ARCHIVES
of
ACOUSTICS

QUARTERLY

WARSAW

ARCHIVES of ACOUSTICS

In Memoriam

Research Papers

Technical Note

In Memoriam

Professor Andrzej Stepnowski

1940 – 2024

Professor Andrzej Stepnowski was born in Biała Podlaska on 4 January 1940, the son of a lawyer couple who settled in Gdańsk after the war. Between 1958 and 1964, he studied at the then Faculty of Communications of Gdańsk University of Technology, and was a prominent member of the board of the Polish Students' Association. In 1964, he earned his master's degree in electronics with a focus on radiolocation, and became an assistant at the Department of Radionavigation, where he conducted hydroacoustics research. This research resulted in the development of a world-class computer-based system for fish stock estimation, which was installed aboard the RV *Profesor Siedlecki* as part of the FAO research programme. Professor Andrzej Stepnowski was awarded the title of Master of Polish Technology for his work. After passing his doctoral examination with distinction in 1974, he went on to obtain his habilitation degree (postdoctoral degree) in 1992. His research in both cases was related to the hydroacoustics of fishing. In 2004, he was awarded the title of professor.

Between 1983 and 1985, he was the head of the Department of Hydroacoustics, and from 1994 to 1999, he chaired the Department of Environmental Acoustics, and then became the head of the Department of Telemonitoring Systems and Department of Geoinformatics Systems – he held these positions until 2011.

From 1996 to 1999, he was the Vice-Dean for Research at the Faculty of Electronics, Telecommunications and Informatics, and served as the Vice-President for research of the Gdańsk University of Technology between 2002 and 2005.

Professor Stepnowski was a world-renowned expert on underwater acoustics, and participated in research internship programmes at the Massachusetts Institute of Technology in the US and the IRPeM Research Institute of the National Research Council of Italy in Ancona. He was also a visiting professor at four universities around the world – in the US, Canada, Turkey and Indonesia – he also served as a UN expert on food security and agriculture, editor for the *Acta Acustica* journal and vice president for Europe of the IEEE National Security Committee.

In the 2000s, Professor Stepnowski's research interests also included geoinformatics, an important aspect of which is the collection of data using hydroacoustic methods. He was the founder and later also the chair of the supervisory board of C-Map Poland, the country's pioneering geoinformatics business, which became Jeppesen Poland, part of the Boeing corporation, in 2007.

Professor Stepnowski's scholarly achievements include two monographs, more than 250 research papers, eight patents and more than 30 successful research and

implementation projects. He was the advisor for seven successful doctoral students, three of whom would go on to earn habilitation degrees (postdoctoral degrees). The mobile satellite urban navigation system for the visually impaired, Voice Maps, developed under his supervision, was awarded a gold medal at the Technological Innovation Fair in Brussels in 2014.

Professor Stepnowski was also a prominent contributor to several scientific societies. He was a member of the Acoustics Committee of the Polish Academy of Sciences and a member and secretary general of the Polish Acoustics Society, as well as organiser of the Hydroacoustics Symposia and various international acoustics conferences.

He was awarded a Knight's Cross and an Officer's Cross of the Order of Polonia Restituta, a Medal of the Commission of National Education and a Medal of Merit for the Gdańsk University of Technology.

Professor Roman Salamon

Research Paper

Speech Emotion Recognition Using a Multi-Time-Scale Approach to Feature Aggregation and an Ensemble of SVM Classifiers

Antonina STEFANOWSKA, Sławomir K. ZIELIŃSKI*

Faculty of Computer Science, Białystok University of Technology
Białystok, Poland

*Corresponding Author e-mail: s.zielinski@pb.edu.pl

Due to its relevant real-life applications, the recognition of emotions from speech signals constitutes a popular research topic. In the traditional methods applied for speech emotion recognition, audio features are typically aggregated using a fixed-duration time window, potentially discarding information conveyed by speech at various signal durations. By contrast, in the proposed method, audio features are aggregated simultaneously using time windows of different lengths (a multi-time-scale approach), hence, potentially better utilizing information carried at phonemic, syllabic, and prosodic levels compared to the traditional approach. A genetic algorithm is employed to optimize the feature extraction procedure. The features aggregated at different time windows are subsequently classified by an ensemble of support vector machine (SVM) classifiers. To enhance the generalization property of the method, a data augmentation technique based on pitch shifting and time stretching is applied. According to the obtained results, the developed method outperforms the traditional one for the selected datasets, demonstrating the benefits of using a multi-time-scale approach to feature aggregation.

Keywords: speech emotion recognition; feature aggregation; ensemble classification.

1. Introduction

Since the publication of Picard's seminal report in 1995 (PICARD, 1995), "affective computing", which concerns the identification, modelling, and reacting to human emotions by machines, has played an increasingly important role in the development of artificial intelligence algorithms. A growing interest of researchers in the area of affective computing is driven by the demands for emotion-aware applications. For example, the algorithms processing human emotions could be applied in health and safety systems, call centers, marketing recommenders, and forensic software. While human emotions could be recognized using a variety of methods, including facial recognition (JAIN *et al.*, 2019), analysis of body movements (ZACHARATOS *et al.*, 2021), or through the exploration of physiological data (YANG *et al.*, 2023), the scope of this paper is limited to the identification of emotions based solely on speech signals.

The methods applied to computational emotions recognition can be divided into the following two groups. The first one consists of the algorithms using the audio feature extractors combined with the classical machine learning algorithms. The second one is based on modern deep learning algorithms such as the convolutional neural networks. The performance of the speech emotion recognition methods has recently greatly improved, primarily due to the incorporation of the aforementioned deep learning techniques (KHALIL *et al.*, 2019; PANDEY *et al.*, 2019). The main advantage of deep learning techniques, over the traditional methods, is that they normally do not require any feature extraction procedure, typically engineered manually by domain experts. The speech signals are either fed to the inputs of the deep learning algorithms directly (TZIRAKIS *et al.*, 2018) or indirectly through some form of intermediate transformations, most notably spectrograms (ESKIMEZ *et al.*, 2018; CHOI *et al.*, 2018; ZHAO *et al.*, 2018; 2019; ZHANG *et al.*, 2020; GUIZZO *et al.*, 2020; TANG *et al.*, 2021). Nevertheless, the methods based on the deep learning approach still exhibit some limitations. For example, they require relatively large data sets for training. Moreover, they may

suffer from over-learning during memorization of layer-wise information (Khalil *et al.*, 2019). Furthermore, due to their relatively high computational complexity, the optimization of the deep learning algorithms typically consumes more electric energy compared to the traditional techniques (the aspect of energy-efficiency of machine learning algorithms is often overlooked in the scientific literature (García-Martín *et al.*, 2019)). Hence, the traditional algorithms should not yet be considered as obsolete.

In this paper, we present an improved version of the traditional method applied to speech emotion recognition. In traditional speech emotion recognition algorithms, input signals are analyzed using a time window of a constant duration (Omman, Eldho, 2022; Seknedy, Fawzi, 2022; Shahin, 2020). Such an approach is based on an implicit assumption that the features analyzed using a fixed-duration time window capture sonic information equally well at a microscopic level (allophones, phones, syllables) and a macroscopic level (words, sentences). However, the way certain emotions affect the articulation of phonemes may be different from the way they influence the pronunciation of words or sentences (prosodic characteristics). The above assumption motivated these authors to design a method that explicitly takes into account information at multiple time scales. Such a strategy could be referred to as a multi-time-scale (MTS) approach to feature aggregation.

In machine audition, MTS methods are not new. For example, they proved to be effective in the area of respiratory sound classification (Monaco *et al.*, 2020). More recently, Guizzo *et al.* (2020) have redesigned a standard convolutional neural network to take into account multiple time scales, demonstrating the superiority of such an approach compared to the standard convolutional networks when applied to speech emotion recognition. However, to the best of the authors' knowledge, no one has attempted to introduce MTS techniques to the "traditional" classification algorithms in the field of speech emotion recognition.

The main contribution of this work is to demonstrate that the performance of the traditional methods can be improved by aggregating features concurrently using time windows of different lengths (MTS approach). Such an approach could be likened to taking pictures with a camera equipped with a set of different focal lenses, allowing a photographer to simultaneously acquire both micro- and macroscopic views of a photographed scene. The additional novelty of this work is the application of a genetic algorithm to optimize the parameters of the feature extractors. In machine learning, genetic algorithms are typically exploited for the purpose of feature "selection" (Sayed *et al.*, 2019; Jadhav *et al.*, 2018). Application of genetic algorithms for tuning feature extractors is very rare. In this study, a genetic algorithm

was used to optimize the feature extractor responsible for the derivation of the Mel-frequency cepstral coefficients (MFCC). Although the research indicates that the parameters employed in the MFCC extraction algorithm should be optimized for a given task (Sahoo, Routray, 2016), undertaking a comprehensive optimization of MFCC extractors still constitutes an uncommon practice. Unlike most of the researchers, in this study, the authors decided to optimize 13 parameters of the MFCC extraction algorithm. Due to a relatively large number of parameters to be optimized, a popular grid-search optimization technique turned out to be impractical. While a genetic algorithm is commonly regarded as computationally demanding, in this study it proved to be more resourceful compared to the aforementioned grid-search technique.

To enhance the generalization property of the method, a data augmentation technique based on pitch shifting and time stretching was applied. In general, applying pitch shifting and time stretching effects to a speech signal may distort the overall prosody of the utterance, weakening its emotional expression. However, according to the research in the area of speech emotion recognition, the original emotional characteristics of speech signals may still be preserved if the above modulation processes are applied conservatively (Mohino-Herranz *et al.*, 2014; Tao *et al.*, 2023). Therefore, care was taken by the authors in employing pitch shifting and time stretching algorithms to maintain the original emotional characteristics of the speech recordings.

The proposed method was evaluated using five publicly available speech corpora, namely: CREMA-D (Cao *et al.*, 2014), eNTERFACE (Martin *et al.*, 2006), RAVDESS (Livingstone, Russo, 2018), SAVEE (Haq, Jackson, 2011), and TESS (Pichora-Fuller, Dupuis, 2020). The method was tested both under speaker-dependent and speaker-independent conditions. Moreover, its generalization property was also evaluated using cross-corpus tests. According to the obtained results, the developed method outperforms or it is comparable to the traditional ones for the selected datasets, demonstrating the benefits of using the MTS approach to feature aggregation.

The paper is organized as follows. In the next section we give an overview of the work of other researchers in the area of speech emotions recognition. In Sec. 3 we explain the methodology applied in our study. The obtained results are described in Sec. 4. The discussion of the obtained results and the conclusions are provided in Secs. 5 and 6, respectively.

2. Related work

Since the pioneering work of Picard (1995), the topic of the automatic speech emotion recognition has been investigated by many scientists, resulting

in a considerable body of research. Table 1 overviews in chronological order the example studies in this area published over the past thirteen years. They were arbitrarily selected by these authors. The studies presented in the table are limited to the traditional algorithms as they are pertinent to the work presented in this paper. The methods based on deep learning techniques have been omitted from the table. An interested reader is referred to papers by KHALIL *et al.* (2019) and PANDEY *et al.* (2019), for comprehensive reviews of deep learning techniques and their applications to speech emotion recognition.

In the traditional methods used for speech emotion recognition, a classical two-stage machine-learning topology is used. It consists of an audio feature extractor followed by a classification algorithm. The features derived in the feature extractor typically include Mel-frequency cepstral coefficients (MFCC), linear predictive coding (LPC) coefficients, signal energy, fundamental frequency (F0), and zero-crossing rate (ZCR), as exemplified in the third column in Table 1. The classical machine learning algorithms are commonly utilized as classifiers, most notably support vector machines (SVM), random forests (RF), multilayer perceptrons (MLP), Gaussian mixture models (GMM), techniques employing linear discriminant analysis (LDA), hidden Markov models (HMM), dynamic time-warping (DT), and K-nearest neighbors (KNN) (cf. the second column in Table 1). The most recent studies in the area of speech emotion recognition have demonstrated that the performance of the traditional methods could be improved by the incorporation of the ensemble of classifiers (SEKNEDY, FAWZI, 2022; OMMAN, ELDHO, 2022). Moreover, evolutionary algorithms, such as genetic algorithms, could be successfully used to further enhance their performance (WANG, HUO, 2019; LIU *et al.*, 2018). However, as already emphasized in Sec. 1, the genetic algorithms are

Table 1. Overview of the traditional methods used for speech emotions recognition since the year 2005 (in chronological order).

Reference	Model	Model input data	Corpus	Number of emotions	Reported accuracy [%]
LIN, WEI (2005)	HMM	F0, energy, F1-4, MFCC1-2, MBE1-5 with SFS selection	DES	5	99.5
	SVM	MEDC			88.9
MAJKOWSKI *et al.* (2016)	KNN	RMS, energy, MFCC1-12, delta features, ZCR, F0, SCG, SF, SRO with SFS selection	Polish radio broadcasts	6	75.6
	LDA				80.5
	SVM				79.2
GHALEB *et al.* (2019)	SVM	low-level energy descriptors, spectral, vocal delta coefficients	CREMA-D	6	56.2
			eNTERFACE		55.9
SHAHIN (2020)	HMM (two-stage)	MFCC	in Arabic	6	72.8
	GMM				63.3
	SVM				64.5
	VQ				61.5
ABDEL-HAMID (2020)	SVM	pitch, intensity, formants, MFCC, LTAS, wavelet features	EYASE	4	66.8
	KNN				61.7
SEKNEDY, FAWZI (2021)	MLP	RMS, MFCC1-12, ZCR, voicing probability, F0	RAVDESS	8	64.93
	SVM				70.56
	RF				59.31
	LR				62.64
SEKNEDY, FAWZI (2022)	MLP	MFCC1-40, Mel-spectrogram1-128, Chroma1-12, Tonnetz, Contrast1-8, RMS	EYASE	4	62.4
	SVM				50.6
	RF				62.4
	LR				62.9
	MLP + SVM + RT + LR (ensemble)				65.1
OMMAN, ELDHO (2022)	SVM (ensemble)	MFCC, ΔMFCC, ΔΔMFCC, spectral subband centroids, logfbank	RAVDESS	8	80.07
CAO *et al.* (2022)	Hessian-based subspace learning + domain adaption	MFCC, ΔMFCC, ΔΔMFCC, LPC, LAFC, Philips fingerprint, spectral entropy	EMO-DB, NNIME, IEMOCAP, MSP-IMPROV, MSP-PODCAST	4	54.93

predominantly used for feature selection (Kanwal, Asghar, 2021; Yildirim *et al.*, 2021; Sidorov *et al.*, 2014), whereas in our study they were employed to optimize the parameters of the feature extractors.

Note that the emotion recognition accuracy reported by an early work of Lin and Wei (2005) (cf. top row of Table 1) exceeds the accuracy levels reported by many other authors, including the most recent work of Cao *et al.* (2022) (cf. the bottom row of the table). This observation highlights the difficulty in the direct comparison across the studies, caused by the differences in the number of investigated emotions, differences in speech corpora characteristics, or differences in testing procedures (e.g., dissimilar proportions between the train and test sets), just to mention a few factors. Therefore, caution has to be exercised when comparing the methods based on a single accuracy metric or a particular testing procedure.

The speech corpora used for evaluation of the methods can be divided into three groups according to the way the emotions were evoked, namely: acted, elicited, and natural. See the work of Basu *et al.* (2017)

for the differentiation between these three groups. The speech corpora overviewed in Table 1 (fourth column) predominantly represent acted emotions (Lin, Wei, 2005; Ghaleb *et al.*, 2019; Abdel-Hamid, 2020; Seknedy, Fawzi, 2021; 2022; Omman, Eldho, 2022; Cao *et al.*, 2022). In the studies of Ghaleb *et al.* (2019) and Cao *et al.* (2022) in addition to the datasets incorporating acted emotions, the corpora employing elicited emotions were used as well. The remaining studies presented in Table 1 used either private corpora with an unknown type of emotions or corpora in which types of emotions are mixed or hard to verify (e.g., broadcasts). As mentioned earlier, the differences in the characteristics between the speech corpora could constitute a confounding factor when comparing the results. Therefore, it is imperative to employ several corpora when evaluating a given method. One of the most challenging evaluation scenarios involves testing new methods using corpora that were not "seen" during the training procedure (cross-corpus tests), including corpora representing different demographic, social, cultural, or language

Table 2. Overview of the speech corpora employed in this study.

Corpus	Reference	Number of speakers	Number of utterances	Duration of utterances [s]			Emotion categories	Emotion types
				Min.	Mean	Max		
CREMA-D	Cao *et al.* (2014)	92	7441	0.59	2.19	5.00	Happiness (1271)* Sadness (1270) Fear (1271) Anger (1271) Disgust (1271) Neutral (1087)	acted
RAVDESS	Livingstone, Russo (2018)	24	1248	1.00	1.74	4.21	Happiness (192) Sadness (192) Surprise (192) Fear (192) Anger (192) Disgust (192) Neutral (96)	acted
SAVEE	Haq, Jackson (2011)	4	480	0.86	3.22	7.14	Happiness (60) Sadness (60) Surprise (60) Fear (60) Anger (60) Disgust (60) Neutral (120)	acted
TESS	Pichora-Fuller, Dupuis (2020)	2	2800	1.13	1.90	2.86	Happiness (400) Sadness (400) Surprise (400) Fear (400) Anger (400) Disgust (400) Neutral (400)	acted
eNTERFACE	Martin *et al.* (2006)	10	1287	0.71	2.11	6.30	Happiness (212) Sadness (215) Surprise (215) Fear (215) Anger (215) Disgust (215)	elicited

* Number of recordings representing a given emotion category.

characteristics (Su, Lee, 2021; Seknedy, Fawzi, 2021; Tamulevičius *et al.*, 2020; Milner *et al.*, 2019; Kaya, Karpov, 2018; Cao *et al.*, 2022). In line with the abovementioned observations, in the present study, five following corpora were used, namely: CREMA-D (7441 utterances, 44 female and 48 male speakers), RAVDESS (1248 utterances, 12 female and 12 male speakers), SAVEE (480 utterances, 4 male speakers), TESS (2800 utterances, 2 female speakers), and eN-TERFACE (1,287 utterances, 5 female and 5 male speakers). All of these datasets were recorded in English. Only the eNTERFACE dataset contained recordings of elicited emotions, as the other four corpora represented acted emotions obtained from amateur or professional voice actors. Table 2 provides a detailed overview of the five corpora used in this study. In addition to speaker-dependent and speaker-independent tests, a cross-corpus test was also included in the evaluation procedure.

3. Method

The conceptual topology of the proposed algorithm is shown in Fig. 1a. It consists of an ensemble of the feature extractors (FE) coupled with the individual SVM classifiers. The prediction of the emotion category is undertaken using the ensemble voting model. The distinct aspect of the proposed method is that the feature extraction procedure is concurrently undertaken using long-term, mid-term, and short-term time windows, as depicted in the figure. Their duration is adjusted adaptively, depending on the duration of the original excerpts, although it does not exceed 7 s for long-term windows, 2.33 s for mid-term windows, and 0.7 s for short-term windows.

The algorithm depicted in Fig. 1a is computationally inefficient since for the long-term, mid-term, and short-term windows, the same set of the "primary" features has to be calculated. The phrase "primary features" is used in this paper to denote the metrics calculated in the feature extractors such as the zero-crossing rate, whereas the expression "secondary features" represents the statistics derived from the primary features. A computationally optimized topology of the proposed method is illustrated in Fig. 1b. It consists of the single feature extractor (FE), providing a set of primary features, and the ensemble of the feature aggregators (FA) coupled with the individual SVM classifiers. The role of the feature aggregators (FA) is to convert specific parts of primary features into secondary statistical features.

In this study, a computationally optimized version of the algorithm has been implemented (Fig. 1b). Its constituent blocks are described in detail in the subsequent sections. More information on the MTS approach proposed in this study is provided in Subsec. 3.3.

a)

b)
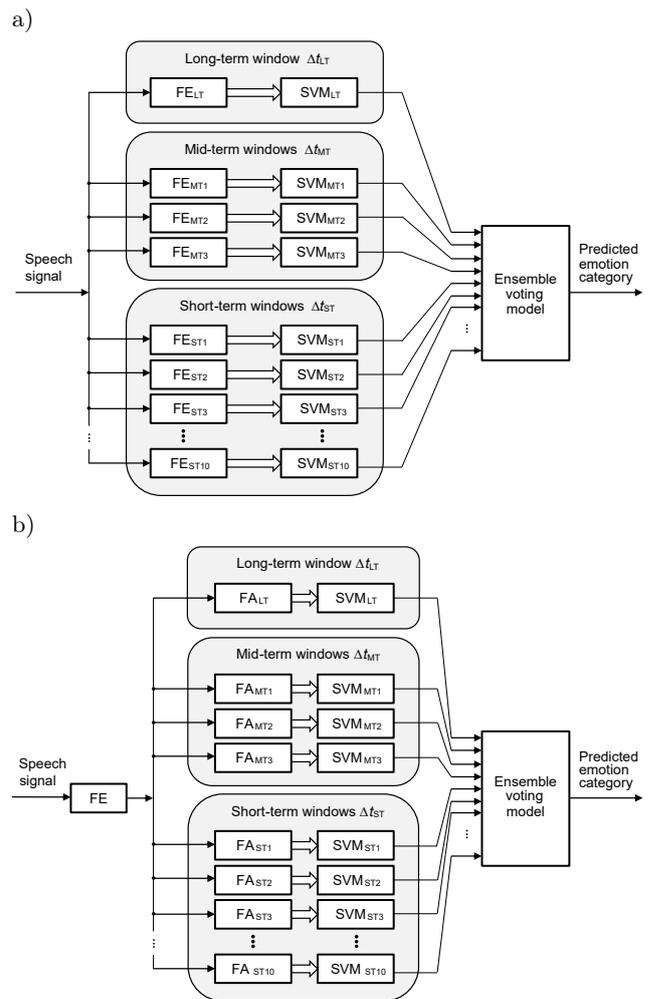
Fig. 1. Multi time-scale speech emotion classification method: a) conceptual algorithm; b) computationally optimized algorithm.

The number of recognizable emotion classes depended on the dataset used in the training stage. In this work, the focus has been put on the Ekman basic emotion set, comprising such emotion categories as anger, joy, disgust, sadness, fear, and surprise (Ekman, 1992), with the addition of the "neutral" emotion class representing utterances that were not emotionally charged, resulting in maximum of 7 classes in total. In some evaluation cases, a subset of this basic emotion set has been taken into consideration due to the dataset limitations (but never being smaller than 5 emotion classes).

3.1. Feature extraction

Prior to undertaking the feature extraction procedure, essential pre-processing tasks were carried out. Namely, the leading and the trailing silence of every recording was trimmed. The silence cutoff point in each recording was the first sample of which the absolute value exceeded 5% of the maximum absolute

value of all samples in the recording. To equalize the audio signals' variance, each recording has been further z-standardized. Since the sample rate in the proposed algorithms was set to 44 100 Hz, all the audio recordings with a miss-matched sample rate were resampled accordingly.

For each recording, the signal was divided into time-frames of 1102 samples each, with an overlap of 827 samples. Such unusual values were adopted in this study, since the employed genetic algorithm (described in Subsec. 3.2) proved those values to be the best in the context of this experiment. Since the sample rate of the audio signals equaled 44 100 Hz, the duration of a single time-frame amounted to approximately 25 ms of which around 19 ms overlapped with the neighboring frame. The Hann window was applied to the signals in each frame. Similarly as before, the choice of the window-type was determined by the genetic algorithm. Finally, the features were extracted for each frame.

The following features have been taken into account: MFCC (20 coefficients), ZCR coefficient, fundamental frequency, and spectral flux, yielding 23 features in total. Additionally, the delta and delta-delta values were computed for ZCR, fundamental frequency, and spectral flux, respectively, as they provide information on abrupt changes and transitions of those features. Ultimately, for each audio frame, a primary feature vector of size 29 was computed. All of the features were calculated using the Essentia toolbox (Bogdanov *et al.*, 2013). For all the configuration parameters unspecified in this paper, default values provided by the toolbox were used.

3.2. Genetic algorithm

While most of the features were relatively straightforward to calculate, the estimation of MFCC turned out to be a more demanding task. The Essentia MFCC extractor takes 13 different parameters, including the number of Mel-frequency coefficients, the number of Mel-frequency bands, upper and lower bounds of the frequency range, discrete cosine transform type, type of spectrum, and the liftering coefficient. Hence, manual tuning proved to be a challenge and a need for an appropriate optimization method arose. The complete list of the optimized parameters is provided in Table 3.

While the popular greedy optimization algorithm Grid Search is usually very effective for parameter tuning, for this exact problem its computational complexity turned out to be impractical. Therefore, an alternative optimization method was utilized, namely the Genetic algorithm (Mitchel, 1996). In the context of this study, it is more computationally efficient than Grid Search as it avoids undertaking checks for every possible solution. According to the literature, genetic algorithms are most often utilized in the feature selection process and classifier hyperparameter optimization (Kanwal, Asghar, 2021; Wang, Huo, 2019). By contrast, in this study, a genetic algorithm has been deployed to tune the parameters of the MFCC extractor. The parameter values determined during this search were subsequently used in the Essentia extractor to calculate the MFCC coefficients. The genetic algorithm was implemented by the first author as a multithreaded Python script. For reproducibility of the research, the developed code is included in the publicly available repository at GitHub (Stefanowska, Zieliński, 2023).

The properties of the implemented genetic algorithm are overviewed in Table 4. A specimen in this problem is understood to be a specific parameter value combination from the set of considered values for each parameter. The fitness value for each specimen is calculated by extracting MFCC using its parameter values, training a single SVM classifier with those extracted coefficients, and checking its accuracy on a validation set. All the fitness values were calculated using the RAVDESS dataset (Livingstone, Russo, 2018)

Table 3. Parameters of the MFCC extraction algorithm optimized by the genetic algorithm.

Parameter	Considered values	Genetic algorithm results
Number of Mel coefficients	10, 13, 20, 40, 80, 120	20
Frame size (in samples)	512, 756, 1024, 1102	1102
Window type	Hamming, Hann	Hann
Mel scale implementation method	Auditory Toolbox (Slaney, 1998), HTK toolkit (Young *et al.*, 2006)	Auditory Toolbox
Logarithmic compression type	Natural, power, magnitudes, logarithmic	Magnitudes
Discrete cosine transform type	II, III	III
Normalization method	Unit sum, unit triangle, unit max	Unit triangle
The upper bound of the frequency range [Hz]	6000, 8000, 16 000, 20 000	16 000
The lower bound of the frequency range [Hz]	0, 50, 100, 200, 500	50
The number of Mel-bands in the filter	26, 128	128
Type of weighting function for determining triangle area	Warping, linear	Warping
Type of spectrum	Magnitude, power	Power
The liftering coefficient	0, 22, 10, 40, 100	40

Table 4. Properties of the genetic algorithm.

Property	Value
Maximum population size	10
Potential parent selection method	3-way tournament
Potential parent number	5
Crossover probability	0.7
Mutation probability	0.5

and then cached to save the computational power in case of reoccurring specimens. The basic properties of the genetic algorithm were picked based on how effectively they seemed to perform in the few initial iterations (Table 4). A relatively high mutation probability proved to help with reaching more effective specimens quicker.

For every parameter, a finite set of possible values was specified (Table 3). Certain parameters were numerical and their possible values were selected empirically, others were categorical (e.g., the Mel scale implementation method), and their possible values were already provided by the toolbox. The best set of final parameter values (see the last column in Table 3) was determined after 120 iterations of the genetic algorithm. A properly tuned MFCC extractor proved to significantly increase the accuracy of the trained model, as illustrated in Fig. 2.
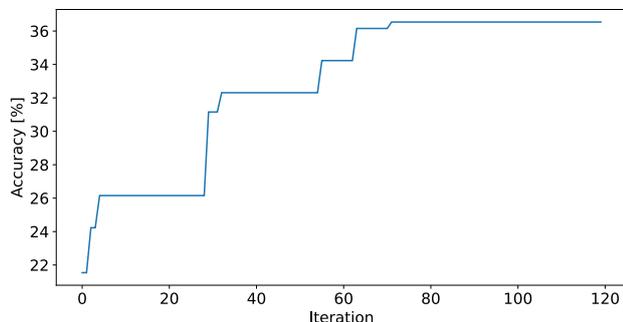
Fig. 2. Example learning curve of the genetic algorithm.

3.3. Multi-time-scale approach to feature aggregation

The primary features calculated by the FE, as described in the previous sections, are then processed using an ensemble of the FA. Each feature aggregator takes a specific "slice" of the primary features, according to the size of the corresponding time window. The following statistics are calculated in the process of the feature aggregation: mean values, standard deviations, minimum and maximum values, as well as lower and upper quartiles, yielding the set of 203 secondary features at the output of each feature aggregator. In accordance with the typical practice in machine audition, the secondary features are further z-standardized (KREYSZIG, 1979).

According to the proposed topology, the algorithm consists of the three blocks signified by the shaded areas in Fig. 1b, each utilizing a different time window length. The top-most block comprises a single feature aggregator FA_{LT} connected to its associated classifier (SVM_{LT}). A long-term window Δt_{LT} of the feature aggregation is used in this block. The duration of the time-window is in this case set to the duration of the whole utterance, constrained to 7 s maximum. In other words, the primary features are aggregated for the initial 7 s of each utterance. If the recording exceeds that limit, it is trimmed to the maximum permissible length of 7 s. It is presumed that the top-most block is responsible for capturing and processing the prosodic features from the whole speech utterance. Note that the way the signal is processed using the top-block (in isolation from the remaining two blocks) could be considered as the standard approach, commonly applied by the researchers in the field of speech emotion recognition (OMMAN, ELDHO, 2022; SEKNEDY, FAWZI, 2022; ABDEL-HAMID, 2020; GHALEB *et al.*, 2019). Therefore, in this study this part of the algorithm is considered as the "baseline" method.

In the middle block depicted in Fig. 1b, the long-term window is divided into the three overlapping mid-term windows of maximum duration equal to $\Delta t_{MT} = 2.33$ s each, with an overlap of approximately 0.1 s. These windows are responsible for dividing all the primary features into the ones representing the initial, middle, and ending part of each utterance, respectively. The primary features from these three mid-term time windows are then processed individually by the three feature aggregators. The statistics calculated by the feature aggregators are the same as the ones described above in the case of the long-term window. In the next step, the secondary features derived by the feature aggregators are fed to the three SVM classifiers. Due to the shorter length of the window of analysis, it could be supposed that the middle block would better utilize information conveyed by individual words.

The finest temporal resolution is exhibited by the bottom-block shown in Fig. 1b. In this case the long-term window of analysis is divided into ten overlapping short-term windows. Consequently, the window of analysis is further reduced down to $\Delta t_{ST} = 0.7$ s at most, with an overlap of 0.05 s. The primary features encompassed by each of the ten short-term windows are processed independently by the ten feature aggregators, and then the ten classifiers. Out of the three blocks included in the algorithm, the bottom one is the most complex, as it consists of the ten feature aggregators combined with the ten associated SVM classifiers. It could be hypothesized that the bottom-block would be particularly efficient in capturing and processing information represented by short words or syllables. In total, each speech utterance is concurrently analyzed

and classified using 14 time windows (one long-term, three mid-term, and ten short-term windows).

3.4. Classification algorithm

The support vector machine (SVM) classifier was selected as the base model for the proposed method. SVM is one of the most commonly used traditional machine learning techniques, which despite being potentially less effective than modern deep learning models, still prove advantageous in certain cases – especially when available datasets are sparse or too small to train effective deep models (Omman, Eldho, 2022; Seknedy, Fawzi, 2022; Abdel-Hamid, 2020; Ghaleb *et al.*, 2019; Shahin, 2020). In the proposed method, the SVM's hyperparameters are optimized using the Grid Search algorithm. The parameters chosen for tuning include the SVM's C coefficient with possible values of 0.1, 1, and 10; gamma coefficient with possible values of $\frac{1}{\text{feature number}} \times 0.1$, $\frac{1}{\text{feature number}}$, $\frac{1}{\text{feature number}} \times 10$; and the Kernel function that might be chosen to be linear, polynomial of the 3rd degree, or radial basis function (RBF).

Another parameter that gets optimized by the Grid Search algorithm, yet does not belong to SVM's hyperparameters, is the number of selected features that are used as the final input vector. The list of possibilities include: $X = 100, 90, 50$ or 25% of all the original features. This optimized value is used at the feature selection stage, which consists of filtering out all the constants and then utilizing the selection method based on the ANOVA F statistic from the scikit-learn library (Pedregosa *et al.*, 2011). A rank of features is created, of which only top X features with the best score get selected (as mentioned previously, the X value is determined by the optimization algorithm). The sequential process of parameter tuning, feature selection and classifier training were managed with the use of the pipeline tool from the scikit-learn toolbox (Pedregosa *et al.*, 2011).

3.5. Ensemble voting model

To make use of all the micro and macro information contained in each of the statistical feature vectors obtained as described in the previous sections, they were used as inputs for separate SVM classifiers which were then combined into an ensemble voting classifier (cf. Fig. 1). The final assembling stage involved building the voting classifier. The soft voting method was utilized. The winning class is the one with the greatest total sum of probability of occurring in each component classifier. Additionally, every probability was weighted based on how well the classifier performed on the validation dataset during the tuning phase. In summary, the score for each class was calculated using the equation:

$$s_c = \sum_{i=0}^{N} w_i * p_{c,i}, \qquad (1)$$

where s_c – score of the emotion class c; w_i – voting weight of the i-th classifier (its accuracy on the validation dataset during the tuning stage); $p_{c,i}$ – probability of the emotion class c in the i-th classifier; N – number of classifiers.

The emotion class with the maximum score is considered to be the final 'decision' of the ensemble voting classifier.

3.6. Data augmentation

In order to enhance the generalization property of the classification model, all the speech recordings went through the data augmentation process. Simple pitch shifting and time stretching operations available in the librosa toolbox (McFee *et al.*, 2015) were applied to enrich the existing datasets. Introducing pitch shifting and time stretching effects to speech signals influences the overall prosody of the utterance. Consequently, such processes may modify emotional expressions. However, the authors assumed that the original emotional characteristics of the speech recordings would be preserved if these effects were applied cautiously, that is using conservative pitch shifting and time stretching limits. This assumption is in accordance with the research in the area of speech emotion recognition (Mohino-Herranz *et al.*, 2014; Tao *et al.*, 2023). In line with the above considerations, the pitch has been shifted up and then down by three semitones whereas the audio signals have been sped up and slowed down by 25%, respectively, resulting in four new audio files for each existing audio file. All the augmented recordings were further used only in the training sets (the test sets comprised solely the original recordings).

The developed method was implemented in Python. The code was made publicly available at GitHub repository (Stefanowska, Zieliński, 2023).

4. Results

The performance of the developed method was evaluated in five experiments. The comparisons were made both against the traditional algorithms as well as the deep learning techniques, published recently in the literature. Three different experimental methodologies have been considered, including speaker-dependent tests, speaker-independent tests, and cross-corpus tests.

4.1. Speaker-dependent tests

In this approach, recordings coming from the same speakers can appear in validation, training, and testing sets. The speaker-dependent tests were conducted

for a single speaker using the TESS dataset. In this case, only samples belonging to the younger actress were utilized. In total, 1400 audio recordings were employed, representing seven emotion categories (200 recordings per emotion). Moreover, additional 5 600 augmented excerpts were utilized in this experiment. The tests followed the methodology from the work of CHATTERJEE *et al.* (2021). The dataset was split with a 65/15/20 percentage ratio in order to obtain the training, validation, and test subsets, respectively. For the sake of comparison, aside from testing only the main proposed method utilizing an ensemble of SVM classifiers with the MTS approach and data augmentation (MTS + Aug), a variant without data augmentation (MTS), as well as variants based on a single SVM classifier with augmentation (SVM + Aug) and without it (SVM), were tested too, which resulted in four test cases. All the experiments were repeated 30 times with different randomization seeds and the final result was the mean value of all the individual accuracy values obtained in the repeated trials. Those results were presented on the mean accuracy chart along with the corresponding standard deviations (Fig. 3). It can be seen that the proposed ensemble classifier with the MTS approach, labeled as MTS in the figure, outperformed the method proposed by CHATTERJEE *et al.* (2021). This outcome was statistically significant, based on the one sample *t*-test (*p*-value was less than 10^{-4} for the dependent *t*-test with a 0.05 alpha level). It also performed better than the single SVM with the use of augmented data ($p < 10^{-4}$). Moreover, it outperformed the standard SVM algorithm without data augmentation (*p*-value was approximately equal to 10^{-4}). Hence, the addition of the augmented data for the training stages did not improve the accuracy for this case. Figure 4 shows the accuracy of recognition of the individual emotions for the TESS dataset. It can be seen
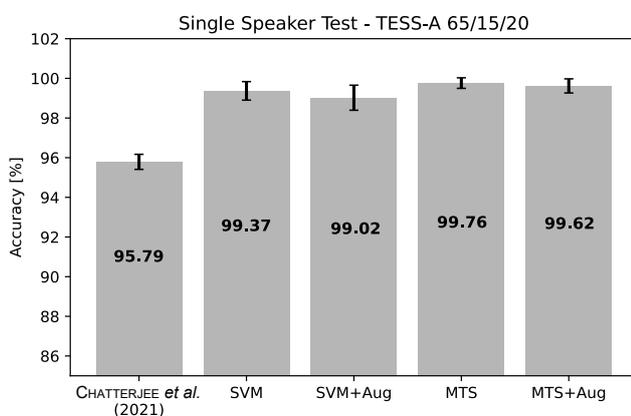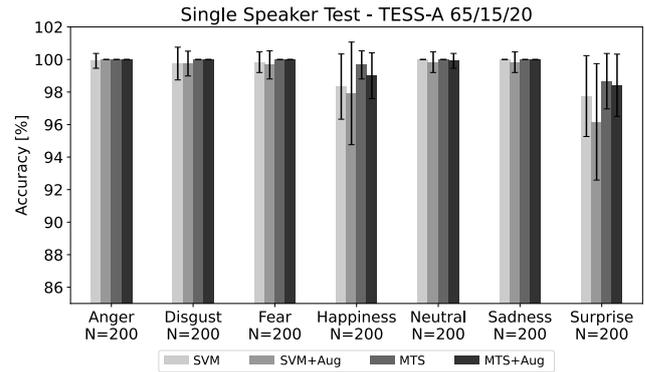
Fig. 4. Accuracy of recognition of individual emotions for the TESS 65/15/20 split ratio experiment with the speaker-dependent testing approach – only younger actress' samples were used. The results represent the mean accuracy values and associated standard deviations.

that all the emotions were identified with almost 100% accuracy using the proposed MTS approach, except for the "surprise" category, which was recognized with 98.67% (SD 1.7%) accuracy.

A separate experiment involving a speaker-dependent test of the proposed MTS method was performed. It was based on the 10-fold cross-validation procedure, conducted using the RAVDESS dataset. It contained 192 recordings per emotion, apart from the neutral state, which was represented by 96 excerpts, giving 1248 audio files in total (plus the addition of 4992 augmented samples). The results showed that the class that was relatively the hardest one to classify was the neutral emotional state (Fig. 5a). It was often mistaken with sadness. Another class often mistaken with sadness was fear. Classes that seemed to be the most recognizable by the proposed method were anger and disgust – they also tended to be mistaken with each other more than with any other emotion.

4.2. Speaker-independent tests

In this experiment, the first test with the speaker-independent constraint was a 10-fold cross-validation and it was conducted using the eNTERFACE dataset with 6 emotion classes. For this repository, each emotion was represented by 215 recordings except happiness which was exemplified by 212 audio excerpts. The reported results are the average values of the accuracies from all the folds. Corresponding standard deviations were also calculated. The literature reference was a method based on a SVM classifier, utilizing multimodal inputs (GHALEB *et al.*, 2019). For comparison purposes, solely audio-only average accuracy was taken into account. The obtained results are presented in Fig. 6. In this case, a single SVM classifier trained with the aid of the augmented data (SVM + Aug) performed better than a single SVM with no augmentation (*p* = 0.0045). Similarly, the proposed MTS model

Fig. 3. Accuracy chart for the TESS 65/15/20 split ratio experiment with the speaker-dependent testing approach – only younger actress' samples were used. The results represent the mean accuracy values and associated standard deviations.

a)

b)

c)
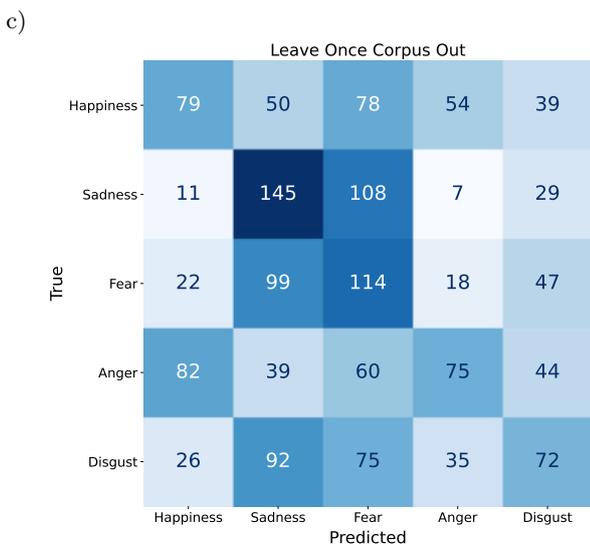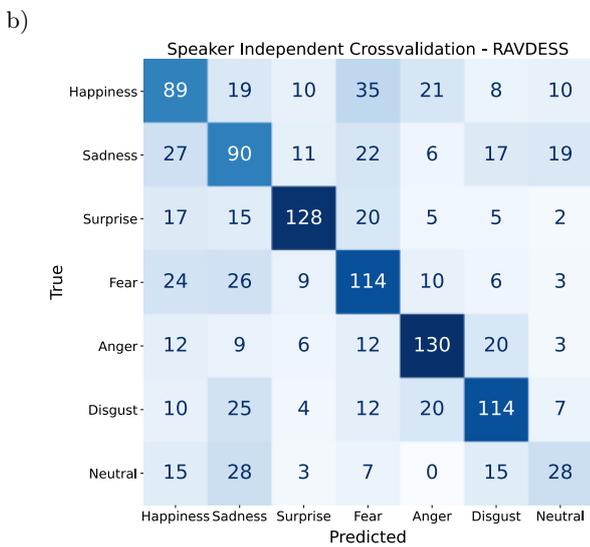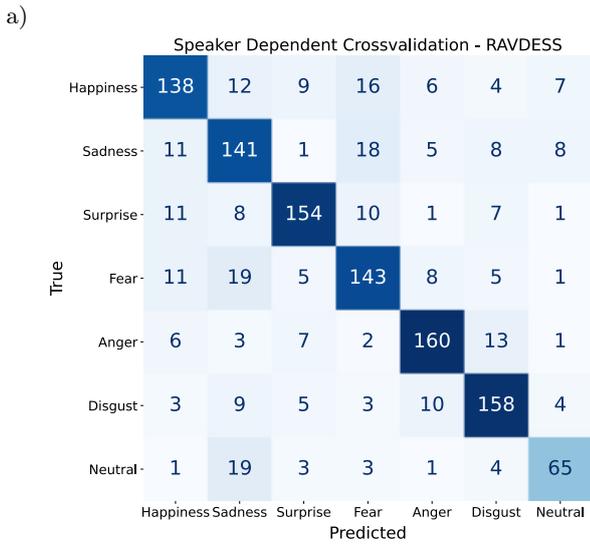
Fig. 5. Classification accuracy tests results for the developed model presented as confusion matrices. Three different testing approaches: a) speaker-dependent; b) speaker-independent; c) corpus-independent.
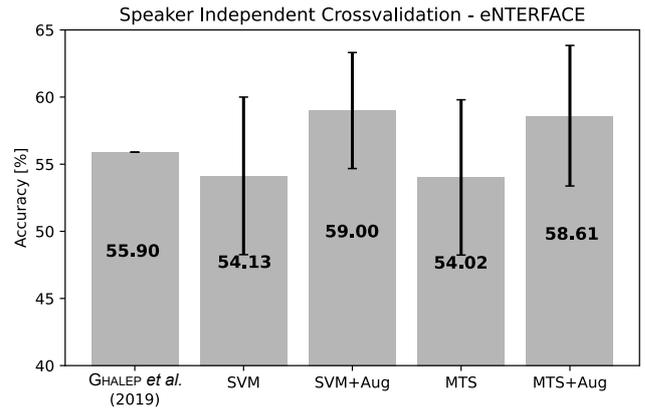
Fig. 6. Accuracy chart for the eNTERFACE cross-validation experiment with the speaker-independent testing approach. The results represent the mean accuracy values and associated standard deviations.

using the augmented data (MTS + Aug) performed better than the same model not utilizing it (MTS) ($p = 0.0036$). The advantage of using MTS model with the augmented data over the literature example could not be verified as the p-value for the one sample t-test was equal to 0.0773, unlike the advantage of exploiting a single SVM classifier with the use of the augmented data (SVM + Aug) ($p = 0.03$). The reasons for such a high accuracy of a single SVM classifier trained using the augmented data could be attributed to properly tuned MFCC extractor and model's hyperparameters. For this amount of data, a properly tuned single classifier proved to be sufficient. As the advantage of the proposed MTS method over the literature example could not be verified ($p = 0.9227$), the results were statistically comparable. According to both the results obtained for the single SVM classification algorithm and the MTS ensemble method, the chosen augmentation procedure substantially improved the accuracy using the selected dataset. Figure 7 shows the accuracy of
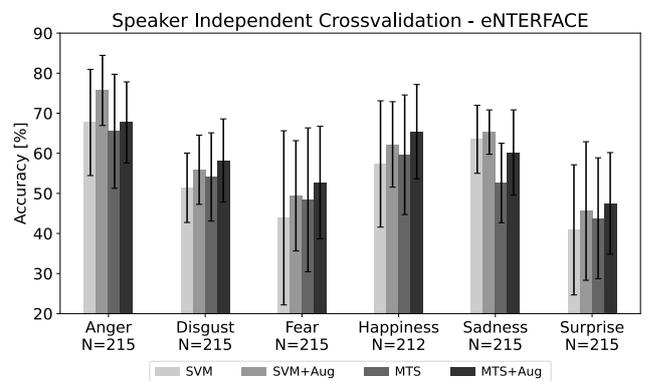
Fig. 7. Accuracy of recognition of individual emotions for the eNTERFACE cross-validation experiment with the speaker-independent testing approach. The results represent the mean accuracy values and associated standard deviations.

recognition of the individual emotions for the eNTER-FACE dataset. It can be seen that the anger category is recognized with the highest accuracy, reaching almost 76%, whereas the fear and surprise categories are identified with the lowest accuracy at a level ranging from 41% to 53%.

The second test set was based on the same literature reference. In this case, it was conducted on the CREMA-D dataset with 6 emotion classes. Due to time constraints, in this experiment 996 samples were randomly chosen from the complete dataset while maintaining the original distribution of speakers and emotions, which resulted in the repository of 166 utterances per emotion class. The results are presented in Fig. 8. It can be seen that a single SVM classifier performed better with the aid of augmented data ($p = 0.0072$) and that the MTS ensemble model with the augmented data performed better than the same model without it ($p = 0.0002$). The advantage of the literature example over the proposed method turned out not to be statistically significant ($p = 0.2385$), therefore two methods seem to be comparable. Figure 9 shows
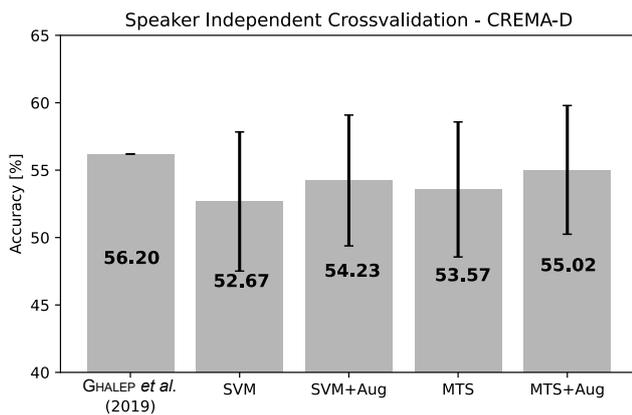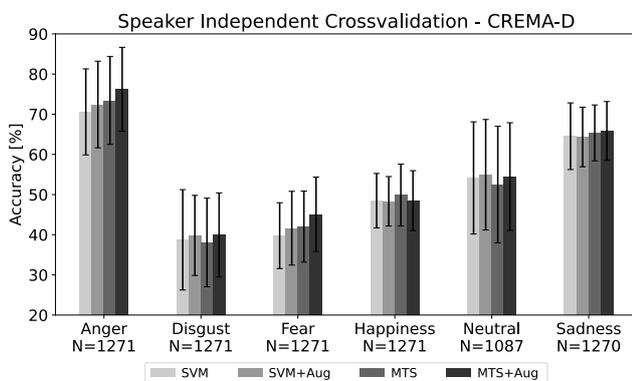
the accuracy of recognition of the individual emotions for the CREMA-D dataset. Similar to the previous experiment employing the eNTERFACE dataset, the anger category is recognized with the highest accuracy, reaching 76%. The disgust and fear categories are identified with the lowest accuracy at a level ranging from 38 to 45%.

The third test set was based on the work of GUIZZO *et al.* (2020), who developed an advanced model employing convolutional neural networks. Their model was trained on the RAVDESS dataset, comprising 192 recordings for each emotion (96 for neutral state), as mentioned before. The results reported by the quoted authors constitute the average accuracy values obtained from a 4-fold cross-validation test. The cited work also utilized the MTS approach by introducing multiple convolution kernels and obtaining differently scaled feature maps that were all used as the model input. The dataset was split with an approximate ratio of 70/20/10 into training, validation, and testing subsets, respectively. In this study, the test was repeated 30 times, and the final result was estimated as the mean value of all the individual accuracy values. According to the obtained results (see Fig. 10), the MTS ensemble model performed better than the single SVM classifier for the case without the use of the augmented data ($p = 8 \times 10^{-4}$). The advantage of using the augmentation process in this case could not be statistically verified. Importantly, the difference between the result reached by GUIZZO *et al.* (2020) and that obtained using the proposed MTS method (MTS + Aug) was not statistically significant ($p = 0.2875$). Consequently, it could be concluded that the accuracy of the proposed MTS method proved to be comparable to the one reached using a state-of-the-art deep learning technique applied to the RAVDESS data set. Figure 11 shows the accuracy of recognition of the individual emotions for the RAVDESS dataset. Similar to the results obtained in the previous two experiments,

Fig. 8. Accuracy chart for the CREMA-D cross-validation experiment with the speaker-independent testing approach. The results represent the mean accuracy values and associated standard deviations.

Fig. 9. Accuracy of recognition of individual emotions for the CREMA-D cross-validation experiment with the speaker-independent testing approach. The results represent the mean accuracy values and associated standard deviations.

Fig. 10. Accuracy chart for the RAVDESS 70/20/10 split ratio experiment with the speaker-independent testing approach. The results represent the mean accuracy values and associated standard deviations.
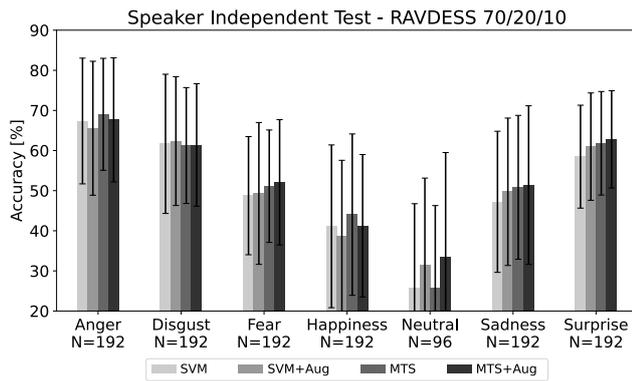
Fig. 11. Accuracy of recognition of individual emotions for the RAVDESS 70/20/10 split ratio experiment with the speaker-independent testing approach. The results represent the mean accuracy values and associated standard deviations.

the anger category was recognized with the highest accuracy, reaching approximately 69%. The neutral category was identified with the lowest accuracy at a level ranging from 26 to 33%.

For the 10-fold speaker-independent cross-validation test using the RAVDESS dataset, similarly to the previously discussed speaker-dependent evaluation (see Subsec. 4.1), a neutral emotional state turned out to be the class with the lowest accuracy (Fig. 5b). Likewise, it was often mistaken with sadness, although this time even more frequently. As the overall accuracy is substantially lower than that obtained for the speaker-dependent case, the system mixes up emotional states much more often. However, the anger emotion still exhibits a relatively high recognition rate.

4.3. Cross-corpus tests

For the cross-corpus test of the proposed MTS method, a leave-one-corpus-out cross-validation experiment was conducted. For this purpose, the following English datasets were used: CREMA-D, RAVDESS, SAVEE, TESS, and eNTERFACE. The selected corpora represented different but overlapping sets of emotions and contained vastly different number of recordings (see Table 2). Due to the above factors, the datasets were unified. Namely, the files with emotions that were not present in every corpus were discarded (neutral and surprise), which resulted in five common emotional classes – happiness, sadness, fear, anger, and disgust. The remaining recordings were sampled (while maintaining the original distribution of emotions and sexes) so that all the datasets were of the same size – 300 recordings each, which was the size of the smallest dataset (SAVEE). Considering that each repository was balanced in terms of the number of recordings representing each emotion, the number of recordings taken from every dataset representing a given emotion equaled 60. With the addition of the augmented files

(1200 for each corpus), the total number of files was equal to 7500. In each iteration one corpus became a testing set, another one was selected as a validation dataset, whereas the remaining three datasets were used for training.

Even with the reduced number of emotional classes, the corpus-independent test turned out to give the lowest overall accuracy so far, being equal to 32.33%. Nevertheless, this value was still statistically greater than the chance level, which in the experiment amounted to 20% ($p = 0.0035$). Despite this outcome, sadness turned out to be recognized comparatively often, with the classification accuracy reaching as much as 48.33%. As shown in Fig. 12, its recognition rate turned out to be comparable to the recognition rate of sadness in the speaker-independent test on a single corpus. It was, however, often mistaken with fear and disgust (Fig. 5c). Unlike in previous tests, for the cross-corpus test, anger and disgust were the hardest emotions to classify.
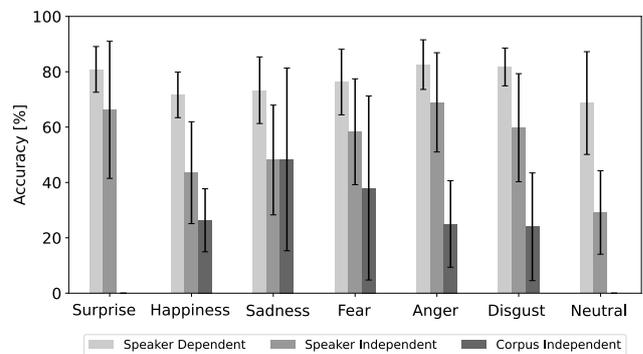
Fig. 12. Classification accuracy tests results obtained using the proposed MTS model under the three testing conditions: speaker-dependent, speaker-independent, and corpus-independent. Note that the corpus-independent approach applies only to five out of seven emotions considered in the remaining tests (see the text for details).

5. Discussion

Out of the four experiments conducted in this study, the two experiments proved that the multi-time scale approach to feature aggregation yields better classification results compared to the conventional techniques. These two experiments were based on the TESS and RAVDESS speech corpora, respectively. Moreover, in one of the experiments, involving the RAVDESS corpus, the MTS method achieved a very similar classification accuracy level compared to the one obtained using the state-of-the-art deep learning technique (GUIZZO *et al.*, 2020), as illustrated in Fig. 10. Furthermore, the acquired results highlight the benefits of using a data augmentation technique (Figs. 6 and 8).

There are two recent studies resembling our work as they also employed the ensembles of classifiers.

In the work by SEKNEDY and FAWZI (2022), an ensemble of four different classification algorithms was used, reaching a maximum classification accuracy of 65.1%. In their work, they utilized an Arabic speech corpus. Therefore, due to the linguistic mismatch, our method cannot be directly (and fairly) compared to the aforementioned one. In their recent work, OMMAN and ELDHO (2022) have employed an ensemble of 20 SVM classifiers. They used a bootstrap aggregating technique to train their ensemble model, reaching an accuracy of 80.07% when tested on the RAVDESS corpus, outperforming our method by 3 percentage points. While this outcome may indicate an inferiority of our method, the cited authors did not provide sufficient details regarding their testing methodology (e.g., whether the tests were speaker-independent), preventing other researchers from a rigorous comparison of the results.

According to the results of the fifth experiment, involving the cross-corpus test (Subsec. 4.3), the classification accuracy of the proposed method was substantially lower than the accuracy levels obtained within the selected corpora using speaker-dependent and speaker-independent tests, respectively. This outcome does not invalidate the proposed method but indicates that its generalization property needs to be improved. Note, that the cross-corpus tests are still very rare in the literature as they constitute the most rigorous way of testing the speech emotions recognition systems (TAMULEVIČIUS *et al.*, 2020). The recent study by CAO *et al.* (2022) confirms that the average accuracy for this testing approach remains relatively low, especially for the simpler methods.

The presented results indicate that the proposed MTS method has an advantage over the baseline technique employing a single classifier with a fixed time-window applied during the feature aggregation. It is, however, more computationally complex, as it utilizes multiple classifiers instead of one. For example, it took 2.73 ms for a single SVM model to classify one recording. Compared to that, using the MTS model for classification took about 5.86 times longer (16 ms). There was also a significant difference between the total duration time of tuning and training. Namely, tuning and training a single SVM classifier on 4160 training files and 260 validation files took in total 8.37 s. In contrast, tuning and training an MTS model on the same dataset took 8 min 33 s and, consequently, it requires more resources. The training of a genetic algorithm itself to tune the feature extractor took 12 h 11 min 47 s, which is the reason why it was not used as a part of a training pipeline but constituted a separate procedure conducted once. All the calculations were carried out using parallel processing on 8 threads of the Intel Core i5 1.6G Hz processor.

There are some limitations of this study that need to be acknowledged. Firstly, the undertaken experiments were based on only five datasets. Broader conclusions could be reached if more corpora were taken into consideration. Secondly, due to the data storage and computation constraints of the hardware used, a subset of the CREMA-D corpus was employed, as described in Subsec. 4.2. Thirdly, the feature extractor tuning procedure was performed using a single speech corpus (RAVDESS), potentially biasing the model towards the selected data set. Fourthly, the duration of the long-term window applied for the feature aggregation was limited to 7 s. In retrospect, the above constraint could be too short for some applications, potentially causing the method to discard important information conveyed by the prosodic characteristics at the ending parts of the sentences. In the present study, this issue affected only one recording belonging to the SAVEE repository (the audio excerpt was trimmed as its duration exceeded the limit). Optimization of the long-term window applied for the feature aggregation as well as the exploration of different optimization strategies for the feature extraction may constitute the subject of future work.

6. Conclusions

This study presents an improved method of speech emotions recognition using an ensemble of SVM classification algorithms. The novelty of the proposed method consists in using a MTS approach to the feature aggregation procedure. Another distinct aspect of the proposed technique is the application of the genetic algorithm to optimize the feature extraction process. Out of the four experiments conducted in this study, the two experiments support the hypothesis that the MTS approach to feature aggregation yields better classification results compared to the conventional way of feature aggregation based on a fixed-duration time window. This implies that the proposed MTS approach is not always superior compared to the conventional technique. Nevertheless, it exhibits satisfactory performance for the selected datasets, matching or outperforming the recently published methods. Interestingly, in one of the experiments conducted within this study, the results obtained using the proposed MTS method proved to be comparable to the ones achieved by means of the state-of-the-art deep learning technique. This outcome indicates that a properly developed traditional classification method could be competitive to a deep learning algorithm. As a side observation, the obtained results exemplified the benefits of data augmentation. The technique of data augmentation is commonly used for the training of deep learning models (MILNER *et al.*, 2019). However, this study demonstrated the advantages of applying such a technique during the development of the traditional model. Future work may be focused on testing the MTS method using a broader scope of speech corpora,

with the aim of gaining knowledge as to how to further optimize the technique within individual data sets while still maintaining a satisfactory cross-corpus generalization property.

Acknowledgments

References

1. ABDEL-HAMID L. (2020), Egyptian Arabic speech emotion recognition using prosodic, spectral and wavelet features, *Speech Communication*, **122**: 19–30, doi: 10.1016/j.specom.2020.04.005.

2. BASU S., CHAKRABORTY J., BAG A., AFTABUDDIN M. (2017), A review on emotion recognition using speech, [in:] *International Conference on Inventive Communication and Computational Technologies (ICICCT)*, doi: 10.1109/ICICCT.2017.7975169.

3. BOGDANOV D. *et al.* (2013), ESSENTIA: An audio analysis library for music information retrieval, [in:] *International Society for Music Information Retrieval Conference (ISMIR'13)*, pp. 493–498.

4. CAO H., COOPER D.G., KEUTMANN M.K., GUR R.C., NENKOVA A., VERMA R. (2014), CREMA-D: Crowd-sourced emotional multimodal actors dataset, *IEEE Transactions on Affective Computing*, **5**(4): 377–390, doi: 10.1109/TAFFC.2014.2336244.

5. CAO X., JIA M., RU J., PAI T. (2022), Cross-corpus speech emotion recognition using subspace learning and domain adaption, *EURASIP Journal on Audio, Speech, and Music Processing*, **2022**: 32, doi: 10.1186/s13636-022-00264-5.

6. CHATTERJEE R., MAZUMDAR S., SHERATT R.S., HALDER R., MAITRA T., GIRI D. (2021), Real-time speech emotion analysis for smart home assistants, *IEEE Transactions on Consumer Electronics*, **67**(1): 68–76, doi: 10.1109/TCE.2021.3056421.

7. CHOI W.Y., SONG K.Y., LEE C.W. (2018), Convolutional attention networks for multimodal emotion recognition from speech and text data, [in:] *Proceedings of Grand Challenge and Workshop on Human Multimodal Language (Challenge-HML)*, pp. 28–34, doi: 10.18653/v1/W18-3304.

8. EKMAN P. (1992), An argument for basic emotions, [in:] *Cognition and Emotion*, **6**(3–4): 169–200.

9. ESKIMEZ S.E., DUAN Z., HEINZELMAN W. (2018), Unsupervised learning approach to feature analysis for automatic speech emotion recognition, [in:] *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5099–5103, doi: 10.1109/ICASSP.2018.8462685.

10. GARCÍA-MARTÍN E., RODRIGUES C.F., RILEY G., GRAHN H. (2019), Estimation of energy consumption in machine learning, *Journal of Parallel and Distributed Computing*, **134**: 75–88, doi: 10.1016/j.jpdc.2019.07.007.

11. GHALEB E., POPA M., ASTERIADIS S. (2019), Metric learning-based multimodal audio-visual emotion recognition, *IEEE MultiMedia*, **27**(1): 37–48, doi: 10.1109/MMUL.2019.2960219.

12. GUIZZO E., WEYDE T., LEVESON J.B. (2020), Multi-time-scale convolution for emotion recognition from speech audio signals, [in:] *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, doi: 10.1109/ICASSP40776.2020.9053727.

13. HAQ S., JACKSON P.J.B. (2011), Multimodal emotion recognition, [in:] *Machine Audition: Principles, Algorithms and Systems*, Wang W. [Ed.], pp. 398–423, IGI Global Press, Hershey, doi: 10.4018/978-1-61520-919-4.

14. JADHAV S., HE H., JENKINS K. (2018), Information gain directed genetic algorithm wrapper feature selection for credit rating, *Applied Soft Computing*, **69**: 541–553, doi: 10.1016/j.asoc.2018.04.033.

15. JAIN D.K., SHAMSOLMOALI P., SEHDEV P. (2019), Extended deep neural network for facial emotion recognition, *Pattern Recognition Letters*, **120**: 69–74, doi: 10.1016/j.patrec.2019.01.008.

16. KANWAL S., ASGHAR S. (2021), Speech emotion recognition using clustering based GA-optimized feature set, *IEEE Access*, **9**: 125830–125842, doi: 10.1109/ACCESS.2021.3111659.

17. KAYA H., KARPOV A.A. (2018), Efficient and effective strategies for cross-corpus acoustic emotion, *Neurocomputing*, **275**: 1028–1034, doi: 10.1016/j.neucom.2017.09.049.

18. KHALIL R.A., JONES E., BABAR M.I., JAN T., ZAFAR M.H., ALHUSSAIN T. (2019), Speech emotion recognition using deep learning techniques: A review, *IEEE Access*, **7**: 117327–117345, doi: 10.1109/ACCESS.2019.2936124.

19. KREYSZIG E. (1979), *Advanced Engineering Mathematics*, 4th ed., Wiley.

20. LIN Y.-L., WEI G. (2005), Speech emotion recognition based on HMM and SVM, [in:] *Fourth International Conference on Machine Learning and Cybernetics*, doi: 10.1109/ICMLC.2005.1527805.

21. LIU Z.-T., XIE Q., WU M., CAO W.-H., MEI Y., MAO J.-W. (2018), Speech emotion recognition based on an improved brain emotion learning model, *Neurocomputing*, **309**: 145–156, doi: 10.1016/j.neucom.2018.05.005.

22. LIVINGSTONE S.R., RUSSO F.A. (2018), The Ryerson audio-visual database of emotional speech and

song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English, *PLOS ONE*, **13**(5): e0196391, doi: 10.1371/journal.pone.0196391.

23. Majkowski A., Kołodziej M., Rak R.J., Korczyński R. (2016), Classification of emotions from speech signal, [in:] *Signal Processing Algorithms, Architectures, Arrangements and Applications* (*SPA*), doi: 10.1109/SPA.2016.7763627.

24. Martin O., Kotsia I., Macq B., Pitas I. (2006), The eNTERFACE'05 audio-visual emotion database, [in:] *22nd International Conference on Data Engineering Workshops* (*ICDEW'06*), doi: 10.1109/ICDEW.2006.145.

25. McFee B. *et al.* (2015), librosa: Audio and music signal analysis in Python, [in:] *14th Python in Science Conference*, pp. 18–25, doi: 10.25080/Majora-7b98e3ed-003.

26. Milner R., Jalal M.A., Ng R.W.N.M., Hain T. (2019), A cross-corpus study on speech emotion recognition, [in:] *IEEE Automatic Speech Recognition and Understanding Workshop* (*ASRU*), doi: 10.1109/ASRU46091.2019.9003838.

27. Mitchel M. (1996), *An Introduction to Genetic Algorithms*, MIT Press, Cambridge.

28. Mohino-Herranz I., Gil-Pita R., Alonso-Diaz S., Rosa-Zurera M. (2014), MFCC based enlargement of the training set for emotion recognition in speech, *Signal & Image Processing: An International Journal* (*SIPIJ*), **5**(1), doi: 10.48550/arXiv.1403.4777.

29. Monaco A., Amoroso N., Bellantuono L., Pantaleo E., Tangaro S., Bellotti R. (2020), Multi-time-scale features for accurate respiratory sound classification, *Applied Sciences*, **10**(23): 8606, doi: 10.3390/app10238606.

30. Omman B., Eldho S.M. (2022), Speech emotion recognition using bagged support vector machines, [in:] *International Conference on Computing, Communication, Security and Intelligent Systems* (*IC3SIS*), doi: 10.1109/IC3SIS54991.2022.9885578.

31. Pandey S.K., Shekhawat H.S., Prasanna M.S. (2019), Deep learning techniques for speech emotion recognition: A review, [in:] *29th International Conference Radioelektronika* (*RADIOELEKTRONIKA*), doi: 10.1109/RADIOELEK.2019.8733432.

32. Pedregosa F. *et al.* (2011), Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, **12**(85): 2825–2830.

33. Picard R.W. (1995), *Affective computing*, M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321.

34. Pichora-Fuller M.K., Dupuis K. (2020), Toronto emotional speech set (TESS) (V1), *University of Toronto Dataverse*, doi: 10.5683/SP2/E8H2MF.

35. Sahoo S., Routray A. (2016), MFCC feature with optimized frequency range: An essential step for emotion recognition, [in:] *2016 International Conference on Systems in Medicine and Biology* (*ICSMB*), doi: 10.1109/ICSMB.2016.7915112.

36. Sayed S., Nassef M., Badr A., Farag I. (2019), A Nested Genetic Algorithm for feature selection in high-dimensional cancer microarray datasets, *Expert Systems with Applications*, **121**: 233–243, doi: 10.1016/j.eswa.2018.12.022.

37. Seknedy M.E., Fawzi S. (2021), Speech emotion recognition system for human interaction applications, [in:] *Tenth International Conference on Intelligent Computing and Information Systems* (*ICICIS*), doi: 10.1109/ICICIS52592.2021.9694246.

38. Seknedy M.E., Fawzi S. (2022), Speech emotion recognition system for Arabic speakers, [in:] *4th Novel Intelligent and Leading Emerging Sciences Conference* (*NILES*), doi: 10.1109/NILES56402.2022.9942431.

39. Shahin I. (2020), Emotion recognition using speaker cues, [in:] *Advances in Science and Engineering Technology International Conferences* (*ASET*), pp. 1–5, doi: 10.48550/arXiv.2002.03566.

40. Sidorov M., Brester C., Minker W., Semenkin E. (2014), Speech-based emotion recognition: Feature selection by self-adaptive multi-criteria genetic algorithm, [in:] *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (*LREC'14*), pp. 3481–3485, http://www.lrec-conf.org/proceedings/lrec2014/pdf/341_Paper.pdf (access: 10.11.2023).

41. Slaney M. (1998), *Auditory Toolbox: A MATLAB Toolbox for Audtiory Modeling Work*, version 2, Interval Research Corporation.

42. Stefanowska A., Zieliński S.K. (2023), Software repository for speech emotion recognition using a multi-time-scale approach to feature aggregation and an ensemble of SVM classifiers, GitHub, https://github.com/antoninastefanowska/MTS-SVM-EmotionRecognition (access: 27.10.2023).

43. Su B.-H., Lee C.-C. (2021), A conditional cycle emotion gan for cross corpus speech emotion recognition, [in:] *IEEE Spoken Language Technology Workshop* (*SLT*), doi: 10.1109/SLT48900.2021.9383512.

44. Tamulevičius G., Korvel G., Yayak A.B., Treigys P., Bernatavičienė J., Kostek B. (2020), A study of cross-linguistic speech emotion recognition based on 2D feature spaces, *Electronics*, **9**(10): 1725, doi: 10.3390/electronics9101725.

45. Tang D., Kuppens P., Geurts L., van Waterschoot T. (2021), End-to-end speech emotion recognition using a novel context-stacking dilated convolution neural network, *EURASIP Journal on Audio, Speech and Music Processing*, **18**(2021), doi: 10.1186/s13636-021-00208-5.

46. Tao H., Shan S., Hu Z., Zhu C., Ge H. (2023), Strong generalized speech emotion recognition based on ef-

fective data augmentation, *Entropy*, **25**(1): 68, doi: 10.3390/e25010068.

47. TZIRAKIS P., ZHANG J., SCHULLER B.W. (2018), End-to-end speech emotion recognition using deep neural networks, [in:] *IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), doi: 10.1109/ICASSP.2018.8462677.

48. WANG Y., HUO H. (2019), Speech recognition based on genetic algorithm optimized support vector machine, [in:] *6th International Conference on Systems and Informatics* (*ICSAI*), doi: 10.1109/ICSAI48974.2019.9010502.

49. YANG K. *et al.* (2023), Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition, [in:] *IEEE Transactions on Affective Computing*, **14**(2): 1082–1097, doi: 10.1109/TAFFC.2021.3100868.

50. YILDIRIM S., KAYA Y., KILIÇ F. (2021), A modified feature selection method based on metaheuristic algorithms for speech emotion recognition, *Applied Acoustics*, **173**: 107721, doi: 10.1016/j.apacoust.2020.107721.

51. YOUNG S. *et al.* (2006), *The HTK Book*, Cambridge University Engineering Department.

52. ZACHARATOS H., GATZOULIS C., CHARALAMBOUS P., CHRYSANTHOU Y. (2021), Emotion recognition from 3D motion capture data using deep CNNs, [in:] *IEEE Conference on Games* (*CoG*), doi: 10.1109/CoG52621.2021.9619065.

53. ZHANG S., CHEN A., GUO W., CUI Y., ZHAO X., LIU L. (2020), Learning deep binaural representations with deep convolutional neural networks for spontaneous speech emotion recognition, *IEEE Access*, **8**: 23496–23505, doi: 10.1109/ACCESS.2020.2969032.

54. ZHAO J., MAO X., CHEN L. (2018), Learning deep features to recognise speech emotion using merged deep CNN, *IET Signal Process*, **12**(6): 713–721, doi: 10.1049/iet-spr.2017.0320.

55. ZHAO J., MAO X., CHEN L. (2019), Speech emotion recognition using deep 1D & 2D CNN LSTM networks, *Biomedical Signal Processing and Control*, **47**: 312–323, doi: 10.1016/j.bspc.2018.08.035.

Research Paper

Snoring Sound Recognition Using Multi-Channel Spectrograms

Ziqiang YE[(1)], Jianxin PENG[(1)]*, Xiaowen ZHANG[(2)], Lijuan SONG[(2)]

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

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

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

Obstructive sleep apnea-hypopnea syndrome (OSAHS) is a common and high-risk sleep-related breathing disorder. Snoring detection is a simple and non-invasive method. In many studies, the feature maps are obtained by applying a short-time Fourier transform (STFT) and feeding the model with single-channel input tensors. However, this approach may limit the potential of convolutional networks to learn diverse representations of snore signals. This paper proposes a snoring sound detection algorithm using a multi-channel spectrogram and convolutional neural network (CNN). The sleep recordings from 30 subjects at the hospital were collected, and four different feature maps were extracted from them as model input, including spectrogram, Mel-spectrogram, continuous wavelet transform (CWT), and multi-channel spectrogram composed of the three single-channel maps. Three methods of data set partitioning are used to evaluate the performance of feature maps. The proposed feature maps were compared through the training set and test set of independent subjects by using a CNN model. The results show that the accuracy of the multi-channel spectrogram reaches 94.18%, surpassing that of the Mel-spectrogram that exhibits the best performance among the single-channel spectrograms. This study optimizes the system in the feature extraction stage to adapt to the superior feature learning capability of the deep learning model, providing a more effective feature map for snoring detection.

Keywords: obstructive sleep apnea-hypopnea syndrome; snoring; convolutional neural network; multi-channel spectrogram.

1. Introduction

Obstructive sleep apnea-hypopnea syndrome (OS-AHS) is a sleep respiratory disease characterized by the repeated collapse and blockage of the upper airway during sleep, resulting in apnea or hypopnea (STROLLO, ROGERS, 1996). Obstructive breathing leads to instinctive body responses, such as brain arousal, sympathetic activation, and decreased blood oxygen saturation. Seriously interrupted and non-restorative sleep will occur, causing most patients with OSAHS to suffer from morning headaches and daytime somnolence. Long-term poor sleep can even lead to a series of complications, such as abnormal metabolism, neurocognitive dysfunction, and cardio-vascular disease (YOUNG *et al.*, 2002). Surveys show that the overall prevalence of OSAHS in the general adult population ranges from 6 to 17%, with the prevalence increasing significantly with age (SENARATNA *et al.*, 2017).

Polysomnography (PSG) is the gold standard for diagnosing OSAHS patients (AHMADI *et al.*, 2009; MENDONÇA *et al.*, 2019). Subjects are required to wear contact-type monitoring instruments throughout the night. The PSG signal obtained from these instruments is used by professional doctors to determine whether the subjects suffer from OSAHS. Although reliable results can be obtained, patients may have to bear the burden of expensive fees and endure discomfort from physically attached sensors (MENDONÇA *et al.*, 2019).

Therefore, there is an urgent need to seek a low-cost, easy-to-operate, and non-contact method to assist in the diagnosis of OSAHS. Snoring is the most distinctive clinical feature of OSAHS, occurring in 70–90% of patients with OSAHS (Karunajeewa *et al.*, 2008; Maimon, Hanly, 2010). The acoustic characteristics of snoring reflect changes in the structure of the upper airway. Moreover, snoring analysis offers the advantages of being non-contact, simple, and reliable, making it feasible to identify patients by analyzing the acoustic characteristics of snoring (Won *et al.*, 2012; Fiz *et al.*, 1996; Pevernagie *et al.*, 2010; Beck *et al.*, 1995; Ip *et al.*, 2002; Perez-Padilla *et al.*, 1993; Sola-Soler *et al.*, 2003; Ng *et al.*, 2008).

In order to improve the initial screening of OSAHS, an increasing number of scientists are dedicated to developing new technologies that can achieve a more accurate clinical diagnosis of OSAHS in a simpler manner (Yadollahi, Moussavi, 2010; Ankışhan, Arı, 2011; Ankışhan, Yilmaz, 2013). So far, there have been numerous studies on the identification technology of OSAHS. Duckitt *et al.* (2006) extracted 39-dimensional Mel-frequency cepstral coefficients (MFCC) from sleep sound recordings of six subjects and classified the signals into snoring, breathing, duvet noise, and other noises based on hidden Markov model (HMM). The recognition rate for snoring can reach the range of 82–89%. Cavusoglu *et al.* (2007) selected recording signals from 18 simple snorers and 12 OSAHS patients to cut the voiced segments by a double threshold method. Then, the authors calculated the sub-band energy distribution of the sound segments and used principal component analysis (PCA) for feature reduction. Finally, robust linear regression was used to classify these sound segments into snoring and non-snoring sounds with an accuracy of 90.2%.

Dafna *et al.* (2013) adopted a feature selection algorithm to filter the 34 most discriminative features from 127 time-domain and frequency-domain features, and then used AdaBoost to construct a snoring recognition model, obtaining an average detection rate of 98.2%, a sensitivity of 98%, and specificity of 98.3% with a cross-validation method. In a study by Cheng *et al.* (2022), a multi-input model based on long short-term memory (LSTM) was proposed, which can accept various audio features to synthesize information for snoring identification. Furthermore, MFCC, Mel filter banks (Fbanks), linear prediction coefficient (LPC), and short-term energy were extracted as the input of the model, finally achieving 95.3% accuracy. With the development of the field of artificial intelligence, deep learning models are gradually applied to the classification task of snoring and non-snoring.

Khan (2019) developed a deep learning model for snoring detection and transferred it to an embedded system that can be connected to a smartphone app using home Wi-Fi. In Khan's study, 1000 sound samples were used to calculate the MFCC images, then the images were fed into a convolutional neural network (CNN) model, resulting in a snoring recognition rate of 96%. The spectrogram, Mel-spectrogram, and constant-Q transformation (CQT) spectrogram collected from the recordings of 15 subjects were used to classify snoring and non-snoring by Jiang *et al.* (2020). The results indicated that the accuracy of Mel-spectrogram in each group reached 95.07%. The advantage of the deep learning model is to learn increasingly complex data samples. Previous studies (Khan, 2019; Jiang *et al.*, 2020; Xie *et al.*, 2021) used single-channel spectrogram as input. However, it is important to note that different feature maps only contain limited frequency-domain information, which could potentially restrict the model's ability to learn diverse representations of audio recordings. Therefore, input features should provide more information about snoring.

In our work, a multi-channel feature map based on the fusion of Mel-spectrogram, spectrogram, and continuous wavelet transform (CWT) is proposed. Three spectrograms of each sound signal are employed as three channels of the red-green-blue (RGB) image to construct the feature map. A CNN model is utilized to perform the classification tasks. In addition, spectrogram, Mel-spectrogram, and CWT are used for comparative experiments. The comparison of the classification performance between the multi-channel spectrogram with that of the single-channel spectrogram is conducted to achieve higher resolution.

2. Methods

2.1. Data acquisition

This study was approved by the Ethics Committee of Guangzhou Medical University (Reference Number 2019-73), and informed consent was obtained from all participants.

Thirty subjects who underwent PSG at the First Affiliated Hospital of Guangzhou Medical University were selected to obtain snoring sounds throughout the night. The recording time for each subject's sleep snoring sounds was not less than 6 hours. The most important indicator for PSG detection to assess the severity of OSAHS is the apnea-hypopnea index (AHI), which is defined as the average number of sleep apnea or hypopnea per hour. It is divided into four categories: simple, mild, moderate, and severe, based on the following ranges: $AHI < 5$, $5 \le AHI < 15$, $15 \le AHI < 30$, and $AHI \ge 30$ (Maimon, Hanly, 2010). Table 1 lists statistical information on the subjects' gender, age, body mass index (BMI), AHI, and the severity of OSAHS for each participant. For recording snoring sounds, a digital audio recorder (Roland, Edirol R-44, Japan), with a frequency response range of 40–20 000 Hz and a microphone (RODE, NTG-3, Sydney, Australia) hanging

Table 1. Statistical information of subjects.

Parameter	Data
Male/female	27/3
Age (years)	44 ± 13 (range: 23–70)
BMI [kg/m^2]	26.7 ± 2.8 (range: 20.8–31.9)
AHI [events/h]	40.8 ± 28.3 (range: 3.2–91.1)
OSAHS [y/n]	28/2

vertically on the heads of patients, positioned about 45 cm above the mouth and nose were used. The original sleep sound signals were recorded by the microphone. PSG device (Alice-5, Pittsburgh, Pennsylvania, USA) was used to monitor PSG signals. The recorded sound was digitized at a sampling rate of 44 100 Hz and a resolution of 16 bits.

2.2. Feature extraction

2.2.1. Spectrogram

A snoring sound is a one-dimensional time-domain signal, making it challenging to observe the frequency conversion pattern. While the frequency distribution of the signal can be viewed by Fourier transform, time-domain information is lost. Many time-frequency analysis methods have emerged to address this problem. Short-time Fourier transform (STFT) is the most classical time-frequency analysis method in speech and audio processing applications and offers minimal calculation and low cost. First, the audio signal is framed into a short time window. In this work, the size of windows is 25 ms with 50% overlap. Next, the Hamming window is applied to each frame signal, and followed by the fast Fourier transform (FFT) to obtain its power spectrum (RABINER *et al.*, 1975). Each frame is then spliced along the time dimension to form a two-dimensional signal map called the spectrogram.

