eISSN 2300-262X (online)

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

ARCHIVES of ACOUSTICS

QUARTERLY

Vol. 48, No. 1, 2023

WARSAW

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

ARCHIVES of ACOUSTICS

QUARTERLY, Vol. 48, No. 1, 2023

Research Papers

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

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

Indexed and abstracted in Science Citation Index Expanded (also known as SciSearch®)
and Journal Citation Reports/Science Edition
Abstracted in Acoustics Abstracts, in Applied Mechanics Reviews
and indexed in Journal of the Acoustical Society of America
Recognised by the International Institute of Acoustics and Vibration, I IAV

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

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

Archives of Acoustics Vol. 48, No. 1, pp. 3–12 (2023), doi: 10.24425/aoa.2022.142906

Research Paper

Sleep Snoring Sound Recognition Based on Wavelet Packet Transform

Li DING[(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@163.com

(*received February 15, 2022; accepted August 31, 2022*)

Snoring is a typical and intuitive symptom of the obstructive sleep apnea hypopnea syndrome (OSAHS), which is a kind of sleep-related respiratory disorder having adverse effects on people's lives. Detecting snoring sounds from the whole night recorded sounds is the first but the most important step for the snoring analysis of OSAHS. An automatic snoring detection system based on the wavelet packet transform (WPT) with an eXtreme Gradient Boosting (XGBoost) classifier is proposed in the paper, which recognizes snoring sounds from the enhanced episodes by the generalization subspace noise reduction algorithm. The feature selection technology based on correlation analysis is applied to select the most discriminative WPT features. The selected features yield a high sensitivity of 97.27% and a precision of 96.48% on the test set. The recognition performance demonstrates that WPT is effective in the analysis of snoring and non-snoring sounds, and the difference is exhibited much more comprehensively by sub-bands with smaller frequency ranges. The distribution of snoring sound is mainly on the middle and low frequency parts, there is also evident difference between snoring and non-snoring sounds on the high frequency part.

Keywords: snoring recognition; wavelet packet transform; feature selection; machine learning.

1. Introduction

The obstructive sleep apnea hypopnea syndrome (OSAHS) is a chronic sleep-related disease affecting the general adult population ranging from 6% to 17% (Senaratna *et al.*, 2017), which is characterized by intermittently partial or complete collapse of the upper airway, resulting in frequently sleep-disordered breathing events (SDB). This kind of disease greatly affects the quality of life and even is an independent risk factor for diseases such as neurocognitive dysfunction, arterial hypertension, metabolic disorders, and cerebrovascular disease (Wang *et al.*, 2017; Hui *et al.*, 2015; Dafna *et al.*, 2013). The traditional and golden standard for clinically diagnosing OSAHS is Polysomnography (PSG) (Jiang *et al.*, 2020) with multiple sensors that must be directly connected to the body to monitor serious biological signals during sleep. However, the complex equipment, professional technicist, time-consuming process, and expensive cost limiting its wide application, makes OSA a significant but underestimated threat to public health (Ayas, 2013). An inexpensive and reliable technology to diagnose OSAHS is urgently needed. Studies have indicated that snoring is a typical and intuitive symptom of OSAHS (Dafna *et al.*, 2013; Hui *et al.*, 2015; Jiang *et al.*, 2020; Ng *et al.*, 2008; Senaratna *et al.*, 2017; Wang *et al.*, 2017) reported in more than 80% of OSAHS patients (Kapur *et al.*, 2002; Young *et al.*, 1997), which has been reported to be a potential method to monitor OSAHS. It is a kind of sleep-related noise caused by oscillations of the soft tissue structures in the upper airways (Lechner *et al.*, 2019) because of a reduction of the muscle tone and slackening of soft tissue narrowing down the upper airways.

Automatic extracting snoring episodes from recorded sleep sounds throughout the night, including breathing, speaking, and other noises, is the first but

the most important step during the whole process of analyzing snoring sounds, which has been studied by many studies (Dafna *et al.*, 2013; Hwang *et al.*, 2015; Lim *et al.*, 2019; Ng *et al.*, 2008; Nonaka *et al.*, 2016; Wang *et al.*, 2017). Most of their studies focused on differencing snoring sounds and non-snoring sounds from acoustic features derived from the time domain, frequency domain, and time-frequency domain. Specifically, Jiang *et al.* (2020) designed a snoring sound detection system based on a non-contact microphone that extracted 127-dimensional time and spectral features to obtain an accuracy of 98.2% in the validation group. The work of Ng *et al.* (2008) achieved a high accuracy in the classification of snoring and non-snoring sounds using formant frequencies. Cavusoglu *et al.* (2007) implemented the snoring and non-snoring classification by the sub-band energy of sound episodes and robust linear regression with an accuracy of 90.2% from the combined dataset of 18 simple snorers and 12 OSAHS patients. Nonaka *et al.* (2016) developed the human auditory image model to extract snoring sounds automatically. These studies mainly focused on the low frequency part or some specific frequency bands rather than analysis of all sub-band, which might ignore the information on the high frequency part.

The wavelet transform (WT) is another way to divide the signal into sub-bands with different frequency ranges, which has been demonstrated to be effective in the speech signal and electroencephalogram processing (Li, Zhou, 2016; Wu *et al.*, 2008; Wang *et al.*, 2020). Wang *et al.* (2020) proposed a novel method of speaker-independent emotion recognition based on the wavelet packet analysis, which performed better than frequency features. Li and Zhou (2016) implemented the classification of electrocardiograms using the wavelet packet entropy and random forests. Wu *et al.* (2008) extracted features of electroencephalogram signals such as the energy of special sub-bands and corresponding coefficients of the wavelet packet decomposition, which had maximal separability according to the Fisher distance criterion. Qian *et al.* (2016; 2017) adopted energy features derived from the wavelet packet transform (WPT) to discriminate snoring sounds from different snoring sites with much better performance than features derived from time and frequency domains. These works indicate that wavelet transform works effectively in the analysis of biological signals.

To explore the relationship between snoring and non-snoring sounds on different frequency bands, an automatic snoring detection system based on WPT features with an XGBoost (Chen, Guestri, 2016) classifier was proposed in this study. The system includes three major steps. Firstly, the recorded sleep-related sounds were enhanced and segmented by a generalization subspace noise reduction algorithm, and signal presence probability based on energy, respectively. Then, WPT features from different wavelet functions and decomposition layers were extracted from segmented sound episodes and selected based on a series of correlation analyses. Finally, snoring sounds were detected from the trained XGBoost classifier. The contribution of the work incorporates: 1) it used WPT to extract sub-band features and yielded comparable accuracy in recognizing snoring sounds compared with existing related studies (Adesuyi *et al.*, 2022; Arsenali *et al.*, 2018; Jiang *et al.*, 2020; Sun *et al.*, 2022); 2) it discovered that signal would be exhibited much more comprehensively by sub-bands with smaller frequency ranges. And the difference between snoring and non-snoring sounds is getting more evident with the frequency range getting smaller, which is more beneficial for classifying; 3) it demonstrated that although the distribution of snoring sound is mainly on the low frequency part, the information on the high frequency part also cannot be ignored, where there is also evident difference between snoring and non-snoring sounds.

2. Material and methods

2.1. Data acquired

In this study, 24 subjects composed of simple snorers and OSAHS patients were selected from the First Affiliated Hospital of Guangzhou Medical University. All subjects have been informed and agreed with the monitoring process during the whole night. The detailed information about these subjects was described in Table 1. During sleeping, a microphone (RODE, NTG-3, Sydney, Australia) and a digital audio recorder (Rowland, Edirol R-44, Japan) were placed approximately 45 cm above the patient's mouth and nose to record the original sleep sound signals for approximately seven hours, with a sampling rate of 44.1 kHz and 16-bit resolution. PSG equipment (Alice-5, Pittsburgh, Pennsylvania, USA) was simultaneously used to monitor the subject's PSG signals.

Table 1. Information (gender, age, apnea/hypopnea (AHI), and body mass index (BMI)) of the subject.

	Simple	Mild	Moderate	Severe
Gender (M/F)	4/0	3/2	5/3	5/2
Age (years)	25 ± 5	35 ± 5.05	46.6 ± 11.58	49.9 ± 9.36
AHI	3.8 ± 0.73	10.8 ± 4.13	21.69 ± 3.46	36.77 ± 3.83
BMI	27.5 ± 4.57	30.24 ± 0.87	35.09 ± 1.05	39 ± 1.48

2.2. Pre-processing

During the process of recording sleep-related sounds, the noncontact nature of signal acquisition is often susceptible to external noise distortion. The additive background noise is inevitably superposed to a snoring sound, which will affect the fidelity of signals. Most studies (Dafna *et al.*, 2013; Jiang *et al.*, 2020; Lim *et al.*, 2019; Wang *et al.*, 2017) conducted the noise reduction process before its further analysis effectively. In the work of Karunajeewa *et al.* (2008), different enhancement algorithms were implemented to yield different snoring recognition results. Different from common noise suppression methods, generalized subspace snoring signal enhancement based on the noise covariance matrix estimation was implemented (Ding *et al.*, 2021). It was verified by our previous work that this algorithm could well update noise in real-time by recursive averaging its past values adjusted by a time-varying smoothing parameter controlled by the snoring signal presence probability during the noise suppression process. Objective quality measurements and the spectrum analysis demonstrated that this method could reduce most background noise with less signal distortion. Moreover, the enhanced snoring signal was detected and segmented by the signal presence probability, which is determined by the ratio between the local energy of the noisy signal and its minimum within a specified time window to detect the sound episode.

Figure 1 shows the process of pre-processing including noise reduction and episode segmentation. The

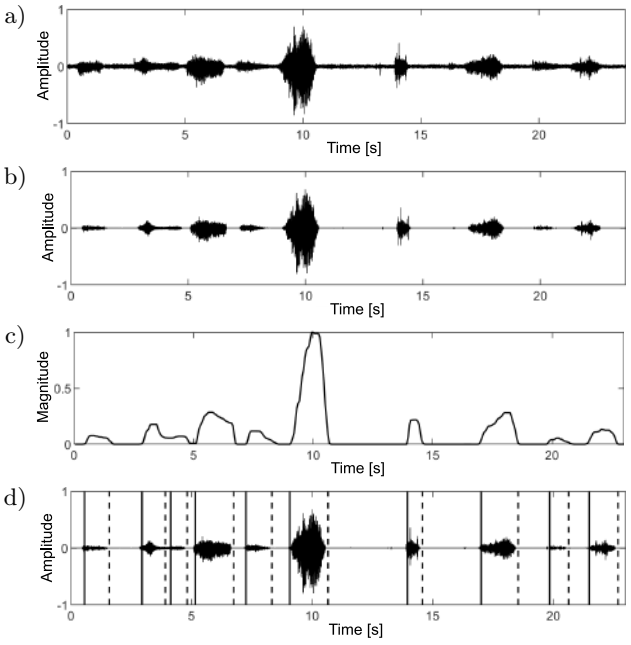


Fig. 1. Example of sound enhancement and detection: a) the original recorded noisy sound; b) the enhanced sound by subspace noise reduction algorithm; c) the signal presence probability of enhanced recorded signal; d) the detection result of sound episodes.

segmented episodes were further labeled as snoring sounds and non-snoring sounds based on PSG signals by ear-nose-throat (ENT) experts. 26561 labeled sound episodes including 17704 snoring sounds and 8857 non-snoring sounds were obtained from all 24 subjects. All labeled sounds were randomly divided into the training and test sets with proportions of 70% and 30%, respectively.

2.3. Feature extraction

The wavelet packet decomposition was applied to divide the signal into sub-bands with different frequency bands. Acoustic features including wavelet packet coefficients, log energy, Shannon entropy, wavelet transform cepstral coefficient, and sound pressure level based on sub-band signals were extracted for further analysis. All signals were framed by the hamming window function with 20 ms frame length and 50% overlap. Amplitude normalization was conducted to eliminate the influence of sound intensities. The statistic functions including the mean and variance of all frames in each signal were calculated to represent each signal. Table 2 shows the detailed information on the features.

Table 2. Information of the extracted features.

Feature	Description	Dimension (layer 4/layer 5/ bark sub-band)
Coefficient	Mean, variance value	32/64/34
Log-energy	Mean, variance value	32/64/34
Shannon entropy	Mean, variance value	32/64/34
Sound pressure level	Mean, variance value	32/64/34
Wavelet transform cepstral coefficient	Mean, variance value	26/26/26
Mel-frequency cepstral coefficient	Mean, variance value	26/26/26

2.3.1. Wavelet packet model

The WT (Sharma *et al.*, 2020) is a typical method to transform the time-domain audio signal into a time-frequency domain consisting of the continuous wavelet transform and discrete wavelet transform. The WT of the signal x at the time y and scale z is defined by the inner product with a wavelet function:

$$W_f(y,z) = \langle x, u_{y,z} \rangle = \frac{1}{\sqrt{y}} \int_{-\infty}^{\infty} f(t) u^* \left(\frac{(t-z)}{y} \right) \mathrm{d}t, \quad (1)$$

where $u(t)^*$ represents the complex conjugate of the wavelet function $u(t)$. WPT applies the transform

step on all frequency bands. It is calculated through time-domain filtering with a sub-signal representation obtained from frequency components with each sub-band. Figure 2 shows an integrated wavelet packet tree of a signal. The original signal is first decomposed into two sub-bands in the first decomposition layer: the low frequency part-1 (1, 0) and the high frequency part-2 (1, 1). Then the low frequency part-1 and part-2 will be further decomposed with increasing decomposition layers to obtain sub-bands with much finer frequency bandwidth. The frequency bandwidth of the k-th sub-band in the j-th decomposition layer is $\frac{fs}{2^{j+1}}$ Hz; fs is the sampling rate with a value of 44.1 kHz in this work. With the increase of the decomposition layer, the finer the frequency is decomposed with much more sub-bands. There are $N_j = 2^j$ sub-bands in the j-th layer. There are 16 sub-bands in layer 4 with a bandwidth of 1378 Hz and 32 sub-bands in layer 5 with a bandwidth of 689 Hz. Moreover, we constructed bark sub-bands from decomposition layers 4 and 5 of WPT with a dimension of 17, which is based on auditory characteristics of humans. The detailed composition of bark sub-bands is shown in the last layer of Fig. 2.

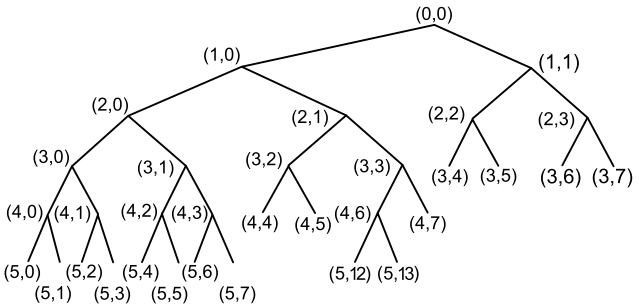


Fig. 2. Decomposition process of WPT and the construction of wavelet bark.

