial data in Table 1 have been omitted. The ultrasonic fingerprints for specimens #3 and #4 are not included in this study as their results were similar to specimen #2 when compared with the target specimen (#1). The results of the identification experiment for each specimen are given in Table 4.

The arithmetic mean \bar{F} and standard deviation σ of the target ultrasonic fingerprints are calculated from Table 1. Subsequently, thresholds in the time and frequency domains are obtained using Eq. (5). These thresholds are given in Tables 2 and 3 to compare with the ultrasonic fingerprints to be identified. Table 2 indicates that six of the seven ultrasonic fingerprints to be identified for specimen #1 fell below the thresholds in both time and frequency domains. Only the third ultrasonic fingerprint could not be identified as the “same item” because the ultrasonic fingerprint value in the



[Fig. 6a-f].

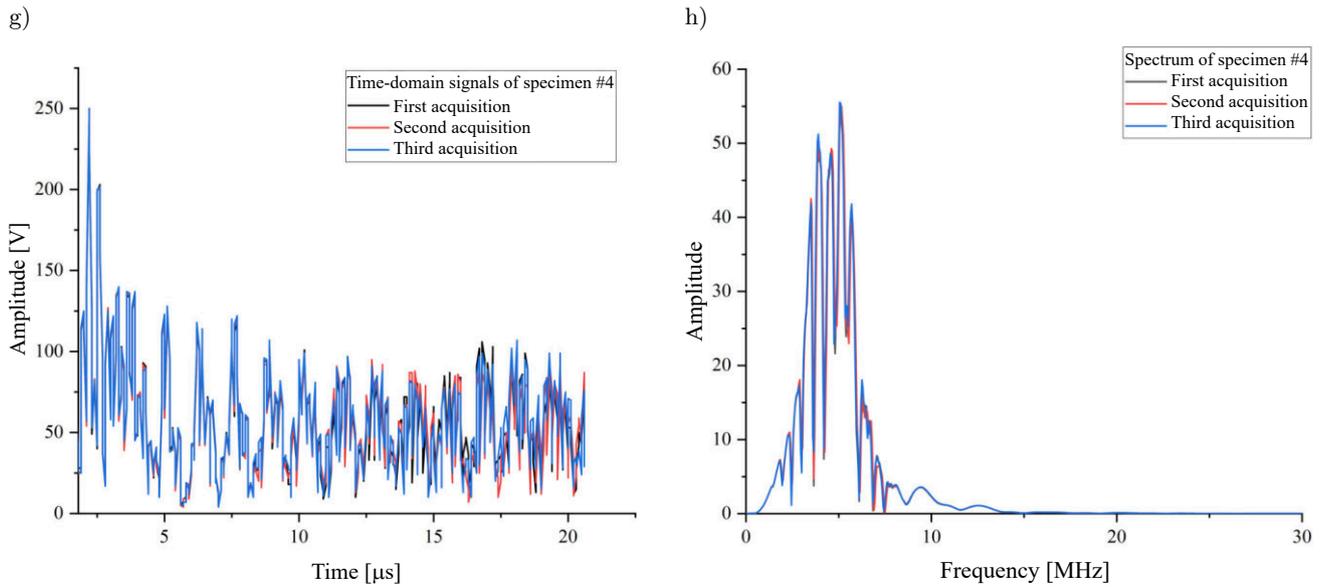


Fig. 6. Time-domain signals and spectra of the specimens at the same position.

Table 1. Target ultrasonic fingerprints of specimen #1.

Domain	Number							
	1	2	3	4	5	...	19	20
Time	4.0507	3.2753	3.8776	3.7949	4.3667	...	4.4440	3.4480
Frequency	24.0818	20.5954	25.5155	25.2757	26.1027	...	28.4524	23.4338

Table 2. Identification of specimen #1 by using ultrasonic fingerprints.

Domain	Number							Threshold
	1	2	3	4	5	6	7	
Time	4.0275	4.5044	6.1990	4.9447	5.0663	4.6045	4.8150	6.1483
Frequency	27.8061	26.4829	31.1543	26.9516	27.7434	28.6894	26.8546	36.6512

Table 3. Identification of specimen #2 by using ultrasonic fingerprints.

Domain	Number							Threshold
	1	2	3	4	5	6	7	
Time	18.4371	18.3059	18.1403	18.4292	18.5042	18.4900	18.4638	6.1483
Frequency	74.2805	72.5703	72.1178	71.8594	72.0542	73.1422	71.8704	36.6512

Table 4. Identification of each specimen by using ultrasonic fingerprints.

Identification specimen	#1 identifies				#2 identifies				#3 identifies				#4 identifies			
	#1	#2	#3	#4	#1	#2	#3	#4	#1	#2	#3	#4	#1	#2	#3	#4
Same	6	0	0	0	0	6	0	0	0	0	7	0	0	0	0	6
Different	0	7	7	7	7	0	7	7	7	7	0	7	7	7	7	0
Result	Same	Different	Different	Different	Different	Same	Different	Different	Different	Different	Same	Different	Different	Different	Different	Same

frequency domain was lower than the threshold while its time domain value was higher than the threshold. Therefore, the result for the third ultrasonic fingerprint was identified as “uncertain”. With 6 results indicating “same item” and 0 indicating “different item”, the final

identification for specimen #1 was concluded as “same item”.

Table 3 indicates that ultrasonic fingerprints to be identified for specimen #2 are much higher than both the time and frequency thresholds. Obviously,

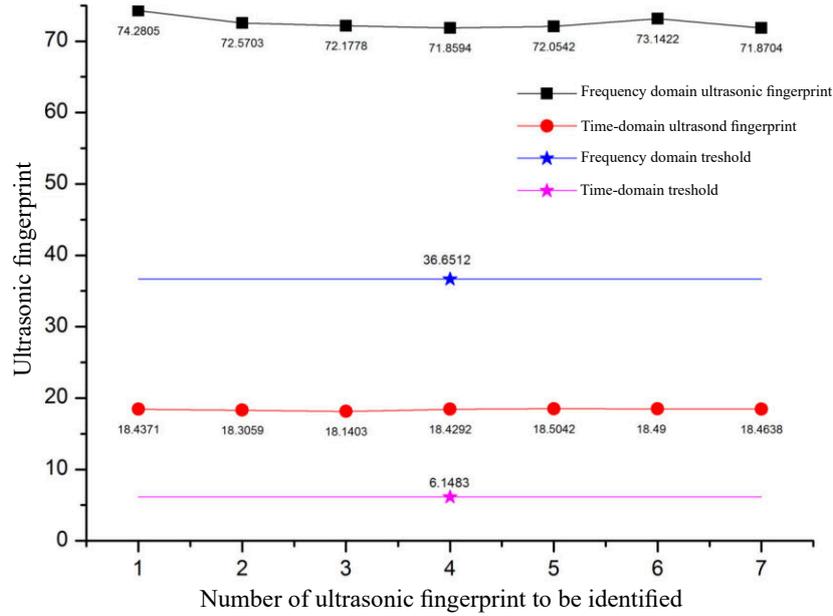


Fig. 7. Ultrasonic fingerprints of specimen #2.

all these seven ultrasonic fingerprints are identified as “different items”. The relationship between ultrasonic fingerprints and their respective thresholds is plotted in Fig. 7. Furthermore, the ultrasonic fingerprints of the four samples were treated as target ultrasonic fingerprints in the proper sequence and then identified with each other. As presented in Table 4, each ceramic sample was accurately identified even when they were mixed up.

4.4. Experiments on other specimens

In addition to the ceramic plate specimens, we also performed experimental verification on three ceramic boxes, three ceramic sinks, and three round

ceramic pots, as shown in Fig. 8. The sinks had a length of 53 cm, a width of 38 cm, a depth of 53 cm, and a thickness of 10 mm. The boxes had a length of 53 mm, a width of 53 mm, and a bottom thickness of 5 mm. The pots had a diameter of 58 mm and a bottom thickness of 3 mm.

The ceramic sinks were labeled as 5–7. Specimen #5 was designated as the target specimen. The time-domain signals and spectra of the ceramic sinks are shown in Fig. 9.

The ultrasonic fingerprints of the target specimen (#5) are listed in Table 5. The ultrasonic fingerprints to be identified for specimens #5 and #6 are listed in Tables 6 and 7, respectively. For brevity, partial data in Table 5 have been omitted.

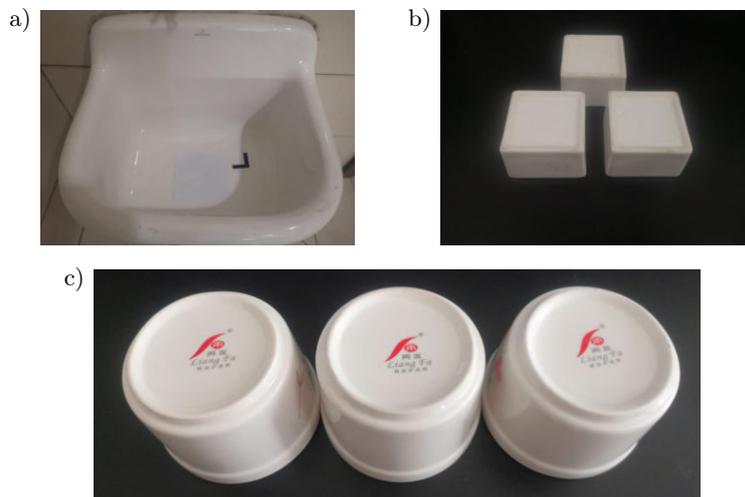


Fig. 8. Additional ceramic samples: a) ceramic sinks; b) ceramic boxes; c) ceramic pots.