2.2.2. Mel-spectrogram

While the frequency of the spectrogram is linearly distributed, the extracted features may not be useful for signals with an inhomogeneous frequency distribution. The Mel-scale filter banks are used to transform the spectrogram into the Mel-spectrogram (PENG *et al.*, 2019; WINURSITO *et al.*, 2018), where the Mel-scale describes the nonlinear characteristics of human ear frequency, and its relationship with frequency can be approximately expressed by the equation:

$$\mathrm{Mel}(f) = 2595 \times \log\left(1 + \frac{f}{700}\right). \qquad (1)$$

In this study, features are calculated using frames of 25 ms frame size with 50% overlap. The Mel-spectrogram is computed using a group of 128 triangular filters in the Mel-scale based on the STFT, and the logarithm of the filtered signal is determined. Figure 1 shows the triangular filter banks used in this study.
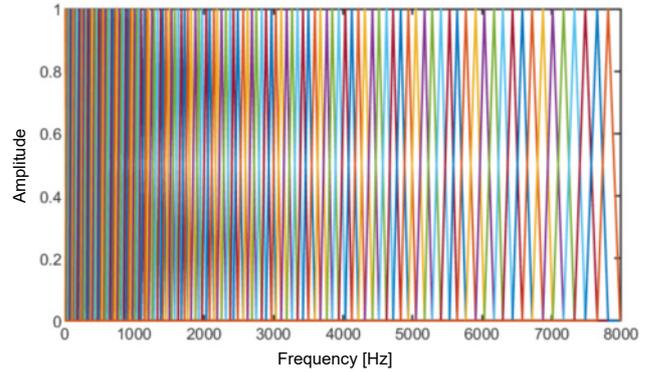
Fig. 1. 128 triangular filters in the Mel-scale applied to the STFT for obtaining the Mel-spectrogram.

2.2.3. Continuous wavelet transform

The time and frequency resolutions of STFT are determined by the size and time shift of the window. A small window size can lead to poorer frequency resolution. Compared to STFT, CWT has the characteristics of window adaptation, enabling high-frequency values to have high-frequency resolution and low time resolution (QIAN *et al.*, 2019).

CWT uses wavelet basis functions to decompose signals, and is defined as:

$$\mathrm{CWT}(\tau s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t)\psi\left(\frac{t-\tau}{s}\right) \mathrm{d}t, \qquad (2)$$

where $x(t)$ is the audio signal, $\psi(x)$ is the mother wavelet (Morlet wavelet in this study), and τ and s, respectively, represent displacement and scale.

Usually, when analyzing time series, it is expected to obtain smooth and continuous wavelet amplitude, so a non-orthogonal wavelet function is more suitable. In addition, to include the information of both amplitude and phase of the time series, a complex-valued wavelet should be selected, because the complex-valued wavelet has an imaginary part and can express the phase very well. The Morlet wavelet is not only non-orthogonal, but also exponential complex-valued wavelet, so it is used in this experiment to obtain the information of both amplitude and phase.

2.2.4. Multi-channel spectrogram

Multi-channel spectrogram has been used in speech recognition with beneficial effects (ADAVANNE *et al.*, 2018; XU *et al.*, 2018; ARIAS-VERGARA *et al.*, 2021).

The spectrogram, Mel-spectrogram, and CWT, each with a size of $224 \times 224 \times 3$, were extracted from each audio segment. Figure 2 shows the above three feature maps of a snore signal. Subsequently, they are normalized to fall between −1 and 1, serving as three channels of the RGB image to construct the multi-channel spectrogram with a size of $224 \times 224 \times 3$. In this construction, the spectrogram is the first chan-

a)

b)

c)
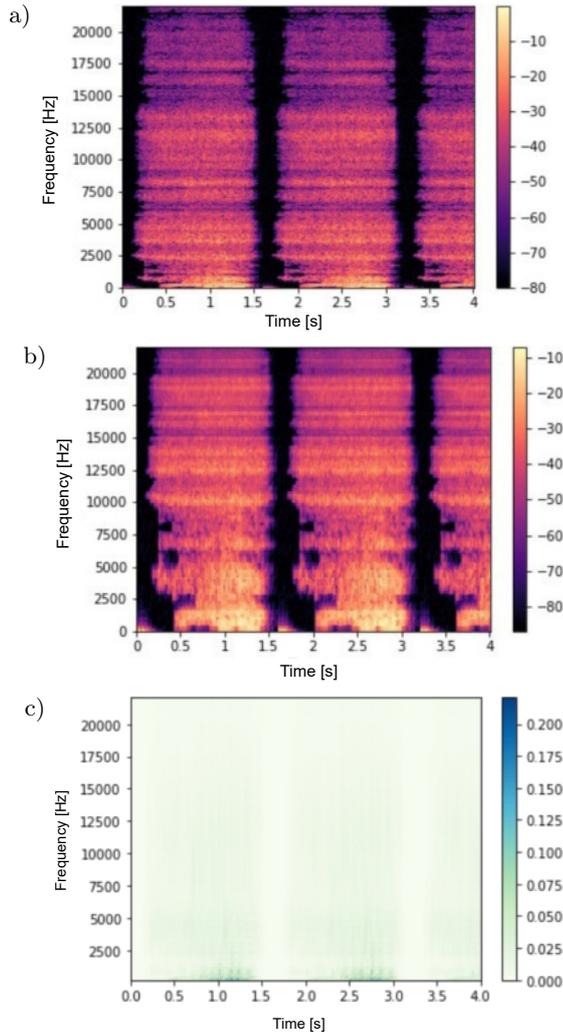
Fig. 2. Feature maps of a snore segment from an OSAHS patient: a) spectrogram; b) Mel-spectrogram; c) CWT.

nel, the CWT is the second channel, and the Mel-spectrogram is the third channel. When the input data contains multiple channels, the number of input channels of the convolutional kernel in the model is the same as that of the input data. In this way, the convolutional kernel of different channels can perform cross-correlation operations with the input data of different channels, and the multi-channel input will enable CNN to supplement information from two other time-frequency representations.

2.3. Model architecture

In order to obtain reasonable results, the classifier must be matched with a suitable input representation. Manual features such as MFCC were used with the traditional machine learning model, which effectively decorrelates features (ADAVANNE *et al.*, 2018). On the contrary, the advantage of CNN lies in their ability to learn spectral time characteristics of the spectrum through weight sharing and pooling technology. Previous studies have applied CNN to speech recognition with good effects (ABDEL-HAMID *et al.*, 2012; 2014). For this experiment, a CNN model was designed, containing an input layer, three convolution layers with rectified linear unit (ReLu) activation functions. The size of the convolution kernel was multiplied layer by layer, leading to 256 neurons activated by ReLu, and the output layer was activated by a softmax function. The incorporated dropout layer will randomly discard some weights in the training process to suppress overfitting, and the dropout ratio is 0.5 (HINTON *et al.*, 2012). Figure 3 shows the process of feeding the multi-channel spectrogram into the CNN. The model parameters are presented in Table 2.
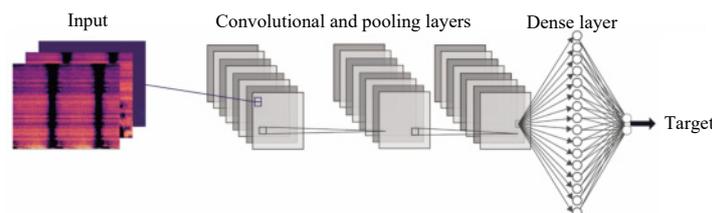
Fig. 3. Process of feeding the multi-spectrogram to a deep learning model (CNN).

Table 2. Structure of CNN.

Layer (type)	Input shape	Output shape	Params
Conv2D	(None, 224, 224, 32)	(None, 222, 222, 32)	896
MaxPooling2D	(None, 222, 222, 32)	(None, 111, 111, 32)	0
Conv2D	(None, 111, 111, 32)	(None, 109, 109, 64)	18 496
MaxPooling2D	(None, 109, 109, 64)	(None, 54, 54, 64)	0
Conv2D	(None, 54, 54, 64)	(None, 52, 52, 128)	73 856
MaxPooling2D	(None, 52, 52, 128)	(None, 26, 26, 128)	0
Flatten	(None, 26, 26, 128)	(None, 86 528)	0
Dense	(None, 86 528)	(None, 128)	11 075 712
Dense	(None, 128)	(None, 2)	258

For excellent training results, the Adam optimizer is used for training, with a learning rate of CNN set to 0.0001. In our experiments, categorical cross-entropy was chosen as the loss function, and each model was trained for 200 epochs on an NVIDIA GTX 1080Ti with a batch size of 128.

2.4. Validation method

In this study, the adaptive threshold method is used to segment the audio sounds from all recording subjects to obtain sound fragments that are subsequently labeled as either snoring or non-snoring under the guidance of ear-nose-throat (ENT) experts. Only sound segments less than 4 seconds long are retained, and two adjacent sound segments less than 0.02 seconds apart are merged. A total of 59 293 sound segments are obtained, consisting of 29 789 snore segments, and 29 504 non-snoring segments, which included sounds of footsteps, speech, breathing, coughing, door closing, and other environmental sounds. In order to evaluate the performance of different spectra, three experiments were designed: independent split training set and test set, leave-one-subject-out cross-validation (LOSOCV), and training set and test set containing all subjects. Table 3 shows the details of the data partition.

Experiment 1: the dataset of 30 subjects was divided into a validation set with 4 subjects, a test set with 4 subjects, and training set with the remaining 22 subjects, and the subjects in the training set, the test set, and the validation set were independent. For the purpose of eliminating the contingency of the experiment, five different partition methods were applied to the data set, and the model was trained on each divided dataset. Finally, the average and standard deviation were taken as the results.

Experiment 2: in a dataset containing 30 subjects, an independent test set and training set were constructed for each participant using the LOSOCV strategy. The data of one subject was selected as the test set, and the data of the remaining 29 subjects were used as the training set. This process is repeated 30 times and the average accuracy is calculated. This maximizes the use of data while ensuring that the subjects in the training set and the test set are from different independent subjects.

Experiment 3: the sound clips of all subjects are combined into a whole dataset, which is then divided into training, validation and test set, with a ratio of 6:1:3.

2.5. Model evaluation

The classification effect of each feature map can be evaluated by multiple indicators, including accuracy, precision, recall, $F1$-score, and the area under the curve (AUC) calculated from the receiver operating characteristic (ROC). Accuracy is the proportion of correct samples to the total number of samples. Precision relates to the ratio of the number of positive samples correctly classified by the classifier to the number of all positive samples classified by the classifier. Recall rate refers to the ratio of the number of positive samples correctly classified by the classifier to the number of all positive samples. $F1$-score is the harmonic mean of precision rate and recall rate. The AUC is meant by the area under the ROC curve, representing the probability that the predicted positive cases rank higher than the negative ones, ranging from 0.5 to 1. The calculation equation is:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (3)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (4)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (5)$$

$$F1_{\text{score}} = \frac{2\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (6)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

3. Results

To evaluate the classification performance, four different feature maps are imported into the model to compare which feature map is more discriminative for snoring. The CNN model is established by the validation set and evaluated on the test set. According to the data set division rules of experiment 1, the classification results are presented in Table 4. In terms of single-channel spectrograms, the classification performance of Mel-spectrogram was superior to those of spectrogram and CWT, with an accuracy of 91.58%, precision of 92.09%, sensitivity of 86.57%, $F1$-score of 88.85%, and AUC of 0.9614. The PPV of the spectrogram and Mel-spectrogram reached more than 90%,

Table 3. Data distribution of training, validation, and test sets in experiments.

	Experiment 1			Experiment 2		Experiment 3		
	Train	Validation	Test	Train	Test	Train	Validation	Test
Subject	22	4	4	29	1		30	
Snore	23 767	3117	2905	LOSOCV		19 133	2872	7784
No-snore	21 971	4094	3439			16 443	3057	10 004

Table 4. Classification results of experiment 1.

Map type	Accuracy [%]	Precision [%]	Recall [%]	F1-score [%]	AUC
Mel-spectrogram	91.58	92.09	86.57	88.85	0.9614
Spectrogram	88.33	91.23	81.56	85.56	0.9448
CWT	85.24	81.78	85.10	83.00	0.9192
Multi-channel spectrogram	94.16	92.64	93.35	92.93	0.9730

indicating that the recognition of the snore fragments was reliable.

Figure 4 shows that the classification of the multi-channel spectrogram is significantly improved compared to that of the single-channel spectrogram, and it has an accuracy of 94.16%, which was 2.58% higher than that of Mel-spectrogram with the best effect in single-channel spectrograms. Other classification indexes were increased, respectively, by 0.55% (PPV), 6.78% (Recall), and 4.08% (F1-score). Although there was little difference in PPV between the two feature maps, the Recall of the multi-channel spectrogram classification was significantly higher than that of the Mel-spectrogram, which is beneficial for detecting the snoring segments of the patients throughout the entire night and further evaluating the severity of OSAHS patients.
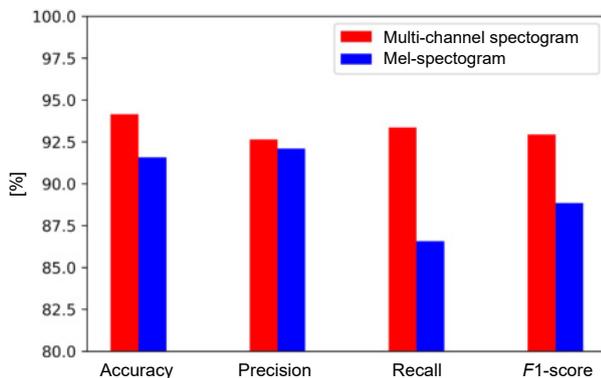
Fig. 4. Comparison histogram of Mel-spectrogram and multi-spectrogram in experiment 1.

Tables 5 and 6 show the classification results for experiments 2 and 3. The results show that the recognition effect of the multi-channel spectrogram is consistently better than that of the single-channel spectrogram when using different dataset partitioning methods.

4. Discussion

In this study, the performance of Mel-spectrogram, spectrogram, CWT, and multi-channel spectrogram in classifying snoring and the non-snoring sound was investigated. The results show that the Mel-spectrogram has the best recognition effect when the single-channel spectrograms were used as input, which is in agreement with the results of the study by JIANG (2020). The energy peak frequency of the snoring sounds mentioned in the study is 250 Hz, and most of the energy is distributed below 1000 Hz, while the energy of respiratory sounds and other noise is distributed above 1000 Hz (PEVERNAGIE *et al.*, 2010; JIANG *et al.*, 2020). The frequency of the spectrogram is linear distribution, which leads to the insufficient frequency resolution in the low-frequency part, making it challenging to detect some weak snoring changes. The Mel-spectrogram converts the linear frequency into the Mel frequency, offering detailed representation of the low-frequency information and rough representation of the high-frequency information, which aligns with the energy distribution of the snoring spectrogram.

Apart from Spectrogram and Mel-spectrogram, which are computed based on STFT, the CWT commonly

Table 5. Classification results of experiment 2.

Map type	Accuracy [%]	Precision [%]	Recall [%]	F1-score [%]	AUC
Mel-spectrogram	90.51	90.83	93.08	91.94	0.9511
Spectrogram	89.36	93.18	88.85	90.97	0.9599
CWT	85.38	89.51	84.82	87.10	0.9191
Multi-channel spectrogram	93.10	92.28	96.66	94.42	0.9774

Table 6. Classification results of experiment 3.

Map type	Accuracy [%]	Precision [%]	Recall [%]	F1-score [%]	AUC
Mel-spectrogram	93.67	98.28	91.44	94.74	0.9817
Spectrogram	91.76	93.03	93.34	93.19	0.9717
CWT	88.99	91.74	89.84	90.78	0.9569
Multi-channel spectrogram	97.80	97.14	99.18	98.15	0.9979

used in speech recognition is also imported into the same CNN model. A study by HUZAIFAH (2017) proved that CWT performs significantly worse than spectrogram and Mel-spectrogram when employed in a CNN to classify various environmental sounds. The same result was obtained when the three feature maps were applied to snoring and non-snoring sound classification. It means that CWT cannot provide more snoring sounds details in the low frequency compared to the other two maps. However, it is premature to conclude that CWT is always inferior to the feature maps based on STFT, because the experiment may be influenced by parameter settings for map extraction and model structure.

It should be pointed out that the peak energy frequency of snoring sound among different people is not consistent, and even the snoring of the same person is different. JIANG et al. (2020) analyzed the energy distributions in snoring and non-snoring sub-bands of subjects and found that 60% of the snoring spectral energy was distributed between 100 and 300 Hz, and 40% of it was also distributed in each frequency band above 300 Hz. The information contained in a single-channel input may be restricted, which can limit the potential of the deep learning model to learn more complicated representations from snoring sound signals. The multi-channel map was used to overcome the limitation of a single-channel input in speech recognition. Various methods were used to construct multi-channel maps in such studies. ADAVANNE et al. (2018) proposed a method where multi-channel could be extracted from the same signal recorded by different microphones. Another approach by FU et al. (2017) involved computing the real and imaginary parts of the STFT to form a 2D-channel spectrogram.

ARIAS-VERGARA et al. (2021) computed CWT, Mel-spectrogram, and gammatone spectrogram from the audio signal and combined them into a 3D-channel spectrogram. Compared with single-channel maps, the performance of these multi-channel maps with a CNN model was improved. In our work, when a multi-channel spectrogram was used as the model input to identify snoring sounds, the result was consistent with the expectation, which was better than the Mel-spectrogram with the best classification effect of single-channel feature maps. This suggests that the multi-channel spectrogram contains more spectrum information than a single spectrum. The CNN model can capture more feature information from the fusion map than from a single-channel feature map through multi-layer convolution layers.

Many researchers have proposed a variety of experimental methods to classify snoring and non-snoring. Table 7 compares the research methods in related fields with the current experiment. KHAN (2019) collected online snoring resources as datasets, extracted MFCC images, and input them into a CNN model training and obtained a 96% accuracy. However, the number of experimental samples was only 1000, and the source of snoring sound was singular. In our experiment, 59 293 sound samples were extracted from 30 subjects with better generalization ability, and three different verification methods were used to evaluate the performance of the feature map, resulting in the generalization of the results. JIANG et al. (2020) used two classifiers, CNNs-DNNs and CNNs-LSTMs-DNNs, to identify snores from sound fragments, including spectrogram, Mel-spectrogram, and CQT-spectrogram. The results demonstrate that the combination of Mel spectrogram and CNNs-LSTMs-DNNs was well suited for the task. However, the input images contained limited information from single-channel spectrogram. Moreover, the data of the training set and the test set are not independent and using this model to detect individual snore fragments throughout entire night may lead to deviation. CHENG et al. (2022) designed a multi-input

Table 7. Summary of previous studies on snoring detection.

Author	Subjects	Datasets	Features	Methods	Result [%]
KHAN (2019)		1000	MFCC image	CNN	Accuracy: 96
JIANG et al. (2020)	15	12 457	Mel-spectrogram	CNN+LSTM+DNN	Accuracy: 95.07 PPV: 94.62 Sensitivity: 95.42
CHENG et al. (2022)	43	15 520	MFCC, Fbanks, Short-time average energy, LPC	A multi-input model based on LSTM	Accuracy: 95.3 PPV: 95.7 Sensitivity: 94.9
DAFNA et al. (2013)	67	281 953	Time-related features, Spectral-related features	AdaBoost	Accuracy: 98.2 Sensitivity: 98.1
CAVUSOGLU et al. (2007)	30	9000	Average normalized energy in each subband	Robust linear regression	Accuracy: 90.2 PPV: 98.7
SUN et al. (2022)	24	36 938	Bark sub-band feature, MFCC, LPC, etc.	XGBoost	Accuracy: 87.22 PPV: 95.09 Sensitivity: 87.16
This work	30	59 293	Multi-spectrogram	CNN	Accuracy: 94.16 PPV: 92.64 Sensitivity: 93.35

model based on LSTM and extracted MFCC, Fbanks, short-term energy, and LPC as four branches of the input layer. After integration, ANN was used as the classifier, and finally, a 95.3% snoring recognition rate was obtained, an improvement compared with a single feature processing network. Nevertheless, the model's input layer has multiple parallel input branches, and the network structure is relatively complex.

In their experiment, the fusion feature maps were employed in feature extraction, and only one entry was needed for model input. In Dafna *et al.* (2013), 127 features from both the time domain and frequency domain were extracted. Using a feature selection method, 34 most effective features were selected objectively, and the AdaBoost classifier was used and yielded a 98.2% recognition rate. However, the extraction process involved various features, making the process of feature extraction complicated.

Cavusoglu *et al.* (2007) divided the frequency range of snoring sounds (0–7500 Hz) into 500 Hz sub-bands and calculated the average normalized energy in each sub-band to obtain spectral characteristics. The linear regression model was used and a 90.2% accuracy was obtained. However, the energy distribution of snoring was mainly concentrated in the low frequency and the band division of equal intervals may lead to insufficient low-frequency resolution. Sun *et al.* (2022) proposed a snoring detection algorithm based on acoustic features and XGBoost. Various training and test data splitting methods were used to evaluate model performance, and the results showed that when the training set and test set are from all subjects, the classification performance was better than that of the training set and test set from different independent subjects.

In terms of experimental accuracy, the method proposed in this work is significantly improved compared with 90.2% reported by Cavusoglu *et al.* (2007) and 92.78% obtained by Sun *et al.* (2022). However, it is important to acknowledge that different research samples are distinct, the subjective standards of labeled samples are different, and the methods of splitting data sets are also different. It is therefore difficult to compare the classification results to make a unified judgment. The multi-channel spectrogram proposed in this study has more than 92% in all evaluation indexes on the CNN model, indicating that this method can effectively detect snoring sound.

5. Conclusion

This study explored a classification method for distinguishing between snoring and non-snoring using a CNN model with a focus on a multi-channel spectrogram with a CNN model. Mel-spectrogram, spectrogram, and CWT were used as three channels for constructing multi-channel maps. The four feature maps of the snoring sound signals of 30 subjects were extracted for training and testing, and the results demonstrate that the classification performance indicators of the multi-channel spectrogram are improved compared with single-channel spectrograms. The main contribution of this work lies in proposing a multi-channel spectrogram based on the fusion of a single-channel spectrogram for snoring detection. The study also compared the classification performance of each feature map under the same network model.

This work focused on improving the feature extraction stage, extracting the feature maps containing more time and frequency domain information, to adapt to the strong fitting ability of the deep learning model. Future work can be carried out in different directions. Firstly, a comparison of diverse types of multi-channel spectrograms combined with various classification networks could be explored to further improve the accuracy of current snoring detection algorithms. Another direction is to explore how snoring sound detection contributes to the task of detecting OSAHS. This experiment can be used as the first step in OSAHS detection because snoring events are closely related to apnea. In addition, the snoring sound identified by this model could be further used to quantitatively evaluate the severity of OSAHS.

However, the snoring data collected in this experiment is limited to a hospital environment. Different recording environments have different background noise, which cannot guarantee the performance of the model in other recording settings. Therefore, more recording data in diverse environments (bedroom, dormitory, hotel, etc.) is needed to obtain a more reliable snoring recognition model and make it more robust and generalized. In addition, it is necessary to pay attention to the computational efficiency and memory overhead of the model to ensure that model meets the requirements for mobile deployment.

Author contributions

All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Informed consent

Informed consent was obtained from all individual participants who were included in the study.

Ethical approval

Consent was obtained from all study participants prior to their participation in the study. Ethical approval was obtained from the Ethics Committee of Guangzhou Medical University (Reference no. 2019-73).

References

1. ABDEL-HAMID O., MOHAMED A., JIANG H., DENG L., PENN G., YU D. (2014), Convolutional neural networks for speech recognition, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **22**(10): 1533–1545, doi: 10.1109/TASLP.2014.2339736.

2. ABDEL-HAMID O., MOHAMED A., JIANG H., PENN G. (2012), Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition, [in:] *2012 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4277–4280, doi: 10.1109/ICASSP.2012.6288864.

3. ADAVANNE S., POLITIS A., VIRTANEN T. (2018), Multi-channel sound event detection using 3D convolutional neural networks for learning inter-channel features, [in:] *2018 International Joint Conference on Neural Networks*, pp. 1–7, doi: 10.1109/IJCNN.2018.8489542.

4. AHMADI N., SHAPIRO G.K., CHUNG S.A., SHAPIRO C.M. (2009), Clinical diagnosis of sleep apnea based on single night of polysomnography vs. two nights of polysomnography, *Sleep Breath*, **13**(3): 221–226, doi: 10.1007/s11325-008-0234-2.

5. ANKIŞHAN H., ARI F. (2011), Snore-related sound classification based on time-domain features by using ANFIS model, [in:] *2011 International Symposium on Innovations in Intelligent Systems and Applications*, pp. 441–444, doi: 10.1109/INISTA.2011.5946113.

6. ANKIŞHAN H., YILMAZ D. (2013), Comparison of SVM and ANFIS for Snore related sounds classification by using the largest Lyapunov exponent and entropy, *Computational and Mathematical Methods in Medicine*, **2013**: 238937, doi: 10.1155/2013/238937.

7. ARIAS-VERGARA T., KLUMPP P., VASQUEZ-CORREA J.C., NÖTH E., OROZCO-ARROYAVE J.R., SCHUSTER M. (2021), Multi-channel spectrograms for speech processing applications using deep learning methods, *Pattern Analysis and Applications*, **24**(2): 423–431, doi: 10.1007/s10044-020-00921-5.

8. BECK R., ODEH M., OLIVEN A., GAVRIELY N. (1995), The acoustic properties of snores, *European Respiratory Journal*, **8**(12): 2120–2128, doi: 10.1183/09031936.95.08122120.

9. CAVUSOGLU M., KAMASAK M., EROGUL O., CILOGLU T., SERINAGAOGLU Y., AKCAM T. (2007), An efficient method for snore/nonsnore classification of sleep sounds, *Physiological Measurement*, **28**(8): 841–853, doi: 10.1088/0967-3334/28/8/007/.

10. CHENG S. *et al.* (2022), Automated sleep apnea detection in snoring signal using long short-term memory neural networks, *Biomedical Signal Processing and Control*, **71**(Part B): 103238, doi: 10.1016/j.bspc.2021.103238.

11. DAFNA E., TARASIUK A., ZIGEL Y. (2013), Automatic detection of whole night snoring events using noncontact microphone, *PLOS ONE*, **8**(12): e84139, doi: 10.1371/journal.pone.0084139.

12. DUCKITT W.D., TUOMI S.K., NIESLER T.R. (2006), Automatic detection, segmentation and assessment of snoring from ambient acoustic data, *Physiological Measurement*, **27**(10): 1047–1056, doi: 10.1088/0967-3334/27/10/010.

13. FIZ J.A. *et al.* (1996), Acoustic analysis of snoring sound in patients with simple snoring and obstructive sleep apnoea, *European Respiratory Journal*, **9**(11): 2365–2370, doi: 10.1183/09031936.96.09112365.

14. FU S., HU T., TSAO Y., LU X. (2017), Complex spectrogram enhancement by convolutional neural network with multi-metrics learning, [in:] *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing*, pp. 1–6, doi: 10.1109/MLSP.2017.8168119.

15. HINTON G.E., SRIVASTAVA N., KRIZHEVSKY A., SUTSKEVER I., SALAKHUTDINOV R.R. (2012), Improving neural networks by preventing co-adaptation of feature detectors, *ArXiv*, doi: 10.48550/arXiv.1207.0580.

16. HUZAIFAH M. (2017), Comparison of time-frequency representations for environmental sound classification using convolutional neural networks, *ArXiv*, doi: 10.48550/arXiv.1706.07156.

17. IP M.S., LAM B., NG M.M., LAM W.K., TSANG K.W., LAM K.S. (2002), Obstructive sleep apnea is independently associated with insulin resistance, *American Journal of Respiratory and Critical Care Medicine*, **165**(5): 670–676, doi: 10.1164/ajrccm.165.5.2103001.

18. JIANG Y., PENG J., ZHANG X. (2020), Automatic snoring sounds detection from sleep sounds based on deep learning, *Physical and Engineering Sciences in Medicine*, **43**(2): 679–689, doi: 10.1007/s13246-020-00876-1.

19. KARUNAJEEWA A.S., ABEYRATNE U.R., HUKINS C. (2008), Silence-breathing-snore classification from snore-related sounds, *Physiological Measurement*, **29**(2): 227–243, doi: 10.1088/0967-3334/29/2/006.

20. KHAN T. (2019), A deep learning model for snoring detection and vibration notification using a smart wearable gadget, *Electronics*, **8**(9): 987, doi: 10.3390/electronics8090987.

21. MAIMON N., HANLY P.J. (2010), Does snoring intensity correlate with the severity of obstructive sleep apnea?, *Journal of Clinical Sleep Medicine*, **6**(5): 475–478, doi: 10.5664/jcsm.27938.

22. MENDONÇA F., MOSTAFA S.S., RAVELO-GARCÍA A.G., MORGADO-DIAS F., PENZEL T. (2019), A review of obstructive sleep apnea detection approaches, *IEEE Journal of Biomedical and Health Informatics*, **23**(2): 825–837, doi: 10.1109/JBHI.2018.2823265.

23. Ng A.K., Koh T.S., Baey E., Lee T.H., Abeyratne U.R., Puvanendran K. (2008), Could formant frequencies of snore signals be an alternative means for the diagnosis of obstructive sleep apnea?, *Sleep Medicine*, **9**(8): 894–898, doi: 10.1016/j.sleep.2007.07.010.

24. Peng P., He Z., Wang L. (2019), Automatic classification of microseismic signals based on MFCC and GMM-HMM in underground mines, *Shock and Vibration*, **2019**: 5803184, doi: 10.1155/2019/5803184.

25. Perez-Padilla J.R., Slawinski E., Difrancesco L.M., Feige R.R., Remmers J.E., Whitelaw W.A. (1993), Characteristics of the snoring noise in patients with and without occlusive sleep apnea, *American Review of Respiratory Disease*, **147**(3): 635–644, doi: 10.1164/ajrccm/147.3.635.

26. Pevernagie D., Aarts R.M., De Meyer M. (2010), The acoustics of snoring, *Sleep Medicine Reviews*, **14**(2): 131–144, doi: 10.1016/j.smrv.2009.06.002.

27. Qian K. *et al.* (2019), A Bag of wavelet features for snore sound classification, *Annals of Biomedical Engineering*, **47**(4): 1000–1011, doi: 10.1007/s10439-019-02217-0.

28. Rabiner L.R., Gold B., Yuen C.K. (1975), Theory and application of digital signal processing, *IEEE Transactions on Systems, Man, and Cybernetics*, **8**(2): 146–146, doi: 10.1109/TSMC.1978.4309918.

29. Senaratna C.V. *et al.* (2017), Prevalence of obstructive sleep apnea in the general population: A systematic review, *Sleep Medicine Reviews*, **34**: 70–81, 10.1016/j.smrv.2016.07.002.

30. Sola-Soler J., Jane R., Fiz J.A., Morera J. (2003), Spectral envelope analysis in snoring signals from simple snorers and patients with obstructive sleep apnea, [in:] *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, **3**: 2527–2530, doi: 10.1109/IEMBS.2003.1280430.

31. Strollo P.J., Rogers R.M. (1996), Obstructive sleep apnea, *New England Journal of Medicine*, **334**(2): 99–104, doi: 10.1056/NEJM199601113340207.

32. Sun X., Peng J., Zhang X., Song L. (2022), Effective feature selection based on Fisher Ratio for snoring recognition using different validation methods, *Applied Acoustics*, **186**: 108483, doi: 10.1016/j.apacoust.2021.108429.

33. Winursito A., Hidayat R., Bejo A. (2018), Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition, [in:] *2018 International Conference on Information and Communications Technology*, pp. 379–383, doi: 10.1109/ICOIACT.2018.8350748.

34. Won T.B. *et al.* (2012), Acoustic characteristics of snoring according to obstruction site determined by sleep videofluoroscopy, *Acta Oto-Laryngologica*, **132**: 13–20, doi: 10.3109/00016489.2012.660733.

35. Xie J. *et al.* (2021), Audio-based snore detection using deep neural networks, *Computer Methods and Programs in Biomedicine*, **200**: 105917, doi: 10.1016/j.cmpb.2020.105917.

36. Xu K. *et al.* (2018), Mixup-based acoustic scene classification using multi-channel convolutional neural network, [in:] *Advances in Multimedia Information Processing – PCM 2018*, pp. 14–23, doi: 10.48550/arXiv.1805.07319.

37. Yadollahi A., Moussavi Z. (2010), Automatic breath and snore sounds classification from tracheal and ambient sounds recordings, *Medical Engineering & Physics*, **32**(9): 985–990, doi: 10.1016/j.medengphy.2010.06.013.

38. Young T., Peppard P.E., Gottlieb D.J. (2002), Epidemiology of obstructive sleep apnea: A population health perspective, *American Journal of Respiratory and Critical Care Medicine*, **165**(9): 1217–1239, doi: 10.1164/rccm.2109080.

Research Paper

Statistical Metrics for the Temporal Acoustics of Durationally Contrastive Vocalics: A Proposal Tested with Data from Arabic and Japanese

Yahya ALDHOLMI

Department of Linguistics, College of Language Sciences
King Saud University
P.O. Box 87907, Riyadh 11652, Saudi Arabia; e-mail: yaldholmi@ksu.edu.sa

Previous research has utilized the *duration ratio* and occasionally the *duration difference* as single-value metrics to measure and compare the temporal acoustics of durationally contrastive vocalics (short vs. long vowels), which allow researchers to reduce two values (short and long) to one, but express a *relationship* instead of representing the vocalic duration values directly. The duration ratio may even be misleading when comparing two languages or dialects, as it is possible to exhibit a similar ratio but differ in durational acoustics, or vice versa. The current study proposes two alternative statistical metrics: a duration metric and a difference metric. The duration metric is an intermediate (mean-like) value between the duration of the short and long vocalics, and the difference metric is a ± value that can be added to or subtracted from the duration metric to obtain the duration of long or short vocalics. We conduct a production experiment on Arabic and Japanese vocalics and analyze the data using both traditional measures and the proposed metrics. The findings show that the proposed metrics better predict the language from which the vocalic duration values were obtained. Such results suggest that the proposed metrics are better candidates for measuring and comparing the temporal acoustics of vocalics.

Keywords: vocalic duration; duration difference; duration ratio; duration contrast; metrics.

1. Introduction

The durationally contrastive vocalics[1] (i.e., short and long vowels) in natural and synthetic speech have been investigated in research on both native vs. non-native and normal vs. impaired production and perception in languages with durationally contrastive vocalics, such as Arabic. Most of the early research took a theoretical perspective and focused on the observation that, as in other Semitic languages such as Hebrew, a vocalic duration is inherently phonemically contrastive in Arabic (e.g., HARRIS, 1942; CANTINEAU, 1956; FERGUSON, 1957; COWAN, 1970). Subsequent studies employed experimental or obser-

vational methods to examine the temporal acoustics of Arabic vocalics acoustically, perceptually, or acoustically-perceptually for both normal and impaired speakers (e.g., AL-ANI, 1970; MITLEB, 1984; ALGHAMDI, 1998; TSUKADA, 2009; 2011; 2012a; 2012b; 2013; AMIR *et al.*, 2014; KALALDEH, 2018; ALDHOLMI *et al.*, 2021). This phenomenon has been explored in Arabic varieties including Iraqi (HASSAN, 1981), Jordanian (MITLEB, 1984), Saudi, Sudanese, Egyptian (e.g., ALGHAMDI, 1998), Moroccan (AL-TAMIMI, BARKAT-DEFRADAS, 2003), Libyan (AHMED, 2008), Cairene (KOTBY *et al.*, 2011), Palestinian (SAADAH, 2011), Syrian (ALMBARK, HELLMUTH, 2015), and Modern Standard Arabic (MSA) (KALALDEH, 2018; ALDHOLMI, 2022), as well as other varieties spoken in non-Arab regions such as the Kassem/Kfar Barra/Jaljulia dialects in Israel (AMIR *et al.*, 2012), Galilee and the Triangle (Muthallath) regions in Israel (AMIR *et al.*, 2014), and Malaysia (ABD ALMISREB *et al.*, 2016).

---

[1]Note that "vocalics" and "temporal acoustics" are used as two general terms throughout the paper. The former refers to vowels regardless of length and type, and the latter refers to all durational aspects regardless of tool (e.g., duration difference, duration ratio...). Specific terms are used instead when applicable.

Most of the aforementioned studies have documented and characterized the temporal acoustics of Arabic vocalics and the short-to-long duration ratio thereof both interdialectally and cross-dialectally. Other studies sought to compare and contrast the Arabic native production of vocalic duration with that of other Semitic languages that exhibit vocalic durational contrast, such as Hebrew (e.g., Amir *et al.*, 2012), and non-Semitic languages that either feature durational contrast, such as Japanese and Thai (Tsukada, 2009), or do not, such as English (Mitleb, 1984). Other studies examined non-native production and perception of vocalic duration by speakers of Arabic as a second language (L2) whose native language (L1) was Japanese (Tsukada, 2012a; 2012b), English (Flege, Port, 1981; Lababidi, Park, 2014), Korean (Hong, Sarmah, 2009), or Hebrew (Zaltz, Segal, 2021).

To this end, researchers have utilized both the *duration difference* and the *duration ratio* in dialect and language comparisons. For instance, some Arabic dialects, specifically Jordanian, have been reported to have a short-to-long duration ratio of 0.65 (Mitleb, 1984), while others have demonstrated a considerably smaller ratio, for example, 0.39 in Palestinian (Saadah, 2011). This discrepancy may not necessarily be due to the use of different stimuli or methods but rather due to actual interdialectal variations, as the duration ratio does *not* truly convey much about vocalic duration in one dialect or another, nor does it allow for a clear cross-dialect or cross-language comparison.

The duration ratio does not directly express vocalic duration in time units (e.g., ms); instead, it shows only how large or small a value is in relation to another value. That is, the duration ratio of 100 to 200 ms is exactly the same as that of 200 to 400 ms (0.5 in both), which makes this measure unhelpful when comparing two language varieties. The duration difference only shows the quantitative relationship between two given vocalics as short in duration and long in duration, rather than reflecting the actual duration acoustics of the segments under investigation. In addition, sometimes, the short version becomes too long or the long version becomes too short, which results in a negative duration difference value when calculating the difference for each minimal pair (e.g., 70 − 100 = −30 ms). Considering that individual value (rather than the overall mean difference), the difference between two positive values should be any nonnegative value (including zero), but a negative duration ratio value will be uninterpretable. There are a few potential solutions to this particular issue, but each has its own problems. For instance, we could transform and normalize data to be at or above zero, but this would increase the overall mean duration difference.

Hence, this study proposes two statistical metrics to allow for *direct* comparison between different vari-eties in terms of vocalic duration. The first section provides background and describes the two metrics, the duration metric and the difference metric, that can be used instead of the duration difference or duration ratio values reported in previous studies. The two metrics are illustrated using available data from relevant literature. In the second section, a production experiment is conducted to test the two alternative metrics, using data from Arabic and Japanese as two languages that have been repeatedly compared and contrasted in the literature (e.g., Tsukada, 2009) because they share similar durationally contrastive vocalics (e.g., Ald-holmi, 2022).

2. Traditional measures vs. proposed metrics

2.1. Traditional measures

As reported in some previous studies, the traditional method for obtaining a short-long duration ratio divides the mean duration of the short vowels by that of the long vowels. For instance, Mitleb (1984) reported a ratio of 0.65, calculated as (Eq. (1)):

$$\text{ratio} = \frac{\text{mean short vowels}}{\text{mean long vowels}},$$
$$0.65 = \frac{83 \text{ ms}}{128 \text{ ms}}. \tag{1}$$

In some cases, the duration difference is reported instead of the duration ratio. The duration difference is simply the difference between the mean duration of the long vowels and that of the short vowels, as shown in Eq. (2). Mitleb (1984) reported a duration difference of 45 ms:

$$\text{difference} = \text{mean long vowels} - \text{mean short vowels},$$
$$45 = 128 \text{ ms} - 83 \text{ ms}. \tag{2}$$

The duration ratio is sometimes reported in qualitative rather than numerical form. For instance, Tsukada (2011) stated that "long [Arabic] vowels are twice as long as their short counterparts" (p. 989), while "long Japanese vowels tend to be more than twice as long as their short counterparts" (p. 990). Regardless, both the duration ratio and the duration difference depend on the range of the two values, specifically the mean short vowel duration and the mean long vowel duration, which on their own are insufficient to precisely quantify the vocalic duration in a given dialect. For example, suppose that in one Arabic dialect the mean duration of two short vocalics (65 + 75 ms) is 70 ms while the mean duration of two long vocalics (165 + 175 ms) is 170 ms, and in another Arabic dialect the mean duration of two short vocalics (115 + 125 ms) is 120 ms while the mean duration of two long vocalics (285 + 295 ms) is 290 ms. In both scenarios, the ratio is

approximately 0.41, but the difference is 100 ms in the first and 170 ms in the second. Using a duration ratio value makes the two dialects seem similar, but the range of the values and the difference do not.

2.2. Proposed metrics

The proposed duration difference is similar to the formant spacing – compact-diffuse (C-D) measure used in some studies in which the first formant ($F1$, a smaller value) is subtracted from the second formant ($F2$, a larger value) (e.g., BLOMGREN *et al.*, 1998; KENT, VORPERIAN, 2018). Although computing the C-D value has a different purpose, namely, to evaluate tongue elevation (e.g., JAKOBSON *et al.*, 1963), it reduces the two values into a single value that can be used for statistical description and inference. Another formant spacing value is the so-called grave-acute (G-A) measure (KENT, VORPERIAN, 2018), which describes tongue advancement (JAKOBSON *et al.*, 1963; BLOMGREN *et al.*, 1998). The G-A value has been computed according to Eq. (3), where X = each individual vocalic, and n = the total number of data points (vocalics):

$$\text{G-A} = \frac{1}{n} \sum \frac{(F1 \text{ of vocalic } X + F2 \text{ of vocalic } X)}{2}. \quad (3)$$

This method can form the basis of a new, alternative metric that can be used to describe the vocalic duration and the vowel difference in languages where the vocalic duration is contrastive. The proposed metric can be calculated by the given equation (Eq. (4)):

$$\text{duration metric} = \frac{1}{n} \sum \frac{(\text{short} + \text{long})}{2}. \quad (4)$$

The output provides us with one value that lies between the original value of the short vowel and that of the long vowel, but it should better inform us about how short or long the two contrastive vocalics are in a given dialect or language. To illustrate this, consider the previous two scenarios, calculated as (a) and (b) for convenience. Note that we treat the mean durations as single data points for two individual vocalics:

a) duration metric = $\frac{1}{2} \left(\frac{65+165}{2} + \frac{75+175}{2} \right)$ = 120 ms,

b) duration metric = $\frac{1}{2} \left(\frac{115+285}{2} + \frac{125+295}{2} \right)$ = 205 ms.

The two obtained values indicate that the first dialect has notably smaller short and long vocalic durations than the second dialect. In other words, the overall duration of vocalics in the second dialect is approximately 42% longer than that in the first dialect. Neither the duration ratio, which is identical in both dialects (0.41), nor the duration difference, which always depends on the distance between the short and long vowel durations, will provide a unified metric that

allows for a direct comparison between the two dialects or languages. Nevertheless, the proposed duration metric here still does not show how far the duration value is from the original short and long durations. Hence, one further step is needed, which is to calculate the difference metric (Eqs. (5) and (6)):

$$\text{difference metric} = \text{duration metric}$$
$$\pm (\text{duration metric} - \text{short vocalic}) \quad (5)$$

or

$$\text{difference metric} = \text{duration metric}$$
$$\pm (\text{long vocalic} - \text{duration metric}). \quad (6)$$

Note that Eqs. (5) and (6) provide the exact value. Consider the vowel difference computed for the aforementioned scenarios:

a) difference metric = 120 ms $\pm \left(\frac{65+75}{2} = 70 \right)$ = 50 ms,

b) difference metric = 205 ms $\pm \left(\frac{115+125}{2} = 120 \right)$ = 85 ms.

The ± value is the difference metric that we can add to or subtract from the vowel duration metric to obtain the duration of the short vocalic or of the long vocalic. In the first scenario, 120 ms ± 50 ms = 70 or 170 to yield the durations of the short and long vocalics, respectively. The difference metric shows that the difference between the short and long vowels is smaller in the first dialect than in the second dialect. The same applies to the second scenario.

Thus, the duration metric provides us with one value that represents both short and long vocalics. This cannot be achieved via the traditional duration difference (where the short duration is subtracted from the long duration) because the short and long vowels can have large values (e.g., 200 and 250 ms, respectively), but the duration difference, which will be 50 ms in this case, cannot be used to calculate the exact duration of either vocalic. Similarly, two smaller values for short and long vowels (e.g., 50 and 110 ms) can have a larger duration difference, calculated here as 60 ms, but this value also indicates nothing about the duration of the short and long vocalics. The proposed duration metric *does* provide information about how long the short and long vocalics are. To illustrate this with a real-world example, we analyze data from TSUKADA (2011).

2.3. An example from Arabic and Japanese

The short vocalic /a/ in trial 1 has a relatively small duration ratio (0.37) compared to its long counterpart, which is below the lowest value reported in the literature on Arabic (0.39), while the short vocalic /u/ in trial 1 has a relatively larger duration ratio (0.51), which is above the frequently reported range (39–45) in the literature (e.g., TSUKADA, 2011). Nevertheless,

Table 1. Duration [in ms], difference [in ms], and ratio in Arabic (data adopted from TSUKADA, 2011, p. 992).

Trial/item	Long vowels	Long vowels duration	Short vowels	Short vowels duration	Duration ratio	Duration difference	Short + long	Duration metric	Difference metric
1	/aa/	307.00	/a/	114.00	0.37	193.00	421.00	210.50	±96.50
2	/aa/	190.00	/a/	96.00	0.51	94.00	286.00	143.00	±47.00
1	/ii/	235.00	/i/	103.00	0.44	132.00	338.00	169.00	±66.00
2	/ii/	173.00	/i/	87.00	0.50	86.00	260.00	130.00	±43.00
1	/uu/	212.00	/u/	109.00	0.51	103.00	321.00	160.50	±51.50
2	/uu/	183.00	/u/	88.00	0.48	95.00	271.00	135.50	±47.50
Mean	**All**	**216.67**	**All**	**99.50**	**0.47**	**117.17**	**316.17**	**158.08**	**±58.58**

Table 2. Duration [in ms], difference [in ms], and ratio in Japanese (data adopted from TSUKADA, 2011, p. 992).

Trial/item	Long vowels	Long vowels duration	Short vowels	Short vowels duration	Duration ratio	Duration difference	Short + long	Duration metric	Difference metric
1	/aa/	190.00	/a/	67.00	0.35	123.00	257.00	128.50	±61.50
2	/aa/	166.00	/a/	65.00	0.39	101.00	231.00	115.50	±50.50
1	/ee/	182.00	/e/	88.00	0.48	94.00	270.00	135.00	±47.00
2	/ee/	159.00	/e/	79.00	0.50	80.00	238.00	119.00	±40.00
1	/ii/	195.00	/i/	80.00	0.41	115.00	275.00	137.50	±57.50
2	/ii/	176.00	/i/	69.00	0.39	107.00	245.00	122.50	±53.50
1	/oo/	181.00	/o/	84.00	0.46	97.00	265.00	132.50	±48.50
2	/oo/	156.00	/o/	75.00	0.48	81.00	231.00	115.50	±40.50
1	/uu/	179.00	/u/	59.00	0.33	120.00	238.00	119.00	±60.00
2	/uu/	152.00	/u/	61.00	0.40	91.00	213.00	106.50	±45.50
Mean	**All**	**173.60**	**All**	**72.70**	**0.42**	**100.90**	**246.30**	**123.15**	**±50.45**

the actual durations of both short vocalics are 114 ms and 109 ms, respectively, which exhibit only a negligible difference (114 – 109 = 5 ms). The overall duration ratio (0.47) and the overall duration difference (117.17 ms) do not indicate the actual magnitude of the durations of short and long vocalics in Arabic. In contrast, the duration metric *does* show that Arabic short vocalics are generally shorter and Arabic long vocalics longer than 158.08 ms and that the distance between short or long vowels and this metric value is ±58.58 ms overall.

Table 2 shows a duration metric in Japanese of 123.15 ms and a difference metric of ±50.45 ms, suggesting that Japanese vocalics generally tend to be shorter than Arabic vocalics (123.15 vs. 158.08 ms, respectively) and that the difference (not the ratio) between short and long vocalics in Japanese is smaller than that in Arabic (50.45 vs. 58.58 ms, respectively). Inspection of the means for both Arabic and Japanese short and long vowels supports this conclusion.

3. Duration metric and difference metric tested: An experiment

The current experiment utilizes the proposed duration and difference metrics for statistical analysis and compares them with the traditional duration difference and the duration ratio measures in Arabic and Japanese.

3.1. Methodology

The stimuli for this study consist of 18 MSA CVCVC vs. CV:CVC words and 18 Japanese CVCV vs. CV:CV words. The Arabic items selected for this experiment were inspired by (but not taken from) HASSAN (2002), while the Japanese items were selected from TSUKADA (2012b). The target vocalic in the stimuli from both languages was the first rather than the second/final vocalic because the final vocalic is subject to certain phonological processes such as shortening and lengthening (see ALDHOLMI, 2022). Following the same protocol by ALDHOLMI (2022), 22 male and 18 female native speakers of Arabic ($n = 40$) produced the items using an Arabic carrier sentence (/ʔanaa ʔaktubu ___ ʔaidˤan/ "I write ___ as well"). The Arabic participants spoke different Arabic dialects, including Najdi (Qassimi, $n = 8$), Hijazi (Jeddah and Madinah, $n = 11$), Southern (Faifa and Abha, $n = 12$), and Northern (Northern Borders, $n = 9$) dialects. The Arabic participants met face-to-face with the experimenter or other linguists who had volunteered to help the researcher collect the data at different Saudi institutions.

Twenty-four male and 16[2] female native speakers of Japanese (n = 40; the initial sample comprised 41 participants, but one was excluded for unclear speech) produced the Japanese items using a Japanese carrier sentence adopted from Tsukada (2011, p. 991) (/tsugiwa ___ to iimasu/ "Next I say the word ___") and performed the task entirely online (using Phonic.ai, 2023). Approximately half of the Japanese participants (n = 19) came from Osaka, while the rest did not specify their origin. The target vocalics were isolated from the adjacent consonants by the experimenter, using both visual and auditory judgements for all items. Praat (Boersma, Weenink, 2021) was used for segmenting and marking the boundaries of segments for all items.

3.2. Results

As shown in Figs. 1 and 2 and Table 3, Arabic has larger means and greater variability for both short and
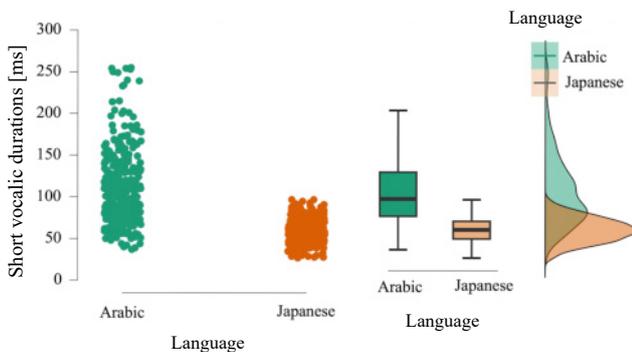

Fig. 1. Short vocalic durations in Arabic and Japanese.
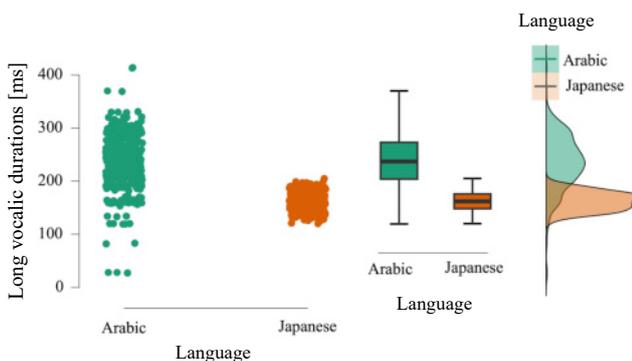
Fig. 2. Long vocalic durations in Arabic and Japanese.

---

[2]It would have been desirable to maintain gender balance for both Arabic and Japanese speakers, had the Japanese speakers been as accessible to the researcher as the Arabic speakers were. Nonetheless, an attempt was made to maintain a similar female-to-male ratio in both groups, although previous studies did not always have gender-balanced groups. For instance, Tsukada had 7 Arabic speakers (4 male and 3 female) in some studies (e.g., Tsukada, 2011) and 9 Arabic speakers (6 male and 3 female) in some other studies (e.g., Tsukada, 2012a). It should also be recalled that speakers maintain duration distinction in languages that exhibit durationally contrastive vocalics such as Hebrew regardless of gender (e.g., Amir *et al.*, 2012).

Table 3. Short and long vocalic means [ms], SDs, minimums and maximums of durations in Arabic and Japanese.

Length	Language	Mean	SD	Minimum	Maximum
Short	Arabic	108.95	42.43	50.00	254.63
	Japanese	60.51	14.56	27.00	97.00
Long	Arabic	237.11	52.55	27.00	413.94
	Japanese	161.88	17.37	120.00	205.00

long vocalics (short: M = 108.95, SD = 42.43; long: M = 237.11, SD = 52.55) than Japanese (short: M = 60.51, SD = 14.56; long: M = 161.88, SD = 17.37).

The repeated-measures ANOVA with the vocalic length (short vs. long) as a within-subject factor and language as a between-subject factor was performed to test mean differences. As detailed in Table 4, the test provided evidence for a statistically significant difference between short and long vocalics with a very large effect size, $F(1, 78) = 2047.16$, $p < 0.001$, $\omega^2 = 0.86$, and between Arabic and Japanese, also with a large effect size, $F(1, 78) = 182.51$, $p < 0.001$, $\omega^2 = 0.53$. There was also a statistically significant interaction between the two factors with an intermediate effect, $F(1, 78) = 27.89$, $p < 0.001$, $\omega^2 = 0.07$.

Table 4. Repeated-measures ANOVA results.

Factors	df	F	p	ω^2
Vocalic length	1	2047.16	<.001	0.86
Language	1	182.51	<.001	0.53
Vocalic length * language	1	27.89	<.001	0.07

Thus, we have strong evidence that Arabic and Japanese differ significantly in terms of duration for both short and long vocalics and that, within each language, short vocalics are shorter than their counterparts. Table 5 presents the duration differences and duration ratios for both languages. The duration difference for Arabic (128.16 ms) and for Japanese (101.82 ms) and the duration ratio for Arabic (0.48) and Japanese (0.37) are similar to those calculated and obtained from the data provided in (Tsukada, 2011). Hence, the duration difference may be misinterpreted as indicative of an overall similarity between the vocalic duration in Arabic and Japanese, which is not precisely the case.

Table 5. Duration differences and duration ratios in Arabic and Japanese.

	Language	Mean	SD	Minimum	Maximum
Duration difference [ms]	Arabic	128.16	54.20	−167.47*	329.84
	Japanese	101.82	19.45	45.00	294.00
Duration ratio	Arabic	0.48	0.29	0.20	3.04
	Japanese	0.37	0.06	0.20	0.54

* Note that the negative value is one of the issues caused by using the duration difference, as explained in the introduction.

Now consider both the proposed duration metric and the difference metric in Table 6. The duration metric for Arabic (173.03 ms) was substantially larger than that for Japanese (111.20 ms). Likewise, the difference metric for Arabic (64.08 ms) was considerably greater than that for Japanese (51.30 ms).

Table 6. Duration metrics and difference metrics in Arabic and Japanese.

	Language	Mean	SD	Minimum	Maximum
Duration metric [ms]	Arabic	173.03	39.32	42.50	283.54
	Japanese	111.20	14.85	74.50	145.50
Difference metric [ms]	Arabic	64.08	42.43	50.00	254.63
	Japanese	51.30	28.31	−52.00	259.00

Thus, based on the aforementioned data, we observe that the duration metric and the difference metric better represent the vocalic duration facts in both languages. The values are re-reported side-by-side in Table 7, which arguably illustrates how the substantial dissimilarity between Arabic and Japanese and between short and long vocalics is reflected more clearly in the duration metric and the difference metric than in the duration difference and the duration ratio. To support this claim, an inverse regression was performed to test which of the four variables (duration difference, duration ratio, duration metric, or difference metric) would most accurately predict the language. We first compare the duration difference and the duration metric, as these two are similar; both inform us about the actual duration of the short vs. long vowels. Next, we compare the duration ratio and the difference metric, as these two are also similar; both inform

us about the relationship between two values. Despite the similarity in purpose between the members of each group, the difference metric and duration metric both have the added benefit of being able to inform us about the mean vocalic duration measures as well.

We fitted an inverse binary logistic regression model, first using the *duration difference* as a predictor variable and the *language* as a predicted variable. The results indicated a significant improvement in fit relative to an intercept-only model, $\chi^2(1) = 79.58$, $p < .001$, and that the duration difference was a statistically significant predictor of language, $\chi^2(1) = 60.60$, $p < .001$. Table 8 shows the −2 log-likelihood (−2LL) and the pseudo-R^2 values of the first model (model 1). As shown, in order from the largest pseudo-R^2 value to the smallest, the Nagelkerke R^2, Tjur R^2, Cox and Snell R^2, and McFadden R^2 exhibited relatively similar, low values. These values become important later when we compare with another predictor variable.

Table 9 shows that the sensitivity of the model was 78.60%, the specificity of the model was 66.70%, and the overall accuracy was 72.60%.

The model was re-fitted using the *duration metric* value as a predictor variable. The results again showed a significant improvement in fit for the second model (model 2) relative to an intercept-only model, $\chi^2(1) = 572.06$, $p < .001$, and that duration metric was a statistically significant predictor of the language, $\chi^2(1) = 159.56$, $p < .001$. When the duration difference was used as a predictor for the language, the −2LL value was lower while the pseudo-R^2 values (Table 10) were higher than those obtained in the previous model, demonstrating the development of better fit in model 2.

Table 7. Vocalic durations, duration differences, duration ratios, duration metrics, and difference metrics in Arabic and Japanese.

	Short duration [ms]		Long duration [ms]		Duration difference [ms]		Duration ratio		Duration metric [ms]		Difference metric [ms]	
	Arabic	Japanese	Arabic	Japanese	Arabic	Japanese	Arabic	Japanese	Arabic	Japanese	Arabic	Japanese
Mean	108.95	60.51	237.11	161.88	128.16	101.82	0.48	0.37	173.03	111.20	64.08	51.30
SD	42.43	14.56	52.55	17.37	54.20	19.45	0.29	0.06	39.32	14.85	42.43	28.31

Table 8. The −2LL and pseudo-R^2 values for model 1.

	χ^2	df	p	−2LL	Nagelkerke R^2	Tjur R^2	Cox and Snell R^2	McFadden R^2
Model	79.58	1	<.001	918.55	0.14	0.12	0.10	0.08
Duration difference	60.60	1	<.001					

Table 9. Confusion matrix (sensitivity and specificity rates) and accuracy rate of model 1.

Observed		Predicted		
		Language		Percentage correct
		Arabic	Japanese	
Language	Arabic	240	120	66.70
	Japanese	77	283	78.60
Overall percentage (accuracy)				72.60

Table 10. The −2LL and pseudo-R^2 values for model 2.

	χ^2	df	p	−2LL	Nagelkerke R^2	Tjur R^2	McFadden R^2	Cox and Snell R^2
Model	572.06	1	<.001	426.06	0.73	0.65	0.57	0.54
Duration metric	159.56	1	<.001					

As shown in Table 11, the sensitivity (85.30%), specificity (92.80%), and overall accuracy (89.00%) all improved in model 2.

Thus, all indicators demonstrated that the duration metric proposed in the current study is a better alternative to the duration difference used in previous studies. We compare the two other indicators (the duration ratio vs. the difference metric), following the same steps used in comparing the duration difference and the duration metric.

An inverse binary logistic regression model (model 3) was performed with the *duration ratio* as a predictor and the *language* as a predicted variable. The output showed that, compared to an intercept-only model, model 3 demonstrated a significant improvement in fit, $\chi^2(1) = 101.92$, $p < .001$, and that duration ratio was a statistically significant predictor of the language, $\chi^2(1) = 68.31$, $p < .001$. The −2LL value (896.20) and the pseudo-R^2 values (Cox and Snell $R^2 = 0.17$, Nagelkerke $R^2 = 0.13$, Tjur $R^2 = 0.12$, and McFadden $R^2 = 0.10$) were very similar (Table 12)

to those obtained when the duration difference was used as an indicator.

The sensitivity (59.7%), specificity (70.8%), and overall accuracy (65.3%) of the model, as shown in Table 13, indicated that this model exhibited poor sensitivity and slightly poor overall accuracy.

Running the model again with the *difference metric* as a predictor, model 4 showed a significant improvement in fit relative to the intercept-only model, $\chi^2(1) = 537.29$, $p < .001$. It also indicated the difference metric as a statistically significant predictor of the language, $\chi^2(1) = 167.56$, $p < .001$. The −2LL value (460.84) and the pseudo-R^2 values (Nagelkerke $R^2 = 0.71$, Tjur $R^2 = 0.60$, Cox and Snell $R^2 = 0.53$, and McFadden $R^2 = 0.52$) were highly similar (Table 14) to those obtained when using the duration metric as an indicator.

The sensitivity (87.2%), specificity (84.2%), and overall accuracy (85.7%) of the model, as shown in Table 15, were notably higher than those in the previous model and indicated good fit. Thus, running the model

Table 11. Confusion matrix (sensitivity and specificity rates) and accuracy rate of model 2.

Observed		Predicted		
		Language		Percentage correct
		Arabic	Japanese	
Language	Arabic	307	53	85.30
	Japanese	26	334	92.80
Overall percentage (accuracy)				89.00

Table 12. The −2LL and pseudo-R^2 values for model 3.

	χ^2	df	p	−2LL	Cox and Snell R^2	Nagelkerke R^2	Tjur R^2	McFadden R^2
Model	101.92	1	<.001	896.20	0.17	0.13	0.12	0.10
Duration ratio	68.31	1	<.001					

Table 13. Confusion matrix (sensitivity and specificity rates) and accuracy rate of model 3.

Observed		Predicted		
		Language		Percentage correct
		Arabic	Japanese	
Language	Arabic	215	145	59.70
	Japanese	105	225	70.80
Overall percentage (accuracy)				65.30

Table 14. The −2LL and pseudo-R^2 values for model 4.

	χ^2	df	p	−2LL	Nagelkerke R^2	Tjur R^2	Cox and Snell R^2	McFadden R^2
Model	537.29	1	<.001	460.84	0.71	0.60	0.53	0.52
Duration metric	167.56	1	<.001					

Table 15. Confusion matrix (sensitivity and specificity rates) and accuracy rate of model 4.

Observed		Predicted		
		Language		Percentage correct
		Arabic	Japanese	
Language	Arabic	314	46	87.20
	Japanese	57	303	84.20
Overall percentage (accuracy)				85.70

again with the *difference metric* as a predictor significantly improved the model's goodness of fit compared to using the *duration ratio* as a predictor.

4. Discussion and conclusion

The findings above agree with a large body of literature that has shown that Arabic and Japanese contrast short and long vowels (e.g., Tsukada, 2013), as well as with previous observations that Arabic short vowels weigh approximately 50% of their long counterparts while Japanese short vowels weigh less than 50% of their long counterparts (e.g., Tsukada, 2011). The duration difference and the duration ratio were, respectively, 128.16 and 0.48 for Arabic vocalics and were, respectively, 101.82 and 0.37 for Japanese vocalics. The duration differences (128.16 and 101.82 ms) do not reflect the short and long durations in Arabic or Japanese; Arabic short vowels are approximately 55% longer than Japanese short vowels, Arabic long vowels are approximately 65% longer than Japanese long vowels, and, overall, Arabic vocalics are approximately 60% longer than Japanese vocalics. Likewise, the duration ratio does not convey much information about vocalic duration within-language (e.g., Arabic or Japanese) or between the two languages (Arabic and Japanese) nor in comparison with other languages. Based on the data we obtained in this experiment, the duration ratios in Arabic and Japanese are relatively similar: 0.48 in Arabic and 0.37 in Japanese. That is, the durations of short and long vowels in Arabic are nearly double those in Japanese, but we cannot deduce this from the duration ratio.

In comparison, the duration metric (173.03 ms) and difference metric in Arabic (±64.08 ms) diverged from the duration metric (111.20 ms) and the difference metric in Japanese (±51.30 ms). The duration metric shows the average length of both short and long vowels; we can see clearly that Arabic vocalics are considerably longer than Japanese vocalics. The duration metric shows the extent to which short vocalics and long vocalics are similar or different within and between Arabic and Japanese, and we can ascertain that the difference between short and long vocalics in Arabic is greater than that in Japanese and that, moreover, that duration is more variable in Arabic than in Japanese. The two metrics together show that the duration metric of short and long vocalics in Arabic (173.3 ms) is

very close to the duration of long vocalics in Japanese (111.20 ± 51.30 = **162 ms**) and that the duration metric of short and long vocalics in Japanese (111.20 ms) is also similar to the duration of short vowels in Arabic (173.03 – 64.08 = **108.95 ms**).

Neither the duration difference nor the duration ratio is a factual duration unit. Unlike the duration metric, the duration difference does not provide the actual duration of vocalics in Arabic vs. Japanese. Likewise, the duration ratio is a completely different measurement unit that no longer expresses the duration in time units and cannot indicate the duration of short vowels relative to long ones in Arabic vs. Japanese. The duration ratio cannot be used to compare vocalic durations between dialects or languages, because two different languages that have two distinct duration measurements for short and long vocalics may still have similar or even identical duration ratios. For instance, the duration ratio for Palestinian vocalics is approximately 0.39 (Saadah, 2011) and for Japanese vocalics in the current experiment was 0.37. These two values are extremely similar, but overall, Palestinian short and long vocalics are both longer than their Japanese counterparts. Using the actual vocalic duration measurements in a statistical test to compare vocalics in Palestinian and Japanese should reveal a significant difference, while using the duration ratio is unlikely to reveal any differences. This is probably the reason why the duration and difference metrics were better predictors of the language.

To summarize, this paper shows how the duration difference and duration ratio measures used in previous studies are not optimal metrics for comparing vocalic duration within and across languages. We propose two alternative metrics: the duration metric and the difference metric. Using data from a previous study (Tsukada, 2011), we illustrate the difference between the duration ratio and the duration difference, on the one hand, and between the duration metric and difference metric, on the other hand. We then conduct an experiment to examine the new metrics. The findings show that short and long vocalic durations differ in both Arabic and Japanese and that Arabic and Japanese also differ in terms of short and long vowel durations. More importantly, the key finding is that the proposed metrics were better predictors of the language than the traditional measures. This finding invites researchers on the vocalic duration, whether pho-

neticians, language acquisitionists, or speech pathologists, to consider using (and testing) of the proposed metrics. We also call for a revisiting of the findings established in previous literature, especially those studies that compared several languages or dialects (e.g., ALGHAMDI, 1998). Future research can survey languages and dialects that have shown similar or dissimilar duration ratios and examine whether the proposed metrics will reveal patterns that differ from those revealed by the traditional duration ratio and duration difference measures.

Acknowledgments

References

1. ABD ALMISREB A., ABIDIN A.F., TAHIR N.M. (2016), An acoustic investigation of Arabic vowels pronounced by Malay speakers, *Journal of King Saud University – Computer and Information Sciences*, **28**(2): 148–156, doi: 10.1016/j.jksuci.2015.08.003.

2. AHMED A.A.M. (2008), *Production and perception of Libyan Arabic vowels*, Ph.D. Thesis, Department of Communication and Language Sciences, Newcastle University.

3. AL-ANI S.H. (1970), *Arabic Phonology: An Acoustical and Physiological Investigation*, De Gruyter Mouton.

4. ALDHOLMI Y. (2022), Medial vowel temporal acoustics in Arabic and Japanese polysyllabic words, *Archives of Acoustics*, **47**(3): 319–329, doi: 10.24425/aoa.2022.142006.

5. ALDHOLMI Y., ALOTAIBI S., ALROUQI M. (2021), Rating nonnativeness in L1-Japanese L2-Arabic speakers' vowels, [in:] *ExLing 2021: Proceedings of the 12th International Conference of Experimental Linguistics*, pp. 9–12.

6. ALGHAMDI M.M. (1998), A spectrographic analysis of Arabic vowels: A cross-dialect study, *Journal of King Saud University – Journal of Arts*, **10**(1): 3–24.

7. ALMBARK R., HELLMUTH S. (2015), Acoustic analysis of the Syrian vowel system, [in:] *Proceedings of the 18th International Congresses of Phonetic Sciences*.

8. AL-TAMIMI J., BARKAT-DEFRADAS M. (2003), Inter-dialectal and inter-individual variability in production and perception: A preliminary study in Jordanian and Moroccan Arabic, [in:] *AIDA 5th Conference Proceedings (Association Internationale de Dialectologie Arabe)*, pp. 171–186.

9. AMIR N., AMIR O., ROSENHOUSE J. (2014), Colloquial Arabic vowels in Israel: A comparative acoustic study of two dialects, *The Journal of the Acoustical Society of America*, **136**(4): 1895–1907, doi: 10.1121/1.4894725.

10. AMIR N., TZENKER O., AMIR O., ROSENHOUSE J. (2012), Quantifying vowel characteristics in Hebrew and Arabic, [in:] *Proceedings of the Afeka-AVIOS Speech Processing Conference*, pp. 37–43.

11. BLOMGREN M., ROBB M., CHEN Y. (1998), A note on vowel centralization in stuttering and nonstuttering individuals, *Journal of Speech, Language, and Hearing Research*, **41**(5): 1042–1051, doi: 10.1044/jslhr.4105.1042.

12. BOERSMA P., WEENINK D. (2021), *Praat: Doing phonetics by computer*, Version 6.2.03, https://www.fon.hum.uva.nl/praat/ (access: 12.03.2022).

13. CANTINEAU J. (1956), The phonemic system of Damascus Arabic, *Word*, **12**(1): 116–124, doi: 10.1080/00437956.1956.11659595.

14. COWAN W. (1970), The vowels of Egyptian Arabic, *Word*, **26**(1): 94–100, doi: 10.1080/00437956.1970.11435584.

15. FERGUSON C.A. (1957), Two problems in Arabic phonology, *Word*, **13**(3): 460–478, doi: 10.1080/00437956.1957.11659647.

16. FLEGE J.E., PORT R. (1981), Cross-language phonetic interference: Arabic to English, *Language and Speech*, **24**(2): 125–146, doi: 10.1177/002383098102400202.

17. HARRIS Z.S. (1942), The phonemes of Moroccan Arabic, *Journal of the American Oriental Society*, **62**(4): 309–318, doi: 10.2307/594035.

18. HASSAN Z.M. (1981), *An experimental study of vowel duration in Iraqi spoken Arabic*, Ph.D. Thesis, Department of Linguistics & Phonetics, University of Leeds.

19. HASSAN Z.M. (2002), Gemination in Swedish and Arabic with a particular reference to the preceding vowel duration: An instrumental and comparative approach, [in:] *Proceedings of Fonetik TMH-QPSR*, **44**(1): 81–84.

20. HONG S.A., SARMAH P. (2009), Korean perception of Arabic phonemes, *Current Issues in Linguistic Interfaces*, **2**: 375–386.

21. JAKOBSON R., FANT C., HALLE M. (1963), *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates*, MIT Press, Cambridge.

22. KALALDEH R. (2018), Acoustic analysis of modern standard Arabic vowels by Jordanian speakers, *International Journal of Arabic-English Studies*, **18**(1): 23–48.

23. KENT R.D., VORPERIAN H.K. (2018), Static measurements of vowel formant frequencies and bandwidths: A review, *Journal of Communication Disorders*, **74**: 74–97, doi: 10.1016/j.jcomdis.2018.05.004.

24. KOTBY M.N. *et al.* (2011), The Arabic vowels: Features and possible clinical application in communication disorders, *Folia Phoniatrica et Logopaedica*, **63**(4): 171–177, doi: 10.1159/000316323.

25. LABABIDI Z., PARK H. (2014), L1-English tense-lax vowel system influence on L2-Arabic short and long vowel learning, [in:] *Perspectives on Arabic Linguistics XXVIII: Papers from the Annual Symposium on Arabic Linguistics*, pp. 63–88, doi: 10.1075/sal.4.03lab.

26. MITLEB F. (1984), Vowel length contrast in Arabic and English: A spectrographic test, *Journal of Phonetics*, **12**(3): 229–235, doi: 10.1016/S0095-4470(19)30879-4.

27. Phonic.ai (2023), *Phonic: Voice and video surveys*, https://www.phonic.ai/ (access: 22.02.2022).

28. SAADAH E. (2011), *The production of Arabic vowels by English L2 learners and heritage speakers of Arabic*, Ph.D. Thesis, Department of Linguistics, University of Illinois at Urbana-Champaign.

29. TSUKADA K. (2009), An acoustic comparison of vowel length contrasts in Arabic, Japanese and Thai: Durational and spectral data, *International Journal of Asian Language Processing*, **19**(4): 127–138.

30. TSUKADA K. (2011), The perception of Arabic and Japanese short and long vowels by native speakers of Arabic, Japanese, and Persian, *The Journal of the Acoustical Society of America*, **129**(2): 989–998, doi: 10.1121/1.3531801.

31. TSUKADA K. (2012a), Comparison of native versus nonnative perception of vowel length contrasts in Arabic and Japanese, *Applied Psycholinguistics*, **33**(3): 501–516, doi: 10.1017/S0142716411000452.

32. TSUKADA K. (2012b), Non-native Japanese listeners' perception of vowel length contrasts in Japanese and Modern Standard Arabic (MSA), *Second Language Research*, **28**(2): 151–168, doi: 10.1177/0267658311435870.

33. TSUKADA K. (2013), Vowel length categorization in Arabic and Japanese: Comparison of native and non-native Japanese perception, *Speech, Language and Hearing*, **16**(4): 187–196, doi: 10.1179/2050572813Y.0000000012.

34. ZALTZ Y., SEGAL O. (2021), The perception of Arabic vowel duration by L1 Hebrew speakers: Can a short training remold the effect of the native phonological system?, *Studies in Second Language Acquisition*, **44**(1): 143–163, doi: 10.1017/S0272263120000728.

Research Paper

Assessment Effects of Humidification of Guitars by Complexity Measures of the Sound Level During Sustain

Robert STĘPIEŃ*, Piotr WRZECIONO

Institute of Information Technology
Warsaw University of Life Sciences
Warsaw, Poland

*Corresponding Author e-mail: robert_stepien@sggw.edu.pl

Air humidity significantly affects the sound of wooden instruments. The sound quality decreases when the instrument is exposed to low humidity for an extended period. Therefore, the instrument is treated with a humidifier to improve sound quality. This study aimed to verify the effectiveness of the humidification process by analyzing the quality of guitar sound with the methods used in signal complexity studies, such as Higuchi's fractal dimension (HFD), symbolic analysis, and empirical mode decomposition (EMD). The sound quality was determined by the sound levels measured before, during, and after the guitars' humidification. The methods used consistently confirmed the improvement of the guitar sound quality after the humidification process. Moreover, it was concluded that the sound quality changes irregularly during the humidification process.

Keywords: guitar; hygroscopicity; complexity parameters; acoustic measurements.

1. Introduction

Sound quality or the quality of an instrument is an essential consideration for musicians and those involved in instrument construction and maintenance. In the case of musical instruments made primarily of wood, the physical condition of this material is essential. It is described by some parameters, among which moisture content is one of the most important. This is because wood contains a high percentage of lignin, which is very hygroscopic (GORDOBIL *et al.*, 2021). A change in wood moisture content causes a change in acoustic parameters, e.g., density, Young's modulus, and other physical parameters. Changes also affect the structure of the wood itself, as very dry wood may crack easily (RATH, STAUDINGER, 2001). The problems mentioned have a significant impact on the musical instrument sound and its overall condition.

The acoustic guitar is an instrument made primarily of wood. The tree species used in the instrument are primarily spruce, cedar, and mahogany (GORE, 2011). Humidity affects not only the sound but also the tuning of this instrument. When the wood is too dry, the neck's geometry changes, causing out-of-tune sounds on the frets (WRZECIONO *et al.*, 2018). In addition, the moisture content of the wood in a guitar may vary with the season. For example, in Poland, wood drying is most noticeable during winter conditions (WRZECIONO *et al.*, 2018).

Luthiers deal with this problem in different ways. One solution is presented in the paper (WRZECIONO *et al.*, 2018), where a parametric analysis of guitar sound before and after humidification was performed. As a result, it was possible to determine several parameters, including the time of sustain, which varied significantly depending on the instrument's condition. However, linking the audible change in the swell to the measurement results remained a challenge. In this paper, we present methods to solve the mentioned problem.

During the experiments, the evaluation of the instrument's sound quality was carried out by its owner. However, the humidification process was conducted in the luthier's workshop, where the instrument was serviced due to excessively low wood moisture content. In such a situation, there is a possibility that

the measured sound parameters do not support the subjective assessment made by the instrument owner. Conducting a double-blind test under such circumstances was impossible. So, measurements and analysis of the results had to be performed to determine whether a trend would correlate with the musicians' perceptions. A detailed study of changes was described in the paper by Wrzeciono *et al.* (2018). Through this study it was found that the most significant changes perceived in sound are related to the sounding time.

The objective of our study closely aligned with the evaluation procedures employed by musicians in assessing the sound quality of an instrument, where in normal operational conditions, musicians leverage the phenomenon of mutual excitation of strings. Consequently, it was necessary to conduct experiments with all strings attached. Therefore, our study aimed not to eliminate this phenomenon but to examine its quantitative changes.

2. Materials and methods

In a pilot study related to the parametric analysis of guitars, about 60 instruments were used before proceeding to a systematic multi-day experiment. All of the instruments tested had new strings attached. The tuning process was carried out using a specialized luthier's device that tunes an empty string to an accuracy of 0.1 cents.

Typically, moisturizing a guitar usually takes a week, and the effects of humidification are monitored daily (Wrzeciono *et al.*, 2018).

The measurement setup included a chamber sound box isolating the guitar from its surroundings, along with a microphone unit. The strings were excited by a free-falling arm containing a handle to prevent the arm from rebounding. Measurements of the sound level obtained by striking the strings with the arm were conducted with a set of microphones placed on the axis of the sound hole, at a distance of 15 cm from the guitar. The primary measurement was made with a PreSonus PRM1 microphone calibrated with a Sonopan KA-50 calibrator. In addition, the Rode NT1 microphone was employed as the second one to record guitar sounds with a low noise level.

The microphones were connected to a Focusrite Scarlett 2i2 2Gen audio interface, and the measurement system was calibrated with a signal from a 94 dB acoustic calibrator (Sonopan KA-50). Then, a recording of the guitar sound was made after striking the strings with the arm. A single recording consisted of ten strokes made every 60 seconds. The recordings were made at a sampling rate of 96 kHz and a bit resolution of 24 bits. Infrasound components were removed from the calibration and measurement signals by the Octave program's digital filter. The signal from the PRM1

microphone was used as a reference to calculate the sound level of the tested guitar. A detailed description of the measurement method and measuring instruments is presented in the paper by Wrzeciono *et al.* (2018).

Several parameters describing the guitar sound were also defined in that work. However, the sounding guitar time, denoted as T_{40}, was the most important one. The T_{40} parameter is the interval of time in which the sound level of the guitar, after impulse excitation, drops by 40 dB. A time window of 10 ms was used in the signal power calculation. In addition, infrasonic components were previously removed from the signal. However, the T_{40} parameter alone does not account for the change in the decay's nature (Wrzeciono *et al.*, 2018).

Therefore, further analytical work was undertaken to reconstruct the auditory impression. The study involved qualitative and quantitative analysis. Higuchi fractal dimension (HFD) and symbolic analysis were chosen as qualitative analysis, while empirical mode decomposition (EMD) was selected as the quantitative analysis.

Both HFD and symbolic method have been used to analyze biomedical signals for medical diagnosis and treatment evaluation efficacy (Gladun, 2020; Gomolka *et al.*, 2018; Pierzchalski *et al.*, 2011; Stojadinović *et al.*, 2020). Therefore, using these methods to analyze a relatively uncomplicated signal, such as the sound level of a guitar, should yield intriguing results. The problem of evaluating the effectiveness of the humidification process is analogous to assess the effectiveness of therapy. At its core, conditioning serves as therapy for the instrument.

The fractal dimension and the characteristics of the symbolic analysis allow, based on the analyzed signal, to determine the state in which the system generating the signal is present, thus enabling the detection of state changes. These methods allow to track changes in long signals through the use of moving window technique. Since the waveforms of sound levels concerning the registration of physiological signals are short, global (for the whole signal) values of fractal dimension and symbolic parameter were calculated. These calculated parameters give general information about the changes in guitar sound.

On the other hand, EMD is currently used in a wide range of topics in geophysics (Huang, Wu, 2008), oceanology (Zhou *et al.*, 2021), biomedicine (Khan, Pachori, 2021; Li *et al.*, 2021; Pierzchalski *et al.*, 2011), and engineering (Zheng *et al.*, 2021). The design of the EMD method gives a broader picture of the changes in the system under study. It is multiparametric and thus less synthetic than the previously discussed methods. The original decomposition into modes and the residue allow us to observe precisely what occurs in the sound level signal.

3. Signal processing

3.1. Brief introduction to methods of analysis

The humidification process significantly changes the shape and complexity of the sound pressure level (SPL) curve. Therefore, employing signal complexity analysis methods is justified. Three methods have been proposed and used here: HFD, symbolic analysis, and EMD. They represent different approaches to signal analysis. HFD is based on scaling law; symbolic analysis uses statistics; EMD is an iterative decomposition procedure. Thus, the convergence of results obtained by these methods confirms the notion of SPL analysis as a complex signal. Furthermore, the agreement of these results with listening evaluations validates the use of these methods for automatic evaluation and control of the humidification process.

HFD and symbolic measure are global. Their values allow determining only the level of complexity of the signal. On the other hand, EMD analysis provides more profound information about the changes that occur in the signal.

3.2. Higuchi's fractal dimension

HFD of the signal curve (HIGUCHI, 1988) measures the signal's waveform complexity and should not be confused with the fractal dimension in phase space (MANDELBROT, 1967). HFD, denoted as D_f, typically ranges from 1.0 (for a straight line or straight Euclidean curve) to 2.0 (for a curve with random amplitudes). The only parameter of Higuchi's algorithm is k_{max}. It is the maximal rescale (time delay) integer parameter, which depends on the sampling frequency and signal length (SPASIĆ et al., 2005). In our study, the optimal value of k_{max} has a value of eight, because D_f has the least variance at this parameter setting.

From sampled in time signal: $X(1), X(2), ..., X(N)$, the algorithm constructs k new series X_m^k: $X(m)$, $X(m+k), ..., X(m+\text{int}((N-m)/k)k)$ for $m = 1, 2, ..., k$, where m is the initial time, k is the delay, and $\text{int}(r)$ is the integer part of a real number r.