In the work of Monson *et al.* (2014) "high frequency" referred to a frequency above about 5 kHz which has traditionally been neglected in speech research. The "middle and low frequency" is defined below 5 kHz in this paper. As Fig. 2 shows, the bark sub-band is constructed by sub-bands in layer 4 and layer 5 which is considered a common division method for audio signal processing that accurately matches the human ear's auditory perception chrematistics (Korniienko, Machusky, 2018). The bark and layer 5 decomposition structure have the same sub-band distributions in the middle and low frequency part (0–5.5 kHz) composed by (5,0)–(5,7), while layer 5 has much finer sub-bands in the high frequency. Studies have indicated that the energy of snoring sounds is mainly concentrated below 2 kHz (Perez-Padilla *et al.*, 1993). The frequency range of the first sub-band at layer 4 is 0–1378 Hz that most snoring information locate in this frequency band. To explore the influence of the frequency bandwidth of the sub-band on the classification result, and the difference between snoring and non-snoring sounds in the middle and low frequency part and high frequency part, acoustic features extracted from layer 4, layer 5, and bark sub-bands are discussed in the work.

2.3.2. Wavelet packet coefficients

The coefficients by WPT can reveal the local characteristics of signals. The mean values of the coefficients of the k-th sub-band in the j-th layer are described as:

$$w_{j,k} = \frac{\sum\limits_{n} v_{j,k,n}}{N_k}, \quad n = 1, 2, ..., N_k, \;\; k = 1, 2, ..., 2^j, \quad (2)$$

where N_k is the number of the coefficient component in the k-th sub-band with the value of 882; $v_{j,k,n}$ represents the n–th coefficient component of the k-th sub-band in the j-th decomposition layer. There are $N_j = 2^j$ coefficients in the j-th decomposition, which are $w_{j,1}, w_{j,2}, ..., w_{j,2^j}$.

2.3.3. Log-energy

The log-energy of the k-th sub-band signal in the j-th level can be calculated by:

$$\log E_{j,k} = \sum_{n=1}^{N_k} v_{j,k,n}^2. \quad (3)$$

2.3.4. Shannon entropy

The probability of the n-th coefficient at its corresponding node can be calculated by:

$$p_{j,k,n} = \frac{E_{j,k,n}}{E_{j,k}} = \frac{v_{j,k,n}^2}{\sum\limits_{n=1}^{N_k} v_{j,k,n}^2}, \quad (4)$$

moreover, Shannon entropy (SE) is defined by the probability distribution of energy $p_{j,k,n}$ as Eq. (4), which is a measure of uncertainty associated with random variables in information theory:

$$\mathrm{SE}_{j,k} = -\sum_{n=1}^{N_k} p_{j,k,n} \cdot \log\left(p_{j,k,n}\right). \quad (5)$$

2.3.5. Wavelet transform cepstral coefficient

WPT can be treated as a filter to divide the frequency to some sub-bands with the equal bandwidth, just like Mel-filter. Then the cepstral coefficient of the signal after the WPT filter can be calculated using discrete cosine transform (DCT), which is called the wavelet transform cepstral coefficient (WTCC):

$$\mathrm{WTCC}_m = \sqrt{\frac{2}{N_j}} \sum_{j=0}^{N_j} \log\left(E_{j,k}\right) \cos\left(\frac{\pi m\left(2j-1\right)}{2N_j}\right), \quad (6)$$

where N_j is the number of sub-bands in the j-th layer; m indicates the m-th DCT spectral line, which was set as 13 in the work.

2.3.6. Mel-frequency cepstral coefficients

The study (Jiang *et al.*, 2020) has indicated that there are obvious differences between snoring and non-snoring sounds via Mel-frequency cepstral coefficients (MFCC). The MFCC of the original signal with 13-dimension is also extracted for snoring sound recognition.

2.4. Wavelet function

As described in Eq. (1), the WPT is based on the wavelet function. Different wavelet functions may result in different WPT features. There are many wavelet function families such as BiorSplines, Coiflets, Daubechies, Symlets, and so on. The Daubechies wavelet family (DB1, DB2, DB3, DB4, DB5, DB6, DB7, DB8, DB9, DB10) has been widely used in the processing of speech and other biological signals. Also, the work of Qian *et al.* (2017) indicated that the Daubechies function (DB3, DB10) performed better on recognition of a snoring site. In this paper, we explore their performance in differentiating snoring and non-snoring sounds.

3. Classification and result

3.1. Classification model

In the study, XGBoost classifier was adopted in this study. XGBoost is an improved algorithm with good performance and high efficiency based on the gradient boosting decision that can construct boosted trees efficiently and operate in parallel. The core of the algorithm is the optimization of the value of the objective function (Torlay *et al.*, 2017). The parameter of the XGBoost classifier is essential for classification performance. Based on a 10-fold cross-validation of the training set, the optimal parameter was obtained. The number of base trees was set as 400, the max depth of trees was 6, and the learning rate was 0.3. Other parameters were set as the default value of XGBoost in Scikit-learn (Pedregosa *et al.*, 2011).

3.2. Feature selection

Feature selection is a vitally important step during the classification task because it can reduce the redundancy of features to improve the robustness of the model and reduce the computation complexity. In this paper, feature selection based on the correlation analysis is conducted to select distinguishing features. Two Pearson correlation coefficients were calculated including correlation coefficients between features and their related labels with a value of P1 and correlation coefficients among features with a value of P2. Features with high correlation with labels and low correlation with other features were selected by thresholds a and b, respectively. There were two steps for feature selection. Firstly, features were reserved if P1 was higher than a. Then, the reserved features were dropped out if P2 was higher than b to obtain relatively independent features. To fully make use of the limited dataset, the 10-fold cross-validation was used in the training set to optimize the model and select features. The threshold a and b were obtained by experiment to set as 0.8 and 0.7, respectively. Moreover, the effect of the decomposition levels and wavelet functions on the classification of snoring and non-snoring sound is explored.

3.3. Model evaluation

To evaluate the performance of the proposed recognition system of snoring sound, evaluating indexes such as sensitivity, accuracy, precision, and F1 score are expressed as follows:

$$\text{Accuracy} = \frac{(\text{TP}+\text{TN})}{(\text{TP}+\text{FP}+\text{TN}+\text{FN})}, \tag{7}$$

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP}+\text{FN})}, \tag{8}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP}+\text{FP})}, \tag{9}$$

$$\text{F1} = \frac{2\text{Precision}\cdot\text{Sensitivity}}{(\text{Precision}+\text{Sensitivity})}, \tag{10}$$

where TP represents the number of snoring sounds classified as snoring sounds (true positive), TN is the number of non-snoring sounds truly detected as non-snoring sounds (true negative), FP represents the number of non-snoring sounds falsely recognized as snoring sounds (false positive), and FN is the number of events corresponding to the false detection of snoring sound as non-snoring sound (false negative).

3.4. Classification results

Figure 3 shows the distribution of coefficient 1 and WTCC13 which have the first and second highest coefficients with labels selected by correlation analysis. 5000 samples were randomly selected from 24 subjects in the training set. It shows that the distribution of coefficient 1 and WTCC13 of snoring sounds is different from non-snoring sounds, which could distinguish snoring sounds to a certain extent. Figure 4 shows the overall accuracy of snoring and non-snoring sounds with different wavelet functions under different decomposition layers. The WPT features extracted from different decomposition layers and different Daubechies wavelet functions could work well with accuracies more than 94%.

It can be seen from Fig. 4a that the overall accuracies of WPT features extracted from layer 5 are slightly 0.5 percentage points higher than accuracies

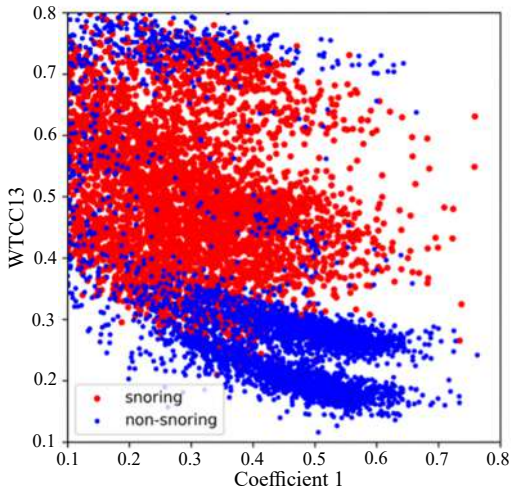


Fig. 3. The distribution of coefficient 1 and WTCC13 of decomposition layer 5 randomly selected from the training set of 5000 samples. The coefficient 1 and WTCC13 are the coefficient and WTCC of the first sub-band and thirteenth sub-band components. Snoring and non-snoring segments are denoted by red circle and blue circle symbols, respectively.

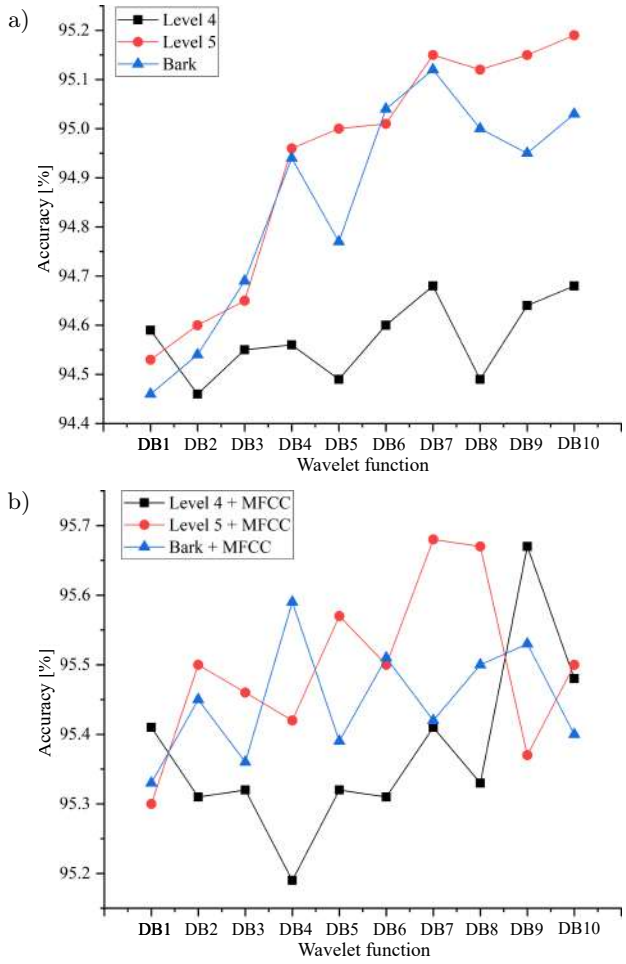


Fig. 4. The snoring recognition result of two kinds of feature sets under decomposition layer 4, 5, and bark and wavelet function DB1 to DB10 in Daubechies family: a) WPT feature set; b) MFCC and WPT combined feature sets.

of layer 4 under most wavelet functions. With the increase of the decomposition level, the frequency of the signal is divided into smaller bandwidths to obtain much more detailed information about signals, which is beneficial for distinguishing snoring and non-snoring sounds, and yield much better performance. Comparing the recognition rate of layer 5 and bark sub-bands, there is little difference between these two decomposition structures. However, layer 5 and the bark sub-band have the same decomposition construction on low frequencies (0–5.5 kHz), and layer 5 has much finer division than the bark sub-band on high frequencies (5.5–44.1 kHz). In other words, the bark sub-band puts much more emphasis on low frequency. The result shows that the energy of snoring sound and the difference between snoring and non-snoring are mainly on low frequencies. The information on the high frequencies part also cannot be ignored.

It also can be observed from Fig. 4a that the wavelet function also influences the final classification result. The overall accuracies are different between wavelet functions in the same decomposition layer. DB7, DB9, and DB10 in level 5, and DB7 in bark yield much higher recognition for WPT features among all test Daubechies wavelet functions, which are 95.2% approximately, indicating that the wavelet function plays an important role in the decomposition of the signal. Comparing Fig. 4, the WPT and MFCC combined features yielded accuracy with an average value of 95.5%, much better than simple WPT features in terms of overall accuracies under all test wavelet functions and decomposition levels. The difference in recognition results of the MFCC and WPT combined features between decomposition levels and wavelet functions is not as obvious as simple WPT features. The first three recognition rates are DB7 in level 5, DB8 in level 5, and DB9 in L4 which are 95.68%, 95.67%, and 95.65% respectively under all test conditions.

Table 3 shows detailed results for snoring and non-snoring recognition of different feature sets including MFCC, WTCC, WPT features, and WPT + MFCC combined features under the selected DB7, DB8, DB9, and DB10 wavelet functions in decomposition layer 5 and bark sub-bands. It can be known from Table 3 that for WPT features, the accuracy of the three kinds of selected feature sets is comparable, which is around 95.2%. However, the features extracted from DB7-Level 5 yield much higher sensitivity with 96.94%, indicating a higher probability of real snoring sound being recalled. For WPT – MFCC combined feature sets, the features extracted from level 5 with the wavelet function DB7 achieved the best performance during all test conditions considering sensitivity, precision, and F1-score, which are 97.27%, 96.48%, and 96.88%, respectively. Compared with MFCC, simple WPT features, the WPT – MFCC features performed best, and there is an average improvement of 1 percent-

Table 3. Detailed classification results and dimensions of selected features of different feature sets including Cepstral coefficients features, WPT features, and WPT + MFCC combined features under the selected wavelet functions and decomposition layers selected from Fig. 4.

Feature-set	Number of features (total)	Accuracy [%]	Sensitivity [%]	Precision [%]	F1 [%]	AUC
Cepstral coefficients						
MFCC	26	94.76	96.58	95.83	96.20	0.98
WTCC	26	93.57	96.17	94.58	95.37	0.98
WPT						
DB7-L5	69 (256)	95.15	96.94	96.21	96.50	0.99
DB9-L5	82 (256)	95.15	96.81	96.18	96.49	0.99
DB10-L5	81 (256)	95.20	96.81	96.23	96.52	0.99
WPT + MFCC						
DB7-L5	98 (282)	95.68	97.27	96.48	96.88	0.99
DB8-L5	96 (282)	95.67	97.44	96.30	96.87	0.99
DB9-L4	99 (162)	95.65	97.27	96.27	96.77	0.99

age points and 0.5 percentage points for WPT – MFCC combined features in terms of overall accuracy, recall, precision, and F1-score mentioned in the work. The dimensions of selected features based on the 10-fold cross-validation are also displayed in Table 3. Features with little contribution to the classification are dropped by the selection technology based on correlation coefficients. Based on the aforementioned discussion, the features extracted from the wavelet function DB7 in level 5 decomposition performed better considering simple WPT features and WPT – MFCC combined features with dimensions of 69 and 98, respectively.

We also compared two kinds of cepstral coefficients MFCC and WTCC, which are derived from the Mel-frequency filter and the wavelet packet transform filter respectively. The MFCC outperformed the WTCC with an average improvement of 1 percentage points in terms of evaluation standards mentioned in the work, which means that the Mel-frequency could carry more important information on the upper airway structure variations than wavelet packet transform does.