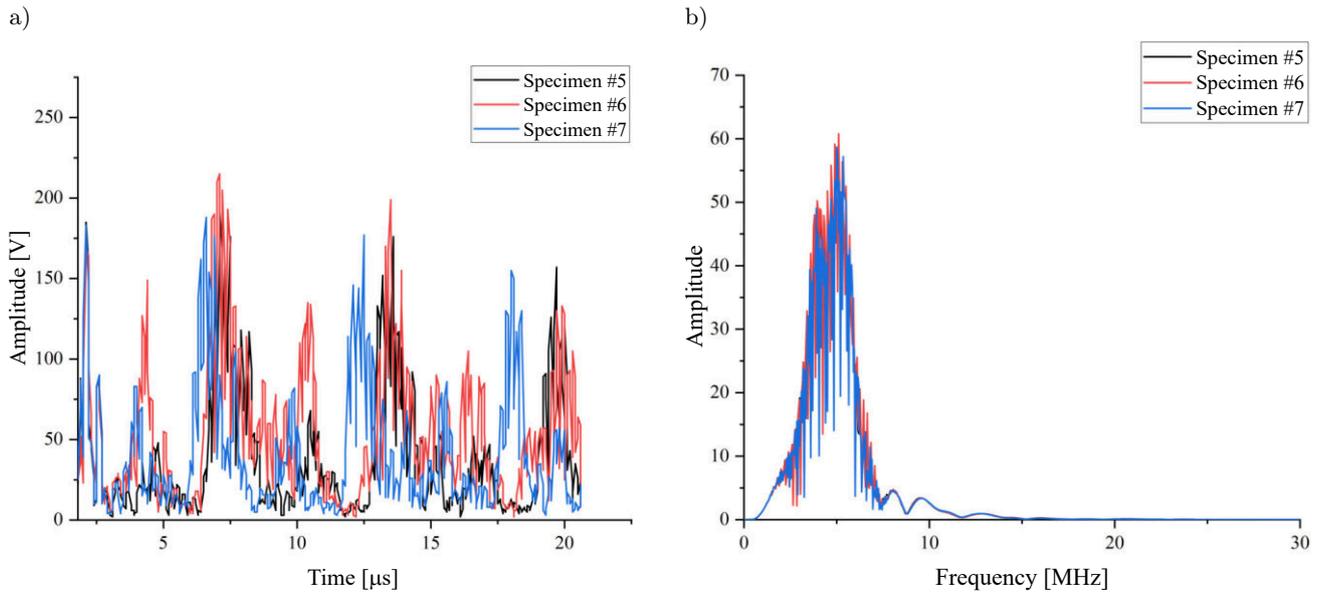


Fig. 9. Time-domain signals and spectra of the ceramic sinks.

Table 5. Target ultrasonic fingerprints of specimen #5.

Domain	Number							
	1	2	3	4	5	...	19	20
Time	4.9865	5.1862	5.0886	5.0006	4.9458	...	6.5132	4.3933
Frequency	43.5764	50.5167	49.7735	50.7363	53.1878	...	55.4088	51.8592

Table 6. Identification of specimen #5 by using ultrasonic fingerprints.

Domain	Number							Threshold
	1	2	3	4	5	6	7	
Time	3.0299	3.4455	4.2011	4.6057	2.7799	2.9915	3.5278	8.1088
Frequency	33.1390	34.4481	42.9251	54.3140	32.4684	33.4754	31.9689	58.9283

Table 7. Identification of specimen #6 by using ultrasonic fingerprints.

Domain	Number							Threshold
	1	2	3	4	5	6	7	
Time	22.7507	25.8529	26.6743	27.4999	28.6314	29.4339	29.7828	8.1088
Frequency	123.8227	128.1348	128.4216	129.5173	135.1324	126.7768	127.1504	58.9283

The ceramic boxes and pots were labeled as 8–10 and 11–13, respectively. The time domain signals and spectra of the ceramic specimens are shown in Figs. 10 and 11.

The ultrasonic fingerprints for the target specimen (#8) are listed in Table 8. The ultrasonic fingerprints to be identified of specimens #8 and #9 are listed in Tables 9 and 10, respectively.

Specimens #8 and #9 are accurately identified as “same item” and “different item”, respectively, according to the relationship between the ultrasonic fingerprints and the threshold.

The ultrasonic fingerprints of the target specimen (#11) are listed in Table 11. The ultrasonic fingerprints to be identified of specimens #11 and #12 are listed in Tables 12 and 13, respectively.

Each of the above ceramic specimens was identified correctly during the experiments. Partial ultrasonic fingerprint data have been omitted for brevity. In addition, the ceramic samples were completely replaced with metal samples during the identification experiments. The identification of metal specimens is also accurate, which means the ultrasonic fingerprinting can also be applied to metallic materials.

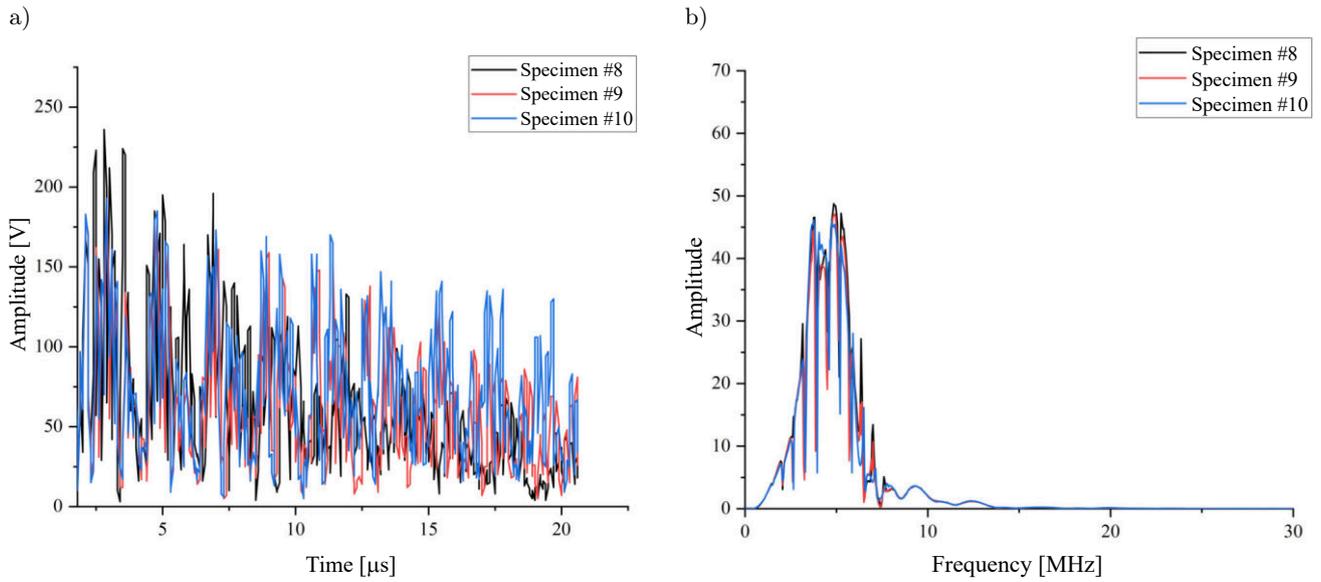


Fig. 10. Time-domain signals and spectra of the ceramic boxes.

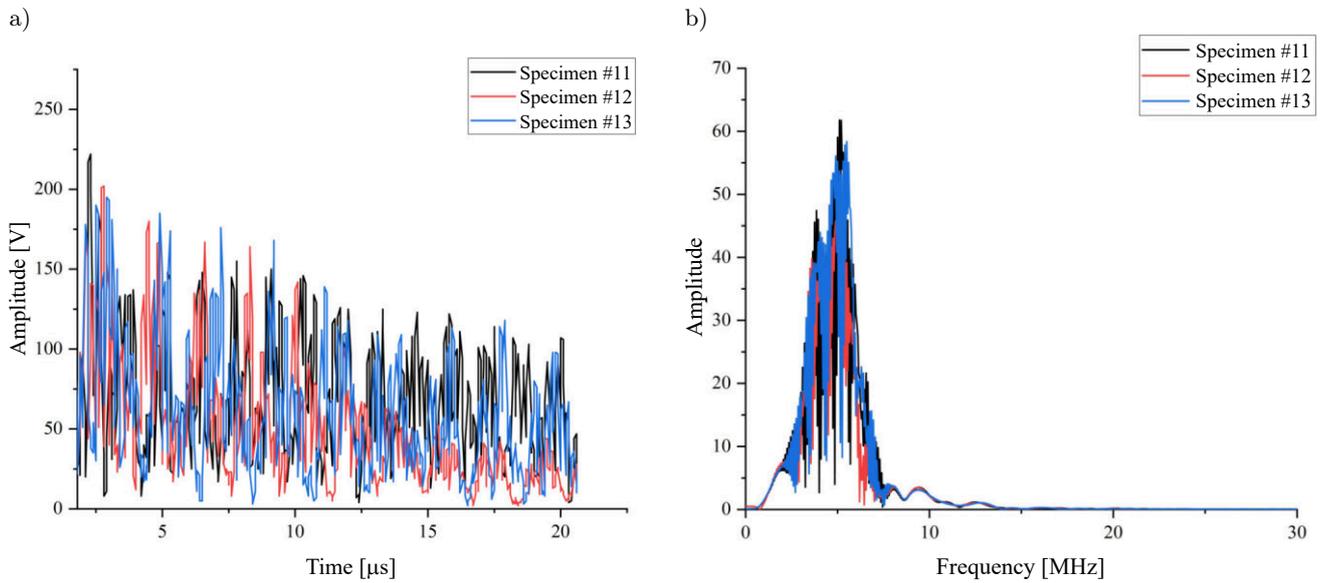


Fig. 11. Time-domain signals and spectra of the ceramic pots.