For every $k = 1, 2, ..., k_{max}$ the difference between shifted samples starting from the following m is calculated as:

$$L_m(k) = \frac{1}{k}\left(\sum_{i=1}^{\text{int}\left(\frac{N-m}{k}\right)} |X(m+ik) - X(m+(i-1)k)|\right)$$
$$\cdot \frac{N-1}{\text{int}\left(\frac{N-m}{k}\right)k}, \tag{1}$$

where N is the total number of samples.

Next, the mean of the k values $L_m(k)$ for $m = 1, 2, ..., k$ is calculated as:

$$L(k) = \frac{1}{k}\sum_{m=1}^{k} L_m(k). \tag{2}$$

$L(k)$ satisfies the scaling law:

$$L(k) \propto k^{-D_f}, \tag{3}$$

where exponent D_f is HFD. This relationship is reduced to linear form:

$$\log(L(k)) \propto D_f \log\left(\frac{1}{k}\right). \tag{4}$$

Hence, the value of the fractal dimension D_f is calculated by a least-squares linear best-fitting procedure.

3.3. Symbolic analysis

The symbolic analysis uses the methodology applied in information theory (STONE, 2022), which defines many parameters of signal complexity, i.e., entropies and related measures (RIBEIRO et al., 2012; 2017). However, in this paper, we propose to use a more specific parameter whose mathematical description is close to the average codeword length (JOHNSON JR et al., 2003).

The construction of the parameter uses the statistical distribution of symbol sequence representing the falling and rising slope of the signal (STEPIEN, 2011). The general idea is to encode the changes in signal between successive samples with symbols "0" and "1":

$$c(i) = \begin{cases} 1 & \text{if } X(i) \geq X(i-1), \\ 0 & \text{if } X(i) < X(i-1). \end{cases} \tag{5}$$

The symbol "1" denotes an amplitude increase, while the symbol "0" denotes an amplitude decrease between successive signal samples. Thus, rising edges of the signal correspond to "1" sequences, and falling edges to "0" sequences. In this way, the monotonicity of the signal is encoded. Hence, sequences comprising only "1" or "0" symbols are called mono-sequences here. We denote the length of the mono-sequence corresponding to the rising slope by $l(\{1\}^*)$, while that of the falling slope by $l(\{0\}^*)$.

To estimate the probabilities $p(l(\{1\}^*))$ and $p(l(\{0\}^*))$ of occurrence of mono-sequences of consecutive lengths let us encode the signal according to the rule (Eq. (5)). Then, count the encoded signal's mono-sequences according to their length and divide by the total number of mono-sequences of a given type.

Our signal characteristic is the sum of mean values of mono-sequences' lengths in the coded signal, which are calculated as:

$$L_1 = \sum_{l=1}^{l_{max}} p(l(\{1\}^*))l(\{1\}^*),$$
$$L_0 = \sum_{l=1}^{l_{max}} p(l(\{0\}^*))l(\{0\}^*). \tag{6}$$

Finally, we obtain a parameter called the sum of mean lengths (SML), which measures the complexity of the signal:

$$\text{SML} = L_0 + L_1. \tag{7}$$

The SML parameter is the sum of the mean values and, as such, is an average measure. However, unlike entropy per symbol, it is not a measure that directly characterizes the source of the signal but rather a measure of the complexity of the signal itself and, indirectly, its source.

In addition to providing overall signal characteristics, this parameter can be used to track the evolution of signal complexity, e.g., using the moving window technique.

3.4. Empirical mode decomposition

EMD decomposes multi-component signals into their mono-components, as proposed by HUANG *et al.* (1998). EMD is a data-driven algorithm that does not depend on any predefined basis function. Such mono-components are called intrinsic mode functions (IMFs). An IMF is a signal that fulfills the following conditions: the number of extrema and the number of zero crossings of the IMF are either the same or their difference is 1; the signal has "zero mean" – meaning the mean value of the envelope determined by the maxima and the envelope defined by the minima is equal to 0 at every point.

The above conditions suggest that EMD – non-stationary signal is decomposed into stationary, symmetric signals (modes) that are easy to analyze.

The crucial step of EMD is extracting extrema from the original signal $x(t)$ end creating the upper envelope $e_{\max}$ and the lower envelope $e_{\min}$ by cubic spline interpolation (DE BOOR, 1978) of the maxima and minima, respectively. Then, the mean value of the two envelopes is computed as:

$$m(t) = \frac{e_{\max} - e_{\min}}{2}. \tag{8}$$

The value $m(t)$ is subtracted from the primary signal $x(t)$ resulting in:

$$\mathrm{imf}_1(t) = x(t) - m(t). \tag{9}$$

This is called the sifting process (Fig. 1).

In an ideal case, $\mathrm{imf}_1(t)$ could be the first mode IMF_1, but it usually remains an asymmetric signal. In such a case, we need to repeat the above procedure, treating $\mathrm{imf}_1(t)$ as the input data for the subsequent sifting process, so the mean value $m(t)$ of the envelopes of $\mathrm{imf}_1(t)$ is calculated, and this value is subtracted from $\mathrm{imf}_1(t)$:

$$\mathrm{imf}_1(t) := \mathrm{imf}_1(t) - m(t). \tag{10}$$

In Eq. (10), the sign ":=" denotes "becomes," that is, in the programming loop, the right-hand side is substituted for the left-hand side. This procedure is repeated until $\mathrm{imf}_1(t)$ meets the conditions of an IMF signal
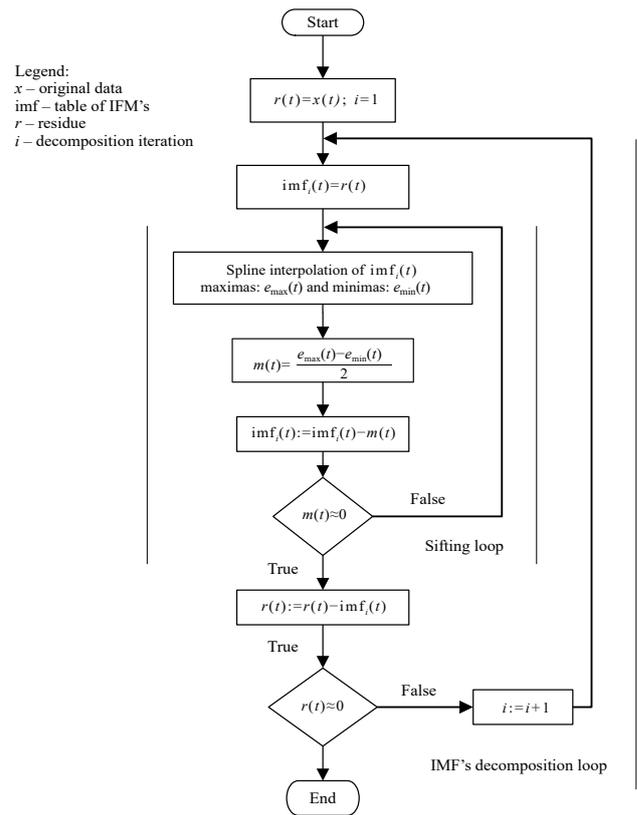

Fig. 1. Block diagram of sifting process
(from (PIERZCHALSKI *et al.*, 2011)).

($m(t) \approx 0$). After the extraction of IMF_1, the original data is reduced by the ultimate value of $\mathrm{imf}_1(t)$:

$$r(t) = x(t) - \mathrm{imf}_1(t). \tag{11}$$

The residue $r(t)$ is treated as input for extracting the subsequent IMF (next sifting loop).
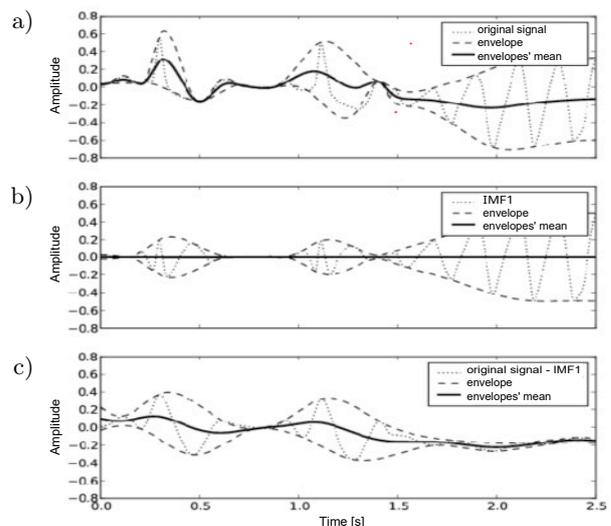

Fig. 2. EMD decomposition: a) start; b) end of the decomposition of the first IMF; c) start of the decomposition for the second IMF (from (PIERZCHALSKI *et al.*, 2011)).

The procedure is looped to obtain all IMFs (Figs. 1 and 2). Decomposition is finished when either the i-th residue $r_i(t) = r_{i-1}(t) - \text{imf}_i(t)$ has less than three extrema or all its points are equal to zero.

The sum of all IMF components (modes) and the residue is equal to the original signal:

$$r_n + \sum_{m=1}^{n} \text{IMF}_m(t) = x(t), \tag{12}$$

where n is the number of modes.

In most analyzes using EMD, researchers focus on the modes themselves, ignoring the residue, which for many complex signals, has tiny amplitudes. However, in this study, the amplitudes of the residues are much higher than the amplitudes of the modes. Thus, the residues give the most crucial information about the signal.

4. Results

The results presented here are for five guitars that underwent the humidification process for nearly a week. The guitar designations are random and do not refer to any specific type or model of guitar. The analyzed data were obtained using the measurement and processing methods described in (Wrzeciono *et al.*, 2018).

4.1. Analysis of sound level during sustain by Higuchi's fractal dimension

Table 1 shows the derived fractal dimension values for the instruments before and after moisturizing. All tested guitars exhibited a higher fractal dimension after the humidification process than before. This means that the SPL curve for the instrument after humidification is more complex than before.

Table 1. HDF for five guitars before and after humidification.

The guitar ID	Before moisturizing	After moisturizing
110	2.12	2.18
111	2.08	2.17
112	2.01	2.2
113	2.08	2.12
114	2.19	2.22

It should be noted that the levels of fractal dimension values attained by guitars depend not only on their initial condition but also on individual features of their construction. For example, the fractal dimension value for guitar 110 before humidification is the same as for the fractal dimension of guitar 113 after humidification. However, after humidification, the fractal dimension for guitar 110 reaches the level of this value for guitar 114.

Moreover, it is observed that the level of increment in fractal dimension depends on the guitar's starting condition and susceptibility to moisturizing. The most significant increases in fractal dimension occurred for guitars 112 and 111, while for guitars 113 and 114, the increment was the smallest.

It is also intriguing that the fractal dimension values exceed the value of 2. This is probably due to the properties of the curve of the sound level during sustain, where a strong nonlinear trend is superimposed on a rapidly varying oscillation. This trend is essential in interpreting how we hear the guitar sound. We write about this further in the results section on EMD analysis.

4.2. Symbolic analysis of sustain curve

The sum of the means lengths of mono-sequences (SML), similarly to HFD, records the difference before and after conditioning the instrument (Fig. 3). For this parameter, we observe a decrease in value after conditioning. This means that the amplitudes of the fast oscillations are statistically shorter and become more uniform after humidification. The complexity of sustain curves grows after moisturizing, which agrees with the results obtained with HFD.
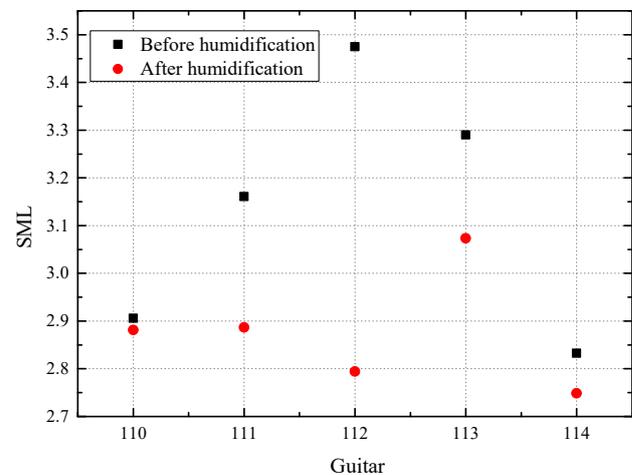

Fig. 3. Values of SML score for five instruments.

Using SML, we tracked the changes in guitars during the humidification process. Figure 4 shows the evolution of the SML scores for guitars from the first before conditioning to the last after an entire humidification cycle.

The effect of the instrument humidification process is irregular – improvement is followed by deterioration. This agrees with listening observations. The evolution during conditioning resembles a fading oscillation, indicating that guitars are moving towards their characteristic equilibrium points.
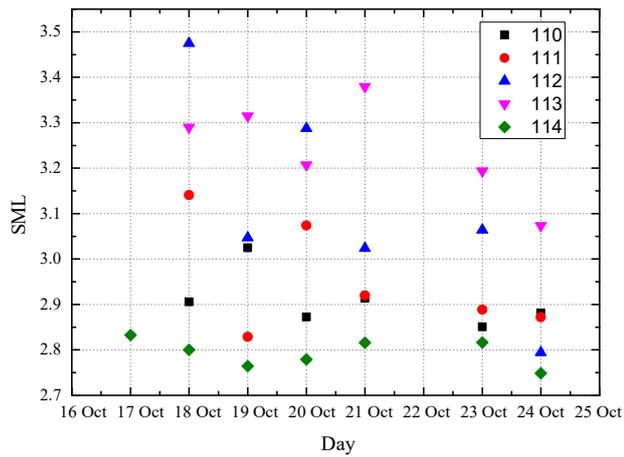
Fig. 4. Evolution of SML scores during instrument humidification.

4.3. EMD decomposition of the sound level in sustain

EMD decomposed the sound level during sustain of the guitar before and during the moisturizing procedure. Figure 5 presents the decomposition of the SPL curve for guitar 112 before and after the complete cycle of moisturizing.

For guitar 112, the number of modes did not change. However, evident changes can be seen in the

shape of the mode waveforms. Here, the shape of the sustain signal is mainly affected by mode IMF_5. The bulges observed in the sustain signal of the guitar before the humidification process, audible as long ripples of sound, are associated with mode IMF_5.

In the seven decompositions, the final number of modes was five; in two cases, it was six, and in one case, it was four. Thus, for two guitars, after conditioning, the number of modes increased from five to six (guitar 111) and four to five (guitar 113, see Fig. 6); for one guitar, it decreased from six to five (guitar 114).

Especially interesting are the results of the residue decomposition of the signal. For example, residues of SPL measurements before and after humidification for guitar 110 are presented in Fig. 7.

Figure 8 shows the changes in the shapes of the residue curves determined for the instruments during the humidification process.

The shape of these curves matches the listening experience and confirms the irregular changes in instrument sound quality during humidification. In addition, an identical irregularity was observed for the previously determined SML parameter.

Through our investigation, we observed significant alterations in the playability of instruments following the humidification process, as indicated by all the methods employed.

a)

b)
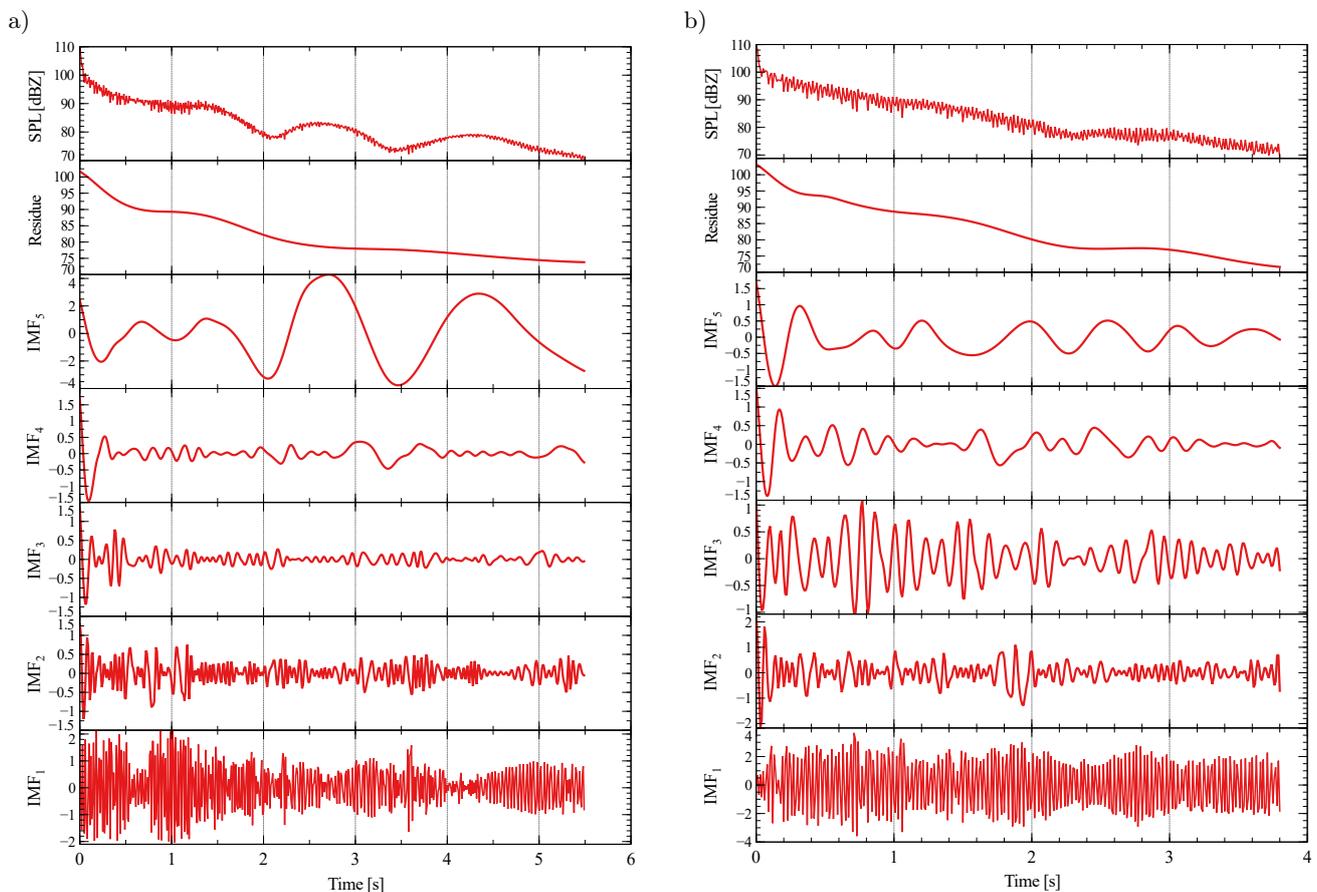
Fig. 5. EMD decomposition of the SPL for guitar 112: a) before moisturizing; b) after moisturizing.
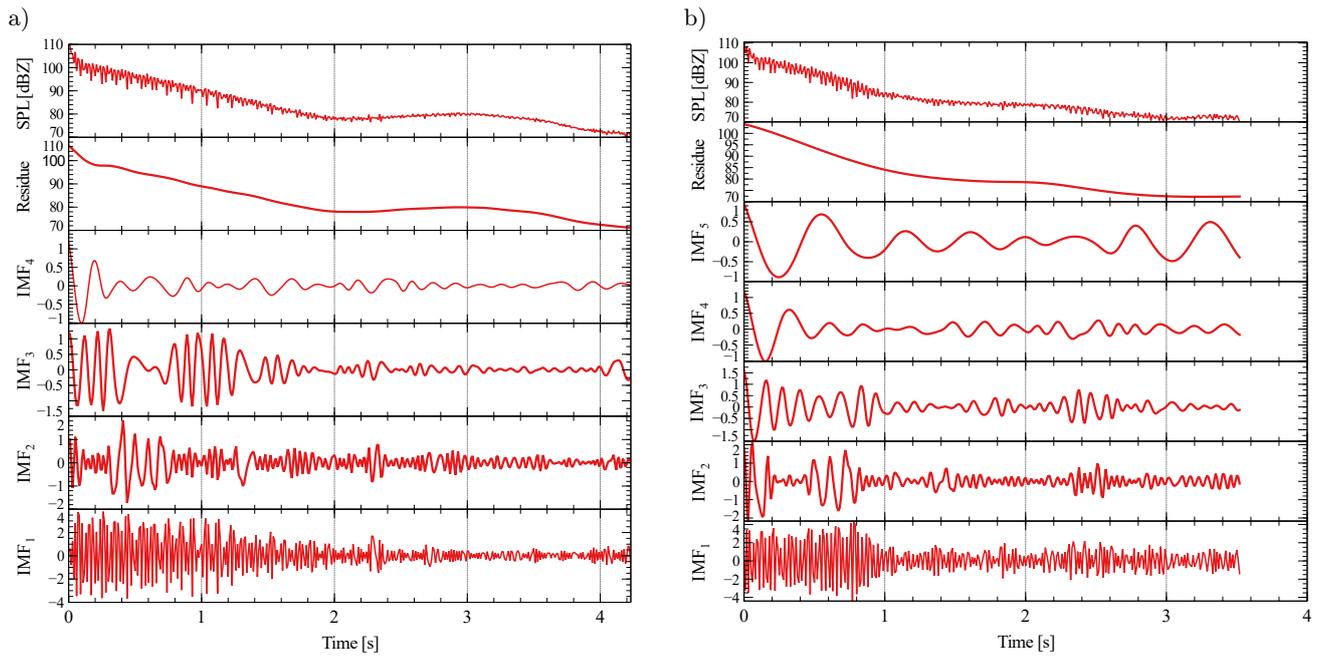
Fig. 6. EMD decomposition of the SPL for guitar 113: a) before moisturizing; b) after moisturizing.
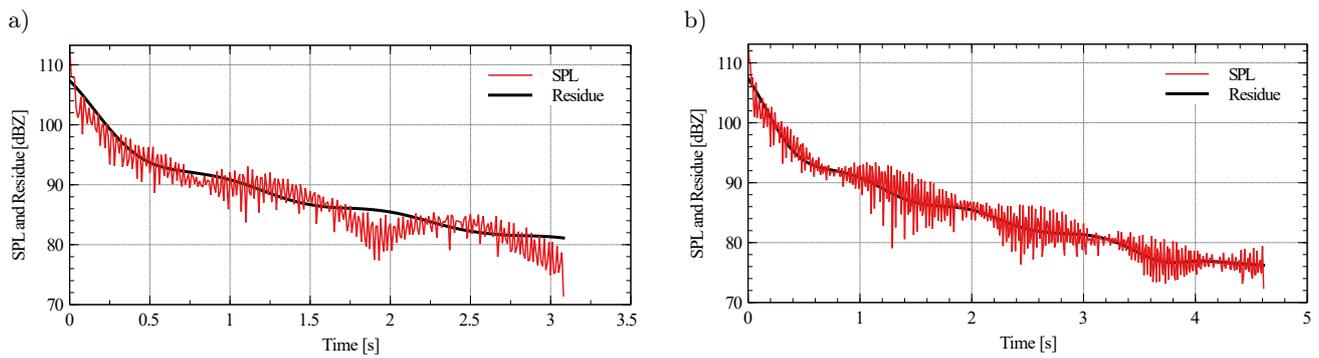

Fig. 7. Residues and sustain of 110 guitar (a) before moisturizing and (b) after moisturizing.


[Fig. 8ab.]

c)

d)

e)
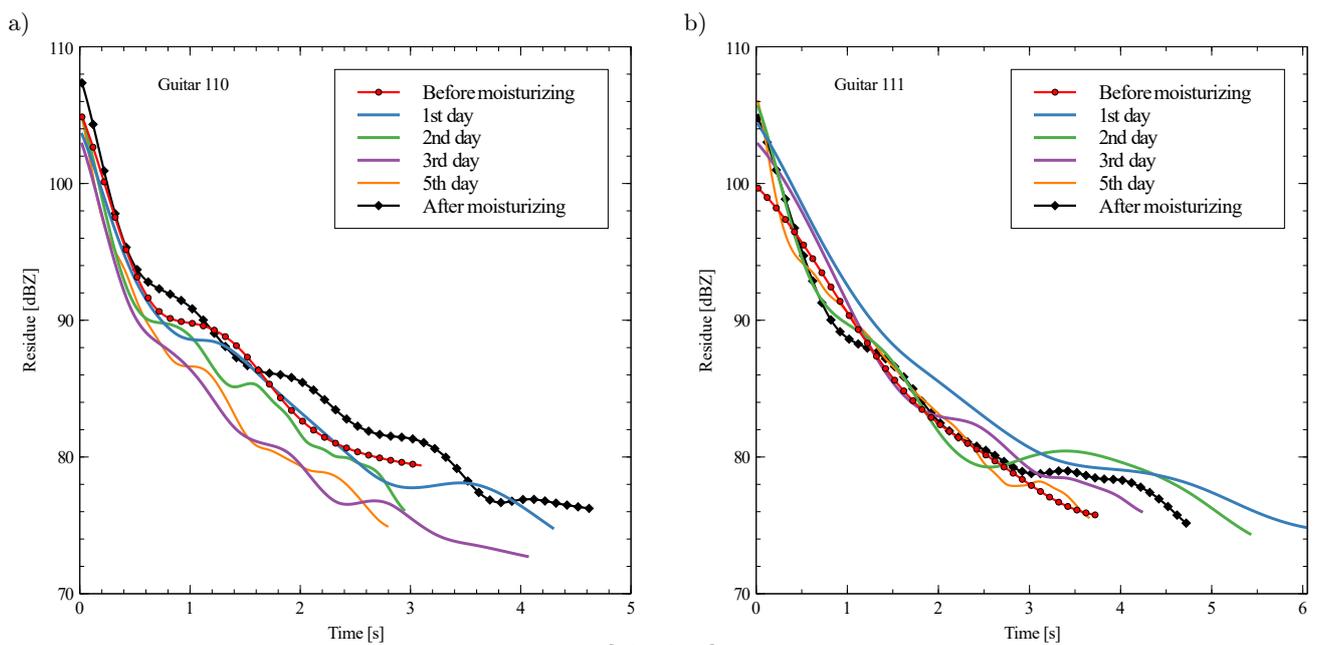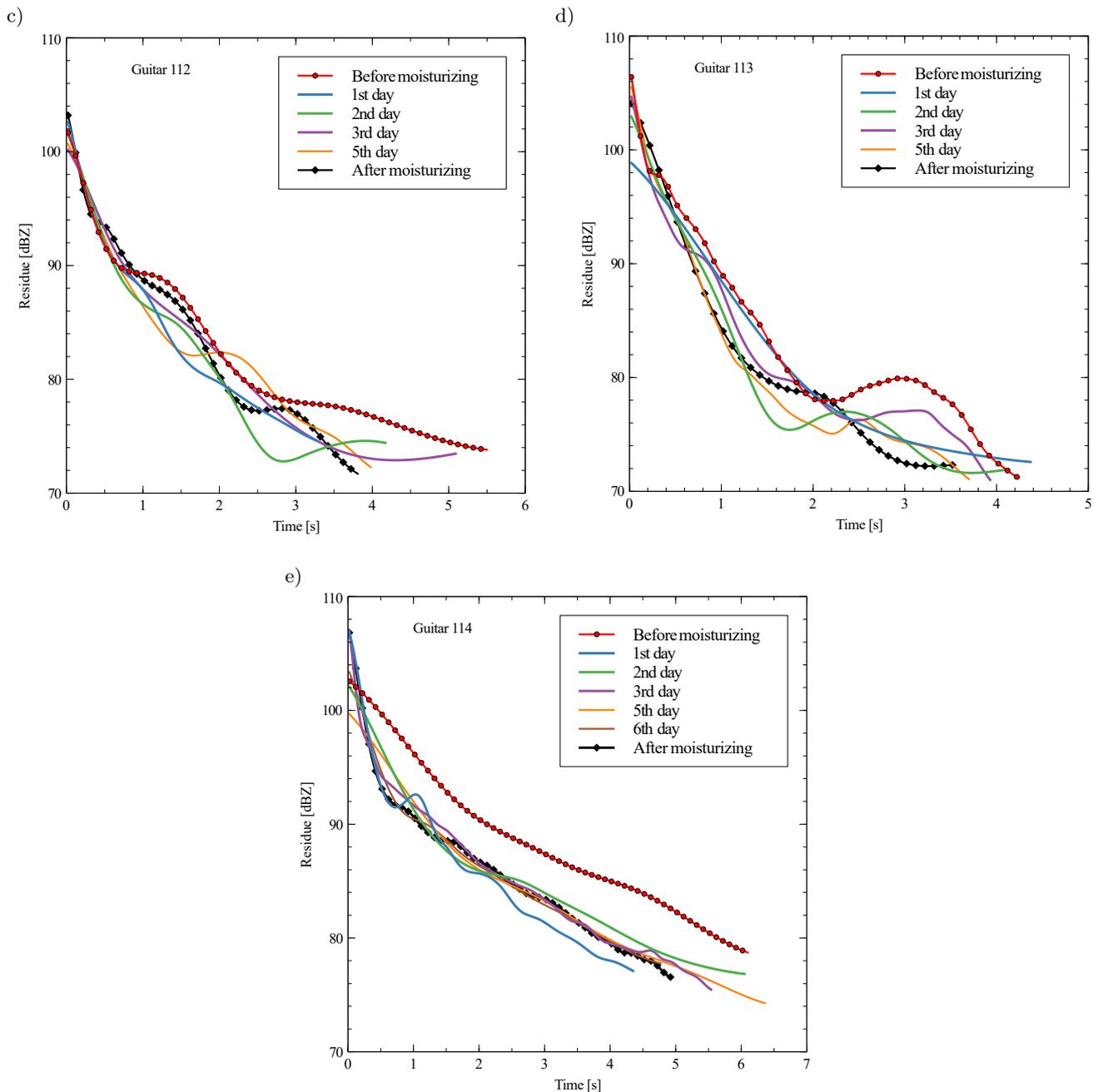
Fig. 8. Residues after the EMD decomposition of SPL measured during moisturizing procedure. The first and last measures are curves with dots.

Previously, a solution to the problem of objective assessment of the influence of the humidification procedure on guitar sound quality was proposed together with the original measurement procedure by WRZE-CIONO *et al.* (2018). Despite having an effective measurement method, the problem turned out to be nontrivial, and the methods of sound level analysis proposed in (WRZECIONO *et al.*, 2018) did not yield entirely satisfactory results. As a result, only the sustain time parameter T_{40} was suitable for guitar condition evaluation. Unfortunately, the disadvantage of this parameter is its excessive sensitivity to changes in the signal level cut-off moment. Therefore, in this work, more global methods that use all measured points on the sound level curve, or as in the case of EMD, generate curves containing deeper information, were used to evaluate sound level changes.

5. Conclusion

The purpose of this study was to find parameters for evaluating guitar humidification performance, and we found that complexity parameters like HFD, symbolic analysis, and EMD provide a consistent and clear

depiction of the changes in guitar sound quality during the humidification process. To the best of the authors' knowledge, this is probably the first application of these methods to evaluate guitar humidification performance.

References

1. DE BOOR C. (1978), *A Practical Guide to Splines*, Springer-Verlag, New York.

2. GLADUN K.V. (2020), Higuchi fractal dimension as a method for assessing response to sound stimuli in patients with diffuse axonal brain injury, *Modern Technologies in Medicine*, **12**(4): 63–70, doi: 10.17691/stm 2020.12.4.08.

3. GOMOLKA R.S., KAMPUSCH S., KANIUSAS E., THÜRK F., SZÉLES J.C., KLONOWSKI W. (2018), Higuchi fractal dimension of heart rate variability during percutaneous auricular vagus nerve stimulation in healthy and diabetic subjects, *Frontiers in Physiology*, **9**: 1162, doi: 10.3389/fphys.2018.01162.

4. GORDOBIL O., HERRERA R., POOHPHAJAI F., SANDAK J., SANDAK A. (2021), Impact of drying process on kraft lignin: Lignin-water interaction mechanism study by 2D NIR correlation spectroscopy, *Journal of Materials Research and Technology*, **12**: 159–169, doi: 10.1016/j.jmrt.2021.02.080.

5. GORE T. (2011), Wood for guitars, [in:] *Proceedings of Meetings on Acoustics 161ASA*, **12**(1): 035001.

6. HIGUCHI T. (1988), Approach to an irregular time series on the basis of the fractal theory, *Physica D: Nonlinear Phenomena*, **31**(2): 277–283, doi: 10.1016/0167-2789(88)90081-4.

7. HUANG N.E. *et al.* (1998), The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **454**(1971): 903–995, doi: 10.1098/rspa.1998.0193.

8. HUANG N.E., WU Z. (2008), A review on Hilbert-Huang transform: Method and its applications to geophysical studies, *Reviews of Geophysics*, **46**(2), doi: 10.1029/2007RG000228.

9. JOHNSON Jr P.D., HARRIS G.A., HANKERSON D. (2003), *Introduction to Information Theory and Data Compression*, 2nd ed., CRC.

10. KHAN S.I., PACHORI R.B. (2021), Automated classification of lung sound signals based on empirical mode decomposition, *Expert Systems with Applications*, **184**: 115456, doi: 10.1016/j.eswa.2021.115456.

11. LI C. *et al.* (2021), Seizure onset detection using empirical mode decomposition and common spatial pattern, [in:] *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **29**: 458–467, doi: 10.1109/TNSRE.2021.3055276.

12. MANDELBROT B. (1967), How long is the coast of Britain? Statistical self-similarity and fractional dimension, *Science*, **156**(3775): 636–638.

13. PIERZCHALSKI M., STEPIEN R.A., STEPIEN P. (2011), New nonlinear methods of heart rate variability analysis in diagnostics of atrial fibrillation, *International Journal of Biology and Biomedical Engineering*, **5**(4): 201–208.

14. RATH J., STAUDINGER G. (2001), Cracking reactions of tar from pyrolysis of spruce wood, *Fuel*, **80**(10): 1379–1389, doi: 10.1016/S0016-2361(01)00016-3.

15. RIBEIRO H.V., JAUREGUI M., ZUNINO L., LENZI E.K. (2017), Characterizing time series via complexity-entropy curves, *Physical Review E*, **95**(6): 062106, doi: 10.1103/PhysRevE.95.062106.

16. RIBEIRO H.V., ZUNINO L., MENDES R.S., LENZI E.K. (2012), Complexity-entropy causality plane: A useful approach for distinguishing songs, *Physica A: Statistical Mechanics and its Applications*, **391**(7): 2421–2428, doi: 10.1016/j.physa.2011.12.009.

17. SPASIĆ S., KALAUZI A., ĆULIĆ M., GRBIĆ G., MARTAĆ L. (2005), Estimation of parameter k_{max} in fractal analysis of rat brain activity, *Annals of the New York Academy of Sciences*, **1048**(1): 427–429, doi: 10.1196/annals.1342.054.

18. STEPIEN R.A. (2011), New method for analysis of non-stationary signals, *Nonlinear Biomedical Physics*, **5**(1): 1–8, doi: 10.1186/1753-4631-5-3.

19. STOJADINOVIĆ G. *et al.* (2020). The effects of Nembutal on the intracerebellar EEG activity revealed by spectral and fractal analysis, *Archives of Biological Sciences*, **72**(3): 425–432.

20. STONE J.V. (2022), *Information Theory: A Tutorial Introduction*, 2nd ed., Sebtel Press.

21. WRZECIONO P., PRZYBYLSKI W., OSTROWSKI A. (2018), Guitar Sound Parameters to assess the impact of atmospheric conditions on the acoustic qualities of the instrument, [in:] *2018 Joint Conference – Acoustics*, pp. 1–5, doi: 10.1109/ACOUSTICS.2018.8502376.

22. ZHENG J., SU M., YING W., TONG J., PAN Z. (2021), Improved uniform phase empirical mode decomposition and its application in machinery fault diagnosis, *Measurement*, **179**: 109425, doi: 10.1016/j.measurement.2021.109425.

23. ZHOU S., BETHEL B.J., SUN W., ZHAO Y., XIE W., DONG C. (2021), Improving significant wave height forecasts using a joint empirical mode decomposition–long short-term memory network, *Journal of Marine Science and Engineering*, **9**(7): 744, doi: 10.3390/jmse9070744.

Research Paper

Research on the Motion Features Model for Underwater Targets with Multiple Highlights and Multiple Micro-Motion Forms

Tongjing SUN*, Zihan ZHOU, Dongliang PENG

Department of Automation
Hangzhou Dianzi University, Xiasha Higher Education Zone
Hangzhou, China

*Corresponding Author e-mail: stj@hdu.edu.cn

Motion characterization, including Doppler and micro-Doppler, is crucial for the detection and identification of high-speed underwater targets. Under high-frequency and short-range conditions, underwater targets cannot be simply regarded as single highlight targets as they exhibit a complex structure with multiple scattering centers accompanied by distinct micro-motions. To address this multi-highlight and multi-micro-motion scenario, a model is proposed to characterize the motion features of underwater targets. Firstly, a mathematical model is established to represent the micro-Doppler features based on the single-highlight model. Subsequently, considering the overlap of multiple highlight echoes caused by the high-speed translation of the target and the long pulse detection signal, precise representation is achieved by setting motion positions and calculating time delays within the model. The results represent the echoes of moving targets with multiple highlights and micro-motions. Finally, a time-frequency analysis method is employed to extract motion features and estimate target parameters, thereby validating the accuracy and effectiveness of the proposed model. This research provides a theoretical foundation for the modeling of underwater moving targets.

Keywords: micro-motion; complex motion; micro-Doppler; underwater micro-motion model; multi-highlight model.

1. Introduction

Existing studies have shown that underwater targets, during navigation or operational activities, exhibit periodic micro-motions such as vibrations and rotations alongside their translation (CLEMENTE *et al.*, 2013; HANIF *et al.*, 2022). Micro-Doppler features, serving as a crucial representation of target motion states, are particularly important in high-speed target detection. By extracting micro-Doppler features from complex echo signals, it is possible to estimate motion parameters of the high-speed target's micro-motion components, which is of great significance for target detection and identification.

In the field of radar detection, extensive research has been conducted on target micro-motion features, micro-motion feature extraction, and classification, leading to valuable achievements (CLEMENTE *et al.*, 2013; HANIF *et al.*, 2022). CHEN *et al.* (2003)

elucidated the frequency modulation effect caused by target micro-motion, known as micro-Doppler. Micro-Doppler features have been used for classifying various human activities. By using raw micro-Doppler signatures as features, detection and recognition of various human activities can be achieved (KIM, LING, 2009; KIM, MOON, 2016). WANG *et al.* (2023) employed time-frequency analysis to extract micro-Doppler features from radar signals to identify the indoor activities of elderly people, enabling an effective assessment of potential risks in their daily routines. ZHAO and SU (2023) decomposed Doppler signals from the echo of small unmanned aerial vehicles based on the micro-Doppler effect of rotating targets. They further extracted motion parameters from the residual rotating signals, achieving efficient identification of LSS UAVs. In addition to civilian applications, relevant research has also been conducted in the military domain. For instance, based on the estimation of

target micro-motion parameters, the identification of warheads has been accomplished, taking into account the different forms of target micro-motions (Gao *et al.*, 2010; Han, Feng, 2020; Zhang *et al.*, 2023).

In underwater acoustics, most research has concentrated on estimating motion parameters and position parameters of targets through Doppler parameters, enabling measurements such as velocity and range estimation (Xu, 2016; Tang *et al.*, 2020). Zhang *et al.* (2018) proposed a method based on frequency-difference-of-arrival measurements to accurately infer the position and velocity of underwater targets by considering the Doppler effect. Gong *et al.* (2020) introduced a low-complexity Doppler estimation algorithm to estimate the Doppler frequency shift and achieve localization of autonomous underwater vehicles. Yang *et al.* (2023) estimated the Doppler frequency shift to obtain the motion states of scatterers and then fused the motion features to accurately identify multiple moving scatterers within the same beam. Regarding micro-Doppler feature research, Kashyap *et al.* (2015) simulated underwater vehicles with rotating propellers using sonar and radar detection. Employing time-frequency analysis to extract micro-Doppler features, the author confirmed that micro-Doppler features can be used for target identification. However, the presented simulation models lack generality. Kulhandjian *et al.* (2020) classified and recognized the swimming postures of humans on the water surface using acoustic micro-Doppler features. Wu *et al.* (2022) simulated the motion features of composite targets with translational and micro-motion components, using a single-highlight model, and obtained the target's motion features through time-frequency analysis. Kou and Feng (2022) focused on targets with various micro-motion forms in a static state. The author separated multi-point echoes and extracted the micro-motion features of each highlight by constructing a redundant dictionary and sparse decomposition. Saffari *et al.* (2023) effectively selected features by extracting micro-Doppler features to distinguish and identify the propellers of various underwater target models (in stationary states) based on their different models and motion states.

Based on the above, most of the existing research models primarily consider micro-motion features in a stationary state and perceive the target as a single highlight structure. However, in actual underwater operational scenarios, targets often exhibit a combined motion pattern of translation and micro-motion, typically presenting complex structures with multiple scattering centers. For instance, in the case of a torpedo moving at high speed underwater, the tail fin undergoes a rotational motion and the engine compartment exhibits vibration. In light of this scenario, this paper builds upon the highlight model and establishes a motion characteristic model that incorporates multi-

ple highlight and micro-motions for combined translation and micro-motion. The feature parameters are extracted using time-frequency analysis methods to verify the correctness and effectiveness of the established model.

2. Mathematical model of micro-motion features based on single highlight

The single-highlight model for underwater targets (Tang, 1994) assumes that the relative distance between a moving target and the sonar remains constant during the pulse width of the transmitted signal. However, when underwater targets exhibit high-speed translation along with micro-motion, such as rotation and vibration, the position of the target's highlights changes not only due to translation but also due to micro-motion. This results in significant changes in the relative distance between the sonar and the target, which should be represented as an instantaneous distance $\mathbf{R}(t)$.

Assuming the transmitted signal is a long pulse signal:

$$p(t) = p_0(t)e^{-j2\pi f t}, \tag{1}$$

where $p_0(t)$ is the envelope of the signal and f is the carrier frequency.

During the pulse width of the transmitted signal illuminating the target, the highlight of the target is established as a dynamic model, represented as a vector indicating the relative distance $\mathbf{R}(t)$.

The vector distance relationship between the sonar and the micro-motion target is illustrated in Fig. 1. The sonar is located at the origin O of the coordinate system (x, y, z) and remains stationary. At time $t = 0$, the highlight of the target is located at the origin O_1 of the coordinate system (x_1, y_1, z_1), which is a translation of the (x, y, z) coordinate system. The initial distance vector between the sonar and the target is denoted as $\mathbf{R}_0$. The highlight moves uniformly with a velocity vector $\mathbf{V}$, and simultaneously, the micro-motion of the highlight introduces a distance change vector $\mathbf{M}(t)$. $\mathbf{M}(t)$ varies depending on the specific micro-motion pattern.
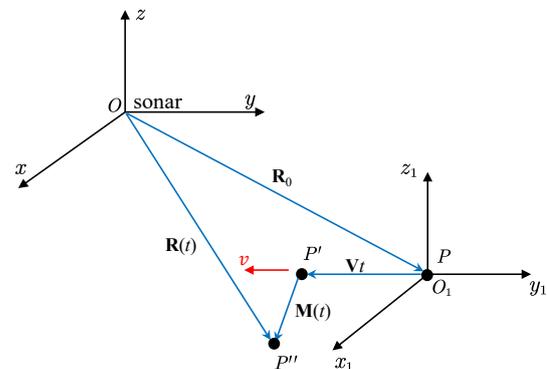

Fig. 1. Sonar and micro-motion target vector relationship.

The motion of the target can be decomposed into two stages: first, the target moves from point P to point P', where the vector distance of this segment is equal to the translational velocity vector multiplied by time; then, based on the micro-motion pattern, it moves from point P' to point P'', where the vector distance of this segment varies based on the form of infinitesimal motion. By adding these vectors together, we can obtain the instantaneous distance $\mathbf{R}(t)$:

$$\mathbf{R}(t) = \|\mathbf{R}_0 + \mathbf{V}t + \mathbf{M}(t)\|. \quad (2)$$

According to the highlight model theory, the time delay τ can be expressed as:

$$\tau = \frac{2\mathbf{R}(t)}{c}. \quad (3)$$

The echo signal expression is:

$$s(t) = Ap_0\left(t - \frac{2\mathbf{R}(t)}{c}\right)e^{-j2\pi f\left(t - \frac{2\mathbf{R}(t)}{c}\right)}e^{j\varphi}$$

$$= Ap_0\left(t - \frac{2\mathbf{R}(t)}{c}\right)e^{-j2\Phi(\mathbf{R}(t))}e^{j\varphi}, \quad (4)$$

where A is the amplitude of the echo from the highlight, φ is the phase change at the formation of the echo, and $\Phi(\mathbf{R}(t))$ is the phase of the echo signal. By taking the derivative of the phase, we can obtain the instantaneous frequency:

$$f_t = \frac{1}{2\pi}\frac{d\Phi(\mathbf{R}(t))}{dt} = f - \frac{2f}{c}\frac{d\mathbf{R}(t)}{dt} = f - f_t', \quad (5)$$

$$f_t' = f_d + f_{md} = \frac{2f}{c}[\mathbf{V}^T \cdot \mathbf{n}] + \frac{2f}{c}\left[\frac{d}{dt}(\mathbf{M}(t))\right]^T \cdot \mathbf{n}, \quad (6)$$

where $\mathbf{n} = (\mathbf{R}_0 + \mathbf{V}t + \mathbf{M}(t))/\|(\mathbf{R}_0 + \mathbf{V}t + \mathbf{M}(t))\|$ is the unit vector of $\mathbf{OP}''$, and since the initial distance is much larger than the distance generated by the target within one pulse period, we can approximate the unit vector of $\mathbf{OP}''$ as the unit vector of $\mathbf{OP}$, denoted as $\mathbf{n} \approx \mathbf{n}_p = \mathbf{R}_0/\|\mathbf{R}_0\|$.

From the above derivation, we can see that $f_d = \frac{2f}{c}[\mathbf{V}^T \cdot \mathbf{n}]$ is the Doppler frequency shift, indicating that the echo signal form with Doppler frequency shift can be obtained through the dynamic representation of time delay. On the other hand, $f_{md} = \frac{2f}{c}\left[\frac{d}{dt}(\mathbf{M}(t))\right]^T \cdot \mathbf{n}$ is the frequency shift caused by the frequency modulation resulting from micro-motion, known as the micro-Doppler frequency shift. Its essence is similar to the Doppler features, as both are caused by the change in the relative distance between the sonar and the target due to the target's motion, which leads to changes in the echo phase.

3. Motion features model for underwater targets with multiple highlights and micro-motion forms

For complex volumetric targets, a single highlight is insufficient to encompass their multifaceted features. Therefore, it is necessary to consider them as a collection of multiple highlights, each exhibiting diverse micro-motion patterns. Taking a high-speed moving torpedo target as an example, a composite motion feature model is constructed to encompass multiple highlights and their different micro-motion patterns.

The torpedo is in a high-speed state underwater. Consider a five-highlight model for the torpedo. The head of the torpedo represents a stationary highlight without any micro-motion. The middle section contains an engine that causes mechanical vibrations in the body. Finally, the tail of the torpedo consists of three propeller blades that can be seen as three highlights rotating around the rear of the body.

The geometric relationship between the torpedo target and the observing sonar is illustrated in Fig. 2. Parts A, B, and C in Fig. 2d correspond to Figs. 2a–c, respectively. The observing sonar is located at the origin O of the sonar coordinate system (x, y, z). Each highlight has its independent coordinate system relative to the sonar coordinate system, denoted as (x_1, y_1, z_1), (x_2, y_2, z_2), and (x_3, y_3, z_3). These coordinate systems are translations of the sonar coordinate system (x, y, z). The distribution and motion of these highlights are as mentioned above, representing the head, middle, and tail of the target, respectively. The highlight representing the middle section undergoes harmonic vibrations, with the origin O_2 of its coordinate system as the center of oscillation. The three highlights representing the tail section rotate around the origin O_3 of their coordinate system, with the y-axis as the rotation axis. They rotate with the same radius and angular velocity. The initial positions of these highlights in the coordinate system (x_3, y_3, z_3) are different. The entire target moves uniformly along the negative y-axis direction with a speed of v.

For a stationary highlight, as shown in Fig. 2a, the position of the highlight at the initial moment is P, which is located at the origin O_1 of the target coordinate system (x_1, y_1, z_1). Its position vector in the sonar coordinate system (x, y, z) is $\mathbf{R}_1 = (X_1, Y_1, Z_1)^T$, and the initial azimuth and pitch angles are α_1 and β_1, respectively. The radial unit vector of the sonar to the target is:

$$\mathbf{n}_p = \mathbf{R}_1/\|\mathbf{R}_1\| = (\cos\alpha_1\cos\beta_1, \sin\alpha_1\cos\beta_1, \sin\beta_1)^T. \quad (7)$$

The velocity vector of the highlight is $\mathbf{V} = (0, v, 0)^T$. The instantaneous distance and echo signal expression of the highlight are:

$$\mathbf{R}_1(t) = \|\mathbf{R}_1 + \mathbf{V}_t\| = \sqrt{X_1^2 + (Y_1 + vt)^2 + Z_1^2}, \quad (8)$$

$$s_1(t) = A_1p_0\left(t - \frac{2\mathbf{R}_1(t)}{c}\right)e^{-j2\omega_c\left(t - \frac{2\mathbf{R}_1(t)}{c}\right)}e^{j\varphi_1}. \quad (9)$$
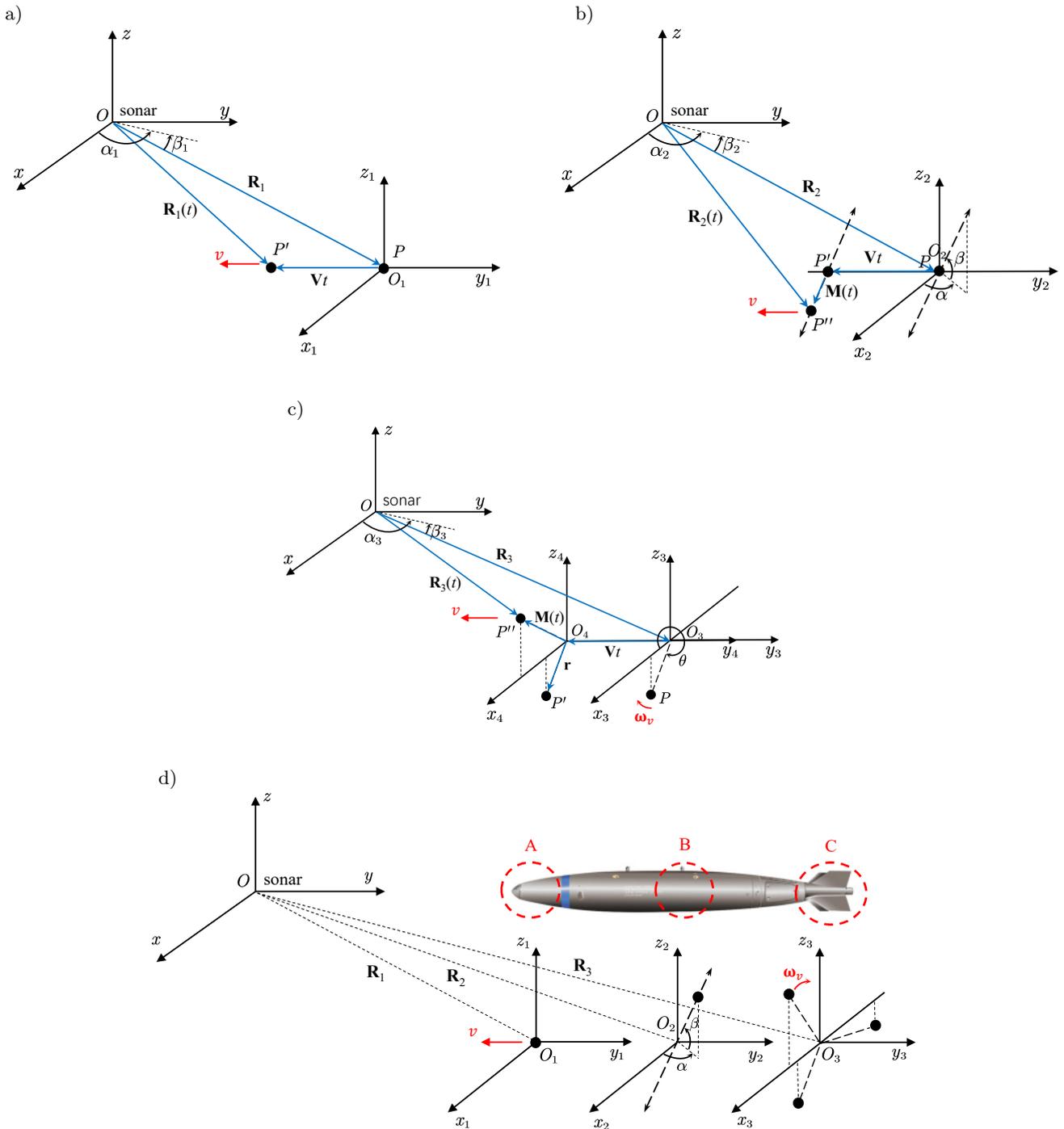
Fig. 2. The geometric relationship in the model: a) stationary high-light at position A; b) vibrating high-light at position B; c) rotating highlight at position C; d) geometric relationship between sonar and target.

It can be seen that there is no micro-Doppler; the only Doppler frequency shift is:

$$f_d = \frac{2f}{c}\left[\mathbf{V}^{\mathrm{T}} \cdot \mathbf{n}_p\right] = \frac{2f}{c}\left(v \sin \alpha_1 \cos \beta_1\right). \quad (10)$$

For the vibrating highlight, as shown in Fig. 2b, assume that the highlight is initially located at position P, which corresponds to the origin O_2 in the target coordinate system (x_2, y_2, z_2). The position vector

of the highlight in the sonar coordinate system (x, y, z) is denoted as $\mathbf{R}_2 = (X_2, Y_2, Z_2)^{\mathrm{T}}$. The initial azimuth angle and pitch angle are represented as α_2 and β_2, respectively. The unit vector in the radial direction from the sonar to the vibrating center O_2 of the target highlight takes the same form as Eq. (7).

The origin O_2 is the vibration center. The harmonic vibration is performed with frequency f_v and amplitude D_v. The azimuth angle of the vi-

bration direction is α, and the pitch angle is β. The unit vector of the vibration direction is $\mathbf{n}_v = (\cos\alpha\cos\beta, \sin\alpha\cos\beta, \sin\beta)^{\mathrm{T}}$. In the form of simple harmonic vibration, the vibration distance is $D_t = D_v\sin(2\pi f_v t)$. And then the instantaneous distance is:

$$\mathbf{R}_2(t) = \|\mathbf{R}_2 + \mathbf{V}t + D_t \cdot \mathbf{n}_v\|$$

$$= \big[(X_2 + D_t\cos\alpha\cos\beta)^2$$

$$+ (Y_2 + vt + D_t\sin\alpha\sin\beta)^2$$

$$+ (Z_2 + D_t\sin\beta)^2\big]^{1/2}. \qquad (11)$$

Echo signal expression of the highlight is:

$$s_2(t) = A_2 p_0\left(t - \frac{2\mathbf{R}_2(t)}{c}\right)e^{-j2\pi f\left(t - \frac{2\mathbf{R}_2(t)}{c}\right)}e^{j\varphi_2}. \quad (12)$$

The Doppler is the same as the stationary highlight, and the micro-Doppler is:

$$f_{md} = \frac{2f}{c}\left[\frac{d}{dt}(D_v\sin(2\pi f_v t)\cdot\mathbf{n}_v)\right]^{\mathrm{T}}\cdot\mathbf{n}_p$$

$$= \frac{4\pi f_v f D_v}{c}\cos(2\pi f_v t)$$

$$\cdot[\cos(\alpha_2 - \alpha)\cos\beta_2\cos\beta + \sin\beta_2\sin\beta]. \quad (13)$$

In summary, the micro-Doppler curve of a vibrating feature can be represented by a sinusoidal curve. The period of the micro-Doppler curve corresponds to the vibration period, and its amplitude is related to the vibration amplitude, vibration period, and the carrier frequency of the transmitted signal.

For rotating highlights, the three highlights differ only in their initial positions. By analyzing the general geometric relationship of a single highlight under motion, we can determine the motion features of three highlights. The rotation radius is l. The rotational angular velocity is $\boldsymbol{\omega} = (0, w, 0)$. The initial rotation angle of the rotation highlight at $t = 0$ is $\theta_i(i = 1, 2, 3)$. Therefore, the initial distance vector of the rotating highlights in the coordinate system (x_3, y_3, z_3) at $t = 0$ can be expressed as:

$$\mathbf{r}_i = (l\cos\theta_i, \ 0, \ l\sin\theta_i)^{\mathrm{T}}. \qquad (14)$$

For the motion analysis of a rotating highlight, its geometric relationship with the sonar system is depicted in Fig. 2c. At $t = 0$, the initial position of the target highlight is located at point P, and its initial rotation angle is denoted as θ. The initial position vector of the target highlight in the target coordinate system (x_3, y_3, z_3) is represented as $\mathbf{r} = (l\cos\theta, 0, l\sin\theta)^{\mathrm{T}}$. The origin O_3 of the target coordinate system (x_3, y_3, z_3) has a position vector $\mathbf{R}_3 = (X_3, Y_3, Z_3)^{\mathrm{T}}$ with respect to the sonar coordinate system (x, y, z). The initial azimuth angle and pitch angle of the target coordinate system

are denoted as α_3 and β_3, respectively. The unit vector in the radial direction from the sonar to the rotating center O_3 of the target highlight takes the same form as Eq. (7).

According to Rodrigues' formula (CHEN *et al.*, 2003), the instantaneous distance is:

$$\mathbf{R}_3(t) = \|\mathbf{R}_3 + \mathbf{V}t + \mathbf{R}_t(t)\mathbf{r}\|$$

$$= \big[(X_3 + l\cos\theta\cos\omega t + l\sin\theta\sin\omega t)^2$$

$$+ (Y_3 + vt)^2$$

$$+ (Z_3 - l\cos\theta\sin\omega t + l\sin\theta\cos\omega t)^2\big]^{1/2}. \quad (15)$$

$\mathbf{R}_t(t)$ is the rotation matrix. The expression of the echo signal is obtained as:

$$s_3(t) = A_3 p_0\left(t - \frac{2\mathbf{R}_3(t)}{c}\right)e^{-j2\omega_c\left(t - \frac{2\mathbf{R}_3(t)}{c}\right)}e^{j\varphi_3}. \quad (16)$$

When placing the target highlight and the sonar in the same plane, i.e., $X_3 = 0$, and applying an approximation method (CHEN, 2014), the specific expression for micro-Doppler can be derived as:

$$f_{md} = \frac{2f\omega l}{c}\sin\beta_3\sin(\omega t + \theta_i). \qquad (17)$$

In summary, the micro-Doppler curve of a rotating highlight follows a sinusoidal pattern, similar to that of a vibrating highlight. The period of the micro-Doppler curve corresponds to the rotational period of the motion. The amplitude of the micro-Doppler curve is influenced by the rotational radius, the rotation period, and the carrier frequency of the transmitted signal.

For the other two rotating highlights, their forms are similar. The final received echo signal form for a multi-highlight moving target is:

$$s(t) = \sum_{i=1}^{5} s_i(t). \qquad (18)$$

According to the traditional underwater target model (DONG *et al.*, 2013), the relative distances between the highlights of a target can result in different temporal distributions of their respective echoes. The relative distances between the highlights can also vary during the period of signal transmission due to the high-speed state of the target. Additionally, to capture finer micro-Doppler features, it is necessary to transmit detection signals with longer pulse widths, which ultimately leads to temporal overlap of the echoes from different highlights.

Since the initial moment corresponds to the time of signal transmission, the starting time of the echo signal is denoted as t_1, and it is given by the equation:

$$t_1 = \tau_1 = \frac{2\mathbf{R}(t_1)}{c}, \qquad (19)$$

where $\mathbf{R}(t_1)$ represents the position of the highlight when the sound wave illuminates it, which corresponds to the initial position in the aforementioned model. The termination time of the echo signal for the highlight is denoted as t_2, and it is given by the equation:

$$t_2 = T + \tau_2 = T + \frac{2\mathbf{R}(t_2)}{c}. \qquad (20)$$

The starting moment of the echo signal corresponds to the position of the highlight. It represents the position where the highlight is located after its motion, compensating for the positional changes caused by the time difference resulting from the relative distances between the highlights.

4. Simulation and analysis

In order to validate the accuracy and effectiveness of the model proposed in this paper, this section simulates the echo signals of moving targets under single-highlight and multi-highlight scenarios with micro-Doppler effects, based on the models established in Secs. 2 and 3. The short-time Fourier transform (STFT) is employed to obtain time-frequency spectrograms, which are then used to extract the motion features of the targets. This enables the estimation of target parameters.

4.1. Single-highlight micro-motion model simulation verification

Simulations and analyses were conducted for stationary highlights, vibrating highlights, and rotating highlights, with the initial moment defined as the time of signal transmission. Partial simulation parameter settings are shown in Table 1.

For the analysis of stationary highlights, the position vector of the target at the moment of sound wave illumination is calculated as $\mathbf{R}' = (0,\ 505.50,\ -200)^{\mathrm{T}}$ m. Taking the initial position vector direction as the incident wave direction, the radial unit vector from the sonar to the target is $\mathbf{n}_p = \mathbf{R}'/\|\mathbf{R}'\|$, and the actual radial velocity component is $v = \mathbf{V}^{\mathrm{T}} \cdot \mathbf{n}_p = 18.597$ m/s. The simulated echo signal pulse width is $T_s' = 0.4879$ s.

Derived from the actual radial velocity, it yields the Doppler factor $\delta = 0.0248$, which then compresses the pulse width to $(1 - \delta)T_s = 0.4876$ s. This value is found to be consistent with the pulse width of the echo signal obtained in this paper's model. The same consistency is observed for vibrating highlights and rotating highlights.

The echo signal is subjected to STFT, resulting in a time-frequency distribution as shown in Fig. 3. The transmitted signal is a rectangular pulse signal, and the echo appears as a straight line on the time-frequency plot. By extracting the maximum value on the frequency axis at the middle moment, the Doppler frequency shift is $\xi = 741.925$ Hz. The radial velocity of the highlight translation is estimated to be $\hat{v} = 18.596$ m/s, which is consistent with the target radial velocity component.
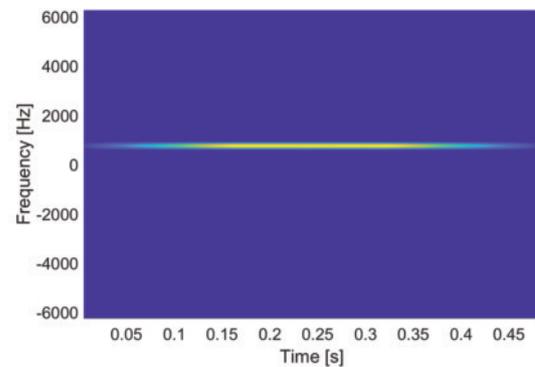
Fig. 3. STFT of the stationary highlight.

To analyze the vibrating highlights, the first step is to obtain the time-frequency representation. Then, peak extraction is performed on the time-frequency representation to obtain the micro-Doppler curve, as shown in Fig. 4.

The time difference between each maximum value is obtained from the micro-Doppler curve, resulting in the micro-Doppler period $\hat{T} = 0.0667$ s. The estimated vibration frequency $\hat{f} = \frac{1}{T} = 15.024$ Hz is found to be in close agreement with the fundamental vibration frequency. Additionally, the micro-Doppler spread $\max f_{md} - \min f_{md} = 1382.6$ Hz is calculated by subtracting the minimum value from the maximum value

Table 1. Highlights simulation parameters.

| Simulation parameters	Value
The initial position vector of each highlight	$\mathbf{R} = (0,\ 520,\ -200)^{\mathrm{T}}$ m
The translational velocity vector of each highlight	$\mathbf{V} = (0,\ -20,\ 0)^{\mathrm{T}}$ m/s
Vibration azimuth angle and pitch angle	$\alpha_v = 0,\ \beta_v = \frac{\pi}{2}$
Vibration frequency	$f_v = 10$ Hz
Vibration amplitude	$D_v = 0.5$ m
Rotational angular velocity vector	$\boldsymbol{\omega} = (0,\ 10\pi,\ 0)$ 1/s
Initial rotation angle	$\theta = 30°$
Rotation radius	$l = 3$ m
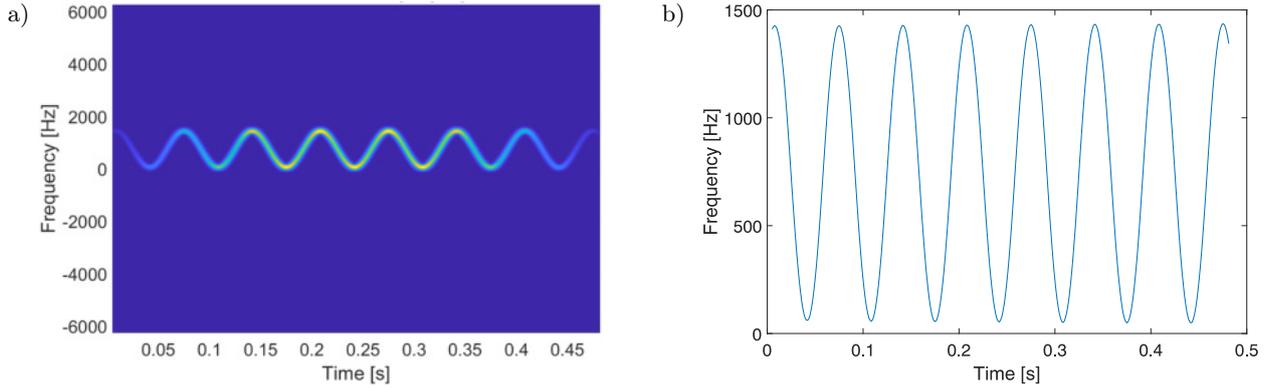
Fig. 4. STFT of the vibrating highlight (a) and the micro-Doppler curve of the vibrating highlight (b).
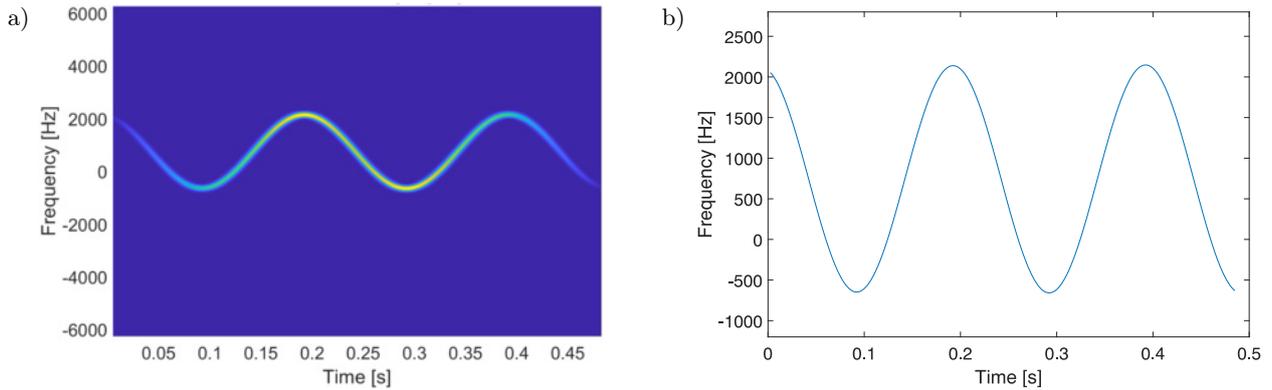
Fig. 5. STFT of the rotating highlight (a) and the micro-Doppler curve of the rotating highlight (b).

of the micro-Doppler curve. Then, the vibration amplitude estimation is $\widehat{D}$ = 0.497 m, which is consistent with the vibration amplitude.

The same analysis is performed on the rotating highlight and the results are shown in Fig. 5. The time difference between each maximum value is obtained from the micro-Doppler curve, resulting in the micro-Doppler period $\widehat{T}$ = 0.1997 s. The estimated angular velocity, $\widehat{\omega}$ = $\frac{2\pi}{\widehat{T}}$ = 10.016π 1/s, is found to be in close agreement with the fundamental angular velocity. Additionally, the micro-Doppler spread max f_{md} – min f_{md} = 2800.8 Hz is calculated by sub-

tracting the minimum value from the maximum value of the micro-Doppler curve. And then the rotation radius estimation is $\widehat{l}$ = 3.017 m, which is basically consistent with the rotation radius.

4.2. Simulation and analysis of multi-highlight moving target models

Simulation analysis is conducted on the multi-highlight model using the size parameters and motion parameters of a torpedo-like object. Some of the simulation parameters are set as shown in Table 2.

Table 2. Simulation parameters of the target.

| Simulation parameters	Value
The initial position of stationary highlight in the head area of the target	$\mathbf{R}_1$ = (0, 500, −200)$^{\mathrm{T}}$ m
The initial position of the vibrating highlight in the middle area of the target	$\mathbf{R}_2$ = (0, 520, −200)$^{\mathrm{T}}$ m
The initial position of the rotating highlight in the tail area of the target	$\mathbf{R}_3$ = (0, 540, −200)$^{\mathrm{T}}$ m
The translational velocity vector of each highlight	$\mathbf{V}$ = (0, −20, 0)$^{\mathrm{T}}$ m/s
Vibration azimuth angle and pitch angle	α_v = 0, β_v = $\frac{\pi}{2}$
Vibration frequency	f_v = 10 Hz
Vibration amplitude	D_v = 0.5 m
Rotational angular velocity vector	$\boldsymbol{\omega}$ = (0, 10π, 0) 1/s
Initial rotation angle	θ = 30°
Rotation radius	l = 3 m
The initial rotation angle of each rotating highlight	θ_1 = 30°, θ_2 = 150°, θ_3 = 270°
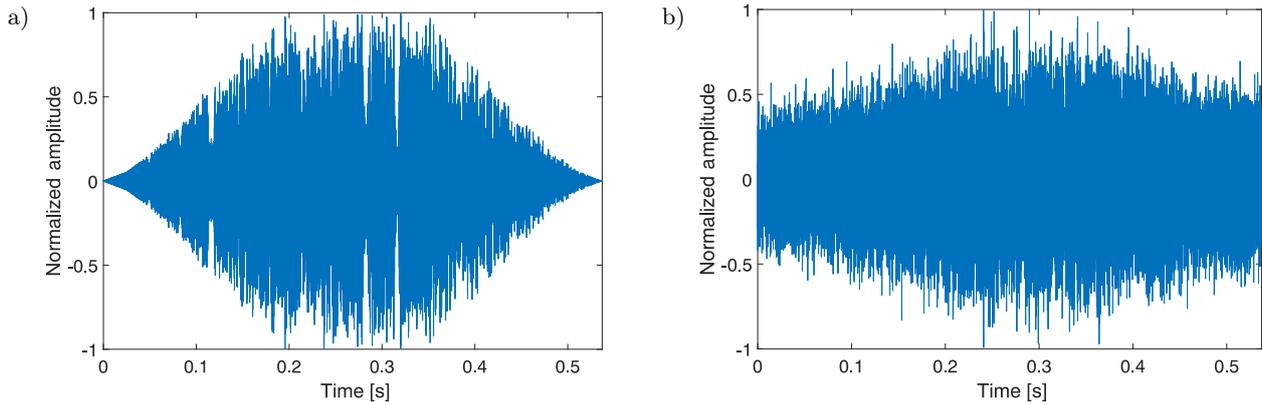
Fig. 6. Final received signal simulation process: a) echo in time domain; b) received signal in time domain.

Taking into account the received echo signals in real underwater environments, it should be noted that these signals are not entirely comprised of valid information due to the presence of environmental noise and reverberation interference in marine environments. The actual received signal is a mixture of the echo signal, noise, and reverberation. By considering the spectral characteristics of marine environmental noise and Gaussian white noise in the frequency domain, the required environmental noise is simulated (HAN *et al.*, 2020). The ocean reverberation is simulated using a unit scattering model (LI, LIU, 2016). The simulated echo time-domain sequence is shown in Fig. 6a. The noise and reverberation are then superimposed on the echo, resulting in the final received signal time-domain sequence as shown in Fig. 6b. The signal-to-noise ratio (SNR) of the final received signal is -1.81 dB, and the signal-to-reverberation ratio (SRR) is -2.51 dB.

The final received signal is processed as follows: first, the analytic signal of the echo signal is obtained, and then it is multiplied by the complex conjugate of the transmitted signal to obtain the baseband signal. Finally, the STFT is applied to the baseband signal, resulting in the time-frequency distribution shown in Fig. 7. By processing the echo signal into a baseband signal, the energy of the reverberation is concentrated
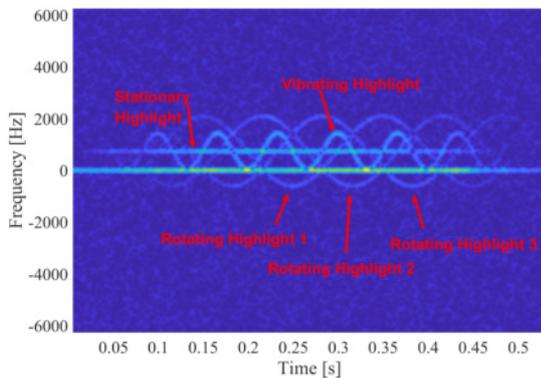
at frequency 0. The bright straight line above the reverberation represents the energy distribution of the stationary highlight at the target's head, characterized by relatively large energy and smaller amplitude. The micro-Doppler curves with larger amplitudes and smaller magnitudes represent the energy distribution of the vibrating highlights at the target's midsection. The three micro-Doppler curves with the highest amplitudes correspond to the energy distribution of the three rotating highlights at the target's tail.

From the time-frequency distribution graph, we can observe that the approximate time difference between the peaks of the micro-Doppler curves for the vibrating highlights is between 0.06 and 0.07 s. This allows us to make a rough estimation of the vibration frequency in the range of 14 to 16 Hz. On the other hand, the approximate time difference between the peaks of the micro-Doppler curves for the rotating highlights is around 0.2 s. This suggests a rough estimation of the rotational angular frequency at approximately 5 Hz.

By modifying the parameters of the micro-motion, we can observe changes in the time-frequency distribution of the target echo. Based on the aforementioned simulation parameters, we change the vibration amplitude of the vibrating highlight, denoted as $D_v = 1$ m, and the rotation radius of the rotating highlight, denoted as $l = 5$ m. The resulting STFT of the echo is shown in Fig. 8. Then, keeping the amplitude and rotation radius unchanged, we modify the vibration frequency of the vibrating highlight ($f_v = 20$ Hz) and the rotation frequency of the rotating highlight ($f_r = 8$ Hz). The resulting STFT of the echo is shown in Fig. 9.

It is observed that the time-frequency distribution exhibits significant changes compared to the previous simulation results. Increasing the vibration amplitude and rotation radius leads to a larger micro-Doppler spread in the echo. Similarly, increasing the vibration frequency and rotation frequency results in a larger micro-Doppler period and an increased micro-Doppler spread in the echo.


Fig. 7. STFT of the echo.
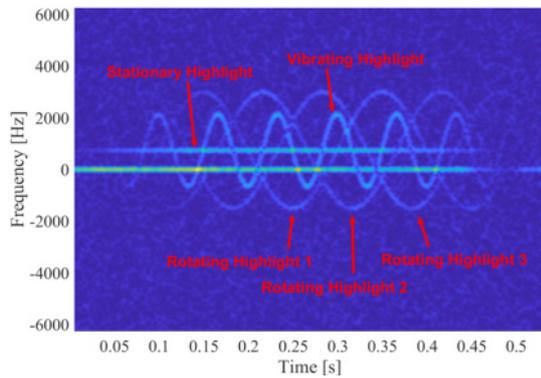
Fig. 8. STFT of the echo with modified vibration amplitude and rotation radius.
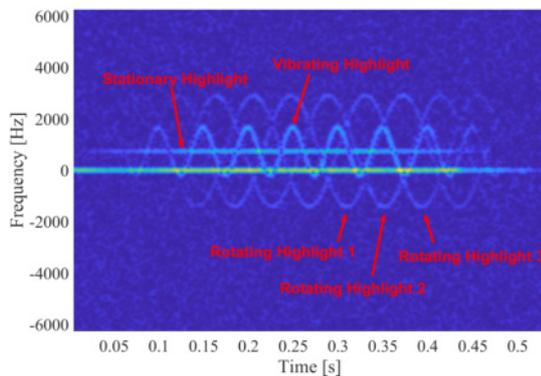

Fig. 9. STFT of the echo with modified vibration frequency and rotation angular velocity.

This indicates that in practical scenarios, the micro-Doppler features are influenced by the micro-motion parameters of the target. Different micro-motion forms of the target or different micro-motion states within the same micro-motion form will result in distinct patterns in the time-frequency domain. Additionally, in theoretical derivations, the micro-Doppler is also influenced by parameters such as the target's initial position and the carrier frequency of the transmitted signal.

Therefore, micro-Doppler can be utilized as a motion feature for target classification and recognition, while the Doppler frequency shift is primarily used to extract the target's radial velocity and can also serve as a motion feature.

5. Conclusion

This paper investigated the underwater motion feature models for the multi-highlight targets at high speed based on hightlight model and the micro-Doppler effect. Firstly, the positional variations of the target highlights were incorporated into the model, and a single-highlight model was established for compound motion patterns. Subsequently, a motion target model was proposed for multiple highlights with micro-

Doppler effects, and the effectiveness of the model was validated through motion parameter estimation using simulated results and the simulation of target echoes in real underwater environments.

The research results demonstrate that the echo pulse width of the composite motion single-highlight model coincides with the pulse width compression (or expansion) caused by translational velocity. Moreover, its micro-doppler characteristics can be fully manifested in the time-frequency distribution, and precise micro-motion parameters such as vibration frequency and rotational velocity can be accurately extracted from the time-frequency distribution. These findings validate the accuracy of the model and its superior representation of motion characteristics compared to traditional underwater point target models.

Additionally, the multi-highlight moving target model addresses the limitations of the single-highlight model in describing complex body targets. The echo simulation results based on this model fully demonstrate the effects of echo superposition caused by differences in the positions of multiple points and translational velocity in the time-frequency distribution, confirming the effectiveness of the model proposed in this study. Furthermore, variations in micro-motion parameters also lead to noticeable differences in the time-frequency distribution, and a rough estimation of the target's micro-motion parameters can be obtained from the time-frequency distribution. This indicates that micro-Doppler can serve as a motion feature for the target. The proposed model in this study provides a theoretical foundation and approach for the investigation of underwater moving targets, serving as a reference for further exploration of the motion features of underwater targets.

References

1. CHEN G. (2014), *Micro-Doppler signature analysis of radar targets and its applications* [in Chinese], Ph.D. Thesis, Xidian University in Xi'an, China, doi: 10.7666/d.D725864.

2. CHEN V.C., LI F., HO S.S., WECHSLER H. (2003), Analysis of micro-Doppler signatures, [in:] *IEE Proceedings-Radar, Sonar and Navigation*, **150**(4): 271–276, doi: 10.1049/ip-rsn:20030743.

3. CLEMENTE C., BALLERI A., WOODBRIDGE K., SORAGHAN J.J. (2013), Developments in target micro-Doppler signatures analysis: Radar imaging, ultrasound and through-the-wall radar, *EURASIP Journal on Advances in Signal Processing*, **2013**: 47, doi: 10.1186/1687-6180-2013-47.

4. DONG Z., LI Y., CHEN X. (2013), Submarine echo simulation method based on highlight model [in Chinese], *Computer Simulation*, **30**(6): 38–41, doi: 10.3969/j.issn.1006-9348.2013.06.009.

5. GAO H., XIE L., WEN S., KUANG Y. (2010), Micro-Doppler signature extraction from ballistic target with

micro-motions, *IEEE Transactions on Aerospace and Electronic Systems*, **46**(4): 1969–1982, doi: 10.1109/TAES.2010.5595607.

6. Gong Z., Li C., Jiang F., Zheng J. (2020), AUV-aided localization of underwater acoustic devices based on Doppler shift measurements, *IEEE Transactions on Wireless Communications*, **19**(4): 2226–2239, doi: 10.1109/TWC.2019.2963296.

7. Han L., Feng C. (2020), Micro-Doppler-based space target recognition with a one-dimensional parallel network, *International Journal of Antennas and Propagation*, **2020**: 8013802, doi: 10.1155/2020/8013802.

8. Han M., Wang C., Sun Q., Wang W., Lu Y. (2020), Measurement and analysis of ambient noise in the South China Sea based on underwater acoustic buoy [in Chinese], *Journal of Applied Acoustics*, **39**(4): 536–542, doi: 10.11684/j.issn.1000-310X.2020.04.006.

9. Hanif A., Muaz M., Hasan A., Adeel M. (2022), Micro-Doppler based target recognition with radars: A review, *IEEE Sensors Journal*, **22**(4): 2948–2961, doi: 10.1109/JSEN.2022.3141213.

10. Kashyap R., Singh I., Ram S.S. (2015), Micro-Doppler signatures of underwater vehicles using acoustic radar, [in:] *2015 IEEE Radar Conference (Radar-Con)*, pp. 1222–1227, doi: 10.1109/RADAR.2015.7131181.

11. Kim Y., Ling H. (2009), Human activity classification based on micro-Doppler signatures using a support vector machine, *IEEE Transactions on Geoscience and Remote Sensing*, **47**(5): 1382–1337, doi: 10.1109/TGRS.2009.2012849.

12. Kim Y., Moon T. (2016), Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks, *IEEE Geoscience and Remote Sensing Letters*, **13**(1): 8–12, doi: 10.1109/LGRS.2015.2491329.

13. Kou S., Feng X. (2022), Angle-micro-Doppler frequency image of underwater target multi-highlight combining with sparse reconstruction, *Applied Acoustics*, **188**: 108563, doi: 10.1016/j.apacoust.2021.108563.

14. Kulhandjian H., Ramachandran N., Kulhandjian M.K., D'Amours C. (2020), Human activity classification in underwater using sonar and deep learning, [in:] *Proceedings of the 14th International Conference on Underwater Networks & Systems*, pp. 1–5, doi: 10.1145/3366486.3366509.

15. Li S., Liu S. (2016), A modeling and simulation based on k-distribution model [in Chinese], *Ship Science and Technology*, **38**(s1): 158–161, doi: 10.3404/j.issn.1672-7619.2016.S1.029.

16. Saffari A., Zahiri S.-H., Khishe M. (2023), Automatic recognition of sonar targets using feature selection in micro-Doppler signature, *Defence Technology*, **20**: 58–71, doi: 10.1016/j.dt.2022.05.007.