4. Discussion

In this work, we proposed a novel system to automatically extract snoring sounds from the recorded sounds during sleep based on WPT features. Based on wavelet packet transform, the snoring sound was decomposed into different sub-bands with the same bandwidth and different frequency ranges. Results of WPT features indicated that the information on snoring sounds and the difference between snoring and non-snoring sounds were mainly in the middle and low frequency. With increasing decomposition layer, the signal was decomposed with much smaller sub-bands, and the difference between snoring and non-snoring sounds was much more obviously accompanied by a higher classification accuracy. The snoring sound detection is the first but vital step during the whole analysis system of snoring sounds. Many studies have detected snoring episodes from different kinds of domains.

In previous studies (Han *et al.*, 2006; Karunajeewa *et al.*, 2011; Qian *et al.*, 2015; Solà-Soler *et al.*, 2007; Sun *et al.*, 2022), acoustic features extracted from frequency sub-band of the signal have been demonstrated effectively and widely used in classifying snoring and non-snoring sounds. Qian *et al.* (2015) used the 1000 Hz sub-band features, and power ratio to detect snoring sound segments. Cavusoglu *et al.* (2007) explored the sub-band energy distribution of snoring and non-snoring segments by dividing the 0–7500 Hz frequency range into 500 Hz sub-bands, which yielded 90.2% accuracy for simple snorers. These works indicated that the information distribution of snoring sounds is different among sub-bands, which mainly focus on middle and low frequency parts. Moreover, since the bark sub-bands focus on the low frequency part and are sparse in the high frequency part. The WPT furtherly divided the high frequency part based on bark sub-bands which makes the difference in the high frequency much more obviously. And recognition accuracies of layer 5 are slightly higher than bark sub-bands under most wavelet function test conditions. Although the distribution of snoring is mainly concentrated in the middle and low frequency parts, the information of snoring sounds in high frequency part also cannot be ignored. And with increasing of decomposition layer, the difference between snoring and non-snoring sounds is getting obvious, because the signal will be exhibited much more comprehensively by sub-bands with smaller frequency ranges.

The results of works (Cavusoglu *et al.*, 2007; Duckitt *et al.*, 2006; Emoto *et al.*, 2018; Jiang

et al., 2020; Lim *et al.*, 2019; Qian *et al.*, 2015; Sun *et al.*, 2022) showed that MFCC and its related features could yield relatively good performance on recognizing snoring sounds. Jiang *et al.* (2020) extracted the Mel-spectrogram of signal and used a convolution neural network (CNN) to classify snoring and non-snoring sounds with good performance. Sun *et al.* (2022) also indicated that different components of MFCC yielded different contributions to the classification result. The same result is also shown in this study. The high effectivity of MFCC and WTCC may be caused by dividing the signal into different subbands with the same frequency range through Mel-filter banks based on the mechanism of human hearing.

There are many studies to classify snoring and non-snoring sounds from other parts shown in Table 4 (Adesuyi *et al.*, 2022; Ankışhan, Tuncer, 2017; Arsenali *et al.*, 2018; Jiang *et al.*, 2020; Lim *et al.*, 2019; Nonaka *et al.*, 2016; Sun *et al.*, 2022) used the processing way of images to analyze snoring and non-snoring sound with the recognition accuracy of 95.1% in Mel-spectrogram. In the work of Nonaka *et al.* (2016), the auditory image model, which has been used to numerically explain the auditory phenomenon of human's auditory, was developed to automatically extract snoring sounds from sleep sounds, which could achieve a sensitivity of 97.2% from 40 subjects. The work of Lim *et al.* (2019) and Adesuyi *et al.* (2022) yielded the highest 99.0% accuracy compared with all other studies. However, these results are not convincing because that there are only 8 and 6 subjects for the two studies. The diversity of samples too small to demonstrate the effectiveness of the proposed algorithm. All these studies are based on subject dependence, which cannot be directly used in practice. It demonstrated evident differences between snoring and non-snoring sounds from all kinds of aspects. The models based on deep learning perform much better than traditional machine learning, which also demonstrates the effectiveness of WPT features proposed in this work. It is worth noticing that these works have yielded competitive results on their own limited dataset. However, there is no sense to compare these accuracies because of the inconsistent dataset used in studies. The dataset of each study is established by its own team with subjects from different counties, different recording equipment, and different labeling standards of snoring and non-snoring sounds. But the result of our work is comparable with previous studies in terms of our own dataset. And it demonstrated that the distribution of snoring and non-snoring sounds in each sub-band of all frequency ranges is obviously different.

In conclusion, there are some contributions to this study. Firstly, it used WPT to extract sub-band features and yielded comparable accuracy in recognizing snoring sounds. Then, it discussed different wavelet functions and decomposition layers, concluding that the difference is getting more evident with the frequency range getting smaller, which is more beneficial for classifying. Thirdly, it demonstrated that although the distribution of snoring sounds is mainly on the low frequency part, there are also differences between snoring and non-snoring sounds in the high frequency part. There are some limitations of the work. The data partition methods mainly included subject dependence and subject independence which greatly influenced the classification performance. Subject dependence is an original data partition method to discuss the features' influence on classification performance, while the result of subject independence is more suitable for use in practice. In this work, the partition of the training set and test set is based on subject dependence because of the limited subjects. The subject independent classification must base on a huge number of training and validation subjects to make up for the influence of individual characteristics. There are only 24 subjects used in the study, which is hard to perform subject independent classification considering individual characteristics. It is the next step of the paper to collect much more snoring sounds from different subjects to implement detecting snoring sounds based on subject independence.

Table 4. Classification results of current studies recorded by ambient microphones. Abbreviations include AdaBoost (adaptive boosting), CNN (convolution neural network), MNLR (multi-nominal logistic regression), STFT (short-time Fourier transform), RNN (recurrent neural network), and LLEs (largest Lyapunov exponents).

Author	Subjects	Features + classifier	Accuracy [%]	Sensitivity [%]
Jiang *et al.* (2020)	15	Mel-spectrogram + CNN	95.1	95.4
Nonaka *et al.* (2016)	40	Audio image model + MNLR	97.3	97.2
Sun *et al.* (2022)	24	Sub-band features + XGBoost	94.3	96.5
Lim *et al.* (2019)	8	MFCC, STFT + RNN	98.5	99.3
Ankışhan, Tuncer (2017)	22	Chaotic features + LLEs	94.4	88.3
Arsenali *et al.* (2018)	20	MFCC + RNN	95.0	92.0
Adesuyi *et al.* (2022)	6	MFCC + CNN	99.0	
This work	24	WPT features + XGBoost	95.15	96.94
		WPT + MFCC + XGBoost	95.68	97.27

5. Conclusion

This study proposed a snoring sounds recognition system based on WPT features and XGBoost classifier. The recorded sleep sounds of 24 subjects, firstly were enhanced and segmented by a subspace noise reduction algorithm and signal presence probability based on the estimation of noise autocorrelation respectively to obtain potential snoring episodes. In the training set, 10-fold cross-validation was implemented to select appreciated features and models. Results of the recognition system showed that features based on sub-bands could well classify snoring and non-snoring sounds with accuracy of 95.65%, sensitivity of 97.27%, and precision of 96.58% in the test set for DB7 function and level 5, the best combination of all test conditions. And the comparison among decomposition layers shows, although the distribution of snoring sounds is mainly in the low frequency part, there is also evident difference between snoring and non-snoring sounds in the high frequency part. However, the MFCC – WPT combined feature set outperformed the simple MFCC and WPT feature sets, with accuracy of 95.68%, sensitivity of 97.27%, and precision of 96.68%. These results have demonstrated that the wavelet packet analysis is effective in recognizing snoring sounds with less computational complexity, which can be further developed to analyze OSAHS at home.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant numbers 11974121, 81570904) and National Youth Foundation of China (grant number 81900927).

Ethical approval

This work was approved by the Ethics Committee of Guangzhou Medical University, and an informed consent was obtained from each participant. This work does not contain any studies with animals performed by any of the authors.

References

1. Adesuyi T.A., Kim B.M., Kim J. (2022), Snoring sound classification using 1D-CNN model based on multi-feature extraction, *International Journal of Fuzzy Logic and Intelligent Systems*, **22**(1): 1–10, doi: 10.5391/IJFIS.2022.22.1.1.
2. Ankişhan H., Tuncer A.T. (2017), A new portable device for the snore/non-snore classification, [in:] *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, doi: 10.1109/ICEngTechnol.2017.8308212.
3. Arsenali B. *et al.* (2018), Recurrent neural network for classification of snoring and non-snoring sound events, [in:] *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 328–331, doi: 10.1109/EMBC.2018.8512251.
4. Ayas N.T. (2013), *Risk factors for obstructive sleep apnea*, [in:] Encyclopedia of Sleep, pp. 212–214, doi: 10.1016/B978-0-12-378610-4.00308-9.
5. 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.
6. Chen T., Guestrin C. (2016), XGBoost: A scalable tree boosting system, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, doi: 10.1145/2939672.293 9785.
7. Dafna E., Tarasiuk A., Zigel Y. (2013), Automatic detection of whole night snoring events using non-contact microphone, *PLOS ONE*, **8**(12): e84139, doi: 10.1371/journal.pone.0084139.
8. Ding L., Peng J., Jiang Y., Song L. (2021), Generalized subspace snoring signal enhancement based on noise covariance matrix estimation, *Circuits, Systems, and Signal Processing*, **40**(7): 3355–3373, doi: 10.1007/s00034-020-01623-3.
9. 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.
10. Emoto T., Abeyratne U.R., Kawano K., Okada T., Jinnouchi O., Kawata I. (2018), Detection of sleep breathing sound based on artificial neural network analysis, *Biomedical Signal Processing and Control*, **41**: 1–89, doi: 10.1016/j.bspc.2017.11.005.
11. Han W., Chan C.F., Choy C.S., Pun K.P. (2006), An efficient MFCC extraction method in speech recognition, [in:] *2006 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 145–148, doi: 10.1109/iscas.2006.1692543.
12. Hui J. *et al.* (2015), Acoustic analysis of snoring in the diagnosis of obstructive sleep apnea syndrome: A call for more rigorous studies, *Journal of Clinical Sleep Medicine*, **11**(7): 765–771, doi: 10.5664/jcsm.4856.
13. Hwang S.H. *et al.* (2015), Polyvinylidene fluoride sensor-based method for unconstrained snoring detection, *Physiological Measurement*, **36**(7): 1399–1414, doi: 10.1088/0967-3334/36/7/1399.
14. Iber C., Ancoli-Israel S., Chesson A., Quan S.F. (2007), *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification*, Westchester, Illinois, American Academy of Sleep Medicine.
15. 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.

16. Kapur V., Strohl K.P., Redline S., Iber C., O'Connor G., Nieto J. (2002), Underdiagnosis of sleep apnea syndrome in U.S. communities, *Sleep and Breathing*, **6**(2): 49–54, doi: 10.1007/s11325-002-0049-5.
17. 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.
18. Karunajeewa A.S., Abeyratne U.R., Hukins C. (2011), Multi-feature snore sound analysis in obstructive sleep apnea-hypopnea syndrome, *Physiological Measurement*, **32**(1): 83–97, doi: 10.1088/0967-3334/32/1/006.
19. Korniienko O., Machusky E. (2018), Voice activity detection algorithm using spectral-correlation and wavelet-packet transformation, *Radioelectronics and Communications Systems*, **61**(5): 185–193, doi: 10.3103/S0735272718050011.
20. Lechner M., Breeze C.E., Ohayon M.M., Kotecha B. (2019), Snoring and breathing pauses during sleep: interview survey of a United Kingdom population sample reveals a significant increase in the rates of sleep apnoea and obesity over the last 20 years – data from the UK sleep survey, *Sleep Medicine*, **54**: 250–256, doi: 10.1016/j.sleep.2018.08.029.
21. Li T., Zhou M. (2016), ECG classification using wavelet packet entropy and random forests, *Entropy*, **18**(8): 1–16, doi: 10.3390/e18080285.
22. Lim S.J., Jang S.J., Lim J.Y., Ko J.H. (2019), Classification of snoring sound based on a recurrent neural network, *Expert Systems with Applications*, **123**: 237–245, doi: 10.1016/j.eswa.2019.01.020.
23. Monson B.B., Hunter E.J., Lotto A.J., Story B.H. (2014), The perceptual significance of high-frequency energy in the human voice, *Frontiers in Psychology*, **5**: 587, doi: 10.3389/fpsyg.2014.00587.
24. 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.
25. Ng A.K., Koh T.S., Puvanendran K., Abeyratne U.R. (2008), Snore signal enhancement and activity detection via translation-invariant wavelet transform, *IEEE Transactions on Biomedical Engineering*, **55**(10): 2332– 2342, doi: 10.1109/TBME.2008.925682.
26. Nonaka R. *et al.* (2016), Automatic snore sound extraction from sleep sound recordings via auditory image modeling, *Biomedical Signal Processing and Control*, **27**: 7–14, doi: 10.1016/j.bspc.2015.12.009.
27. Pedregosa F. *et al.* (2011), Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, **12**(85): 2825–2830.
28. Perez-Padilla J.R., Slawinski E., Difrancesco L.M., Feige R.R., Remmers J.E., Whitelaw W.A. (1993), Characteristics of the snoring noise in patients with and without occlusive sleep apnea, *American Review of Respiratory Disease*, **147**(3): 635–644, doi: 10.1164/ajrccm/147.3.635.
29. Qian K. *et al.* (2017), Snore sound recognition: On wavelets and classifiers from deep nets to kernels, [in:] *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3737–3740, doi: 10.1109/EMBC.2017.8037669.
30. Qian K., Janott C., Zhang Z., Heiser C., Schuller B. (2016), Wavelet features for classification of vote snore sounds, [in:] *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 221–225, doi: 10.1109/ICASSP.2016.7471669.
31. Qian K., Xu Z., Xu H., Wu Y., Zhao Z. (2015), Automatic detection, segmentation and classification of snore related signals from overnight audio recording, *IET Signal Processing*, **9**(1): 21–29, doi: 10.1049/iet-spr.2013.0266.
32. Senaratna C.V. *et al.* (2017), Prevalence of obstructive sleep apnea in the general population: A systematic review, *Sleep Medicine Reviews*, **34**: 70–81, doi: 10.1016/j.smrv.2016.07.002.
33. Sharma G., Umapathy K., Krishnan S. (2020), Trends in audio signal feature extraction methods, *Applied Acoustics*, **158**: 107020, doi: 10.1016/j.apacoust.2019.107020.
34. Solà-Soler J., Jané R., Fiz J.A., Morera J. (2007), Automatic classification of subjects with and without Sleep Apnea through snoring analysis, [in:] *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6093–6096, doi: 10.1109/IEMBS.2007.4353739.
35. Sun X., Peng J., Zhang X., Song L. (2022), Effective feature selection based on Fisher Ratio for snoring recognition using different validation methods, *Applied Acoustics*, **185**: 108429, doi: 10.1016/j.apacoust.2021.108429.
36. Torlay L., Perrone-Bertolotti M., Thomas E., Baciu M. (2017), Machine learning–XGBoost analysis of language networks to classify patients with epilepsy, *Brain Informatics*, **4**: 159–169, doi: 10.1007/s40708-017-0065-7.
37. Wang C., Peng J., Song L., Zhang X. (2017), Automatic snoring sounds detection from sleep sounds via multi-features analysis, *Australasian Physical and Engineering Sciences in Medicine*, **40**: 127–135, doi: 10.1007/s13246-016-0507-1.
38. Wang K., Su G., Liu L., Wang S. (2020), Wavelet packet analysis for speaker-independent emotion recognition, *Neurocomputing*, **398**: 257–264, doi: 10.1016/j.neucom.2020.02.085.
39. Wu T., Yan G.-Z., Yang B.-H., Sun H. (2008), EEG feature extraction based on wavelet packet decomposition for brain computer interface, *Measurement: Journal of the International Measurement Confederation*, **41**(6): 618–625, doi: 10.1016/j.measurement.2007.07.007.
40. Young T., Evans L., Finn L., Palta M. (1997), Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women, *Sleep*, **20**(9): 705–706, doi: 10.1093/sleep/20.9.705.