Table 8. Target ultrasonic fingerprints of specimen #8.

Domain	Number							
	1	2	3	4	5	...	19	20
Time	2.0344	1.1269	1.1505	2.1904	1.3621	...	1.2235	1.1005
Frequency	24.1381	11.4562	11.7572	25.3966	12.2876	...	13.9796	13.7000

Table 9. Identification of specimen #8 by using ultrasonic fingerprints.

Domain	Number							Threshold
	1	2	3	4	5	6	7	
Time	1.3398	1.2174	1.4260	1.5649	1.4916	1.3487	1.2371	2.7580
Frequency	17.2821	14.0739	15.7114	18.2949	14.5042	15.8238	17.0235	21.4454

Table 10. Identification of specimen #9 by using ultrasonic fingerprints.

Domain	Number							Threshold
	1	2	3	4	5	6	7	
Time	5.8814	6.2411	7.1876	7.5086	8.1488	7.5317	7.8776	2.7580
Frequency	60.7242	65.6821	74.9572	78.9511	84.5907	86.5673	88.2513	21.4454

Table 11. Target ultrasonic fingerprints of specimen #11.

Domain	Number							
	1	2	3	4	5	...	19	20
Time	4.1651	3.2998	5.1291	3.2866	3.7905	...	4.0025	4.2389
Frequency	47.9999	31.7008	48.6838	34.2596	38.4055	...	39.5561	47.6875

Table 12. Identification of specimen #11 by using ultrasonic fingerprints.

Domain	Number							Threshold
	1	2	3	4	5	6	7	
Time	3.6012	3.5054	3.1888	3.1953	3.4661	3.3552	3.8833	6.0462
Frequency	31.2349	32.8217	30.9763	31.0628	31.9996	31.7118	39.2756	52.0035

Table 13. Identification of specimen #12 by using ultrasonic fingerprints.

Domain	Number							Threshold
	1	2	3	4	5	6	7	
Time	20.4706	22.0224	24.1944	24.6466	21.3210	23.9550	22.1081	6.0462
Frequency	130.9253	125.4786	128.0756	130.5101	116.5375	132.6178	132.0513	52.0035

5. Conclusions

In this study, the notion of ultrasonic fingerprints was presented to identify and protect ceramics. An algorithm to extract ultrasound signal features was developed as an identification program. The ultrasonic flaw detection card and the computer were assembled into a virtual prototype to integrate the ultrasonic fingerprint acquisition system and the identification system. Then, experiments were conducted to identify a variety of ceramic specimens. The experimental results indicated that the ceramic specimens can be identified and distinguished accurately. The development of the virtual prototype also provides a good foundation for advancing the intelligence, automation, integration, and miniaturization of ultrasonic fingerprint identification systems.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 12174241).

References

- BADIDI B.A., LEBAILI S., BENCHALA A. (2003), Grain size influence on ultrasonic velocities and attenuation, *NDT & E International*, **36**(1): 1–5, doi: [10.1016/S0963-8695\(02\)00043-9](https://doi.org/10.1016/S0963-8695(02)00043-9).
- BUENOS A.A., JR P.P., MEI P.R., SANTOS A.A. (2014), Influence of grain size on the propagation of L_{CR} waves in low carbon steel, *Journal of Nondestructive Evaluation*, **33**: 562–570, doi: [10.1007/s10921-014-0252-x](https://doi.org/10.1007/s10921-014-0252-x).
- DEJOIE C., TAMURA N., KUNZ M., GOUDEAU P., SCIAU P. (2015), Complementary use of monochromatic and white-beam X-ray micro-diffraction for the investigation of ancient materials, *Journal of Applied Crystallography*, **48**: 1522–1533, doi: [10.1107/S1600576715014983](https://doi.org/10.1107/S1600576715014983).
- FIGUEIREDO E., SILVA R.J.C., ARAÚJO M.F., MARTINEZ J.C.S. (2010), Identification of ancient gilding technology and late bronze age metallurgy by EDXRF, Micro-EDXRF, SEM-EDS and metallographic techniques, *Microchimica Acta*, **168**: 283–291, doi: [10.1007/s00604-009-0284-6](https://doi.org/10.1007/s00604-009-0284-6).
- HIRAO M., AOKI K., FUKUOKA H. (1987), Texture of polycrystalline metals characterized by ultrasonic velocity measurements, *Journal of the Acoustical Society of America*, **81**(5): 1434–1440, doi: [10.1121/1.394495](https://doi.org/10.1121/1.394495).
- LAUX D., CROS B., DESPAUX G., BARON D. (2002), Ultrasonic study of UO_2 : Effects of porosity and grain size on ultrasonic attenuation and velocities, *Journal of Nuclear Materials*, **300**(2–3): 192–197, doi: [10.1016/S0022-3115\(01\)00747-4](https://doi.org/10.1016/S0022-3115(01)00747-4).

7. LI J., YANG L., ROKHLIN S.I. (2014), Effect of texture and grain shape on ultrasonic backscattering in polycrystals, *Ultrasonics*, **54**(7): 178–1803, doi: [10.1016/j.ultras.2014.02.020](https://doi.org/10.1016/j.ultras.2014.02.020).
8. MURTHY G.V.S., GHOSH S., DAS M., DAS G., GHOSH R.N. (2008), Correlation between ultrasonic velocity and indentation-based mechanical properties with microstructure in Nimonic 263, *Materials Science and Engineering: A*, **488**(1–2): 398–405, doi: [10.1016/j.msea.2007.11.017](https://doi.org/10.1016/j.msea.2007.11.017).
9. OLINGER C.T., LYON M.J., STANBRO W.D., MULLEN M.F., SINHA D.N. (1993), Acoustic resonance spectroscopy in nuclear safeguards, [in:] *34th Annual Meeting of the Institute of Nuclear Materials Management*, <https://www.osti.gov/biblio/61350>.
10. ÖZKAN V., SARPÜN İ.H., EROL A., YÖNETKEN A. (2013) Influence of mean grain size with ultrasonic velocity on microhardness of B₄C-Fe-Ni composite, *Journal of Alloys and Compounds*, **574**(15): 512–519, doi: [10.1016/j.jallcom.2013.05.097](https://doi.org/10.1016/j.jallcom.2013.05.097).
11. PADELETTI G., FERMO P. (2010), A scientific approach to the attribution problem of renaissance ceramic productions based on chemical and mineralogical markers, *Applied Physics A*, **100**: 771–784, doi: [10.1007/s00339-010-5689-x](https://doi.org/10.1007/s00339-010-5689-x).
12. PALANICHAMY P., JOSEPH A., JAYAKUMAR T., RAJ B. (1995), Ultrasonic velocity measurements for estimation of grain size in austenitic stainless steel, *NDT & E International*, **28**(3): 179–185, doi: [10.1016/0963-8695\(95\)00011-L](https://doi.org/10.1016/0963-8695(95)00011-L).
13. SARPÜN H.I., KILIÇKAYA S.M. (2005), Mean grain size determination in marbles by ultrasonic first backwall echo height measurements, *NDT & E International*, **39**(1): 82–86, doi: [10.1016/j.ndteint.2005.06.010](https://doi.org/10.1016/j.ndteint.2005.06.010).
14. SCIAU P., LEON Y., GOUDEAU P., FAKRA S.C., WEBBD S., MEHTA A. (2011), Reverse engineering the ancient ceramic technology based on X-ray fluorescence spectromicroscopy, *Journal of Analytical Atomic Spectrometry*, **26**(5): 969–976, doi: [10.1039/C0JA00212G](https://doi.org/10.1039/C0JA00212G).
15. SHI S., LIU Z.G., SUN J.T., ZHANG M., DU G.S., LI D. (2015), Study of errors in ultrasonic heat meter measurements caused by impurities of water based on ultrasonic attenuation, *Journal of Hydrodynamics*, **27**: 141–149, doi: [10.1016/S1001-6058\(15\)60466-8](https://doi.org/10.1016/S1001-6058(15)60466-8).
16. SMITH R.L. (1982), The effect of grain size distribution on the frequency dependence of the ultrasonic attenuation in polycrystalline materials, *Ultrasonics*, **20**(5): 211–214, doi: [10.1016/0041-624X\(82\)90021-X](https://doi.org/10.1016/0041-624X(82)90021-X).
17. VIJAYALAKSHMI K., MUTHUPANDI V., JAYACHITRA R. (2011), Influence of heat treatment on the microstructure, ultrasonic attenuation and hardness of SAF 2205 duplex stainless steel, *Materials Science and Engineering A*, **529**(25): 447–451, doi: [10.1016/j.msea.2011.09.059](https://doi.org/10.1016/j.msea.2011.09.059).
18. YANG L., LI J., LOBKIS O.I., ROKHLIN S.I. (2012), Ultrasonic propagation and scattering in duplex microstructures with application to titanium alloys, *Journal of Nondestructive Evaluation*, **31**: 270–283, doi: [10.1007/s10921-012-0141-0](https://doi.org/10.1007/s10921-012-0141-0).