17. Tang W. (1994), Highlight model of echoes from sonar targets, *Acta Acustica*, **19**(2): 92–100.

18. Tang Y., Wang X., Li H., Gao C., Miao X. (2020), Experimental research on interior field noise and the vibration characteristics of composite reinforced sheet-beam structures, *Applied Acoustics*, **160**: 107154, doi: 10.1016/j.apacoust.2019.107154.

19. Wang Z., Ren A., Zhang Q., Zahid A., Abbasi Q.H. (2023), Recognition of approximate motions of human based on micro-Doppler features, *IEEE Sensors Journal*, **23**(11): 12388–12397, doi: 10.1109/JSEN.2023.3267820.

20. Wu Y., Luo M., Li S. (2022), Measurement and extraction of micro-Doppler feature of underwater rotating target echo, [in:] *OCEANS 2022 – Chennai*, pp. 1–5, doi: 10.1109/OCEANSChennai45887.2022.9775247.

21. Xu L. (2016), *Study on Doppler parameter estimation of underwater acoustic signal and its application* [in Chinese], Ph.D. Thesis, Northwestern Polytechnical University in Xi'an, China, doi: 10.7666/d.D01304090.

22. Yang Y., Fan J., Wang B. (2023), Research on scattering feature extraction of underwater moving cluster targets based on the highlight model, *Archives of Acoustics*, **48**(2): 235–247, doi: 10.24425/aoa.2023.145235.

23. Zhang B., Hu Y., Wang H., Zhuang Z. (2018), Underwater source localization using TDOA and FDOA measurements with unknown propagation speed and sensor parameter errors, *IEEE Access*, **6**: 36645–36661, doi: 10.1109/ACCESS.2018.2852636.

24. Zhang R., Wang Y., Yeh C., Lu X. (2023), Precession parameter estimation of warhead with fins based on micro-Doppler effect and radar network, *IEEE Transactions on Aerospace and Electronic Systems*, **59**(1): 443–459, doi: 10.1109/TAES.2022.3182635.

25. Zhao Y., Su Y. (2023), Estimation of micro-Doppler parameters with combined null space pursuit methods for the identification of LSS UAVs, *IEEE Transactions on Geoscience and Remote Sensing*, **61**: 1–11, doi: 10.1109/TGRS.2023.3264643.

Research Paper

A Hybrid Finite Element Method – Kirchhoff Approximation Method for Modeling Acoustic Scattering from an Underwater Vehicle Model with Alberich Coatings with Periodic Internal Cavities

Fan YANG[(1)], Zilong PENG[(1), (3)]*, Hao SONG[(2)], Yuhang TANG[(3)], Xuhong MIAO[(3)]

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

[(2)] *Systems Engineering Research Institute*
Beijing, China

[(3)] *PLA Unit 92578*
Beijing, China

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

Anechoic tiles can significantly reduce the echo intensity of underwater vehicles, thereby increasing the difficulty of detecting such vehicles. However, the computational efficiency of conventional methods such as the finite element method (FEM) and the boundary element method (BEM) has its limitations. A fast hybrid method for modeling acoustic scattering from underwater vehicles with anechoic tiles with periodic internal cavities, is developed by combining the Kirchhoff approximation (KA) and FEM. The accuracy and rapidity of the KA method were validated by FEM. According to the actual situation, the reflection coefficients of rubber materials with two different structures under rigid backing are simulated by FEM. Using the KA method, the acoustic scattering characteristics of the underwater vehicle with anechoic tiles are obtained by inputting the reflection coefficients and the target's geometric grid. Experiments on the monostatic target strength (TS) in the frequency range of 1 to 20 kHz and time domain echo characteristics of acoustic scattering on a benchmark scale model with anechoic tiles are conducted. The research results indicate that the TS values and echo characteristic curves of the KA solutions closely approximate the experimental results, which verifies the accuracy of the KA method in calculating the TS and echo characteristics of underwater vehicles with anechoic tiles.

Keywords: anechoic tile; Kirchhoff approximation; target strength; echo characteristic.

1. Introduction

Anechoic tiles are widely used in underwater combat platforms such as underwater vehicles. In the underwater acoustic environment, considering the influence of temperature, salinity, pressure, and viscosity (WITOS, 2019; ESFAHANI *et al.*, 2023; ESFAHANI, SUN, 2023), laying the anechoic tiles on the submarine's surface not only absorbs incident sound waves but also reduces structural noise. Therefore, anechoic tiles play an important role in the acoustic stealth of underwater vehicles (YAO *et al.*, 2007; HUANG *et al.*, 2015; XU *et al.*, 2004). With the development of modern sonar technology and in-depth research of acoustic stealth technology around the world, studying acoustic scattering and echo characteristics simulation of underwater vehicles equipped with anechoic tiles has become a crucial mission.

Many scholars have studied the scattered sound field of underwater complex targets using numerical methods such as finite element method (FEM) and boundary element method (BEM). In 2013, WEI *et al.*

simulated the acoustic scattering of rigid spheres and infinite cylinders based on BEM. In 2015, Xu *et al.* used the BASIS method to simulate the scattering of the BeTSSi model at a frequency of 200 Hz. In 2018, Feng *et al.* employed BEM to analyze the target strength (TS) of the rigid boundary benchmark model in the low-frequency range. In the same year, Chen and Luo (2018) deduced the Helmholtz formula for the scattered sound field of underwater targets based on the boundary element theory.

FEM and BEM require extremely dense meshes to ensure accurate calculation, and this mesh density increases with higher computing frequencies (Liu *et al.*, 2019). As the mesh density increases, the computational time for both methods at high frequencies progressively extends. To solve this problem, Fan and Tang (1999) proposed a Kirchhoff approximation (KA) method for calculating TS. Subsequently, Fan and Zhuo (2006) introduced a KA method for graphic visualization, considering occlusion, secondary scattering (Zheng *et al.*, 2011; Marston, Sun, 1995), ray tracing, and multiple iterations (Wang *et al.*, 2021).

Lee and Seong (2009) derived the time domain solution of Kirchhoff's formula for impedance polygons. Pignier *et al.* (2015) used the KA formula to study multiple scattering problems of external radiation noise from moving ground vehicles through objects such as roads, buildings, and noise-shields. Abawi (2016) used KA to establish a method for solving the frequency and time domain solutions of scattered sound fields from arbitrarily shaped targets. To improve the computational speed of the KA method, Lavia *et al.* (2018) proposed the concept of replacing the flat facets with curved triangles. Additionally, Liu *et al.* (2023) established a highlight model that can reflect the real lines and scattering characteristics of underwater vehicles.

The traditional KA method is primarily applied to calculate the echo intensity of rigid targets. While FEM can solve the acoustic scattering problem of the targets laying anechoic tiles, the modeling process becomes increasingly time-consuming, especially when dealing with complex anechoic tiles and the varying shapes of underwater vehicles. Therefore, whether it is modeling or calculation, FEM will require a substantial time investment. In this paper, the reflection coefficients of anechoic tiles with periodic cavities are calculated by FEM at different angles and frequencies. To overcome the time constraints associated with FEM, we integrate the target grid information and reflection coefficient matrix into the KA method, originally designed for rigid targets. This enables the quick and accurate calculation of the TS for models with anechoic tiles containing periodic internal cavities. By combining FEM and the KA method, the modeling work of laying anechoic tile on the submarine model

can be omitted, and the spatial characteristics of the scattered sound field are studied.

2. Theoretical research and calculation model

2.1. Physical method of geometrical acoustic scattering of underwater target

As shown in Fig. 1, the plane wave is incident from M_1 to the surface s, then the scattered acoustic potential function satisfies the following Helmholtz integral equation (Fan *et al.*, 2012):

$$\phi_s = \frac{1}{4\pi} \int_s \left[\phi_s \frac{\partial}{\partial \mathbf{n}} \left(\frac{e^{ik\mathbf{r}_2}}{\mathbf{r}_2} \right) - \frac{\partial \phi_s}{\partial \mathbf{n}} \frac{e^{ik\mathbf{r}_2}}{\mathbf{r}_2} \right] \mathrm{d}s, \qquad (1)$$

where s is the surface of the scatterer, $\mathbf{n}$ is the external normal of the surface, $k = 2\pi f/c$, where the sound velocity c is a nonlinear function of temperature, ambient pressure, and salinity. In the case of free sound field and horizontal detection, where the target, transducer, and hydrophone are all situated at the same depth, the sound velocity under the circumstances can be considered constant; $\mathbf{r}_1$ and $\mathbf{r}_2$ are the incident point radius vector and the scattering point radius vector, respectively; θ_1 is the angle between $\mathbf{r}_1$ and $\mathbf{n}$, θ_2 is the angle between $\mathbf{r}_2$ and $\mathbf{n}$.
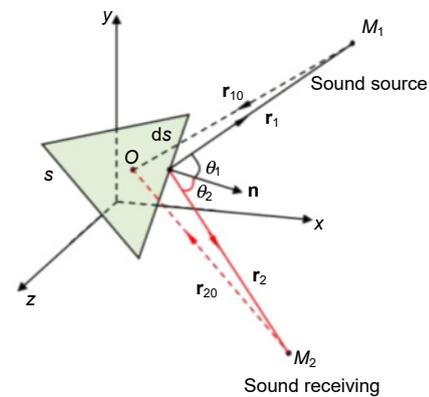

Fig. 1. Schematic diagram of the integral region, where O is the origin of coordinates, generally considered to be the geometric center of the target. The light green triangle $\mathrm{d}s$ represents the integral grid. Point M_1 represents the sound source, and point M_2 represents the sound receiving point.

In the monostatic configuration, where $\mathbf{r}_1 = \mathbf{r}_2 = \mathbf{r}$, $\theta_2 = \theta_1 = \theta$, and A is the amplitude, the far-field scattered sound field ϕ_s can be approximated as (Fan *et al.*, 2012):

$$\phi_s = -\frac{A}{2\pi} \int_s e^{ik2\mathbf{r}} \left(\frac{ik\mathbf{r} - 1}{\mathbf{r}^3} \cos\theta \right) \mathrm{d}s. \qquad (2)$$

For near-field acoustic scattering, considering the plate shown in Fig. 2, point o is the target's center and origin of coordinates. Point c, selected as a reference point on the plate, has a vector diameter denoted
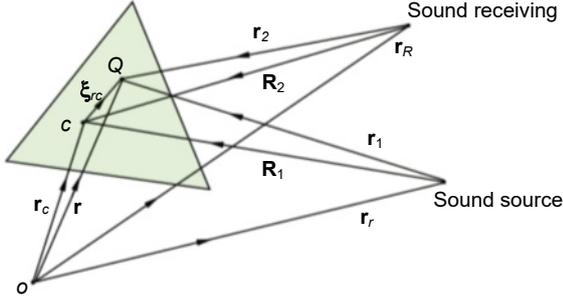
Fig. 2. Near-field sound scattering.

as $\mathbf{r}_c$, commonly representing the geometric center of the plate. Point Q is an arbitrary point on the plate. The vector diameter from point o to point Q is $\mathbf{r}$, the vector diameter of the transmitting transducer is $\mathbf{r}_T$, and the vector diameter of the receiving transducer is $\mathbf{r}_R$. There are $\mathbf{r}_1 = \mathbf{r} - \mathbf{r}_T$, unit vector $\mathbf{r}_{10} = (\mathbf{r} - \mathbf{r}_T)/r_1$; $\mathbf{r}_2 = \mathbf{r} - \mathbf{r}_R$, unit vector $\mathbf{r}_{20} = (\mathbf{r} - \mathbf{r}_R)/r_2$; $\mathbf{R}_1 = \mathbf{r}_c - \mathbf{r}_T$, unit vector $\mathbf{R}_{10} = (\mathbf{r}_c - \mathbf{r}_T)/R_1$; $\mathbf{R}_2 = \mathbf{r}_c - \mathbf{r}_R$, unit vector $\mathbf{R}_{20} = (\mathbf{r}_c - \mathbf{r}_T)/R_2$.

Where $\mathbf{n}$ is the unit normal vector of the plate. For small plates, it can be considered that the incident sound wave is a plane wave, allowing for the following approximations: $r_1 \approx R_1 = |\mathbf{r}_c - \mathbf{r}_T|$, $\mathbf{r}_{10} \approx \mathbf{R}_{10}$, $r_2 \approx R_2 = |\mathbf{r}_c - \mathbf{r}_T|$, $\mathbf{r}_{20} \approx \mathbf{R}_{20}$; $r_1 \approx R_1 - \mathbf{R}_{10} \cdot \boldsymbol{\xi}_{rc}$, $r_2 \approx R_2 - \mathbf{R}_{20} \cdot \boldsymbol{\xi}_{rc}$, $\boldsymbol{\xi}_{rc} = \mathbf{r} - \mathbf{r}_c$.

In the monostatic configuration, the near-field scattered sound field ϕ_{sm} is

$$\phi_{sm} = -\frac{A}{2\pi} e^{i2k\mathbf{R}_1} \frac{ik\mathbf{R}_1 - 1}{\mathbf{R}_1^3} \mathbf{n} \cdot \mathbf{R}_{10} \int_s e^{-2ik\mathbf{R}_{10} \cdot \boldsymbol{\xi}_{rc}} \, \mathrm{d}s. \tag{3}$$

For the near-field KA method, the key to the aforementioned formula is to find the area integral:

$$I = \int_s e^{-2ik\mathbf{R}_{10} \cdot \boldsymbol{\xi}_{rc}} \, \mathrm{d}s. \tag{4}$$

Assuming that the integral plane is a plane polygon s_0 in the XOY plane, and the normal of the plane is $\mathbf{R}_{10} = u'\mathbf{i} + v'\mathbf{j} + w'\mathbf{k}$, $\boldsymbol{\xi}_{rc} = x\mathbf{i} + y\mathbf{j}$, then the integral operation above can be converted to a summation for an exact expression of the result by the following equation (LIU, 2020):

$$I_p = \int_{s_0} e^{2ik(u'x + v'y)} \, \mathrm{d}x \, \mathrm{d}y$$

$$= \sum_{n=1}^{3} \frac{e^{-i(x_n u + y_n v)}(p_{n-1} - p_n)}{(u + P_{n-1}v)(u + P_n v)}, \tag{5}$$

and $u' = 2ku$, $v' = 2kv$, $w' = w$, $p_1 = \frac{y_2 - y_1}{x_2 - x_1}$, $p_2 = \frac{y_3 - y_2}{x_3 - x_2}$, $p_3 = \frac{y_1 - y_3}{x_1 - x_3}$, $p_0 = \frac{y_1 - y_3}{x_1 - x_3}$, where n is the number of polygon vertices and (x_n, y_n) is the coordinates of polygon vertices.

The coordinates of points in the original coordinate system are transformed into a new coordinate system. A coordinate transformation is performed for points 1, 2, 3 to obtain their coordinates in the new coordinate system, which are $x2(n)$, $y2(n)$, $z2(n)$ ($n = 1, 2, 3$). In the new coordinate system, the unit vector from the receiving point to the center of the plate is represented as $u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$, with $u' = 2ku$, $v' = 2kv$, $w' = w$, $k = 2\pi f/1500$.

According to the transformation of $x2(n)$, $y2(n)$, $z2(n)$ ($n = 1, 2, 3$),

$$\phi = w \sum_{n=1}^{3} e^{-i(x2(n) + y2(n))} \frac{p_{n-1} - p_n}{(u' + p_{n-1}v')(u' + p_n v')}, \tag{6}$$

and $p_1 = \frac{y2(2) - y2(1)}{x2(2) - x2(1)}$, $p_2 = \frac{y2(2) - y2(2)}{x2(3) - x2(2)}$, $p_3 = \frac{y2(1) - y2(3)}{x2(1) - x2(3)}$, $p_0 = p_3$.

2.2. Research on time domain echo characteristics of target based on the linear transfer network model

In the framework of linear acoustics, the problem of target scattering can be described using 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 $h(\tau, \mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\rho})$, where τ is the delay, $\mathbf{r}_1$ and $\mathbf{r}_2$ are the radius vector of the incident and scattering point, respectively, and $\boldsymbol{\rho}$ is the radius vector of the target. The frequency domain transfer function $H(f, \mathbf{r}_1, \mathbf{r}_2, \boldsymbol{\rho})$ of the transfer network is denoted as the ratio of the scattering wave potential function to the incident wave potential function, where f is the frequency. In the monostatic configuration, $\mathbf{r}_1 = \mathbf{r}_2$, and the transfer functions in the time domain and frequency domain are $h(\tau, \mathbf{r}, \boldsymbol{\rho})$ and $H(f, \mathbf{r}, \boldsymbol{\rho})$, 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 (LIU, 2020):

$$y(t) = x(t) \otimes h(\tau, \mathbf{r}, \boldsymbol{\rho}), \tag{7}$$

where $\otimes$ denotes the convolution operation. Then, the echo signal in the frequency domain is given by:

$$Y(f) = X(f) \cdot H(f, \mathbf{r}, \boldsymbol{\rho}), \tag{8}$$

$$y(t) = F^{-1}[Y(f)]. \tag{9}$$

The time-domain echo signal can be obtained by taking the inverse Fourier transform of the frequency-domain signal. This framework allows for the indirect extraction of the time-domain echo characteristics of underwater targets by calculating the frequency-domain signals.

2.3. Materials and model

Achieving semi-infinite air and water by adding a perfectly matched layer, the materials used are shown in Table 1.

In Fig. 3, the benchmark model (Nell, Gilroy, 2003) is chosen for conducting the acoustic scattering and TS prediction of the target after laying different anechoic tiles.

The acoustic wave is incident horizontally, and the specific parameters of the benchmark model are shown in Fig. 3b. To satisfy the far field condition, the simulated distance is 10 km.

2.4. Application of Floquet theory in solving acoustic scattering from elastomers

The challenge of the FEM in solving the acoustic scattering problem lies in the inability to create infinitely long waveguides. As shown in Fig. 4a, Floquet-Bloch theory imposes boundary conditions on the com-

Table 1. Material parameters.

| Material	Density [kg/m³]	Young's modulus [Pa]	Poisson ratio	Sound velocity [m/s]	Thickness [mm]
Rubber (Xu *et al.*, 2004)	1090	$3 \times 10^7 (1 + 0.249i)$	0.49	–	50
Steel	7800	2.13×10^{11}	0.3	–	8
Water	1000	–	–	1500	Semi-infinite
Air	1.21	–	–	343	Semi-infinite

a)

b)
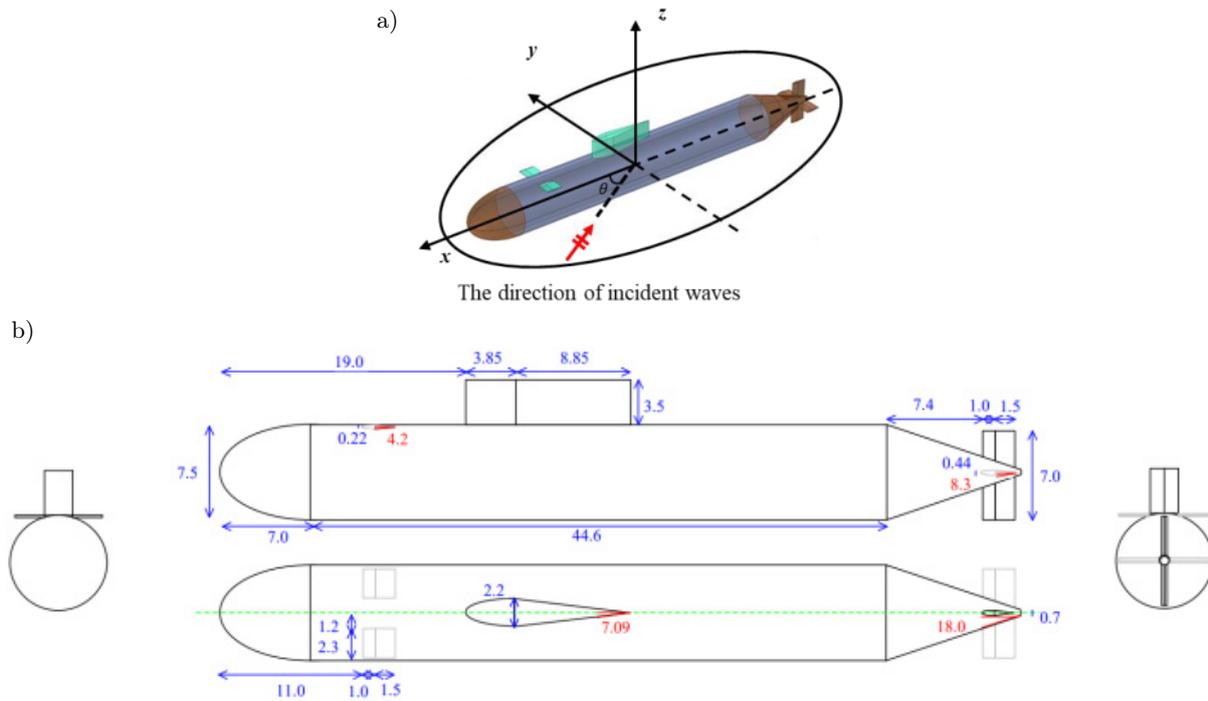
Fig. 3. a) Benchmark model and incident direction of the plane wave; b) specific dimensions of the benchmark submarine.
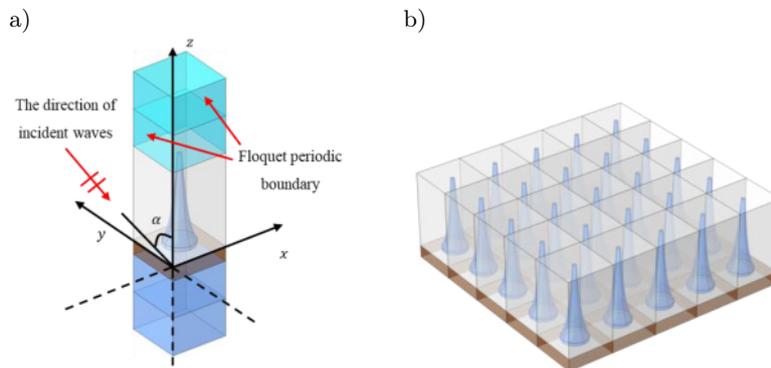

Fig. 4. a) Computational unit; b) schematic diagram of multiple units.

putational unit to simulate infinitely long waveguides. In the FEM, the equation is (Tong *et al.*, 2020):

$$\mathbf{u}_{\mathrm{dst}}(x,y) = \mathbf{u}_{\mathrm{src}}(x,y)\mathrm{e}^{-i\mathbf{k}_F(\mathbf{r}_{\mathrm{dst}}-\mathbf{r}_{\mathrm{src}})}, \qquad (10)$$

where $\mathbf{u}_{\mathrm{src}}$ and $\mathbf{r}_{\mathrm{src}}$ are the displacement and coordinate of the source boundary, and $\mathbf{u}_{\mathrm{dst}}$ and $\mathbf{r}_{\mathrm{dst}}$ are the displacement and coordinate of the target boundary. Floquet periodic boundary conditions are applied on both sides of the model, and different traveling wave numbers $\mathbf{k}_F$ are input to simulate the finite length region as an infinite length waveguide. The anechoic tile model after incorporating the Floquet periodic condition along with a schematic diagram of multiple units are shown in Fig. 4.

2.5. Validation of the KA method

To determine the accuracy and efficiency of the KA method in calculating the acoustic scattering from the target equipped with anechoic tiles with air cavities, a verification study was conducted. In Fig. 5a, a cylindrical shell with a radius of 0.2 m, height of 1 m, and thickness of 0.008 m is selected. The cylindrical shell is made of steel with air inside, there are several anechoic tiles, each 0.05 m thick with a grating spacing of 0.03 m outside the cylindrical shell. The center of each anechoic tile contains a longitudinally cylindrical cavity with a radius of 0.005 m and a height of 0.025 m. The domain being studied is meshed with tetrahedral elements, with a maximum mesh size of $\lambda/6$ (Lu, 2014), where $\lambda = \frac{2\pi f}{c}$, with f representing the frequency and c representing the velocity.

In Fig. 5b, a comparison of the TS for the verification model, covered with anechoic coatings with cylindrical cavities, is presented between the FEM results and KA solutions, when α is 0°. In the frequency range of 200 Hz–3 kHz, the average values of the results calculated by the FEM and the KA method are −13.75 and −13.61 dB, respectively. This validates the accuracy of the KA method in modeling the acoustic scattering of underwater vehicles with anechoic coatings containing cavities. The total calculation time of the FEM is 25.391 s, while the KA method only takes 15.35 s, indicating that the computational speed of the KA method is roughly 1654 times faster than the FEM method. This significant difference highlights the rapidity of the KA method. The research process of this study is shown in Fig. 6.

Fig. 5. a) Verification model; b) comparison of the FEM results and the KA solutions.
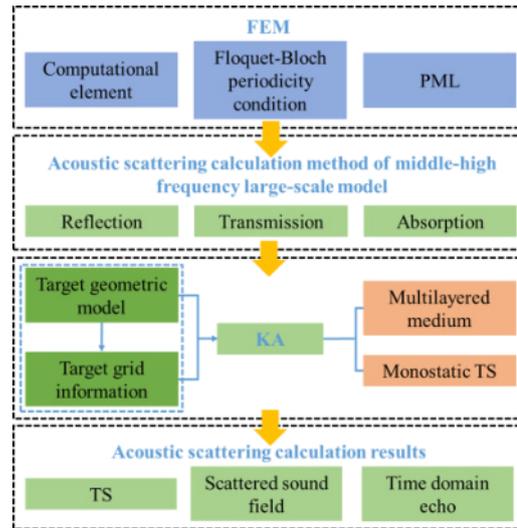
Fig. 6. Flow chart of acoustic scattering characteristics of underwater vehicle with Alberich coatings.

3. Numerical simulation of underwater vehicle with anechoic tiles

3.1. Reflection coefficients of different rubber structures

The reflection coefficients of homogeneous medium and rubber with air cavities at different angles and frequencies (f) are calculated by FEM, as shown in Fig. 7.

As illustrated in Fig. 7, the calculated frequency is in the range of 200 Hz–20 kHz with a step of 200 Hz. The cavity significantly impacts the acoustic characteristics of the anechoic tile. The resonance effect of
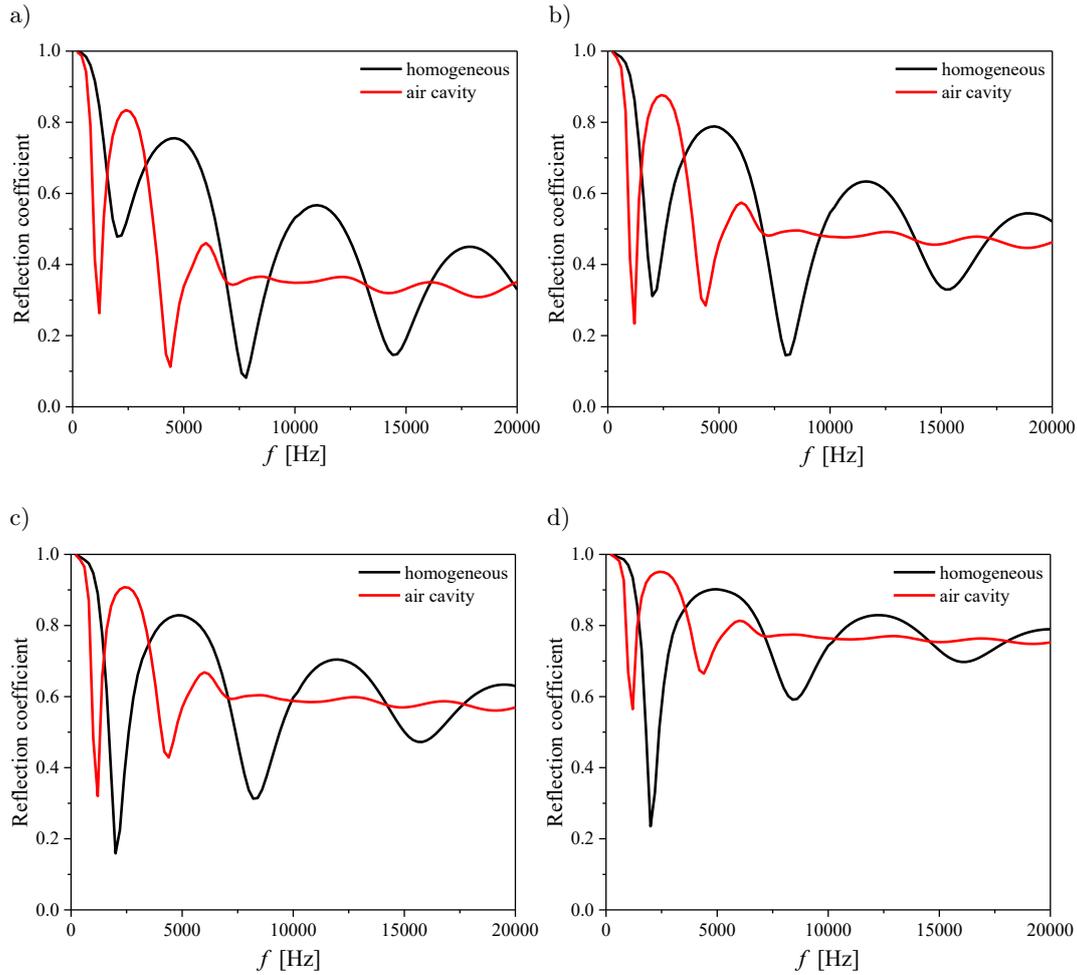
Fig. 7. Comparison of reflection coefficients of different anechoic tiles at different angles:
a) $0°$; b) $45°$; c) $60°$; d) $75°$.

the cavity not only accelerates the convergence speed of the reflection coefficient but also makes the reflection peak to move towards the low-frequency direction, which enhances the middle and low-frequency performance of the structure. The cavity structure makes the reflection coefficient more stable in the high frequency range. However, in practical applications, anechoic tiles with air cavities generally require more complex processing and manufacturing procedures compared to homogeneous medium anechoic tiles. Additionally, as the depth of the submarine increases, the pressure of the seawater medium also rises, leading to alterations in the acoustic performance of anechoic tiles with air cavities.

3.2. TS of the submarine model with homogeneous rubber and rubber with periodic internal cavities

The reflection coefficients, as shown in Fig. 7, are substituted into the KA method to obtain the TS of the benchmark model after applying different kinds of anechoic tiles. This process yields a TS angle-frequency spectrum, with the horizontal axis representing the angle, and the vertical axis representing the frequency.

As shown in Fig. 8, different types of anechoic tiles exhibit varying degrees of suppression on the TS of underwater vehicles at different incident angles and frequencies. Similar to the trends observed in the reflection coefficient, anechoic tiles with periodic internal cavities demonstrate superior sound absorption in the low and specific middle frequency ranges compared to homogeneous mediums under the resonant interaction of the air cavity.

In Fig. 9, the directivity diagrams of TS for the benchmark model are presented at different frequencies before and after the applications of anechoic tiles, and the figure also illustrates the impact of different types of anechoic tiles on the TS of the model.

In Fig. 9, the variation trend of the TS presents a butterfly shape with the azimuth angle changing. The highest TS values are observed in the abeam direction of the underwater vehicle, primarily due to mirror reflections on the hull surface and acoustic reflections from the conning tower. With the existence
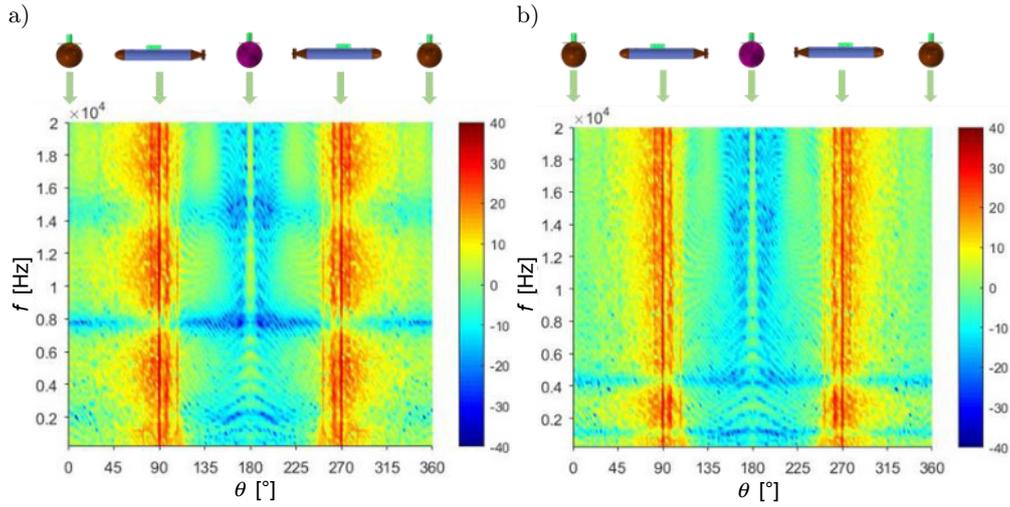
Fig. 8. TS angle-frequency spectrum of benchmark model with different anechoic tiles:
a) homogeneous medium anechoic tiles; b) anechoic tiles with air cavities.
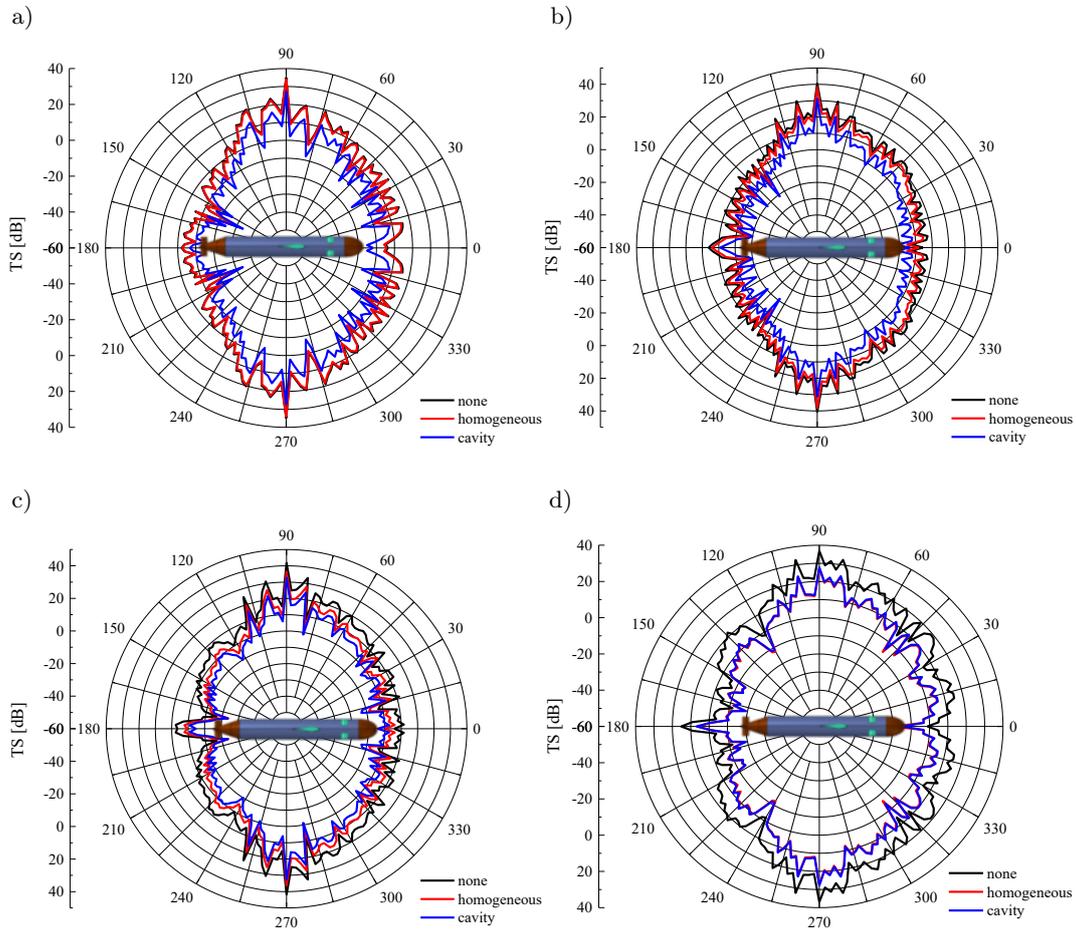
Fig. 9. TS directional diagrams of benchmark model with different anechoic tiles at different frequencies:
a) 1 kHz; b) 5 kHz; c) 10 kHz; d) 20 kHz.

of cavity resonance, anechoic tiles with air cavities exhibit a more effective sound absorption effect compared to anechoic tiles with homogeneous mediums, particularly in the middle and low frequency ranges. Under the condition of high frequency, the silencing effect of different anechoic tiles becomes consistent.

4. Comparisons of acoustic scattering characteristics between the KA solutions and experimental results

4.1. Acoustic scattering characteristic test instrument and arrangement

By conducting a lake test of the benchmark scale model with anechoic tiles, the variation of echo characteristics concerning the horizontal azimuth and frequency of the incident wave is quantitatively analyzed, providing data support for verifying the accuracy of the KA method. Figure 10 shows the benchmark scale model and the test site.

Fig. 10. Benchmark scale model and test site photos.

According to Fig. 11, the Agilent 33220A signal generator is primarily employed to produce acoustic signals with specified parameters, including frequency, amplitude, and pulse width. The JYH1000A power amplifier amplifies the acoustic signals generated by the signal generator to ensure that the transmitted signals meet the requirements for underwater acoustic detection. The transducer converts the electrical signal into the acoustic signal and transmits it for target detection. The hydrophone is the device that converts the acoustic signal into an electrical signal, utilized for receiving the echo signal from the underwater model and converting it into an electrical representation. The charge amplifier amplifies the echo signal from the hydrophone. The China Orient Institute of Noise & Vibration INV3062A1 data acquisi-
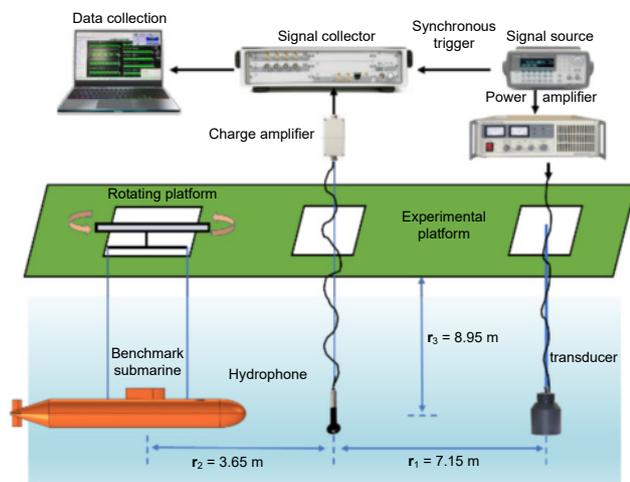

Fig. 11. Schematic diagram of the test equipment arrangement.

tion system is responsible for receiving and storing the output echo signal from the charge amplifier, allowing for subsequent analysis and interpretation. The specific parameters of the benchmark model are shown in Fig. 3b. The scaling ratio used in this experiment, relative to the original model, is 1:15. The thickness of the shell is 3 mm, the length is 4.13 m, and the radius of the midship is 0.25 m. The signal pulse width is set to 1.5 ms, the signal length is 500 ms, and the sampling frequency is 80 kHz. Anechoic tiles with a homogeneous medium are laid on the surface of the experimental model. The thickness of each anechoic tile is 10 mm, and its Young's modulus, density and Poisson's ratio are 4.41×10^6 Pa, 1100 kg/m^3 and 0.48, respectively. The transmitting transducer, hydrophone, and research model are aligned in a straight line. The distance between the transducer and the hydrophone is 7.15 m, and the distance between the hydrophone and the research model is 3.65 m. All three are located at an underwater depth of 8.95 m. At the beginning of the test, it is ensured that the research model is facing the transducer in the abeam direction, and a horizontal uniform rotation of 360° during the rotation process is maintained.

4.2. Benchmark scale model monostatic TS experimental data analysis

The TS azimuth characteristics of the benchmark scale model obtained by the experiment and the KA method are shown in Fig. 12.

As illustrated in Fig. 12, the TS is highest in the abeam direction, which is $-7 \sim -3$ dB. In the bow direction it is $-14 \sim -6$ dB, and in the stern direction it is the smallest, measuring $-16 \sim -9$ dB. It can be seen that except for the range of 210–330°, the KA solutions are basically consistent with the experimental results in trend and magnitude, demonstrating the reliability of the KA method in predicting the monostatic TS of underwater vehicles with anechoic tiles. In the range of 210–330°, because of the model processing and suspension problems, there will be some discrepancies between the experimental results and the KA solutions.

4.3. Experimental data analysis of the time domain echo spreading characteristics of the benchmark scale model

The benchmark scale model with anechoic tiles used in the test has the directions of incident waves within the angle range of 0–360° (where the abeam direction corresponds to incident directions of 0 and 180°, and the bow and stern correspond to 90 and 270° incident directions, respectively). The experimental results of time domain echo characteristics in different frequency ranges are compared to the KA solutions.

a)

b)

c)

d)
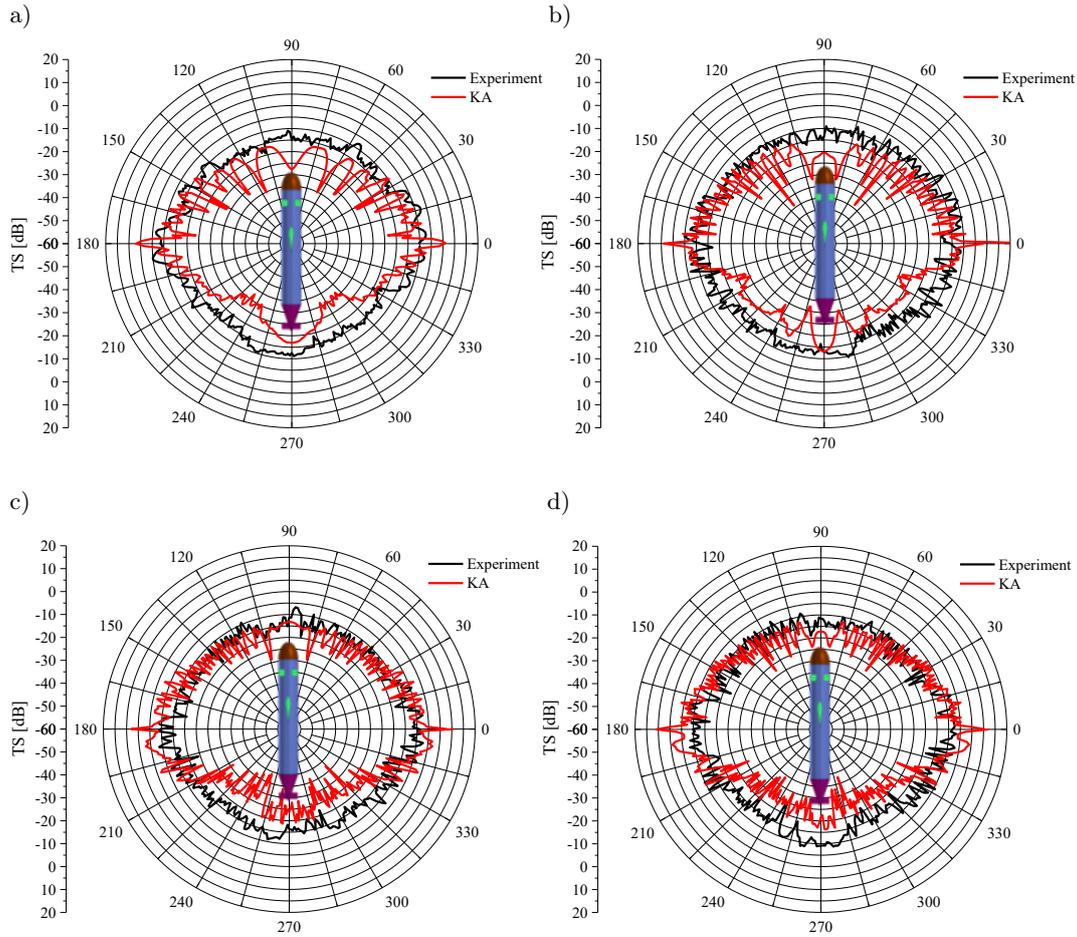
Fig. 12. Comparisons of TS between the experimental results and the KA solutions at different frequencies:
a) 5 kHz; b) 10\;kHz; c) 15 kHz; d) 20 kHz.

The comparisons of acoustic scattering echo characteristics between the benchmark submarine scale model' experimental results and KA solutions are shown in Figs. 13 and 14. Echo characteristics of the bow, elevator, conning tower and stern of the model can be clearly observed in the experimental results and KA solutions, and the characteristics of each component can be accurately reflected in the echo time. However, there are still some differences between the experimental results and KA solutions. For example,

a)

b)
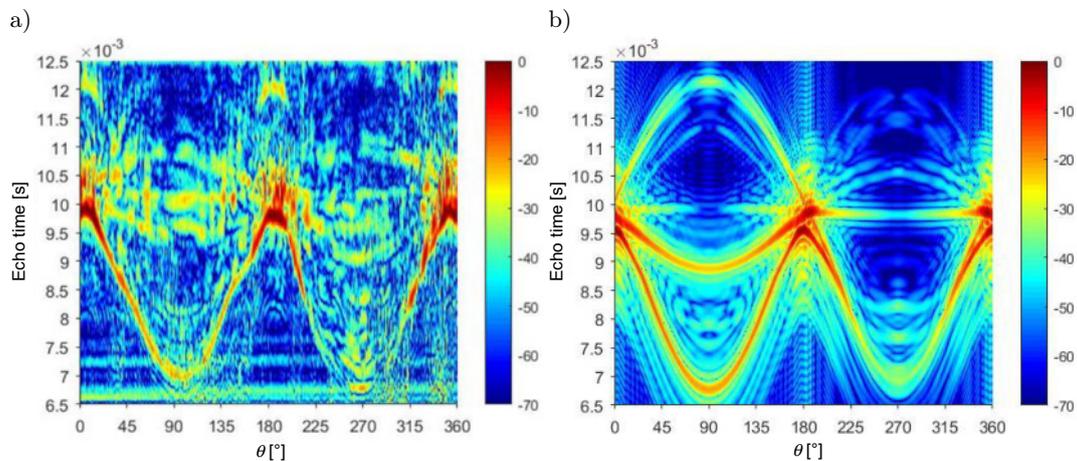
Fig. 13. Comparisons of the echo characteristics of the benchmark submarine scale model with anechoic tiles in the frequency band of 1–10 kHz between the experimental results and KA solutions: a) experimental results; b) KA solutions.
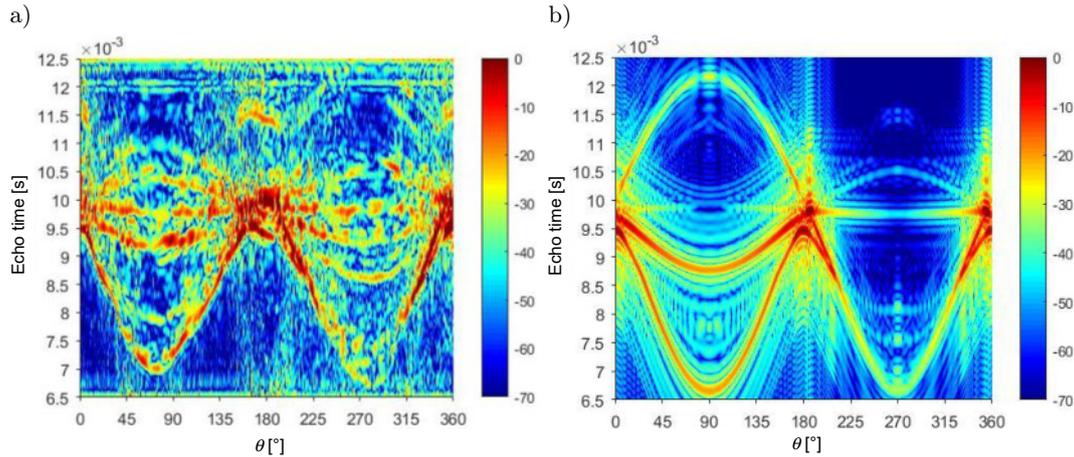
Fig. 14. Comparisons of the echo characteristics of the benchmark submarine scale model with anechoic tiles in the frequency band of 10–20 kHz between the experimental results and the KA solutions: a) experimental results; b) KA solutions.

the experimental results exhibits obvious highlights in the position of the elevator, and there are several more characteristic curves near the abeam direction which are presumed to be the echo characteristics from ropes, buckles and the beam used when hoisting the model. The above factors make the experimental results have more echo characteristics than the KA solutions.

5. Conclusions

In this study, the KA method is used to evaluate the TS and echo characteristics of a target after laying anechoic tiles. The rapidity and accuracy of the KA method are confirmed through comparisons with FEM and experimental results. This investigation achieves fast acoustic scattering prediction for targets with anechoic tiles, which makes a significant contribution to improving the acoustic stealth technology of the underwater vehicle.

From the research on underwater vehicles with anechoic tiles, we can come to the following conclusions.

The cavity significantly affects the resonance characteristics of anechoic tiles, especially in the case of vertical incidence, which not only reduces the reflection coefficient peaks, but also makes the reflection coefficient peaks move to the low frequency direction. This trend becomes more pronounced at frequencies below 10 kHz. However, acoustic tiles with internal cavities also require more complex processing and manufacturing processes, and the acoustic performance exhibits greater sensitivity to changes in environmental conditions, such as sound pressure. Therefore, in practical applications, a thorough assessment of the advantages and disadvantages of both methods is imperative to ensure the selection of an acoustic tile design that aligns with specific requirements.

In the experimental results and KA solutions, the TS and echo characteristics of the benchmark submarine model with anechoic tiles at different incident frequencies are approximately consistent. The benchmark submarine model with anechoic tiles can produce clear echo characteristics at the bow, elevator, conning tower, hull, and empennage, and several echo curves of the experimental results and KA solutions exhibit good agreement in terms of echo time, brightness, and shape. This alignment accurately reflects the external characteristics of the submarine model with anechoic tiles.

By investigating the sound absorption characteristics of anechoic tiles with cavities, we have gained profound insights into detection and stealth capabilities of underwater vehicles. The KA method enables rapid and precise evaluation of the TS after the application of anechoic tiles with periodic internal cavities. The significant reduction in computational time provides an efficient tool for the acoustic parameter design and optimization of anechoic tiles, thereby advancing underwater vehicle acoustic stealth and detection technology.

References

1. ABAWI A.T. (2016), Kirchhoff scattering from non-penetrable targets modeled as an assembly of triangular facets, *The Acoustical Society of America*, **140**(3): 1878–1886, doi: 10.1121/1.4962735.

2. CHEN X., LUO Y. (2018), Simulation of scattering acoustic field of underwater target in low frequency based on ANSYS and SYSNOISE [in Chinese], *Journal of Ordnance Equipment Engineering*, **39**(5): 103–107, doi: 10.11809/bqzbgcxb2018.05.022.

3. ESFAHANI I.C., JI S., SUN H. (2023), A drop-on-micropillars (DOM) based acoustic wave viscometer for high viscosity liquid measurement, *IEEE Sensors Journal*, doi: 10.1109/JSEN.2023.3309757.

4. ESFAHANI I.C., SUN H. (2023), A droplet-based micropillar-enhanced acoustic wave (μPAW) device for viscosity measurement, *Sensors and Actuators A: Physical*, **350**: 114121, doi: 10.1016/j.sna.2022.114121.

5. FAN J., TANG W.L. (1999), The planar element method for computing target strength (TS) of sonar [in Chinese], *Proceedings of the Acoustical Society of China 1999 Youth Conference*, pp. 40–41.

6. FAN J., TANG W.L., ZHUO L.K. (2012), Planar elements methods for forecasting the echo characteristics from sonar targets, *Ship Mechanics*, **16**(1–2): 171–180, doi: 10.3969/j.issn.1007-7294.2012.01.020.

7. FAN J., ZHUO L.K. (2006), Graphical acoustics computing method for echo characteristics calculation of underwater targets, *Acta Acustica*, **31**(6): 511–516, doi: 10.3321/j.issn:0371-0025.2006.06.006.

8. FENG X.L., CHEN N.R., LI X.W., LI J. (2018), Analyzing the target strength of Benchmark submarine by boundary element method at low and middle frequencies [in Chinese], *Technical Acoustic*, **37**(05): 418–424, doi: 10.16300/j.cnki.1000-3630.2018.05.003.

9. HUANG L.Z., XIAO Y., WEN J.H., YANG H.B., WEN X.S. (2015), Analysis of decoupling mechanism of an acoustic coating layer with horizontal cylindrical cavities [in Chinese], *Acta Physica Sinica*, **64**(15): 154301, doi: 10.7498/aps.64.154301.

10. LAVIA E., GONZALEZ J.D., BLANC S. (2018), Modelling high-frequency backscattering from a mesh of curved surfaces using Kirchhoff Approximation, *Journal of Theoretical and Computational Acoustics*, **27**(04): 17, doi: 10.1142/S2591728518500573.

11. LEE K., SEONG W. (2009), Time-domain Kirchhoff model for acoustic scattering from an impedance polygon facet, *The Acoustical Society of America*, **126**(1): 14–21, doi: 10.1121/1.3141887.

12. LIU B. (2020), *Research on acoustic scattering characteristics of typical structures of MUUV* [in Chinese], Shanghai Jiao Tong University, doi: 10.27307/d.cnki.gsjtu.2020.002589.

13. LIU H., PENG Z.L., FAN J., WU K. (2019), Numerical and experimental research on acoustic scattering time-frequency characteristics of dock landing ship [in Chinese], *Technical Acoustics*, **38**(02): 147–152, doi: 10.16300/j.cnki.1000-3630.2019.02.006.

14. LIU J.W., PENG Z.L., FAN J., LIU Y., KONG H.M. (2023), Acoustic scattering prediction method of underwater vehicles based on slice-parameterized multi-highlight model, *Acta Armamentarii*, **44**(02): 517–525, doi: 10.12382/bgxb.2021.0764.

15. LU D. (2014), *Researches on acoustic scattering of elastic target on finite element methods* [in Chinese], Harbin Engineering University.

16. MARSTON P.L., SUN N.H. (1995), Backscattering near the coincidence frequency of a thin cylindrical shell: Surface wave properties from elasticity theory and an approximate ray synthesis, *The Acoustical Society of America*, **92**(6): 777–783, doi: 10.1121/1.412124.

17. NELL C.W., GILROY L.E. (2003), *An improved BASIS model for the BeTSSi submarine*, DRDC Atlantic TR, Technical report.

18. PIGNIER J.N., O'REILLY J.C., BOIJ S. (2015), A Kirchhoff approximation-based numerical method to compute multiple acoustic scattering of a moving source, *Applied Acoustics*, **96**: 108–117, doi: 10.1016/j.apacoust.2015.03.016.

19. TONG Y.Z., FAN J., WANG B. (2020), Application of Floquet-Bloch theory in dipersion curve calculation [in Chinese], *Technical Acoustics*, **39**(01): 11–14, doi: 10.16300/j.cnki.1000-3630.2020.01.002.

20. WANG W.H., WANG B., FAN J., ZHOU J. (2021), An iterative planar elements method for calculating multiple acoustic scattering from concave targets, *Proceedings of the 18th Symposium on Underwater Noise of Ships*, pp. 121–126, doi: 10.26914/c.cnkihy.2021.056714.

21. WEI K.N., LI W., LEI M., CHAI Y.B. (2013), Simulation research on acoustic scattering characteristics of underwater targets based on boundary element method [in Chinese], *Proceedings of the Fourteenth Symposium on Underwater Noise of Ships*, pp. 452–462.

22. WITOS F. (2019), Properties of amplitude distributions of acoustic emission signals generated in pressure vessel during testing, *Archives of Acoustics*, **44**(3): 493–503, doi: 10.24425/aoa.2019.129264.

23. XU H.T., AN J.Y., LIU C.F. (2004), Acoustic characteristics of anechoic coatings containing air cavities in water [in Chinese], *Technical Acoustics*, **23**(z1): 345–347, doi: 10.3969/j.issn.1000-3630.2004.z1.101.

24. XU Z.C., ZHANG M.M., WANG L. (2015), Numerical simulation of acoustic scattering at low frequency for the BeTSSi submarine [in Chinese], *Computer & Digital Engineering*, **43**(04): 551–553+575, doi: 10.3969/j.issn1672-9722.2015.04.003.

25. YAO X.L., ZHANG Y., QIAN D.J., HUANG C., ZHANG H.H. (2007), Characteristics analysis of acoustic insulation and decoupled tiles by FEM and experiment [in Chinese], *Chinese Journal of Ship Research*, **2**(6): 9–15, doi: 10.3969/j.issn.1673-3185.2007.06.003.

26. ZHENG G.Y., FAN J., TANG W.L. (2011), A modified planar elements method considering occlusion and secondary scattering [in Chinese], *Acta Acustica*, **36**(4): 377–383, doi: 10.15949/j.cnki.0371-0025.2011.04.010.

Research paper

Analysis of Bottom Reverberation Intensity Under Beam-Controlled Emission Conditions in Deep Water

Guangying ZHENG[(1), (2), (3)*], Xiaowei GUO[(1), (2), (3)**], Fangwei ZHU[(1), (2), (3)],
Fangyong WANG[(1), (2), (3)], Linlang BAI[(1), (2), (3)]

[(1)] *Science and Technology on Sonar Laboratory*
Hangzhou, China

[(2)] *Hangzhou Applied Acoustics Research Institute*
Hangzhou, China

[(3)] *Hanjiang National Laboratory*
Wuhan, China

Corresponding Authors' e-mails: [*]276454158@qq.com, [**]645494944@qq.com

A comprehensive understanding of the characteristics and the formation mechanism of reverberation is the key to improving the performance of the active target detection. In response to the challenge of analyzing the intensity of bottom reverberation in typical deep-sea environments, this study proposes a prediction method for the bottom reverberation intensity under beam-controlled emission conditions. It explains the variation law of bottom reverberation intensity under beam-controlled emission conditions in typical deep-sea environments of the South China Sea through theoretical and simulation analyses. Reverberation intensity of the deep-sea bottom under beam-controlled emission conditions exhibits significant fluctuations during the duration of reverberations in the direct sound zone of the seabed. This phenomenon is closely related to the directionality of the source emission, leading to intermittent reverberation masking and detectable areas in the active sonar detection. In addition, the duration of the high-reverberation zone near the cutoff distance of the direct sound from the seabed is longer under the beam-controlled emission conditions of the emission array located within the surface waveguide layer of the deep sea during winter.

Keywords: bottom reverberation; deep water; beam controlled emission; ray theory.

1. Introduction

Reverberation is the primary background interference in the active sonar target detection, and a comprehensive understanding of its characteristics and formation mechanism is the key to improving the performance of the active target detection (Cui *et al.*, 2023; Hao *et al.*, 2023). Meanwhile, reverberation carries hidden ocean information that can be used for environmental parameter inversion, leading to the increasing attention toward reverberation research in the field of marine acoustics. At present, theoretical and experimental studies on shallow water reverberation are relatively more common than those on deep water.

With regard to the reverberation prediction theory, domestic and foreign scholars have established the theory of normal mode reverberation (Zhang *et al.*, 1987), the theory of ray reverberation (Lupien *et al.*, 1995), and the theory of parabolic equation reverberation (Collins, Evans, 1992). The intensity attenuation characteristics and spatial correlation characteristics of reverberation signals combined with experimental research have been simultaneously analyzed. Hence, the current study no longer develops this topic.

The characteristics of deep sea reverberation are significantly different from those of shallow sea reverberation, and their analysis and numerical modeling methods are different. A large number of the-

oretical and experimental studies on deep-sea reverberation have been conducted in the last century. Ellis and Crowe (1991) proposed a 3D seabed scattering function that included backscattering and lateral scattering based on the Lambert scattering model. This function was used in the numerical simulation of deep-sea bistatic reverberation and compared with experimental results. Mackenzie (1961) calculated the deep-sea bottom reverberation of near-bottom sound sources and receivers at specific frequencies and explained the applicable angle range of the scattering formula. Urick and Saling (1962) calculated the seabed backscatter excited by an explosive sound source and obtained a scattering intensity curve with angle. Ellis, Haller (1987), and Ellis, Crowe (1991) combined the Lambert scattering model with the surface scattering function based on the Kirchhoff approximation to propose a 3D seabed scattering function that included backscattering and lateral scattering. They used it in the numerical simulation of deep-sea bistatic reverberation and compared it with experimental measurement results. Williams and Jackson (1998) used the Kirchhoff approximation and the perturbation theory to describe the seafloor backscattering while discussing the effects of seafloor sediment layers and substrates on scattering.

With the support of relevant national plans, significant improvements have been achieved in recent years in experimental methods and equipment for deep-sea acoustics in China. A large number of deep-sea experiments have been conducted, promoting theoretical and experimental research on deep-sea bottom reverberation. Weng *et al.* (2014) conducted numerical simulations of local deep-sea bottom reverberation by using the ray method and provided preliminary explanations and analyses of experimental data. Guo *et al.* (2009) proposed an incoherent bottom reverberation signal model based on ray theory; this model simplified the calculation of the reverberation signal prediction. Xu *et al.* (2016) calculated deep-sea reverberation generated by the first bottom reflection of sound waves and obtained numerical results that were consistent with experimental data. Qin *et al.* (2019) proposed a deep-sea bottom reverberation model based on the ray theory for calculating local and bistatic reverberation and then compared the experimental data with the simulation results to obtain the seabed scattering coefficient of the experimental sea area. Xue *et al.* (2021) described interface reverberation as an incoherent superposition of different multipath reverberation fading processes and combined it with the physical mechanism of interface scattering. They established a reverberation intensity model with the physical parameters of the sea surface and seabed as variables.

The current study focuses on deep sea bottom reverberation characteristics under beam-controlled emission conditions in the actual work of active sonar.

It introduces a prediction method for deep sea bottom reverberation under beam-controlled emission conditions and explains the formation mechanism of deep sea bottom reverberation fluctuation laws through theoretical and simulation analyses.

2. Methodology

2.1. Model of deep-sea bottom reverberation intensity under beam-controlled emission conditions

The description of sound propagation in accordance with the ray theory is simple and intuitive, and thus, it is extremely helpful for explaining the results of other sound propagation models. Simultaneously, it directly establishes the relationship between the sound propagation distance and propagation time; hence, it can be used to predict reverberation intensity in high-frequency situations in the deep sea. Accordingly, this study uses the ray theory to establish a deep-sea bottom reverberation model for simulation analysis.

Reverberation signals can be expressed as the process of propagating sound signals excited by a sound source to a seafloor scatterer and then scattering them back to the receiver. Figure 1 shows a diagram of the formation of deep-sea bottom reverberation in the case of a combined transmitter and receiver. Considering that beam control in practical active sonar applications is generally the pitch angle, i.e., the vertical beam control, this study abstracts it as a vertical dimension emission array to study deep-sea bottom reverberation intensity characteristics under beam-controlled emission conditions.
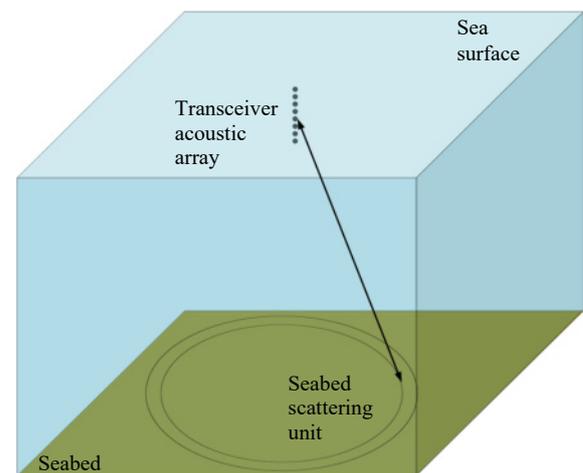

Fig. 1. Schematic of seabed scatterer division under beam-controlled emission conditions.

In the case of directional emission from a transceiver acoustic array, p_{inc} denotes the sound pressure transfer function from the sound source to the seabed scattering element under directional sound source radiation, p_{scatt} represents the sound pressure transfer

function from the seabed scattering element to the transceiver array under omnidirectional sound source radiation, $\mathbf{r}_i$ denotes the i-th eigenray, and the intensity of the incident wave propagating along the incident eigenray to the seabed scattering element $\mathrm{d}s$ is represented as $p_{\mathrm{inc}}^2(\mathbf{r}_i)$. The scattering wave intensity scattered by the seabed scattering unit $\mathrm{d}s$ can be expressed as

$$p_{\mathrm{inc}}^2(\mathbf{r}_i)\, f(\theta_{\mathrm{inc}}, \theta_{\mathrm{scatt}})\, \mathrm{d}s, \tag{1}$$

where $f(\theta_{\mathrm{inc}}, \theta_{\mathrm{scatt}})$ represents the backscatter function of the seabed, which is affected by the incident and scattering grazing angles, and $\mathrm{d}s$ is the area of the scattering unit. The reverberation intensity incident along the i-th eigenray to the scattering element returning to the receiving point along the j-th eigenray can be expressed as (Xue et al., 2021):

$$I_{\mathrm{scatt}(ij)} = \int p_{\mathrm{inc}}^2(\mathbf{r}_i)\, f(\theta_{\mathrm{inc}}, \theta_{\mathrm{scatt}})\, p_{\mathrm{scatt}}^2(\mathbf{r}_j)\, \mathrm{d}s. \tag{2}$$

The total bottom reverberation intensity at the receiving array is the sum of the reverberations that arrive along all the propagation paths (Xue et al., 2021):

$$I_{\mathrm{scatt}} = \sum_{i=1}^{N} \sum_{j=1}^{M} \int p_{\mathrm{inc}}^2(\mathbf{r}_i)\, f(\theta_{\mathrm{inc}}, \theta_{\mathrm{scatt}})\, p_{\mathrm{scatt}}^2(\mathbf{r}_j)\, \mathrm{d}s, \tag{3}$$

where N represents the number of incident eigenrays, and M represents the number of scattered eigenrays that correspond to the i-th incident eigenray, I_{scatt} is the total reverberation intensity.

The prerequisite for predicting the intensity of deep-sea bottom reverberation under beam-controlled emission conditions is the prediction of the deep-sea sound field transfer function under beam-controlled conditions and seabed scattering characteristics. The current study utilizes the ray model sound field calculation program BELHOP to predict the eigenray and the corresponding grazing angle, time delay, and the transfer function under directional source conditions. Meanwhile, a small-slope approximation (SSA) model is used to predict the seabed scattering characteristic.

2.2. SSA model

In this study, the small slope formalism is adopted for bottom interface scattering. This lowest-order SSA (Thorsos, Broschat, 1995; Broschat, Thorsos, 1997) models interface scattering strength in all orders of the surface height h and through the first-order derivatives of h (surface slope). Using the local SSA instead of the standard first-order perturbation approximation improves prediction accuracy at the cost of moderately increasing numerical complexity.

The SSA result for an incoherent component of the scattering cross section per unit area (per unit solid angle) for a random, rough interface is as follows (Grauss et al., 2002):

$$\sigma \frac{1}{8} \left| \frac{\beta}{|Q_h|\, Q_z} \right|_{\mathrm{int}}^2, \tag{4}$$

where β is an algebraic form that depends on the boundary conditions that are prevailing at the interface. $|Q_h|$ and Q_z are given by (Grauss et al., 2002):

$$|Q_h| = k_0 \sqrt{a^*}, \tag{5}$$

$$Q_z = -k_0 (\sin\theta_{\mathrm{inc}} + \sin\theta_{\mathrm{scatt}}), \tag{6}$$

where $a^* = \cos^2\theta_{\mathrm{inc}} + \cos^2\theta_{\mathrm{scatt}} - 2\cos\theta_{\mathrm{inc}}\cos\theta_{\mathrm{scatt}}\cos\phi_{bi}$ and $k_0 = \frac{2\pi f}{c_0}$ denotes the acoustic wavenumbers; θ_{inc} denotes the incident grazing angle; θ_{scatt} denotes the scattered grazing angle; ϕ_{bi} denotes the bistatic angle, which is defined as the difference in azimuth between the incident and scattered directions; I in Eq. (7) denotes the integral that involves the spatial spectrum of roughness, and it is given by (Grauss et al., 2002):

$$I(\alpha) = \int_0^\infty J_0(y)\, y \exp\left(-\alpha y^{2\nu}\right) \mathrm{d}y, \tag{7}$$

where $\nu \equiv \frac{(\gamma_2 - 2)}{2}$, γ_2 denotes the roughness spectral exponent, with $\gamma_2 \in (2, 4)$, J_0 is the 0-th-order Bessel function of the first kind, and α is given by (Grauss et al., 2002):

$$\alpha = \frac{(h_{\mathrm{rms}} Q_z)^2}{(2h_0\, |Q_h|)^{2\nu}} \frac{\Gamma(1-\nu)}{\Gamma(1+\nu)}, \tag{8}$$

where h_{rms}^2 denotes the mean-square roughness, which is given by (Grauss et al., 2002):

$$h_{\mathrm{rms}}^2 = \frac{\pi w_2}{h_0^2 \nu}, \tag{9}$$

where w_2 is the input rough spectral intensity, and h_0 is a normalizing reference distance of 1 m.

Notably, once the real or imaginary part of a sound speed acquires dependence on frequency, then α, β, and σ_{int} acquire complicated frequency dependencies.

3. Analysis of deep-sea sound field characteristics under beam-controlled conditions

This section conducts a simulation analysis of the sound field distribution under the conditions of omnidirectional and directional sound sources as a prerequisite for the simulation analysis of deep-sea bottom reverberation under beam-controlled emission conditions.

The simulation analysis focuses on the typical deep-sea hydrological environment of the South China Sea,

with a depth of 4000 m. The historically measured summer and winter sound speed profiles are presented in Fig. 2. The material of the seabed is fine sand, with a sound speed of 1753 m/s, a density of 1.957 g/cm^3, and an attenuation coefficient of 0.51 dB/m·Hz. To compare the changes in the sound field distribution caused by beam-controlled emission, simulation calculations were conducted using the BELLHOP (PORTER, 2011) ray model sound field calculation program, and sound field distributions were given for omnidirectional and directional source emission cases. The sound source array under beam-controlled emission conditions is an eight-element vertical array, with a center frequency of 1 kHz. The array elements are arranged at a half-wavelength spacing of 0.75 m, and the center depth of the array is arranged at a depth of 10 m underwater, such that the depth of the entire sound source array is within 50 m. The array is arranged inside the surface waveguide to illustrate sound field differences during winter when a surface waveguide is present. The natural directionality of the eight-element transceiver combined with a vertical array is shown in Fig. 3. Its main lobe corresponds to 0°, with the first side lobe appear-

Fig. 3. Natural directivity of eight-element source array.

ing at an angle of ±21°, the second side lobe appearing at an angle of ±38°, and the third side lobe appearing at an angle of ±61°.

The distribution of sound fields excited by directional and omnidirectional sources in the deep sea of the South China Sea during summer is shown in Fig. 4. In contrast with the excited sound field of omnidirectional sources, the main energy emitted by directional sources leads to a clear high-sound-intensity region at 18–19 km, which corresponds to the main lobe with a grazing angle of 0° in the directional pattern. The leakage of the side-lobe energy leads to intermittent high-sound-intensity regions within the range of 0–15 km. The high-sound-intensity region at 10 km corresponds to the first side lobe with a grazing angle of ±21°. The high-sound-intensity region at 5 km corresponds to the second side lobe with a grazing angle of ±38°. The high-sound-intensity region at 2.2 km corresponds to the third side lobe with a grazing angle of ±61°. Notably, the cutoff distance of the deep-sea bottom direct sound zone in the summer hydrological environment is about 18–19 km, and the sound field excited by directional sources does not significantly change the cutoff distance of the deep-sea bottom direct sound zone.

The distribution of sound fields excited by directional and omnidirectional sources during winter (with the presence of a 50 m surface waveguide) is shown in Fig. 5. In contrast with the excited sound field of an omnidirectional source, the main energy emitted by a directional source leads to a significant high-sound-intensity region at 18–23 km, which corresponds to the main lobe with a grazing angle of 0° in the directivity pattern. The leakage of the same side-lobe energy leads to intermittent high-sound-intensity regions within the range of 0–15 km. The high-sound-intensity region at 10 km corresponds to the first side lobe with a grazing angle of ±21°. The high-sound-intensity region at 5 km corresponds to the second side


Fig. 2. Sound speed profile (SSP) in deep water. The blue solid line represents the summer sound speed profile, while the red dashed line represents the winter sound speed profile. The subplot of Fig. 2 gives the variation of the SSP over a depth range of 100 m, emphasizing the difference between the summer and winter SSP, i.e., a surface isothermal layer of 50 m occurs during winter in blue solid line.

a)
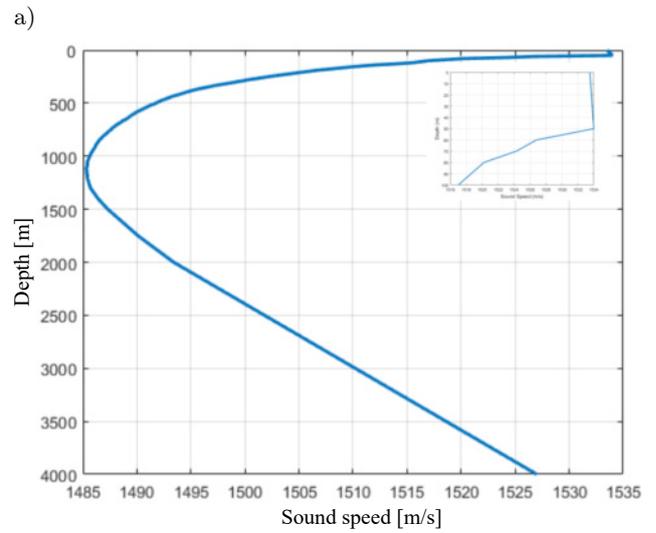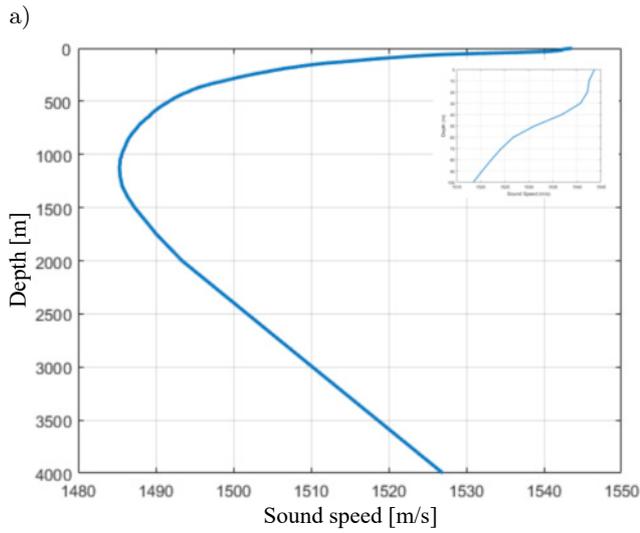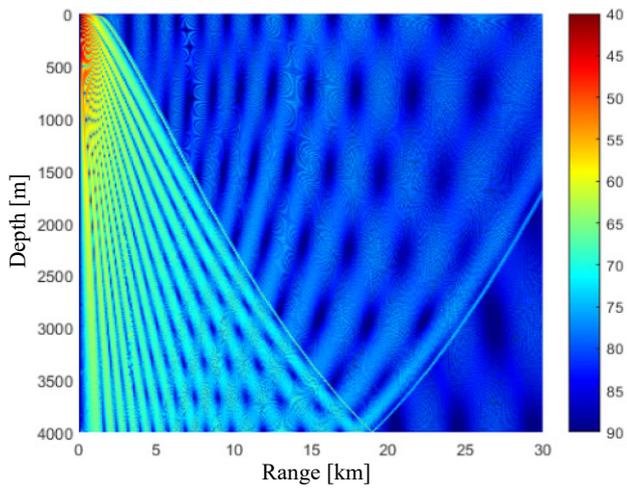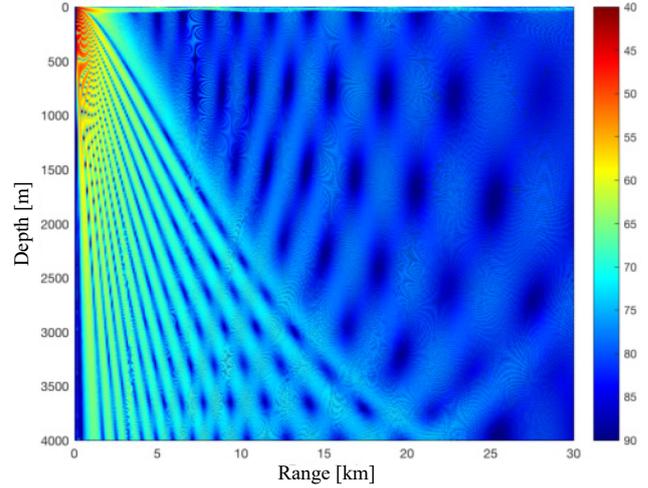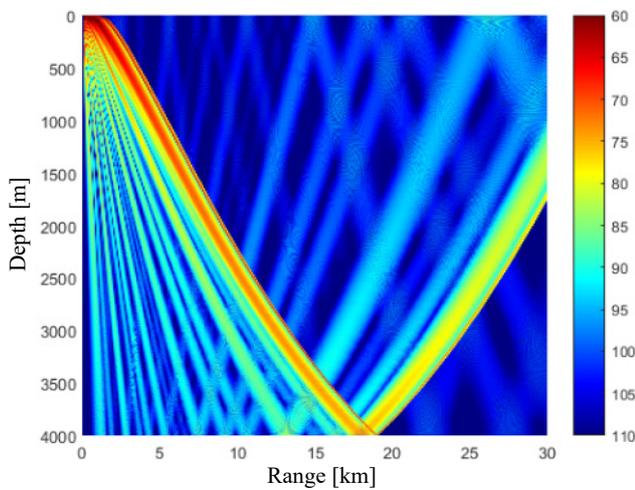
b) Freq = 1000 Hz, Sd = 10 m

c) Freq = 1000 Hz, Sd = 10 m
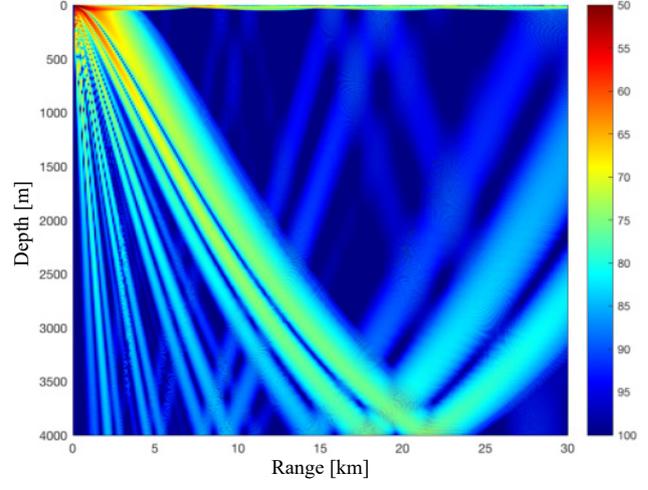
Fig. 4. Transmission loss during summer in the South China Sea: a) sound speed profile in summer; b) omnidirectional sources; c) directional sources.

a)

b) Freq = 1000 Hz, Sd = 10 m


c) Freq = 1000 Hz, Sd = 10 m

Fig. 5. Transmission loss during winter in the South China Sea: a) sound speed profile in winter; b) omnidirectional sources; c) directional sources.

lobe with a grazing angle of ±38°. The high-sound-intensity region at 2.2 km corresponds to the third side lobe with a grazing angle of ±61°. Notably, under the condition of omnidirectional source radiation, the cutoff distance of the deep-sea bottom that directly reaches the sound zone is about 22 km in the winter hydrological environment, which is far from the 18 km in the summer hydrological environment, because the surface waveguide leaks energy, causing it to propagate further. The cutoff distance of the sound field excited by the directional source from the seabed to the sound zone extends to 23–24 km, because the emission array beam-controlled emission within the surface waveguide layer causes more energy to concentrate on the surface waveguide layer. Moreover, the amount of energy leaked from the surface waveguide layer to the seabed to the direct sound zone increases.

On the basis of the analysis of the sound field characteristics in the direct sound zone of the deep-sea bottom under beam-controlled conditions, the next section analyzes the bottom reverberation characteristics in typical deep-sea environments.

4. Analysis of bottom reverberation intensity under beam-controlled emission conditions

Before analyzing and calculating deep-sea bottom reverberation intensity, obtaining the backscatter intensity of the seabed is necessary. For the simulation environment in Sec. 3, the seabed material is fine sand, and detailed parameters can be found in Sec. 3. On the basis of SSA to calculate backscatter intensity, the input rough spectral intensity of the model is 0.0004, and the rough spectral exponent is 2.6. On the basis of the aforementioned model parameters, the variation of the backscattering intensity of the deep-sea seabed with grazing angle is calculated at a frequency of 1 kHz. As shown in Fig. 6, the backscattering intensity of the rough interface increases with an increase

in the grazing angle. This rough interface backscattering intensity is the input of the model for calculating bottom reverberation intensity.

On the basis of this scattering model and Eq. (3), combined with the sound field transfer function predicted by the ray model, the deep-sea bottom reverberation intensity excited by omnidirectional and directional sources under summer and winter conditions in the South China Sea is calculated.

The variation in deep-sea bottom reverberation intensity over time caused by omnidirectional and directional sources in the South China Sea during summer is shown in Fig. 7. For omnidirectional and directional sources, bottom reverberation occurs after 5 s, which corresponds to the time when bottom vertical reflected reverberation occurs. The reverberation intensity excited by omnidirectional sources monotonically decreases within 5 to 27 s, while the reverberation intensity excited by directional sources exhibits significant fluctuations within 5 to 27 s. This phenomenon exerts a significant effect on the active sonar target detection, resulting in intermittent reverberation masking and detectable areas.

Fig. 7. Variation of bottom reverberation intensity with time: a) omnidirectional source; b) directional source.