Archives of Acoustics Vol. 48, No. 1, pp. 13–24 (2023), doi: 10.24425/aoa.2022.142905

Research Paper

Comparative Analysis of Classifiers for the Assessment of Respiratory Disorders Using Speech Parameters

Poonam SHRIVASTAVA[(1)], Neeta TRIPATHI[(1)], Bikesh Kumar SINGH[(2)], Bhupesh Kumar DEWANGAN[(3)*]

[(1)] *Department of Electronics and Telecommunication, SSTC*
Bhilai, India

[(2)] *Department of Biomedical Engineering, National Institute of Technology*
Raipur, India

[(3)] *Department of Computer Science and Engineering, School of Engineering, OP Jindal University*
Raigarh, India

*Corresponding Author e-mail: bhupesh.dewangan@gmail.com

(*received September 9, 2021; accepted August 31, 2022*)

Non-invasive techniques for the assessment of respiratory disorders have gained increased importance in recent years due to the complexity of conventional methods. In the assessment of respiratory disorders, machine learning may play a very essential role. Respiratory disorders lead to variation in the production of speech as both go hand in hand. Thus, speech analysis can be a useful means for the pre-diagnosis of respiratory disorders. This article aims to develop a machine learning approach to differentiate healthy speech from speech corresponding to different respiratory disorders (affected). Thus, in the present work, a set of 15 relevant and efficient features were extracted from acquired data, and classification was done using different classifiers for healthy and affected speech. To assess the performance of different classifiers, accuracy, specificity (Sp), sensitivity (Se), and area under the receiver operating characteristic curve (AUC) was used by applying both multi-fold cross-validation methods (5-fold and 10-fold) and the holdout method. Out of the studied classifiers, decision tree, support vector machine (SVM), and k-nearest neighbor (KNN) were found more appropriate in providing correct assessment clinically while considering 15 features as well as three significant features ($Se > 89\%$, $Sp > 89\%$, $AUC > 82\%$, and accuracy $> 99\%$). The conclusion was that the proposed classifiers may provide an aid in the simple assessment of respiratory disorders utilising speech parameters with high efficiency. In the future, the proposed approach can be evaluated for the detection of specific respiratory disorders such as asthma, COPD, etc.

Keywords: healthy speech; affected speech; machine learning; classification techniques; respiratory disorders; speech analysis.

1. Introduction

The sound produced by humans to express language orally is called speech. As the respiratory system is the power source, thus, for the production of speech, to produce vibration in the vocal cord, sufficient airflow is required (Dogan *et al.*, 2007). Different breathing patterns depend on the purpose and nature of speech production. As speaking occurs only during exhalation, to increase the time available for speech production, breath out is slower and breath in is fast. Any kind of airway inflammation can affect the sound of voice quality. Disorders of the respiratory system may affect any of the structures and organs which have to do with breathing (Mohamed, El Maghraby, 2014). Human speech analysis is a wide research area that helps in medical condition diagnosis affecting the speech parameters (Dixit *et al.*, 2014). As per the Global Initia-

tive for Chronic Obstructive Lung Diseases (GOLD), respiratory diseases are affecting 400 million worldwide (Halpin *et al.*, 2020). To treat respiratory disorders promptly and appropriately, correct diagnosis is essential. Initially, diagnosis involves auscultation, i.e., the use of a stethoscope for examining lung sounds. Analog filtering and sound amplification is the basic requirement of standard stethoscopes to be interpreted by trained professionals. Then the pulmonary function test (PFT) which measures lungs volume and capacity, airflow rates, and gas exchange is performed. PFT can be done by two methods: spirometry or plethysmography. Along with these, chest X-rays and CT scans are performed. Still, misdiagnosis, under-diagnosis, and delayed diagnosis may occur in the treatment. The reason may involve the expertise needed for performing PFT and auscultation, overlapping among the disease, and the complexity of the disease. The prediction of the disease in the initial stage is very important in the medical field, as death may occur if a proper treatment got delayed.

Thus, we proposed a computerised analysis of speech signals for normal individuals and patients affected by different respiratory disorders.

1.1. Literature review

A comparative study between parametric and non-parametric methods, involving the mathematical transformations in the analysis of speech for the detection of disease has been demonstrated in (Sonu, Sharma, 2012). Mel-frequency cepstral coefficient (MFCC) for feature extraction and dynamic time warping (DTW) for feature matching were used. Although it was a time-consuming process, voice signal could be an alternative approach for respiratory disorder analysis.

A comparison of the performances of various classifiers such as Gaussian mixture model (GMM), multilayer perceptron (MLP) neural networks, support vector machine (SVM), and hierarchical fuzzy signature (HFS) along with the usage of a hybrid classifier, which also reduced the dimensionality, was reported in (Alghowinem *et al.*, 2013). It was observed that the best performance was given by using SVM with GMM as the hybrid classifier. Out of the three fusion methods, it was observed that while associating with HFS, MLP, and GMM, the performance of score fusion was better, while for SVM, the performance of decision fusion was the best. Feature fusion resulted in very poor performance as compared to other methods.

An acoustic analysis for asthmatic and normal persons in which jitter, shimmer, noise to harmonic ratio (NHR), and harmonic to noise ratio (HNR) showed significant variation was reported in (Teixeira, Fernandes, 2014). Jitter, shimmer, and HNR values for males and females were recorded. The result obtained on vowel comparison was found to have no difference between jitter values but there was a difference for shimmer and HNR values.

For the diagnosis of chronic diseases, different algorithms of classification have been applied to the database of diseases and the results are very promising. Still, a novel classification technique is needed. The different methods of acoustic feature extraction and classification that can help in detecting the disease in the prior stage are to be developed so that the process of diagnosis can be simplified.

The different methods of acoustic feature extraction and classification that can help in detecting the disease in the prior stage leading to the discrimination between the voice of healthy and unhealthy persons were discussed in (Saloni *et al.*, 2014). Digital signal processing (DSP) techniques were used for feature extraction whereas vector quantization (VQ), DTW, SVM, GMM, and artificial neural network (ANN) were used for feature classification. It was observed that different classification techniques may not be compared directly due to being measured on a different database.

An automatic disease diagnosis system that was adaptive based on SVM was developed in (Gürbüz, Kiliç, 2014). For the detection of disease in a better way, a new kind of SVM, "Adaptive SVM" has been introduced, showing 100% correct classification rates. The result showed that the proposed method demonstrated a higher success rate than an adaptive method as compared to non-adaptive methods. The method was not disease-specific and as practical as it separates the bias parameters space into subfragments.

To determine the level of asthma, a numerical formula was demonstrated in (Walia, Sharma, 2016). Voice parameters like jitter, shimmer, fundamental frequency, and maximum phonation time were used for generating the formula. On analysis, it was found that the jitter value was low for healthy and high for asthmatic patients while maximum phonation time was vice versa.

The myAirCoach decision support systems design aspects were proposed in (Kocsis *et al.*, 2017) with the focus on the analysis of three machine learning approaches (SVM, random forests, AdaBoot) as support tools. In comparison with SVM and AdaBoot, the random forests algorithm shows better accuracy.

SVM, Naïve Bayes, decision tree, and ANN are considered to be the most widely used classifiers for chronic disease prediction, but Jain and Singh (2018) put focus on adaptive and parallel classification systems that enhance the rate of success and reduce the time taken in making the decision. In this proposed method only for feature selection, the filter method was found to be more efficient. However, by applying hybrid approaches to disease databases, redundant, noisy, and insignificant features may be reduced.

After applying machine learning to self-management asthma, it was found that both Naïve Bayes and logistic regression-based classifiers provided the highest accuracy (AUC > 0.87) in (Tsang *et al.*, 2020). Asthma Mobile Health Study (AMHS) dataset was used. Several prevailing machine learning classifiers, both probabilistic and deterministic models and linear and non-linear were used. Along with AUC, geometric mean accuracy (GMA) was employed due to the skewed nature of the data.

A feasible method of disease detection using analysis of voice was proposed in (Gore *et al.*, 2020). For feature extraction, MFFC and feature matching DTW were applied. Various voice analyses were presented and verified to track characteristics variation in patients' voices.

For the passive assessment of pulmonary functions, two algorithms were proposed in (Chun *et al.*, 2020). One of them was used for distinguishing between healthy and affected with pulmonary disease and the other one to estimate the FEV1/FVC (Forced Expiratory Volume to Forced Vital Capacity) ratio using speech features. Data sets from the research study and in-clinic study were used to develop and validate the algorithms. It was observed that the classification accuracy was obtained to be 73.7% while the F1 score was 84.5%. Also, a mean absolute error of 8.6% was observed with FEV1/FVC ratio in regression analysis.

Even though so much work has already been done in the speech analysis area and respiratory disorders but still, less work has been done for combining both. In the speech area, recognition and emotional patterns have been considered while the respiratory function is generally assessment done using lung sounds. So, this paper deals with the assessment of respiratory disorder using speech parameters by comparing different classifiers on the same dataset. As per the literature survey, the main drawback was the use of various datasets which makes the comparison even more complicated.

Therefore, for the comparison of classifiers, there is a need of using the same dataset.

In this article, we have compared 5 classifiers on the basis of multifold cross-validation and the hold-out method, and 15 features were extracted from the speech of each of the 20 participants.

Summary of the study contribution:

- in this study, we have proposed a speech signal-based detection of affected speech. Different speech features were extracted from the speech signal and evaluated using classification techniques to detect abnormalities in the speech parameters;
- we also implemented and evaluated different machine learning classifiers capable of differentiating healthy speech and affected speech.

2. Materials and methods

This section presents different steps involved in the systematic classification of speech features.

2.1. Data collection

The dataset comprises speech samples of 20 individuals aged between 24–65 years, 10 healthy (6 males and 4 females), and 10 patients (10 males and 0 females). All the participants have given their written consent.

Samples were recorded using Goldwave software with the sampling frequency of 11 025 Hz by a microphone located 2–3 cm in front of the participant's lips. The participants were asked to repeat specified words in Hindi while sitting and to adopt pitch and loudness with which they were usually comfortable. Each individual recorded the speech for two minutes in a continuous manner. Only one recording was obtained from each patient.

The database consisting of a sustained phonation was created. The input signal waveforms of healthy and affected people' speech are shown in Fig. 1.

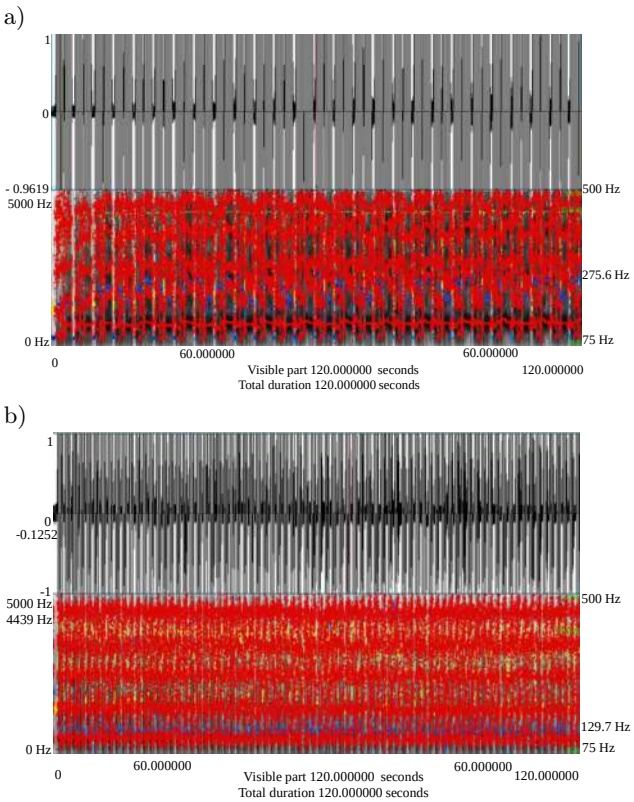


Fig. 1. (a) Healthy person's input signal waveform; (b) affected person's input signal waveform.

From Fig. 1 it was observed that for the healthy person, the waveform was uniform but for the affected person, the waveform contains deformities.

For the lung assessment, a spirometry test had been performed before speech recording for all participants.

This test estimates the amount of air that can be breathed in and breathed out of the lungs, as well as if the air can be blown out of the lungs, fast and easily. In spirometry, participants were asked to inhale deeply and hold for 6 seconds then exhale completely. Thus, the forced expiratory volume in 1 second (FEV1), defined as the amount of air that can be forced from the lungs in one second, and peak expiratory flow rate (PEFR) which measures how much air flows through the bronchi and displays the level of obstruction in the lungs, were recorded.

All the participants were subjected to a spirometry test before categorising them as healthy or affected under clinical supervision. The values of FEV1 and PEFR were used as the gold standard to differentiate the two groups. A participant was considered affected if the values of FEV1 and PEFR were observed to be less than 60.

The vital information such as age, gender, height, weight, occupation, and medical history was also noted down and included in the final analysis other than spirometry.

2.2. Feature extraction

A set of 15 speech features, namely, formant frequencies, F1, F2, and F3, pitch, intensity, jitter (rap, %), mean autocorrelation, jitter (local, %), mean NHR, shimmer (local, %), mean HNR, amplitude mean (Pascal), total energy ($\mathrm{Pa}^2 \cdot \mathrm{s}$), mean power (intensity), and standard deviation in a channel (Pascal) was extracted with the help of PRAAT software.

The following speech parameters are explained as follows:

- Formant frequencies: speech producing different frequency components of the sound signal is called formant frequencies.
- Pitch: pitch is defined as the ordering of sound property on a scale that is frequency-related. Thus, the relative highness or lowness of a tone is considered pitch.
- Intensity: the power of sound per unit area is known as intensity. Also defined as the amplitude of the vibrations that affect loudness.
- Jitter: jitter may be defined as frequency parameters variation from cycle to cycle.
  Relative jitter or local jitter is defined as the ratio between the average differences between consecutive periods, relative to the overall average period. It is given in percentage:

$$\text{Jitter (relative)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \cdot 100, \quad (1)$$

  where T_i is extracted F0 period lengths, N is number of extracted F0 periods.