Research Paper

**Excitation of the Secondary Modes by the Broad Spectrum Sound
in a Liquid with Relaxation Losses**

Anna PERELOMOVA

*Faculty of Applied Physics and Mathematics
Gdansk University of Technology*Gdańsk, Poland; e-mail: anna.perelomova@pg.edu.pl*(received June 30, 2023; accepted April 16, 2024; published online July 11, 2024)*

Features of nonlinear phenomena and, in particular, acoustic excitation of the entropy and relaxation modes in a liquid electrolyte with a chemical reaction are examined. The total range of frequencies of an exciter is considered, and the instantaneous dynamic equations are derived which govern perturbations in the secondary modes. The instantaneous leading-order acoustic forces of the secondary modes are evaluated. Examples of harmonic and nearly harmonic acoustic exciter are considered in detail. The difference in the nonlinear acoustic phenomena in an electrolyte and gases with relaxation mechanisms are specified and discussed.

Keywords: relaxation; dispersive media; acoustic heating.



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

The frequency-dependent attenuation of sound in fluids is of importance in many technical and medical applications, especially in medical imaging and tissue specification. It is well-known that the Newtonian attenuation often does not match the experimental data of attenuation in tissues (PARKER, 1983; DUCK *et al.*, 1998). There are variety of mechanisms of relaxation leading to dispersion and attenuation of sound, such as chemical reactions and vibrational relaxation of the molecules. The comprehensive review concerning various mechanisms of absorption in gases can be found in the book written by PIERCE (1981).

The mechanisms of relaxation and absorption in liquids in the context of wave dynamics, is still unresolved issue since the problem was first mentioned (LIEBERMANN, 1948). The composition of a liquid has a key role in a variety of relaxation processes. Typically, several kinds of relaxation take place in a liquid (PIERCE, MAST, 2021). Sound absorption in the sea water is dominated by chemical relaxations with contribution of magnesium sulfate at high frequencies of oscillations and boric acid at low frequencies of oscillations. There is an intermediate-frequency relaxation due to magnesium and carbonic acid (YEA-GER, FISHER, 1973; MELLEN *et al.*, 1979). A detail

description of wave processes in the sea water is of importance in view of fishery, naval, and biological applications. The experimental studies of wave processes in water, aqueous solutions and other liquids, including electrolytes, confidently indicate the kinds of relaxation in a liquid and its composition. The dispersion which follows relaxation in fluids, makes the sound speed and attenuation frequency-dependent. Theoretical foundations of spatially distributed relaxation processes with the frequency power law attenuation have been discussed recently by PIERCE and MAST (2021). Shear viscosity and thermal conduction also contribute to the total attenuation.

Whereas, attenuation due to relaxation is the most important case of attenuation in liquid tissues, the first justified wave equation considering several relaxation mechanisms, was published by NACHMAN *et al.* (1990). It relies on absorption dominated by relaxation. Until this, the studies have focused mostly on attenuation in the presence of relaxation, but did not develop the linear wave theory (all the more so the nonlinear wave theory). NACHMAN *et al.* (1990) made use of a set of classical dynamic equations supplemented by thermodynamic and constitutive relations for an electrolyte (LIEBERMANN, 1949; EIGEN, TAMM, 1962). This requires involving into consideration quantities specifying chemical relaxation and enlargement of number of

equations and variety of modes in a flow. NACHMAN *et al.* (1990) considered spatial inhomogeneities in material compressibility, density, and parameters of relaxation and pointed the difference of relaxation processes and wave features in liquids and gases. The different kind of relaxation in liquids provides especial features of nonlinear distortions of a waveform, nonlinear excitation of the non-wave modes (that is, the entropy and the vortex modes) by the intense wave perturbations (PERELOMOVA, 2015).

An exact solution to the dynamic nonlinear equation with frequency-dependent relaxation in the wide range of wave frequencies is still unresolved issue (HAMILTON *et al.*, 1998). As for dynamic equations which describe secondary nonlinear effects induced by sound in electrolytes, they have been derived in the limiting cases of high and low frequencies of an exciter. This concerns excitation of the relaxation mode and acoustic streaming and heating, i.e., nonlinear excitation of vorticity and entropy modes due to losses in momentum and energy of the intense sound (PERELOMOVA, 2015). In general, the nonlinear effects of sound in a relaxing fluid are poorly studied theoretically and rely only on a few papers (MOLEVICH, 2001; PERELOMOVA, PELC-GARSKA, 2010). This theoretical study considers the nonlinear effects of sound in liquid electrolytes with relaxation due to chemical reactions in all frequency regimes.

The mathematical concept of derivation of coupling nonlinear equations for interacting modes is explained and discussed in details by LEBLE and PERELOMOVA (2018). In essence, it is immediate projection of the initial equations onto dynamic equations governing different modes. The first step is to determine modes as specific types of motion in a fluid. That allows to establish projecting operators and to derive the leading-order coupling equations governing every mode in a weakly nonlinear flow. These equations take into account interaction of modes. As usual, some wave modes are considered as dominative and hence contribute predominantly to the “forces” in excitation of the secondary modes. The method was proposed, tested and applied by the author in analysis of nonlinear phenomena in fluids with the standard attenuation and different mechanisms of relaxation (LEBLE, PERELOMOVA, 2018). This method operates on instantaneous quantities, does not require averaging over the sound period at any stage and is valid for both periodic and aperiodic sound. It allows to distribute the nonlinear terms between dynamic equations for the individual modes correctly. This study investigates weakly nonlinear dynamics of the secondary entropy and relaxation modes in electrolyte with a chemical reaction in the total frequency range of the acoustic exciting perturbations. The peculiarity of this type of relaxation in comparison with relaxation in gases, is explained and discussed in the Concluding Remarks. Relaxation of this kind

brings also unusual nonlinear relations between acoustic perturbations.

2. Weakly nonlinear dynamics of a flow

2.1. Starting points and basic equations

We start from consideration of governing equations for a fluid flow with relaxation:

$$\frac{D\rho}{Dt} + \rho(\nabla \cdot \mathbf{v}) = 0, \quad (1)$$

(conservation of mass; ρ and \mathbf{v} are density and particle velocity, and $\frac{D}{Dt}$ designates the total time derivative operator),

$$\rho \frac{D\mathbf{v}}{Dt} = -\nabla p, \quad (2)$$

(conservation of momentum; p is the total pressure), and the energy balance equation. The Gibbs relation for the rate of variation of entropy s is as follows:

$$T \frac{Ds}{Dt} = \frac{Du}{Dt} - \frac{p}{\rho^2} \frac{D\rho}{Dt} + \sum_l A_l \frac{Dn_l}{Dt}, \quad (3)$$

where u is the internal energy per unit mass, A_l are affinities, and n designates the number of molecules of species l per unit mass (LIEBERMANN, 1949; EIGEN, TAMM, 1962). The entropy is a sum of an equilibrium part s^e and of an irreversible part (EIGEN, TAMM, 1962; NACHMAN *et al.*, 1990; PIERCE, 2021):

$$s(p, \rho, n) = s^e(p, \rho) + \frac{C_p}{\beta T} \sum_l \kappa_l \Delta \xi_l, \quad (4)$$

where

$$\Delta \xi_l = \frac{n - n_l^e(p, T)}{\partial n_l^e(p, T) / \partial p},$$

κ_l designates the contribution of the dissolved molecules of species l to the isothermal compressibility (reciprocal of the bulk modulus), and the appropriate thermodynamic coefficients (the heat capacity under constant pressure and the volumetric coefficient of thermal expansion) are determined as

$$C_p = T_0 \left(\frac{\partial s^e}{\partial T} \right)_p, \quad \beta = -\frac{1}{\rho_0} \left(\frac{\partial \rho}{\partial T} \right)_p,$$

where T_0 and ρ_0 denote temperature and density of a medium in equilibrium. Finally, the entropy balance equation takes the leading-order form as follows (NACHMAN *et al.*, 1990):

$$\frac{Dp}{Dt} - c^2 \frac{D\rho}{Dt} - \frac{B}{2\rho_0^2} \frac{D\rho^2}{Dt} + \rho c^2 \sum_l \kappa_l \frac{D\Delta \xi_l}{Dt} = \sum_l \frac{\kappa_l \beta c^2}{\tau_l C_p} \Delta \xi_l^2, \quad (5)$$

where c is the equilibrium speed of infinitely-small magnitude sound, τ_l is the relaxation time for the process in species l and B is determined by an equality