Fig. 6. Variation of seabed backscatter intensity with grazing angle.
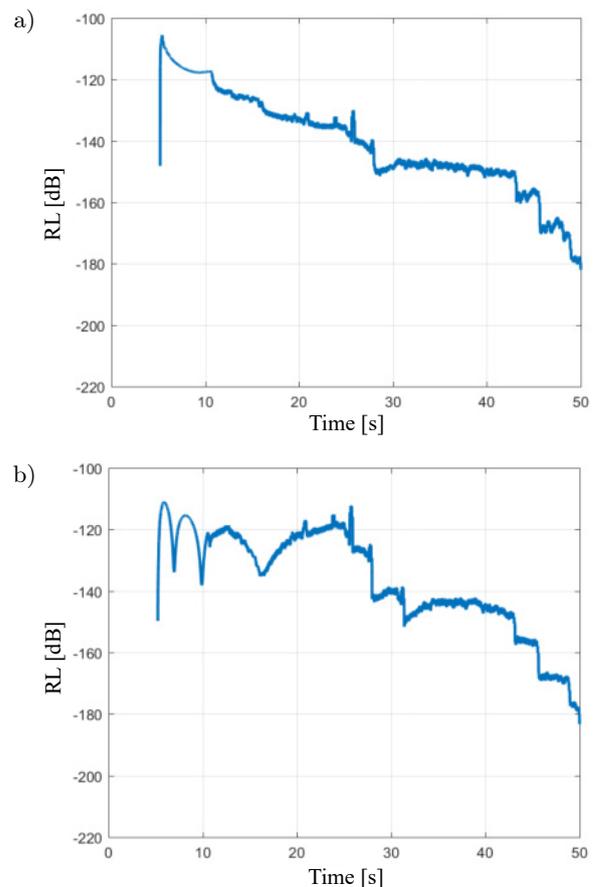
By converting horizontal axis time into the acoustic path that corresponds to the active sonar, the variation

of deep sea bottom reverberation intensity excited by omnidirectional and directional sources in the South China Sea during summer with an acoustic path can be obtained, as shown in Fig. 8. Evident peaks are observed in the reverberation intensity at 4, 6, 9, and 18 km. This finding is related to the emission directionality of the eight-element vertical emission array.

a)

b)
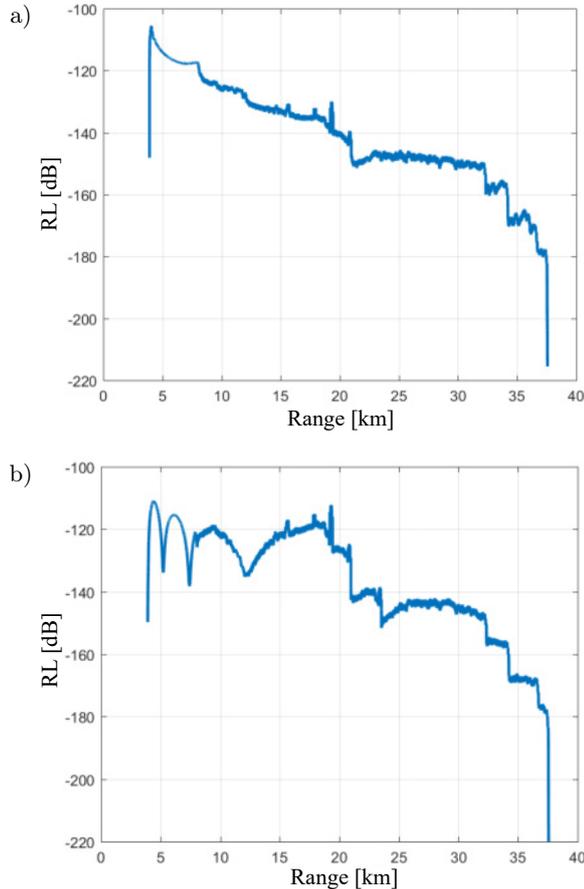
Fig. 8. Variation of bottom reverberation intensity with acoustic path: a) omnidirectional source; b) directional source.

To further illustrate the corresponding relationship among the peak values of intermittent bottom reverberation intensity, the bottom scattering area, and emission directionality, the acoustic path is converted into the horizontal distance from the transceiver array. The variation in deep-sea bottom reverberation intensity excited by omnidirectional and directional sources in the South China Sea during summer with a horizontal distance can be obtained as shown in Fig. 9. At a depth of 3995 m, the variation in propagation loss with horizontal distance is also given, as shown in Fig. 10. Comparing the variation of reverberation intensity in Fig. 9 with the variation of propagation loss in Fig. 10, four high-energy regions of reverberation intensity are observed, corresponding to the four high-energy regions of transmission loss and correspond-

a)

b)
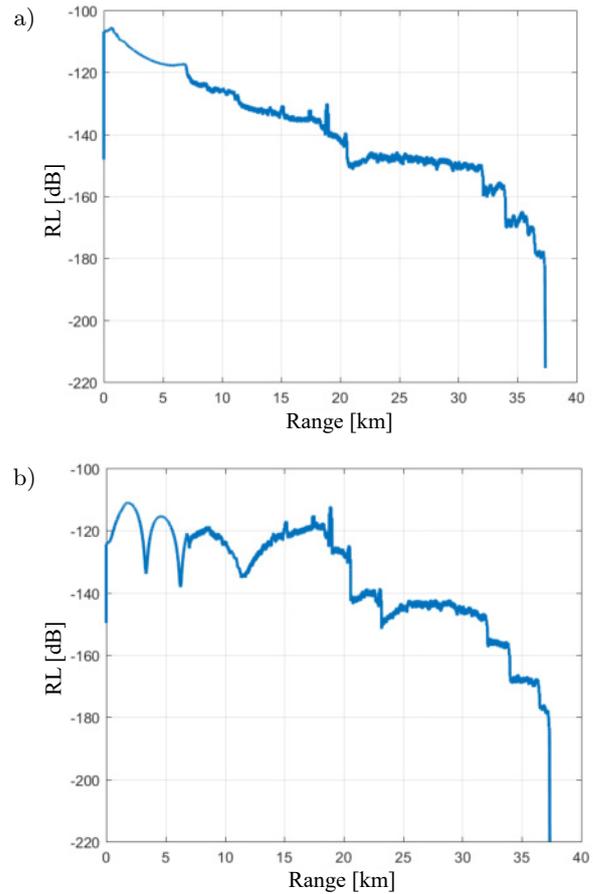
Fig. 9. Variation of bottom reverberation intensity with horizontal range: a) omnidirectional source; b) directional source.
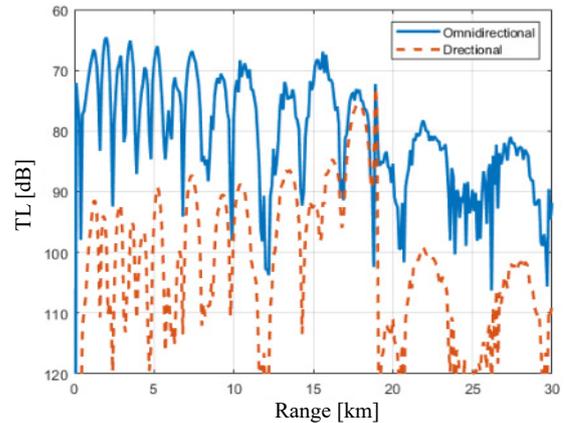
Fig. 10. Variation of transmission loss with horizontal range for a receiver depth of 3995 m. The blue solid line represents the transmission loss of sound field excited by an omnidirectional point source, while the red dashed line represents the transmission loss of sound field excited by a directional transmitter array.

ing to the main lobe and the three side lobes of directional sound sources. The high-reverberation zone that appears at 15–19 km corresponds to the main lobe with a grazing angle of 0° in the directivity pat-

tern. The leakage of side-lobe energy leads to intermittent high-reverberation zones within the range of 0–15 km. The high-reverberation zone at 10 km corresponds to the first side lobe with a grazing angle of ±21°. The high-reverberation zone at 5 km corresponds to the second side lobe with a grazing angle of ±38°. The high-reverberation zone at 2.2 km corresponds to the third side lobe with a grazing angle of ±61°.

The variation of deep-sea bottom reverberation intensity over time caused by omnidirectional and directional sources in the South China Sea during winter is shown in Fig. 11. Similar to that during summer, bottom reverberation occurs after 5 s, which corresponds to the time when vertically reflected reverberation occurs on the seabed. The reverberation intensity excited by omnidirectional sources monotonically decreases within 5 to 30 s, while the reverberation intensity excited by directional sources exhibits significant fluctuations within 5 to 30 s. This phenomenon exerts a significant effect on the active sonar target detection, resulting in intermittent reverberation masking and detectable areas.
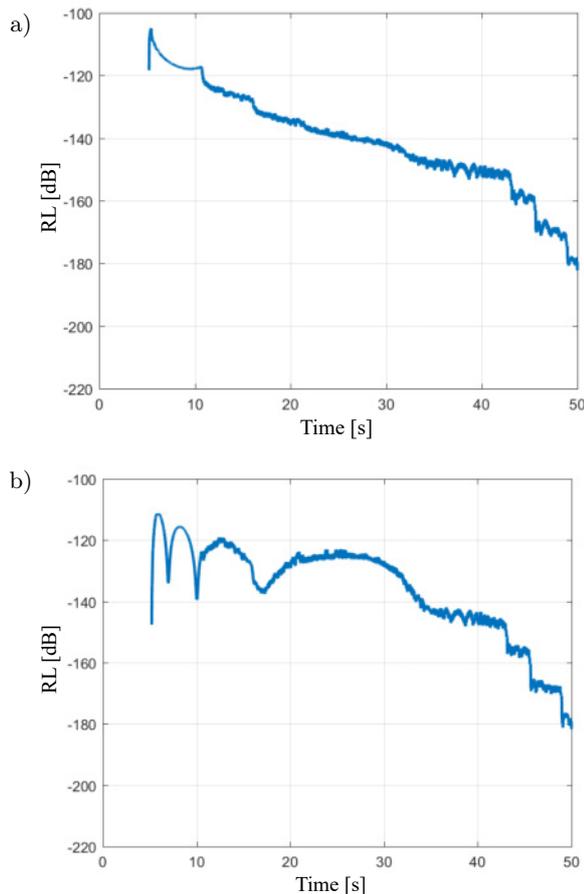

Fig. 11. Variation of bottom reverberation intensity with time: a) omnidirectional source; b) directional source.

By converting horizontal axis time into the acoustic path that corresponds to the active sonar, the variation of the deep-sea bottom reverberation intensity excited by omnidirectional and directional sources in the South China Sea during winter can be obtained with respect to the acoustic path, as shown in Fig. 12. Evident peaks can be seen in the reverberation intensity at 4, 6, 9, and 20 km. These peaks are related to the emission directionality of the eight-element vertical emission array.
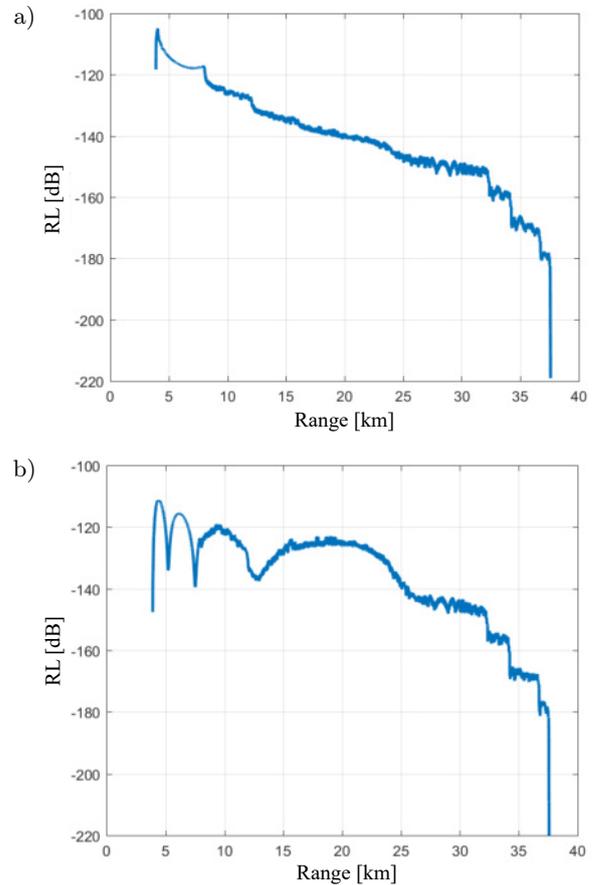

Fig. 12. Variation of bottom reverberation intensity with acoustic path: a) omnidirectional source; b) directional source.

To further illustrate the corresponding relationship among the peak value of intermittent bottom reverberation intensity, bottom scattering area, and emission directionality, the acoustic path is converted into the horizontal distance from the transceiver array. The variation of deep-sea bottom reverberation intensity excited by omnidirectional and directional sources in the South China Sea during winter with horizontal distance can be obtained as shown in Fig. 13. At a depth of 3995 m, the variation in transmission loss with the horizontal distance is also given, as shown in Fig. 14. Comparing the variation of reverberation intensity in Fig. 13 with the variation of transmission loss in Fig. 14, four high-energy regions of reverberation intensity can also be observed, corresponding to the four high-energy regions of transmission loss and corresponding to the main lobe and three side lobes

a)

b)
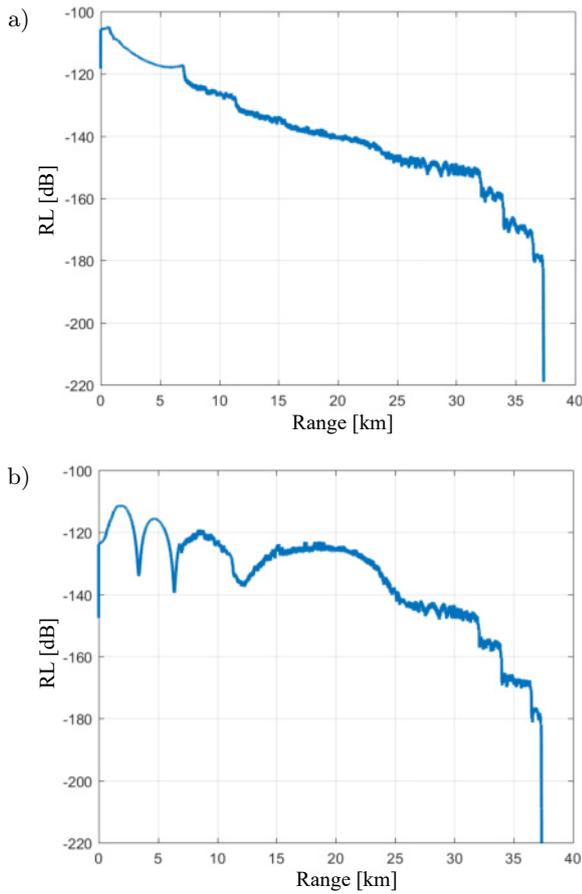
Fig. 13. Variation of bottom reverberation intensity with horizontal range: a) omnidirectional source; b) directional source.
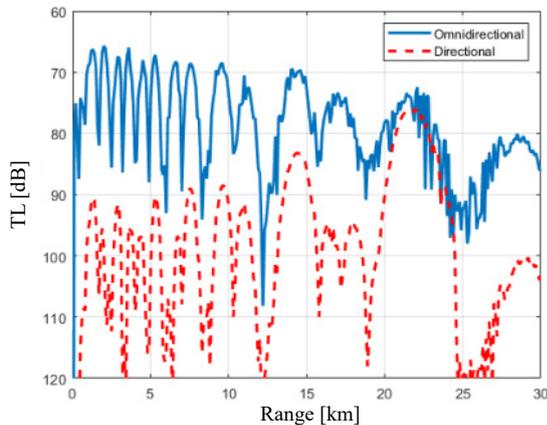

Fig. 14. Variation of transmission loss with horizontal range for a receiver depth of 3995 m. The blue solid line represents the transmission loss of sound field excited by an omnidirectional point source, while the red dashed line represents the transmission loss of sound field excited by a directional transmitter array.

of directional sound sources. The high reverberation zone appearing at 15–23 km corresponds to the main lobe with a grazing angle of 0° in the directivity pat-

tern. The leakage of side-lobe energy leads to intermittent high-reverberation zones within the range of 0–15 km. The high-reverberation zone at 10 km corresponds to the first side lobe with a grazing angle of ±21°. The high-reverberation zone at 5 km corresponds to the second side lobe with a grazing angle of ±38°. The high-reverberation zone at 2.2 km corresponds to the third side lobe with a grazing angle of ±61°.

To compare the differences in the distribution of bottom reverberation intensity caused by beam-controlled emission in the South China Sea during summer and winter, Figs. 10 and 14 were drawn together, as shown in Fig. 15. Except for the differences in hydrological conditions (sound speed profile), the two curves are plotted under the same simulation conditions. The variation trend of reverberation intensity within a range of 12 km is nearly consistent, including the high-reverberation zone caused by side-lobe energy leakage. However, significant differences exist in the high-reverberation zone caused by the directional main lobe emission. The major issues are:

1) For the high-reverberation zone near a range of 18 km excited by the main lobe energy, bottom reverberation intensity during winter is weaker than that during summer, because the existence of an isothermal layer on the surface during winter results in most of the energy radiated by small grazing angles being bound to the surface waveguide layer. The energy reaching the direct sound zone on the seabed beyond 15 km is lower than that in the absence of a surface waveguide layer.

2) For the high-reverberation zone excited by the main lobe energy, the duration (corresponding horizontal distance) of the main lobe's high-reverberation zone is longer during winter, and the effect on the active target detection in deep-sea environments is more significant. The reason for this finding is that after the beam-controlled
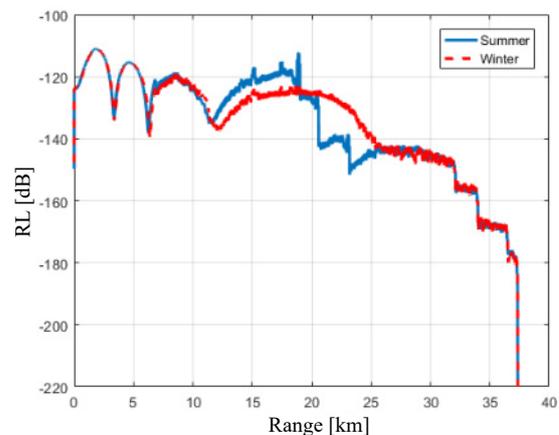
Fig. 15. Bottom reverberation during summer and winter under beam-controlled emission conditions.

emission of the emission array inside the surface waveguide layer, the sound energy leakage of the surface waveguide layer expands the direct sound zone of the deep sea during winter, and the acoustic ray with a small grazing angle can reach up to 23 km, resulting in an increase in the duration of the main lobe's high-reverberation zone.

5. Conclusions

In response to the challenge of analyzing the intensity of bottom reverberation in typical deep-sea environments, this study proposes a prediction method for the intensity of bottom reverberation under beam-controlled (shaded) emission conditions and explains the variation law of bottom reverberation intensity under beam-controlled emission conditions through theoretical and simulation analyses in typical deep-sea environments of the South China Sea with a seabed material of fine sand. The conclusions drawn are:

1) Deep-sea bottom reverberation intensity under beam-controlled emission conditions exhibits significant fluctuations during the duration of reverberation in the direct sound zone of the seabed. This phenomenon is closely related to the directionality of source emission, leading to intermittent reverberation masking and detectable areas in the active sonar detection.

2) For the high-reverberation zone near the cutoff distance of the direct sound from the seabed excited by the main lobe energy of the directional source, the reverberation intensity during winter is weaker than that during summer. The reason for this finding is that the existence of a surface isothermal layer during winter results in most of the energy emitted by the small grazing angle being bound to the surface waveguide layer. Meanwhile, the energy reaching the cutoff distance of the direct sound from the seabed is lower than that without surface waveguide layer.

3) Under the beam-controlled emission conditions of the emission array located within the surface waveguide layer of the deep sea during winter, the duration of the high-reverberation zone near the cutoff distance of the direct sound from the seabed is longer, because the sound energy leakage from the surface waveguide layer expands the direct sound zone of the deep sea during winter. The acoustic ray with a small grazing angle can reach further distance, resulting in an increase in the duration of the high-reverberation zone of the main lobe. This phenomenon will have a more significant effect on the active target detection in deep-sea environments.

The analysis of the distribution characteristics of deep-sea bottom reverberation intensity in this study provides guidance for the suppression of deep-sea active sonar reverberation, and the evaluation and reasonable application of deep-sea active sonar detection performance in actual combat.

References

1. Broschat S.L., Thorsos E.I. (1997), An investigation of the small slope approximation for scattering from rough surfaces. Part II. Numerical studies, *The Journal of the Acoustical Society of America*, **101**(5): 2615–2625, doi: 10.1121/1.418502.

2. Collins M.D., Evans R.B. (1992), A two-way parabolic equation for acoustic backscattering in the ocean, *The Journal of the Acoustical Society of America*, **91**(3): 1357–1368, doi: 10.1121/1.402465.

3. Cui X., Chi C., Li S., Li Z., Li Y., Huang H. (2023), Waveform design using coprime frequency-modulated pulse trains for reverberation suppression of active sonar, *Journal of Marine Science and Engineering*, **11**(1): 28, doi: 10.3390/jmse11010028.

4. Ellis D.D., Crowe D.V. (1991), Bistatic reverberation calculations using a three-dimensional scattering function, *The Journal of the Acoustical Society of America*, **89**(5): 2207–2214, doi: 10.1121/1.400913.

5. Ellis D.D., Haller D.R. (1987), A scattering function for bistatic reverberation calculations, *The Journal of the Acoustical Society of America*, **82**(1): 124, doi: 10.1121/1.2024654.

6. Grauss R.C., Gragg R.F., Wurmser D., Fialkowski J.M., Nero R.W. (2002), *Broadband models for predicting bistatic bottom, surface and volume scattering strengths*, Naval Research Laboratory Report.

7. Guo X.Y., Su S.J., Wang Y.K. (2009), Research on the signal modeling method for sea bottom reverberation based on ray theory [in Chinese], *Technical Acoustics*, **28**(3): 203–207.

8. Hao Y. *et al.* (2023), Underwater reverberation suppression via attention and cepstrum analysis-guided network, *Journal of Marine Science and Engineering*, **11**(2): 313, doi: 10.3390/jmse11020313.

9. Lupien V.H., Bondaryk J.E., Baggeroer A.B. (1995), Acoustical ray-tracing insonification software modeling of reverberation at selected sites near the

Mid-Atlantic Ridge, *The Journal of the Acoustical Society of America*, **98**(5): 2987, doi: 10.1121/1.413929.

10. Mackenzie K.V. (1961), Bottom reverberation for 530- and 1030-cps sound in deep water, *The Journal of the Acoustical Society of America*, **33**(11): 1498–1504, doi: 10.1121/1.1908482.

11. Porter M.B. (2011), *The Bellhop Manual and User's Guide: Preliminary Draft*.

12. Qin J.X., Wang L.H., Li Z.L. (2019), Theory and experiment of large-depth reverberation in deep water [in Chinese], *Technical Acoustics*, **38**(5): 95–96.

13. Thorsos E.I., Broschat S.L. (1995), An investigation of the small slope approximation for scattering from rough surfaces. Part I. Theory, *The Journal of the Acoustical Society of America*, **97**(4): 2082–2093, doi: 10.1121/1.412001.

14. Urick R.J., Saling D.S. (1962), Backscattering of explosive sound from the deep-sea bed, *The Journal of the Acoustical Society of America*, **34**(11): 1721, doi: 10.1121/1.1909106.

15. Weng J.B., Li F.H., Liu J.J. (2014), The preliminary study on bottom reverberation model in deep water [in Chinese], *Technical Acoustics*, **33**(S2): 71–73.

16. Williams K.L., Jackson D.R. (1998), Bistatic bottom scattering: Model, experiments, and model data comparison, *The Journal of the Acoustical Society of America*, **103**(1): 169–181, doi: 10.1121/1.421109.

17. Xu L., Yang K., Guo X., Li H. (2016), Bistatic bottom reverberation in deep ocean: Modeling and data comparison, [in:] *OCEANS 2016 – Shanghai*, pp. 1–5, doi: 10.1109/OCEANSAP.2016.7485385.

18. Xue R.Z., Duan R., Yang K.D., Ma Y.L., Guo Y. (2021), Modeling and analysis of monostatic incoherent boundary reverberation intensity in deep water, *Acta Acoustica*, **46**(6): 926–938, doi: 10.15949/j.cnki.0371-0025.2021.06.014.

19. Zhang R.H., Jin G.L. (1987), Normal-mode theory of the average reverberation intensity in shallow water, *Journal of Sound and Vibration*, **119**(2): 215–223, doi: 10.1016/0022-460X(87)90450-0.

Research Paper

Numerical Study on the Dynamics of a Charged Bubble in the Acoustic Field

Liang LV[(1)], Fei LIU[(2)], Yawei LI[(3)]*

[(1)] *School of Mechano-Electronic Engineering, Suzhou Vocational University*
Suzhou, China

[(2)] *Department of Sports Health and Art Education, Hebei Petroleum University of Technology*
Chengde, China

[(3)] *Department of the Party and the Mass, Hebei Petroleum University of Technology*
Chengde, China

*Corresponding Author e-mail: liyawei_123321@163.com

In this paper, the dynamics of an acoustic bubble with a constant charge in compressible liquid are investigated numerically, which is based on the Gilmore-NASG model to estimate the radial oscillations. The cavitation effects are enhanced due to the presence of the charge on the bubble surface. The obtained results from the present model are compared with that calculated by the previous model within a wide range of parameters (e.g., charge, acoustic pressure amplitude, ultrasound frequency, and liquid temperature). The similar influences of these parameters on bubble collapse intensity can be observed from both models. Since the present model fully considers the compressibility of gas and liquid, it can be applied to a wider parameter range and leads to the larger predicted values. The research in this paper can provide important insights about the effects of charge on bubble dynamics and the acoustic cavitation applications (e.g., sonochemistry, water treatment, and food industry).

Keywords: charged bubble dynamics; acoustic cavitation; bubble collapse intensity; Gilmore-NASG model.

1. Introduction

Cavitation bubbles exist widely in nature, e.g., the snapping shrimp uses cavitation bubbles formed by the rapid closure of its claws to stun its prey; in a fast flowing system, cavitation bubbles are prone to occur if the channel suddenly narrows and then widens. The phenomenon of cavitation in the liquid is concerned because of its damage to hydraulic machinery and ship propellers (SEZEN *et al.*, 2021; WANG *et al.*, 2022). Due to unique physical and chemical phenomena (CLEVE *et al.* 2019; DEHANE *et al.*, 2022; KERBOUA *et al.*, 2021; LV *et al.*, 2019; LV, LIU, 2023; TIAN *et al.*, 2023) (e.g., liquid jet, free radicals formation, radiation pressure, and acoustic microstreaming) caused by the bubble oscillations, cavitation has been contributed to various applications such as water treatment (FERKOU *et al.*, 2015), petroleum hydrocarbons

degradation (LEI *et al.*, 2020), nanoparticle synthesis (POKHREL *et al.*, 2016), and so on.

The charges carried on the bubble surface have been reported by many research groups. The zeta potential of the bubble measured experimentally by using the microelectrophoresis technique indicated that the effect of the PH solution on the variation of the bubble zeta potential depended not only on the type of the metal ions but also on the electrolyte concentration (YANG *et al.*, 2001). They also found the charge polarity varied at different solutions (e.g., the bubble was the negative charge in NaCl solutions, while its charge polarity reversed in multivalent metal ions solutions). TAKAHASHI (2005) found that the bubbles were negatively charged under a wide range of the PH condition and positively charged under strongly acidic conditions. In (LEE, CHOI, 2020), the stable light emission of a single bubble sonoluminescence (SBSL bubble)

with charge in water was investigated for the first time. The results revealed that the SBSL bubble was positively charged and suggested that it was necessary for analyses of the SBSL bubble to take the electrical properties into consideration. The behaviour of laser-generated bubbles in an electric field was studied by PHUKAN *et al.* (2023). They found the maximum bubble radius increased with the increase of the electric field intensity. This effect was more pronounced in the presence of an acetone medium and decreased successively in ethanol and water media owing to their varying magnitudes of electrical conductivity. In addition to experimental researches of the bubbles in an electric field, theoretical studies have also made great progress. The volume mode and shape model dynamics were examined in a weakly viscous dielectric fluid under the uniform and the axisymmetric straining electric field (OH *et al.*, 2001). A model was built describing the violent collapse of the bubble in the homogeneous, irrotational, solenoidal, and unsteady electric field (SPELT, MATAR, 2006). The research group optimized the bubble model in the electric field, and further analyzed the bubble dynamics of violent collapse, translation and shape deformations (SHAW *et al.*, 2009).

Nevertheless, there are few studies on the charged bubbles that undergo acoustic cavitation, which has been applied in many fields (e.g., ultrasonic cleaning, drug delivery, and inactivate viruses). Based on the study of the stability of a charged bubble in the dielectric liquid in (GRIGOR'EV, ZHAROV, 2000), the model of a charged bubble excited by ultrasound wave has been deduced by HONGRAY *et al.* (2014; 2015). The study results proved that the effective surface was reduced due to the presence of charge. Compared with the uncharged bubble oscillations under acoustic excitation, the bubble expanded to a larger radius and compressed to a smaller size, which in turn caused the bubble collapse to be stronger. The bifurcation diagrams have also been studied to some extent, with the presence of charge leading to advance bifurcations.

The theoretical and experimental researches have been proved a lot on the charge bubbles in liquid. The theoretical study of bubble dynamics in the acoustic field is mainly based on uncharged bubbles, which will lead a bias in calculation results. The model of charged bubbles was derived from the Keller-Miksis equation (HONGRAY *et al.*, 2014; 2015). When the effect of charge is taken into account, the bubble collapse will be enhanced resulting in the significant increase of the speed of a bubble wall, and the Mach number can easily reach 1, which is the critical value that the Keller-Miksis equation is safely adopted (ZILONOVA *et al.*, 2018). Therefore, it is necessary to establish a model that can be used in large parameter intervals, especially the high pressure amplitude and low frequency excitation. In this paper, based on the Gilmore-NASG model (DENNER *et al.*, 2021), a more applicable dynamic model of charged bubbles is established and compared with the model (HONGRAY *et al.*, 2014; 2015) in detail within a wide range of parameters. Subsequently, the sections of present paper are organized as follows. In Sec. 2, the model of a charged bubble and the numerical method are introduced. In Sec. 3, the calculation results of the present model are quantitatively compared with those of the previous model. In Sec. 4, the main findings of the present paper are summarized.

2. Mathematical model and simulation method

For simplicity, the following assumptions are used in physical models: (1) the bubble is spherically symmetric; (2) the fluid is Newtonian and compressible; (3) the buoyancy force and gravity are neglected; (4) the bubble-bubble interaction is neglected; (5) the thermal conductivity, phase change and mass transport across the bubble-liquid interface are neglected. The radial dynamics of the bubble is governed by the Gilmore equation (DENNER, 2021):

$$\left(1 - \frac{\dot{R}}{c_l}\right) R\ddot{R} + \frac{3}{2}\left(1 - \frac{\dot{R}}{3c_l}\right)\dot{R}^2 = \left(1 + \frac{\dot{R}}{c_l}\right) H$$
$$+ \left(1 - \frac{\dot{R}}{c_l}\right) R\frac{\dot{H}}{c_l}, \quad (1)$$

where R is the instantaneous radius of the bubble, the overdot denotes the time derivative, c_l is the speed of sound in the liquid at the bubble wall, H is the difference between the enthalpy of the liquid at the bubble wall and at infinity. The state of gas and vapor inside the bubble, and the liquid outside the bubble are described by the NASG equation of state, and the expression is (DENNER, 2021):

$$p(v, T) = \frac{(\Gamma - 1)C_V T}{v - b} - B, \quad (2)$$

where p is the pressure, v is the specific volume, T is the temperature, Γ is the polytropic exponent, C_v is the heat capacity at a constant volume, b is the co-volume that represents the volume occupied by the individual molecules, B is a pressure constant that models molecular attraction.

H and c_l are defined as Eqs. (3) and (4), respectively:

$$H = \frac{\Gamma_l}{\Gamma_l - 1}\left(\frac{p_l + B_l}{\rho_l} - \frac{p_\infty + B_l}{\rho_\infty}\right) - b_l \frac{p_l - p_\infty}{\Gamma_l - 1}, \quad (3)$$

$$c_l = \sqrt{\Gamma_l \frac{p_l + B_l}{\rho_l(1 - b_l\rho_l)}}, \quad (4)$$

where Γ_l is the liquid polytropic exponent, p_l is the pressure in the liquid at the bubble wall, B_l is the liquid

pressure constant, p_∞ is the liquid pressure at infinity, ρ_l and ρ_∞ are the densities of liquid at the bubble wall and at infinity, respectively. The expressions of p_l, p_∞, ρ_l, and ρ_∞ are given as Eqs. (5)–(8):

$$p_l = p_g - \frac{2\sigma}{R} - 4\mu\frac{\dot{R}}{R} + \frac{Q^2}{8\pi\varepsilon R^4}, \tag{5}$$

$$p_\infty = P_{l,0} - P_a\sin(2\pi ft), \tag{6}$$

$$\rho_l = \frac{K_l(p_l + B_l)^{\frac{1}{\Gamma_l}}}{1 + b_l K_l(p_l + B_l)^{\frac{1}{\Gamma_l}}}, \tag{7}$$

$$\rho_\infty = \frac{K_l(p_\infty + B_l)^{\frac{1}{\Gamma_l}}}{1 + b_l K_l(p_\infty + B_l)^{\frac{1}{\Gamma_l}}}, \tag{8}$$

where σ is the surface tension coefficient of the liquid, μ is the viscosity of the liquid, Q is the charge at a bubble surface, ε (i.e., $\varepsilon = 85\varepsilon_0$, ε_0 is the vacuum permittivity) is the liquid permittivity, $P_{l,0}$ is an ambient pressure in the liquid, P_a is the acoustic pressure, f is the ultrasonic frequency, K_l and p_g are the constants representing the liquid reference state and gas pressure inside the bubble, respectively, and expressions are defined as Eqs. (9) and (10):

$$K_l = \frac{\rho_{l,0}}{(\rho_{l,0} + B_l)^{\frac{1}{\Gamma_l}}(1 - b_l\rho_{l,0})}, \tag{9}$$

$$p_g = (P_{g,0} + B_g)\left[\frac{\rho_{g,0}\left(\frac{R_0}{R}\right)^3(1 - b_g\rho_{g,0})}{\rho_{g,0}\left(1 - b_g\rho_{g,0}\left(\frac{R_0}{R}\right)^3\right)}\right]^{\Gamma_g} - B_g, \tag{10}$$

where $\rho_{l,0}$ is the predefined reference liquid density, $P_{g,0}$ is the predefined reference gas pressure, $\rho_{g,0}$ is the predefined reference gas density, R_0 is the initial bubble radius, Γ_g is the gas polytropic exponent, B_g is the gas pressure constant.

The temperature of the gas and the liquid at the bubble wall can be calculated by Eq. (11):

$$T = T_0\left(\frac{p + B}{P_0 + B}\right)^{\frac{\Gamma - 1}{\Gamma}}, \tag{11}$$

where T_0 and P_0 are the reference temperature and pressure, respectively.

Equations (1)–(11) constitute the model of present study, which is called G-M-N-C model. The bubble radial dynamics, sound velocity, liquid density at the bubble wall, gas pressure, gas temperature and liquid temperature are obtained from Eqs. (1), (4), (7), (10), (11), respectively. The fourth term on the right side of Eq. (5) is introduced by considering the charge on the bubble surface. If this term is ignored, the G-M-N-C model is reduced to the model in (Denner, 2021), which is called the G-M-N model in this paper.

To check the validation of the simulation results, the bubble dynamics obtained by the G-M-N-C model is compared with previous models. A single gas bubble with an initial radius of 3.5 μm oscillating in the water is considered. If not specified, the parameters in present studies are used in Table 1. The models are solved using the method of the Runge-Kutta 4–5 order formula with a variable step length. To obtain the results satisfying the precision requirement, both of the absolute error and relative error are 1×10^{-12}.

Table 1. Parameters keep constant during simulations (Denner, 2021; Hongray et al., 2014; 2015).

Name	Notations	Value	Unit
Gas reference density	$\rho_{g,0}$	1.2	kg/m³
Liquid reference density	$\rho_{l,0}$	998	kg/m³
Gas ambient pressure	$P_{g,0}$	1	atm
Liquid ambient pressure	$P_{l,0}$	1	atm
Gas ambient temperature	$T_{g,0}$	300	K
Liquid ambient temperature	$T_{l,0}$	300	K
Gas molecule co-volume	b_g	0	–
Liquid molecule co-volume	b_l	6.7212×10^{-4}	m³/kg
Gas pressure constant	B_g	0	–
Liquid pressure constant	B_l	6.2178×10^8	Pa
Gas polytropic exponent	Γ_g	1.67	–
Liquid polytropic exponent	Γ_l	1.19	–
Vacuum permittivity	ε_0	8.85×10^{-12}	F/m
Surface tension	σ	7.2×10^{-2}	N/m
Liquid viscosity	μ	1×10^{-3}	Pa·s
Charge on the bubble	Q	0.3	pC

3. Results and discussion

Before showing the dynamics obtained with the G-M-N-C model under various conditions, the other models (i.e., G-M-N in (Denner, 2021); K-M-C in (Hongray et al. 2014; 2015)) are adopted to calculate the bubble dynamics under a high amplitude driving acoustic wave in order to conduct the comparative analysis.

Figure 1 shows the predictions obtained by three models, i.e., G-M-N-C, K-M-C, and G-M-N. Most of the variables associated with the bubble are periodic oscillations, except that the sound velocity, the density and temperature of liquid at the bubble wall remain constant calculated by the model of K-M-C. From Fig. 1a, the bubble grows slowly in the initial stage and then rapidly expands to its maximum radius (R_{max}). Correspondingly, the gas temperature and pressure inside the bubble reach the minimum values (Figs. 1c and 1d). At this time, the bubble begins to collapse due to the difference between the inside and outside of the bubble having a maximum value. When the bubble collapse to the minimum radius (R_{min}), according to Figs. 1b–1g, the variable values (i.e., bubble wall velocity, gas temperature and pressure, the temperature, density and sound velocity of liquid on the bubble wall) reach the maximum. After the first collapse, the bubble oscillates slightly several times.
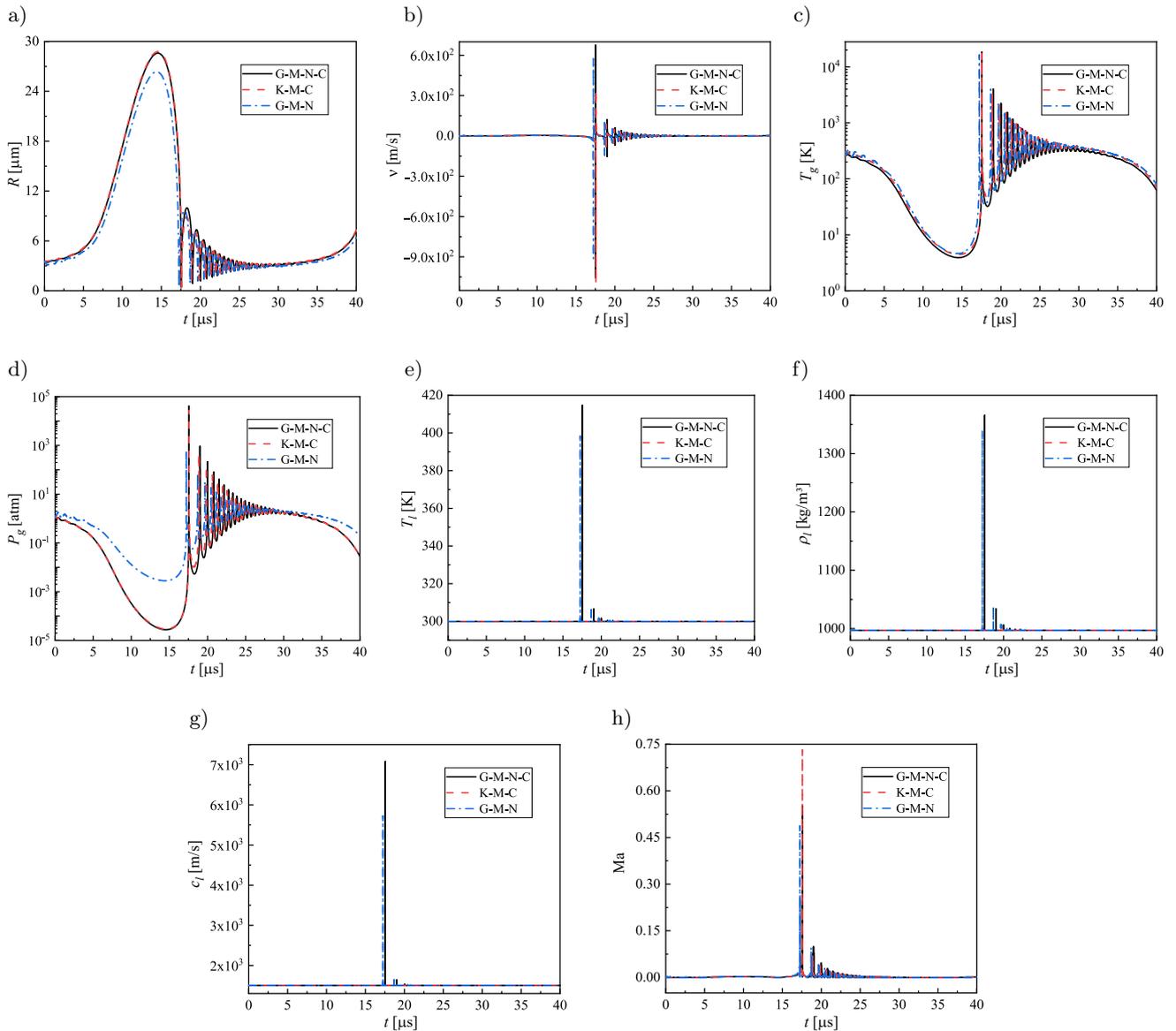
Fig. 1. Time developments of the bubble radius (a), the bubble wall velocity (b), the gas temperature (c) and pressure (d) inside the bubble, the liquid temperature (e), density (f) and sound velocity (g) at the bubble wall, and the Mach number of the bubble wall (h) as retrieved by G-M-N-C model (solid line), K-M-C model (dash line) and G-M-N model (dash dot line). The bubble driven by an ultrasound wave with frequency of 30 kHz and amplitude of 1.35 atm.

Comparing to the predictions obtained by the model of G-M-N, the maximum variable values (i.e., the gas temperature and pressure, bubble wall velocity, liquid temperature, density and sound speed) calculated by the model of G-M-N-C are larger. This owes to the larger absolute value of the maximum pressure difference at the bubble wall in the model of G-M-N-C with the consideration of the charge on the bubble surface. The bubble can absorb more energy during its expansion, reaching a larger size, and a smaller size will be acquired as the bubble collapses.

The dynamics of the charged bubble are commonly predicted by the K-M-C model, which is compared in detail with the G-M-N-C model. As reported in Fig. 1a,

the lower R_{min} is achieved from the G-M-N-C model (0.42 μm, compared to 0.44 μm for the K-M-C model). Since NASG equation of state is used in G-M-N-C model to describe dynamic features of the gas in the bubble and the liquid at the bubble wall, and the compressibility of the gas in the bubble is fully considered. It is believed that the covolumes of gas molecules should be different with temperature and pressure. The high temperature and pressure environment caused by the bubble collapse results in a smaller covolume of gas molecule, meaning that there is more space for the gas to compress. Therefore, the bubble collapse depth is deeper, and the minimum bubble radius is smaller. Consequently, the bubble can be compressed more,

yielding a much higher gas temperature and pressure inside the bubble, as indicated in Figs. 1c and 1d. This result matches the finding in (Nazari-Mahroo *et al.*, 2018; 2020).

The G-M-N-C model gives the maximum liquid temperature and density at the bubble wall reaching 414.67 K and 1.37×10^3 kg/m^3, respectively, whereas these two variables keep constant in the K-M-C model from Figs. 1e and 1f. The calculations of the Mach number (Ma) for two models are shown in Fig. 1h. It can be seen that the peak Ma calculated by the G-M-N-C model is significantly lower than that predicted by the K-M-C model. The liquid velocity of sound at the bubble wall varies with time in the G-M-N-C model (Fig. 1g), at the instant of the first bubble collapse, the bubble wall velocity and the liquid velocity of sound at the wall are both large (Figs. 1b and 1g), reducing Ma (e.g., Ma = 0.55 from Fig. 1h), and the reliability of the model can be guaranteed. While the liquid velocity of sound at the bubble wall remains constant in the K-M-C model leading to a larger Ma during the first bubble collapse (e.g., Ma = 0.73 from Fig. 1h).

The K-M-C model evolves from the Keller-Miksis equation taking the liquid compressibility into account, which is accurate at Ma < 1 (Zilonova *et al.*, 2018). The G-M-N-C model is derived from the Gilmore equation, which is obtained basing on the variation of liquid sound velocity and integrating the liquid enthalpy directly instead of the liquid pressure. Hence, the bubble dynamics with charge estimated by the G-M-N-C model is accurate for Ma ≤ 2.2 (Zilonova *et al.*, 2018). Figure 2 summarizes the maximum absolute value of Ma in P_a (1.2–3 atm) – f (20–400 kHz) plane for two models. Within the parameters studied, the Ma calculated by the K-M-C model range from 0.02 to 10.60, while the values obtained by the G-M-N-C model are 0.02–4.76. Under a certain parameter, Ma calculated by the latter model is smaller than that from the former model. The example can be seen in Fig. 1h. The positions of the blue curves are Ma = 1 in Fig. 2a and 2.2 in Fig. 2b, respectively, which are the critical values of Ma for the K-M-C model and the G-M-N-C model applied in numerical studies. Under the excitation parameters on the upper left of the curves, the application of the model is safe, while the lower right is not. It can be seen that the G-M-N-C model has a wider application range than the K-M-C model.

In application researches, evaluating the cavitation intensity by calculating the gas temperature or the emitted sound pressure, which is a difficult task, because of the need to obtain the bubble wall velocity, or the acceleration. To this end, it is necessary to measure the cavitation intensity from the radial dynamics of the bubble. There exist several approaches in the literatures, e.g., the compression ratio $(R_{\max}/R_{\min})$ (Hongray *et al.*, 2014; Kalmár *et al.*, 2020; Nazari-
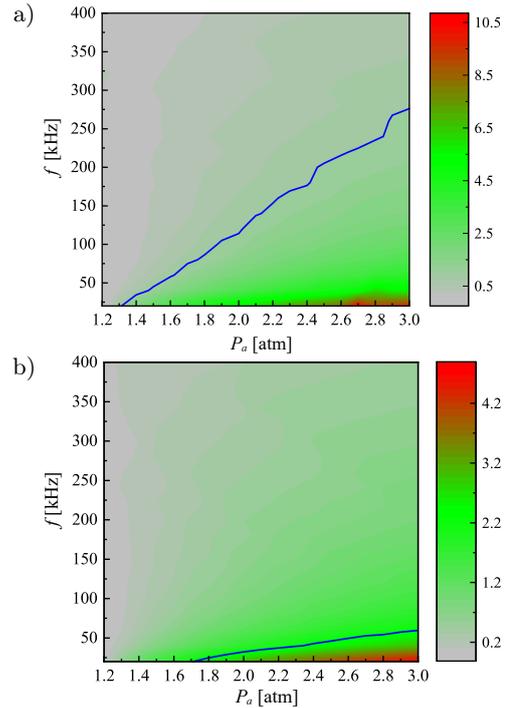
Fig. 2. Maximum absolute value of Mach number of the bubble wall calculated by K-M-C model (a) and G-M-N-C model (b). The numerical results are obtained from more than 200 combinations of ultrasound frequency (20–400 kHz) and acoustic amplitude (1.2–3.0 atm).

Mahroo *et al.*, 2018), the expansion-compression ratio $((R_{\max}-R_0)/(R_0-R_{\min}))$ (Hongray *et al.*, 2015), and the quantity of $\frac{R_{\max}^3}{t_c}$ (Kalmár *et al.*, 2020), where t_c is the bubble collapse time. In the present study, the compression ratio is used to describe the cavitation intensity and is represented by C_s.

The effect of charge Q on the compression ratio C_s is shown in Fig. 3. As can be seen, C_s increases linearly with increasing Q from 0 to 0.6 pC for both models. At the coordinate value of (0.49, 77.36), the two prediction curves intersect. At this time, the minimum bubble radii calculated by two models are similar, and the
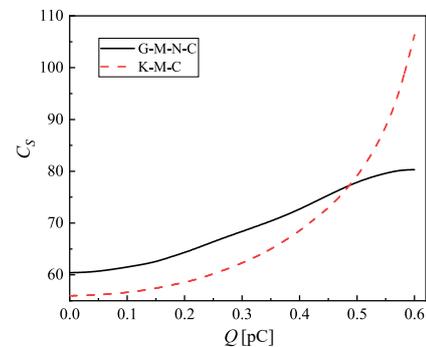
Fig. 3. Compression ratio C_s vs. charge on the bubble, calculated by G-M-N-C model (solid line) and K-M-C model (dash line) for ultrasound frequency of 30 kHz and amplitude of 1.35 atm.

values are about 0.41 μm, which is near the hard-core radius for the bubble with an initial radius of 3.5 μm. Therefore, the variation curves of C_s predicted by two models can only intersect at this point. Also, this result proves the accuracy of the present model. When Q value is away from 0.49 pC, the difference of C_s calculated by two models is increasing. At $Q < 0.49$ pC, the C_s obtained by the G-M-N-C model is larger than that predicted by the K-M-C model. It is worth noting that in the K-M-C model, when Q is greater than 0.53 pC, bubble collapse intensity increases, resulting in Ma > 1 (e.g., at $Q = 0.55$ pC, Ma = 1.03), so that the accuracy of the model is lost. Nevertheless, the range of Q is 0–0.6 pC, and the value of Ma is 0.49–0.61 lying in the application range of the G-M-N-C model.

In order to seek the correlation between the compression ratio C_s and the pressure amplitude P_a, numerical studies have been performed for various P_a (1.2–3.0 atm). The observed results are presented in Fig. 4. With the increase of P_a, C_s calculated by two models increases monotonically. The predicted C_s of the G-M-N-C model is always higher than that of the K-M-C model. The difference of C_s calculated by two models is about 10.02% at P_a from 1.2 to 2.2 atm. When P_a is between 2.2–3.0 atm, the difference increases significantly, and the difference value reaches the maximum at $P_a = 2.4$ atm, which is 51.51%.
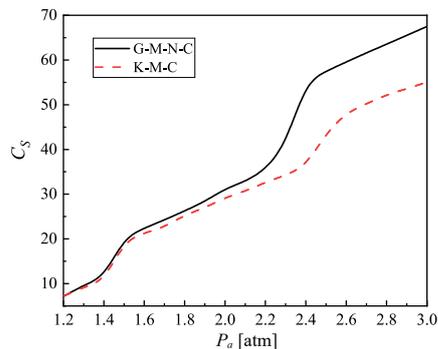
Fig. 4. Compression ratio C_s vs. acoustic amplitude, calculated by G-M-N-C model (solid line) and K-M-C model (dash line) for ultrasound frequency of 30 kHz and charge of 0.3 pC on the bubble.

The effect of ultrasound frequency f on the compression ratio C_s is explored as presented in Fig. 5. It is observed that C_s decreases by increasing f. It is a well-known trend of cavitation bubble dynamics. With the increase of f, there has no enough time to grow for the bubble, resulting in the decrease of C_s. The cavitation intensity is weakened, and the bubble temperature and pressure are reduced. For the two considered models, C_s in the G-M-N-C model is larger than that in the K-M-C model. The difference of C_s calculated by two models reaches the maximum value of 140.51% at the frequency of around 700 kHz.

Other reports in the literatures supported the variation trends of the compression ratio C_s with the pres-
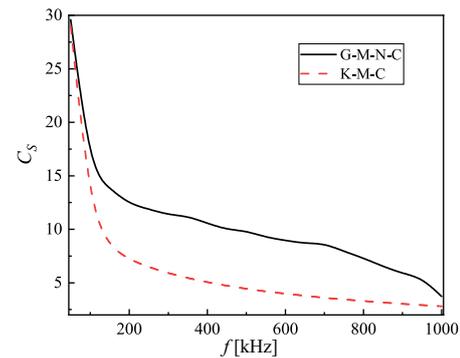
Fig. 5. Compression ratio C_s vs. ultrasound frequency, calculated by G-M-N-C model (solid line) and K-M-C model (dash line) for ultrasound amplitude of 1.35 atm and charge of 0.3 pC on the bubble.

sure amplitude P_a and ultrasound frequency f. FER-KOU *et al.* (2015) examined the effect of ultrasound frequency and acoustic intensity on the sonolytic degradation of naphthol blue black in water operating acoustic intensity in the range of 0.44–5.22 W/cm² with the frequency in the range of 585–1140 kHz. The observed results indicated that the sonochemical degradation rate increased with increasing acoustic intensity and decreasing the frequency. The similar results were obtained for efficiency of sonochemical degradation Bisphenol A (TORRES *et al.*, 2008), the hydrogen production from a collapsing Ar bubble in water (KER-BOUA *et al.*, 2021; DEHANE *et al.*, 2021a) and a collapsing Ar-O₂ bubble in methanol (DEHANE *et al.*, 2022).

Figure 6 shows the variation of the compression ratio C_s as a function of liquid temperature T_∞. The cavitation intensity can be enhanced by the increase of T_∞. As can be seen from the figure, the difference (around 10%) between predicted values by two models is not strongly affected by T_∞ from 283 to 333 K. The similar results obtained by Merouani and co-workers (CHADI *et al.*, 2018; DEHANE *et al.*, 2021b; MEROUANI *et al.*, 2014). They found that cavitation intensity and the production ratio of ·OH generated by a single bub-
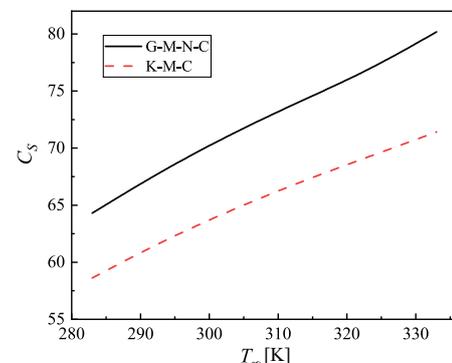

Fig. 6. Compression ratio C_s vs. liquid temperature, calculated by G-M-N-C model (solid line) and K-M-C model (dash line) for ultrasound frequency of 30 kHz, acoustic amplitude of 1.35 atm and charge of 0.3 pC on the bubble.

ble collapse was enhanced with the increase of liquid temperature (Chadi *et al.*, 2018). But the overall production ratio and sonochemical degradation of nonvolatile organic pollutants in aqueous media reached their maximum values at $T_\infty = 50°$. The cavitation effect is mainly affected by the single bubble collapse intensity and the number of bubbles, and the latter decreased with the increase in liquid temperature. Dehane *et al.* (2021b) found the production of RCS, ·OH, ·H, HCl and HOCl increased proportionally with the increase of liquid temperature (from 10° to 50°) on carbon tetrachloride sono-conversion under acoustic excitation with the intensity of 0.7 W/cm² and the ultrasonic frequency of 355 kHz.

Based on the Gilmore-NASG model, the acoustic bubble dynamic model is established in this paper considering the effect of the charge on the bubble surface. The paper analyze the bubble dynamics and cavitation intensity, along with the accuracy of the present model by comparing with the results in the literature. The paper proves the model has a wider range of application parameters. In the future work, present model can be optimized. For example, the effects of bubble-bubble interaction (Zhang *et al.*, 2016) should be taken into account to investigate the secondary Bjerknes force between two gas bubbles, that coupled with a viscous drag force to analyze the translational motions of the bubbles. The detail results are refered to the works in (Wang *et al.*, 2023). In addition, heat exchange, mass transport and chemical reactions on the dynamics of the bubble with a constant charge are also worth considering (Dehane *et al.*, 2021a; 2021b; 2022; Kerboua *et al.*, 2021; Lv, Liu, 2023).

4. Conclusion

In the present work, the model for a charged bubble under acoustic excitation is proposed. Compared with the previous model, the proposed model can be applied to a wider parameter range, especially high acoustic amplitudes and low ultrasound frequencies excitation. The bubble collapse intensity calculated by two models is enhanced with increasing charge on the bubble surface, acoustic amplitude and liquid temperature, and the decrease of ultrasound frequency. Except for the charge on the bubble surface, the bubble collapse intensity by the present model is larger under the studied parameter range, mainly because the effect of liquid and gas compressibility is fully considered. When the charge is 0.49 pC, the minimum bubble radius calculated by two models is near the hard-core radius, and the bubble collapse intensity is equal. The differences between predicted values by two models reaches the maximum at $P_a = 2.4$ atm or $f = 700$ kHz, and the liquid temperature has little effect on the difference (around 10%).

References

1. Chadi N.E., Merouani S., Hamdaoui O., Bouhelassa M. (2018), New aspect of the effect of liquid temperature on sonochemical degradation of nonvolatile organic pollutants in aqueous media, *Separation and Purification Technology*, **200**: 68–74, doi: 10.1016/j.seppur.2018.01.047.

2. Cleve S., Guédra M., Mauger C., Inserra C., Blanc-Benon P. (2019), Microstreaming induced by acoustically trapped, non-spherically oscillating microbubbles, *Journal of Fluid Mechanics*, **875**: 597–621, doi: 10.1017/jfm.2019.511.

3. Denner F. (2021), The Gilmore-NASG model to predict single-bubble cavitation in compressible liquids, *Ultrasonics Sonochemistry*, **70**: 105307, doi: 10.1016/j.ultsonch.2020.105307.

4. Dehane A., Merouani S., Chibani A., Hamdaoui O., Ashokkumar M. (2022), Influence of processing conditions on hydrogen sonoproduction from methanol sono-conversion: A numerical investigation with a validated model, *Chemical Engineering and Processing – Process Intensification*, **179**: 109080, doi: 10.1016/j.cep.2022.109080.

5. Dehane A., Merouani S., Hamdaoui O., Alghyamah A. (2021a), A complete analysis of the effects of transfer phenomenons and reaction heats on sonohydrogen production from reacting bubbles: Impact of ambient bubble size, *International Journal of Hydrogen Energy*, **46**(36): 18767–18779, doi: 10.1016/j.ijhydene.2021.03.069.

6. Dehane A., Merouani S., Hamdaoui O., Abdellattif M.H., Jeon B.-H., Benguerba Y. (2021b), A full mechanistic and kinetics analysis of carbon tetrachloride (CCl₄) sono-conversion: Liquid temperature effect, *Journal of Environmental Chemical Engineering*, **9**(6): 106555, doi: 10.1016/j.jece.2021.106555.

7. Ferkou H., Merouani S., Hamdaoui O., Rezgui Y., Guemini M. (2015), Comprehensive experimental and numerical investigations of the effect of frequency and acoustic intensity on the sonolytic degradation of naphthol blue black in water, *Ultrasonics Sonochemistry*, **26**: 30–39, doi: 10.1016/j.ultsonch.2015.02.004.

8. Grigor'ev A.I., Zharov A.N. (2000), Stability of the equilibrium states of a charged bubble in a dielectric liquid, *Technical Physics*, **45**(4): 389–395, doi: 10.1134/1.1259640.

9. HONGRAY T., ASHOK B., BALAKRISHNAN J. (2014), Effect of charge on the dynamics of an acoustically forced bubble, *Nonlinearity*, **27**(6): 1157–1179, doi: 10.1088/0951-7715/27/6/1157.

10. HONGRAY T., ASHOK B., BALAKRISHNAN J. (2015), Oscillatory dynamics of a charged microbubble under ultrasound, *Pramana – Journal of Physics*, **84**(4): 517–541, doi: 10.1007/s12043-014-0846-y.

11. KALMÁR C., KLAPCSIK K., HEGEDŰS F. (2020), Relationship between the radial dynamics and the chemical production of a harmonically driven spherical bubble, *Ultrasonics Sonochemistry*, **64**: 104989, doi: 10.1016/j.ultsonch.2020.104989.

12. KERBOUA K., HAMDAOUI O., ISLAM M.H., ALGHYAMAH A., HANSEN H.E., POLLET B.G. (2021), Low carbon ultrasonic production of alternate fuel: Operational and mechanistic concerns of the sonochemical process of hydrogen generation under various scenarios, *International Journal of Hydrogen Energy*, **46**(53): 26770–26787, doi: 10.1016/j.ijhydene.2021.05.191.

13. LV L., LIU F. (2023), Numerical investigation of sonochemical production in a single bubble under dual-frequency acoustic excitation, *Physica Scripta*, **98**(11): 115240, doi: 10.1088/1402-4896/acfeb1.

14. LV L., ZHANG Y.X., ZHANG Y.N., ZHANG Y.N. (2019), Experimental investigations of the particle motions induced by a laser-generated cavitation bubble, *Ultrasonics Sonochemistry*, **56**: 63–73, doi: 10.1016/j.ultsonch.2019.03.019.

15. LEE H.-B., CHOI P.-K. (2020), Electrification of sonoluminescing single bubble, *Journal of Physical Chemistry B*, **124**(15): 3145–3151, doi: 10.1021/acs.jpcb.0c00956.

16. LEI Y.-J., ZHANG J., TIAN Y., YAO J., DUAN Q.-S., ZUO W. (2020), Enhanced degradation of total petroleum hydrocarbons in real soil by dual-frequency ultrasound-activated persulfate, *Science of the Total Environment*, **748**: 141414, doi: 10.1016/j.scitotenv.2020.141414.

17. MEROUANI S., HAMDAOUI O., REZGUI Y., GUEMINI M. (2014), Energy analysis during acoustic bubble oscillations: Relationship between bubble energy and sonochemical parameters, *Ultrasonics*, **54**(1): 227–232, doi: 10.1016/j.ultras.2013.04.014.

18. NAZARI-MAHROO K., PASANDIDEH K., NAVID H.A., SADIGHI-BONABI R. (2018), Influence of liquid compressibility on the dynamics of single bubble sonoluminescence, *Physics Letters A*, **382**(30): 1962–1967, doi: 10.1016/j.physleta.2018.04.058.

19. NAZARI-MAHROO K., PASANDIDEH K., NAVID H.A., SADIGHI-BONABI R. (2020), Influence of liquid density variation on the bubble and gas dynamics of a single acoustic cavitation bubble, *Ultrasonics*, **102**: 106034, doi: 10.1016/j.ultras.2019.106034.

20. OH J.M., KIM P.J., KANG I.S. (2001), Chaotic oscillation of a bubble in a weakly viscous dielectric fluid under electric fields, *Physics of Fluids*, **13**(10): 2820–2830, doi: 10.1063/1.1400135.

21. POKHREL N., VABBINA P.K., PALA N. (2016), Sonochemistry: Science and engineering, *Ultrasonics Sonochemistry*, **29**: 104–128, doi: 10.1016/j.ultsonch.2015.07.023.

22. PHUKAN A., KHARPHANBUH S.M., NATH A. (2023), An empirical experimental investigation on the effect of an external electric field on the behaviour of laser-induced cavitation bubbles, *Physical Chemistry Chemical Physics*, **25**: 2477–2485, doi: 10.1039/d2cp05561a.

23. SEZEN S., UZUN D., TURAN O., ATLAR M. (2021), Influence of roughness on propeller performance with a view to mitigating tip vortex cavitation, *Ocean Engineering*, **239**: 109703, doi: 10.1016/j.oceaneng.2021.109703.

24. SHAW S.J., SPELT P.D.M., MATAR O.K. (2009), Electrically induced bubble deformation, translation and collapse, *Journal of Engineering Mathematics*, **65**(4): 291–310, doi: 10.1007/s10665-009-9314-y.

25. SPELT P.D.M., MATAR O.K. (2006), The collapse of a bubble in an electric field, *Physical Review E*, **74**(4): 046309, doi: 10.1103/PhysRevE.74.046309.

26. TAKAHASHI M. (2005), ζ potential of microbubbles in aqueous solutions: Electrical properties of the gas-water interface, *Journal of Physical Chemistry B*, **109**(46): 21858–21864, doi: 10.1021/jp0445270.

27. TIAN L., ZHANG Y.-X., YIN J.-Y., LV L., ZHANG J.-Y., ZHU J.-J. (2023), Study on the liquid jet and shock wave produced by a near-wall cavitation bubble containing a small amount of non-condensable gas, *International Communications in Heat and Mass Transfer*, **145**(Part A): 106815, doi: 10.1016/j.icheatmasstransfer.2023.106815.

28. TORRES R.A., PÉTRIER P., COMBET E., CARRIER M., PULGARIN C. (2008), Ultrasonic cavitation applied to the treatment of bisphenol A. Effect of sonochemical parameters and analysis of BPA by-products, *Ultrasonics Sonochemistry*, **15**(4): 605–611, doi: 10.1016/j.ultsonch.2007.07.003.

29. WANG C., ZHANG Y., LIU Z., YUAN Z. (2022), Multi-objective optimization of a two-stage liquefied natural gas cryogenic submerged pump-turbine in pump mode to reduce flow loss and cavitation, *Journal of Energy Storage*, **52**(Part C): 105064, doi: 10.1016/j.est.2022.105064.

30. WANG X., NING Z., LV M., WU P., SUN C., LIU Y. (2023), Transition mechanisms of translational motions of bubbles in an ultrasonic field, *Ultrasonics Sonochemistry*, **92**: 106271, doi: 10.1016/j.ultsonch.2022.106271.

31. YANG C., DABROS T., LI D., CZARNECKI J., MASLIYAH J.H. (2001), Measurement of the zeta potential of gas bubbles in aqueous solutions by microelectrophoresis method, *Journal of Colloid and Interface Science*, **243**(1): 128–135, doi: 10.1006/jcis.2001.7842.

32. ZHANG Y., ZHANG Y., LI S. (2016), The secondary Bjerknes force between two under dual-frequency acoustic excitation, *Ultrasonics Sonochemistry*, **29**: 129–145, doi: 10.1016/j.ultsonch.2015.08.022.

33. ZILONOVA E., SOLOVCHUK M., SHEU T.W.H. (2018), Bubble dynamics in viscoelastic soft tissue in high-intensity focal ultrasound thermal therapy, *Ultrasonics Sonochemistry*, **40**(Part A): 900–911, doi: 10.1016/j.ultsonch.2017.08.017.

Research Paper

Effect of Acoustic Enclosure on the Sound Transmission Loss of Multi-Layered Micro-Perforated Plates

Brahim EL KHARRAS*[iD], Mohammed GAROUM[iD], Abdelmajid BYBI[iD]

Higher School of Technology in Salé, Material, Energy and Acoustics Team (MEAT)
Mohammed V University in Rabat
Salé, Médina, Morocco

*Corresponding Author e-mail: brahim_elkharras@um5.ac.ma

This study presents an examination of the transmission properties of multilayered partitions made up of multiple micro-perforated plates (MPPs) coupled to acoustic enclosures with general impedance boundaries. Multi-layered MPPs can lower the transmission while minimizing reflection in the source and receiving enclosure. Previous research has mainly focused on the double MPPs or triple MPPs partition itself. However, it is vital to analyze the in-situ sound transmission loss of the multi-layered MPP and their efficiency in a complex vibro-acoustic environment. The case when the multilayered MPPs are coupled to a receiving enclosure or coupled to both a source and receiving enclosure is investigated. The objective is to provide an analytical method to evaluate the transmission properties of multilayered MPPs coupled to acoustic enclosures while being computationally more efficient than the finite element method (FEM). Using the modified Fourier series for the acoustic pressure, a variational form for the acoustic and structure medium yields a completely coupled vibroacoustic system. A comparison between the sound transmission loss of the double MPPs, when mounted on an impedance tube and coupled to acoustics enclosures, shows the modal effect of the enclosures. The effect of enclosure shape, impedance boundary, perforation ratio, air gap thickness on the sound transmission properties of the double MPPs structure is examined for both cases. Finally, in both situations, the performance of triple MPP structure insulation is evaluated.

Keywords: micro-perforated plate (MPP); sound transmission loss; noise insulation; coupled structural-acoustic; surface impedance; modal analysis.

1. Introduction

In recent decades, micro-perforated panels (MPP) have become more popular as a means of noise reduction. In the initial design, the arrangement involved placing perforated panels in front of a solid wall. The purpose was to increase particle velocity through the perforations and dissipate acoustic energy, as documented by MAA (1998). Remarkably, without the need for an additional layer of porous materials, the sub-millimeter-sized holes were capable of providing relatively broad-spectrum absorption. This phenomenon results from the shear forces generated by air vibrations as they pass through these tiny openings. As a result, this design approach facilitates the construction of sound-absorbing walls that are both

lightweight and free of fibers, all while maintaining high functionality. MPPs may be utilized as acoustic liners at the intake and exhaust of aircraft nacelles to minimize fan and jet engine noise because of their resistance to degradation and their ability to survive harsh settings where porous materials can degrade. These materials are great for use as outdoor noise barriers, and may even be made out of transparent materials. Therefore, they might be particularly helpful to architects who utilize interior or exterior glass structures (FUCHS, ZHA, 1997), for example, to reduce the long reverberation durations that contribute to poor intelligibility in glass-enclosed spaces. And since environmental considerations are taken into account while creating standards, it is anticipated that in the future, these lightweight, non-polluting, soundproof alterna-

tives, which may be made from recyclable materials, would replace the use of porous barriers.

Micro-perforated panels have also been used in a vast array of other technical applications. Micro-perforated cylinder silencers were the subject of analytical and experimental study by Bravo *et al.* (2016) in the high-sound pressure and low-frequency domains. They examined the nonlinear behavior of a micro-perforated cylinder liner as it dissipated energy and transmitted sound at high pressure. Alisah *et al.* (2021) studied the potential of an expansion chamber coupled micro-perforated cylindrical panel using the boundary element method to enhance the acoustic attenuation for in-duct noise control issues. Yu *et al.* (2015) studied the effectiveness of hybrid silencers made from MPPs and inner partitions, finding that the balance between dissipative and reactive noise attenuation effects could be modified by varying the ideal hole size and perforation ratio, among other aspects. Allam and Åbom (2011) included MPPs in the design of vehicle exhaust mufflers, indicating that mufflers made using MPPs rather than porous materials provide the advantage of a non-fibrous, lighter alternative. Micro-perforated insertion units were developed by Pfretzschner *et al.* (2006) and may be used to protect the acoustic properties of MPPs while reinforcing their fragile thin plates or foils. To increase the frequency range, Liu *et al.* (2017) investigated layered sound absorbers on which a 3D-printed MPP was mounted using a porous sound absorbent material. Yang and Cheng (2016) conducted research on the sound absorption properties of MPPs when supported by either an air volume or a honeycomb structure within small enclosures. Their study revealed that the interaction between the backing cavity and the enclosure had a notable influence on the MPP's ability to absorb sound. Kang and Brocklesby (2005) investigated the viability of utilizing a window arrangement with transparent micro-perforated absorbers and proved that noise could be decreased while enabling much greater airflow compared with traditional window systems. An optimization of the MPP with a multi-depth cavity was carried out by Falsafi and Ohadi (2018) to extend the absorption bandwidth. MPPs have been utilized in honeycomb and corrugated constructions as low-frequency sound-absorbing structures (Meng *et al.*, 2017; 2019; Tang *et al.*, 2017; 2019).

Although sound absorption has been the primary focus of most investigations on MPPs, just a few have studied their sound-isolating properties. Toyoda and Takahashi (2008) examined the acoustic properties of an infinite MPP structure with a back wall. To reduce mid-frequency transmission loss, they performed an analysis of the problem in only two dimensions and proposed an air gap subdivision, namely the use of a structure based on honeycomb. Micro-perforation enhances soundproofing effectiveness at the mass-spring resonance in the two studies of the soundproofing properties of infinite double and triple windows and panels with micro-perforations (Mu *et al.*, 2011a; 2011b). Sound transmission loss (STL) results of single and double-layered construction with and without micro-perforations were compared by Dupont *et al.* (2003) in their study of lightweight MPP systems. Bravo *et al.* (2012) developed a fully coupled modal approach that predicts the absorption and transmission characteristics of flexible MPP-cavity-plate partitions. The study was expanded (Bravo *et al.*, 2014) and a comparison was conducted between MPP-MPP-plate and MPP-porous-plate partitions. Kim *et al.* (2020a) conducted a comprehensive analysis using the transfer matrix technique to study the impact of micro-perforations on the sound transmission loss (STL) of multi-layered infinite MPPs across the entire frequency spectrum. They employed the concept of equivalent impedance, which combines the impedances associated with both the inertia term and the micro-perforations, as a means to characterize the effects of these perforations. In a related study, Kim *et al.* (2020b) employed the transfer matrix approach while assuming conditions of plane waves in the low-frequency domain. Their investigation centered on analyzing the STL of multi-layered flexible MPPs, which were positioned within a rectangular cross-sectioned impedance tube. Their particular focus was on understanding the behavior of these panels at resonance frequencies.

Analytical and numerical techniques have been utilized to characterize the insulating characteristics of the dividing partition and the parameters that lead to observed differences in the low-frequency range of these qualities to fully explain how finite cavities affect the findings. Kihlman (1967) examined inaccuracies in sound transmission measurement using modal analysis. He noticed that systematic differences may develop in the low-frequency spectrum if the source and receiving rooms have identical dimensions. Other authors (Mulholland, Lyon, 1973; Gagliardini *et al.*, 1991; Kropp *et al.*, 1994; Osipov *et al.*, 1997a; 1997b) have adopted comparable strategies to investigate the impact on the measured low-frequency sound transmission loss such as room size, source location, reverberation duration, interface geometry, or mounting conditions. Bravo and Elliott (2004) introduced a coupled model to investigate systems comprising cavity-panel-cavity and cavity-panel configurations, focusing on the evaluation of low-frequency sound transmission loss. Cheng *et al.* (2005) studied the transmission of energy within double-wall partitions that were mechanically coupled to an acoustic cavity. They specifically examined how the noise insulation characteristics of these structures were affected by factors such as the air gap and mechanical connections, utilizing a fully coupled vibro-acoustic formulation.

We have characterized the double and triple MPPs configurations, as they are of relevance when considering the reduction of the noise transmitted through building walls while providing boundary absorption in the source room and receiving room, lowering their reverberation times and improving the speech intelligibility. In a typical application, multi-layered MPPs might be utilized to reduce transmissions and reflections from both sides compared to the double plate or the plate-cavity-MPP partitions. Multi-layered MPPs behavior must be evaluated in such applications by considering it as a component of the complete system. There are two significant limitations in existing research: (a) the sound transmission of multi-layered MPPs were assessed in simple acoustic environment such as impedance tube or free-field conditions, and (b) the effect of enclosures and boundary conditions on transmission loss the double and triple MPPs structures, when coupled to a receiving enclosure or both the source and receiving enclosure, was not considered. Since these characteristics can greatly alter the transmission properties of multi-layered MPPs, comprehensive knowledge of vibro-acoustic phenomena and their sensitivity to changes in system parameters is of paramount importance. Additionally, the development of a versatile and precise methodology capable of addressing various aspects of systems would greatly benefit both academic researchers and engineers. While numerical approaches such as the finite element method (FEM) are effective for obtaining relevant acoustic data, there is a preference for analytical techniques due to their computational efficiency. Analytical methods are particularly suited for parametric investigations, sensitivity analyses, and optimization. They also serve as reliable benchmarks for the development of acoustic engineering software.

To address the limitations observed in previous models and offer a clearer understanding of the transmission properties of double MPP coupled to acoustic enclosures and their sensitivity to changes in various system parameters, we propose a vibro-acoustic formulation for modeling the behavior of the coupled system. In this proposed approach, we use the modified Fourier series to express the acoustic pressure within the enclosures. Our model assumes a clamped boundary condition, which, while more complex to handle analytically than the simply supported boundary condition, better reflects real-world engineering scenarios in many cases. To obtain solutions for displacement and acoustic pressure, we apply a modified variational principle for the coupled system. Our approach efficiency and accuracy is validated by comparing it with the FEM and impedance tube results. We investigate the impact of key factors on the sound transmission loss of a multi-layered MPP structure when coupled to a receiving acoustic enclosure or both the receiving and source enclosure. These factors include the dimensions of the acoustic enclosure, air gap thickness, acoustic enclosure impedance, plate micro-perforation, and the presence of absorbent materials within the gap. Finally, the case of triple MPPs is also investigated.

2. Theoretical formulations

2.1. Description of the coupled system

Figure 1 illustrates the coupled system under investigation in the present study. This system is a double MPP that is separated by an air gap and coupled to an enclosure cavity. The double MPP structure is composed of two homogeneous and isotropic rectangular micro-perforated plates. The two MPPs are identical. The clamped boundary condition for the MPP is assumed. A uniform, oblique plane sound wave with an incidence elevation angle ϕ and an azimuth angle θ is applied to the top MPP, designated 1, whilst the lower MPP denoted 2, is coupled to a receiving enclosure in the first case. To study the vibro-acoustic behavior of double MPP when coupled to both a source and receiving enclosure, an acoustic point source is placed in a secondary enclosure and coupled to the double MPPs structure in the second case. The air gap and the acoustic enclosures have acoustically rigid walls except for the regions occupied by the two MPPs. Nonetheless, it is worth noting that we can readily introduce an arbitrary impedance surface condition when necessary.
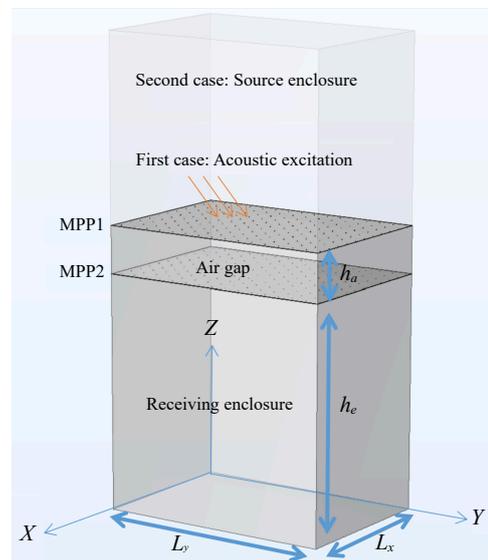

Fig. 1. Schematic of a double micro-perforated plate structure coupled to an enclosure.

The acoustic pressure in the acoustic gap should satisfy both the wave equation and the corresponding boundary conditions.

The wave equation is given as:

$$\nabla^2 p_g + k^2 p_g = 0, \tag{1}$$

where p_g is the acoustic pressure within the acoustic gap, $k = \frac{\omega}{c_0}$ is the wavenumber, and $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$.

The boundary conditions on the air gap boundaries:

$$\frac{\partial p_g}{\partial z} = j\omega\rho_0\overline{v}_1, \qquad \text{on MPP1}, \qquad (1)_1$$

$$\frac{\partial p_g}{\partial z} = -j\omega\rho_0\overline{v}_2, \qquad \text{on MPP2}, \qquad (1)_2$$

$$\frac{\partial p_g}{\partial n} = 0, \qquad Z(\omega) = \infty, \qquad \text{on the rigid wall}, \quad (1)_3$$

where ρ_0 and c_0 are the density of air and the speed of sound, respectively, and $\overline{v}_j$ is the average velocity of the j-th MPP.

Similarly, the acoustic pressure within the enclosure satisfies the wave equation including the continuity conditions on the boundaries:

$$\nabla^2 p_e + k^2 p_e = 0,$$

$$\frac{\partial p_e}{\partial z} = j\omega\rho_0\overline{v}_2, \qquad \text{on MPP2}, \qquad (2)$$

$$\frac{\partial p_e}{\partial z} = -j\omega\rho_0\frac{p_e}{Z}, \qquad \text{on the walls},$$

where p_e is the acoustic pressure inside the enclosure.

As can be seen in Fig. 2, the average velocity $\overline{v}$ of the MPP is related directly to the velocity of the plate v_p as well as the velocity of the fluid v_f as it passes through the hole:

$$\overline{v} = v_p(1 - \sigma) + v_f\sigma, \qquad (3)$$

where σ denotes the perforation ratio, and it is defined as $\sigma = \pi d^2/4U^2$, where d denotes the hole's diameter and U represents the distance between the holes of the MPP.
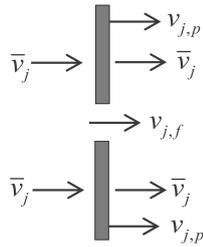

Fig. 2. Average velocity on the surface of the j-th MPP.

As was shown in prior research (Takahashi, Tanaka, 2002), the impedance of the MPP is related to the pressure difference:

$$Z_{\text{resist}}(v_f - v_p) + Z_{\text{react}}v_f = \Delta p. \qquad (4)$$

According to Maa (1998), the impedance of the hole, $Z = Z_{\text{resist}} + Z_{\text{react}}$, which consist of a resistive part and an imaginary part. The former corresponds to the resistive force between the fluid and the inner surface of the hole and the latter deals with the inertia force (motion) of the fluid in the hole. They are given as:

$$Z_{\text{resist}} = \frac{8\eta_0 h}{(d/2)^2}\left(\sqrt{1 + \frac{X^2}{32}} + \frac{\sqrt{2}\,dX}{32h}\right), \qquad (5)$$

$$Z_{\text{react}} = j\rho\omega h\left(1 + \frac{1}{\sqrt{9 + \frac{X^2}{2}}} + \frac{8d}{3\pi h}\right), \qquad (6)$$

where η_0 represents the air's viscosity coefficient, h denotes the thickness of the MPP, and $X = \left(\frac{d}{2}\right)\sqrt{\frac{\rho\omega}{\eta}}$.

When we eliminate $v_{j,f}$ from Eqs. (3) and (4), we get the relation:

$$\overline{v} = \gamma v_p + \frac{\sigma\Delta p}{Z}, \qquad (7)$$

in which

$$\gamma = 1 - \sigma\left(\frac{Z_{\text{react}}}{Z}\right). \qquad (8)$$

For the first MPP, $\Delta p_1 = p_0 - p_g$.

The governing equation of the first MPP is given as a function of MPP displacement w_1 (and $v_{1,p} = j\omega w_1$):

$$D_1\nabla^4 w_1 - M_1\omega^2 w_1 = p_0 - p_g, \qquad (9)$$

where M_1 is the mass per unit area and D_1 is the flexural rigidity of the first MPP, it can be written as:

$$D_1 = E_1 h_1^3/12(1 - v_1^2), \qquad (10)$$

where E_1 represent Young's modulus, v_1 is Poisson's ratio, and h_1 is the thickness.

In Eq. (10), damping of the structure is taken into consideration by inserting complex Young's modulus $E(1 + j\eta)$, where η is the loss factor.

The sound pressure acting on the incident panel consists of three distinct pressures: the incident pressure, the reflected pressure, and the radiated pressure. Among these three components, the radiated pressure is notably negligible. This is primarily due to the incident plate's impedance is comparable to that of a rigid boundary when subjected to air loading. Consequently, it is reasonable to assume that the magnitudes of the incident and reflected pressure waves are equal (Carneal, Fuller, 2004; Chazot, Guyader, 2007).