  Jitter (rap): the average absolute difference between a period and the average of it and its two neighbors, divided by the average period is defined as jitter (rap). It is expressed as a percentage:

$$\text{Jitter (rap)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1} T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \cdot 100. \quad (2)$$

- Shimmer: it may be represented as the parameters associated with the variation of the amplitude of the sound wave.
  Shimmer relative: the ratio between the average absolute difference between the amplitudes of consecutive periods and the average amplitude is defined as shimmer relative, given in percentage:

$$\text{Shimer (relative)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \cdot 100, \quad (3)$$

  where A_i is extracted peak-to-peak amplitude data, N is number of extracted fundamental frequency periods.
- Mean autocorrelation: the relationship between the current values of variables and their past values is measured by autocorrelation. The correlation coefficient is usually denoted ρ. For variables, x and y, each contains N values:

$$\rho = \frac{\sum_i (x_i - \mu_x)(y_i - \mu_y)}{N\sigma_x\sigma_y}, \quad (4)$$

  where the means of x and y are given by μ_x and μ_y, and their standard deviations are given as σ_x and σ_y.
- Harmonic to noise ratio (HNR): the periodic components of speech sound divided by non-periodic components is represent by harmonic to noise ratio:

$$\text{HNR} = 10 \cdot \log_{10} \frac{AC_V(T)}{AC_V(0) - AC_V(T)}, \quad (5)$$

  where the autocorrelation coefficient consisting of all signal energy at the origin is given by $AC_V(0)$, and the autocorrelation component related to the fundamental period is given by $AC_V(T)$.
- Noise to harmonic ratio (NHR): hoarseness can be measured effectively by noise to harmonic ratio:

$$\text{NHR} = 1 - \text{autocorrelation}. \quad (6)$$

- Amplitude mean: it is the amplitude of the vibrations that affects loudness which is the size of oscillations of the vocal folds.

2.3. Statistical significant analysis

Now, to determine the statistical significance of extracted features, statistically significant analysis using Statistical Package for Social Sciences (SPSS) was applied. Each study has a confidence level of 95% and a p-value of <0.05 considering being statistically significant.

2.4. Classification

Classification is a process in which the class of given data points is predicted and along with targets, input data are also provided. This type of classification comes under the category of supervised learning.

Thus, to understand the given input variables related to the class, a classifier utilises some training data.

There are several classification algorithms varying in the nature and application of available data.

In the present study, the following classification algorithms were evaluated:

- KNN with all the kernels,
- SVM with all the kernels,
- decision tree,
- logistic regression,
- linear discriminant.

As these algorithms represent a variety of classifiers' algorithms, they were chosen (Kuncheva, 2014) and also some of them had performed well in previous studies (Caruana, Niculescu-Mizil, 2006).

After all the statistically significant features were obtained, they were used for the classification of healthy and affected speech. For this, different supervised machine learning techniques were applied, explained as follows:

- Decision tree: as the name specifies, the classification or regression models are built in the form of a tree structure in the decision tree method. The method applied an if-then rule set that is both mutually exclusive and exhaustive for the classification. One at a time the training data are used for learning the rule sequentially.
- SVM: for classification and regression problems, support vector machines are widely used. Figure 2 represents the SVM in a two-dimensional space. By constructing a hyperplane, it separates two classes of a sample to distinguish class members from non-members (Byun, Lee, 2002). A hyperplane is constructed as the decision plane in SVM, separating the positive (+1) and negative (–1) classes with the largest margin. The maximum margin of separation between the two classes is an optimal hyperplane, where the margin is the sum of the distances from the hyperplane to the closest data points of each of the two classes. These closest data points are called support vectors (SVs). The optimal separating hyperplane is represented by the solid line in Fig. 2.

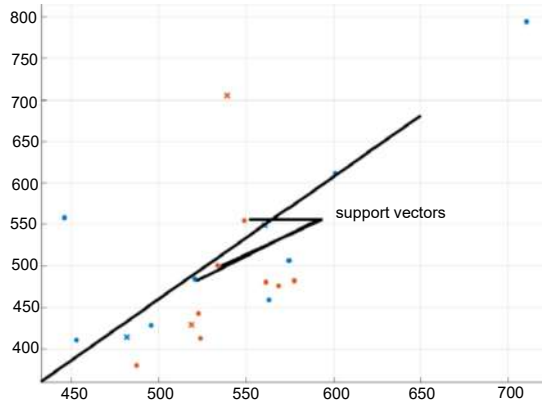


Fig. 2. Representation of SVM in a two-dimensional space.

In a variety of classification and regression theories, SVM has been successfully used (Sapankevych, Sankar, 2009):

- Logistic regression: to model the conditional probability, the logistic function is used by the statistical model and this is known as logistic regression. Basically when the target variable's value is categorical then this classification algorithm is used. Thus, most commonly used when the data in question has binary output.
  The formula is given:

$$P = \frac{1}{1 + e^{-(a+bX)}}, \tag{7}$$

  where P is the probability of 1 (the proportion of 1 s), e is the base of the natural algorithm, a and b are parameters of the models, and X is the independent variable related to the logistic curve.
- Linear discriminant: linear discriminant analysis is a simple and effective supervised classification method, used to create machine learning models (Amaral *et al.*, 2012). It is used for modelling differences in groups, i.e., separating two or more classes.
- KNN: KNN method that uses data and classifies new data points based on similarity measures is considered to be the simplest method applied for regression and classification problems (Amaral *et al.*, 2013). Classification is done by a maximum vote from its neighbors. KNN was briefly defined in the previous works, as KNN calculates the distance between data points. For this, the simple Euclidean distance formula is generally used:

$$\mathrm{d}(p,q) = \mathrm{d}(q,p) = \sqrt{\sum_{i=1}^{n}(\mathbf{q}_i - \mathbf{p}_i)^2}, \tag{8}$$

where p and q are two points in Euclidean n-space, $\mathbf{q}_i$ and $\mathbf{p}_i$ are Euclidean vectors, starting from the origin of the space (initial point), n is n-space.

The block diagram and overall proposed strategy are depicted in Fig. 3.

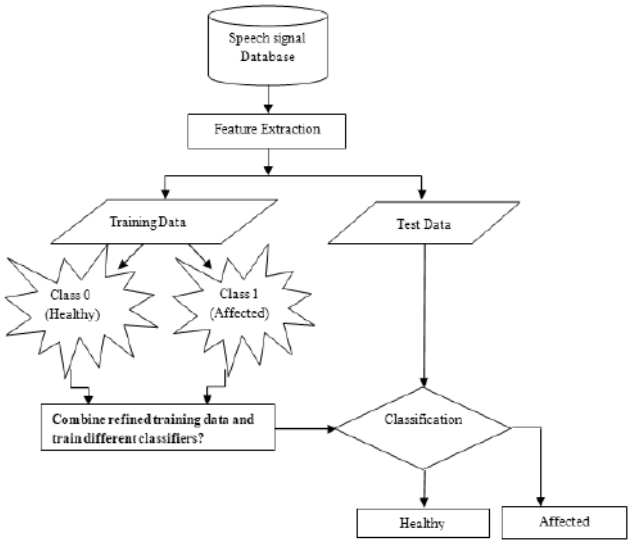


Fig. 3. Flowchart of the proposed classification approach.

The dataset has been divided into 67% training and 33% testing parts in the case of the holdout method, while in the case of the multifold cross-validation process both 5-fold and 10-fold have been used. After classification, the healthy speech was distinguished from the affected speech signals.

2.5. Performance evaluation

2.5.1. Data division protocol

Two groups of the dataset were formed to evaluate the performance of the proposed classifier models. One group was used for training purposes while the other was used for testing, and k-fold cross-validation and the holdout were applied as two data division protocols. The holdout is the most common method out of the several existing in which the given dataset has two groups divided randomly. The train set will be used to train the data set, and the unseen test data will be used to test its predictive power. 67% of the samples were used for training while 33% were used for testing. The common problem that generally occurs in most of the models in machine learning is over-fitting. So, to verify that the model is not overfit, k-fold cross-validation can be conducted in which random partition of the data set into k manually exclusive groups each approximately of the same size is made. For testing, one is kept, while for training others are used (Refaeilzadeh *et al.*, 2009). The experiments were conducted with the 5-fold and 10-fold cross-validation.

2.5.2. Performance measures

After training the data, the testing was performed and the following performance parameters were used for evaluation. Accuracy, recall, specificity, true positive rate, sensitivity, false positive rate, precision, and the area under the receiver operating characteristic (ROC), and the area under curve (AUC) are some well known performance criteria (Fawcett, 2006). In this paper, we have selected accuracy, sensitivity, specificity, and AUC for ROC curves as they are generally applied in medical diagnoses:

– Accuracy: accuracy is obtained by dividing all the correct predictions by the total number of predictions:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{9}$$

where TP – true positive, TN – true negative, FP – false positive, FN – false negative.

– Sensitivity: sensitivity is calculated by dividing true positive by true positive plus false negative:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{10}$$

– Specificity: true negative divided by the sum of true negative and false positive is defined as specificity:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{11}$$

– Area under curve (AUC): to evaluate the performance, a single value metric is used, known as area under ROC curve, plotted between true positive rate (TPR) on the y-axis and false positive rate (FPR) on the x-axis at the various thresholds.

3. Results and discussion

Table 1 presents an independent sample t-test applied to the database of 15 features. The detailed statistics of various features were divided into two classes, i.e., class 0 (healthy) and class 1 (affected) using SPSS software.

It has been observed that t-values and df-values did not provide any significant variances but sig (2-tailed) provided values of mean autocorrelation to be 0.011 and 0.012 when equal variance was assumed and when it was not assumed, respectively, as presented in Table 1. Similarly, mean NHR values were achieved to be 0.009 and 0.010, and mean HNR values were 0.000 for both the cases, i.e., when equal variance was assumed and when it was not assumed. This test shows that only the mean autocorrelation, mean NHR, and mean HNR were statistically evident in the healthy and affected speech (p-value < 0.05 and confidence interval of 95%), while the rest features were not statistically

Table 1. *t*-test for equality means.

Independent samples test		
Speech features	Conditions	*t*-test for equality of means
		sig (2-tailed)
F1	Equal variances assumed	0.838
	Equal variances not assumed	0.839
F2	Equal variances assumed	0.137
	Equal variances not assumed	0.145
F3	Equal variances assumed	0.425
	Equal variances not assumed	0.426
Pitch	Equal variances assumed	0.448
	Equal variances not assumed	0.451
Intensity	Equal variances assumed	0.973
	Equal variances not assumed	0.973
Jitter (local)	Equal variances assumed	0.190
	Equal variances not assumed	0.190
Jitter (rap)	Equal variances assumed	0.319
	Equal variances not assumed	0.333
Shimmer (local)	Equal variances assumed	0.365
	Equal variances not assumed	0.367
mean Auto Correlation	Equal variances assumed	0.011
	Equal variances not assumed	0.012
mean NHR	Equal variances assumed	0.009
	Equal variances not assumed	0.010
mean HNR	Equal variances assumed	0.000
	Equal variances not assumed	0.000
Amplitude mean [Pa]	Equal variances assumed	0.617
	Equal variances not assumed	0.621
Total energy [$Pa^2 \cdot s$]	Equal variances assumed	0.672
	Equal variances not assumed	0.672
mean Power (intensity) in air [dB]	Equal variances assumed	0.971
	Equal variances not assumed	0.971
Standard deviation in the channel [Pa]	Equal variances assumed	0.763
	Equal variances not assumed	0.764

significant. Hence these three features were considered for further classification.

Four performance measures, namely: accuracy, sensitivity, specificity, and AUC were used for evaluation under three data division protocols, namely: 5-fold, 10-fold, and the holdout. MATLAB software was used to perform this step.

Table 2 shows the performance of various classifier models using 5-fold, 10-fold, and the holdout cross-validation when 15 features were considered. It was observed that in 5-fold and 10-fold, the decision tree achieved the highest values of classification accuracy of 90%. The other performance measures, namely, sensitivity and specificity were found to be 90%, and AUC was 0.82. Whereas logistic regression and linear discriminant achieved the lowest values. Cubic KNN also shows 90% classification accuracy in the case of the 5-fold method. Also, coarse KNN achieved the least value of classification accuracy and AUC, i.e., 0.5, but achieved the highest value for specificity, i.e., 100%, and zero value for sensitivity.

It was found that in the holdout method, decision tree, linear discriminant, all the kernels of SVM (except quadratic SVM), and weighted KNN achieved 100% sensitivity. Zero specificity was observed in fine Gaussian SVM and fine KNN. Also, zero sensitivity was again observed in the case of coarse KNN along with 50% of accuracy and 86% of specificity. The decision tree achieved the highest accuracy of 83%.

Table 3 shows the performance evaluation of three significant features obtained by statistical analysis using SPSS, namely: mean NHR, mean HNR, and mean

Table 2. Performance of various classifiers using 15 features under different data division protocols.

Data division protocol	Classification techniques	Performance measures			
		Accuracy [%]	Sensitivity [%]	Specificity [%]	AUC
5-fold	Decision tree	90	90	90	0.82
	Linear discriminant	45	40	50	0.45
	Logistic regression	45	40	50	0.45
	Linear SVM	75	80	70	0.70
	Quadratic SVM	65	70	60	0.67
	Cubic SVM	45	40	50	0.68
	Fine Gaussian SVM	45	60	30	0.41
	Medium Gaussian SVM	80	80	80	0.80
	Coarse Gaussian SVM	75	80	70	0.75
	Fine KNN	60	60	60	0.60
	Medium KNN	85	90	80	0.89
	Coarse KNN	50	0	100	0.50
	Cosine KNN	85	80	90	0.86
	Cubic KNN	90	90	90	0.89
	Weighted KNN	75	90	60	0.83
10-fold	Decision tree	90	90	90	0.82
	Linear discriminant	55	50	60	0.49
	Logistic regression	40	50	30	0.42
	Linear SVM	65	70	60	0.66
	Quadratic SVM	75	90	60	0.82
	Cubic SVM	60	60	60	0.67
	Fine Gaussian SVM	50	80	20	0.41
	Medium Gaussian SVM	80	80	80	0.81
	Coarse Gaussian SVM	80	80	80	0.49
	Fine KNN	55	60	50	0.55
	Medium KNN	70	90	50	0.84
	Coarse KNN	50	0	100	0.50
	Cosine KNN	80	80	80	0.84
	Cubic KNN	70	80	60	0.78
	Weighted KNN	75	90	60	0.81
Holdout	Decision tree	83	100	71	0.83
	Linear discriminant	67	67	75	0.67
	Logistic regression	67	67	75	0.67
	Linear SVM	67	100	60	0.56
	Quadratic SVM	50	67	67	0.56
	Cubic SVM	67	100	60	0.78
	Fine Gaussian SVM	50	100	0	0.67
	Medium Gaussian SVM	67	100	60	0.89
	Coarse Gaussian SVM	67	100	60	0.44
	Fine KNN	50	100	0	0.50
	Medium KNN	67	67	75	0.67
	Coarse KNN	50	0	86	0.50
	Cosine KNN	50	67	67	0.56
	Cubic KNN	50	33	80	0.61
	Weighted KNN	67	100	60	0.56

Table 3. Performance of various classifiers using 3 significant features under different data division protocols.