(RUDENKO, SOLUYAN, 1977; MAKAROV, OCHMANN, 1996):

$$\left(\frac{\partial^2 p}{\partial \rho^2}\right)_{se} = \frac{B}{\rho_0^2}.$$

Our primary objective is to derive dynamic equations valid at order M^2 , where the Mach number M is a ratio of magnitude of a fluid velocity to the sound speed. The Mach number measures degree of nonlinearity of a flow. A linear flow corresponds to the infinitely small Mach number. Assuming constant equilibrium quantities ρ_0 and p_0 , and a lack of the bulk flow and a planar geometry of a flow, we arrive to the system (PERELOMOVA, 2015):

$$\begin{aligned} \frac{\partial v}{\partial t} + \frac{1}{\rho_0} \frac{\partial p'}{\partial x} &= -v \frac{\partial v}{\partial x} + \frac{\rho'}{\rho_0^2} \frac{\partial p'}{\partial x}, \\ \frac{\partial p'}{\partial t} + \rho_0 c^2 \frac{\partial v}{\partial x} + \rho_0 c^2 \sum_l \kappa_l \frac{\partial}{\partial t} \Delta \xi_l &= -v \frac{\partial p'}{\partial x} \\ &\quad - \frac{c^2 \rho_0 + B}{\rho_0} \rho' \frac{\partial v}{\partial x} + \frac{\beta c^2}{C_p} \sum_l \frac{\kappa_l}{\tau_l} \Delta \xi_l^2 \\ &\quad - \rho' c^2 \sum_l \kappa_l \frac{\partial}{\partial t} \Delta \xi_l - \rho_0 c^2 v \frac{\partial}{\partial x} \sum_l \kappa_l \Delta \xi_l, \\ \frac{\partial \rho'}{\partial t} + \rho_0 \frac{\partial v}{\partial x} &= -\rho' \frac{\partial v}{\partial x} - v \frac{\partial \rho'}{\partial x}, \\ \frac{\partial \Delta \xi_l}{\partial t} + \frac{\Delta \xi_l}{\tau_l} + \frac{\partial p'}{\partial t} &= -v \frac{\partial p'}{\partial x} - v \frac{\partial}{\partial x} \Delta \xi_l. \end{aligned} \quad (6)$$

The linear version of this system was derived by NACHMAN *et al.* (1990).

2.2. Dispersion relations and modes of a linear flow

Studies of motions of infinitely-small magnitudes begin usually by representing all perturbations as a sum of planar waves:

$$f'(x, t) = \int \tilde{f}(k) \exp(i\omega(k)t - ikx) dk,$$

where k is the wave vector, $\tilde{f}(k) \exp(i\omega(k)t)$ denotes the Fourier transform of $f'(x, t)$, $\tilde{f}(k, t) = \frac{1}{2\pi} \int f(x, t) e^{ikx} dx$. The number of dispersion relations, i.e., number of modes, coincides with the number of initial equations in partial derivatives which include the first derivative with respect to time. Only one relaxation process is considered for simplicity. There are two sound modes ($i = 1$ and $i = 2$), one entropy mode ($i = 3$) and one relaxation mode ($i = 4$). The dispersion relations ω_i ($i = 1, \dots, 4$) take the leading-order forms:

$$\begin{aligned} \omega_{1,2} &= \pm ck + i \frac{c^4 k^2 \rho_0 \tau}{2(1 + ick\tau)} \kappa, & \omega_3 &= 0, \\ \omega_4 &= \frac{i}{\tau} \left(1 + \frac{\kappa c^2 \rho_0}{1 + c^2 k^2 \tau^2} \right). \end{aligned} \quad (7)$$

We consider weakly dispersive flows which are characterized by a small dimensionless parameter responsible for relaxation:

$$\alpha = \kappa c^2 \rho_0 \ll 1.$$

All formulae are derived with an accuracy up to α^1 . It is convenient to express all Fourier transforms in terms of the Fourier transform of excess density $\tilde{\rho}$ for the first three modes:

$$\begin{aligned} \tilde{v}_i &= \frac{\omega_i \tilde{\rho}_i}{k \rho_0}, \\ \tilde{p}_i &= \frac{\omega_i^2 \tilde{\rho}_i}{k^2}, \\ \tilde{\Delta \xi}_i &= -i \frac{\omega_i^3 \tilde{\rho}_i}{k^2 (i\omega_i + 1/\tau)}, \quad (i = 1, 2, 3), \end{aligned} \quad (8)$$

and to make use of $\tilde{\Delta \xi}_4$ as the reference variable for the relaxation mode:

$$\begin{aligned} \tilde{\rho}_4 &= \frac{k^2 (i - \tau \omega_4)}{\tau \omega_4^3} \tilde{\Delta \xi}_4, \\ \tilde{p}_4 &= \frac{i - \tau \omega_4}{\tau \omega_4} \tilde{\Delta \xi}_4, \\ \tilde{v}_4 &= \frac{k (i - \tau \omega_4)}{\tau \rho_0 \omega_4^2} \tilde{\Delta \xi}_4. \end{aligned} \quad (9)$$

In particular, the leading-order links for the Fourier transforms of perturbations in the first sound mode, the entropy mode and the relaxation mode are as follows:

$$\begin{aligned} \tilde{p}_1 &= c^2 \tilde{\rho}_1 + \frac{i\alpha c^3 k \tau}{1 + ick\tau} \tilde{\rho}_1, \\ \tilde{v}_1 &= \frac{c}{\rho_0} \left(1 + \frac{i\alpha ck\tau}{2(1 + ick\tau)} \right) \tilde{\rho}_1, \\ \tilde{\Delta \xi}_1 &= -\frac{ic^3 k \tau}{1 + ick\tau} \left(1 + \frac{\alpha(3ick\tau - 2c^2 k^2 \tau^2)}{2(1 + ick\tau)^2} \right) \tilde{\rho}_1, \\ \tilde{p}_3 &= 0, \quad \tilde{v}_3 = 0, \quad \tilde{\Delta \xi}_3 = 0, \\ \tilde{\rho}_4 &= \frac{\alpha k^2 \tau^2}{1 + c^2 k^2 \tau^2} \tilde{\Delta \xi}_4, \\ \tilde{p}_4 &= -\frac{\alpha}{1 + c^2 k^2 \tau^2} \tilde{\Delta \xi}_4, \\ \tilde{v}_4 &= \frac{i\alpha k \tau}{(1 + c^2 k^2 \tau^2) \rho_0} \tilde{\Delta \xi}_4. \end{aligned} \quad (10)$$

Equations (10) determine relations of perturbations in x space at any moment t unequivocally. The total small-magnitude perturbations and their Fourier transforms are the sums of all specific ones:

$$\begin{aligned} \tilde{v} &= \sum_{i=1}^4 \tilde{v}_i, & \tilde{p} &= \sum_{i=1}^4 \tilde{p}_i, \\ \tilde{\rho} &= \sum_{i=1}^4 \tilde{\rho}_i, & \tilde{\Delta\xi} &= \sum_{i=1}^4 \tilde{\Delta\xi}_i. \end{aligned} \tag{11}$$

An equation governing excess density in an acoustic planar wave which propagates in the positive direction of the axis OX , $\rho_1(x, t)$, may be reconstructed by the use of ω_1 (Eq. (7)):

$$\frac{\partial \rho_1}{\partial t} + c \frac{\partial \rho_1}{\partial x} - \frac{\alpha c}{2} \int_x^\infty \exp\left(\frac{x-x'}{c\tau}\right) \frac{\partial^2 \rho_1}{\partial x'^2} dx' = 0. \tag{12}$$

The linear equation specifying perturbations in the relaxation mode, takes the form

$$\frac{\partial \Delta\xi_4}{\partial t} + \frac{\Delta\xi_4}{\tau} + \frac{\alpha}{2c\tau^2} \int_{-\infty}^\infty \exp\left(-\frac{|x-x'|}{c\tau}\right) \Delta\xi_4(x', t) dx' = 0. \tag{13}$$

In derivation of Eqs. (12) and (13), we make use of conformity of operators in k and x spaces:

$$\begin{aligned} -ik &\Leftrightarrow \frac{\partial}{\partial x}, \\ \frac{1}{1+ick\tau} &\Leftrightarrow \frac{1}{c\tau} \int_x^\infty dx' \exp\left(\frac{x-x'}{c\tau}\right), \\ \frac{1}{1-ick\tau} &\Leftrightarrow \frac{1}{c\tau} \int_{-\infty}^x dx' \exp\left(-\frac{x-x'}{c\tau}\right), \\ \frac{1}{1+c^2k^2\tau^2} &\Leftrightarrow \frac{1}{2c\tau} \int_{-\infty}^\infty dx' \exp\left(-\frac{|x-x'|}{c\tau}\right). \end{aligned} \tag{14}$$