Hence, in cases involving light fluid-loading and sub-millimetric holes with substantial resistive effects, neglecting the radiated pressure directed outward and assuming equivalence in magnitude between the incident and reflected pressures, we can conclude that the blocked pressure, often referred to as the excitation pressure, is twice the magnitude of the incident wave:

$$p_0(x, y, t) = 2p_{\text{inc}}\exp\left(j\omega t - jkz\cos(\phi)\right.$$

$$\left. - jky\sin(\phi)\sin(\theta) - jkx\sin(\phi)\cos(\theta)\right). \quad (11)$$

In the same manner, the governing equation of the second MPP is given as a function of the MPP displacement w_2 (and $v_{2,p} = j\omega w_2$):

$$D_2 \nabla^4 w_2 - M_2 \omega^2 w_2 = p_g - p_e, \qquad (12)$$

where M_2 is the mass per unit area and D_2 is the flexural rigidity of the second MPP.

2.2. Solution procedure of the coupled system

Under the assumption that the double MPP are fully clamped, the transverse deflection and moment rotation of each panel are constrained to remain zero along their edges. Consequently, their transverse displacements can be expressed as:

$$w_{j,p}(x,y) = \sum_{m,n} q_{j,mn} \varphi_{mn}(x,y), \qquad (13)$$

where the modal functions φ_{mn} or, more strictly speaking, the basic functions take the following forms (XIN *et al.*, 2008):

$$\varphi_{mn}(x,y) = \left(1 - \cos\left(\frac{2m\pi x}{L_x}\right)\right)\left(1 - \cos\left(\frac{2n\pi y}{L_y}\right)\right). \quad (14)$$

Note that the clamped modal function of Eq. (13) is different from the simply supported modal function $\varphi_{mn}(x,y) = \sin\left(\frac{2m\pi x}{L_x}\right)\sin\left(\frac{2n\pi y}{L_y}\right)$.

To ensure differential continuity at the structural-acoustic coupling interface, the sound field inside the enclosure and the air gap may be represented (DU *et al.*, 2011):

$$\begin{aligned}
p_g(x,y,z) &= \sum_{m_x=0}^{M_x} \sum_{m_y=0}^{M_y} \sum_{m_z=0}^{M_z} A^g_{m_x m_y m_z} \\
&\quad \cdot \cos(\lambda_{m_x} x) \cos(\lambda_{m_y} y) \cos(\lambda_{m_z} z) \\
&\quad + \sum_{m_x=0}^{M_x} \sum_{m_y=0}^{M_y} (\xi_{1L_z}(z) a^g_{m_x m_y} + \xi_{2L_z}(z) b^g_{m_x m_y}) \\
&\quad \cdot \cos(\lambda_{m_x} x) \cos(\lambda_{m_y} y), \qquad (15)
\end{aligned}$$

$$\begin{aligned}
p_e(x,y,z) &= \sum_{m_x=0}^{M_x} \sum_{m_y=0}^{M_y} \sum_{m_z=0}^{M_z} A^e_{m_x m_y m_z} \\
&\quad \cdot \cos(\lambda_{m_x} x) \cos(\lambda_{m_y} y) \cos(\lambda_{m_z} z) \\
&\quad + \sum_{m_x=0}^{M_x} \sum_{m_y=0}^{M_y} \xi_{2L_z}(z) b^e_{m_x m_y} \\
&\quad \cdot \cos(\lambda_{m_x} x) \cos(\lambda_{m_y} y), \qquad (16)
\end{aligned}$$

where $\lambda_{m_s} = \frac{m_s \pi}{L_s}$, $(s = x, y, z)$. The supplemental functions $\xi_{1L_s}(s)$ and $\xi_{2L_s}(s)$ can be found in (DU *et al.*, 2011).

Theoretically, an acoustic analysis of a cavity can be formulated using a variational approach. This approach can yield a solution that is more advantageous compared to simply solving the Helmholtz equation. To achieve this objective, a modified variational approach (CHIEN, 1983; QU *et al.*, 2013a; 2013b) is employed to define the distribution of sound pressure. This technique involves seeking the minimum value of the corresponding modified variational function:

$$\iiint_{V_g} \frac{1}{2}\left[p_g \frac{jk}{\rho_0 c_0} p_g - (\nabla p_g)^{\mathrm{T}} \frac{j}{\rho_0 \omega}(\nabla p_g)\right] \mathrm{d}V$$

$$+ \iint_{S_{\mathrm{MPP1}}} p_g \overline{v}_1 \,\mathrm{d}S - \iint_{S_{\mathrm{MPP2}}} p_g \overline{v}_2 \,\mathrm{d}S = 0. \quad (17)$$

Using Eq. (17) we get:

$$\iiint_{V_g} \frac{1}{2}\left[p_g \frac{jk}{\rho_0 c_0} p_g - (\nabla p_g)^{\mathrm{T}} \frac{j}{\rho_0 \omega}(\nabla p_g)\right] \mathrm{d}V$$

$$+ \iint_{S_{\mathrm{MPP1}}} p_g \left(\gamma_1 v_{1,p} + \frac{\sigma_1 \Delta p_1}{Z_1}\right) \mathrm{d}S$$

$$- \iint_{S_{\mathrm{MPP2}}} p_g \left(\gamma_2 v_{2,p} + \frac{\sigma_2 \Delta p_2}{Z_2}\right) \mathrm{d}S = 0, \quad (18)$$

where $\Delta p_1 = p_0 - p_g$ and $\Delta p_2 = p_g - p_e$.

The characteristic equation may be found by inserting the admissible function specified in Eqs. (13), (15), and (16) into Eq. (18), and then carrying out the variational operation in terms of the generalized coordinate vector.

The resultant equation for the air gap is as:

$$[K_g + j\omega Z_g + \omega^2 M_g]P_g + \omega^2 C_{p_g,w_1} W_{\mathrm{MPP1}}$$

$$- \omega^2 C_{p_g,w_2} W_{\mathrm{MPP2}} + j\omega C_{p_g,p_e} P_g = j\omega\{P_{0,w_1}\}. \quad (19)$$

In the same manner, for the acoustic enclosure, we get:

$$\left[K_e + j\omega Z_e + \omega^2 M_e\right]P_e + \omega^2 C_{p_e,w_2} W_{\mathrm{MPP2}}$$

$$+ j\omega C_{p_e,p_g} P_g = \{0_e\}. \quad (20)$$

By setting the integral of a weighted residual of the modal function to zero, an arbitrarily accurate double series solution can be obtained. For the current double MPPs partition system, the integral equations are:

$$\iint_{S_{\mathrm{MPP_1}}} (D_1 \nabla^4 w_1 - M_1 \omega^2 w_1)\varphi_{mn}(x,y) \,\mathrm{d}x\,\mathrm{d}y$$

$$= \iint_{S_{\mathrm{MPP_1}}} p_0 \varphi_{mn}(x,y) \,\mathrm{d}x\,\mathrm{d}y$$

$$- \iint_{S_{\mathrm{MPP_1}}} p_g \varphi_{mn}(x,y) \,\mathrm{d}x\,\mathrm{d}y, \qquad (21)$$

$$\iint\limits_{S_{\mathrm{MPP}_2}} (D_2 \nabla^4 w_2 - M_2 \omega^2 w_2) \varphi_{mn}(x,y) \,\mathrm{d}x\,\mathrm{d}y$$

$$= \iint\limits_{S_{\mathrm{MPP}_2}} p_g \varphi_{mn}(x,y) \,\mathrm{d}x\,\mathrm{d}y$$

$$- \iint\limits_{S_{\mathrm{MPP}_2}} p_e \varphi_{mn}(x,y) \,\mathrm{d}x\,\mathrm{d}y. \tag{22}$$

By substituting Eqs. (13), (15), and (16) into Eqs. (21) and (22) and subsequently engaging in meticulous yet straightforward algebraic manipulations, the matrix equations for the MPPs are derived:

$$[K_{p_1} + \omega^2 M_{p1}] W_{\mathrm{MPP1}}$$
$$+ C_{p_g, w_1}^{\mathrm{T}} P_g = \{ P'_{0, w_1} \}, \tag{23}$$

$$[K_{p_2} + \omega^2 M_{p2}] W_{\mathrm{MPP2}} - C_{p_g, w_2}^{\mathrm{T}} P_g$$
$$+ C_{p_e, w_2}^{\mathrm{T}} P_e = \{0\}. \tag{24}$$

The enclosure and the air gap Eqs. (19) and (20) and the MPPs Eqs. (23) and (24) form a set $(N_e + N_g + M_{\mathrm{MPP1}} + M_{\mathrm{MPP2}})$ of coupled equations that may be put in matrix form as:

$$[\mathbf{K} + j\omega C + \omega^2 \mathbf{M}]\mathbf{X} = \mathbf{Q}, \tag{25}$$

where $\mathbf{M}$ and $\mathbf{K}$ are diagonal matrices representing the mass and stiffness of the whole system, respectively, and $\mathbf{X}$ is the vector of the unknown complex modal amplitudes. The MPP and cavity modal components that couple with the external pressure are correspondingly represented by the generalized excitation vector $\mathbf{Q}$.

To study the effect of a source room on the transmission properties of multi-layered MPPs. The previous theoretical formulations could also be extended to include a secondary acoustic enclosure as a source room.

The acoustic pressure in the emitting enclosure satisfies the wave equation including the continuity conditions on the boundaries:

$$\nabla^2 p_s + k^2 p_s = j\omega \rho_0 Q \delta(x - x_0)(y - y_0)(z - z_0),$$
$$\frac{\partial p_s}{\partial z} = j\omega \rho_0 \overline{v}_1, \tag{26}$$

where p_s is the acoustic pressure inside the enclosure, Q is the source strength, and (x_0, y_0, z_0) is the position of the acoustic source.

Similarly making use of the modified Fourier series for the acoustic pressure and then carrying out the variational operation in terms of the generalized coordinate vector. A new set $(N_s + N_e + N_g + M_{\mathrm{MPP1}} + M_{\mathrm{MPP2}})$ of coupled equations can be obtained.

2.3. Sound transmission loss

The transmission loss or sound reduction index of a double MPP connected to a receiving enclosure is defined by:

$$\mathrm{TL} = 10 \log_{10} \left(\frac{\Pi_{\mathrm{inc}}}{\Pi_{\mathrm{rad}}} \right) \text{ [dB]}, \tag{27}$$

where Π_{inc} and Π_{rad} are the sound power incident and radiated by the double MPP structure, respectively, at a given frequency.

The incident acoustic power can be defined as:

$$\Pi_{\mathrm{inc}} = \frac{1}{2} \mathrm{Re} \iint\limits_A p_i . v_i^* \,\mathrm{d}A. \tag{28}$$

The conjugate of a complex variable is denoted by $*$, v_i is the acoustic velocity, and p_i the incident sound pressure. When the incident wave is plane and the acoustic medium is air, the incident power is (Chazot, Guyader, 2007; Xin *et al.*, 2008):

$$\Pi_{\mathrm{inc}} = \frac{p_i^2 \cos(\varphi) . S}{2 \rho_0 c_0}. \tag{29}$$

The potential energy in the receiving enclosure is used to compute the partition's radiated power.

In the second case where the multilayered MPPs are connected to both a receiving and the source room, we define the sound transmission loss as (Løvholt *et al.*, 2017):

$$\mathrm{STL} = 20 \log_{10} \left(\frac{|p_{\mathrm{re}}|}{|p_{\mathrm{so}}|} \right), \tag{30}$$

where $|p_{\mathrm{re}}|$ and $|p_{\mathrm{so}}|$ are the absolute values of the pressure averaged over the receiving room and the source room, respectively.

3. Numerical model

The numerical model is constructed using COMSOL Multiphysics. In this model, a plane wave is applied to the incident section. The air in the incident field, receiving enclosure, and air gap is considered to be compressible, with no consideration for thermal conductivity or viscosity. Therefore, the pressure acoustics module within COMSOL, which is suitable for all frequency-domain simulations, is employed.

During the simulation, the MPP are represented as isotropic linear elastic materials using COMSOL's Solid Mechanics module. It is important to note that the simulation accounts for the thermal conduction and viscosity of the air within the small pores. To handle this, the thermal-acoustics module is utilized.

In the case where a source enclosure is coupled to the multilayered MPPs, the incident field is replaced by an enclosure and the pressure acoustics module is employed.

3.1. Boundary conditions

In the finite element (FE) model, it is assumed that at the interface between the pressure acoustic

field and the solid panel, the normal accelerations of the air and the panel are equal. Furthermore, the FE model enforces continuous normal stress, acceleration, and adiabatic conditions at the interface between the thermal acoustic field and the pressure acoustic field. Additionally, the air velocity at the interface between the thermal acoustic field and the solid panel coupling boundary is set to be equal to the velocity of the panel, and temperature variation is treated as isothermal.

The FE model also takes into account the boundary conditions of the MPP, the acoustic enclosure, and the air gap. Tetrahedral elements are used to mesh the FE model, as depicted in Fig. 3. It is important to note that the dimensions of each part have an impact on the size of the elements used in the model.
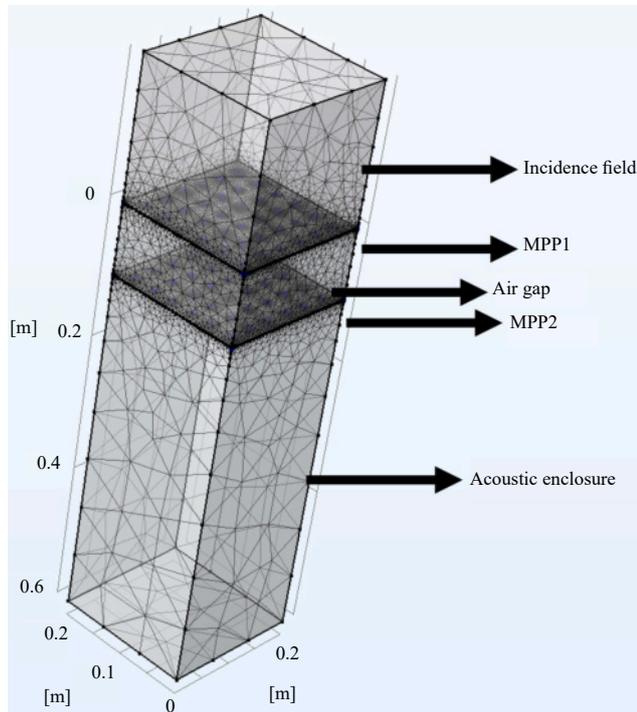

Fig. 3. FE model of a double MPP coupled to an acoustic enclosure.

4. Numerical results and discussions

In this section, the model is validated and a parametric study is conducted by the theoretical modeling and solution approach outlined in the preceding sections of the paper. Comparison between the STL of the double MPP structure, in both cases, obtained using the current approach and FEM is carried out first to assess the limit of applicability and verify its reliability and accuracy. Then, parametric research on the effect of the coupled system parameters on the sound transmission loss is also carried out, including acoustic enclosure dimension and impedance, micro-perforation of the MPPs, air gap thickness, and the influence of the absorbent material introduced in the gap.

4.1. Validation of the analytical formulation

The numerical precision and accuracy of the analytical formulation are assessed on a double MPP connected to a receiving enclosure and both a source and receiving enclosure.

In this study, we consider MPPs made of aluminum plates with the following material properties: Young's modulus $E = 7.2 \times 10^{10}$ Pa, Poisson's ratio $v = 0.34$, density $\rho_p = 2700$ kg/m^3, and the loss factor is assumed to be $\eta = 0.01$. The sidewalls around the air gap cavity and the enclosure are perfectly rigid. The properties of air are $\rho_0 = 1.2$ kg/m^3 and $c_0 = 343$ m/s, and the viscosity is $\eta_0 = 17.9 \times 10^{-6}$ kg/m·s. The flexible MPPs have a dimension of 0.2 m × 0.2 m, and 1 mm thick. The depth of the air gap (h_a) and the receiving enclosure (h_e) considered are 0.1 and 0.5 m, respectively. In the case when the double MPPs structure is connected to both a receiving and source enclosure, a point source of strength $Q_0 = 10^{-4}$ m^3/s placed at $(0.15, 0.12, 0.4)$ in the source enclosure. The dimensions of the source enclosure are: $L_x \times L_y \times h_{e2} = 0.2$ m × 0.2 m × 0.6 m.

The clamped boundary conditions for both the flexible MPPs are considered in the present solution technique. The MPP parameters are $\sigma_1 = \sigma_2 = 0.1\%$ and $d_1 = d_2 = 0.8$ mm.

The primary factor influencing the accuracy of the solution is the number of modes employed for decomposing both displacement and acoustic pressure. To ensure the calculation accuracy, the number of modal truncations (N) and (M) is checked for both the structural and acoustic domains based on the FE calculation results. Typically, precision can be enhanced by increasing the number of modes until convergence is obtained within the desired frequency range. For the dimensions of the coupled system, when $N_1 = N_2 = 7$ and $M_x = M_y = M_z = 4$, a satisfactory level of accuracy is obtained.

Figure 4 illustrates the sound transmission loss results of a double MPP structure under sound excita-
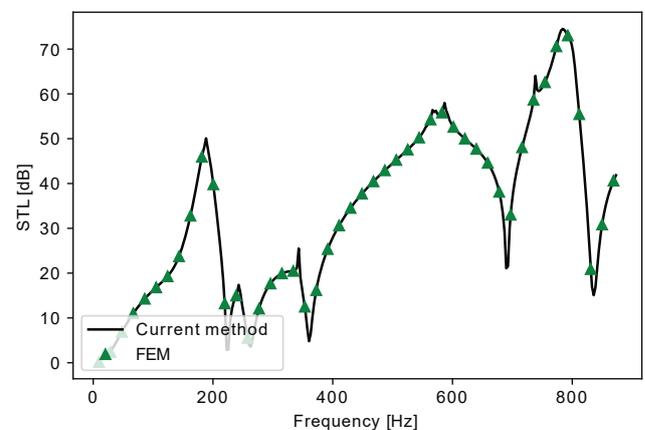
Fig. 4. Comparison of calculated STLs using the present approach and FEM results for double MPP coupled to a receiving enclosure.

tion using the current approach and FEM simulation. As depicted, the STL predictions obtained through the present technique align closely with the results obtained through FEM simulation, demonstrating good agreement between the two methods.

Figure 5 illustrates the STL results of a double MPP structure when connected to a receiving and source enclosure was examined using the current approach alongside FEM simulations. As illustrated, the STL derived from the current method employed in this study agrees with the outcomes obtained through the FEM simulation.
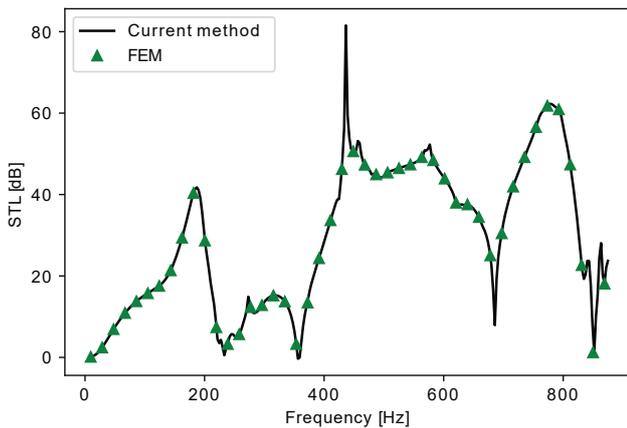

Fig. 5. Comparison of calculated STLs using the present approach and FEM results for double MPP coupled to both the receiving and source enclosure.

4.2. Comparison with impedance tube

Comparisons are made between the predicted TL when the finite partition is mounted on a rectangular impedance tube, and when it is coupled to acoustic enclosures. Particularly, we have selected the study provided by KIM *et al.* (2020b), which examines the propagation of two-dimensional plane waves over an insulating partition of a finite size 0.26×0.21 m consisting of two flexible MPPs mounted on the rectangular Kundt tube separated by an air gap of 0.03 m with clamped supported boundaries. Simulations were conducted using the same physical parameters as in the selected study (KIM *et al.*, 2020b): the MPPs are made of aluminium with the following physical properties; Young's modulus of 7.2×10^{10} Pa, the density of 2700 kg/m³, Poisson's ratio of 0.34, structural damping ratio of 0.01, 1 mm thickness, 0.8 mm diameter holes, and a perforation ratio of 0.05%. The chosen dimensions for the MPPs are 0.26×0.21 m. The dimension of the receiving acoustic enclosure is $L_x \times L_y \times h_e = 0.26$ m $\times 0.21$ m $\times 0.4$ m.

Figure 6 displays the findings for the TL when the double MPPs are excited by a plane wave coupled to a receiving enclosure, and when an emitting room is connected to the coupled system. Comparing these
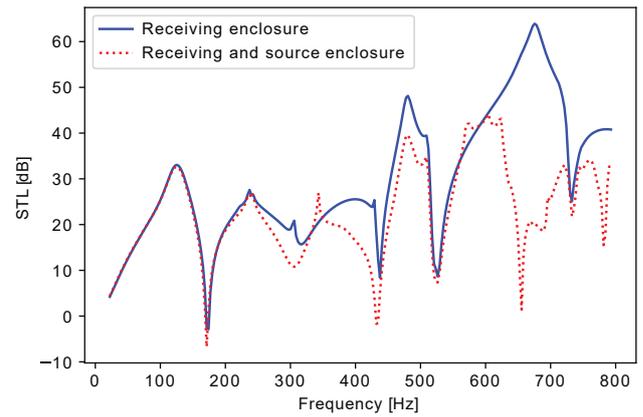
Fig. 6. Computed STL for double MPP coupled to acoustic enclosures using the same physical parameters as in (KIM *et al.*, 2020b).

findings to those of the finite double MPP partition put on an impedance tube reveals that the suggested modal formulation can predict the TL's general trend with high accuracy over a broad range of frequencies. Also, the STL is improved at plate-cavity-plate resonance in the coupled configuration similarly to the impedance tube case. It can be observed that there are visible and separated natural frequencies of the receiving room. As a result, intense dips and fluctuations marked the TL in addition to those presenting the modal behavior of the structure. However, it can be remarked that the STL is slightly degraded when coupled to an acoustic enclosure. The acoustic resonances present in the enclosure offer further information on the modal effects on the sound transmission. The TL drops turn negative (Fig. 6), while the TL remains positive in Kundt's tube example. This paradoxical feature is due to the resonance of the coupled system. To further study the effect of the source room on the transmission properties of the structure, a secondary enclosure is coupled to the multilayered MPPs as an emitting room. The dimensions of the source enclosure are $L_x \times L_y \times h_{e2} = 0.26$ m $\times 0.21$ m $\times 0.3$ m with a source strength $Q_0 = 10^{-4}$ m³/s placed at $(0.15, 0.12, 0.2)$. From Fig. 6, the predicted TL for the multi-layered MPPs when it is connected to a source and receiving an enclosure. It can be seen that the modal behavior of the emitting room is still present at very low frequencies presented by the dips and fluctuations on the STL curve corresponding to the excited acoustic modes in both the source and receiving enclosures.

4.3. Parametric analysis

Numerical modeling is used to investigate the impact of various system parameters on the sound transmission of the double MPPs structure. Numerical analysis is employed to assess various critical system parameters. These include factors such as the dimensions

and impedance of the acoustic enclosure, the perforation ratio of the MPP, the thickness of the air gap, and the incorporation of absorbent materials within the gap. In the first case, the upper MPP is under acoustic excitation and an acoustic point source of strength $Q_0 = 10^{-4}$ m^3/s is placed at $(0.15, 0.12, 0.4)$ in the source room in the second. In numerical simulations, the dimensions of the coupled enclosures are defined as follows: the dimension of the enclosure is $L_x \times L_y \times h_{e1} = 0.5$ m $\times 0.35$ m $\times 0.6$ m, the dimension of the source enclosure when coupled to the system is $L_x \times L_y \times h_{e2} = 0.5$ m $\times 0.35$ m $\times 0.7$ m, and the physical properties of the MPPs are kept the same as in the previous section. The air gap depth (h_a) is 70 mm and the hole diameter is 0.3 mm.

4.3.1. Effect of acoustic enclosure dimensions on STL

To investigate the modal effects of both the receiving enclosure and emitting enclosure on STL, a set of numerical simulations with varying h_e is conducted. Figure 7 displays the impact of different receiving enclosure depths on the sound transmission loss of the double MPP structure when connected to an acoustic enclosure under acoustic excitation.
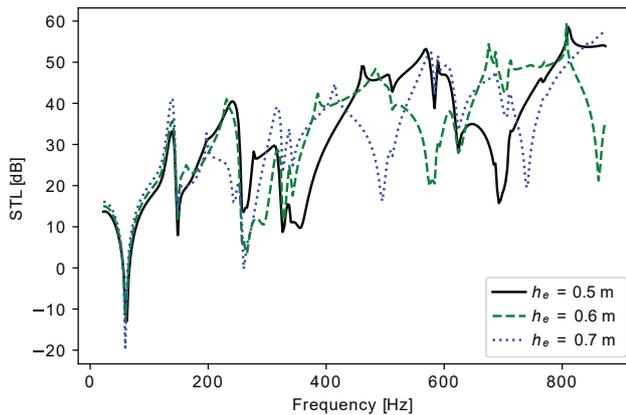
Fig. 7. Sound transmission loss of double MPP for different dimensions of receiving acoustic enclosure, $\sigma_1 = \sigma_2 = 0.1\%$.

Using the potential energy in the receiving chamber for calculation, Fig. 7 illustrates the corresponding transmission loss. The introduction of the receiving enclosure has notably influenced the previous results, primarily due to its modal properties playing a role. Below the first mode controlled by the enclosure, energy transmission is primarily influenced by structural resonances. The observable difference in the curves arises because the modes of the coupled enclosure amplify the mean square pressure within it for a given panel excitation. Additionally, changes in the enclosure size lead to modifications in the resonance frequencies, resulting in shifts in the dips corresponding to excited enclosure modes. It is worth noting that a decrease in the cavity depth results in an increase in the associated frequencies of the acoustical cavity, as seen in

Fig. 7. For instance, the dip generated by the acoustic mode at 678 shifts to 592 Hz when the depth changes from 0.5 to 0.6 m. In the second case when the double MPPs is connected to a source room, the depth of the emitting enclosure is varied (h_{e2} = 0.6, 0.7, and 0.8 m) while maintaining the receiving enclosure depth constant and equal to 0.6 m. It can be seen from Fig. 8 that additional dips and peaks are present. This can be explained by the sensitivity of the diffuseness of the acoustic field to many parameters such as the source position and the room geometry. The presence of receiving enclosure or a source and receiving enclosure wield a significant influence on the STL of double MPP structures. Hence, the transmission loss of double MPPs connected to acoustic enclosures necessitates thorough examination in each case.
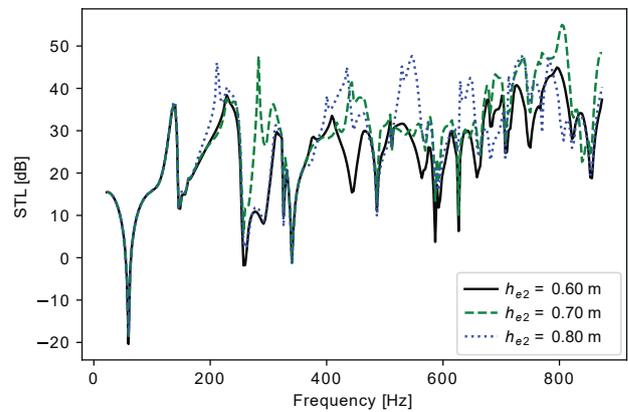
Fig. 8. Sound transmission loss of double MPP for different dimensions of source acoustic enclosure, $\sigma_1 = \sigma_2 = 0.1\%$.

4.3.2. Effect of acoustic enclosure impedance on STL

The coupled system transmits the acoustic excitation into the enclosure. The enclosure's boundary conditions may affect the acoustic properties of the enclosure and further influence the transmission of energy via the double MPP structure. Consequently, it is crucial to examine the influence of wall impedance on energy transmission. Figure 9 shows how the coupled
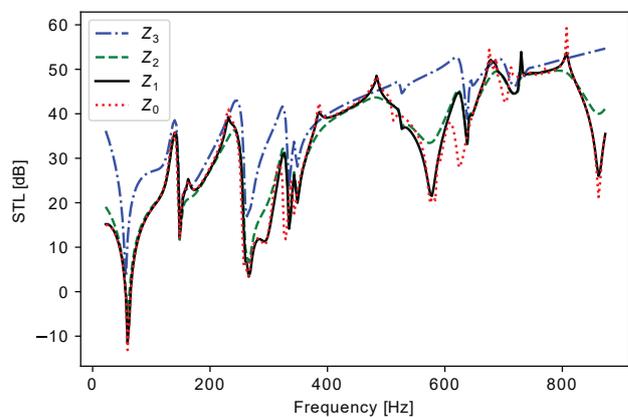
Fig. 9. Effect of impedance on STL, $\sigma_1 = \sigma_2 = 0.1\%$.

system's sound insulation properties change when the impedance is modified. Note that impedance boundary conditions are accounted for by modifying Eq. (2). The impact of the impedance's real component is investigated.

The influence of the real component of the impedance on the STL is shown in Fig. 9. STLs are computed with specified values of impedance $Z_0 = j \times 10^8$, $Z_1 = \rho_0 c(200 - 2j)$, $Z_2 = \rho_0 c(20 - j2)$, $Z_3 = \rho_0 c(2 - 2j)$. These three numbers represent the corresponding sound absorption coefficients: rigid, 0.04, 0.58, and 0.86. As can be observed in Fig. 9, the STL improves whenever there is a general rise in the sound absorption coefficient. In cases with very low absorption coefficients, the STL over the whole frequency spectrum that was investigated, is practically the same as that of rigid walls. More than 0.04 sound absorption, STL occurs mostly at the resonance frequencies of the enclosure but is largely unchanged at the initial resonant frequencies of the system. This is because the real component of the impedance increases sound transmission capabilities by enabling the cavity to dissipate energy more efficiently. At the resonant acoustic modes, it is thus plausible to conclude that the actual component of wall impedance has the greatest effect on sound transmission.

4.3.3. Effect of perforation ratio of the MPPs on STL

To investigate the effect of the perforation ratio on the multilayered MPPs for both cases (when the MPPs are coupled to the receiving room and when the MPP is coupled to both the receiving and emitting room), the STLs are plotted for different perforation ratio values. The micro-perforation ratios of the two MPPs are equal: $\sigma_1 = \sigma_2 = 0.0$, 0.05, 0.1, and 0.2%. As shown in Figs. 10 and 11, as the micro-perforation ratios grow, the STL deteriorates for both cases with a large gain at resonances.
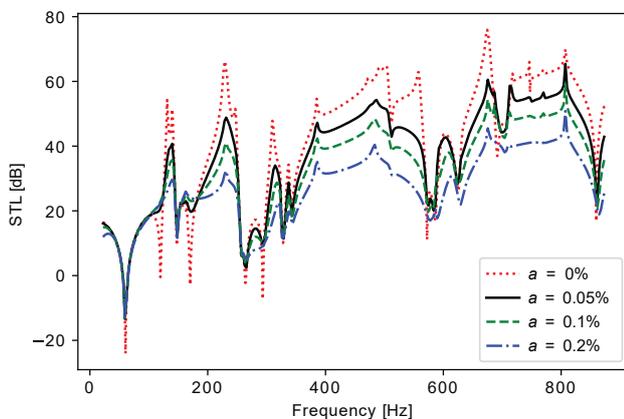
Fig. 10. STL of double MPPs connected to a receiving enclosure for various perforation ratios when $\sigma_1 = \sigma_2$.

Perforation ratios of the two MPPs are normally varied in a typical arrangement to reduce reflections
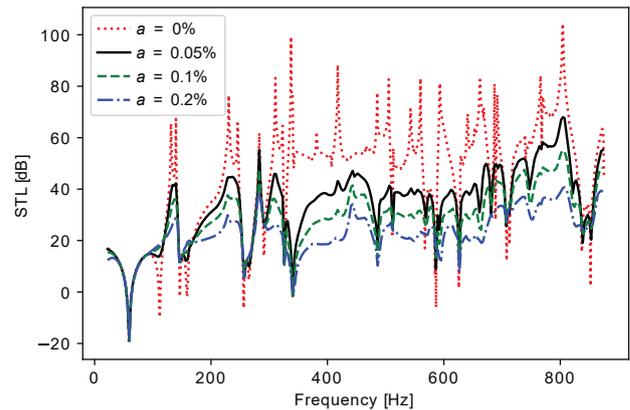
Fig. 11. STL of double MPPs connected to an emitting and receiving enclosure for various perforation ratios when $\sigma_1 = \sigma_2$.

and transmissions from internal noise sources such as those produced by the engine and sent into the passenger compartment or in adjacent rooms. Figure 12 illustrates the calculated STLs for various perforation ratios in the first case.
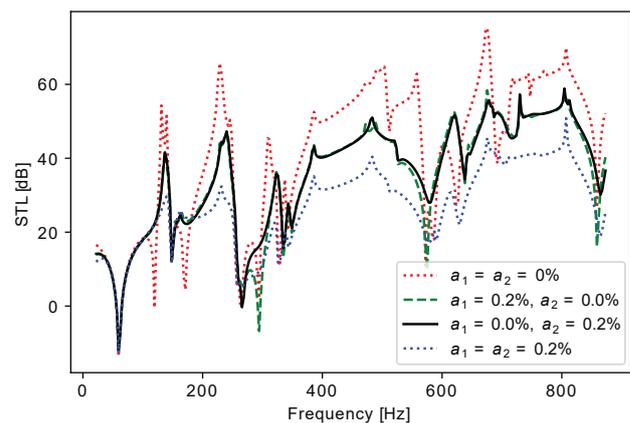
Fig. 12. STL of double MPPs connected to a receiving enclosure for various perforation ratios when $\sigma_1 \neq \sigma_2$.

Similar to the case of equal micro-perforation, the STLs deteriorate when the plate is perforated, as shown in Fig. 12. In addition, the STL is dependent on the variation of perforation ratios of the two MPPs, as illustrated in Fig. 12, for the case of (0.0, 0.2%) and (0.2, 0.0%). However, a large gain is obtained at resonances with the arrangement (0.0, 0.2%).

Figure 13 displays the sound transmission loss when the double MPPs is coupled to both the source and receiving enclosures. Acoustic resonances in the receiving chamber (dips) are attenuated when the perforations are located on the receiving side (0.0, 0.2%). Similarly, acoustic resonances in the source enclosure (peaks) are also damped when the perforations are positioned on the source side (0.2, 0.0%). For the arrangement (0.2, 0.2%), the resonances of both the receiving and source enclosure are damped. However, the sound transmission loss is highly degraded.
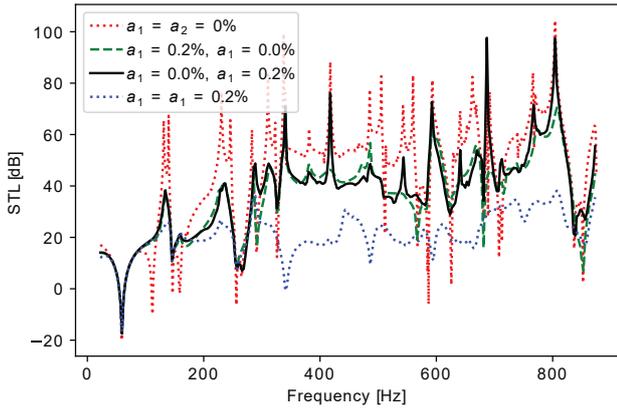
Fig. 13. STL of double MPPs connected to an emitting and receiving enclosure for various perforation ratios when $\sigma_1 \neq \sigma_2$.

4.3.4. Effect of the gap thickness

To assess the influence of varying air gap depths on STL, a series of numerical simulations were conducted. Figures 14 and 15 illustrate how the STL of a double MPP structure responds to different air gap depths
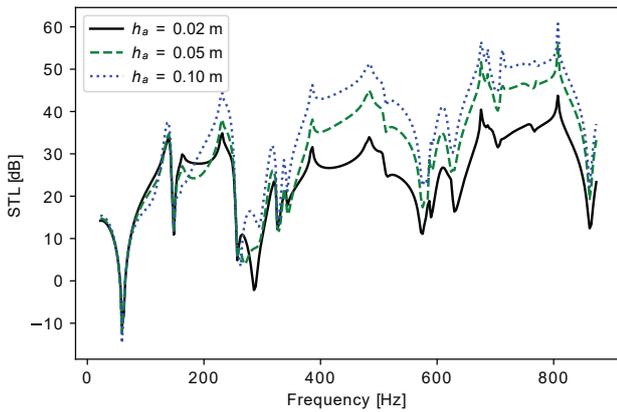
Fig. 14. Influence of air gap on STL of double MPP partition, $\sigma_1 = \sigma_2 = 0.1\%$.
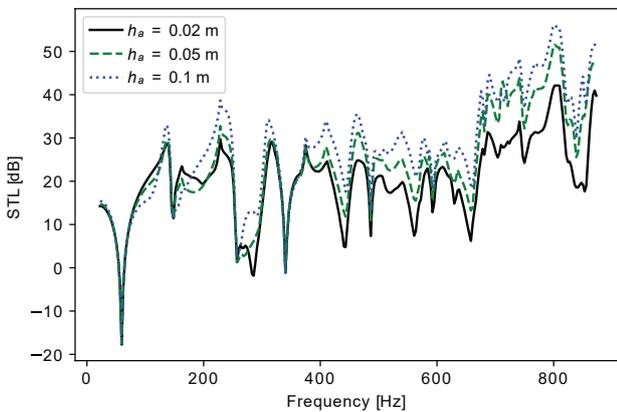
Fig. 15. Influence of air gap on STL of double MPP partition when coupled to both source and receiving enclosure, $\sigma_1 = \sigma_2 = 0.1\%$.

(0.02, 0.05, and 0.10 m) when subjected to acoustic excitation and when it is coupled to both the source and receiving room. As depicted in Figs. 14 and 15, as the air gap thickness increases, the STL values increase. This phenomenon is attributed to an increase of system coupling resulting from the reduced air gap thickness. As the air gap thickness rises, Figs. 14 and 15 demonstrate that the transmission characteristics in the case when coupled to a receiving chamber increase greater than those of the second case. Thus, each case should be carefully considered. In conclusion, the thickness of the air gap significantly affects the sound insulation capabilities of the structure. In the case, when the partition is coupled to both the receiving and source enclosure the dissipation properties are significant when the air gap is increased.

4.3.5. Influence of porous material on sound transmission loss

To enhance the sound transmission loss at resonant frequencies, we introduce sound-absorbing material with a specific flow resistivity into the space between the MPPs. This absorbent material is represented as an equivalent fluid, and its material characteristics are determined using the empirical model developed by DELANY and BAZLEY (1970). The associated weak form (Eq. (15)) may account for airborne propagation through low frame stiffness insulating materials filling the cavity, by replacing c_0 and ρ_0 with c_a and ρ_a which are complex and frequency-dependent. It is important to note that for a comprehensive description, poroelastic modeling should be considered. However, as demonstrated by BERANEK in 1947, when sound propagates through soundproofing materials with relatively low frame stiffness, longitudinal elastic vibrations attenuate significantly more than acoustic compression waves.

Figure 16 presents a comparison of computational results for a gap filled with fiberglass as opposed to air for the case when the partition is coupled to both the
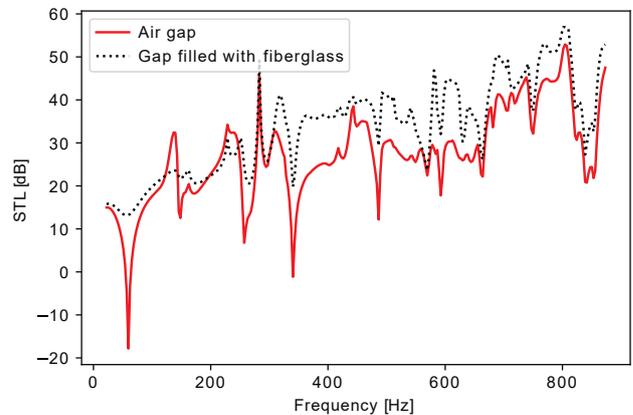
Fig. 16. Computed STLs for gap filled with fiberglass and gap filled with air.

receiving and source enclosure. The findings reveal an increase in transmission loss. Notably, significant gains are observed at resonance points. The system's response is predominantly influenced by the damping introduced by the absorbent material. This damping is primarily attributed to acoustic dissipation resulting from viscous drag forces and thermal interactions between the air and the material.

4.4. Sound transmission loss of triple MPPs

In Figs. 17 and 18, we have plotted the sound transmission loss for different perforation ratios for the two cases. The respective thicknesses and hole diameters of the three MPPs are consistent at $h_1 = h_2 = h_3 = 1$ mm and $d_1 = d_2 = d_3 = 0.3$ mm, with cavity depths equal to $h_{a1} = h_{a2} = 40$ mm. The dimensions of the enclosures remain unchanged.
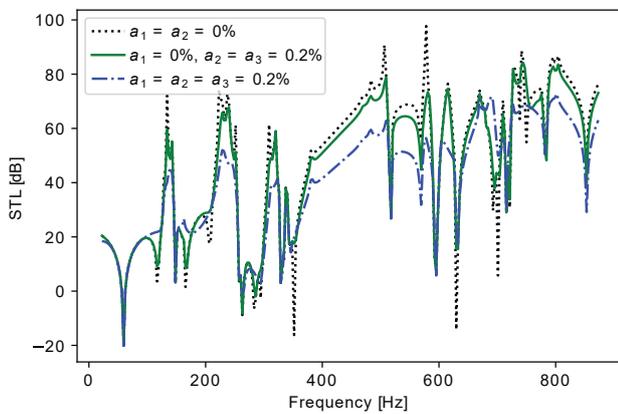
Fig. 17. STL of triple MPPs for various perforation ratios of the MPP when coupled to the receiving enclosure.
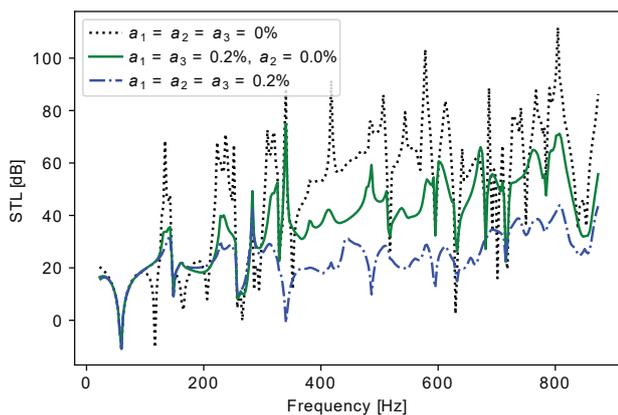
Fig. 18. STL of triple MPPs for various perforation ratios of the MPP when coupled to both receiving and source enclosure.

As observed with double MPPs, the modal effect of the acoustic enclosure is present and the STL values tend to decrease as the micro-perforation ratio increases, except at the resonant frequencies of the coupled system where large gain is obtained.

5. Conclusion

This study presents the results of a vibro-acoustic investigation of a double MPP structure coupled to receiving enclosure or both the source and receiving enclosure. The transmission loss of this structure was predicted using an analytical methodology. Using a modified variational model, the equations governing the coupled system were derived. The acoustic pressure within the enclosures is described by the 3D-enhanced Fourier series. Furthermore, the validity and precision of this model were confirmed through FE computations, resulting in a substantial level of agreement. Then a parametric analysis was carried out aimed at discerning the influence of enclosure size, air gap thickness, acoustic enclosure impedance, perforation ratio of the MPPs, and insertion of absorbent material in the gap on the sound transmission of double MPPs in the coupled configurations.

The results summarized here are the most significant. It has been shown that the presence of an acoustic enclosure on the receiving side or both sides may have a significant effect on the sound transmission loss of double MPP structures, resulting in additional dips and peaks associated with the excited acoustic modes of both the source and receiving enclosure. In addition, the real component of the wall impedance dissipates the energy of the enclosure to affect the sound transmission over the double MPP structure, and thus primarily influences the sound transmission at the natural frequencies of the enclosure. As expected, the STLs deteriorate with the increase of the micro-perforation ratio of the plate with a gain at resonances. In the case where the double MPP is coupled to both the receiving and emitting enclosure, the sound transmission loss is damped at acoustic resonances in the two enclosures. Furthermore, it was found that as the thickness of the air gap increases, the Sound Transmission Loss values also increase pointing out that there is difference in the increase in STL between the two situations. Additionally, the insertion of absorbent material into the gap, in the case where the partition is coupled to both an emitting enclosure and receiving enclosure, leads to an increase in transmission loss, especially at resonance points. In the case of triple MPPs, we observe consistent behavior. Furthermore, when excited by a plane wave and connected to an acoustic enclosure, the plate-MPP-MPP structure demonstrated optimal performance. On the other hand, the MPP-plate-MPP configuration exhibited good performance when connected to both receiving and source enclosures.

The introduction of an acoustic enclosure on the receiving side or both sides can have a substantial effect on the sound transmission loss of double MPP structures. Hence, the transmission loss of double MPPs connected to acoustic enclosures necessitates thor-

ough examination in each case. Compared to earlier methods, this approach offers greater ease in defining various structural and acoustic boundary conditions. Moreover, the suggested method allows the direct application to more complex geometric models that only require coordinate transformation to transform irregular shapes into rectangles.

References

1. ALISAH M.I., OOI L.E., RIPIN Z.M., YAHYA A.F., HO K. (2021), Acoustic attenuation performance analysis and optimisation of expansion chamber coupled micro-perforated cylindrical panel using response surface method, *Archives of Acoustics*, **46**(3): 507–517, doi: 10.24425/aoa.2021.138143.

2. ALLAM S., ÅBOM M. (2011), A new type of muffler based on microperforated tubes, *Journal of Vibration and Acoustics*, **133**(3): 031005, doi: 10.1115/1.4002956.

3. BERANEK L.L. (1947), Acoustical properties of homogeneous, isotropic rigid tiles and flexible blankets, *The Journal of the Acoustical Society of America*, **19**(4): 556–568, doi: 10.1121/1.1916521.

4. BRAVO T., ELLIOTT S.J. (2004), Variability of low frequency sound transmission measurements, *The Journal of the Acoustical Society of America*, **115**(6): 2986–2997, doi: 10.1121/1.1738452.

5. BRAVO T., MAURY C., PINHÈDE C. (2012), Sound absorption and transmission through flexible microperforated panels backed by an air layer and a thin plate, *The Journal of the Acoustical Society of America*, **131**(5): 3853–3863, doi: 10.1121/1.3701987.

6. BRAVO T., MAURY C., PINHÈDE C. (2014), Optimising the absorption and transmission properties of aircraft microperforated panels, *Applied Acoustics*, **79**: 47–57, doi: 10.1016/j.apacoust.2013.12.009.

7. BRAVO T., MAURY C., PINHÈDE C. (2016), Optimisation of micro-perforated cylindrical silencers in linear and nonlinear regimes, *Journal of Sound and Vibration*, **363**: 359–379, doi: 10.1016/j.jsv.2015.11.011.

8. CARNEAL J.P., FULLER C.R. (2004), An analytical and experimental investigation of active structural acoustic control of noise transmission through double panel system, *Journal of Sound and Vibration*, **272**(3–5): 749–771, doi: 10.1016/S0022-460X(03)00418-8.

9. CHAZOT J.-D., GUYADER J.-L. (2007), Prediction of transmission loss of double panels with a patch-mobility method, *The Journal of the Acoustical Society of America*, **121**(1): 267–78, doi: 10.1121/1.2395920.

10. CHENG L., LI Y.Y., GAO J.X. (2005), Energy transmission in a mechanically-linked double-wall structure coupled to an acoustic enclosure, *The Journal of the Acoustical Society of America*, **117**(5): 2742–2751, doi: 10.1121/1.1886525.

11. CHIEN W.Z. (1983), Method of high-order Lagrange multiplier and generalized variational principles of elasticity with more general forms of functional, *Applied Mathematics and Mechanics*, **4**: 143–157, doi: 10.1007/BF01895439.

12. DELANY M.E., BAZLEY E.N. (1970), Acoustical properties of fibrous materials, *Applied Acoustics*, **3**(2): 105–116, doi: 10.1016/0003-682X(70)90031-9.

13. DU J.T., LI W.L., LIU Z.G., XU H.A., JI Z.L. (2011), Acoustic analysis of a rectangular cavity with general impedance boundary condition, *The Journal of the Acoustical Society of America*, **130**(9): 807–817, doi: 10.1121/1.3605534.

14. DUPONT T., PAVIC G., LAULAGNET B. (2003), Acoustic properties of lightweight micro-perforated plate systems, *Acta Acustica united Acustica*, **89**(2): 201–212.

15. FALSAFI I., OHADI A. (2018), Optimisation of multi-step cavity configuration to extend absorption bandwidth of micro perforated panel absorber, *Archives of Acoustics*, **43**(2): 187–195, doi: 10.24425/122366.

16. FUCHS H.V., ZHA X. (1997), Acrylic-glass sound absorbers in the plenum of the deutscher bundestag, *Applied Acoustics*, **51**(2): 211–217, doi: 10.1016/S0003-682X(96)00064-3.

17. GAGLIARDINI L., ROLAND J., GUYADER J.L. (1991), The use of a functional basis to calculate acoustic transmission between rooms, *Journal of Sound and Vibration*, **145**(3): 457–478, doi: 10.1016/0022-460X(91)90114-Y.

18. KANG J., BROCKLESBY M.W. (2005), Feasibility of applying micro-perforated absorbers in acoustic window systems, *Applied Acoustics*, **66**(6): 669–689, doi: 10.1016/j.apacoust.2004.06.011.

19. KIHLMAN T. (1967), Sound radiation into a rectangular room. Applications to airborne sound transmission in buildings, *Acta Acustica*, **18**(1): 11–20.

20. KIM H.-S., KIM S.-R., KIM B.-K., MA P.-S., SEO Y.-H. (2020a), Sound transmission loss of multilayered infinite micro-perforated plates, *The Journal of the Acoustical Society of America*, **147**(1): 508–515, doi: 10.1121/10.0000600.

21. KIM H.-S., MA P.-S., KIM B.-K., LEE S.-H., SEO Y.-H. (2020b), Sound transmission loss of multi-layered elastic micro-perforated plates in an impedance tube, *Applied Acoustics*, **166**: 107348, doi: 10.1016/j.apacoust.2020.107348.

22. KROPP W., PIETRZYK A., KIHLMAN T. (1994), On the meaning of the sound reduction index at low frequencies, *Acta Acustica (Les Ulis)*, **2**(5): 379–392.

23. LIU Z., ZHAN J., FARD M., DAVY J.L. (2017), Acoustic properties of multilayer sound absorbers with a 3D printed micro-perforated panel, *Applied Acoustics*, **121**: 25–32, doi: 10.1016/j.apacoust.2017.01.032.

24. LØVHOLT F., NORÈN-COSGRIFF K., MADSHUS C., ELLINGSEN S.E. (2017), Simulating low frequency sound transmission through walls and windows by a two-way coupled fluid structure interaction model, *Journal of Sound and Vibration*, **396**: 203–216, doi: 10.1016/j.jsv.2017.02.026.

25. MAA D.-Y. (1998), Potential of microperforated panel absorber, *The Journal of the Acoustical Society of America*, **104**(5): 2861–2866, doi: 10.1121/1.423870.

26. Meng H., Galland M.A., Ichchou M., Bareille O., Xin F.X., Lu T.J. (2017), Small perforations in corrugated sandwich panel significantly enhance low-frequency sound absorption and transmission loss, *Composite Structures*, **182**: 1–11, doi: 10.1016/j.compstruct.2017.08.103.

27. Meng H., Galland M.A., Ichchou M., Xin F.X., Lu T.J. (2019), On the low-frequency acoustic properties of novel multifunctional honeycomb sandwich panels with micro-perforated faceplates, *Applied Acoustics*, **152**: 31–40, doi: 10.1016/j.apacoust.2019.02.028.

28. Mu R.L., Toyoda M., Takahashi D. (2011a), Improvement of sound insulation performance of multilayer windows by using microperforated panel, *Acoustical Science and Technology*, **32**(2): 79–81, doi: 10.1250/ast.32.79.

29. Mu R. L., Toyoda M., Takahashi D. (2011b), Sound insulation characteristics of multi-layer structures with a microperforated panel, *Applied Acoustics*, **72**(11): 849–855, doi: 10.1016/j.apacoust.2011.05.009.

30. Mulholland K.A., Lyon R.H. (1973), Sound insulation at low frequencies, *The Journal of the Acoustical Society of America*, **54**(4): 867–878, doi: 10.1121/1.1914340.

31. Osipov A., Mees P., Vermeir G. (1997a), Numerical simulations of airborne sound transmission at low frequencies: The influence of the room and the partition parameters, [in:] *Proceedings of InterNoise 97*, pp. 759–762.

32. Osipov A., Mees P., Vermeir G. (1997b), Low-frequency airborne sound transmission through single partitions in buildings, *Applied Acoustics*, **52**(3–4): 273–288, doi: 10.1016/S0003-682X(97)00031-5.

33. Pfretzschner J., Cobo P., Simón F., Cuesta M., Fernández A. (2006), Microperforated insertion units: An alternative strategy to design microperforated panels, *Applied Acoustics*, **67**(1): 62–73, doi: 10.1016/j.apacoust.2005.05.005.

34. Qu Y.G., Hua H.X., Meng G. (2013a), A domain decomposition approach for vibration analysis of isotropic and composite cylindrical shells with arbitrary boundaries, *Composite Structures*, **95**: 307–321, doi: 10.1016/j.compstruct.2012.06.022.

35. Qu Y.G., Chen Y., Long X.H., Hua H.X., Meng G. (2013b), A modified variational approach for vibration analysis of ring-stiffened conical–cylindrical shell combinations, *European Journal of Mechanics – A/Solids*, **37**: 200–215, doi: 10.1016/j.euromechsol.2012.06.006.

36. Takahashi D., Tanaka M. (2002), Flexural vibration of perforated plates and porous elastic materials under acoustic loading, *Journal of the Acoustical Society of America*, **112**(4): 1456-1464, doi: 10.1121/1.1497624.

37. Tang Y., Xin F., Huang L., Lu T.J. (2017), Deep subwavelength acoustic metamaterial for low-frequency sound absorption, *Europhysics Letters*, **118**(4): 44002, doi: 10.1209/0295-5075/118/44002.

38. Tang Y., Xin F., Huang L., Lu T.J. (2019), Sound absorption of micro-perforated sandwich panel with honeycomb-corrugation hybrid core at high temperatures, *Composite Structures*, **226**: 111285, doi: 10.1016/j.compstruct.2019.111285.

39. Toyoda M., Takahashi D. (2008), Sound transmission through a microperforated-panel structure with subdivided air cavities, *The Journal of the Acoustical Society of America*, **124**(6): 3594–3603, doi: 10.1121/1.3001711.

40. Xin F.X., Lu T.J., Chen C.Q. (2008), Vibroacoustic behavior of clamp mounted double-panel partition with enclosure air cavity, *The Journal of the Acoustical Society of America*, **124**(6): 3604–3612, doi: 10.1121/1.3006956.

41. Yang C., Cheng L. (2016), Sound absorption of microperforated panels inside compact acoustic enclosures, *Journal of Sound and Vibration*, **360**: 140–155, doi: 10.1016/j.jsv.2015.09.024.

42. Yu X., Cheng L., You X. (2015), Hybrid silencers with micro-perforated panels and internal partitions, *The Journal of the Acoustical Society of America*, **137**(2): 951–962, doi: 10.1121/1.4906148.

Research Paper

Modeling of Acoustic Coupling of Ultrasonic Probes for High-Speed Rail Track Inspection

Sławomir MACKIEWICZ, Zbigniew RANACHOWSKI*, Tomasz KATZ,
Tomasz DĘBOWSKI, Grzegorz STARZYŃSKI, Przemysław RANACHOWSKI

Institute of Fundamental Technological Research
Polish Academy of Sciences
Warsaw, Poland

*Corresponding Author e-mail: zranach@ippt.pan.pl

The paper presents the modeling of transmission of the ultrasonic plane wave through an uniform liquid layer. The considered sources of the ultrasonic wave were normal (straight) beam longitudinal wave probes and angle beam sheer waves probes commonly used in non-destructive testing. Coupling losses (CL) introduced by the presence of the coupling layer are discussed and determined applying the numerical procedure. The modeling applies to both monochromatic waves and short ultrasonic pulses with a specified frequency bandwidth. Model implementation and validation was performed using a specialized software. The predictions of the model were confirmed by coupling losses measurements for a normal beam longitudinal wave probe with a delay line made of polymethyl methacrylate (PMMA). The developed model can be useful in designing ultrasonic probes for high-speed rail track inspections, especially for establishing the optimal thickness of the water coupling layer and estimation of coupling losses, due to inevitable changes of the water gap during mobile rail inspection.

Keywords: non-destructive testing; ultrasonic examination; plane wave propagation.

1. Introduction

Rail track failures create a significant problem enforcing the railroad administration to permanent control of the integrity of exploited infrastructure. Both International Union of Railways [UIC] as well as Federal Railroad Administration in USA issue the legislative rules for regulating the safety of the subject railroad system (UIC, 2022; Federal Railroad Administration, 2015). A detailed codes of appropriate procedures of rail track testing can be found in European Union Standards (EN 16729-3, 2018; EN 17397-1, 2021). A remarkable challenge related to the testing procedure is the total length of the railway network, which is by UIC estimated at 260 000 kilometers.

Railway rails are exposed to high mechanical loads and challenging environmental conditions such as rolling contact fatigue, thermal stresses and corrosive environment. Due to such operational conditions different types of defects can develop in the rail head,

web, and foot (BRAY, 2000). The rolling contact defects, like wear, stripping, crushing, and fatigue cracks are distributed mainly on the surface and in the upper part of the rail head. Many of these defects can steadily grow and finally cause rail breakage, leading to derailments or more catastrophic events. Such extreme consequences can be avoided, provided that proper inspection procedures are performed and all unacceptable defects are detected before they cause the catastrophic failure.

A variety of equipment is applied for ultrasonic inspection (PAPAELIAS *et al.*, 2008). The simplest solution is a push-trolley. In this case the operator is moving across the track with his instrumentation at a walking pace, simultaneously interpreting the test data on a flaw detector. When a suspect defect is identified, the operator stops and manually verifies the defect type and location. The most efficient solution used for ultrasonic rail inspection is the use of specialized wagons or entire inspection trains, as it was described in

(Heckel *et al.*, 2018). The detailed investigation has revealed that the efficient operation of the SPZ1 train is possible below 80 km/h and the practical inspection speed is highly influenced by the local quality of the track. The further progress in ultrasonic testing of railway rails depends heavily on computer modeling and simulations of all aspects of the ultrasonic inspections process (Heckel *et al.*, 2019). One of these aspects is quality and stability of acoustic coupling between the testing probe, and the rail surface. No extensive research on this specific subject has been carried to date. One of the few published works (Zulian, 2022) explores the effect of the coupling media type and surface roughness on contact transfer losses. Unfortunately, in the case of automatic railway rail inspection, the only possible coupling agent is water, and the main reason for the fluctuation of transfer losses is the instability of the water layer thickness rather than roughness of the rail surface.

The inspection probes are mounted in specially designed probe holders enabling the fixed position of the probes over the tested rail. The holders are equipped with water bleeders, that provide acoustic coupling between the ultrasonic probes and the rail surface.

The quality of the acoustic coupling between the probe and the tested rail is one of the key issues, related to the high speed ultrasonic inspection of railway tracks. Under real test conditions – the gap between the rail surface and testing probes may vary due to waviness and dents on the rail surface. To ensure the continuous transmission of ultrasonic waves to the rail body, the gap must be constantly filled with water without any air bubbles or cavitation. To fulfill this critical condition, the water coupling system for high speed scanning must be designed in much greater detail than for ordinary ultrasonic inspections, where the water gap is usually undefined and uncontrolled. An example of such an ordinary probe holder, used in moderate speed inspection wagons (i.e., 30–40 km/h) is presented in Fig. 1. In such a solution – both the lower surface of the probe holder and the surfaces of ultrasonic probe wedges are pressed directly to the wetted rail surface, without any distance. It means that the water supplied by the dispensers located at the be-

ginning of the probe holder, can be wiped off the rail surface before it fills the gap between the active surface of the probe and the rail. Such a method has turned out to be impractical in the case of high speed rail inspection systems. In that case, other solution prefer train probe holders equipped with some abrasion resistive slides – to set the gap between the probe surface and rail surface to 0.2 mm (Heckel*et al.*, 2009). Such an arrangement ensures better water coupling and enhances the probe lifespan.
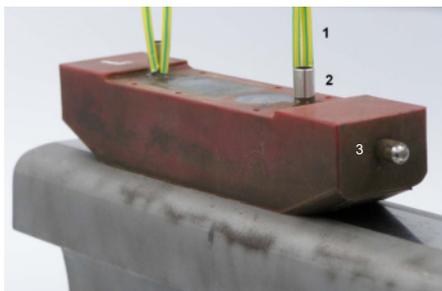
The objective of the presented paper is to analyze in detail the influence of the probe – rail gap filled with water on the transfer loses of ultrasonic energy emitted and received by ultrasonic transducers. The practical purpose of these research is to establish optimal thickness of the water coupling layers for ultrasonic probes of different types (angle probes, normal probes), operating at different frequencies and refraction angles. It is expected that the optimal gap thickness heavily depends on ultrasonic waves frequency as well, as on the angle of incidence on wedge – rail contact. The analysis should consider not only the absolute minimization of the transfer losses but, also the minimization of ultrasonic signal fluctuation, due to inevitable variability of the water gap thickness under practical conditions. Minimization of signal fluctuations due to coupling variations is particularly important to maintain the constant sensitivity of ultrasonic inspection along the whole rail length.

In order to achieve the intended goals, the new theoretical model for calculation of transmission losses through the coupling layer was developed. The model is more general and comprehensive than simplified traditional solutions, used up to now in ultrasonic non-destructive testing and described in (Krautkrämer, Krautkrämer, 1990; Obraz, 1983). The first novelty is going beyond the case of perpendicular incidence of the wave on the coupling layer. Due to this, the model is applicable not only to normal beam longitudinal wave probes, but also to angle beam shear wave probes, which are commonly used in non-destructive testing. The second novelty is going beyond the case of monochromatic wave, and taking into consideration the wideband nature of modern ultrasonic probes.

The developed model was implemented in the computer program and used for example calculations, showing its compatibility with existed analytical formulas for normal incidence on the coupling layer. A very good agreement was also achieved with experimental results obtained for the typical ultrasonic probes used in non-destructive testing of railway rails.

2. Theoretical model

2.1. General considerations

In this section the general theoretical model of transmission of ultrasonic plane wave through the uni-

Fig. 1. General view of ultrasonic probe holder used for inspection in Polish Railways: 1 – signal cable; 2 – water coupling bleeder connector; 3 – positioning bracket.

form liquid layer is developed. It is more general, than widely known formulas presented in ultrasonic textbooks (KRAUTKRÄMER, KRAUTKRÄMER, 1990), which consider only the normal incidence of ultrasonic wave on the contact layer. It assumes the longitudinal wave incidence at an arbitrary angle, as it actually takes place in angle beam ultrasonic probes. The theoretical treatment follows the one used by FOLDS and LOGGINS (1977) in their paper on transmission and reflection of ultrasonic waves in layered media. Compared to the mentioned theoretical work there is one important change. Folds and Loggins considered transmission of plane wave from one semi-infinite liquid media to another semi-infinite liquid media, through a system of plane-parallel solid layers. In this work, transmission of the longitudinal (*L*-type) wave from one semi-infinite solid media to another semi-infinite solid media through a plane parallel liquid layer, is considered. Due to this change it is possible to model, not only direct transmission of longitudinal wave, but also the transmission with transformation from a longitudinal to a transversal wave – as it actually takes place in shear wave angle beam probes, widely used in ultrasonic testing of railway rails.

The transmission of ultrasonic waves through the system of plane-parallel layers was extensively investigated, in the context of underwater sound applications for optimization of sonar domes, underwater transducer windows and reflectors. One of the first theoretical works which considered the transmission of ultrasonic plane wave through the system of plane-parallel layers at oblique incidence was by BREKHOVSKIKH (1980). As there were some restrictions concerning the validity of equations presented in that paper, the other authors (BARNARD *et al.,* 1975;

FOLDS, LOGGINS, 1977) improved the solution to be valid for a system of solid layers with arbitrary parameters. However, in all these treatments, it was assumed that both the initial and final medium is liquid – as it was natural for underwater applications.

In the case of ultrasonic angle beam probes, used for nondestructive testing, the ultrasonic wave is transmitted from a solid wedge made of PMMA or Rexolite to a solid rail material (steel), through the liquid coupling layer (water). Therefore, the considered problem is somewhat different from the mentioned hydro-acoustic problems. The general scheme of a wave propagation in a coupling layer problem is shown in Fig. 2.

In the probe wedge (medium 3) there is one incident longitudinal (*L*-type), a wave with an arbitrary incidence angle θ_3 and two reflected waves, *L*-type and transversal (*T*-type), with reflection angles, respectively – θ_3 and θ_3', given by the Snell law. In the coupling layer (medium 2) there is only one refracted *L*-type wave and one reflected *L*-type wave, both with angles θ_2 to the normal. In the tested material (medium 1), there is generally one refracted *L*-type wave and one refracted *T*-type wave, with refraction angles, respectively, θ_1 and θ_1'. But, if the incidence angle θ_3 is between the first and second critical angle there is only one, *T*-type wave – which propagates in medium 1. This is actually the case encountered in angle beam shear wave probes, used in nondestructive testing. However, it should be noted that instead of a sinusoidal *L*-type wave propagating in the tested material, there exists so called evanescent *L*-type wave, exponentially decaying from the material surface as in the work (SCHMERR JR., 2016). It must be included in the theoretical model, despite the fact that it has no practical significance.
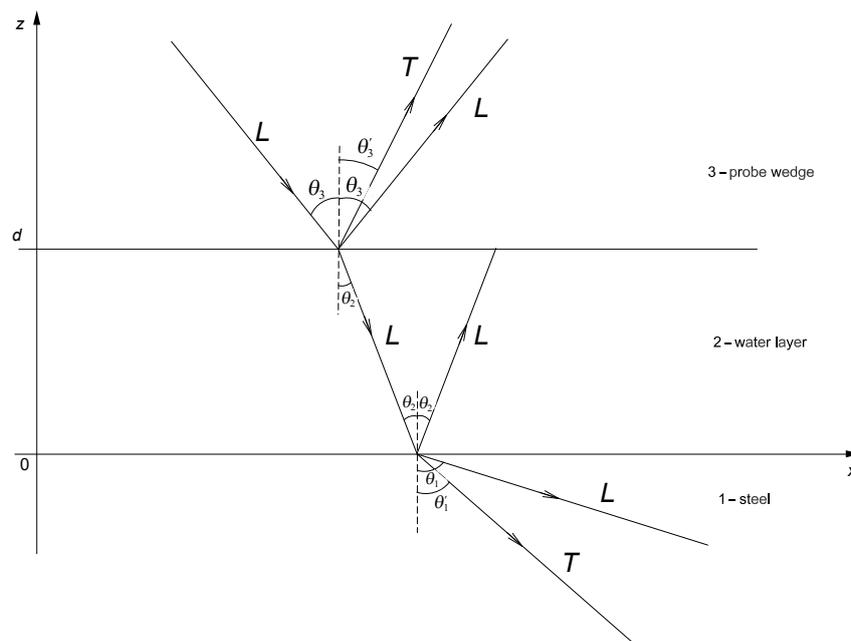

Fig. 2. General scheme of a coupling layer problem.

The all three media are assumed to be perfectly elastic and isotropic with Lame moduli: λ_i, μ_i, mass density ρ_i, and ultrasonic velocities of longitudinal waves $Cl_i = \sqrt{(\lambda_i + 2\mu_i)/\rho_i}$, and transversal waves $Ct_i = \sqrt{\mu_i/\rho_i}$, respectively. Index i denotes the number of medium (1, 2 or 3), as indicated in Fig. 2. For harmonic waves the time dependence for all considered waves is given by a factor $\exp{(i\omega t)}$, which is neglected in further considerations.

The problem is solved using potential functions for the particle velocity φ_i – for longitudinal waves and ψ_i – for transversal waves. The Cartesian coordinate z is normal to the coupling layer, the coordinate x lies in the plane of incidence, and the coordinate y is normal to the plane of incidence. For the geometrical configuration shown in Fig. 2 the explicit forms of potential functions in media 1, 2, and 3 are:

– in medium 1:
$$\begin{aligned} \varphi_1 &= B_1 e^{-i(\alpha_1 z - \sigma x)}, \\ \psi_1 &= D_1 e^{-i(\beta_1 z - \sigma x)}; \end{aligned} \tag{1}$$

– in medium 2:
$$\begin{aligned} \varphi_2 &= B_2 e^{-i(\alpha_2 z - \sigma x)} + A_2 e^{i(\alpha_2 z - \sigma x)}, \\ \psi_2 &= 0; \end{aligned} \tag{2}$$

– in medium 3:
$$\begin{aligned} \varphi_3 &= B_3 e^{-i(\alpha_3 z - \sigma x)} + A_3 e^{i(\alpha_3 z - \sigma x)}, \\ \psi_3 &= C_3 e^{i(\beta_3 z + \sigma x)}, \end{aligned} \tag{3}$$

where α_i are the z-coordinates of wave vectors of the longitudinal wave, β_i are the z-coordinates of the wave vectors of the transversal waves, A_i are the amplitudes of potential functions for L-type waves traveling in the positive z-direction, B_i are the amplitudes of potential functions for L-type waves traveling in the negative z-direction, C_i are the amplitudes of potential functions for T-type waves traveling in the positive z-direction, and D_i are the amplitudes of potential functions for T-type waves traveling in the negative z-direction.

The σ symbol indicates the x-coordinate of all the wave vectors. They have to be equal to meet the continuity conditions at the interfaces. Actually, this is a condition equivalent to the Snell law, and can be expressed as:

$$\begin{aligned} \sigma &= k_{l1} \sin\theta_1 = k_{t1} \sin\theta'_1 = k_{l2} \sin\theta_2 \\ &= k_{l3} \sin\theta_3 = k_{t3} \sin\theta'_3, \end{aligned} \tag{4}$$

where $k_{li} = \frac{\omega}{Cl_i}$ is the wave number of longitudinal wave in the i-th medium and $k_{ti} = \frac{\omega}{Ct_i}$ is the wave number of transversal wave in the i-th medium.

Consequently, the z-coordinates of the wave vectors in the three media are given by the formulas:

– in medium 1:
$$\alpha_1 = \sqrt{k_{l1}^2 - \sigma^2}, \qquad \beta_1 = \sqrt{k_{t1}^2 - \sigma^2}; \tag{5}$$

– in medium 2:
$$\alpha_2 = \sqrt{k_{l2}^2 - \sigma^2}; \tag{6}$$

– in medium 3:
$$\alpha_3 = \sqrt{k_{l3}^2 - \sigma^2}, \qquad \beta_3 = \sqrt{k_{t3}^2 - \sigma^2}. \tag{7}$$

It should be noted that z-coordinate of wave vector of longitudinal wave in medium 1 (tested material) can be a real value – for incidence angles below the 1st critical angle or an imaginary value – for incidence angles above the 1st critical angle. The z-coordinates of all other wave vectors are real, as we consider only practical cases, where the incidence angle is below the 2nd critical angle.

The particle velocities and stresses of all considered waves can be determined from potential functions by the following formula. For simplicity we neglect the media indexes i, since the form of these formulas is the same for all media:

$$v_x(z) = \frac{\partial\varphi}{\partial x} - \frac{\partial\psi}{\partial z}, \tag{8}$$

$$v_z(z) = \frac{\partial\varphi}{\partial z} + \frac{\partial\psi}{\partial x}, \tag{9}$$

$$T_{zz}(z) = \frac{i}{\omega}\left(\lambda\frac{\partial v_x}{\partial x} + \lambda\frac{\partial v_z}{\partial z} + 2\mu\frac{\partial v_z}{\partial z}\right), \tag{10}$$

$$T_{xz}(z) = \frac{i}{\omega}\left(\mu\frac{\partial v_x}{\partial z} + \mu\frac{\partial v_z}{\partial z}\right). \tag{11}$$

Substituting into these formulas potential functions given by Eqs. (1), (2), (3) and doing some ordering, the particle velocities and stresses in each medium can be expressed as the linear combinations of the amplitudes of relevant potential functions.

In medium 1:

$$v_x^1(z) = i\sigma e^{-i\alpha_1 z}e^{i\sigma x}B_1 + i\beta_1 e^{-i\beta_1 z}e^{i\sigma x}D_1, \tag{12}$$

$$v_z^{(1)}(z) = -i\alpha_1 e^{-i\alpha_1 z}e^{i\sigma x}B_1 + i\sigma e^{-i\beta_1 z}e^{i\sigma x}D_1, \tag{13}$$

$$T_{zz}^{(1)}(z) = -ie_1 e^{-i\alpha_1 z}e^{i\sigma x}B_1 + ig_1\beta_1 e^{-i\beta_1 z}e^{i\sigma x}D_1, \tag{14}$$

$$T_{xz}^{(1)}(z) = ig_1\alpha_1 e^{-i\alpha_1 z}e^{i\sigma x}B_1 + ie_1 e^{-i\beta_1 z}e^{i\sigma x}D_1, \tag{15}$$

where $e_1 = \left(\lambda_1 k_{l1}^2 + 2\mu_1\alpha_1^2\right)/\omega$ and $g_1 = 2\mu_1\sigma/\omega$.

In medium 2:

$$v_x^{(2)}(z) = i\sigma e^{i\alpha_2 z}e^{i\sigma x}A_2 + i\sigma e^{-i\alpha_2 z}e^{i\sigma x}B_2, \tag{16}$$

$$v_z^{(2)}(z) = i\alpha_2 e^{i\alpha_2 z}e^{i\sigma x}A_2 - i\alpha_2 e^{-i\alpha_2 z}e^{i\sigma x}B_2, \tag{17}$$

$$T_{zz}^{(2)}(z) = -ie_2 e^{i\alpha_2 z}e^{i\sigma x}A_2 - ie_2 e^{-i\alpha_2 z}e^{i\sigma x}B_2, \tag{18}$$

$$T_{xz}^{(2)}(z) = 0, \qquad (19)$$

where $e_2 = \lambda_2 k_{l2}^2 / \omega$ and $T_{xz}^{(1)}(z) = 0$ because medium 2 is a liquid and does not transfer shear stresses.

In medium 3:

$$\begin{aligned} v_x^{(3)}(z) &= i\sigma e^{i\alpha_3 z} e^{i\sigma x} A_3 + i\sigma e^{-i\alpha_3 z} e^{i\sigma x} B_3 \\ &\quad - i\beta_3 e^{i\beta_3 z} e^{i\sigma x} C_3, \end{aligned} \qquad (20)$$

$$\begin{aligned} v_z^{(3)}(z) &= i\alpha_3 e^{i\alpha_3 z} e^{i\sigma x} A_3 - i\alpha_3 e^{-i\alpha_3 z} e^{i\sigma x} B_3 \\ &\quad + i\sigma e^{i\beta_3 z} e^{i\sigma x} C_3, \end{aligned} \qquad (21)$$

$$\begin{aligned} T_{zz}^{(3)}(z) &= -ie_3 e^{i\alpha_3 z} e^{i\sigma x} A_3 - ie_3 e^{-i\alpha_3 z} e^{i\sigma x} B_3 \\ &\quad - ig_3 \beta_3 e^{i\beta_3 z} e^{i\sigma x} C_3, \end{aligned} \qquad (22)$$

$$\begin{aligned} T_{xz}^{(3)}(z) &= -ig_3 \alpha_3 e^{i\alpha_3 z} e^{i\sigma x} A_3 + ig_3 \alpha_3 e^{-i\alpha_3 z} e^{i\sigma x} B_3 \\ &\quad + ie_3 e^{i\beta_3 z} e^{i\sigma x} C_3, \end{aligned} \qquad (23)$$

where $e_3 = \left(\lambda_3 k_{l3}^2 + 2\mu_3 \alpha_3^2\right)/\omega$ and $g_3 = \frac{2\mu_3 \sigma}{\omega}$.

Now, we consider the boundary conditions for particle velocities and stresses at the borders $z = 0$ and $z = d$. Both borders are between solid and liquid medium and we neglect viscosity of the liquid medium – in our case water. It follows that there is no continuity of tangential displacements and particle velocities at the borders. Consequently the boundary conditions are:

– at the border $z = 0$:
  continuity of normal particle velocities

$$v_z^{(1)}(0) = v_z^{(2)}(0), \qquad (24)$$

  continuity of normal stresses

$$T_{zz}^{(1)}(0) = T_{zz}^{(2)}(0), \qquad (25)$$

  zeroing of tangential stresses in the solid media

$$T_{xz}^{(1)}(0) = 0; \qquad (26)$$

– at the border $z = d$:
  continuity of normal particle velocities

$$v_z^{(3)}(d) = v_z^{(2)}(d), \qquad (27)$$

  continuity of normal stresses

$$T_{zz}^{(3)}(d) = T_{zz}^{(2)}(d), \qquad (28)$$

  zeroing of tangential stresses in the solid media

$$T_{xz}^{(3)}(d) = 0. \qquad (29)$$

Substituting Eqs. (12)–(23) to the above boundary conditions there are obtained six complex equations, for six unknown amplitudes of potential functions: A_3, C_3, A_2, B_2, B_1, D_1. The amplitude B_3 is the amplitude

of incident longitudinal wave, so it is a known value which can be set to 1.

The obtained set of six linear equations can be conveniently expressed in a matrix form to facilitate further numerical processing:

$$\begin{bmatrix} \alpha_3 e^{i\alpha_3 d} & \sigma e^{i\beta_3 d} & -\alpha_2 e^{i\alpha_2 d} & \alpha_2 e^{-i\alpha_2 d} & 0 & 0 \\ -e_3 e^{i\alpha_3 d} & -g_3 \beta_3 e^{i\beta_3 d} & e_2 e^{i\alpha_2 d} & e_2 e^{-i\alpha_2 d} & 0 & 0 \\ -g_3 \alpha_3 e^{i\alpha_3 d} & e_3 e^{i\beta_3 d} & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_2 & -\alpha_2 & \alpha_1 & -\sigma \\ 0 & 0 & -e_2 & -e_2 & e_1 & -g_1 \beta_1 \\ 0 & 0 & 0 & 0 & g_1 \alpha_1 & e_1 \end{bmatrix} \begin{bmatrix} A_3 \\ C_3 \\ A_2 \\ B_2 \\ B_1 \\ D_1 \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_3 e^{-i\alpha_3 d} \\ e_3 e^{-i\alpha_3 d} \\ -g_3 \alpha_3 e^{-i\alpha_3 d} \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad (30)$$

or

$$[M]\,[X] = [Y]. \qquad (31)$$

By inverting this matrix equation, all unknown A_3, C_3, A_2, B_2, B_1, D_1 amplitudes of potential functions in media 1, 2 and 3 are obtained:

$$[X] = [M]^{-1}\,[Y], \qquad (32)$$

or

$$\begin{bmatrix} A_3 \\ C_3 \\ A_2 \\ B_2 \\ B_1 \\ D_1 \end{bmatrix} = [M]^{-1} \begin{bmatrix} \alpha_3 e^{-i\alpha_3 d} \\ e_3 e^{-i\alpha_3 d} \\ -g_3 \alpha_3 e^{-i\alpha_3 d} \\ 0 \\ 0 \\ 0 \end{bmatrix}. \qquad (33)$$

In this work the main subject of interest is the transmission coefficient of the ultrasonic wave from medium 3 (probe wedge) to medium 1 (tested material). For the standard angle beam shear wave probes, i.e., for cases when the incident angle is between the 1st and the 2nd critical angle, this coefficient is given by:

$$T_{3-1}^{L-T} = \frac{D_1}{B_3} = \frac{D_1}{1} = D_1. \qquad (34)$$

For angle beam longitudinal wave probes, i.e., for cases where the incident angle is below the 1-st critical angle, this coefficient is given by:

$$T_{3-1}^{L-L} = \frac{B_1}{B_3} = \frac{B_1}{1} = B_1. \qquad (35)$$

In non-destructive testing procedures, the ultrasonic pulse passes through the coupling layer twice, first

when it is introduced to the tested material, and second – on its way back, after reflection from the defect. So, from practical point of view, the most interesting is the double transmission coefficient, i.e., the product of transmission coefficients from medium 3 to 1 and from medium 1 to 3.

For shear wave probes it is given by:

$$T_{3-1-3}^{L-T-L} = T_{3-1}^{L-T} \cdot T_{1-3}^{T-L}. \tag{36}$$

For longitudinal wave probes it is given by:

$$T_{3-1-3}^{L-L-L} = T_{3-1}^{L-L} \cdot T_{1-3}^{L-L}. \tag{37}$$

The inverse transmission coefficients T_{1-3}^{T-L} and T_{1-3}^{L-L} could be determined from properly redefined models similar to the one given in Fig. 2. However, most easily they can be calculated from the direct transmission coefficients using the so-called Stokes' relations (SCHMERR JR., 2016). Applying these relations we obtain, respectively:

$$T_{1-3}^{T-L} = T_{3-1}^{L-T} \cdot \frac{Ct_1\rho_1\cos\theta_1'}{Cl_3\rho_3\cos\theta_3} \tag{38}$$

and

$$T_{1-3}^{L-L} = T_{3-1}^{L-L} \cdot \frac{Cl_1\rho_1\cos\theta_1}{Cl_3\rho_3\cos\theta_3}. \tag{39}$$

Finally, substituting (38) and (39) to (36) and (37), the sought formulas for double transmission coefficients, for both types of ultrasonic probes used in nondestructive testing are obtained. The formulas take into account the liquid coupling layer of thickness d, between the probe wedge and tested material:

$$\begin{aligned} T_{3-1-3}^{L-T-L} &= \left(T_{3-1}^{L-T}\right)^2 \cdot \frac{Ct_1\rho_1\cos\theta_1'}{Cl_3\rho_3\cos\theta_3} \\ &= D_1^2 \frac{Ct_1\rho_1\cos\theta_1'}{Cl_3\rho_3\cos\theta_3} \end{aligned} \tag{40}$$

and

$$\begin{aligned} T_{3-1-3}^{L-L-L} &= \left(T_{3-1}^{L-L}\right)^2 \cdot \frac{Cl_1\rho_1\cos\theta_1}{Cl_3\rho_3\cos\theta_3} \\ &= B_1^2 \frac{Cl_1\rho_1\cos\theta_1}{Cl_3\rho_3\cos\theta_3}. \end{aligned} \tag{41}$$

The aforementioned formulas, for double transmission coefficients, are derived for harmonic plane waves with a strictly defined frequency. In practice, the ultrasonic pulses generated by ultrasonic probes have a specified frequency spectrum, characterized by the so-called relative bandwidth parameter defined as:

$$\text{WB} = \frac{f_u - f_l}{f_0} \cdot 100\%, \tag{42}$$

where f_u – upper frequency of the −6 dB frequency spectrum, f_l – lower frequency of the −6 dB frequency spectrum, $f_0 = \frac{f_u - f_l}{2}$ – center frequency of the −6 dB frequency spectrum.