Data division protocol	Classification techniques	Performance measures			
		Accuracy [%]	Sensitivity [%]	Specificity [%]	AUC
5-fold	Decision tree	90	90	90	0.82
	Linear discriminant	75	70	80	0.76
	Logistic regression	75	70	80	0.69
	Linear SVM	85	90	80	0.87
	Quadratic SVM	80	80	80	0.84
	Cubic SVM	70	70	70	0.67
	Fine Gaussian SVM	80	70	90	0.89
	Medium Gaussian SVM	80	80	80	0.85
	Coarse Gaussian SVM	85	90	80	0.85
	Fine KNN	85	90	80	0.85
	Medium KNN	85	100	80	0.79
	Coarse KNN	50	0	100	0.50
	Cosine KNN	85	90	80	0.84
	Cubic KNN	85	90	80	0.79
	Weighted KNN	85	90	80	0.85
10-fold	Decision tree	90	90	90	0.82
	Linear discriminant	80	80	80	0.92
	Logistic regression	75	70	80	0.81
	Linear SVM	85	90	80	0.86
	Quadratic SVM	85	90	80	0.81
	Cubic SVM	70	70	70	0.63
	Fine Gaussian SVM	75	70	80	0.87
	Medium Gaussian SVM	80	80	80	0.84
	Coarse Gaussian SVM	85	90	80	0.85
	Fine KNN	80	80	80	0.80
	Medium KNN	85	90	80	0.85
	Coarse KNN	50	0	100	0.50
	Cosine KNN	85	90	80	0.90
	Cubic KNN	85	90	80	0.82
	Weighted KNN	85	90	80	0.84
Holdout	Decision tree	100	100	100	1.00
	Linear discriminant	80	67	100	1.00
	Logistic regression	80	67	100	0.83
	Linear SVM	100	100	100	1.00
	Quadratic SVM	100	100	100	1.00
	Cubic SVM	80	100	50	0.50
	Fine Gaussian SVM	80	67	100	0.83
	Medium Gaussian SVM	100	100	100	1.00
	Coarse Gaussian SVM	80	100	67	1.00
	Fine KNN	80	67	100	0.83
	Medium KNN	60	33	100	1.00
	Coarse KNN	40	0	100	0.50
	Cosine KNN	100	100	100	1.00
	Cubic KNN	60	33	100	1.00
	Weighted KNN	80	67	100	1.00

autocorrelation that showed significant variance. The same procedure was adopted and the result was shown.

Table 3 summarises the performance measures of various classifier methods when only three significant features were considered using different data division protocols. It was found that 90% of classification accuracy was shown by a decision tree in 5-fold as well as 10-fold cross-validation. The other performance measures, namely, sensitivity and specificity were 90%, and AUC was 0.82. Again, coarse KNN showed the smallest value of classification accuracy, i.e., 50%, and AUC to be 0.5 but gave the highest value for specificity, i.e., 100%, and zero value for sensitivity.

Whereas in the holdout method it was found that decision tree, linear SVM, quadratic SVM, medium Gaussian SVM, and cosine KNN have shown classification accuracy, sensitivity, and specificity to be 100%, and AUC to be 1. Specificity was found to be 100% in almost all the classifiers except cubic SVM and coarse Gaussian SVM. Coarse KNN achieved zero sensitivity, 40% accuracy, 0.5 AUC along with 100% specificity.

On analysing Tables 2 and 3, it was observed that sensitivity, accuracy, specificity, and AUC were found to be high in the case of the decision tree as compared to the rest of the classifiers. Similarly, the holdout method was found to be the best performer out of all the three data division protocols considering both 15 features and three significant features.

It was also observed that, out of all the kernels of KNNs, coarse KNN, and out of all the kernels of SVMs, fine Gaussian SVM was giving poor performance throughout the experiment.

The diagnostic capability of classifiers and features can be determined using TPR and TNR. A plot between TPR and TNR is called receiver operating characteristics (ROC). Figure 4 shows the result of the

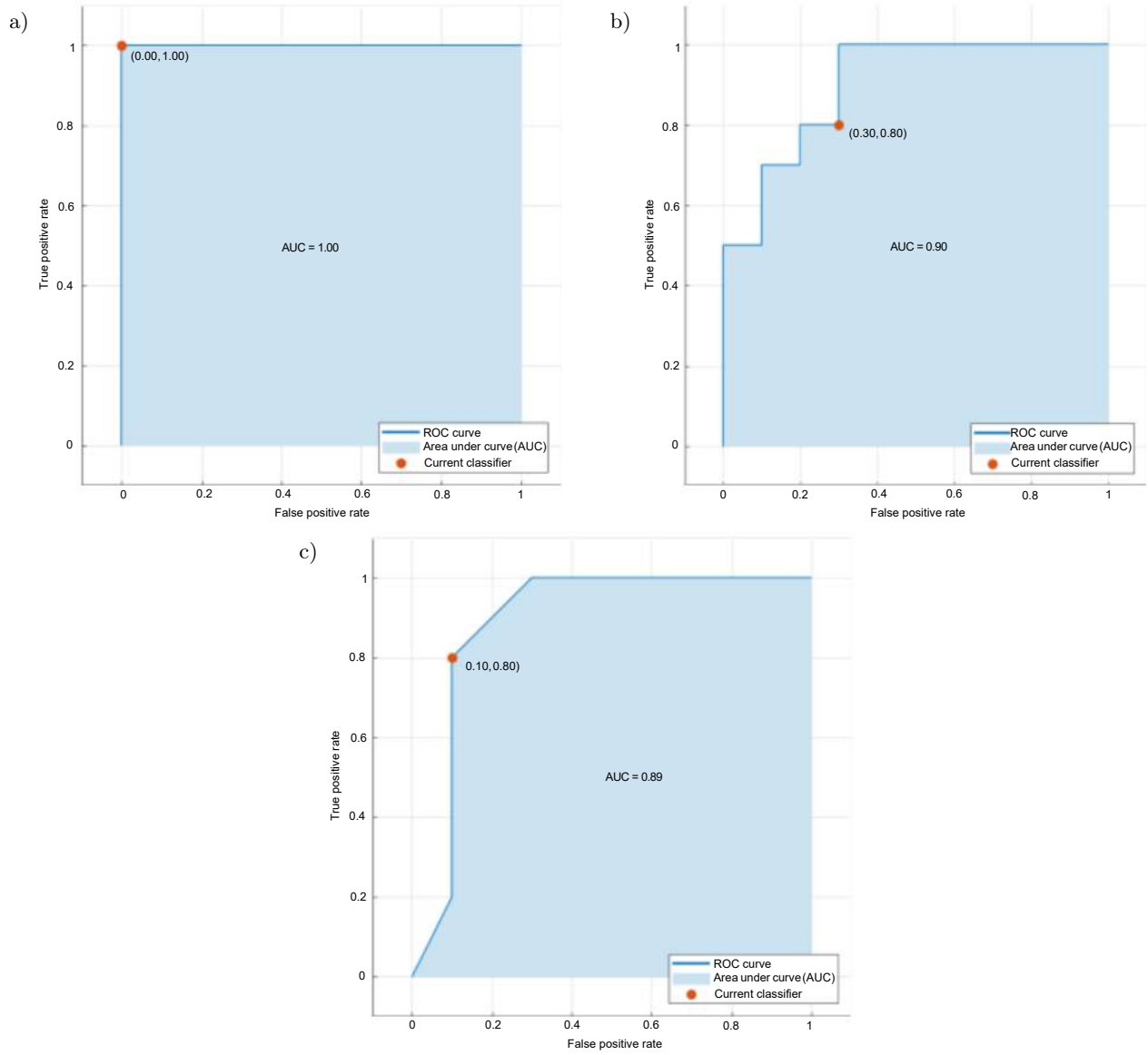


Fig. 4. ROC plots for various classifiers: a) decision tree (hold and method); b) cosine KNN (10-fold); c) cubic KNN (5-fold).

best three ROC analyses. It was observed that the AUC of 1 was obtained by the holdout method for the decision tree and the AUC of 0.90 and 0.89 were achieved by k-fold methods for KNN classifiers and verified in the results of Tables 2 and 3.

The best three ROC curves observed during the analysis were represented in Fig. 4.

4. Conclusion

It has been observed that on comparing different classifiers for all features, accuracy, specificity, sensitivity, and area under the curve performance measures for 5-fold, 10-fold cross-validation, and the holdout method, the decision tree achieved 90–100% classification accuracy. Further, SVMs and KNNs achieved the lowest accuracy between 40 and 70%. The holdout method had given a promising result.

Similarly, on comparing the accuracy with three significant features almost the same result was shown by both the 5-fold and 10-fold cross-validation methods for all the various classifiers. For the decision tree, it was again 90%, for SVM and KNN it was found to be 90–100%. But for logistic regression and linear discriminant, the performance improved to 80–90%. Again, the holdout method had given almost perfect results.

Thus, it is concluded that using different machine learning techniques, a comparative analysis of classifiers shows that the decision tree was effective as classification accuracy achieved 90% along with the holdout data division protocol for classification of speech of healthy and affected individuals.

Acknowledgments

The authors express their sincere thanks to the Director and Management of the research center, Shri Shankara College of Engineering and Technology, Bhilai for permitting to use the facilities of their Institute.

The authors extend their appreciation to the Dean and the supporting staff of the Pulmonary department of the O.P Jindal Fortis Hospital and Research Center for permitting to use their facilities for creating the database.

References

1. Alghowinem S. *et al.* (2013), A comparative study of different classifiers for detecting depression from spontaneous speech, [in:] *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8022–8026, doi: 10.1109/ICASSP.2013.6639227.

2. Amaral J.L., Lopes A.J., Jansen J.M., Faria A.C., Melo P.L. (2012), Machine learning algorithms and forced oscillation measurements applied to the automatic identification of chronic obstructive pulmonary disease, *Computer Methods and Programs in Biomedicine*, **105**(3): 183–193, doi: 10.1016/j.cmpb.2011.09.009.

3. Amaral J.L., Lopes A.J., Jansen J.M., Faria A.C., Melo P.L. (2013), An improved method of early diagnosis of smoking-induced respiratory changes using machine learning algorithms, *Computer Methods and Programs in Biomedicine*, **112**(3): 441–454, doi: 10.1016/j.cmpb.2013.08.004.

4. Byun H., Lee S.W. (2002), Applications of support vector machines for pattern recognition: A survey, [in:] *International Workshop on Support Vector Machines*, pp. 213–236, Springer, Berlin, Heidelberg, doi: 10.1007/3-540-45665-1_17.

5. Calverley P.M.A. (2020), Defining airflow obstruction: More data, further clarity, *American Journal of Respiratory and Critical Care Medicine*, **202**(5): 649–650, doi: 10.1164/rccm.202005-1551ED.

6. Caruana R., Niculescu-Mizil A. (2006), An empirical comparison of supervised learning algorithms, [in:] *Proceedings of the 23rd International Conference on Machine Learning*, pp. 161–168, doi: 10.1145/1143844.1143865.

7. Chun K.S. *et al.* (2020), Towards passive assessment of pulmonary function from natural speech recorded using a mobile phone, [in:] *2020 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–10, doi: 10.1109/PerCom45495.2020.9127380.

8. Dixit, Mittal V., Sharma Y. (2014), Voice parameter analysis for the disease detection, *IOSR Journal of Electronics and Communication Engineering*, **9**(3): 48–55.

9. Dogan M., Eryuksel E., Kocak I., Celikel T., Sehitoglu M.A. (2007), Subjective and objective evaluation of voice quality in patients with asthma, *Journal of Voice*, **21**(2): 224–230, doi: 10.1016/j.jvoice.2005.11.003.

10. Fawcett T. (2006), An introduction to ROC analysis, *Pattern Recognition Letters*, **27**(8): 861–874, doi: 10.1016/j.patrec.2005.10.010.

11. Gore S.M., Salunke M.M., Patil S.A., Kemalkar A.K. (2020), Disease detection using voice analysis, *International Research Journal of Engineering and Technology*, **7**(5): 7655–7659.

12. Gürbüz E., Kiliç E. (2014), A new adaptive support vector machine for diagnosis of diseases, *Expert Systems*, **31**(5), 389–397, doi: 10.1111/exsy.12051.

13. Halpin D.M. *et al.* (2021), Global initiative for the diagnosis, management, and prevention of chronic obstructive lung disease. The 2020 GOLD science committee report on COVID-19 and chronic obstructive pulmonary disease, *American Journal of Respiratory and Critical Care Medicine*, **203**(1): 24–36, doi: 10.1164/rccm.202009-3533SO.

14. Jain D., Singh V. (2018), Feature selection and classification systems for chronic disease prediction: A re-

view, *Egyptian Informatics Journal*, **19**(3): 179–189, doi: 10.1016/j.eij.2018.03.002.

15. Kocsis O. *et al.* (2017), Assessing machine learning algorithms for self-management of asthma, [in:] *2017 E-Health and Bioengineering Conference (EHB)*, pp. 571–574, doi: 10.1109/EHB.2017.7995488.
16. Kuncheva L.I. (2014), *Combining Pattern Classifiers: Methods and Algorithms*, John Wiley & Sons.
17. Mohamed E.E., El Maghraby R.A. (2014), Voice changes in patients with chronic obstructive pulmonary disease, *Egyptian Journal of Chest Diseases and Tuberculosis*, **63**(3), 561–567, doi: 10.1016/j.ejcdt.2014.03.006.
18. Refaeilzadeh P., Tang L., Liu H. (2009), Cross-validation, [in:] *Encyclopedia of Database Systems*, Liu L., Özsu M. [Eds.], Springer, doi: 10.1007/978-0-387-39940-9_565.
19. Saloni, Sharma R.K., Gupta A.K. (2014), Disease detection using voice analysis: A review, *International Journal of Medical Engineering and Informatics*, **6**(3): 189–209, doi: 10.1504/IJMEI.2014.063173.
20. Sapankevych N.I., Sankar R. (2009), Time series prediction using support vector machines: A survey, *IEEE Computational Intelligence Magazine*, **4**(2): 24–38, doi: 10.1109/MCI.2009.932254.
21. Sonu, Sharma R.K. (2011), Disease detection using analysis of voice parameters, [in:] *5th IEEE International Conference on Advanced Computing and Communication Technologies (ICACCT-2011)*, Choudhary R.K., Verma M., Saini S. [Eds.], pp. 416–420.
22. Teixeira J.P., Fernandes P.O. (2014), Jitter, shimmer and HNR classification within gender, tones and vowels in healthy voices, *Procedia Technology*, **16**: 1228–1237, doi: 10.1016/j.protcy.2014.10.138.
23. Tsang K.C., Pinnock H., Wilson A.M., Shah S.A. (2020), Application of machine learning to support self-management of asthma with mHealth, [in:] *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 5673–5677, doi: 10.1109/EMBC44109.2020.9175679.
24. Walia G.S., Sharma R.K. (2016), Level of asthma: mathematical formulation based on acoustic parameters, [in:] *2016 Conference on Advances in Signal Processing (CASP)*, pp. 24–27, doi: 10.1109/CASP.2016.7746131.
25. Wiechern B., Liberty K.A., Pattemore P., Lin E. (2018), Effects of asthma on breathing during reading aloud, *Speech, Language and Hearing*, **21**(1): 30–40, doi: 10.1080/2050571X.2017.1322740.