The approximate solutions to an equation similar to Eq. (12) are discussed by HAMILTON *et al.* (1998). Apart from the dispersion relations and links specifying every mode, we are able to evaluate operators which distinguish the specific perturbation from the vector of total disturbances in a linear flow. Solving algebraic equations:

$$\begin{aligned} \tilde{P}_1(\tilde{v} \ \tilde{p} \ \tilde{\rho} \ \tilde{\Delta\xi})^T &= \tilde{\rho}_1, \\ \tilde{P}_3(\tilde{v} \ \tilde{p} \ \tilde{\rho} \ \tilde{\Delta\xi})^T &= \tilde{\rho}_3, \\ \tilde{P}_4(\tilde{v} \ \tilde{p} \ \tilde{\rho} \ \tilde{\Delta\xi})^T &= \tilde{\Delta\xi}_4, \end{aligned}$$

one arrives at operator rows which distinguish every specific Fourier transform of excess quantities for these three modes:

$$\begin{aligned} \tilde{P}_1 &= \begin{pmatrix} \frac{\rho_0}{2c} + \frac{\alpha ck^2\tau^2\rho_0}{4(1+ick\tau)^2} \\ \frac{1}{2c^2} - \frac{i\alpha k\tau(1+2ick\tau)}{4c(1+ick\tau)^2} \\ 0 \\ \frac{\alpha}{2c^2(1+ick)} \end{pmatrix}^T, \\ \tilde{P}_3 &= \begin{pmatrix} 0 \\ -\frac{1}{c^2} \\ 1 \\ -\frac{\alpha}{c^2} \end{pmatrix}^T, \\ \tilde{P}_4 &= \begin{pmatrix} \frac{ic^2k\rho_0\tau}{1+c^2k^2\tau^2} + \frac{4i\alpha c^4k^3\rho_0\tau^3}{(1+c^2k^2\tau^2)^3} \\ \frac{c^2k^2\tau^2}{1+c^2k^2\tau^2} + \frac{\alpha c^2k^2\tau^2(3c^2k^2\tau^2-1)}{(1+c^2k^2\tau^2)^3} \\ 0 \\ 1 + \frac{2\alpha c^2k^2\tau^2}{(1+c^2k^2\tau^2)^2} \end{pmatrix}^T. \end{aligned} \tag{15}$$

They are evaluated with accuracy up to terms proportional to α^0 and α^1 but without any restrictions concerning spatial spectrum of perturbations. The elements of \tilde{P}_3 and \tilde{P}_4 determine some integro-differential spatial operators P_3 and P_4 in the x space accordingly to Eq. (14). When P_3, P_4 apply at the system of conservation Eq. (6), they distinguish equations for the excess specific densities of the entropy and relaxation modes, respectively.

2.3. Nonlinear dynamics of sound

Equation (12) may be expanded by inclusion of the nonlinear term. For this purpose, it is sufficient to apply P_1 on the system Eq. (6) with the leading-order result:

$$\begin{aligned} \frac{\partial \rho_1}{\partial t} + c \frac{\partial \rho_1}{\partial x} - \frac{\alpha c}{2} \int_x^\infty \exp\left(\frac{x-x'}{c\tau}\right) \frac{\partial^2 \rho_1}{\partial x'^2} dx' \\ + \frac{\varepsilon c}{\rho_0} \rho_1 \frac{\partial \rho_1}{\partial x} = 0. \end{aligned} \tag{16}$$

The nonlinear distortion of a wave is conditioned by the positive parameter of nonlinearity ε :

$$\varepsilon = \frac{B}{2c^2\rho_0} + 1.$$

Links (Eq. (10)) specifying acoustic perturbations, may be corrected by involving nonlinear terms. As for links connecting p_1, v_1 and ρ_1 , they are as follows:

$$\begin{aligned}
p_1 &= c^2 \rho_1 - c^2 \alpha \int_x^\infty \exp\left(\frac{x-x'}{c\tau}\right) \frac{\partial \rho_1}{\partial x'} dx' + \frac{B}{2\rho_0^2} \rho_1^2, \\
v_1 &= \frac{c}{\rho_0} \rho_1 + c^2 \int_x^\infty \exp\left(\frac{x-x'}{c\tau}\right) \frac{\partial \rho_1}{\partial x'} dx' + \frac{B-2c^2\rho_0}{4c\rho_0^3} \rho_1^2.
\end{aligned} \tag{17}$$

The nonlinear correction to ξ_1 may be readily derived in two limiting cases:

$$\text{low-frequency: } \Delta\xi_1 = c^3\tau \frac{\partial \rho_1}{\partial x} + \frac{Bc\tau}{2\rho_0^2} \rho_1 \frac{\partial \rho_1}{\partial x}, \tag{18}$$

$$\text{high-frequency: } \Delta\xi_1 = -c^2 \rho_1. \tag{19}$$

Hence, nonlinearity introduces nothing in “high-frequency” links (if $c\tau \left|\frac{\partial \rho_1}{\partial x}\right| \gg |\rho_1|$), but contributes in the “low-frequency” ones (if $c\tau \left|\frac{\partial \rho_1}{\partial x}\right| \ll |\rho_1|$). The nonlinear correction is unusual. The links which determine v_1 and p_1 are proportional to ρ_1^2 in the Riemann wave.

3. Acoustic heating

We consider the dominant first acoustic mode as an exciter of secondary perturbations. The perturbations specifying all other modes are considered as negligibly small at the beginning of evolution. The leading-order linear links for the perturbations in this mode take the forms (Eqs. (10) and (14)):

$$\begin{aligned}
p_1 &= c^2 \rho_1 - \alpha c^2 \int_x^\infty \exp\left(\frac{x-x'}{c\tau}\right) \frac{\partial \rho_1}{\partial x'} dx', \\
v_1 &= \frac{c}{\rho_0} \rho_1 - \frac{\alpha c}{2\rho_0} \int_x^\infty \exp\left(\frac{x-x'}{c\tau}\right) \frac{\partial \rho_1}{\partial x'} dx',
\end{aligned} \tag{20}$$

$$\Delta\xi_1 = c^2 \int_x^\infty \exp\left(\frac{x-x'}{c\tau}\right) \frac{\partial \rho_1}{\partial x'} dx'.$$

They make possible to express an acoustic source in terms of perturbation of density attributable to the acoustic mode. Application of P_3 (Eq. (15)) at the system Eq. (6) reduces all acoustic and relaxation terms in the left-hand linear side and yields the leading-order equation which governs acoustic heating:

$$\begin{aligned}
\frac{\partial \rho_3}{\partial t} &= -\frac{\alpha}{2cC_p\rho_0^2\tau} \left(2\beta c^3 \rho_0 \left(\int_x^\infty e^{\frac{x-x'}{c\tau}} \frac{\partial \rho_1(x',t)}{\partial x'} dx' \right)^2 \right. \\
&+ 2cC_p\rho_0\rho_1 \int_x^\infty e^{\frac{x-x'}{c\tau}} \frac{\partial \rho_1(x',t)}{\partial x'} dx' \\
&+ C_p\tau\rho_1 \left(\frac{B}{c\tau} \int_x^\infty e^{\frac{x-x'}{c\tau}} \frac{\partial \rho_1(x',t)}{\partial x'} dx' \right. \\
&\left. \left. - (B+2c^2\rho_0) \frac{\partial \rho_1(x',t)}{\partial x} \right) \right) = Q_{a,e}.
\end{aligned} \tag{21}$$

The leading-order low- and high- frequency limits of $Q_{a,e}$ have been preliminarily considered by PERELOMOVA (2015). They are the limiting cases of Eq. (21) if $c\tau \left|\frac{\partial \rho_1}{\partial x}\right| \ll |\rho_1|$, $c\tau \left|\frac{\partial \rho_1}{\partial x}\right| \gg |\rho_1|$, respectively:

low-frequency:

$$Q_{a,e} = -\frac{\alpha\tau}{2C_p\rho_0^2} \left(2\beta c^4 \rho_0 \left(\frac{\partial \rho_1}{\partial x} \right)^2 + BC_p\rho_1 \frac{\partial^2 \rho_1}{\partial x^2} \right),$$

high-frequency:

$$\begin{aligned}
Q_{a,e} &= -\frac{\alpha}{2c^2\tau^3 C_p\rho_0^2} \left(2\beta c^2 \rho_0 \left(\int \rho_1(x,t) dx \right)^2 \right. \\
&+ 2c(2\beta c^2 - C_p)\rho_0\tau \left(\int \rho_1(x,t) dx \right) \rho_1 \\
&- \tau^2 \rho_1 \left((BC_p - 2\beta c^4 \rho_0 + 2c^2 C_p \rho_0) \rho_1 \right. \\
&\left. \left. + cC_p(B + 2c^2 \rho_0)\tau \frac{\partial \rho_1}{\partial x} \right) \right).
\end{aligned}$$

Equation (21) may be solved by the integration of $Q_{a,e}$ over time.