It means that the actual drop of amplitude of an ultrasonic pulse, traveling through the coupling layer from the probe to the tested material and vice versa, is a certain average of the double transmission coefficients for all frequencies represented in the pulse spectrum. To include this effect in the discussed model, a certain frequency spectrum is assumed for ultrasonic pulse incident on the boundary between the probe wedge and the coupling layer. Such an initial spectrum distribution can be reasonably approximated by the Gaussian function, given by:

$$G_i(f) = \exp\left(\frac{-(f-f_0)^2}{2\sigma^2}\right), \tag{43}$$

where f_0 is the center frequency of the ultrasonic probe. The parameter $\sigma = \text{WB}/235$ is defined so that the modeled spectrum has a bandwidth equal to the actual bandwidth parameter WB of the ultrasonic probe. The WB parameter is usually presented in the probe certificate or can be measured according to EN ISO 22232-2 (2020).

The signal waveform of the initial pulse in the time domain is given by the inverse Fourier transform of its spectrum:

$$h_i(t) = \int_{-\infty}^{\infty} G_i(f)e^{i2\pi ft}\,\mathrm{d}f. \tag{44}$$

After the double passage of the ultrasonic pulse through the coupling layer its spectrum is modified by the double transmission coefficient in the following way:

$$G_t(f) = G_i(f) \cdot T_{3-1-3}^{L-T-L}(f) \tag{45}_1$$

for the shear wave probes, and

$$G_t(f) = G_i(f) \cdot T_{3-1-3}^{L-L-L}(f) \tag{45}_2$$

for the longitudinal wave probes.

Knowing the spectrum of the double transmitted ultrasonic pulse, its signal waveform may be calculated using the inverse Fourier transform:

$$h_t(t) = \int_{-\infty}^{\infty} G_t(f)e^{i2\pi ft}\,\mathrm{d}f. \tag{46}$$

Then, the double transmission coefficient for the ultrasonic pulse in a time domain may be calculated according to the formula:

$$T_{3-1-3}^{L-T(L)-L}(f_0, \text{WB}) = \frac{\max\{|h_t(t)|\}}{\max\{|h_i(t)|\}}. \tag{47}$$

The designation $T^{T(L)}$ denotes here that the formula is valid for T^{L-T-L}, and for T^{L-L-L} configurations. The above definition of the double transmission coefficient in the time domain, corresponds to the experimental measurements of this value, where one compares the maximum amplitudes of the ultrasonic pulse

before and after double passage through the coupling layer. The arguments f_0 and WB, specified to the double transmission coefficient in the time domain, indicate that its value depends not only on the central frequency of the ultrasonic probe (as it is the case in the monochromatic model), but also on the probe bandwidth defined by its WB parameter.

2.2. Model implementation and validation

The described theoretical model was implemented in the prepared computer program TransmissionLoss 1.x of which the main purpose was to facilitate the design of ultrasonic probes for the new, high speed inspection wagon for the Polish Railways. In contrast to the known analytical solutions, the program allowed for calculation of double transmission coefficients and related transfer losses. It was possible not only for normal beam longitudinal wave probes, but also for angle beam shear wave probes and angle beam longitudinal wave probes, which are used in ultrasonic testing of railway rails. Moreover, the program took into account the finite bandwidth of modern ultrasonic probes, what considerably changes the dependence of transfer losses on the coupling layer thickness.

To use effectively the formulas quoted in Sec. 2 in the computer program they have to be discretized. In particular, the continuous initial pulse spectrum defined by Eq. (43) was replaced by a discretized spectrum given by dependency:

$$G_i\left(n\Delta f\right) = \exp\left(\frac{-\left(\frac{n-32}{32}\right)^2}{2\sigma^2}\right), \qquad (48)$$

where $\Delta f = f_0/32$ was specified as a step in the frequency domain and integer n was changing from 1 to 63, to embrace frequency spectrum from $\frac{1}{32}f_0$ to near $2f_0$. The bandwidth of majority of ultrasonic probes, used in non-destructive testing, lies between 30% and 80% of f_0, so their −6 dB spectrum is within the range from $0.6f_0$ to $1.4f_0$. It means that the assumed discretization range is sufficient for that application.

After calculation of the discretized initial spectrum, the program numerically inversed the matrix Eq. (30) for every discreet frequency $n\Delta f$ – within the probe spectrum – and calculates complex amplitudes $D_1\left(n\Delta f\right)$ and $B_1\left(n\Delta f\right)$. Based on these amplitudes, the double transmission coefficients $T_{3-1-3}^{L-T-L}\left(n\Delta f\right)$ and $T_{3-1-3}^{L-L-L}\left(n\Delta f\right)$ were calculated from Eqs. (40) and (41), for every discreet frequency in the probe spectrum.

Then the discreet inverse Fourier transforms were calculated for the initial pulse spectrum $G_i\left(n\Delta f\right)$ and for the pulse spectrum after double transmission through the coupling layer $G_t\left(n\Delta f\right) = T_{3-1-3}^{L-T-L}\left(n\Delta f\right)G_i\left(n\Delta f\right)$ using the FFT algorithm for

N = 1024 point. As a result, discreet time waveforms for initial pulse $h_i\left(k\Delta t\right)$ and after transmission pulse $h_t\left(k\Delta t\right)$ were obtained, where Δt is the time step in the time domain, and k is integer from 0 to N. The time domain step Δt is related to the Δf step and the number of points in the FFT transform by relation:

$$\Delta t = \frac{1}{N \cdot \Delta f} = \frac{1}{1024 \cdot \Delta f}. \qquad (49)$$

For example, for a typical ultrasonic probe with central frequency f_0 = 2 MHz and time period T_0 = 0.5 μs the frequency domain step $\Delta f = f_0/32 = 0.0625$ MHz, and the time domain step $\Delta t = 0.015625$, μs $= 1/32T_0$. It means that the discretization of waveform functions $h_i(t)$ and $h_t(t)$ can produce quantization error not greater than 0.5%, when estimating maxima of these functions from its discreet representations in Eq. (47). In non-destructive testing 0.5% error in evaluation of the ultrasonic signal amplitude is negligible – so the implemented calculation algorithm is sufficient for the intended application. On the other hand, it is fast enough to be executed on a typical personal computer.

In ultrasonic testing practice signal amplitudes relations were commonly expressed using a logarithmic scale, so the signal amplitude drop, caused by its double transmission through the coupling layer, can be conveniently expressed in decibels by the transfer losses (TL) defined as:

$$\text{TL} = -20\log_{10}\left(T_{3-1-3}^{L-T(L)-L}\right). \qquad (50)$$

The TL defined in Eq. (50) include the signal amplitude drops caused not only by the presence of the coupling layer, but also by the impedance mismatch between the material of ultrasonic probe wedge and the material tested. In many practical applications, such as ultrasonic testing of railway rails, the impedance mismatch is the same throughout the entire inspection process, and the only changing factor is the thickness of the coupling layer between the probe and the tested material. To focus attention on this dependency a more suitable parameter called coupling losses CL can be defined:

$$\text{CL}(d) = \text{TL}(d) - \text{TL}(0), \qquad (51)$$

where d is the thickness of the coupling layer and TL(0) are the transfer losses, calculated for the thickness d set to 0.

Due to such a definition, the coupling losses were zero for the best case scenario (the zero coupling layer) and allow for analyzing changes in the transfer losses, due to fluctuations of the coupling layer thickness. This is an important aspect of every mechanized ultrasonic inspection, as the changes in the transfer losses – due to fluctuations of the coupling layer thickness – cause uncontrolled changes in testing sensitivity during ultrasonic scanning. Knowing the characteristic of sensitivity changes as a function of coupling layer thickness, some scanning gain corrections can be applied to

compensate for the predicted sensitivity drops. This way, the coupling layer thickness fluctuations during actual examination can only increase testing sensitivity, which is a more secure situation than uncontrolled sensitivity drops.

The developed model was checked against the known analytic solution for monochromatic longitudinal plane wave incident on the coupling layer at an incident angle $\theta_3 = 0°$ (see Fig. 2). The analytic formula for the transmission coefficient through the layer was taken from work (OBRAZ, 1983) and rewritten using the notations defined in this work:

$$T_{3-1}^{L-L} = 2\left[\left(1 + \frac{\rho_3 Cl_3}{\rho_1 Cl_1}\right)^2 \left(\cos\frac{2\pi\,\mathrm{d}f}{Cl_2}\right)^2\right.$$
$$\left. + \left(\frac{\rho_2 Cl_2}{\rho_1 Cl_1} + \frac{\rho_3 Cl_3}{\rho_2 Cl_2}\right)^2 \left(\sin\frac{2\pi\,\mathrm{d}f}{Cl_2}\right)^2\right]^{-\frac{1}{2}}. \quad (52)$$

The reverse transmission coefficient T_{1-3}^{L-L} can be calculated from Eq. (52) by interchanging indexes 1 and 3. Then the double transmission coefficient from medium 3 to 1 and back can be calculated as $T_{3-1}^{L-L} \cdot T_{1-3}^{L-L}$ in the same way as in Eq. (37) of the model developed in this work.

The example calculations executed using the above analytic formula and our more general numerical model, in which we assumed the monochromatic wave and incident angle $\theta_3 = 0°$, were shown in Table 1. The calculations were performed for typical conditions encountered in a railway rail inspection – i.e., for 2 MHz L-type probe with the PMM wedge and assuming water coupling layer changes from 0.0 to 0.5 mm.

Table 1. Comparison of coefficient T_{3-1-3}^{L-L-L}, calculated for 2 MHz longitudinal monochromatic wave and incident angle $\theta_3 = 0°$, applying the proposed model (Eq. (41)) and by the analytic Eq. (52).

d [mm]	T_{3-1-3}^{L-L-L}	
	model	analytic formula
0.00	0.242355	0.242355
0.05	0.156465	0.156465
0.10	0.085844	0.085844
0.15	0.061231	0.061231
0.20	0.057934	0.057934
0.25	0.072430	0.072430
0.30	0.120302	0.120302
0.35	0.221854	0.221854
0.40	0.201053	0.201053
0.45	0.106440	0.106440
0.50	0.067703	0.067703

The aforementioned calculations were performed using double precision arithmetic. As can be seen from Table 1, the results of the presented numerical model and the known analytic solution match up to six decimal places. It could be expected, as both approaches

assume the same problem geometry and boundary conditions. It is a prove however, that the rather complicated model implementation, based on the numerical solution of a complex 6×6 equation set, does not contain any errors in the program algorithms.

3. Experimental verification of modeling results

The preliminary experimental verification of the developed model was performed using the commercial ultrasonic probe Panametrics A106 integrated with a delay line (probe wedge) made of PMMA. The probe was of the longitudinal wave and of the straight $(\theta_3 = 0°)$ beam type, with the nominal frequency 2.25 MHz and the transducer diameter of 12.5 mm. The spectral characteristic of the probe with the PMMA delay line was determined using the laboratory system comprising Panametrics Epoch 650 Pulser/Receiver/Digitizer, Calibration Block No 1 (according to ISO 2400), and proprietary software implementing the Fourier transform using the FFT algorithm.

The details of the experimental setup are presented in Fig. 4. The central frequency f_0 and –6 dB bandwidth BW were determined according to EN 12668-2:2010 applying the measured transmitting-receiving spectral characteristic. The lower band frequency was $f_l = 1.17$ MHz, and the upper $f_u = 2.58$ MHz. Applying the abovementioned data, the central frequency and relative bandwidth of the probe coupled with the delay line, were calculated using the following standard formulas:

$$f_0 = \frac{(f_l + f_u)}{2} = \frac{1.17 + 2.58}{2} = 1.88\text{ MHz}, \quad (53)$$

$$\text{BW} = \frac{2(f_u - f_l)}{(f_u + f_l)}100\% = \frac{1.41}{1.88}100 = 75\%. \quad (54)$$

The parameters BW and f_0 were introduced to the presented model, that takes into account the bandwidth of the probe. It was performed applying Eq. (43) to calculate CL introduced by a thin water layer placed between the delay line of the probe and the steel block representing the tested material. The conceptual scheme of this experimental setup is shown in Fig. 3.

According to this diagram, the piezoelectric transducer of the ultrasonic probe transmits to the PMMA delay line a short pulse of longitudinal wave, which travels to the coupling layer of a thickness d on the border between the delay line and the steel block. Part of the ultrasonic energy reflects on the border layer, creating on the screen of the ultrasonic receiver the first ultrasonic echo (so called interface echo with amplitude $A0$), which is normally of no importance for ultrasonic testing. The other part of the ultrasonic energy passes through the coupling layer to the steel block and reflects from its bottom giving a so-called back wall echo of amplitude $A1$. The amplitude of this echo
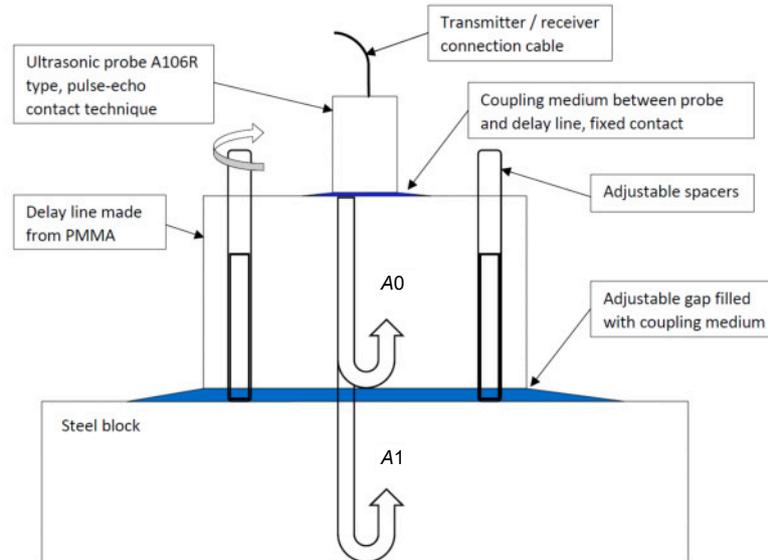
Fig. 3. Scheme of the experimental setup used for verification of theoretical models for calculation of CL, introduced by water coupling layer between PMMA delay line and the steel block (look for a description in the text).

depends on the double transmission coefficient through the coupling layer, but also on all other factors affecting the pulse on the way from the sending transducer to the bottom surface of the steel block, and after reflection on the way back to the receiving transducer:

$$A1 = A0\, S_{a3} S_{d3} S_{a1} S_{d1} R_1\, T_{3-1-3}^{L-L-L}, \qquad (55)$$

where $A0$ – the initial amplitude of the ultrasonic pulse near the sending transducer, T_{3-1-3}^{L-L-L} – the two way transmission coefficient through the coupling layer, S_{a3} – attenuation of the signal amplitude in the PMMA delay line (medium 3), S_{d3} – diffraction divergence losses in the PMMA delay line, S_{a1} – attenuation of the signal amplitude in the steel block (medium 1), S_{d1} – diffraction divergence losses in the steel block, R_1 – reflection coefficient at the bottom surface of the steel block.

The double transmission coefficient, and in consequence the amplitude of the back wall echo $A1$, depends on the thickness of the coupling layer d. For practical applications, this dependence is most conveniently expressed in terms of CL, defined by Eqs. (51) and (50). In the presented experiment, coupling losses are determined by the ratio of back wall echo amplitudes, obtained for the coupling layer of a thickness d to the back wall echo amplitude for the coupling layer thickness equal to zero. It is expressed in decibels by the formula:

$$\mathrm{CL}(d) = -20 \log_{10}\left(\frac{A1(d)}{A1(0)}\right). \qquad (56)$$

It should be noted, that in an experimental approach described here, it can be determined the coupling losses without considering the ultrasonic attenuation and beam divergence in the PMMA delay line and the steel block as well as considering the reflection coefficient of the ultrasonic pulse at the bottom of the steel block. All these factors obviously affect the measured amplitudes of back wall echoes but are independent of the thicknesses of the coupling layer and are reduced in the fractional expression of Eq. (56).

The photograph of the experimental setup, used for determination of CL, depending on the coupling layer thickness d, is shown in Fig. 4. The distance d between the PMMA delay line and the steel block was adjusted using three fine-pitch screws. This distance was controlled using a dial micrometer with accuracy of ± 0.001 mm. During measurements the steel block and a lower part of the delay line were immersed in
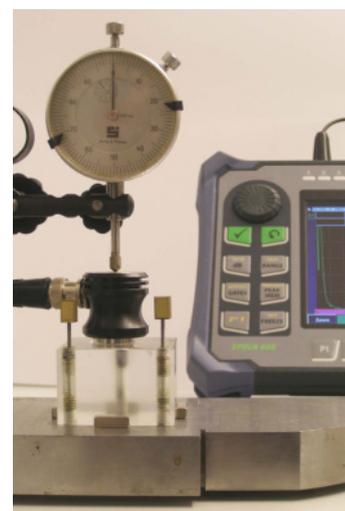
Fig. 4. Photograph of the experimental setup to determine CL, introduced by water coupling layer existing between a PMMA delay line and a steel block.

water to ensure complete filling with water of the gap between these objects.

The first step of the performed procedure was to determine the back wall echo amplitude for the coupling layer of zero thickness – $A1(0)$. During this step, the distancing screws were released and the PMMA plate was slightly pressed against the steel block to remove any remaining water between them. Within the next steps the distance between the coupled objects, representing the thickness of the water coupling layer d, was steadily increased, and back wall echo amplitudes $A1(d)$ were measured. Finally, the measured echo amplitudes were substituted to Eq. (56) and CL in dependence of d, were determined. The measurements were performed for a water layer thickness in the range from 0.0 to 0.8 mm, which embrace the coupling layer thickness fluctuations, that can be reasonably expected in actual rail testing.

The results of experimental measurements of $CL(d)$ are presented in Fig. 5 with discreet points (squares). The range of measurement errors was determined based on the spread of twelve measurements for each point. The continuous line presents a theoretical curve calculated for the considered setup using the wideband model, and the dashed line depicts the theoretical curve calculated from the basic monochromatic model. It can be seen that in an initial layer thickness range (up to ca. 0.15 mm), the agreement between experimental results and both theoretical curves is very good. Then the experimental results and model predictions start to diverge. The curve calculated from the monochromatic model presents much larger deviations from the experimental data than the curve calculated applying the wideband model. This is especially visible for the layer thicknesses above 0.2 mm, where the more precise model still gives reasonable approximation of experimental data, while the basic monochromatic model completely fails showing nonexistent minima and maxima of coupling losses. In general, the wideband model predicts slightly higher values of coupling losses then experimental data, however, the difference does not exceed 1.0 dB at any measurement point. This can be assessed as a quite sufficient
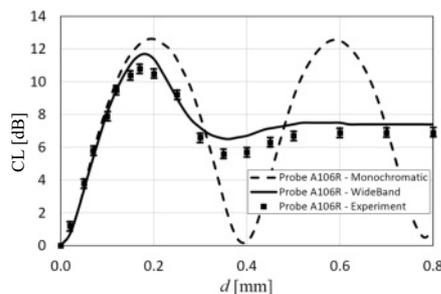

Fig. 5. Comparison of experimentally determined CL introduced by water coupling layer with theoretical calculations using the proposed wideband model and a commonly known monochromatic model.

modeling accuracy from the point of view of practical applications in automated ultrasonic examinations.

4. Exploration of model implications for railway rail testing

One of the most important factors, affecting the reliability of ultrasonic inspections of railway rails, is dependence of testing sensitivity on the random changes of coupling between scanning probes and the rail surface. Fluctuations of coupling layer thickness during high speed ultrasonic scanning are unavoidable due to waviness of the rail surface and spring suspension of ultrasonic probes.

According to widely known monochromatic Eq. (52) dependence of CL on the water layer thickness d is periodic – as shown in Fig. 6 – with a black color curve. The results were calculated for the probe of a longitudinal wave and of $\theta_3 = 0°$ beam type. In this approximation the minima of the coupling losses occur periodically at $d = n\frac{\lambda_2}{2}$ that is, at multiples of half the wavelength in the coupling layer. It is rather a theoretical result not observed in the ultrasonic testing practice.
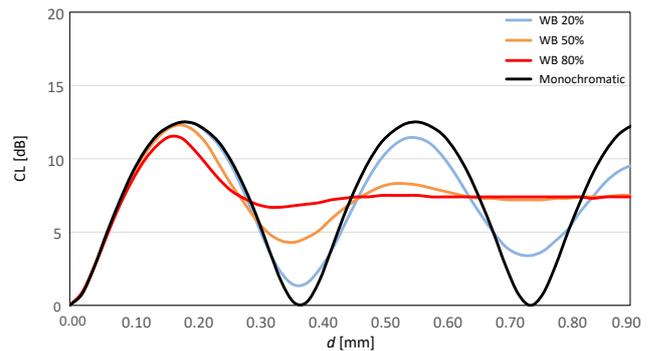
Fig. 6. Dependences of CL versus the water layer thickness, calculated for 2 MHz L-type probes, with different relative bandwidths WB.

To investigate this problem in more detail, there was calculated, applying the wideband model, coupling losses characteristics for the same 2 MHz frequency, but assuming three typical bandwidths of commercial ultrasonic probes, ranging from 20 to 80%. The calculation results are shown in Fig. 6 with colored curves. For the typical narrow band probe (WB = 20%) the characteristic is similar to the monochromatic case, but successive minima are getting shallower and coupling losses reach zero value only for the layer thickness equal to zero. For most common medium band probes (WB = 50%), there is only one additional minimum at $d = \frac{\lambda_2}{2}$ but it is much shallower than the minimum observed at the zero thickness layer. For wide band probes (WB = 80%), there is no additional minima and only one maximum of coupling losses observed at $d = \frac{\lambda_2}{4}$. After this maximum, the characteristic flattens out and shows no significant changes in coupling losses.

The presented results confirm the statement, that the bandwidth of ultrasonic probes has a significant effect on the coupling losses during ultrasonic scanning, and should be taken into account when designing and during calibration of ultrasonic inspection systems. For a typical 2 MHz *L*-type probe with medium bandwidth (parameter WB = 50%), the maximum coupling losses of 12.3 dB occur at the coupling layer thickness of 0.18 mm, and falls rapidly to zero with reduction of the layer thickness. It corresponds to 12.3 dB fluctuations in testing sensitivity between different sections of the tested rail, regarding the worst case scenario. To avoid such big changes in testing sensitivity, the minimum thickness of the coupling layer should be limited to ca. 0.1 mm, by fixing a distancing pins made of the hard material in the probe scanning surface. In this way, the fluctuations of testing sensitivity could be reduced from 12.3 to about 8 dB.

In addition to the analysis of the operation of 2 MHz *L*-type probes, it is interesting to determine how the coupling layer thickness affects the transfer losses determined for angle beam shear wave probes, commonly used in railway rail inspections.

As can be seen from Fig. 7, coupling losses characteristics are almost the same for *L*-type probe and *T*-type probes with refraction angles of 45° and 70°, assuming they have the same central frequency and bandwidth. However, if the central frequency of the probe would change, the coupling losses characteristic also change considerably, as can be seen in Fig. 8. Increase of the probe frequency from 2 to 4 MHz contracts the characteristic curve by a factor 2 on the layer thickness axis. It means that for 4 MHz probes the minimum separation between the probe faces and the rail surface could be reduced to 0.05 mm, without negative consequences for testing sensitivity fluctuations. The reduced fluctuations of CL of the order of 8 dB are rather high and still can cause problems during rail inspections.
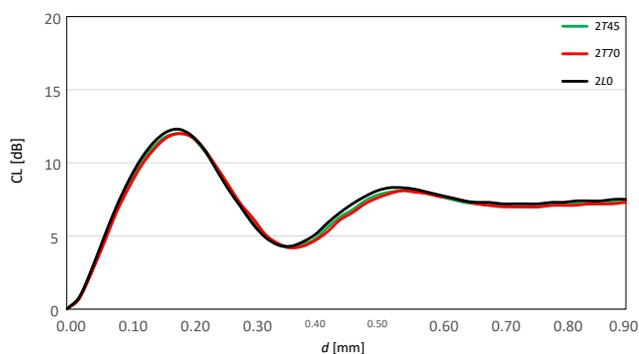

Fig. 7. Dependences of coupling losses on the coupling layer thickness for 2 MHz *L*-type probe and *T*-type probes with the same bandwidth (WB = 50%) but with different refraction angles; 2*T*45 – probe with transversal wave and refraction angle $\theta_3 = 45°$; 2*T*70 – probe with transversal wave and refraction angle $\theta_3 = 70°$; 2*L*0 – probe with longitudinal wave and refraction angle $\theta_3 = 0°$.
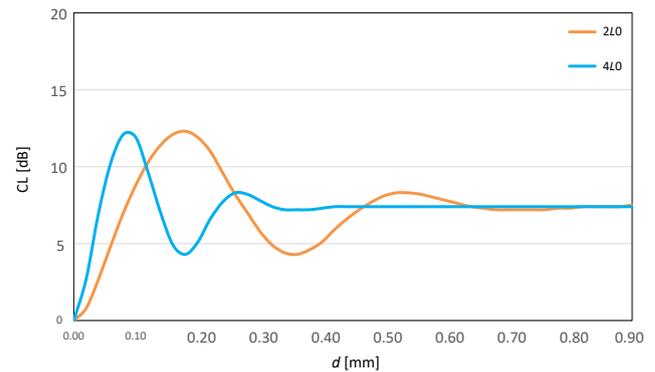

Fig. 8. Dependences of coupling losses on the coupling layer thickness for 2 MHz *L*-type probe and 4 MHz *L*-type probe with the same bandwidth (WB = 50%); 2*L*0 – 2 MHz probe with longitudinal wave and refraction angle $\theta_3 = 0°$; 4*L*0 – 4 MHz probe with longitudinal wave and refraction angle $\theta_3 = 0°$.

Using the presented model, one can search for other methods to reduce fluctuations of coupling losses. One of the possibilities is replacement of water with a coupling medium of higher acoustic impedance, e.g., glycerin. Unfortunately, this solution is impractical for a high speed ultrasonic inspection of railway rails due to environmental and technical constraints. Instead, one can investigate another solution consisting in replacing the conventional probe wedges made of PMMA with probe wedges made of Rexolite, that is a relatively new material which already entered ultrasonic applications. The Rexolite has lower acoustic impedance than PMMA (2.4 versus 3.2 Rayls) which is closer to water used as a coupling medium.

The coupling losses characteristic for 2 MHz probes with wedges made of PMMA and Rexolite are shown in Fig. 9. The maximum of the coupling losses for the Rexolite wedge is about 4.5 dB lower than the maximum for the PMMA wedge. After restriction of the minimal coupling layer to 0.10 mm, the amplitude of the testing sensitivity fluctuations will be reduced from 8 dB for PMMA wedges to 5 dB for Rexolite wedges.
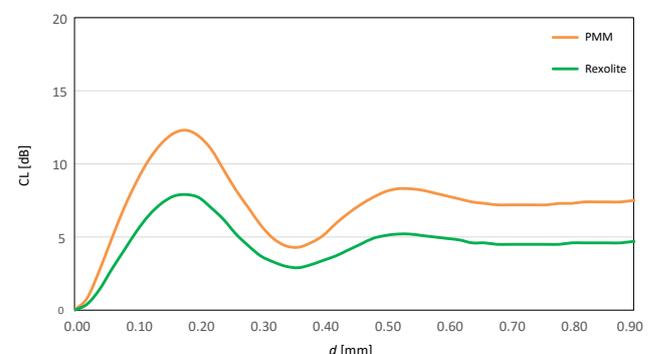

Fig. 9. Dependences of coupling losses on the coupling layer thickness for 2 MHz, WB = 50% probes with wedges made of PMMA and Rexolite.

The latter is a reasonably low value which may be compensated by scanning gain correction.

5. Conclusions

Losses determined for the case of beam transmission through the coupling layer between the ultrasonic probe emitting longitudinal or transverse waves and tested object, are presented in the article. As a standard, formulas for double transmission coefficients have so far been derived for harmonic waves of a strictly defined frequency. However, in practice, the pulses generated by modern ultrasonic probes used in non-destructive testing have a relatively wide frequency bandwidth. This means that the actual decrease in the amplitude of the ultrasonic pulse passing through the coupling layer from the probe to the tested material and vice versa is a certain average of the double transmission coefficients for all frequencies represented in the pulse spectrum. The numerical procedure presented in the paper takes into account the finite bandwidth of modern ultrasonic probes, which significantly changes the dependence of coupling losses on the thickness of the coupling layer. Contrary to the known analytical solutions, the model and program presented in the work allowed for precise calculation of coupling losses not only for normal beam probes producing longitudinal waves, but also for angle beam probes producing transversal waves, which are commonly used in non-destructive testing. Therefore, the developed model can be a significant improvement to the testing methodology for high speed ultrasonic inspection of railway rails, which can also be applied to other kinds of ultrasonic inspections, where fluctuation of the coupling layer thickness is an issue.

References

1. Barnard G.R., Bardin J.L., Whiteley J.W. (1975), Acoustic reflection and transmission characteristics for thin plates, *The Journal of the Acoustical Society of America*, **57**(3): 577–584, doi: 10.1121/1.380486.

2. Bray D.E. (2000), Historical review of technology development in NDE, [in:] *Proceedings of the 15th World Conference on NDT*, https://www.ndt.net/search/docs.php3?id=1212 (access: 28.07.2023).

3. Brekhovskikh L.M. (1980), *Waves in Layered Media*, 2nd ed., Academic Press, New York.

4. EN 16729-3 (2018), *Railway applications – Infrastructure – Non-destructive testing on rails in track – Part 3: Requirements for identifying internal and surface rail defects*, European Standards, https://www.en-standard.eu/csn-en-16729-3-railway-applications-infrastructure-non-destructive-testing-on-rails-in-track-part-3-requirements-for-identifying-internal-and-surface-rail-defects/ (access: 28.07.2023).

5. EN 17397-1 (2021), *Railway applications – Rail defects – Part 1: Rail defect management*, European Standards, https://www.en-standard.eu/din-en-17397-1-railway-applications-rail-defects-part-1-rail-defect-management/ (access: 28.07.2023).

6. EN ISO 22232-2 (2020), *Non-destructive testing – Characterization and verification of ultrasonic test equipment – Part 2: Probes*, European Standards, https://standards.iteh.ai/catalog/standards/cen/c360d5ef-82bb-46cb-8868-c046f44b1c9d/en-iso-22232-2-2020 (access: 28.07.2023).

7. Federal Railroad Administration (2015), *Track Inspector Rail Defect Reference Manual*, https://railroads.dot.gov/elibrary/track-inspector-rail-defect-reference-manual (access: 28.07.2023).

8. Folds D.L., Loggins C.D. (1977), Transmission and reflection of ultrasonic waves in layered media, *The Journal of the Acoustical Society of America*, **62**(5): 1102–1109, doi: 10.1121/1.381643.

9. Heckel T., Casperson R., Ruthe S., Mook G. (2018), Signal processing for non-destructive testing of railway tracks, [in:] *Proceedings of 44th Annual Review of Progress in Quantitative Nondestructive Evaluation*, **1949**(1): 030005, doi: 10.1063/1.5031528.

10. Heckel T., Thomas H.M., Kreutzbruck M., Rühe S. (2009), High speed non-destructive rail testing with advanced ultrasound and Eddy-current testing techniques, [in:] *Proceedings of the National Seminar & Exhibition on Non-Destructive Evaluation*, pp. 261–265.

11. Heckel T., Wack Y., Mook G. (2019), Simulation of an instrumented ultrasonic test run with a rail inspection train, [in:] *Review of Progress in Quantitative Nondestructive Evaluation*, https://www.iastatedigitalpress.com/qnde/article/id/8743/ (access: 28.11.2023).

12. International Union of Railways [UIC] (2022), *UIC Safety Report 2022*, https://safetydb.uic.org/IMG/pdf/uic_safety_report_2022.pdf (access: 28.07.2023).

13. Krautkrämer J., Krautkrämer H. (1990), *Ultrasonic Testing of Materials*, 4th ed., Springer Berlin, Heidelberg.

14. Obraz J. (1983), *Ultrasound in Measurement Technique* [in Polish: *Ultradźwięki w Technice Pomiarowej*], Wydawnictwa Naukowo-Techniczne, Warszawa.

15. Papaelias M.P., Roberts C., Davis C.L. (2008), A review on non-destructive evaluation of rails: State-of-the-art and future development, *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, **222**(4): 367–384, doi: 10.1243/09544097JRRT209.

16. Schmerr Jr. L.W. (2016), *Fundamentals of Ultrasonic Nondestructive Evaluation. A Modelling Approach*, 2nd ed., Springer Cham.

17. Zulian D. (2022), Effect of ultrasonic coupling media and surface roughness on contact transfer loss, *Cogent Engineering*, **9**(1): 2009092, doi: 10.1080/23311916.2021.2009092.

Research Paper

Improving Sound Insulation in Low Frequencies by a Three-Component Cladding Acoustic Metamaterial Panel

Na TAN, Xiaowei YAN, Bingfei LIU*

Aeronautical Engineering College, Civil Aviation University of China
Tianjin, China

*Corresponding Author e-mail: bingfeiliu2@126.com

In this paper, a three-component cladding acoustic metamaterial panel with good sound insulation effect in the low-frequency range is proposed. The sound transmission loss of metamaterial panels under different structural configurations and different material parameters is investigated by combining finite element simulation calculations with experimental research. The results show that the closer the center of gravity of the scatterer is to the substrate, the better the stability of the resonance unit, the wider the range of effective sound isolation frequencies, and the higher the degree of normalization. The filling rate of the scatterer is maintained at about 0.5 to obtain a better sound insulation effect. At the same time, choosing lower density materials for the substrate and metal materials with high density and high modulus of elasticity for the scatterer can maximally widen the bandgap and allows for low-frequency sound insulation below 600 Hz. This approach improves the low-frequency sound insulation efficiency of acoustic metamaterials. The results provide important explanations and references for a deeper understanding of the sound insulation mechanism and the effects of different parameters on sound insulation.

Keywords: acoustics; acoustic metamaterial panels; sound insulation properties; local resonance; low frequency sound insulation.

1. Introduction

The aerospace and automotive industries use acoustic damping and noise reduction devices to reduce low-frequency noise, a subject considered in many studies (JIANG *et al.*, 2021; CHEN *et al.*, 2021; IANNAC *et al.*, 2021; ATMOJO *et al.*, 2021). In these industries, beyond simply enhancing passenger comfort, the advantages also help to minimize errors in precision instruments caused by vibration (ZUO *et al.*, 2016; SONG, 2015).

The composite columnar local resonant unit is a valuable structural form for engineering purposes (ZHAO *et al.*, 2015; LI *et al.*, 2016). In the studies of WEN *et al.* (2005) and (2008), acoustic crystal plates were constructed with a periodic array of columnar local resonant units attached to it. These studies demonstrated the presence of a local resonant bandgap, effectively suppressing the vibration transmission of the acoustic crystal plate. PENNEC *et al.* (2008) designed

and calculated a crystal consisting of a periodic array of cylindrical dots deposited on a thin layer of uniform material, with the number of bandgaps increasing with the height of the cylinders.

OUDICH *et al.* (2010; 2011) created an acoustic metamaterial plate using two types of local resonant units: a single-layer rubber column and a composite rubber column. They analyzed how the unit parameters affected the bandgap characteristics. Additionally, they investigated an acoustic metamaterial plate with regularly attached single-layer rubber columns and confirmed the presence of local resonance bandgaps in this structure. HSU (2011) designed an acoustic metamaterial plate with a stepped local resonant unit and a two-dimensional phononic crystal composed of a stepped resonator array. BADREDDINE ASSOUAR and OUDICH (2012), and BADREDDINE *et al.* (2012) conducted a study on the impact of a double-sided arrangement of columnar resonant units on a thin plate. Their findings suggest that this arrangement effectively

broadens the damping frequency band. Also, they designed two-dimensional acoustic crystals utilizing short cut-off plates on both sides.

Hsu *et al.* (2013) explored the band gap and waveguide properties of cylindrical acoustic metamaterial plates with a stepped structure. Similarly, Yu *et al.* (2013) put forward the idea of two-dimensional phonon crystals composed of stepped resonator arrays. Zhao *et al.* (2015) examined the vibration isolation properties of a raised phonon crystal sheet in three different structural forms: single-sided single oscillator, single-sided double oscillator, and double-sided single oscillator. Meanwhile, Li *et al.* (2015) designed two-dimensional binary local resonant phonon plates. He and Wen (2018) studied the sound insulation properties of acoustic metamaterial plates that contained columnar local resonance units. They also explored how the sound insulation properties were affected by different lattice constants.

Zhou *et al.* (2020) proposed the use of multilayer rubber cylinders and metal cylinders attached to a thin plate. By combining units with different geometrical parameters, the resulting metamaterial demonstrated improved low-frequency sound insulation efficiency. In their respective studies, Yang *et al.* (2020) and Zhou *et al.* (2021) proposed novel solutions for low-frequency vibration and noise reduction problems in engineering. Yang *et al.* (2020) introduced a two-dimensional conical scatterer phonon crystal plate, while Zhou *et al.* (2021) designed a double-sided composite resonator structure and demonstrated a hybrid phonon crystal plate through simulations and experiments. Nakayama *et al.* (2021) presented a practical design for sheet acoustic metamaterials that could be potentially used in industrial applications. Their aim was to develop lightweight and compact materials that could effectively insulate against noise generation effects.

Some properties of phononic crystals hold good application prospects, such as exploring acoustic focusing properties and defect state properties of phononic crystals to guide the design of acoustic functional components. Qiu and Liu (2006) used phonon crystals to obtain a directional sound source. Liang *et al.* (2009) developed an acoustic diode model utilizing phonon crystals. Building upon this work, Maldovan (2013) further investigated the practical applications of phonon crystals, demonstrating the theoretical and experimental feasibility of using them for acoustic diodes and cloaks. At the same time, programmable smart metamaterials have also emerged (Yin *et al.*, 2022). The use of phononic crystals in designing of automotive, marine, and aircraft structures has also gained traction as phonon crystal research continues to advance (Ma *et al.*, 2018; Zhang *et al.*, 2016a).

Numerous studies highlight the attention given by scholars to the noise control problems in plate structures, resulting in important achievements. However, the law governing low-frequency sound insulation of cladding acoustic metamaterials (especially below 600 Hz) remains underexplored. However, the influence of low-frequency noise cannot be ignored in practical production and engineering applications. Therefore, the three-component cladding acoustic metamaterial plate structure was selected as the research object in this paper, aiming to explore the law of low-frequency wideband sound insulation by changing the structural and material parameters of the metamaterial plate. The paper is organized as follows: Sec. 1 provides the introduction, Sec. 2 presents the models and research methods, and Sec. 3 shows the results and discussion. Finally, conclusions are given in Sec. 4.

2. Models and methods

2.1. Experimental methods

The material chosen for the cladding in the experiments is the methyl vinyl polymer (VMQ) silicone rubber, the substrate material is epoxy resin, and the scatterer material is aluminum. The relevant properties of the materials are given in Table 1. The connection between the epoxy resin plate and the rubber, as well as that between the rubber and the cladding, were realized by superglue with a tight fit. In order to facilitate experimental tests with the ZK1030 circular impedance tube, the plate-type acoustic metamaterial was designed as a circular sample (Figs. 1a and 1b) during the preparation of the acoustic isolation test samples. The circular sample is 100 mm in diameter and contains seven hexagonal acoustic metamaterial cells. The thickness of the substrate is $e = 1$ mm, the single-cell lattice constant $a = 30$ mm, and the radius and height of the cladding layer remain unchanged at $R = 6$ mm and $H = 9$ mm, respectively. For comparative experiments, the radius of the scatterer is selected to be $r_1 = 4$ mm and $r_2 = 5$ mm. The corresponding heights of the scatterer are $h_1 = 4.75$ mm and $h_2 = 4.32$ mm, respectively.

Table 1. Material constants of the components.

Materials	Mass density [kg/m^3]	Young's modulus [Pa]	Poisson's ratio
Tungsten	19 100	3.54×10^{11}	0.35
Lead	11 600	4.08×10^{10}	0.42
Copper	8960	1.1×10^{11}	0.35
Steel	7780	2.1×10^{11}	0.30
Aluminum	2700	7×10^{11}	0.33
Epoxy resin	1180	4.35×10^{9}	0.38
Silicone rubber	1300	1.37×10^{5}	0.47

This experiment uses the ZK1030 impedance tube test system, which includes the B&K-23750 power amplifier, Type-3160-A-042 data acquisition instrument,

a)

b)

c)
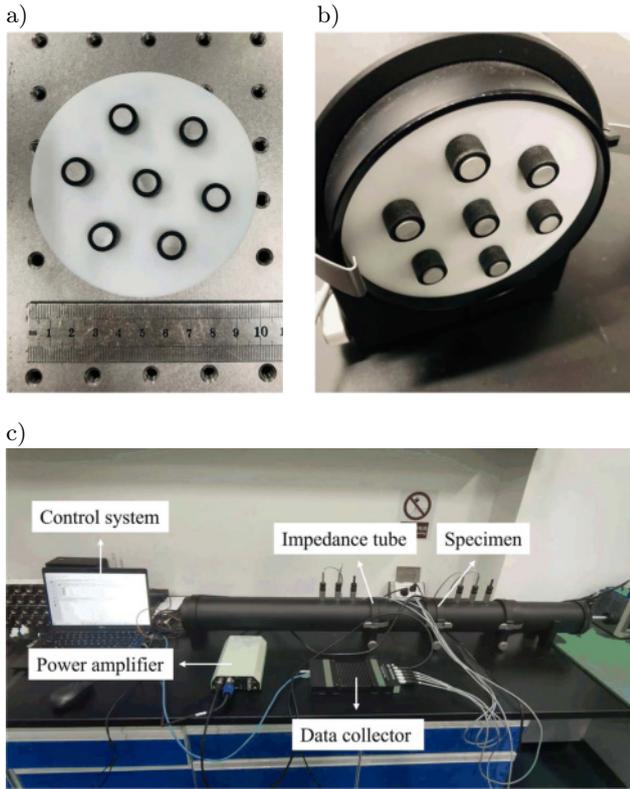
Fig. 1. Circular acoustic metamaterial samples (a); placement of the specimens in the experiment (b); sound insulation experimental test systems (c).

computer (installed with B&K acoustic test software) and impedance tube and other equipment (Fig. 1c). The impedance tube is 100 mm in diameter and consists of a large standing wave tube (including a loudspeaker and sample holder), a sample tube (for 0–100 mm wide samples), and a receiver tube. The test procedure is as follows: firstly, the relevant components of the impedance tube test system are installed and fixed in accordance with the requirements, ensuring that the whole impedance tube is on a horizontal line. Then, the B&K Acoustic Test Software is started and opened, and the relevant settings for the sound insulation items are set up. After no warnings have been given for the signal-noise ratio measurement, the sample tube is opened and the sample is placed into it, keeping the sample in a vertical plane as far as possible. Then, the sound insulation test is conducted and the sound insulation of the specimen can be obtained by substituting the sound pressure value into the transfer function.

2.2. Finite element simulation methods

In engineering practice, the sound insulation performance of structures is generally evaluated by sound transmission loss (STL). Noise reduction materials utilize various components, structures, or systems to hinder the spread of sound and diminish its energy once

it has passed through the material. To elucidate the sound insulation properties, the STL of the plate-type acoustic metamaterials is calculated.

In the x–z plane, there is an infinite homogeneous thin plate with thickness h, the top and bottom of which are in contact with the air, the speed of sound in the air is c, and the density of the air is ρ_0 (Fig. 2). There is a simple harmonic plane wave P_{inc} incident from the z side, with acoustic wave amplitude P_0, incident angle θ, ω is the incident angular frequency, and the wave vector of the incident wave $k_0 = \omega/c$. By decomposing k_0 into two-dimensional coordinates, the projected components in the x and z squares are, respectively, k_x and k_z. According to Eq. (1), it can be observed that the plate vibrates under acoustic excitation, and the surface couples with the air to emit transmitted acoustic wave P_{tr} and reflected acoustic wave outward P_{ref}:

$$
\begin{aligned}
k_x &= k_0 \sin \theta, \\
k_z &= k_0 \cos \theta.
\end{aligned}
\tag{1}
$$
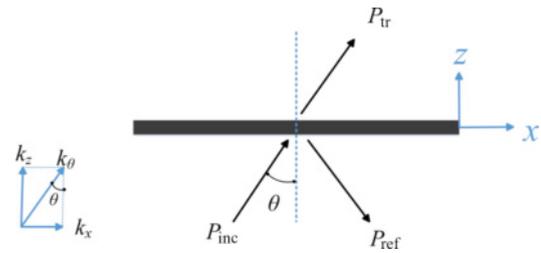
Fig. 2. Infinite homogeneous thin plate in the x–z plane.

According to the small amplitude one-dimensional plane wave equation P_{inc} can be assumed as:

$$
P_{\text{inc}} = P_{i0} e^{-i(k_x x + k_z z)} e^{i\omega t} e^{i\varphi_0},
\tag{2}
$$

where P_{i0} is the incident wave amplitude and φ_0 is the initial phase of the incident wave at $t = 0$. Similarly, the reflected wave P_{ref}, transmitted wave P_{tr} and plate displacement w can be obtained as:

$$
\begin{aligned}
P_{\text{ref}} &= P_{r0} e^{-i(k_x x - k_z z)} e^{i\omega t} e^{i\varphi}, \\
P_{\text{tr}} &= P_{t0} e^{-i(k_x x + k_z z)} e^{i\omega t} e^{i\varphi}, \\
w &= W_0 e^{-i k_x x} e^{i\omega t} e^{i\varphi}.
\end{aligned}
\tag{3}
$$

The equation of motion for an ideal fluid medium leads to Eq. (4):

$$
\frac{\partial p}{\partial z} = -\rho_0 \frac{\partial v}{\partial t} = -\rho_0 \frac{\partial^2 w}{\partial t^2}.
\tag{4}
$$

According to the bending equation of a homogeneous thin plate, it can be obtained as:

$$
D \frac{\partial^4 w}{\partial x^4} - \rho h \omega^2 w = P_{\text{inc}}\big|_{z=0} + P_{\text{ref}}\big|_{z=0} - P_{\text{tr}}\big|_{z=0},
$$

$$
D = \frac{E h^3}{12(1 - \gamma)},
\tag{5}
$$
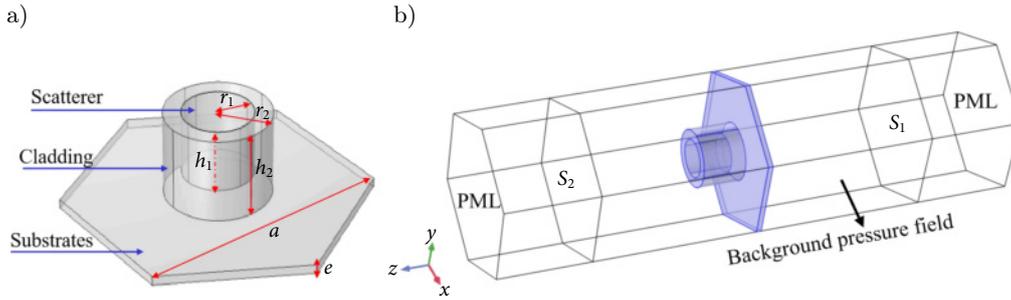
Fig. 3. Schematic illustrations of the unit cells of the plate-type acoustic metamaterial models (a)
and schematic illustration of the calculation of the STL (b).

where E is Young's modulus of the plate material, D is the bending stiffness of the plate, and γ is Poisson's ratio.

Through Eqs. (2)–(4), the relationship between acoustic amplitude and plate amplitude can be obtained as:

$$P_{r0} - P_{i0} = -\frac{\rho_0 i\omega^2 W_0}{k_z},$$
$$P_{t0} = \frac{\rho_0 i\omega^2 W_0}{k_z}. \tag{6}$$

Then, substituting Eq. (6) into Eq. (5) to obtain P_{i0} as:

$$P_{i0} = \frac{W_0 \left[D\left(k_0 \sin\theta\right)^4 - \rho h\omega^2 \right]}{2} + \frac{\rho_0 i\omega^2 W_0}{k_0 \cos\theta}. \tag{7}$$

The acoustic transmission coefficient τ and the STL can be obtained as:

$$\tau_{(\theta)} = \frac{|P_{t0}|^2 / 2\rho_0 c}{|P_{i0}|^2 / 2\rho_0 c}$$
$$= \frac{2\rho_0 i\omega^2}{\left[D\left(k_0 \sin\theta\right)^4 - \rho h\omega^2 \right] k_0 \cos\theta + 2\rho_0 i\omega^2}, \tag{8}$$

$$\mathrm{STL}_{(\theta)} = 10 \log_{10}\left(\frac{1}{\tau_{(\theta)}}\right). \tag{9}$$

An acoustic metamaterial type plate cell model was developed in COMSOL Multiphysics software (Fig. 3a). The basic parameters of the cell are: substrate thickness $e = 1$ mm, and individual cell lattice constant $a = 30$ mm. The radius and height of the cladding panel are $r_2 = 6$ mm and $h_2 = 9$ mm. In addition, the radius and height of the scatterer vary depending on the research scenario. Based on the acoustic metamaterial cell, we conducted an acoustic isolation simulation (Fig. 3b). The acoustic metamaterial type plate cell under study is shown in dark color at the center. There are two layers of air domains above and below the monocell. The uppermost layer serves as a perfect matching layer to absorb sound waves and simulate a non-reflective sound field. The lower air domain

of the wall plate is set as the background pressure field for acoustic excitation. The simulation uses acoustic-structure interaction and applies the perfectly matched layer (PML) to the model. A background pressure field (P_0) is applied on the back of the plate. In this study, Eq. (8) is expanded to apply to the sound insulation simulation model structure, i.e., Eq. (10):

$$\tau_\theta = \frac{F_{\mathrm{in}}}{F_{\mathrm{out}}}, \tag{10}$$

$$F_{\mathrm{in}} = \int_{S_{\mathrm{in}}} \frac{P_{\mathrm{inc}}^2 \cos\theta}{2\rho_0 c}\, \mathrm{d}S, \tag{11}$$

$$F_{\mathrm{out}} = \int_{S_{\mathrm{out}}} \frac{P_{\mathrm{out}c}^2 \cos\theta}{2\rho_0 c}\, \mathrm{d}S, \tag{12}$$

where F_{in} and F_{out} are the acoustic energies on the two surfaces S_1 and S_2, S_{in} and S_{out} are the areas of S_1 and S_2, respectively, P_{inc} and $P_{\mathrm{out}c}$ are the sound pressures of S_1 and S_2, respectively, ρ_0 is the air density, c is the velocity of sound waves propagating in the air, and θ is the pitch angle of sound waves incident on the sound waves. By combining Eq. (9) and Eq. (10), we can calculate the STL.

3. Results and discussion

3.1. Sound insulation (simulation and experimental)

In order to verify the simulation results, we carried out comparative experiments on acoustic metamaterial plates with scatterer radius of 4 and 5 mm. The obtained comparison curves between experimental and simulation results are presented in Fig. 4.

It can be seen in Fig. 4a that certain inconsistencies exist between the outcomes obtained from the experimental analysis and the finite element simulation results at the peaks of the sound insulation curve. Specifically, the frequencies of the peaks are 288 and 309 Hz, with corresponding sound insulation values of 52 and 45 dB, respectively. However, at the valleys of the curve, both the experimental results and

a)

b)
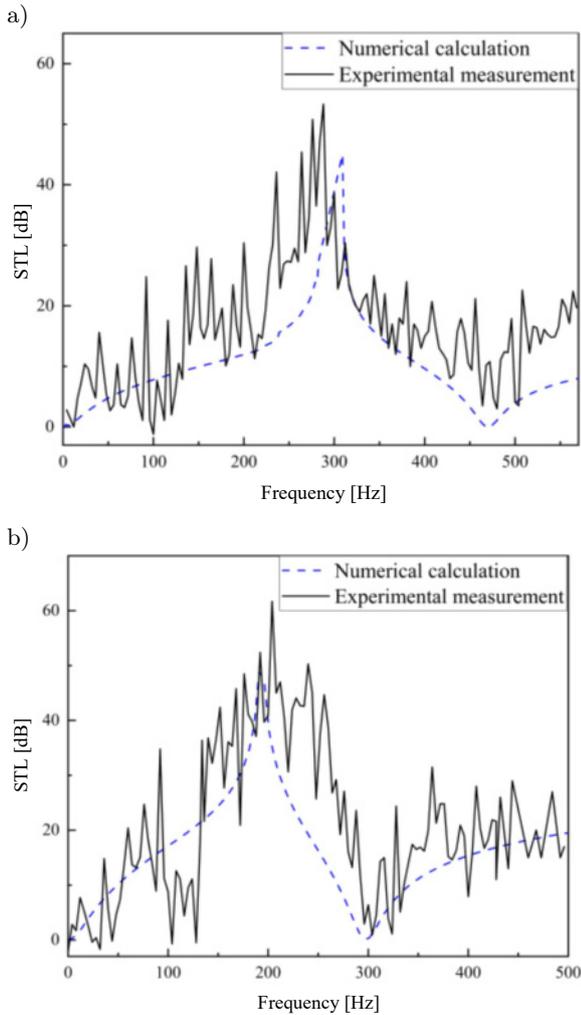
Fig. 4. Comparison curves between experiment and simulation for scatterer radius of 4 mm (a) and 5 mm (b).

finite element simulation results exhibit a consistent trend. The sound wave frequencies at these points are 475 and 463 Hz, with corresponding sound insulation values of 1 and 3 dB, respectively.

As can be seen from the curves in Fig. 4b, the growth or reduction trend of the sound insulation curves is basically consistent with the simulation curves. During the rising phase of sound insulation, the experimental results curve exhibits a faster growth than the simulation results curve, and there are some cliff-like decrease places, but the overall trend is still growing. The frequencies corresponding to the peaks of the experimental curves are slightly higher than the simulation results, with frequencies of the peaks of 209 and 198 Hz, and the corresponding sound insulation values of 61 and 49 dB, respectively. Additionally, the frequencies corresponding to the valleys of the experimental curves match the simulation results.

During the sound insulation experiment, errors in the measured sound insulation values and corresponding frequency bands may occur due to various factors

such as test equipment, sample processing, and the experimental environment. The specific reasons for the errors are: (1) the sound insulation values and corresponding frequency bands may have been affected due to limitations in the experimental equipment. This is because the actual tested specimen was a circular multi-cell specimen, which is different from the hexagonal single-cell specimen used in the simulation test; (2) the experimental measurements of sound insulation may vary due to potential errors in the manufacturing of the tested parts.

Overall, although there are some deviations between the experimental results and the simulation results, the finite element simulation results effectively predicts the sound insulation performance of acoustic metamaterial panels, and there is a certain reference significance in the overall trend of the sound insulation curve.

3.2. Limiting the total mass of the scatterers

We classify and discuss the effects of different scatterer radii (Fig. 5a) and sinking depth of the scatterer (Fig. 5b) on sound insulation, when the total mass of the limited scatterers is maintained constant, with the unit cell lattice constant $a = 30$ mm, substrate thickness $e = 1$ mm, and the cladding radius $r_2 = 6$ mm, and the total height $h_2 = 9$ mm kept unchanged.

a)

b)
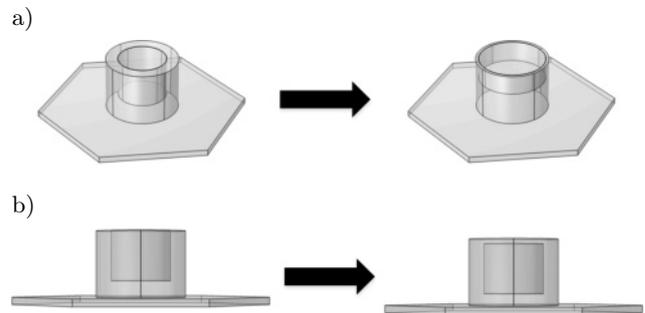
Fig. 5. The total mass of the scatterer is kept constant, changing the radius of the scatterer, with the height of the scatterer adjusting accordingly (a) and the radius and height of the scatterer remain unchanged, only the sinking depth Z of the scatterer is changed (b).

In order to quantify the advantages and disadvantages of different lattice constants, the effective sound insulation band is defined here as the band range in which the sound insulation is 5 dB higher than that of a homogeneous plate with the same mass (ZHANG *et al.*, 2016b). Therefore, we choose a range of sound insulation higher than 30 dB. The red dots marked as f_1 and f_2 represent the start and cut-off frequencies when the sound insulation is 30 dB (Figs. 6a and 6b). The normalization is calculated using Eq. (13):

$$\Delta f = \frac{f2 - f1}{(f2 + f1)/2}. \tag{13}$$

First of all, with the total mass of the scatterer held constant and only changing the radius of the scatterer, the height of the scatterer adjusts accordingly. We choose the radius r_1 of the scatterer as 4, 4.5, 5, 5.5 mm, respectively, and the height h_1 corresponds to 4.75, 5.3, 4.32, 3.57 mm. It can be clearly seen that the effective sound isolation band gradually shifts to the low frequency with an increase scatterer radius, the onset
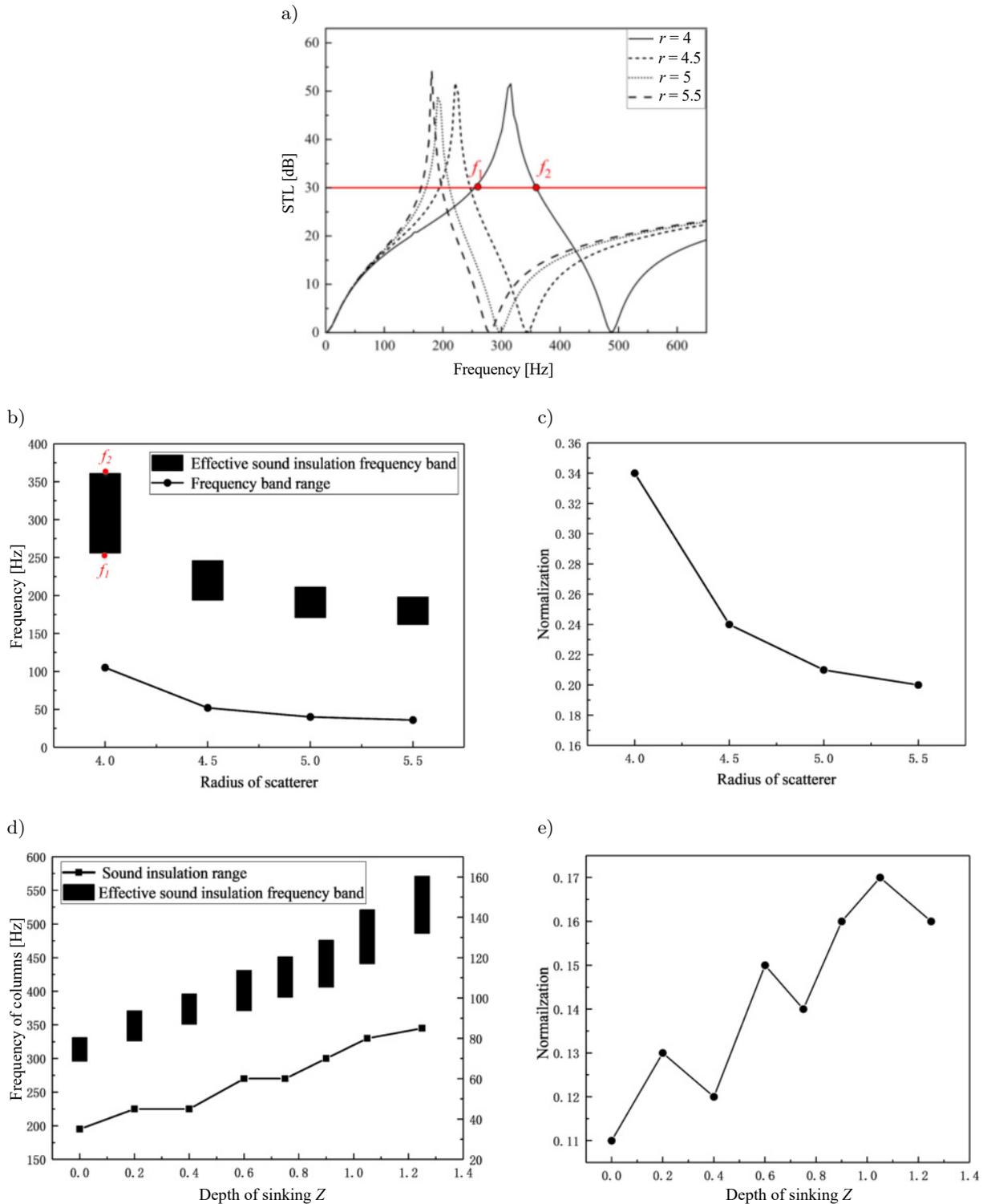
Fig. 6. a) Sound insulation curves corresponding to scatterers with radius of 4, 4.5, 5, and 5.5 mm; b) effective sound insulation frequency band and frequency band range corresponding to different scatterer radius; c) normalization curve for different scatterer radius; d) effective sound insulation frequency band and frequency band range of different sinking depth Z of scatterers; e) normalization curve for different sinking depth.

frequency decreases from 255 to 163 Hz, the cutoff frequency decreases from 363 to 198 Hz, and the sound isolation range decreases from 108 to 35 Hz (Fig. 6b). It is shown that the normalization is 0.34 when the radius of the scatterer is 4 mm, but it decreases to 0.2 when the radius increases to 5.5 mm (Fig. 6c).

Then, keeping the scatterer radius r_1 = 4 mm and height h_1 = 6 mm unchanged, the scatterer sinking depth Z is changed. Here, Z represents the distance between the upper surface of the scatterer and the upper surface of the cladding, with the selection range of 0 ~1.2 mm. The substrate material is aluminum, the cladding material is silicone rubber, and the diffuser material is copper (Table 1). As the sinking depth Z increases, the upper and lower limits of the effective sound absorption frequency band shift to high frequencies, and the frequency range gradually increases from 35 to 85 Hz (Fig. 6d). As the sinking depth increases, the normalization shows a stepwise growth trend, from 0.11 to 0.17, and then reaches a maximum of 0.17 at a sinking depth of $Z = 1$ mm (Fig. 6e).

In summary, when limiting the additional mass of the resonance unit, the smaller the radius of the scatterer or the deeper the sinking depth Z, the closer the center of gravity of the scatterer is to the substrate, and the resonance unit as a whole is more stable. The range of effective sound isolation frequency band is obviously widened, and the frequency band is gradually shifted to high frequency, and the degree of normalization is also gradually increased. The band gap can be maximally broadened, and the sound insulation at low frequencies below 600 Hz can be realized, thus improving the low-frequency sound insulation efficiency of acoustic metamaterials.

3.3. Not limited to additional mass of the resonance unit

We classify and discuss the effects of different filling rates f and different materials of different components on sound insulation, when there are no restrictions on the additional mass of the resonance unit, with a cell constant a = 30 mm, substrate thickness e = 1 mm, the radius and height of the cladding are r_2 = 6 mm, h_2 = 9 mm and, additionally, the radius and height of the scatterer vary depending on the research scenario.

Initially, while maintaining the materials for the substrate, cladding and scatterer as aluminum, silicon rubber and copper, respectively, we consider a filling rate range from 0.3 to 0.6. The filling rate f is calculated from Eq. (7):

$$f = \frac{V_S}{V_C}, \tag{14}$$

where V_S and V_C are the volumes of the scatterer and the cladding, respectively. Then, with a constant filling rate of 0.5, the materials of the scatterer and the substrate are changed. The scatterer materials are tungsten, lead, copper and steel, and the substrate materials are aluminum and epoxy resin (Table 1).

The effective sound insulation frequency band, corresponding to different filling rates f, shows an overall trend of increasing and then decreasing. When the filling rates is in the range of 0.3 to 0.5, the sound insulation range increases from 46 to 155 Hz, and the effective sound insulation frequency band gradually shifts to the high frequency. The cutoff frequency also increases from 252 to 436 Hz. As the filling rates increases from 0.5 to 0.6, the sound isolation range decreases and eventually settles at 107 Hz. However, the sound isolation range remains 27 Hz higher than at the filling rates of 0.4. This shows that the sound insulation effect is better when the filling rate is in the range of 0.45~0.6, and the sound insulation effect is best when the filling rate is 0.5 (Fig. 7a). As the filling rates increase, the normalization shows a trend of first increasing and then remaining stable, indicating that the filling rate is more stable in the range of 0.4~0.6, and the filling rates selected in this range are preferrable in practical applications (Fig. 7b).

When the substrate is aluminum, increasing the density of the scatterer material causes a gradual decrease in the upper and lower boundary frequencies of the effective sound insulation band. However, when the substrate is epoxy resin, increasing the density of the scatterer material results in the upper and lower boundary frequencies of the effective sound insulation band shifting towards higher and lower frequencies, respectively (Fig. 7c). The scatterer material is arranged in the following order: steel, copper, lead, and tungsten, with a gradual increase in density from 7780 to 19 100 kg/m^3. Regardless of whether the substrate material is aluminum or epoxy resin, the frequency band range and normalization both exhibit an increasing trend. However, when the substrate material is epoxy resin, the advantages are more obvious, resulting in a larger frequency band range and normalization compared to aluminum as substrate material (Fig. 7d).

The scatterer acts like the mass in a spring-mass oscillator. The density of the scatterers directly affects the total equivalent mass of the cylinder. A denser scatterer not only increases the unit weight of the acoustic metamaterial, but also increases the resulting bandgap of frequencies. Therefore, when selecting scatterers, it is recommended to use high-density and high-elastic modulus metals such as lead and tungsten to obtain a wider bandgap range. However, it is important to consider the impact of the quality factor in practical applications, making metals such as copper and steel viable choices as scatterers. Additionally, low density materials such as epoxy resins should be chosen as substrates when constructing phononic crystals.

a)

b)

c)

d)
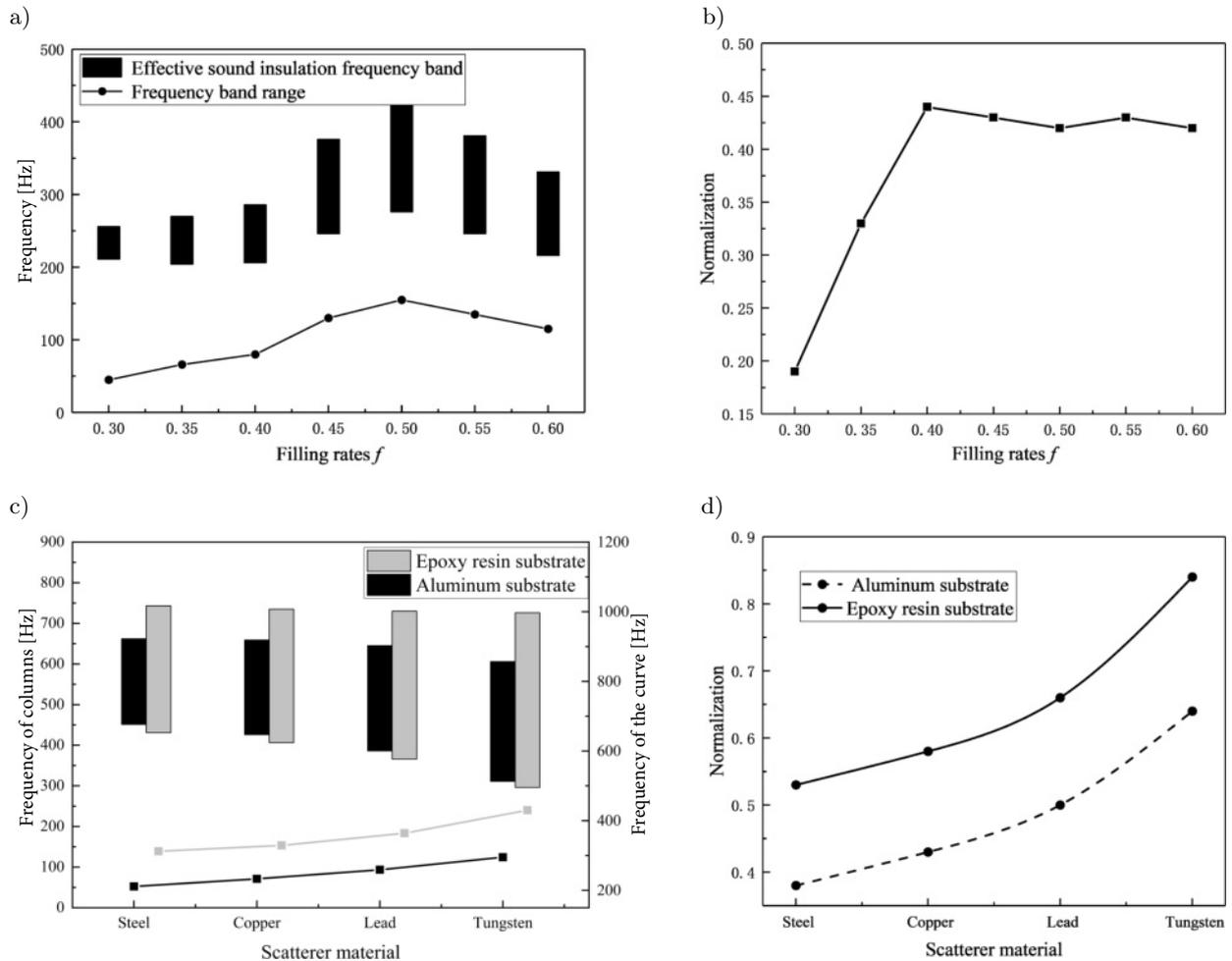
Fig. 7. a) Effective sound insulation frequency band and frequency band range corresponding to different filling rates; b) normalization curve for different filling rates; c) bar chart representing the effective sound insulation band and the curve representing the band range – the black color corresponds to the aluminum substrate and the grey color to the epoxy resin substrate; d) normalization curve for different materials.

4. Conclusion

This paper proposed a plate-type acoustic metamaterial with good sound insulation in low-frequency ranges. In the study, we used finite element simulation to analyze the sound insulation characteristics of the three-component cladding acoustic metamaterial panel, and carried out experimental verification. The results show that the three-component cladding acoustic metamaterial plate can significantly suppress the propagation of noise across a wide frequency band below 600 Hz. Both theoretical and experimental findings demonstrate that the acoustic metamaterial plate provide excellent sound insulation, particularly at the resonant frequency of the local resonance unit. The effects of structural parameters of the scatterer, the filling rate of the resonant unit, and the materials of the components on the sound insulation properties of the acoustic metamaterials were subsequently investigated. The following conclusions are obtained.

When limiting the additional mass of the resonance unit, the smaller the radius of the scatterer or the deeper the sinking depth, the effective sound insulation frequency band shifts to high frequencies, and the frequency band range and normalization tend to increase. When the additional mass of the resonance unit is not limited, as the filling rate f increases, the effective sound insulation frequency band and frequency band range first increase and then decrease. Therefore, in order to obtain a better sound insulation effect, the filling rate should be kept at around 0.5. When selecting scatterer materials, it is recommended to use metals with high density and high-elastic modulus to obtain a wider bandgap range. Additionally, it is advisable to opt for a substrate material with a lower density.

Aknowledgments

Data availability

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

Conflicts of interest

The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. ATMOJO A.Y. *et al.* (2021), ANC Gen1: BBTA3-BPPT 1st prototype of active noise control for vehicle cabin noise, [in:] *Journal of Physics: Conference Series*, **1951**(1): 012030, doi: 10.1088/1742-6596/1951/1/012030.

2. BADREDDINE ASSOUAR M., OUDICH M. (2012), Enlargement of a locally resonant sonic band gap by using double-sides stubbed phononic plates, *Applied Physics Letters*, **100**(12): 123506, doi: 10.1063/1.3696050.

3. BADREDDINE ASSOUAR M., SENESI M., OUDICH M., RUZZENE M., HOU Z. (2012), Broadband plate-type acoustic metamaterial for low-frequency sound attenuation, *Applied Physics Letters*, **101**(17): 173505, doi: 10.1063/1.4764072.

4. CHEN D., ZI H., LI Y., LI X. (2021), Low frequency ship vibration isolation using the band gap concept of sandwich plate-type elastic metastructures, *Ocean Engineering*, **235**: 109460, doi: 10.1016/j.oceaneng.2021.109460.

5. HE X.D., WEN J.H. (2018), Study on the effect of lattice constants on the sound insulation properties of acoustic metamaterial plates, *Noise and Vibration Control* (*S1*), pp. 51–55.

6. HSU J.C. (2011), Local resonances-induced low-frequency band gaps in two-dimensional phononic crystal slabs with periodic stepped resonators, *Journal of Physics D: Applied Physics*, **44**(5): 055401, doi: 10.1088/0022-3727/44/5/055401.

7. HSU J.C., WU T.T., HSU H.S. (2013), Measurement of frequency gaps and waveguiding in phononic plates with periodic stepped cylinders using pulsed laser generated ultrasound, *Journal of Applied Physics*, **113**(8): 083511, doi: 10.1063/1.4793491.

8. IANNAC G., CIABURRO G., TREMATERRA A. (2021), Metamaterials acoustic barrier, *Applied Acoustics*, **181**: 108172, doi: 10.1016/j.apacoust.2021.108172.

9. JIANG C., MOREAU D., FISCHER J., DOOLAN C. (2021), Additively manufactured sound-absorbing porous structures for airfoil trailing-edge noise control, *Journal of Aerospace Engineering*, **34**(5): 04021068, doi: 10.1061/(ASCE)AS.1943-5525.0001317.

10. LI S., CHEN T., WANG X., LI Y., CHEN W. (2016), Expansion of lower-frequency locally resonant band gaps using a double-sided stubbed composite phononic crystals plate with composite stubs, *Physics Letters A*, **380**(25–26): 2167–2172, doi: 10.1016/j.physleta.2016.03.027.

11. LI Y., CHEN T., WANG X., XI Y., LIANG Q. (2015), Enlargement of locally resonant sonic band gap by using composite plate-type acoustic metamaterial, *Physics Letters A*, **379**(5): 412–416, doi: 10.1016/j.physleta.2014.11.028.

12. LIANG B., YUAN B., CHENG J.C. (2009), Acoustic diode: Rectification of acoustic energy flux in one-dimensional systems, *Physical Review Letters*, **103**(10): 104301, doi: 10.1103/PhysRevLett.103.104301.

13. MA C., SHAO C., WAN Q., WANG X., CHENG Y., LIU X. (2018), A locally-resonant phononic crystal for low-frequency vibration control of vehicle [in Chinese], *Journal of Applied Acoustics*, **37**(1): 152–158, doi: 10.11684/j.issn.1000-310X.2018.01.022.

14. MALDOVAN M. (2013), Sound and heat revolutions in phononics, *Nature*, **503**(7475): 209–217, doi: 10.1038/nature12608.

15. NAKAYAMA M. *et al.* (2021), A practically designed acoustic metamaterial sheet with two-dimensional connection of local resonators for sound insulation applications, *Journal of Applied Physics*, **129**(10): 105106, doi: 10.1063/5.0041738.

16. OUDICH M. *et al.* (2011), Experimental evidence of locally resonant sonic band gap in two-dimensional phononic stubbed plates, *Physical Review B*, **84**(16): 165136, doi: 10.1103/PhysRevB.84.165136.

17. OUDICH M., LI Y., ASSOUAR B.M., HOU Z. (2010), A sonic band gap based on the locally resonant phononic plates with stubs, *New Journal of Physics*, **12**(8): 083049, doi: 10.1088/1367-2630/12/8/083049.

18. PENNEC Y., DJAFARI-ROUHANI B., LARABI H., VASSEUR J.O., HLADKY-HENNION A.C. (2008), Low-frequency gaps in a phononic crystal constituted of cylindrical dots deposited on a thin homogeneous plate, *Physical Review B*, **78**(10): 104105, doi: 10.1103/PhysRevB.78.104105.

19. QIU C., LIU Z. (2006), Acoustic directional radiation and enhancement caused by band-edge states of two-dimensional phononic crystals, *Applied Physics Letters*, **89**(6): 063106, doi: 10.1063/1.2335975.

20. SONG Y.B., WEN J., YU D., WEN X. (2015), Suppression of vibration and noise radiation in a flexible floating raft system using periodic structures, *Journal of Vibration and Control*, **21**(2), 217–228, doi: 10.1177/1077546313488156.

21. WEN J., WANG G., YU D., ZHAO H., LIU Y. (2005), Theoretical and experimental investigation of flexural wave propagation in straight beams with periodic structures: Application to a vibration isolation structure, *Journal of Applied Physics*, **97**(11): 114907, doi: 10.1063/1.1922068.

22. WEN J.H., WANG G., YU D.L., ZHAO H.G., LIU Y.Z. WEN X.S. (2008), Study on the vibration band gap and vibration attenuation property of phononic crystals, *Science in China Series E: Technological Sciences*, **51**: 85–99,doi: 10.1007/s11431-008-0008-x.

23. YANG Q. *et al.* (2020), Simulations on the wide bandgap characteristics of a two-dimensional tapered scatterer phononic crystal slab at low frequency, *Physics Letters A*, **384**(35): 126885, doi: 10.1016/j.physleta.2020.126885.

24. YIN J. *et al.* (2022), Review on research progress of mechanical metamaterials and their applications on vibration and noise control [in Chinese], *Advance in Mechanics*, **52**(3): 508–586, doi: 10.6052/1000-0992-22-005.

25. YU K., CHEN T., WANG X. (2013), Band gaps in the low-frequency range based on the two-dimensional phononic crystal plates composed of rubber matrix with periodic steel stubs, *Physica B: Condensed Matter*, **416**: 12–16, doi: 10.1016/j.physb.2013.02.011.

26. ZHANG J., YAO H., DU J., ZHAO J., DONG Y., QI P. (2016a), Low frequency sound insulation characteristics of the locally resonant phononic crystals in the large aircraft cabin [in Chinese], *Journal of the Chinese Ceramic Society*, **44**(10): 1440–1445, doi: 10.14062/j.issn.0454-5648.2016.10.08.

27. ZHANG H., XIAO Y., WEN J., YU D., WEN X. (2016b), Ultra-thin smart acoustic metasurface for low-frequency sound insulation, *Applied Physics Letters*, **108**(14): 141902, doi: 10.1063/1.4945664.

28. ZHAO H. J., GUO H.W., GAO M.X., LIU R.Q., DENG Z.Q. (2016), Vibration band gaps in double-vibrator pillared phononic crystal plate, *Journal of Applied Physics*, **119**(1): 014903, doi: 10.1063/1.4939484.

29. ZHAO H.J., GUO H.W., LI B.Y., DENG Z.Q., LIU R.Q. (2015), Flexural vibration band gaps in a double-side phononic crystal plate, *Journal of Applied Physics*, **118**(4): 044906, doi: 10.1063/1.4927627.

30. ZHOU P., WAN S., WANG X., FU J., ZHU Y. (2021), A novel hybrid composite phononic crystal plate with multiple vibration band gaps at low frequencies, *Physica B: Condensed Matter*, **623**: 413366, doi: 10.1016/j.physb.2021.413366.

31. ZHOU X., WANG L., QIN L., PENG F. (2020), Improving sound insulation in low frequencies by multiple band-gaps in plate-type acoustic metamaterials, *Journal of Physics and Chemistry of Solids*, **146**: 109606, doi: 10.1016/j.jpcs.2020.109606.

32. ZUO K.C. (2016), Current status of research on aircraft cabin noise, *Journal of Aviation*, **37**(8): 2370–2384.

Research Paper

Comparative Analysis of Acoustic Emission Signals from On-Load Tap-Changers for Potential Detecting of Non-Simultaneous Operations

Andrzej CICHOŃ, Sebastian BORUCKI, Michał WŁODARZ*

Department of Electric Power Engineering and Renewable Energy
Faculty of Electrical Engineering, Automatic Control and Informatics
Opole University of Technology
Opole, Poland

*Corresponding Author e-mail: michal.wlodarz@student.po.edu.pl

The research reported in this paper deals with the potential of detecting non-simultaneous operation in on-load tap-changer (OLTC) using an acoustic emission method. Tests conducted under laboratory conditions were carried out using an OLTC model. Three transducers with different characteristics were used: WD 17 AH, D9241A, and R15α, alongside oscillography as the reference method. The use of two new descriptors in the time domain was proposed. The feasibility of detecting the defect with different piezoelectric transducers was investigated.

As a result of the analysis of the results, it was found that each piezoelectric transducer can identify non-simultaneous operation of the switch. The most significant changes in descriptor values occurred in the time domain, and the most effective transducer turned out to be R15α.

Keywords: acoustic emission; on-load tap-changer; piezoelectric transducer.

1. Introduction

Maintaining the appropriate quality parameters of transmission networks constitutes an important task for electricity distributors. One of the critical parameters is the voltage level, which affects how reactive power flows through the network. Voltage levels in an electric power system can fluctuate for various reasons, including changes in energy demand. The on-load tap-changer (OLTC) enables to adjust the voltage levels in power lines by changing the transformer's turn ratio. This is achieved by changing the number of active turns in the secondary winding. The OLTC switching mechanism is designed to work while the transformer is in use, without the need to shut the unit down.

Power transformers, crucial components of the transmission network, significantly impact the stability of power system operation. Although these devices have a relatively low failure rate, the potential cost to power utilities is very high in the event of an incident.

The OLTC is a component of transformers and it has the highest failure rate, making its diagnosis an important issue (MAJCHRZAK *et al.*, 2016). The causes of OLTC damage can be divided into three groups (JONGEN *et al.*, 2014):

– failure of the mechanical system, mainly related to the torque transmission system;
– damage to the main circuit, due to wear or damage to the contacts;
– damage to the insulation system.

During OLTC operation, the mechanical energy generated by the drive is stored in a mechanical energy accumulator in the form of two parallel springs. If one spring breaks, the switching process is extended, which can lead to resistor overheating (DUAN, WANG, 2015).

The contacts of the switch are affected by degradation due to arcing. Excessive contact wear can lead to an increase in contact resistance, which raises the temperature of these components and accelerates their

degradation (Kang, Birthwhistle, 2001a; 2001b; Schellhase *et al.*, 2002). If their contact degradation varies between phases, asymmetrical contact switching may occur.