Archives of Acoustics Vol. 48, No. 1, pp. 25–38 (2023), doi: 10.24425/aoa.2023.144263

Research Paper

Numerical Simulation of Breast Cancer in the Early Diagnosis with Actual Dimension and Characteristics Using Photoacoustic Tomography

Maryam AHANGAR DARBAND[(1)], Esmaeil NAJAFI AGHDAM[(1)*], Arash GHARIBI[(2)]

[(1)] *Department of Electrical Engineering, Sahand University of Technology*
Tabriz, Iran

[(2)] *Institute of Modern Physics, Shanxi Normal University*
Linfen, China

*Corresponding Author e-mail: najafiaghdam@sut.ac.ir

(*received August 26, 2021; accepted August 30, 2022*)

A numerical study and simulation of breast imaging in the early detection of tumors using the photoacoustic (PA) phenomenon are presented. There have been various reports on the simulation of the PA phenomenon in the breast, which are not in the real dimensions of the tissue. Furthermore, the different layers of the breast have not been considered. Therefore, it has not been possible to rely on the values and characteristics of the resulting data and to compare it with the actual state. Here, the real dimensions of the breast at three-dimensional and different constituent layers have been considered. After reviewing simulation methods and software for different stages of the PA phenomenon, a single suitable platform, which is commercially available finite element software (COMSOL), has been selected for simulating. The optical, thermal, elastic, and acoustic characteristics of different layers of breast and tumor at radiated laser wavelength (800 nm) were accurately calculated or obtained from a reliable source. Finally, by defining an array of 32 ultrasonic sensors on the breast cup at the defined arcs of the 2D slices, the PA waves can be collected and transmitted to MATLAB software to reconstruct the images. We can study the resulting PA wave and its changes in more detail using our scenarios.

Keywords: photoacoustic (PA); photoacoustic imaging (PAI); photoacoustic tomography; breast cancer; early diagnosis.

1. Introduction

Breast cancer is one of the most common forms of cancer among females, but it is less common among men. In 2019, the estimated 268,600 new cases of invasive breast cancer were diagnosed among women and approximately 2,670 cases were diagnosed in men. Breast cancer typically has no symptoms when the tumor is small and can be easily treated, therefore screening is vital for early detection (American Cancer Society, 2019). There are several methods for carefully screening breast tissue to identify cancerous tumors. However, a low-risk and non-invasive way that can detect cancerous tumors early is a priority. Among the available imaging modalities for breast tomography, we discuss the advantages and disadvantages of only two types of these methods in this paper which are based on pure optical and pure acoustic imaging. The photoacoustic (PA) approach combines these two methods and takes the best of both. The main limitation of the pure optical imaging modalities for breast cancer detection, such as diffuse optical tomography (DOT), fluorescence molecular tomography (FMT) (Corlu *et al.*, 2007), and optical coherence tomography (Boppart *et al.*, 2004) is the scattering of light in the environment. It lowers the spatial resolution in deep tissue imaging. Light photons can penetrate to a depth of up to 1 mm of living tissue without scattering, as they are based on ballistic and quasi-ballistic photons (Wang, 2008). The scattering of ul-

trasonic waves in living tissue is approximately 2–3 orders of magnitude weaker than in optical waves; therefore, pure ultrasonic imaging has a better resolution than optical imaging. However, in pure acoustic imaging, the image's contrast is determined by the mechanical and elastic properties of the living tissue. Therefore, the contrast of the images may not be suitable for the early detection of a cancerous tumor (Wang, 2004). Photoacoustic imaging (PAI) has become an attractive, low-risk, and non-invasive tool in imaging biological tissues using non-ionized short laser pulse absorption and receiving PA signals. The PAI is based on the PA phenomenon discovered in 1880 by Alexander Graham Bell. It refers to the generation of acoustic waves by the absorption of electromagnetic energy such as optical or radiofrequency. The use of this phenomenon in medical imaging had not been formed before the advent of the laser light. To generate an acoustic pressure wave in the PA phenomenon, the temperature changes in the sample must be variable which can be achieved by the pulsed laser light or continuous laser light with the power intensity modulated at a constant or variable frequency. Due to the high signal-to-noise ratio of pulsed laser light, we use it in PAI, by considering the maximum permissible pulse energy and the maximum permissible pulse repetition rate governed by the American National Standards Institute (ANSI) laser safety standards. To use the PA phenomenon in tissue imaging and to obtain information about the density, sound speed, and geography of the tissue, a short (instead of long) homogenized pulsed laser light should be used (Wang, 2017). Then, by sampling information and signals of the ultrasonic detector and choosing appropriate image reconstruction algorithms, the tissue can be imaged. The use of PAI for breast tissue in the early diagnosis of breast tumors and cancer has been expanded in recent decades because it combines the high optical contrast of tissue chromophores and the high spatial resolution of pure ultrasonic imaging (Lin *et al.*, 2018). According to angiogenesis of breast cancer and subsequent mapping of endogenous breast chromophore concentration including oxy- (HbO_2) and deoxyhemoglobin (HHb), PAI could show high spatial resolution and sensitivity in the early detection of tumors, including dense breas (Laufer *et al.*, 2006). Additionally, tissue imaging using the PA phenomenon (Silverman *et al.*, 2010), especially in the early detection of cancerous tumors, has been extensively studied. It is also possible to detect and track different materials in various environments such as liquid and gas by using the PA phenomena (Pogorzelski *et al.*, 1999; Ponikwicki *et al.*, 2019).

2. Materials and methods

Following this section: we describe the different simulation software that we have used in various steps of the PA phenomenon and finally choose the most suitable simulation software. Later the governing equations of the PA phenomenon are briefly discussed. We show the simulated shape of tissue in our COMSOL simulation by expressing the actual structure of breast tissue and its different parts. Different modules of COMSOL software and their necessary parameters have also been stated by relevant mathematical relations. In addition, the image reconstruction method has been discussed using MATLAB code for the reconstruction of 2D tomographic image slices of breast tissue. The remaining sections concern the study of reports resulting from simulation.

2.1. Simulation of the photoacoustic phenomenon using different software

To have a risk-free environment and insight into dynamics, visualization, increasing accuracy, handling uncertainty, and saving money and time, we first decided to use PAI simulation modeling. There are some methods for simulating different parts of the PA phenomenon. In the following subsections, we briefly describe some of the techniques that we have used and choose the best one.

2.1.1. Software for modeling the homogenizing system of laser light

To transfer the laser energy uniformly to the 2D tomographic slices, in the 3D imaging process of the tissue, it is critical to shape and homogenize the laser light with suitable optical systems. For this purpose, various methods such as refractive, diffractive, and microlens (imaging and non-imaging) beam homogenizers (Ai *et al.*, 2017) are available for which VirtualLab Fusion has been used. This part of the simulation is essential in PA system building.

2.1.2. Simulation of light tissue interaction and PA generation and propagation

The Monte Carlo method is an approach used to solve forward modeling problems in many different fields of physics. It is also used to simulate how photons are transported and absorbed in turbid media. The Monte Carlo model of steady-state light transport in multilayered tissue (MCML) helps us determine the amount of light absorption in the studied tissue (Wang, Jacques, 1992; Jacques, Wang, 1995). After acquiring light tissue interaction data, for simulation of PA generation and propagation, we must import the absorption data matrix into the K-wave MATLAB toolbox and simulate the time-domain PA propagating for homogeneous or heterogeneous medium in one, two, and three dimensions (Treeby, Cox, 2010).

Furthermore, a finite element (FE) – based simulation model has been developed incorporating light propagation, PA signal generation, and sound wave prop-

agation, in soft tissues using a commercial FE simulation package, COMSOL Multiphysics. By using the simulation mentioned in the first part, MCML for light and tissue interaction, and K-wave MATLAB toolbox for photoacoustic propagation, the simulations are intermittent and incoherent. However, MCML is very time-consuming (Cassidy *et al.*, 2018; Wang *et al.*, 2012). In the case of MCML and K-wave toolbox, we cannot maneuver on the actual shape, size, and geometry of the tissue (Akhlaghi *et al.*, 2019). According to the above-mentioned reasons and the fact that the PA phenomena is multiphysics, it seems that the FEM should be a suitable option for this simulation. The FEM method is an efficient and accurate approach for PA phenomena simulation. Thus, we switch to a single platform, COMSOL, to carefully model and solve the problems expressed.

2.1.3. The final software selected for this research

This research has used commercially available finite element software (COMSOL) as a single platform for simulating PAI in the breast tissue. Based on the latest reports (Sowmiya, Thittai, 2017) concerning numerical modeling of PA, the tissue should be performed in tiny dimensions and an unrealistic shape to reduce the calculations and complexity of simulations. With regards to our modeling, the actual dimensions of breast tissue with its various components, including muscle, mammary glands, and fat, have been considered. Each stage of the PA signal generation process, including the effect of penetration of the radiant laser, the conversion of absorbed light into heat, the expansion of the tissue in volume and stress generation, and consequently mechanical expansion generating acoustic pressure wave propagating in the tissue would be fully modeled in COMSOL. Moreover, by changing the tumor location to different areas of breast tissue and its diameter, we maneuver the PA signal resulting from different conditions and positions. Besides all this, the arrangement of ultrasonic sensors is very similar to the real state. Finally, the PA signal obtained from an array of the detector on breast tissue in COMSOL software is converted into images by image reconstruction algorithm codes written in MATLAB.

2.2. A brief description of mathematical equations

Using linear fluid dynamics equations, a pair of coupled differential equations for temperature and pressure can be developed. According to Morse and Ingard, changes in temperature (T) and pressure (P) from their ambient values can be expressed as follows (Wang, 2017):

$$\frac{\partial}{\partial t}\left(T - \frac{\gamma - 1}{\gamma\alpha}P\right) = \frac{k}{\rho C_p}\nabla^2 T + \frac{H}{\rho C_p} \tag{1}$$

and

$$\left[\nabla^2 - \frac{\gamma}{v_s^2}\frac{\partial^2}{\partial t^2}\right]P = -\frac{\alpha\gamma}{v_s^2}\frac{\partial^2}{\partial t^2}T, \tag{2}$$

where γ is the specific heat ratio, α (pressure expansion coefficient) $= \left(\frac{\partial P(\text{pressure})}{\partial T(\text{temperature})}\right)_{V(\text{volume})}$, k is the thermal conductivity, C_p is the specific heat, p is the ambient density, v_s is the sound speed, t is the time, and H is the energy per unit volume and time deposited by the optical radiation beam (Wang, 2017).

The aforementioned equations did not include the effects of viscosity or energy relaxation. Assuming $\gamma = 1$, which for most fluids is a common and acceptable assumption, and assuming heat conductivity is set to zero ($\nabla T = 0$), then Eqs. (1) and (2) reduce to an equation for heat diffusion and a wave equation relating pressure to the speed of sound:

$$\left[\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2}\right]p = -\frac{\beta}{C_p}\frac{\partial H}{\partial t}, \tag{3}$$

where $\beta = \left(\frac{1}{V}\right)\left(\frac{\partial V}{\partial T}\right)_p$ and $\alpha = \frac{\rho\beta v_s^2}{\gamma}$, V is volume, and T is temperature (Wang, 2017).

To improve the optimal feature of the PA signal, the duration of pulsed laser light radiation (τ) must be a few nanoseconds to include the following two conditions:

– thermal confinement, which represents the neglect of thermal diffusion during laser light irradiation (Li, Wang, 2009):

$$\tau < \tau_{th} = \frac{d_c^2}{4D_T}, \tag{4}$$

where τ_{th} is the thermal confinement threshold, D_T is the thermal diffusivity, and d_c is the desired special resolution;

– stress confinement means neglecting the volume expansion of the absorber during laser light irradiation:

$$\tau < \tau_{st} = \frac{d_c}{V_S}, \tag{5}$$

where τ_{st} is the stress confinement threshold and V_S is the speed of sound.

Under two conditions, irradiation time can be treated as a delta function (Li, Wang, 2009).

Therefore, we are able to rewrite Eq. (3) in this form:

$$\left[\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2}\right]p(r,t) = -\frac{p_0}{v_s^2}\frac{\partial\delta(t)}{\partial t}. \tag{6}$$

The acoustic wave $p(r,t)$ at the position (r) and time (t), prompted by an initial source, $p_0(r) = \Gamma(r)\,A_e(r)$, where $A_e(r)$ is a spatial electromagnetic (EM) absorption function and $\Gamma(r) = v_s^2\beta/C_p$ is the Grüneisen parameter (Wang, 2017).

2.3. FEM model

2.3.1. Breast model in simulation

Breast tissue consisting of some similar sections is called the lobe, which is made up of many smaller parts called lobules. The milk produced from the lobe is passed to the nipple through tiny vessels named ducts. Fibrous tissue and fat fill the spaces between the lobules and ducts. In addition, fat is present in different regions: subcutaneous, retro-mammary, and intra-glandular. Figure 1 shows a cross-section of the human mammary in detail. Most of the initial development of cancer occurs inside the glandular tissue. Breast cancer can be in either lobes or ducts. Based on the starting location, ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) are the two main types of in situ breast cancer (American Cancer Society, 2019). This study aims to model breast tissue in real dimensions for accurate screening of PA signals. In modeling, different parts of the breast tissue are separated to discriminate the distinct characteristics of each part required in the steps of the PA signal propagation.

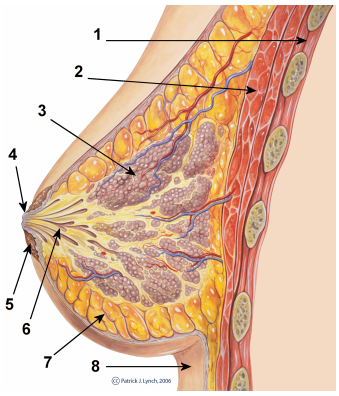
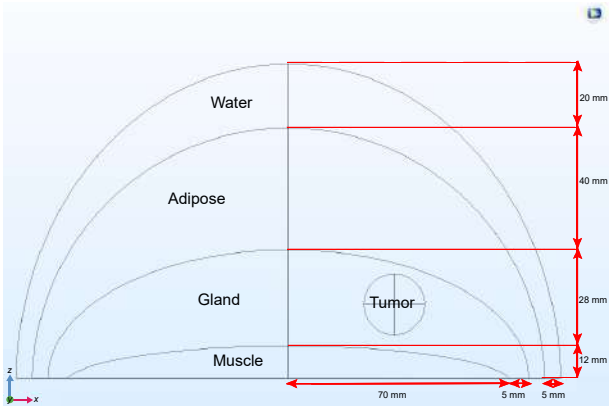


Fig. 1. Cross-sectional image of the normal breast: 1 – chest wall; 2 – pectoralis muscles; 3 – lobules; 4 – nipple; 5 – areola; 6 – milk duct; 7 – fatty tissue; 8 – skin (Lynch, Jaffe, 1987).