3.1. Nearly harmonic exciter

The most interesting case is the median domain of frequencies. Let the exciting perturbation in density takes the form

$$\begin{aligned}
\rho_1 &= M\rho_0 \exp\left(-\frac{\alpha c^2 k^2 \tau}{1+c^2 k^2 \tau^2} t\right) \sin \\
&\cdot \left(kx - \frac{ck(1+(1+0.5\alpha)c^2 k^2 \tau^2)}{1+c^2 k^2 \tau^2} t \right),
\end{aligned} \tag{22}$$

where M is the Mach number, and $\frac{ck(1+(1+0.5\alpha)c^2 k^2 \tau^2)}{1+c^2 k^2 \tau^2}$ is the quasi-frequency of an exciter. The amplitude of excess density decreases with time. The leading-order acoustic source averaged over period equals

$$\begin{aligned}
\langle Q_{a,e} \rangle &= \exp\left(-\frac{\alpha c^2 k^2 \tau}{1+c^2 k^2 \tau^2} t\right) \\
&\cdot \frac{\alpha M^2 k^2 \tau (BC_p + 2c^2(C_p - \beta c^2)\rho_0)}{4C_p(1+c^2 k^2 \tau^2)}.
\end{aligned} \tag{23}$$

High-frequency limit $ck\tau \gg 1$ results in

$$\langle Q_{a,e} \rangle = \exp\left(-\frac{\alpha t}{\tau}\right) \frac{\alpha M^2 (BC_p - 2c^2 C_p \rho_0 - 2c^4 \beta \rho_0)}{4c^2 C_p \tau}. \tag{24}$$

This leads to uniform generation of perturbation of density which starts at $t=0$ ($\rho_3(0)=0$):

$$\rho_3 = \left(1 - \exp\left(-\frac{\alpha t}{\tau}\right)\right) \frac{M^2 (BC_p - 2c^2 C_p \rho_0 - 2c^4 \beta \rho_0)}{4c^2 C_p}. \tag{25}$$

The maximum absolute value of excess density which may be achieved in the course of acoustic heating, does not depend on α and τ (these parameters determine the rate of exciting of the entropy mode).

4. Excitation of the relaxation mode. Case of harmonic exciter

Recalling relations (Eq. (10)) and applying P_4 at Eq. (6), one arrives at the leading-order dynamic equation for the perturbations in the relaxation mode:

$$\frac{\partial \Delta \xi_4}{\partial t} + \frac{1}{\tau} \Delta \xi_4 + \frac{\alpha}{c\tau^2} \int_{-\infty}^{\infty} \exp\left(-\frac{|x-x'|}{c\tau}\right) \cdot \Delta \xi_4(x', t) dx' = Q_{a,r}. \quad (26)$$

In view of the complex form of operators in P_4 corresponding to \tilde{P}_4 (Eq. (15)), we use the solution of the equation without account of dispersion and non-linearity,

$$\rho_1 = M\rho_0 \sin(\Omega(t - x/c)). \quad (27)$$

The averaged form of $Q_{a,r}$ in the case of the harmonic exciter is as follows:

$$\begin{aligned} \langle Q_{a,r} \rangle &= \frac{\alpha \Omega^2 M^2}{64 C_p \tau (1 + \Omega^2 \tau^2)^3} (8 B C_p (1 + \Omega^2 \tau^2)^2 \\ &+ c^2 \rho_0 (C_p (16 + 93 \Omega^2 \tau^2 + 46 \Omega^4 \tau^4 + 9 \Omega^6 \tau^6)) \\ &- 18 \beta (1 + \Omega^2 \tau^2)^2 c^2). \end{aligned}$$

The limiting forms of coupling equations are

low-frequency:

$$\frac{\partial \Delta \xi_4}{\partial t} + \frac{1 + \alpha}{\tau} \Delta \xi_4 = \frac{\alpha M^2 \Omega^2 \tau}{8 C_p} \cdot (B C_p + 2c^2 (C_p \rho_0 - \beta c^2) \rho_0), \quad (28)$$

high-frequency:

$$\frac{\partial \Delta \xi_4}{\partial t} + \frac{1}{\tau} \Delta \xi_4 = \frac{9}{64} \alpha M^2 \rho_0 c^2 \Omega^2 \tau. \quad (29)$$

The efficiency of low-frequency harmonic excitation is fairly low. As for the high-frequency excitation, the absolute value of an acoustic force does not tend to any finite limit but enlarges with frequency of an exciter as Ω^2 in contrast to an acoustic source of the entropy mode. Equations (28) and (29) may be readily solved by direct integration over time making the use of the substitutions $\Delta \xi_4 = f(x, t) \exp(-t(1 + \alpha)/\tau)$ and $\Delta \xi_4 = f(x, t) \exp(-t/\tau)$, respectively. There is non-zero variation in temperature which associates with the relaxation mode:

$$T_4 = \frac{\gamma}{\beta} p_4 - \frac{1}{\rho_0 \beta} \rho_4,$$

where

$$\gamma = \frac{1}{\rho_0} \left(\frac{\partial \rho}{\partial p} \right)_T$$

designates the isothermal compressibility. These variations are of the order $\alpha^2 M^2$ in all ranges of frequencies. They are much smaller than in the entropy mode (the latter are of the order αM^2).

5. Concluding remarks

In this study, we consider weakly nonlinear effects of sound in an electrolyte. The total range of sound frequencies is considered. Nonlinearity of a flow and relaxation of thermodynamic processes are necessary conditions for interaction of modes and excitation of the secondary entropy and relaxation modes in the field of intense sound. The instantaneous equations which govern excitation of the relaxation and entropy modes in the field of intense sound wave, are derived (Eqs. (21), (26)). Thermodynamic parameters of the nonlinear phenomena slowly vary with time since they do not represent the wave processes. They are readily measured and may be useful in indication of relaxation processes in liquids. Theoretical conclusions may be employed in modeling of nonlinear scattering in applications such as image reconstruction in a liquid. The theory is of interest from the standpoint of understanding wave processes in the sea and in biological systems, and as a tool in studies of colloidal systems. In addition, studies of sound absorption in various fluids such as distilled water and electrolytic solutions, are of crucial importance in the understanding of their molecular structure. The nonlinear phenomena indicate the kind of relaxation in a fluid, its equilibrium thermodynamic properties and frequency of exciting wave perturbations. The mechanic viscosity and heat conduction are not considered in this study. In gases, these effects are well-studied (RUDENKO, SOLUYAN, 1977; MAKAROV, OCHMANN, 1996), and they are small compared to relaxation effects in liquids and biological tissues (MANDELSHTAM, LEONTOWICH, 1937; HERTZFELD, LITOWITZ, 1959; NYBORG, 1978).

There is an apparent difference of the considered case of relaxation in liquids and relaxation due to other processes of deviation from the thermodynamic equilibrium in gases, such as excitation of vibrational degrees of a molecule's freedom or exothermic chemical reactions. That concerns dispersion relations specifying wave and non-wave modes and, as a result, linear dynamic equations describing perturbations in these modes. The links of specific perturbations in wave and non-wave mode are also different. This has been indicated by PERELOMOVA (2015). In particular, the dispersion relation for the relaxation mode in a gas flow with vibrational relaxation looks different as compared with $\omega_{4,V}$ from Eq. (7) (τ_V is the characteristic time of vibrational relaxation, and

T_0 designates the equilibrium temperature of a gas) (PERELOMOVA, 2010):

$$\omega_{4,V} = \frac{i}{\tau_V} + \frac{i(\gamma - 1)(\gamma + c^2 k^2 \tau_V^2) T_0 C_v}{c^2 \tau_V (1 + c^2 k^2 \tau_V^2)}.$$

It describes different behavior especially at intermediate characteristic frequencies of exciting perturbations. The relaxation mode is isobaric in the case of relaxation processes in gases. An excess pressure in the relaxation mode in electrolytes is not zero and relates to excess density in the following manner

$$\rho_4 = \tau^2 \frac{\partial^2 p_4}{\partial x^2}.$$

Also, the dispersion relations for the high-frequency sound in a liquid which readily follow from Eq. (7):

$$\omega_{1,2} = \pm (ck + 0.5\kappa c^3 k \rho_0),$$

fundamentally differ from that in a gas with vibrational relaxation (OSIPOV, UVAROV, 1992; MOLEVICH, 2001):

$$\omega_{1,2} = \pm ck + 0.5i(\gamma - 1)^2 C_v T_0 / (\tau_V c^2).$$

There is an attenuation independent of frequency in the case of vibrational relaxation, while there is only pure dispersion without attenuation in the case of relaxation in electrolytes. The Newtonian attenuation is proportional to k^2 . The links between acoustic pressure and excess density also look different. In particular:

$$p_1 = c^2 \rho_1 - \alpha c^2 \int_x^\infty e^{\frac{x-x'}{c\tau}} \frac{\partial \rho_1(x', t)}{\partial x'} dx'$$

in the case of relaxation in electrolytes, and

$$p_1 = c^2 \rho_1 - \frac{\theta c}{\tau} \int_x^\infty e^{\frac{x-x'}{c\tau}} \rho_1(x', t) dx'$$

in the case of vibrational relaxation in gases, where θ is a dimensionless parameter reflecting vibrational relaxation in gases in view of inflow of energy (may be negative) (PERELOMOVA, 2019). The similar link specifies a gas flow with a chemical reaction (LEBLE, PERELOMOVA, 2018). The different behavior conditions different hysteresis acoustic curves (PERELOMOVA, 2013). The nonlinear effects of sound also look different in gases and liquids. The leading-order acoustic force is proportional to the squared Mach number M^2 and α , and is frequency-dependent. The high-frequency sound is effective in excitation of the secondary perturbations in presence of all kinds of relaxation, but the low-frequency exciters are not effective (PERELOMOVA 2010; 2019).

The nonlinear links of specific acoustic variables attract special attention. We face with the unusual

relations (Eq. (18)) for low-frequency sound perturbations with a nonlinear term proportional to $\rho_1 \frac{\partial \rho_1}{\partial x}$, not ρ_1^2 which is specific for the flows over uniform fluids (RUDENKO, SPYAN, 1977). Hence, nonlinear links reveal the frequency-dependent character and may indicate sound frequency, characteristic time of relaxation and equilibrium parameters of an electrolyte.