The oscillographic method has been established and widely used for assessing the condition of OLTCs. It uses the characteristic changes in the current flowing through the OLTC during switching to determine the degree of wear in current path components and the mechanical system (Boricic *et al.*, 2019; Jongen *et al.*, 2012). Dynamic resistance measurement takes a similar course to the oscillographic method. The main difference is that the resistance between the OLTC terminals is measured during switching instead of the current (Aziz *et al.*, 2014; Osmanbasic, Skelo, 2017). A disadvantage of these methods is that the unit must be taken out of service. Due to the strategic importance of transformers in electricity grids, shutting them down for diagnostic purposes is problematic and it is associated with additional costs for energy distributors.

Acoustic signals are generated during the OLTC switching process. Their source may be associated with the switch's mechanical system, contacts operation, or electrical discharges. The switches are mounted in containers filled with oil, and the acoustic waves generated are transmitted to the metal walls of the tank. This makes it possible to record acoustic emission (AE) signals using piezoelectric transducers. The recorded AE signals provide information characterizing the operation of the OLTC, enabling the diagnosis of the device using these signals (Cichoń *et al.*, 2011a; 2011b; Li *et al.*, 2012).

The most important advantage of the AE method is that diagnostic tests can be carried out online, eliminating the need to shut down the transformer. In addition, the AE method can be used simultaneously to determine the mechanical condition of the switch and to detect partial discharges (Cichoń *et al.*, 2011b; 2012; Secic, Kuzle, 2017; Seo *et al.*, 2017). However, acoustic interference generated by the transformer and surrounding devices can lead to a reduction in the effectiveness of this diagnostic method. Besides, the time waveforms of the AE signals generated during power switch operation are difficult to interpret. A correct diagnosis requires expert knowledge, so artificial intelligence (AI) tools are proposed (Wotzka, Cichoń, 2020; Wotzka *et al.*, 2019).

This article, which constitutes a follow-up of research conducted for several years at Opole University of Technology, presents the results of measurements using the oscillographic and AE methods. Tests were carried out on two systems: one without defects and another where non-simultaneous operation occurred. The main goal of the research was to determine the possibility of detecting non-simultaneous operation of the OLTC using piezoelectric transducers with different transmission characteristics. In addition, a compar-

ative analysis of the used transducers was performed to determine the most effective means of detecting the asymmetry in system operation. A proposal for two new descriptors describing AE signals in the time domain is presented. The article presents the differences in these descriptors between the normal system and the modelled defect. Also, methods for determining descriptors based on the transducer used are proposed.

2. Experimental setup

The research focused on the analysis of the OLTC's operating stage when changes occurred in the selector tap position under the control of the power switch. Tests were conducted in a laboratory setting using an OLTC model with a separate selector and a VEL-110 power switch. An actual OLTC system with a selector shortened to six taps was used to create the test bench. The switch, together with the selector, was placed in a tank filled with insulating oil. There were pin-outs on the top cover of the tank to allow testing OLTC with the current flow. The measuring system was equipped with a motor that allowed switching. Three single-phase transformers were used to simulate the impedance of the transformer windings. The setup utilized in this study provides a range of defects that can be modeled: contact wear, non-simultaneous switching, and spring failure. A PLC was also installed to automate switching operations, thus speeding up measurements.

The non-simultaneous operation of the OLTC was simulated by changing contacts from new to worn ones. The degree of wear was simulated by milling the appropriate thickness of the original contact. The changes in the contact thickness used during research were as follows:

– phase A – 2 mm;
– phase B – 3 mm;
– phase C – 0 mm.

During oscillographic measurements, DC flows through the OLTC. During switching, there occur changes in the value of the current. The degree of wear of the device can be determined based on waveform. The results obtained with this method were used as a reference for the AE method during the tests.

MT-3, an instrument measuring basic transformer parameters, was used for oscillographic measurements. It can measure the dynamic change in current passing through OLTC during switching. MT-3 samples the current signal at a frequency of 8120 Hz. The manufacturer of the MT-3 also provides OLTC.exe software to assist diagnosticians in assessing the condition of the transformer or OLTC (Energo-Complex, 2008).

During switching, the OLTC generates sounds that are transmitted to the metal tank due to the presence of insulating oil. The AE waveforms carry information

about the switching process. The main goal of the research was to determine the feasibility of using these data for the OLTC diagnosis.

Laboratory tests were carried out using an OLTC model. Transducers with different characteristics were used to determine the diagnostic capability of OLTC in various frequency bands. The summary of their technical data is presented in Table 1. The transducers are referred to as 1, 2, and 3. The possibility of using different transducers for AE diagnostics will be evaluated during the study.

Table 1. Technical data of the transducers used (MISTRAS Group, n.d.a; n.d.b; n.d.c).

No.	Type	Frequency band [kHz]	Peak sensitivity, Ref V/[m/s]	Peak sensitivity, Ref V/μbar
1.	WD 17 AH	100–900	56 dB	−61 dB
2.	D9241A	20–60	82 dB	–
3.	R15α	50–400	80 dB	−63 dB

All transducers were mounted by means of magnetic holders. They were connected to amplifiers via pre-amplifiers. The gains of both components are presented in Table 2.

Table 2. Preamplifiers and amplifiers gains.

No.	Preamplifier gain [dB]	Amplifier gain [dB]
1.	20	15
2.	20	3
3.	20	9

For recording AE waveforms, the Acquitek CH3160 measuring card was used, operating at a sampling frequency of 350 kHz. It was coupled with a laptop with the installed AcquiFlex software. Figure 1 depicts the measurement procedure.
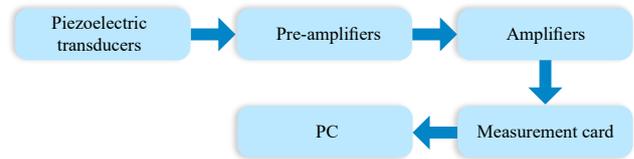

Fig. 1. Measurement procedure.

3. Results

The oscillographic method was used to verify whether an asymmetry of operation occurred in the system following its modification. The results obtained with the AE method will be compared to this method. The results obtained for normal system performance are presented in Fig. 2 and the results for the system with the modeled non-simultaneous operation are shown in Fig. 3.

Characteristic points of the waveforms corresponding to individual switching stages are marked with red lines. For the normal system, it can be observed that for each phase, the switching steps coincide in time, while for the modified system, they occur at different moments. Based on the results obtained, it can be concluded that, after the modification, there is a non-simultaneous operation in the system. A summary of the characteristic times and the differences between each phase is given in Table 3.
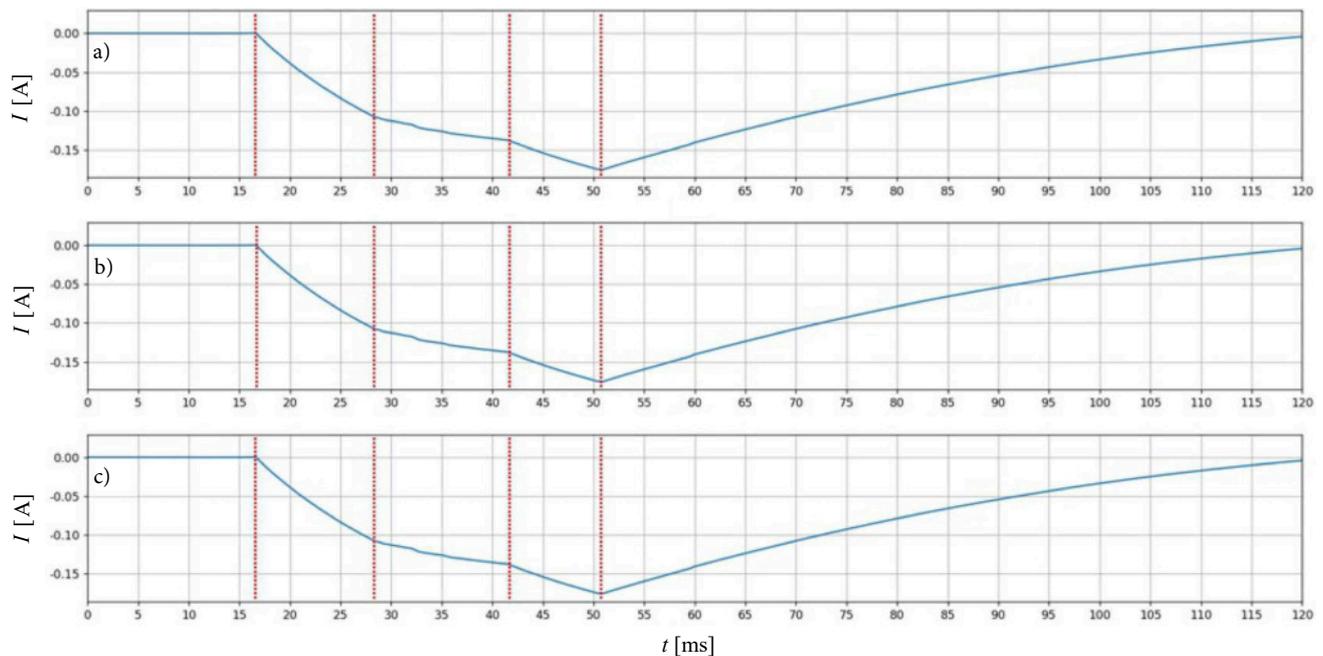

Fig. 2. Oscillographic waveforms for a symmetric system: a) phase A; b) phase B; c) phase C.
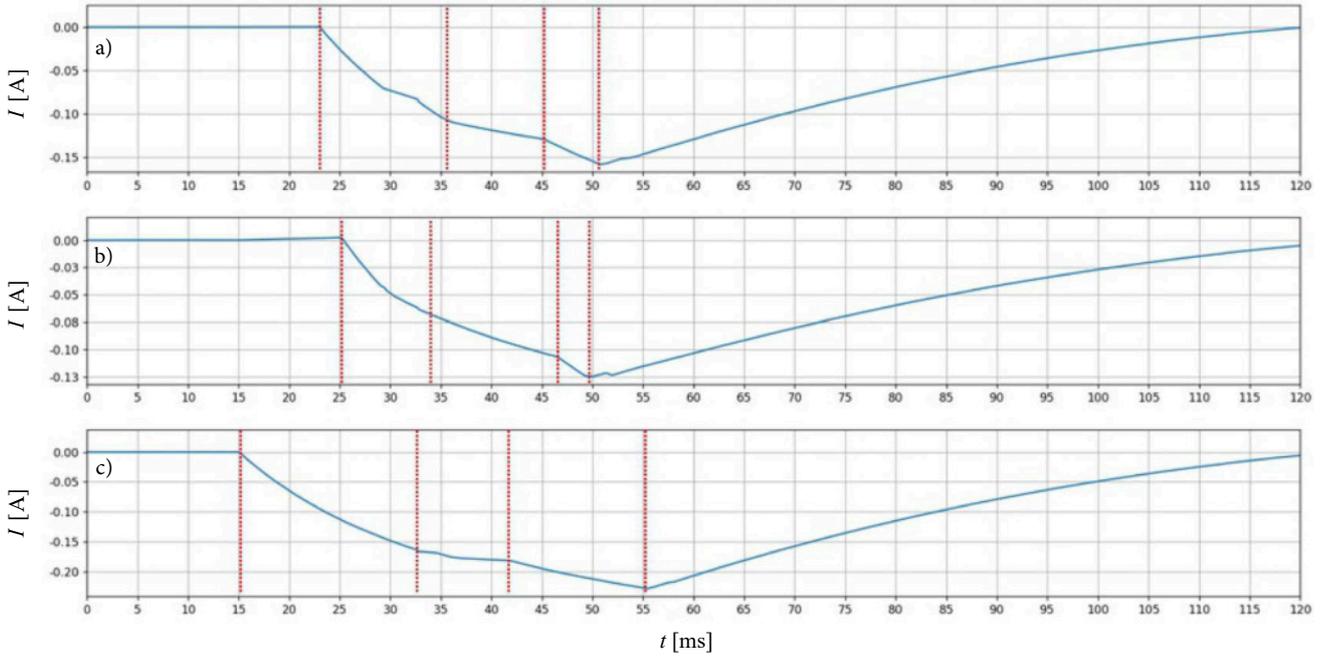
Fig. 3. Oscillographic waveforms for an asymmetric system: a) phase A; b) phase B; c) phase C.

Table 3. Characteristic times read and time differences between the phases.

Simultaneous operation					
Characteristic times [ms]			Time difference [ms]		
A	B	C	A–B	B–C	C–A
16.90	16.80	16.90	0.10	0.10	0.00
28.90	28.90	29.10	0.00	0.20	0.20
41.50	41.40	41.50	0.10	0.10	0.00
50.70	50.70	50.80	0.00	0.10	0.10
Non-simultaneous operation					
Characteristic times [ms]			Time difference [ms]		
A	B	C	A–B	B–C	C–A
23.10	25.00	15.00	1.90	10.00	8.10
36.00	37.40	32.80	1.40	4.60	3.20
45.20	46.60	41.90	1.40	4.70	3.30
50.70	49.60	55.40	1.10	5.80	4.70

Subsequently, the waveforms for the standard system and the modified one were juxtaposed to determine the possibility of diagnosing the modeled fault using the AE method. All signals were normalized by dividing by the maximum value. No filter was used during signal analysis. The results for each of the three transducers are given in Figs. 4–6.

A clear difference can be observed between the symmetrical and asymmetrical systems. The interval describing contact switching is between 40 and 90 ms. In this interval, significantly more acoustic events with smaller amplitudes can be observed for the system with non-simultaneous operation compared to the case of the normal system. This is due to the non-

simultaneous contact closure between the phases. For the normal system, these events occur simultaneously for each phase.

For the accurate determination of the technical feasibility of individual transducers, two descriptors were determined. Firstly, the envelope of the AE signal was determined and then the time at which the envelope is above the threshold of 0.05 was determined. The envelope was obtained by determining the local maxima for which polynomial interpolation was used. The threshold was adjusted by analyzing several time courses and conducting simulations for different values. An example of the determination process is shown in Fig. 7. In this way, the duration of the switching operation was
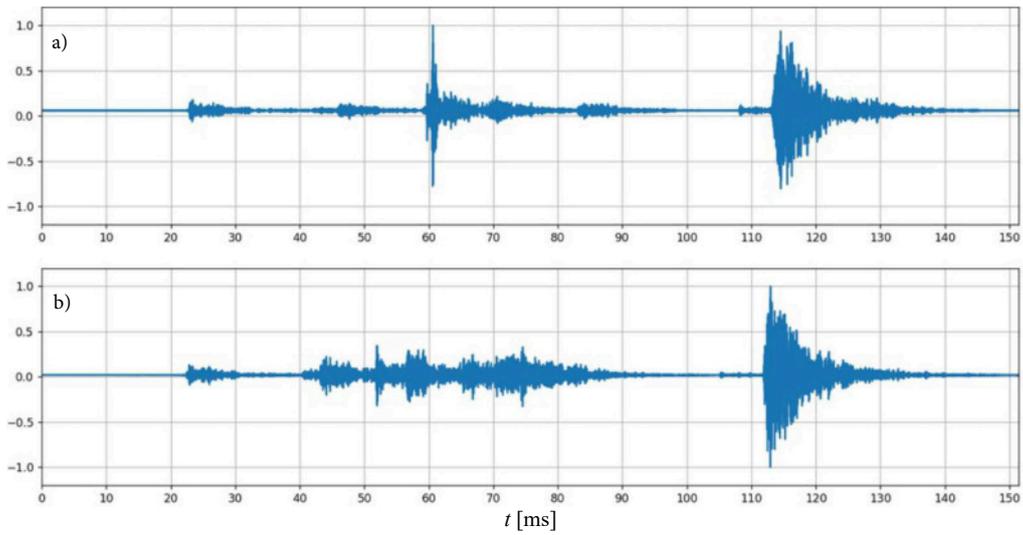
Fig. 4. AE signal generated by OLTC recorded with transducer 1 for:
a) system without asymmetry; b) system with asymmetry.
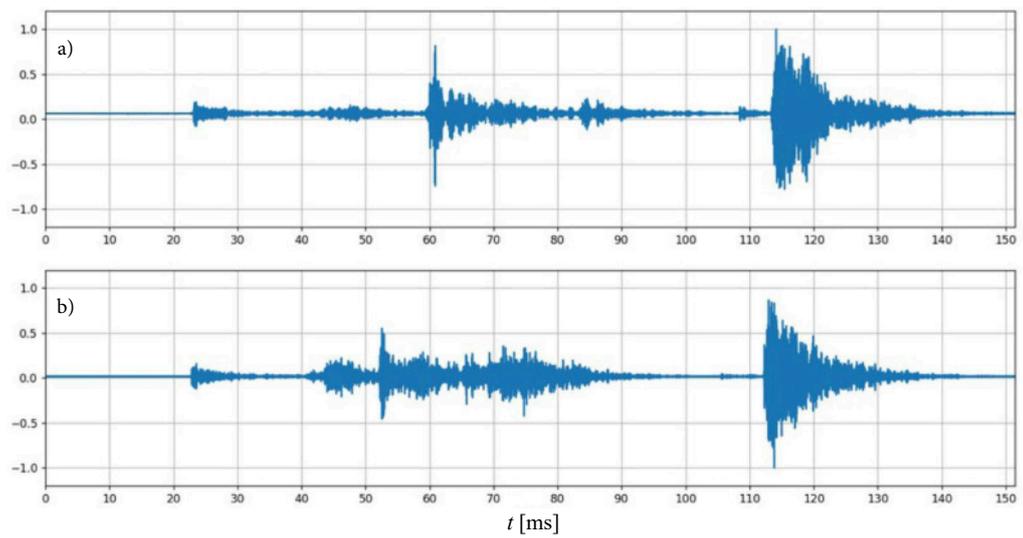
Fig. 5. AE signal generated by OLTC recorded with transducer 2 for:
a) system without asymmetry; b) system with asymmetry.
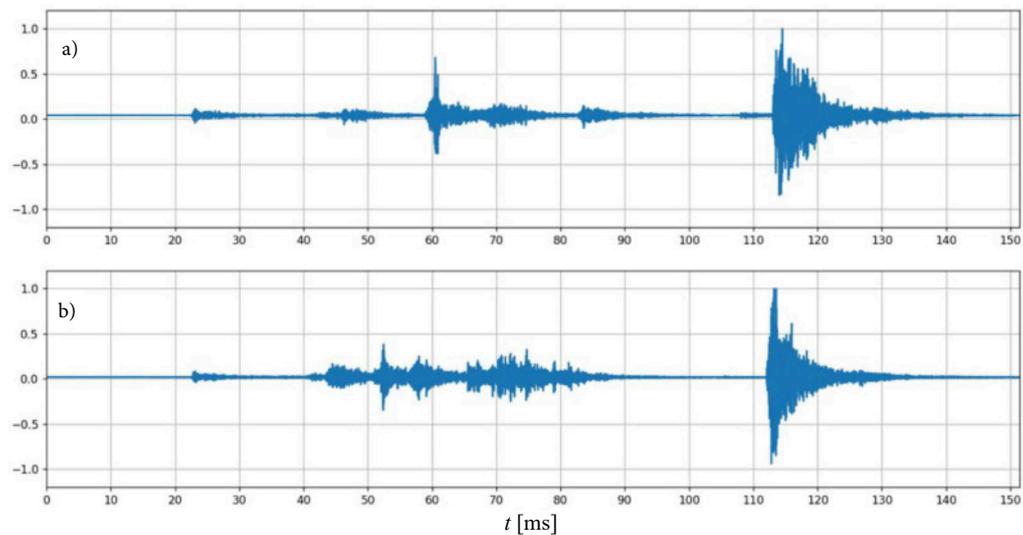
Fig. 6. AE signal generated by OLTC recorded with transducer 3 for:
a) system without asymmetry; b) system with asymmetry.
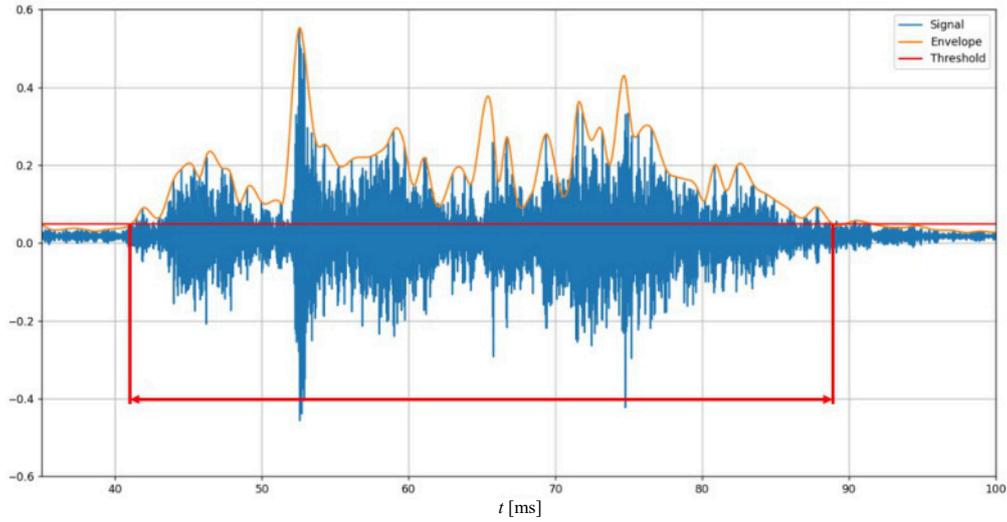
Fig. 7. AE signal generated by OLTC with plotted envelope. The switching time is determined as the time during which the envelope values are above the threshold.
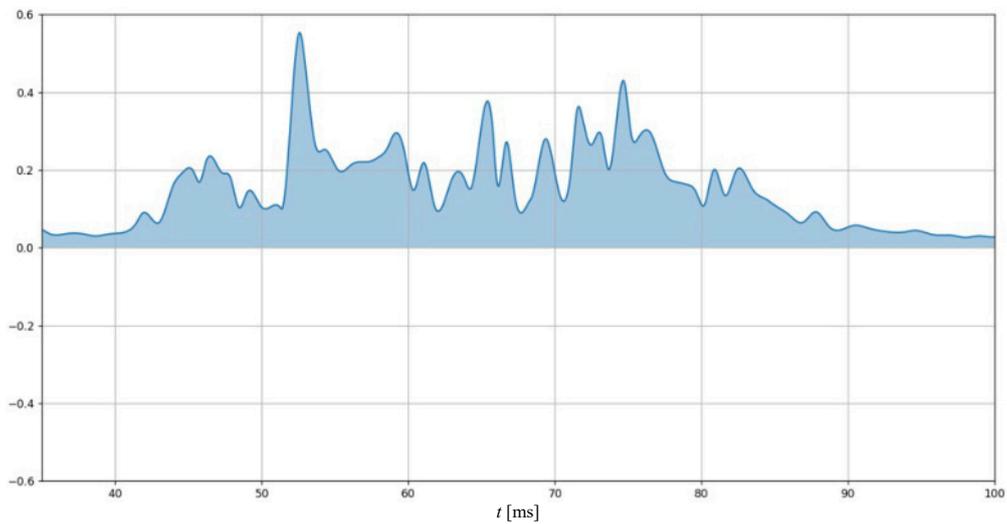

Fig. 8. Graphical representation of the descriptor describing the field under the acoustic emission signal envelope.

calculated for all cases. The second designated descriptor covers the area under the envelope, and a visualization of its determination is shown in Fig. 8.

All calculated descriptors and the relative differences between the values obtained for the symmetric and asymmetric systems are presented in Table 4. The switching duration increased significantly for the system with non-simultaneous operation, while the area under the envelope decreased. The time-frequency analysis constituted the next stage in the analysis of the results. It was utilized not only to illustrate the differences in the frequency response of the different transducers but also to visualize the differences in frequencies found in systems with a defect relative to normal systems.

Figures 9–11 present the results of the time-frequency analysis of AE signals generated by OLTC with original and modified contacts. The presented

Table 4. Descriptors calculated in the time domain.

	Transducer 1			Transducer 2			Transducer 3		
	Symmetry	Asymmetry	The relative difference [%]	Symmetry	Asymmetry	The relative difference [%]	Symmetry	Asymmetry	The relative difference [%]
Duration [ms]	8.19	31.54	321.73	13.71	39.43	187.60	7.76	35.74	360.57
Area under envelope	4.14	2.6	37.20	4.49	2.93	34.74	2.83	2.13	24.73
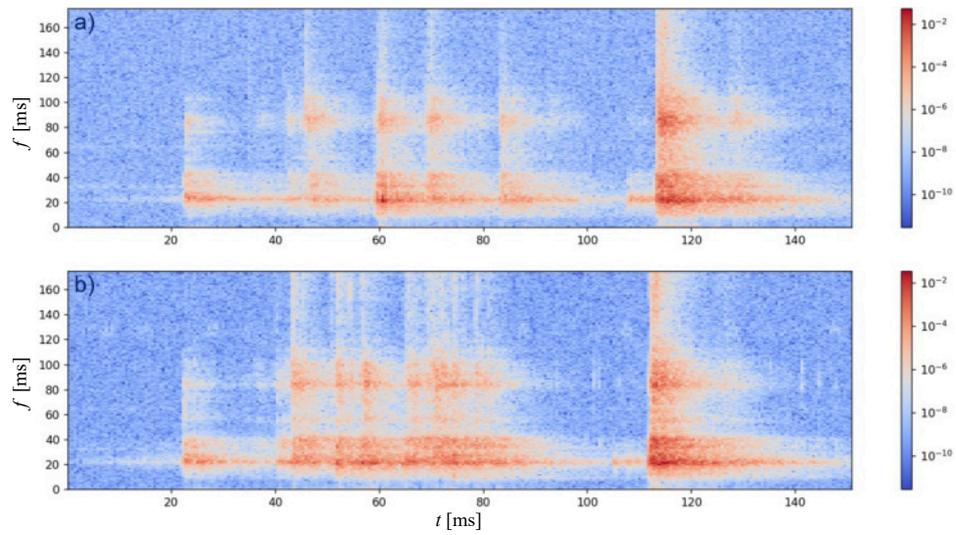
Fig. 9. Spectrogram of AE signal generated by OLTC recorded with transducer 1 for:
a) system without asymmetry; b) system with asymmetry.
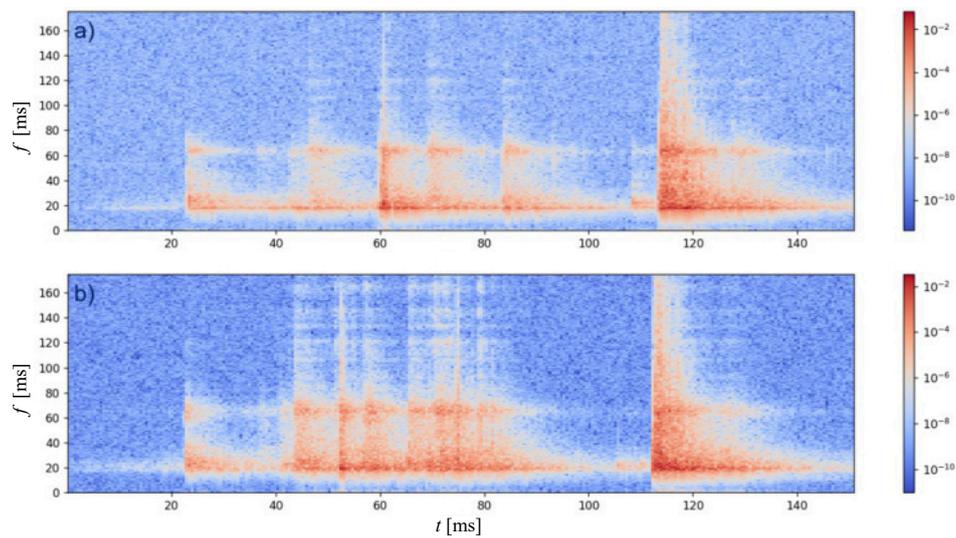

Fig. 10. Spectrogram of AE signal generated by OLTC recorded with transducer 2 for:
a) system without asymmetry; b) system with asymmetry.
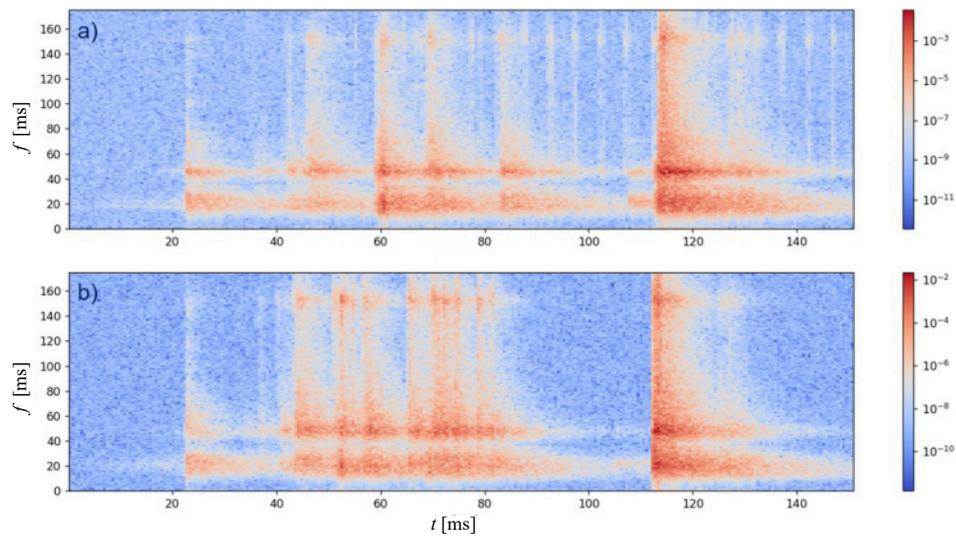
Fig. 11. Spectrogram of AE signal generated by OLTC recorded with transducer 3 for:
a) system without asymmetry; b) system with asymmetry.

results cover the frequency band up to 175 kHz. This is due to the Nyquist frequency. Since OLTC tests are performed for mechanical damage, there is no need to analyze higher frequencies. For the modified system, the presence of a large number of acoustic events in the 40–90 ms range can be seen. These events correspond to individual switch strikes, occurring non-simultaneously for each phase.

In the case of an unmodified system, more discrete structures are visible and can be distinguished in time. In the case of the faulty device, an extension in time of the individual frequency structures is visible. They are relatively continuous in time, which makes it possible to identify a much larger number of unevenly occurring acoustic events resulting from unevenly switching contacts of the individual phases. In the case of uniform switching, we have a relatively synchronized switching cycle because the acoustic signals can be uniquely isolated from the recording. In the case of a modeled fault, this cannot be done. It should be noted that such a phenomenon is visible regardless of the type of transducer used. The clear differences observed in the spectrograms allow an unambiguous assessment of the presence of the defect under investigation.

Each of the transducers used allows the simultaneity of the switch to be assessed. However, the differences are most pronounced for transducer 3. For this transducer, the frequency structures describing the individual beats are most evident at higher frequencies;

thus, it is best suited when diagnosing OLTC using spectrograms of acoustic emission signals.

To avoid relying solely on visual analysis of the obtained waveforms and spectrograms during the diagnosis of OLTC, several descriptors were identified and used:

- maximum value (max);
- root mean squared (RMS);
- median;
- peak factor;
- form factor;
- frequency of highest amplitude.

These descriptors were applied in the analysis of amplitude and power density spectrum. For each of the descriptors analyzed, changes were observed after system modification. The results obtained make it possible to diagnose the switching asymmetry based on analytical data rather than visual inputs. The same procedure was performed on the power density spectrum. Values calculated for the amplitude spectrum are shown in Table 5, while in Table 6 values for the power density spectrum are presented.

The changes between the normal and modified systems are significant, confirming that it is possible to diagnose asymmetric performance based on the proposed descriptors. It can be seen that time series descriptors undergo the most significant changes. The system with asymmetric operation had a significantly

Table 5. Amplitude spectrum descriptors.

	Transducer 1			Transducer 2			Transducer 3		
	Symmetry	Asymmetry	The relative difference [%]	Symmetry	Asymmetry	The relative difference [%]	Symmetry	Asymmetry	The relative difference [%]
Max	4.003E–06	5.435E–06	35.76	1.830E–05	8.889E–06	51.41	3.793E–06	1.248E–06	67.09
RMS	1.269E–07	1.583E–07	24.74	3.020E–07	1.961E–07	35.05	9.936E–08	5.535E–08	44.29
Median	3.594E–10	3.498E–10	2.66	2.129E–10	4.017E–10	88.70	5.661E–10	1.468E–09	159.40
Peak factor	3.154E+01	4.333E+00	86.26	6.059E+01	4.532E+01	25.19	3.817E+01	2.255E+01	40.92
Form factor	6.207E+00	6.421E+00	3.45	9.573E+00	6.459E+00	32.53	5.754E+00	3.744E+00	34.94
F_{max}	21.00 kHz	21.13 kHz	0.60	17.86 kHz	18.97 kHz	6.17	45.53 kHz	46.85 kHz	2.90

Table 6. Power density spectrum descriptors.

|	Transducer 1			Transducer 2			Transducer 3		
	Symmetry	Asymmetry	The relative difference [%]	Symmetry	Asymmetry	The relative difference [%]	Symmetry	Asymmetry	The relative difference [%]
Max	1.927E+02	2.245E+02	16.52	4.119E+02	2.871E+02	30.30	1.876E+02	1.076E+02	42.63
RMS	1.377E+01	1.512E+01	9.80	1.710E+01	1.678E+01	1.89	1.266E+01	1.171E+01	7.47
Median	1.826E+00	1.801E+00	1.34	1.404E+00	1.930E+00	37.46	2.292E+00	3.690E+00	61.05
Peak factor	1.399E+01	1.485E+01	6.11	2.408E+01	1.711E+01	28.96	1.482E+01	9.187E+00	38.00
Form factor	2.279E+00	2.343E+00	2.84	2.562E+00	2.282E+00	10.92	2.118E+00	1.696E+00	19.91
F_{max}	17.86 kHz	18.97 kHz	6.17	1.786E+04	1.897E+04	6.17	45.53 kHz	46.85 kHz	2.90
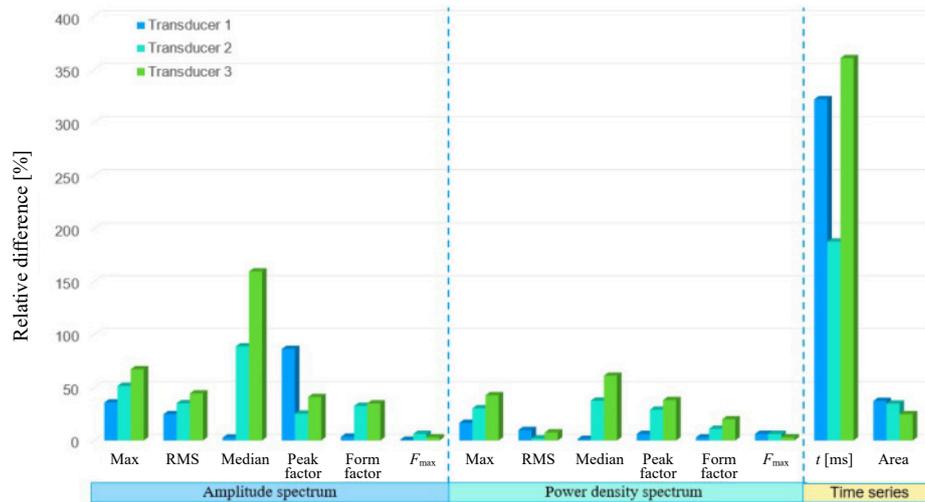
Fig. 12. Relative differences between normal and modified systems for each transducer.

longer switching duration than the normal system. The changes recorded in the descriptors determined based on the amplitude spectrum and power density are also significant.

In Fig. 12, all relative differences in values are plotted. By analyzing the graph, it is possible to determine the tested transmitters for the case yielding the best results in terms of diagnosing the studied defect.

4. Conclusion

On the basis of the analysis of the results obtained in the study, it was found that it is possible to diagnose non-simultaneous operation using each of the piezoelectric transducers subjected to testing. The differences are also clearly visible on the spectrograms. The differences between the time courses for the modified and unmodified systems are discernible, allowing the OLTC condition to be assessed through visual inspection. The assessment of the occurrence of non-simultaneous operations was simplified by calculating the descriptors given earlier. The differences in designated values are substantial, ranging up to 300%. This allows diagnosis to be carried out even by staff members who do not have specialized knowledge. In addition, the ability to carry out OLTC diagnostics without the need to shut down the transformer offers more frequent measurements to be carried out. It is also possible to use the AE method for continuous measurements. The results demonstrate that transducers with different characteristics can be used for AE diagnostics. This opens up the possibility of using transducers to detect partial discharges to monitor the mechanical state of the OLTC.

By comparing relative differences between individual descriptors, we are able to identify the transducer that can be the most useful for diagnosing non-simultaneous operation. The number of descriptors for which the transducer achieved the highest relative difference value:

– transducer 1: 4 differences;
– transducer 2: 2 differences;
– transducer 3: 9 differences.

For transducer 3, the highest number of differences was registered. Therefore, it was concluded that it is the most suitable one for diagnostic purposes.

Two new descriptors, calculated from time series, were introduced, one to describe the duration of the switching event and the other to establish the field under the envelope. These have proven valuable in assessing the symmetry of OLTC operations. Thus, their use in developing an expert system for OLTC diagnostics will allow more efficient identification of the defect under investigation.

The results presented in this paper indicate differences in the AE signals generated by normal system and asymmetrically switching systems. The significant advantage of the AE diagnostic method is to perform diagnosis in a non-destructive way, as it is not necessary to take the unit out of service. Correct interpretation of time courses can be quite a challenging task, so future work will focus on evaluating the potential for detecting different kinds of damages using artificial intelligence.

The research presented in the paper forms one of the phases of work aimed at creating an expert system for online diagnosis of OLTC. The descriptors analyzed in this paper have demonstrated their feasibility in diagnosing non-simultaneity and will, therefore, be used as one of the input parameters for neural network. For other defects, a similar analysis will be carried out. This will allow the creation of a set of descriptors that fully describe the performance of OLTC. Further work will also focus on examining the possibility of using the presented method to determine the technical condition of other types of OLTC. It is expected that after

determining the values of the described descriptors for undamaged OLTCs of other types, they will be successfully used as a reference point for determining defects occurring in the described devices.

References

1. Aziz M.A.A., Talib M.A., Arumugam R. (2014), Diagnosis of on-load tap changer (OLTC) using dynamic resistance measurement, [in:] *IEEE 8th International Power Engineering and Optimization Conference (PEOCO2014)*, pp. 494–497, doi: 10.1109/PEOCO. 2014.6814479.

2. Boricic A. *et al.* (2019), Dynamic resistance measurements and result interpretation for various on-load tap changers, [in:] *IEEE Milan PowerTech*, pp. 1–6, doi: 10.1109/PTC.2019.8810530.

3. Cichoń A., Borucki S., Boczar T. (2011a), Diagnosis of the non-concurrent operation of the on-load tap changer contacts by the acoustic emission method, *Archives of Acoustics*, **36**(4): 823–830, doi: 10.2478/v10168-011-0054-4.

4. Cichoń A., Borucki T., Boczar T., Zmarzły D. (2011b), Characteristic of acoustic emission signals generated by electric arc in on load tap changer, [in:] *Proceedings of 2011 International Symposium on Electrical Insulating Materials*, pp. 437–439, doi: 10.1109/ISEIM.2011.6826306.

5. Cichoń A., Borucki T., Boczar T., Zmarzły D. (2012), The possibilities of using acoustic signals generated by the on load tap changer, [in:] *Proceedings of the 41st International Congress and Exposition on Noise Control Engineering 2012*.

6. Duan R., Wang F. (2015), Mechanical condition monitoring of on-load tap-changers using chaos theory & fuzzy C-means algorithm, [in:] *IEEE Power and Energy Society General Meeting*, doi: 10.1109/PESGM. 2015.7286077.

7. Energo-Complex (2008), MT-3 v.3 Technical description, https://energo-complex.pl/wp-content/uploads/2018/03/ENERGO-COMPLEX-Przyrzad-do-badania-transformatorow-MT-3.pdf (access: 02.04.2023).

8. Jongen R., Gulski E., Siodła K., Parciak J., Erbrink J. (2014), Diagnosis of degradation effects of on-load tap changer in power transformers, [in:] *ICHVE International Conference on High Voltage Engineering and Application*, doi: 10.1109/ICHVE.2014.7035469.

9. Jongen R., Seitz P.P., Smit J., Strehl T., Leich R., Gulski E. (2012), On-load tap changer diagnosis with dynamic resistance measurements, [in:] *IEEE International Conference on Condition Monitoring and Diagnosis*, pp. 485–488, doi: 10.1109/CMD.2012.6416184.

10. Kang P., Birthwhistle D. (2001a), Condition monitoring of power transformer on-load-tap-changers. Part 1: Automatic condition diagnostics, *IEE Proceedings – Generation, Transmission and Distribution*, **148**(4): 301–306, doi: 10.1049/ip-gtd:20010389.

11. Kang P., Birthwhistle D. (2001b), Condition monitoring of power transformer on-load-tap-changers. Part 2: Detection and aging from vibration signature, *IEE Proceedings – Generation, Transmission and Distribution*, **148**(4): 307–311, doi: 10.1049/ip-gtd: 20010388.

12. Li Q., Zhao T., Zhang L., Lou J. (2012), Mechanical fault diagnostics of on-load tap changer within power transformers based on hidden Markov model, [in:] *IEEE Transactions on Power Delivery*, **27**(2): 596–601, doi: 10.1109/TPWRD.2011.2175454.

13. Majchrzak H., Cichoń A., Borucki S. (2016), Application of the acoustic emission method for diagnosis in on-load tap changer, *Archives of Acoustics*, **42**(1): 29–35, doi: 10.1515/aoa-2017-0004.

14. MISTRAS Group (n.d.a), WD Sensor, Product brochure, https://www.physicalacoustics.com/content/literature/sensors/Model_WD.pdf (access: 02.04.2023).

15. MISTRAS Group (n.d.b), D9241A Sensor, Product brochure, https://pdf.directindustry.com/pdf/physical-acoustics/d9241a-sensor/27111-396871.html (access: 02.04.2023).

16. MISTRAS Group (n.d.c), R15α Sensor, Product brochure, https://pdf.directindustry.com/pdf/physical-acoustics/r15-alpha/27111-396891.html (access: 02.04.2023).

17. Osmanbasic E., Skelo G. (2017), Tap changer condition assessment using dynamic resistance measurement, *Procedia Engineering*, **202**: 52–64, doi: 10.1016/j.proeng.2017.09.694.

18. Schellhase H.-U., Pollock R.G., Rao A.S., Korolenko E.C., Ward B. (2002), Load tap changers: Investigations of contacts, contact wear and contact coking, [in:] *Proceedings of the Forty-Eighth IEEE Holm Conference on Electrical Contacts*, pp. 259–272, doi: 10.1109/HOLM.2002.1040850.

19. Secic A., Kuzle I. (2017), On the novel approach to the on load tap changer (OLTC) diagnostics based on the observation of fractal properties of recorded vibration fingerprints, [in:] *IEEE EUROCON 2017 – 17th International Conference on Smart Technologies*, pp. 720–725, doi: 10.1109/EUROCON.2017.8011205.

20. Seo J., Ma H., Saha T.K. (2017), A joint vibration and arcing measurement system for online condition monitoring of onload tap changer of the power transformer, [in:] *IEEE Transactions on Power Delivery*, **32**(2): 1031–1038, doi: 10.1109/TPWRD.2016. 2531186.

21. Wotzka D., Cichoń A. (2020), Study on the influence of measuring AE sensor type on the effectiveness of OLTC defect classification, *Sensors*, **20**(11): 3095, doi: 10.3390/s20113095.

22. Wotzka D., Cichoń A., Manowski P. (2019), Classification of OLTC defects based on AE signals measured by two different transducers, *20th International Scientific Conference on Electric Power Engineering (EPE)*, pp. 1–5, doi: 10.1109/EPE.2019.8778140.

Technical Note

A Side Lobe Level Reduction Method Using Simulated Annealing Algorithm in a Uniform Arc Array

Song-Il KANG*, Kyong-Sim U, Kyong-Chol CHOE, Yong-Kwang RI, Hyok-Il KYE

Institute of Electronic Materials, High Tech and Development Centre
Kim Il Sung University
Pyongyang, Democratic People's Republic of Korea

*Corresponding Author e-mail: si.kang0604@ryongnamsan.edu.kp

In general, the amplitude-weighting method for an acoustic transducer array is widely used to improve the array directivity and reject disturbances. This paper presents a method to effectively reduce the side lobe level while minimizing the main lobe width increase. This is done using the simulated annealing algorithm (SAA) for a uniformly spaced arc array of omnidirectional underwater acoustic transducers, even at low signal-to-noise ratio (SNR). We propose a new cost function for the SAA and obtain the weighting coefficients for all array elements using the SAA, and next compare them with various amplitude weighting methods.

Through simulation and comparison, it is verified that the proposed method is effective in beamforming of the uniform arc array of underwater acoustic transducers.

Keywords: underwater acoustic transducer array; beamforming; simulated annealing algorithm; side lobe level reduction.

1. Introduction

Beamforming is used to increase the transmit power and reduce losses in array antennas. Minimizing the side lobe level and enhancing the main lobe level is very important to improve target detection accuracy and jamming stability for underwater acoustic transducer arrays.

For an underwater acoustic transducer array transmitting and receiving a sound wave, a time delay for individual elements steers the acoustic beam in a certain direction, while array amplitude weighting is used to reduce side lobe level, improving the SNR, and jamming suppression capability. Weighting methods of the two-dimensional arrays include the window method (NOFAL *et al.*, 2013; DESSOUKY *et al.*, 2006; 2007; SARKER *et al.*, 2016; SCHMERR JR. 2015; RUCKSANA BEGUM, RAMARAO, 2015) and various heuristic optimization methods (ALBAGORY, ALRADDADY, 2021; SINGH, SALGOTRA, 2018; LI *et al.*, 2017; VAN LUYEN, VU BANG GIANG, 2017). These optimal search weighting methods generally result in optimal position as well as magnitude weight values for array elements

if a cylindrical array is used. The window weighting method with the array elements placed at equal spacing suits the requirements of problem we consider in this paper.

Conventional weighting methods effectively reduce the side lobe level. However, it is inevitable that a decrease in the level of the side lobe is accompanied by an increase in the main lobe width. So, we cannot reduce the side lobe level indefinitely, and we must also pay due attention to the degradation of the system resolution due to the widening of the main lobe.

Originally, designed for the annealing heat process in metals, the simulated annealing algorithm (SAA) is now applied to optimal designs. The SAA seeks the solution that minimizes the value of a cost function among a number of possible solutions, which is analogous to the process of finding a stable state with the lowest free energy during annealing. There exist many examples of the SAA used in various optimization problems (ZANGENE *et al.*, 2014; CHEN *et al.*, 2019; HONG *et al.*, 1991; CRETU *et al.*, 2010; RASDI RERE *et al.*, 2015; GINTARAS *et al.*, 2019; CARDONE *et al.*, 2002). For instance, ZANGENE *et al.* (2014) reduced the

side lobe level while minimizing the main lobe width increase by applying the SAA to non-uniform circular arrays used in wireless communication. The authors optimized the position of elements placed in the circumference and the element amplitude weight vector to reduce the main lobe level, and demonstrated that the proposed method is more efficient than genetic algorithm and uniform weighting methods. An application of SAA for acoustic sensors can be found in (CHEN et al., 2019), where it is used to effectively reduce the sound pressure level in passenger trains. Many previous methods in which the SAA was used to reduce the side lobe level of the array antenna mostly focused on enhancing the beamforming effect in wireless communication systems and adopted the unequally spaced array.

The SAA was rarely used in underwater signal processing and in the case of arrays used for underwater signal processing, in particular, circular arrays and cylindrical arrays, it is difficult for engineers to arrange acoustic elements with non-uniform spacing at any location. This is because the size of the elements becomes very large due to the fact that the sound velocity is five times higher, and the available frequencies are lower in water than in air.

To overcome this drawback and improve the transmitter and receiver directivity in the acoustic sensor array, the SAA is introduced into the beamforming of uniform arc arrays of underwater acoustic transducers. We propose a new cost function for the SAA and use it to obtain the weight coefficients for the array elements. Thus, we reduce the side lobe level while minimizing the main lobe width increase of the array directivity, even for low signal-to-noise ratio (SNR).

The effectiveness of the proposed method is verified by comparing it with various array weighting methods, such as the cosine weighting method, Hanning weighting method, Hamming weighting method, Blackman weighting method, triangular weighting method, etc. (SCHMERR JR., 2015).

The paper is organized as follows: Sec. 2 briefly describes the array weighting method and the SAA; Sec. 3 presents a new side lobe level reduction method using the SAA; in Sec. 3, the new cost function of SAA and the weighting coefficients of the array elements derived using this SAA are presented; Sec. 4 presents simulation results, underwater test results, and analysis; Sec. 5 provides the conclusion.

2. Theoretical fundamentals

2.1. Array weighting

Arc arrays are a special case of circular arrays, commonly used in underwater acoustic transducer arrays. The array of interest is a uniform arc array consisting of 12 omnidirectional elements, each positioned with identical central angles between them. An illustration of this arrangement is shown in Fig. 1.
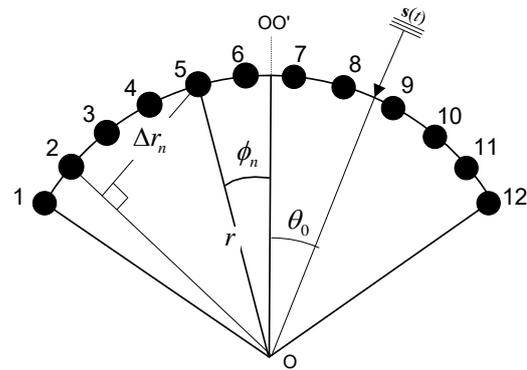

Fig. 1. Uniform arc array with 12 elements.

We assume the signal incident on the array is $\mathbf{s}(t)$, the steering matrix of the array is $\mathbf{A}(\theta, \mathbf{I})$, and the output of the array beamforming is expressed as (ZANGENE et al., 2014):

$$B(\theta, \mathbf{I}) = |\mathbf{A}(\theta, \mathbf{I})\mathbf{s}(t)|, \tag{1}$$

$$\mathbf{A}(\theta, \mathbf{I}) = [a_1(\theta) \quad a_2(\theta) \quad \cdots \quad a_N(\theta)], \tag{2}$$

$$\mathbf{s}(t) = [s_1(t) \quad s_2(t) \quad \cdots \quad s_N(t)]^{\mathrm{T}}, \tag{3}$$

$$a_n(\theta) = I_n e^{j2\pi f_0 r \cos(\theta - \phi_n)/c}, \quad n = 1, 2, ..., N, \tag{4}$$

$$s_n(t) = A_m e^{j2\pi f_0(t - t_n)}, \tag{5}$$

$$t_n = r_n/c = r\cos(\theta_0 - \phi_n)/c, \quad n = 1, 2, ..., N, \tag{6}$$

where θ is the angle of interest, $\mathbf{I}$ is the weight vector, I_n is the n-th weight coefficient of the weight vector, f_0 is the center frequency of the signal, θ_0 is the angle of incidence of the signal, r is the radius of the array, c is the propagation velocity of the signal, ϕ_n is the angle from the center line (OO′) to the n-th element, N is the number of elements, and A_m is the signal amplitude. The sign of the angle is positive when its orientation is counterclockwise from the center line and vice versa.

2.2. Simulated annealing algorithm

The SAA yields good results, although it is rather time-consuming. This algorithm, deriving its name from metallurgy, was first proposed by KIRKPATRICK et al. (1983). The SAA used to improve the directivity of the underwater transducer array minimizes the cost function obtained from the directivity function of the array by setting the initial temperature, final temperature, and the initial weight vector and decreasing the temperature according to certain rules. In subsequent iterations, the weight vector is updated such that the

value of cost function decreases with decreasing temperature. The algorithm prevents the value of the cost function from being trapped in a local minimum and allows it to reach a global minimum. The global minimum is derived by accepting with a certain probability the state in which the cost function increases as well and setting the current state to the state in which the cost function decreases. During the minimization process of cost function, the probability of accepting the state in which the value of the cost function increases gradually drops to zero.

The expression shows the probability of accepting a new state:

$$P\left(\Delta F(\mathbf{I})\right) = \begin{cases} e^{-\frac{\Delta F(\mathbf{I})}{T}} & \text{if} \quad \Delta F(\mathbf{I}) > 0, \\ 1 & \text{if} \quad \Delta F(\mathbf{I}) \le 0, \end{cases} \tag{7}$$

where $\Delta F(\mathbf{I}) = F_{\text{new}}(\mathbf{I}+\mathbf{I}) - F_{\text{current}}(\mathbf{I})$ is the difference between the cost functions for the new and current states and T is the temperature.

The optimization procedure by the SAA is given as:

1) *Generation of the initial solution vector*: generate the initial vector $\mathbf{I}_0$ and compute the value of the cost function.

2) *Setting of the initial temperature*: in the algorithm, the initial temperature T_0 is very important; if the temperature is too high, the system cannot converge to the minimum state, whereas if it is too low, the global minimum cannot be reached.

3) *Generation of a new solution*: at the temperature T, the new solution $\mathbf{I}_0 + \mathbf{I}$ is generated.

4) *Evaluation of the new solution*: calculate the value of the cost function for the new solution and, based on Eq. (7), accept or reject the new state according to the difference ΔF.

5) *Decrease of the temperature*: decrease the temperature so that the probability of accepting the state in which the value of the cost function increases is reduced.

6) *Repeat of the above steps*: repeat steps 2–5 until the temperature value reaches the final temperature set.

3. The side lobe level reduction method using the SAA

The proposed cost function for reducing the side lobe level to the maximum while minimizing the beam width increase of array directivity is defined as follows:

$$F(\mathbf{W}) = \alpha \left| \frac{W_d - W_I}{W_d} \right| + \beta \frac{\sum\limits_{i=1}^{M} B\left(\theta_i, \mathbf{I}\right)}{B\left(\theta_0, \mathbf{I}\right)}, \tag{8}$$

where θ_i is the position of the i-th side lobe, M is the number of side lobes, $\mathbf{I}$ is the weight vector,

W_d is the zero beam width of interest by the weight vector, and W_I is the zero beam width by the weight vector.

While the first term on the right-hand side in Eq. (8) expresses the deviation between the beam width of interest and the beam width by the weight vector, the second term is the ratio of the sum of all the side lobe levels for Eq. (1) to the main lobe level, and the cost function is divided by W_d to be dimensionless.

In Eq. (8), α and β are constants that determine the contributions of the main lobe width and side lobe level, respectively. While the classical cost function (ZANGENE *et al.*, 2014) considered only the third-order lobe level, the proposed method considers all side lobe levels. Given an array, it is impossible to simultaneously make both the side lobe level and the main lobe width small. Therefore, to optimize the lobe level while minimizing the main lobe width increase, α and β are introduced to reflect the characteristics of the side lobe level and the main lobe width, respectively. Thus, depending on whether the main lobe width or the side lobe level is considered, the values of α and β can be set differently.

In this paper, we theoretically consider the convergence of the proposed cost function for the SAA. The cost function is related to the directivity function of the arc array, which is rather complicated and can only be obtained from numerical calculations. Therefore, we use the approximate equation for the directivity function expressed by the zero-order Bessel function as (LI, 2011):

$$B(\theta) \approx \left| J_0 \left(\frac{4\pi r}{\lambda} \sin \left(\frac{\theta - \theta_0}{2} \right) \right) \right|. \tag{9}$$

Equation (9) is satisfied for the uniform arc array under the condition:

$$\frac{1}{\alpha_0} \ge \frac{2r}{\lambda} + \frac{1}{\pi}, \tag{10}$$

where λ is the signal wavelength and α_0 is the central angle between adjacent elements in the radian. Because the cost function is expressed by the Bessel function, as in Eq. (9), the proposed cost function can converge to maintain the main lobe width and simultaneously reduce the side lobe level with changing the weighting coefficients.

Let F_0 be the initial value of cost function, T_0 the initial temperature, T_e the final temperature, $\mathbf{I}_0$ the unit vector, MaxTryT the maximum number of trials at a given temperature, MaxSucT the maximum number of successes at a given temperature, MaxRej the maximum exclusion number, and trialCountT the number of trials at a given temperature. First, we find the difference $\Delta F = F_{\text{new}} - F_{\text{current}}$ between the cost functions for the new and current states by randomly changing the weight vector of the array beamforming output.

The procedure to randomly change the weight vector is as follows:

1) *Generation of random numbers*: generate a random number vector of size $N/2$.

2) *Sorting of random numbers*: sort the generated random number vectors in ascending order.

3) *Symmetrizing of the random number vector*: make a random number vector of size N, symmetrizing the sorted random number vector.

4) *Update of the weight vector*: update the weight vector by adding the random number vector to the current weight vector.

5) *Normalization of the weight vector*: normalize so that the maximum value of the weight vector is equal to 1.

Then, the cost function is updated by introducing the new solution with the transition probability by Eq. (13). This process is repeated while the trial number (trialCountT) is less than the maximum number of trials at a given temperature (MaxTryT) or the success number (sucCountT) is less than the maximum number of successes (MaxSucT) that is to be accepted as a new state. Next, the above process is repeated by changing the temperature in a certain way.

If the temperature reaches the final temperature T_e set or the number of states rejected (countRej) is greater than the maximum rejection number (MaxRej) set, the calculation is ended. The algorithm used in the simulation is shown in Fig. 2.

The solution obtained by the proposed algorithm deviates from the local minima and converges to the global minimum, thus giving good results.
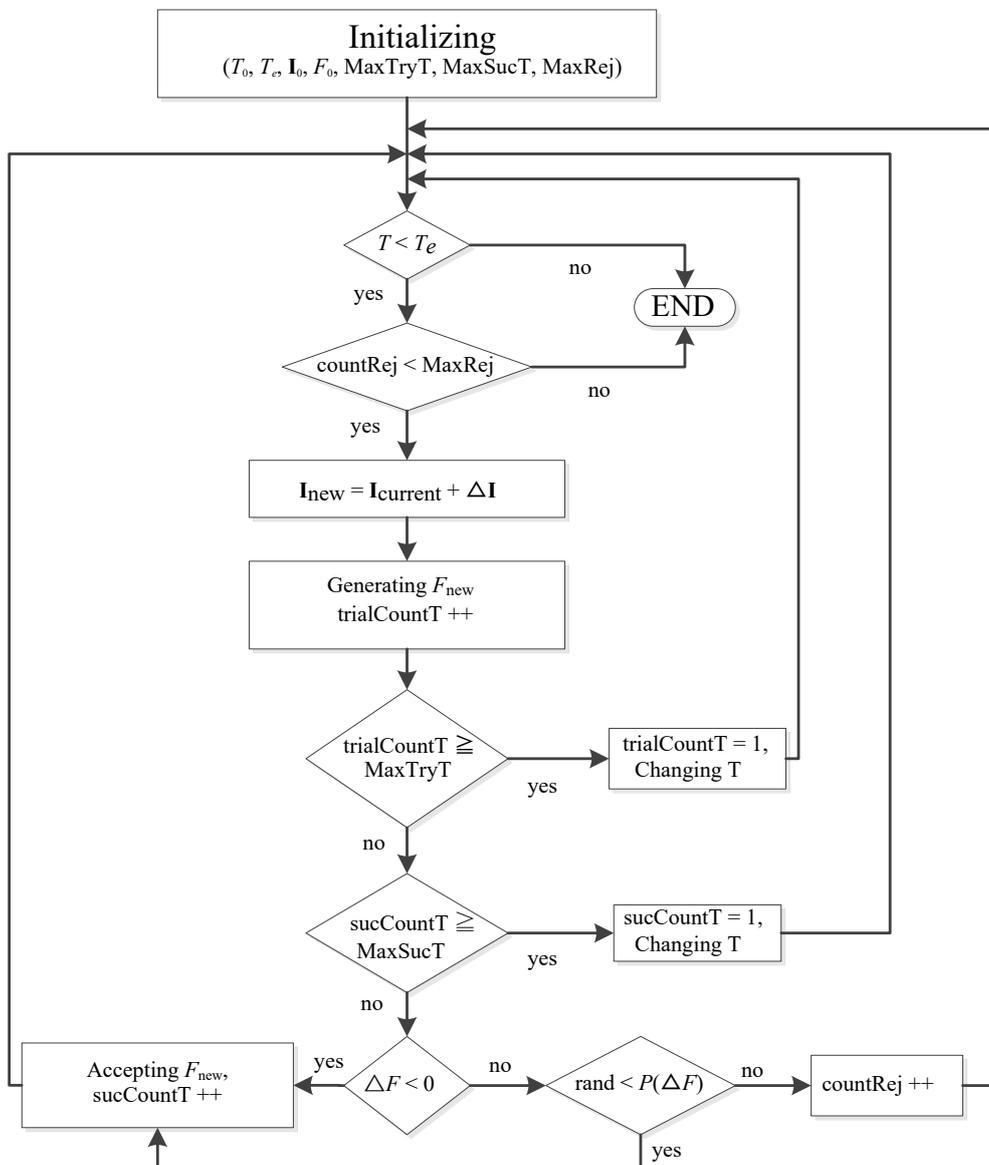

Fig. 2. Computational flow diagram using the proposed algorithm.

4. Simulation and discussion

In the simulation, all the array elements are considered to be omnidirectional. The center angle between each element is 10°, the center frequency of the signal is f_0 = 50 kHz, the speed of sound in water is c = 1500 m/s, and the radius of the arc is r = 0.08 m. For simplicity, let us consider the directivity of the array as the incidence direction of the signal to be set at 0° and the range of angles as [−60°, 60°], while the SNR is varied to −5, 0, 5, 10, 20, and 30 dB.

The initial values of parameters used in the simulation are shown in Table 1.

Figure 3 shows the convergence of the cost function and the temperature throughout the iterative process for the simulation using the proposed method. The temperature drastically reduces until 2000-th iteration and then gradually converges to 0. It can be seen that the convergence process of the cost function exhibits a local minimum around 1445-th iteration, and this is due to the nature of the SAA, which deviates from the local minimum and then converges to the global minimum.

Figure 4 shows the directivity of the arc array obtained by applying various weighting methods and compares them with the proposed method. The side lobe level of the directivity function using the proposed method is −16.6 dB, even for an SNR of −5 dB, but it is about −10 dB for the other methods. It can be seen that the proposed method is effective in side lobe level reduction and powerful over noise suppression, even at low SNRs.

Table 2 presents the weighting factor vectors calculated by the proposed method and the weighting factor values calculated by different methods. Table 3 provides a comparison of different weighting methods with the proposed method for the main lobe width and side lobe level of array directivity at different SNRs. In Table 3, the front values in each column are the 6 dB beam width in degrees, followed by the maximum side lobe levels in dB. From Table 3, it can be seen that at low SNRs below 0 dB, the proposed method minimizes the main lobe width increase, while the side lobe level is the smallest among the seven methods. Furthermore, the main lobe width and the side lobe level do not change significantly at various SNRs. This indicates that the proposed method exhibits high noise robustness.

Table 4 shows the main lobe width (6 dB beam width) and the maximum lobe level of the array directivity with varying values of α and β. From Table 4, it is evident that the main lobe width decreases and the maximum side lobe level increases as β increases.

Figure 5 shows the comparison between the proposed method and various weighting methods using the cost function for different SNRs. Only the maximum side lobe level is used. As shown, the proposed method maintains a constant cost function value, regardless of the SNR change.

Table 1. The initial values of parameters.

| Parameters	Initial values
Initial temperature T_0	1
Final temperature T_e	1e-8
Maximum number of trials at a given temperature (MaxTryT)	3000
Maximum Success at a given temperature (MaxScT)	40
Maximum rejection number (MaxRej)	1000
[α, β]	[1, 1]
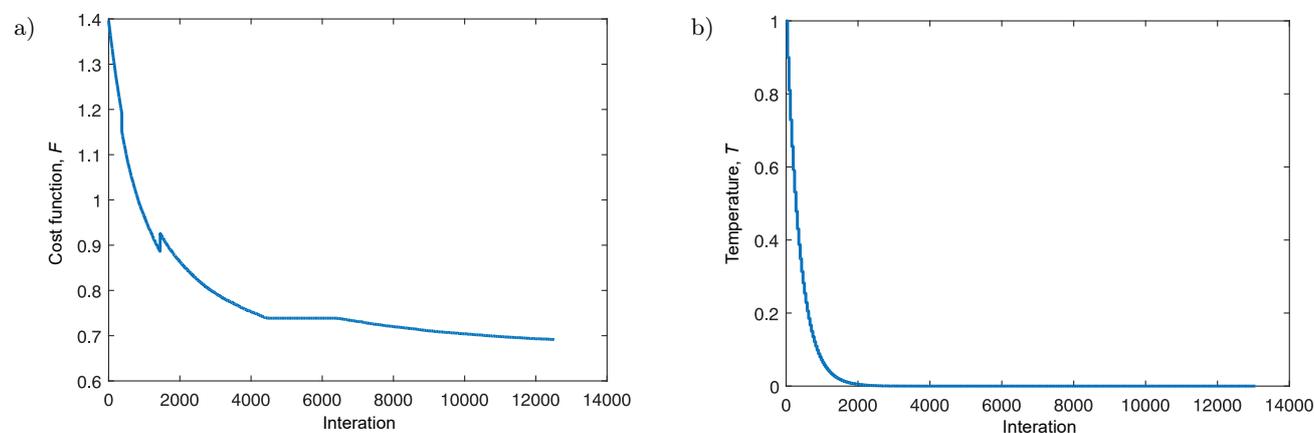
Fig. 3. Convergence of the cost function and the temperature throughout the iterative process:
a) convergence; b) temperature.
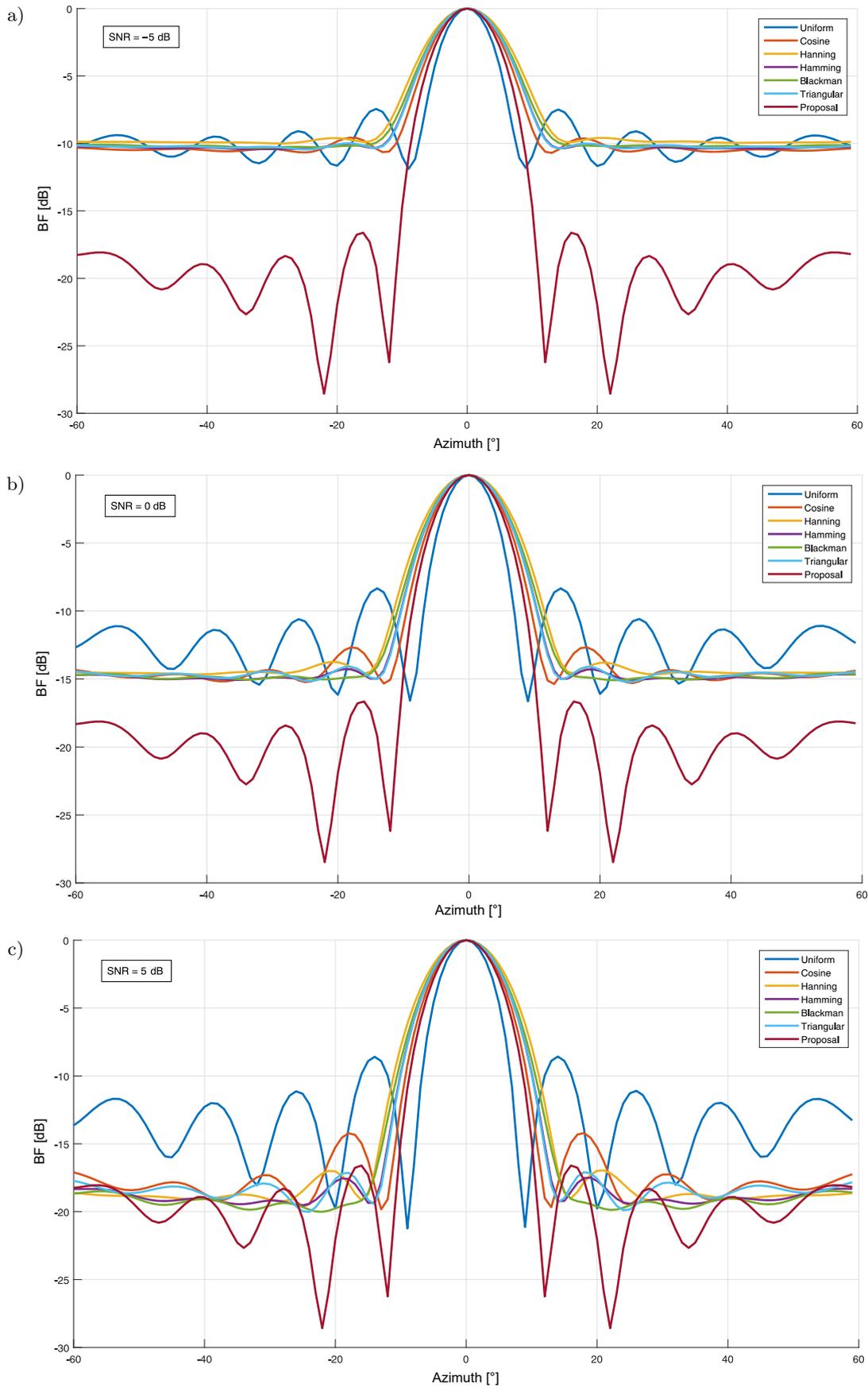
[Fig. 4abc.]
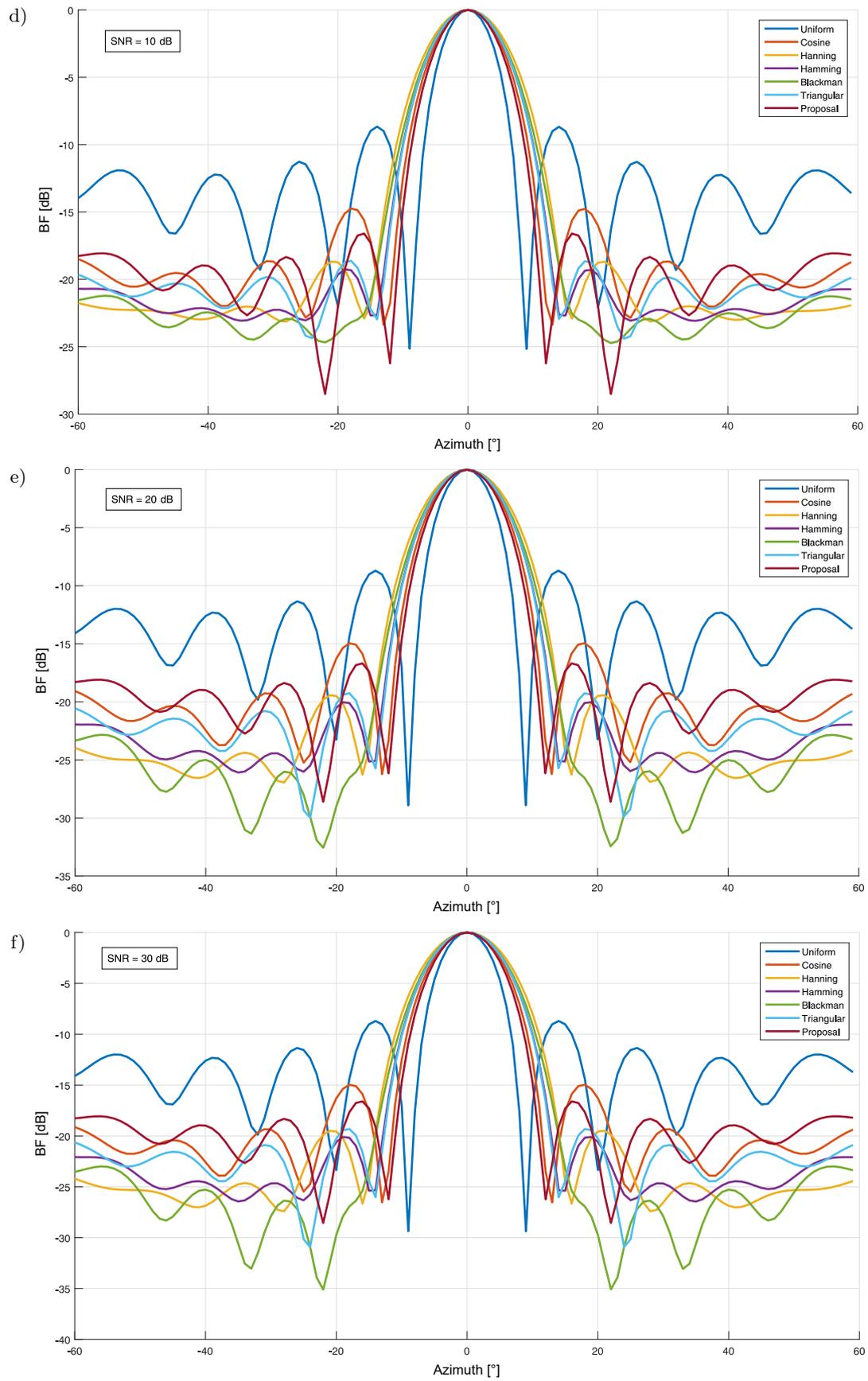
Fig. 4. Comparison between the proposed method and various weighting methods for the different SNRs:
a) −5 dB; b) 0 dB; c) 5 dB; d) 10 dB; e) 20 dB; f) 30 dB.

Table 2. Weighting vectors by various weighting methods (SNR = 10 dB).

Method	Weighting vector
Uniform	[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
Cosine	[0, 0.2817, 0.5406, 0.7557, 0.9096, 0.9898, 0.9898, 0.9096, 0.7557, 0.5406, 0.2817, 0]
Hanning	[0.0000, 0.0794, 0.2923, 0.5712, 0.8274, 0.9797, 0.9797, 0.8274, 0.5712, 0.2923 0.0794, 0.0000]
Hamming	[0.0800, 0.1530, 0.3489, 0.6055, 0.8412, 0.9814, 0.9814, 0.8412, 0.6055, 0.3489, 0.1530, 0.0800]
Blackman	[0.1200, 0.1526, 0.2799, 0.5344, 0.8560, 1.0870, 1.0870, 0.8560, 0.5344, 0.2799, 0.1526, 0.1200]
Triangular	[0, 0.1818, 0.3636, 0.5455, 0.7273, 0.9091, 0.9091, 0.7273, 0.5455, 0.3636, 0.1818, 0]
Proposal	[0.2686, 0.4147, 0.5613, 0.7073, 0.8537, 1.0000, 1.0000, 0.8537, 0.7073, 0.5613, 0.4147, 0.2686]

Table 3. Main lobe width and side lobe level for various weighting methods.

Method	SNR [dB]					
	−5	0	5	10	20	30
Uniform	11.6, −7.4	11.4, −8.3	11.2, −8.6	11.2, −8.6	11.0, −8.7	11.0, −8.7
Cosine	16.0, −9.6	15.2, −12.6	15.0, −14.2	14.8, −14.7	14.8, −15.0	14.8, −15.0
Hanning	19.2, −9.6	18.0, −13.7	17.6, −17.0	17.6, −18.6	17.6, −19.4	17.6, −19.5
Hamming	17.2, 10.0	16.6, −14.2	16.4, −17.5	16.2, −19.2	16.0, −20.0	16.0, −20.1
Blackman	18.0, −10.1	17.0, −14.8	16.6, −19.4	16.6, −22.0	16.6, −22.8	16.6, −23.0
Triangular	17.2, −9.9	16.4, −14.1	16.2, −17.1	16.2, −18.6	16.0, −19.2	16.0, −19.3
Proposal	14.0, −16.6	14.0, −16.6	14.0, −16.6	14.0, −16.7	14.0, −16.7	14.0, −16.7

Table 4. Main lobe width and maximum side lobe level for α and β (SNR = 10 dB).

[α, β]	[1,1]	[1,3]	[1,5]	[1,7]	[1,9]
Main lobe width [°]	14.0	13.8	13.6	13.5	13.4
Maximum side lobe level [dB]	−16.7	−16.5	−16.2	−15.9	−15.5

Therefore, it is obvious that the proposed method is effective in improving the directivity pattern of the array and increasing the azimuthal resolution ability by significantly reducing the side lobe level while minimizing the main lobe width increase even for low SNRs.
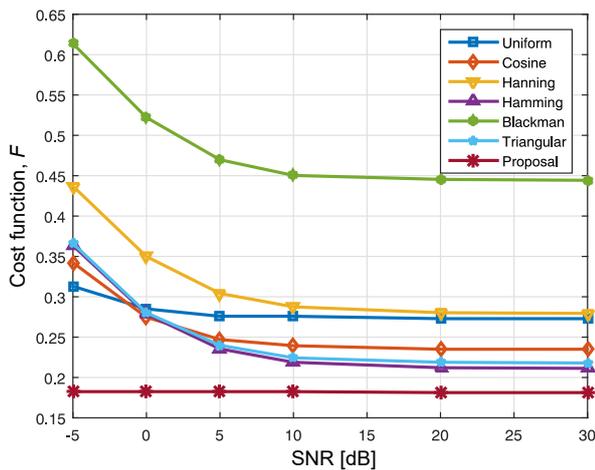

Fig. 5. Comparison of array directivity using the cost function for different SNRs.

5. Conclusion

The array weighting by the proposed method can reduce the side lobe level by about 8 dB lower than the uniform method while minimizing the main lobe width increase for low SNRs. Other methods, except the cosine method, increase the main lobe width considerably, although those may lower the side lobe level more than the proposed method. Therefore, the proposed method can effectively suppress the noise while maintaining the resolution of the underwater acoustic transducer array.

According to the evaluation of directivity by various methods based on the cost function, for the SNR equal to −5 dB, the value of the cost function by the proposed method is 1.55, 1.44, 1.77, 1.39, 2.69, and 1.41 times lower than those by the uniform method, cosine method, Hanning method, Hamming method, Blackman method, and triangular method, respectively.

For uniform arc arrays with underwater acoustic transducers, it is confirmed that array element weighting with weight vectors obtained by using the proposed cost function and simulated annealing based on weight factor updating can effectively reduce the side lobe level while minimizing the main lobe width increase of array directivity even at low SNRs.

References

1. ALBAGORY Y., ALRADDADY F. (2021), An efficient approach for side lobe level reduction based on recursive sequential damping, *Symmetry*, **13**(3): 480, doi: 10.3390/sym13030480.

2. Cardone G., Cincotti G., Pappalardo M. (2002), Design of wide band arrays for low side-lobe level beam patterns by simulated annealing, [in:] *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **49**(8): 1050–1059, doi: 10.1109/TUFFC.2002.1026017.

3. Chen W., Gu Z., Ma X., Zhang S., Han Z., Zhong Y. (2019), Robust optimization of vibro-acoustics system based on an elliptical basis function neural network, *Applied Acoustics*, **145**: 41–51, doi: 10.1016/j.apacoust.2018.09.013.

4. Cretu N., Pop M.I., Rosca I.C. (2010), Acoustic design by simulated annealing algorithm, *Physics Procedia*, **3**(1): 489–495, doi: 10.1016/j.phpro.2010.01.064.

5. Dessouky M., Sharshar H., Albagory Y. (2006), Efficient side lobe reduction technique for small-sized concentric circular arrays, *Progress In Electromagnetics Research – PIER*, **65**: 187–200, doi: 10.2528/PIER06092503.

6. Dessouky M., Sharshar H., Albagory Y. (2007), An approach for Dolph-Chebyshev uniform concentric circular arrays, *Journal of Electromagnetic Waves and Applications*, **21**(6): 781–794, doi: 10.1163/156939307780749075.

7. Gintaras P., Tomkevičius A., Ostreika A. (2019), Hybridizing simulated annealing with variable neighborhood search for bipartite graph crossing minimization, *Applied Mathematics and Computation*, **348**: 84–101, doi: 10.1016/j.amc.2018.11.051.

8. Hong G., Zuckermann M., Harris R., Grant M. (1991), A fast algorithm for simulated annealing, *Physica Scripta*, **1991**(T38): 40–44, doi: 10.1088/0031-8949/1991/T38/010.

9. Kirkpatrick S., Gelatt Jr. C.D., Vecchi M.P. (1983), Optimization by simulated annealing, *Science*, **220**(4598): 671–680, doi: 10.1126/science.220.4598.671.

10. Li H., Liu Y., Sun G., Wang A., Liang S. (2017), Beam pattern synthesis based on improved biogeography-based optimization for reducing sidelobe level, *Computers and Electronical Engineering*, **60**: 161–174, doi: 10.1016/j.compeleceng.2017.01.003.

11. Li Q. (2011), *Digital Sonar Design in Underwater Acoustics: Principles and Applications*, pp. 238–242, Zhejiang University Press, Hangzhou.

12. Nofal M., Aljahdali S., Albagory Y. (2013), Tapered beamforming for concentric ring arrays, *AEU – International Journal of Electronics and Communications*, **67**(1): 58–63, doi: 10.1016/j.aeue.2012.06.005.

13. Rasdi Rere L.M., Mohamad I.F., Aniati M.A. (2015), Simulated annealing algorithm for deep learning, *Procedia Computer Science*, **72**: 137–144, doi: 10.1016/j.procs.2015.12.114.

14. Rucksana Begum S., Ramarao G. (2015), Synthesis of non uniform linear arrays using Dolph-Chebyshev polynomial by reducing side lobe level based on modulating parameter array factor, *International Journal of Advance Research in Science and Engineering*, **4**(8): 65–72.

15. Sarker M.A., Hosaain M.S., Masud M.S. (2016), Robust beamforming synthesis technique for low side lobe level using taylor excited antenna array, [in:] *Proceedings of the 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering*, doi: 10.1109/ICECTE.2016.7879566.

16. Schmerr Jr. L.W. (2015), *Fundamentals of Ultrasonic Phased Arrays*, Springer Cham, pp. 60–65.

17. Singh U., Salgotra R. (2018), Synthesis of linear antenna array using flower pollination algorithm, *Neural Computing & Applications*, **29**: 435–445, doi: 10.1007/s00521-016-2457-7.

18. Van Luyen T., Vu Bang Giang T. (2017), Interference suppression of ULA antennas by phase-only control using bat algorithm, [in:] *IEEE Antennas and Wireless Propagation Letters*, **16**: 3038–3042, doi: 10.1109/LAWP.2017.2759318.

19. Zangene A., Dalili Oskouei H.R., Nourhoseini M. (2014), Reduction of side lobe level in non-uniform circular antenna arrays using the simulated annealing algorithm, *Journal of Electrical and Electronics Engineering Research*, **6**(2): 6–12, doi: 10.5897/JEEER2014.0504.