References

1. DUCK F.A., BAKER A.C., STARRITT H.C. (1998), *Ultrasound in Medicine*, Institute of Physics Publishing, Bristol, Philadelphia.
2. EIGEN M., TAMM K. (1962), Sound absorption in electrolyte solutions as a sequence of chemical reactions, *Zeitschrift fuer Elektrochemie*, **66**(2): 93–121.
3. EIGEN M., DE MAYER L. (1963), Relaxation methods, [in:] *Techniques of Organic Chemistry*, Freiss S.L., Lewis E.S., Weissberger A. [Eds.] Interscience Publishers, New York.
4. HAMILTON M., IL'INSKII Yu., ZABOLOTSKAYA E. (1998), Dispersion, [in:] *Nonlinear Acoustics*, Hamilton M., Blackstock D. [Eds.], Academic Press.
5. HERTZFELD K.F., LITOWITZ T.A. (1959), *Absorption and Dispersion of Ultrasonic Waves*, Academic Press, New York.
6. LEBLE S., PERELOMOVA A. (2018), *The Dynamical Projectors Method: Hydro and Electrodynamics*, CRC Press.
7. LIEBERMANN L.N. (1948), The origin of sound absorption in water and in sea water, *The Journal of the Acoustical Society of America*, **20**(6): 868–873, doi: [10.1121/1.1906450](https://doi.org/10.1121/1.1906450).
8. LIEBERMANN L.N. (1949), Sound propagation in chemically active media, *Physical Review*, **76**(10): 1520, doi: [10.1103/PhysRev.76.1520](https://doi.org/10.1103/PhysRev.76.1520).
9. MAKAROV S., OCHMANN M. (1996), Nonlinear and thermoviscous phenomena in acoustics, Part I, *Acustica*, **82**(4): 579–606.
10. MANDELSHTAM L.I., LEONTOWICH M.A. (1937), To the theory of sound absorption in liquids, *Zhurnal Éksperimental'noi i Teoreticheskoi Fiziki*, **7**(3): 438.
11. MELLEN R.H., SIMMONS V.P., BROWNING D.G. (1979), Sound absorption in sea water: A third chemical relaxation, *The Journal of the Acoustical Society of America*, **65**(4): 923–925, doi: [10.1121/1.382595](https://doi.org/10.1121/1.382595).
12. MOLEVICH N.E. (2001), Amplification of vortex and temperature waves in the process of induced scattering of sound in thermodynamically nonequilibrium media, *High Temperature*, **39**(6): 884–888, doi: [10.1023/A:1013147207446](https://doi.org/10.1023/A:1013147207446).
13. NACHMAN A., SMITH J.F., WAAG R.C. (1990), An equation for acoustic propagation in inhomogeneous media with relaxation losses, *The Journal of the Acoustical Society of America*, **88**(3): 1584–1595, doi: [10.1121/1.400317](https://doi.org/10.1121/1.400317).

14. NYBORG W.L. (1978), *Physical Mechanisms for Biological Effects of Ultrasound*, The Bureau of Radiological Health, Rockville.
15. OSIPOV A.I., UVAROV A.V. (1992), Kinetic and gasdynamic processes in nonequilibrium molecular physics, *Soviet Physics Uspekhi*, **35**(11): 903, doi: [10.1070/PU1992v035n11ABEH002275](https://doi.org/10.1070/PU1992v035n11ABEH002275).
16. PARKER K.J. (1983), Ultrasonic attenuation and absorption in liver tissue, *Ultrasound in Medicine & Biology*, **9**(4): 363–369, doi: [10.1016/0301-5629\(83\)90089-3](https://doi.org/10.1016/0301-5629(83)90089-3).
17. PERELOMOVA A. (2010), Nonlinear generation of non-acoustic modes by low-frequency sound in a vibrationally relaxing gas, *Canadian Journal of Physics*, **88**(4): 293–300, doi: [10.1139/P10-011](https://doi.org/10.1139/P10-011).
18. PERELOMOVA A. (2013), Hysteresis curves and loops for harmonic and impulse perturbations in some non-equilibrium gases, *Central European Journal of Physics*, **11**(11): 1541–1547, doi: [10.2478/s11534-013-0305-2](https://doi.org/10.2478/s11534-013-0305-2).
19. PERELOMOVA A. (2015), The nonlinear effects of sound in a liquid with relaxation losses, *Canadian Journal of Physics*, **93**(11): 1391–1396, doi: [10.1139/cjp-2014-0676](https://doi.org/10.1139/cjp-2014-0676).
20. PERELOMOVA A. (2019), Excitation of non-wave modes by sound of arbitrary frequency in a chemically reacting gas, *Acta Acustica united with Acustica*, **105**(6): 918–927, doi: [10.3813/AAA.919373](https://doi.org/10.3813/AAA.919373).
21. PERELOMOVA A., PELC-GARSKA W. (2010), Efficiency of acoustic heating produced in the thermoviscous flow of a fluid with relaxation, *Central European Journal of Physics*, **8**(6): 855–863, doi: [10.2478/s11534-010-1015-y](https://doi.org/10.2478/s11534-010-1015-y).
22. PIERCE A.D. (1981), *Acoustics: An Introduction to its Physical Principles and Applications*, McGraw-Hill, New York.
23. PIERCE A.D., MAST T.D. (2021), Acoustic propagation in a medium with spatially distributed relaxation processes and a possible explanation of a frequency power law attenuation, *Journal of Theoretical and Computational Acoustics*, **29**(2): 2150012, doi: [10.1142/S2591728521500122](https://doi.org/10.1142/S2591728521500122).
24. RUDENKO O.V., SOLUYAN S.I. (1977), *Theoretical Foundations of Nonlinear Acoustics*, Plenum, New York.
25. YEAGER E., FISHER F.H. (1973), Origin of the low-frequency sound absorption in sea water, *The Journal of the Acoustical Society of America*, **53**(6): 1705–1707, doi: [10.1121/1.1913523](https://doi.org/10.1121/1.1913523).

FROM THE EDITORIAL BOARD

The Editorial Committee wishes to express many thanks and appreciation to all the reviewers, listed below, for their valuable opinions concerning papers submitted to the Archives of Acoustics in 2024:

Szymon	BANASZAK	Witold	MICKIEWICZ
Agnieszka	BETKOWSKA CAVALCANTE	Witold	MIKULSKI
Stefan	BRACHMAŃSKI	Akhilesh	MIMANI
Rafał	BURDZIK	Andrzej	MIŚKIEWICZ
Damian	CETNAROWICZ	Barłomiej	MRÓZ
Aleksandra	CHOJAK	Mats	NILSSON
Piotr	CUPIAŁ	Elżbieta	NOWICKA
Szymon	CYGAN	Artur	NOWOŚWIAT
Ireneusz	CZAJKA	Lu-Ean	OOI
Marcin	DĄBROWSKI	Krzysztof J.	OPIELIŃSKI
Paul	DEVOS	Janusz Piotr	PIECHOWICZ
Antonino	DI BELLA	Przemysław	PLASKOTA
Szymon	DRGAS	Dariusz	PLEBAN
Paweł	DZIECHCIŃSKI	Diana Ioana	POPESCU
Jan	FELCYN	Prasant Kumar	PRADHAN
Adrian	FOLTYN	Anna	PREIS
Fatma Nafaa	GAAFER	Tadeusz	PUSTELNY
Barbara	GAMBIN	Zbigniew	RANACHOWSKI
Luis	GOMEZ-AGUSTINA	Wojciech	RDZANEK
Łukasz	GORAZD	Jens Holger	RINDEL
Dorota	GÓRNICKA	Krzysztof	RUDNO-RUDZIŃSKI
Marcin Andrzej	GROCHOWINA	Kamil	STAWIARSKI
Andrzej	GRZĄDZIELA	Bartłomiej	STĘPIEŃ
Tadeusz	GUDRA	Shaul Hameed	SYED
Honorata Zofia	HAFKE-DYS	Krzysztof	SZCZUROWSKI
Dan	HE	Krzysztof	SZKLANNY
Adam	JABLONSKI	Malina	SZYCHOWSKA
Bartosz	JAKUBEK	Ayşe	TIRYAKIOĞLU
Wiesław	JAKUBIK	Jesus Alejandro	TORRES
Maciej	KŁACZYŃSKI	Piero	TORTOLI
Jedrzej	KOCIŃSKI	Jan	WARCZEK
Iwona	KOMORSKA	Stefan	WEYNA
Tomasz	KORBIEL	Jacek	WIERZBICKI
Krzysztof	KOSAŁA	Grzegorz	WOJNAR
Henryk	LASOTA	Piotr	WRZECIONO
Vincent	LAUDE	Tadeusz	WSZOLEK
Bogumił	LINDE	Wiesław	WSZOLEK
Karol	LISTEWNİK	Adam	ZAGUBIEŃ
Ewa	ŁUKASIK	Maciej	ZAWISZA
Sylwia	MAKOMASKA	Marcin	ZIELIŃSKI
Mirosław	MEISSNER	Sławomir Krzysztof	ZIELIŃSKI
Mykhaylo	MELNYK	Mariusz	ZIÓŁKO
Andrzej	MICHALSKI	Jakub	ŻMIGRODZKI