The breast has been modeled as a hemisphere with different layers, as shown in Fig. 2. A tumor at a diameter under 20 mm with a spherical shape has been supposed to investigate early-stage cancer (Hammer *et al.*, 2008). Breast tissue is considered as a hemisphere with a diameter of 80 mm. The size of the breast is based on the actual available sizes (Bengtson, Glicksman, 2015). The entire tissue is enclosed inside a semi-ellipsoid chamber of water. The inner layers of breast tissue are simulated as semi-ellipsoid layers consisting of the gland, adipose, and muscle of appropriate size in realistic dimensions. The laser source is located as a point source near tumor side of the breast tissue in the aqueous environment, the distance of laser source to the tissue is about 3 mm in coordinates as follows: $x = 77$ mm, $y = 36$ mm, $z = 53$ mm (Fig. 2b).

a)



b)

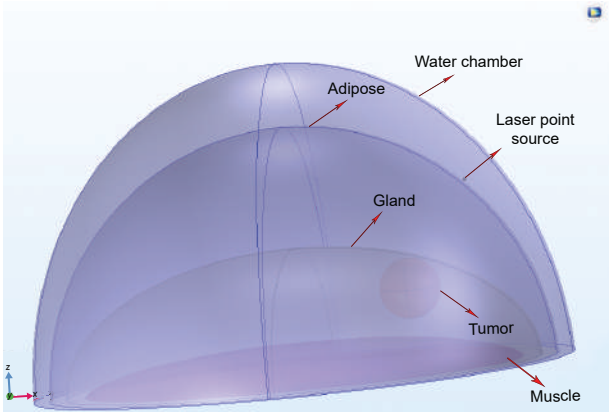


Fig. 2. a) Schematic of the cross-sectional view of the breast model at LOT1 scenario with dimensional details; b) 3D computational domain of breast model in COMSOL Multiphysics.

For a complete study of simulation and subsequently PA signal, we have considered various scenarios for breast tumor, different positions, and sizes, which change in each scenario as abbreviated and listed in Table 1. In all scenarios mentioned, the location and power of the laser light has not changed. Three scenarios for tumor location change were defined: 1) entirely inside the glandular tissue (LOT1); 2) entirely inside the adipose (LOT2); 3) between the gland and adipose (LOT3). Also, compared to the base scenario: (LOT1), in LOT2 and LOT3, the tumor location just along the z-axis increased by 10 and 25 mm, respectively.

Table 1. Various scenarios considered in the simulation.

Scenario name	Location of tumor	Diameter of tumor [mm]
LOT1	Inside gland	19
LOT2	Inside adipose	19
LOT3	Between gland and adipose	19
DOT1	Between gland and adipose	40

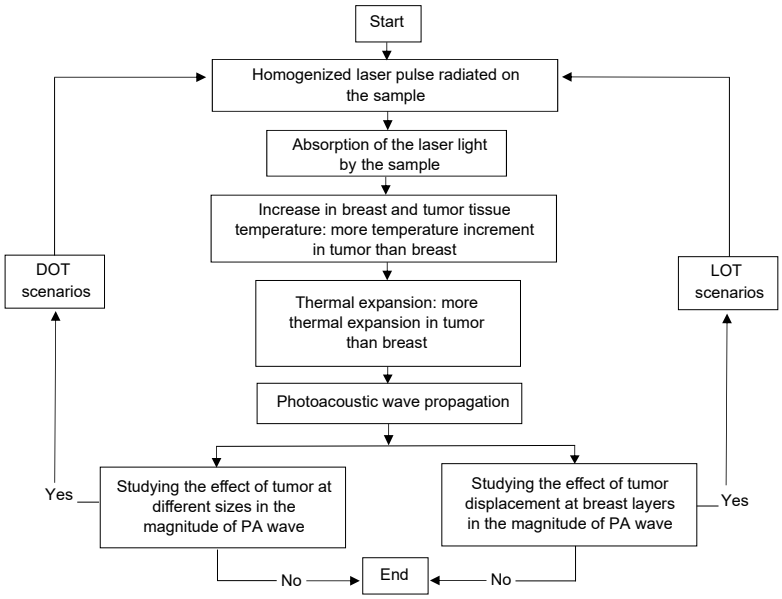


Fig. 3. Flowchart description of the procedure of the PAI and the simulations.

In the scenario of increasing the tumor size (DOT1), the tumor radius compared to previously defined scenarios has increased from 9.5 mm to 20 mm without changing the coordinate position of the tumor. However, due to the increase in tumor diameter, it will be located between adipose tissue and glands. For a better conception of the PAI process and the procedure of the simulations, a schematic flowchart is illustrated in Fig. 3.

We used four modules of COMSOL Multiphysics to simulate PA propagation in the breast. We described all of these in the following.

2.3.2. Light propagation

The initial local acoustic pressure wave (p_0) could also be estimated with the following equation to evaluate the amount of EM energy absorption effectively (Wang, 2017):

$$p_0(r) = \Gamma \eta_{th} \mu_a F, \tag{7}$$

where Γ is the Grüneisen parameter, η_{th} is the percentage of optical absorption that is converted into heat, μ_a is the optical absorption coefficient [cm^{-1}], and F is the optical fluence [J/cm^2]. The optical fluence of radiated laser light is directly related to the amplitude of acoustic pressure waves. Therefore, to obtain a strong PA signal, we must maximize the laser fluence on tissue. In the diffusion regime:

$$F(r) = \frac{F_0}{4\pi D r}\, e^{-(\mu_{\text{eff}})r}, \tag{8}$$

where r is the distance from the source, F_0 is the fluence at the source, D is the diffusion constant ($D = 1/(3(\mu_a + \mu_s'))$), and $\mu_{\text{eff}} = (\sqrt{\mu_a/D})$ is the effective attenuation coefficient. According to Eq. (8), the penetration depth of the laser is defined by (μ_{eff}) which is dependent on the laser wavelength. In PAI, red, and near-infrared (NIR) wavelengths usually are preferred because, in this wavelength range, the optical attenuation of biological tissue is at its lowest level. HbO_2, HHb, and water are assumed to be the main absorbers in this study. In particular, a wavelength of around 800 nm is the isosbestic point of the molar extinction spectra of HbO_2 and HHb (Wang, 2017). Here, we modeled a radiated laser source as the Gaussian pulse in the 800 nm wavelength, with power, $W_p = 8$ mJ/cm^2, pulse duration $\tau_p = 10$ ns, and $\tau_{\text{center}} = 30$ ns. The radiated laser fluence is modeled by the "coefficient form PDE" interface in COMSOL. The optical properties of the tumor and glandular tissue depend on the assumed magnitude of absorbers that have a dramatic effect at the given wavelength. Once the optical properties are obtained at each wavelength, we can calculate the absorption coefficient of light, μ_a [cm^{-1}], of an absorber by using one of these equations (Jacques, 2013):

$$\mu_a = -\frac{1}{T}\frac{\partial T}{\partial L}, \tag{9}$$

$$T = e^{-\mu_a L} = 10^{-\varepsilon C L} = e^{-4\pi n'' L/\lambda}, \tag{10}$$

where T (dimensionless) is transmitted or surviving fraction of the incident light after an incremental path length ∂L [cm], n'' is the imaginary refractive index of the medium, ε is the extinction coefficient [cm^{-1}M^{-1}], and C(M) is the concentration of chromophore. The absorption coefficient of a tissue is the sum of all main absorbers:

$$\mu_a = \ln(10) \sum_i C_i \varepsilon_i. \tag{11}$$

Table 2. Related parameters of the breast tissue sections and tumor at 800 nm in the first module of COMSOL.

Materials	Density ρ [kg/m^3]	Refractive index	μ_a at 800 nm [mm^{-1}]	μ_s' at 800 nm [mm^{-1}]
Adipose	930 [a, b]	1.455 [c]	0.00193	1.52
Glandular	1050 [a, b]	1.4 [c]	0.00436	1.12
Muscle	1030 [d]	1.4 [d]	0.230 [d]	0.785 [d]
Water	994 [b]	1.3290 [e]	0.00196 [f]	0.0000402 [e, f]
Tumor	1050 [a]	~1.4 [g]	0.00302	0.625

[a](Soltani *et al.*, 2019); [b](Hasgall *et al.*, 2018); [c](Dehghani *et al.*, 2005); [d](Wang *et al.*, 2012); [e](Hale, Querry, 1973; Polyanskiy, 2016); [f](Downing, 2008); [g](Metwally *et al.*, 2014).

Table 3. Required parameters in the "bioheat transfer" module.

Materials	Thermal conductivity k [W·m^{-1}·k^{-1}]	Specific heat capacity C_p [J·kg^{-1}·K^{-1}]	Coefficient of thermal expansion β [K^{-1}]
Adipose	0.21 [a]	2770 [a]	3E-5 [a]
Glandular	0.48 [a]	3770 [a]	4.5E-5 [a]
Muscle	0.49 [b]	3421 [b]	26.73E-5 [c]
Water	0.60 [d]	4178 [d]	~24E-5 [e]
Tumor	0.54 [a]	3852 [a]	6.5E-5 [a]

[a](Soltani *et al.*, 2019); [b](Singh, Repaka, 2018); [c](Wilkie, 1953, Lan *et al.*, 2019); [d](Hasgall *et al.*, 2018); [e](Oraevsky *et al.*, 2001).

Also, the reduced scattering (or transport scattering) coefficient of light, μ_s' [mm^{-1}], the spectrum of tissue has been shown to fit well to an empirical approximation to the Mie scattering theory given by (Dehghani *et al.*, 2009):

$$\mu_s' = a\lambda^{-b}, \tag{12}$$

where a and b are the scatter amplitude and the scatter power, respectively, at any wavelength in μm.

The percentage of absorbent ingredients of tumor, glandular tissue, and adipose of breast is based on (Dehghani *et al.*, 2009). The spectral characteristics of HHb, HbO_2, and water at 800 nm are taken from references (Hale, Querry, 1973; Prahl, 2017). The optical parameters of breast tissue sections and tumor used in "coefficient form PDE" module at COMSOL are listed in Table 2.

We entered the effect of radiant laser light as a point source in the finite element-based numerical model irradiated onto breast tissue surrounded by an aqueous medium. The boundary conditions are also based on references (Wang *et al.*, 2012).

2.3.3. Heat transfer

The second part of the acoustic pressure wave production in the PA phenomenon process is converting the penetrating energy of the radiant laser into heat of which the tumor will have the highest absorption amount according to its material properties. This part can be described by the "bioheat transfer" module of COMSOL. We can connect this module by the heat source that can be inserted in an equation by multiplying the optical absorption coefficient and the optical fluence rate estimated previously. The required thermal properties of breast tissue and tumor used in the heat transfer module are listed in Table 3.

2.3.4. Thermal expansion

Due to the increase in tumor temperature, the tumor subsequently undergoes thermoelastic expansion, which was modeled using a "structural mechanics" module. All related parameters entered in this module are listed in Table 4.

Table 4. Related parameters in the "structural mechanics" module.

Materials	Poisson's ratio
Adipose	0.49 [a]
Glandular	0.49 [a]
Muscle	0.45 [b]
Water	0.4995 [c]
Tumor	0.49 [d]

[a](Bhatti, Sridhar-Keralapura, 2012; Gefen *et al.*, 2007; Soltani *et al.*, 2019); [b](Grimal *et al.*, 2005); [c](Gefen *et al.*, 2007); [d](Soltani *et al.*, 2019).

2.3.5. PA signal generation

Finally, the pressure wave propagation can be studied using the "pressure acoustics, transient" module in COMSOL. Table 5 lists all related parameters in the module of this section.

Table 5. All related parameters in the last module of COMSOL.

Materials	Speed of sound C_s [m/s]	Young's modulus [Pa]
Adipose	1440.2 [a]	18E3 [b]
Glandular	1505.0 [a]	50E3 [b]
Muscle	1588.4 [a]	675E3 [c]
Water	1482.3 [a]	Bulk modulus = 2.1790576 GPa [d]
Tumor	1540 [e]	106E3 [b]

[a](Hasgall *et al.*, 2018); [b](Soltani *et al.*, 2019); [c](Grimal *et al.*, 2005); [d](Zhutovsky, Kovler, 2015); [e](Li *et al.*, 2009).

2.3.6. Image reconstruction

In this section, we describe the algorithm used in the image reconstruction section. There are various algorithms and their reformation for reconstruction of a proper image such as universal back-projection, time reversal, etc. (Dobrucki, Opieliński, 2000; Tasinkevych *et al.*, 2018, Wang, 2017). For PAI reconstruction, a phase-controlled algorithm was used in this study. For each reconstructed pixel in the image, $P_{\mathbf{m}}$, we have the following formula (Zhou *et al.*, 2011):

$$P_{\mathbf{m}} = \sum_{k=1}^{K} D(\theta)_k P(t_{mk}), \tag{13}$$

where $\mathbf{m}$ is the spatial vector of the PA source, k is the position of the k-th detector, and K is the total number of working detectors, $D(\theta)$ is defined as the projection intensity weight function of θ, the projection angle which is no more than the maximal acceptance angle of the array element, $P(t_{mk})$ is the signal value collected by the k-th detector at position $\mathbf{m}$. The t_{mk} represents the time when PA pulses spread from position $\mathbf{m}$ to k. The r_{mk} is the distance between the PA source at position $\mathbf{m}$ and the k-th detector. The v is the average velocity of the acoustic wave in tissue. According to Eq. (13), the volumetric data collected by the plane transducer array in a single laser pulse can be used to reconstruct 2D projection image slices at various z-axis depths. To have the final 3D image of the breast, proper processing software could merge these slices.

3. Results and discussion

Using the finite element (FE) based simulation models, light propagation, absorption, conversion to heat, PA wave formation from the target (tumor), and its propagation in breast tissue were successfully simulated. For a better demonstration of our simulation process, captured PA signals in the L1 model of COMSOL simulation have been converted to 2D tomography image slices using MATLAB code. For 3D data visualization, processing, and analysis we can convert these slices to 3D images using special software. Converting 2D image slices to 3D breast tissue have not been done by the relevant software in this project. Finally, comparative results of all simulations concerning the tumor location and size were studied in detail.

3.1. Simulation model in detail

A laser source as the Gaussian pulse radiated on tissue as a point source is showed in Fig. 4. The coordinates and geometric position of discussed points in this section are shown in Fig. 5. We have chosen 6 points, two points inside and outside of the tumor near the wall to study the accuracy of the tumor edge detection in our PAI simulation, a point inside the adipose region, a point in the glandular region, and the two other points inside the tumor on the other axis for studying stress.

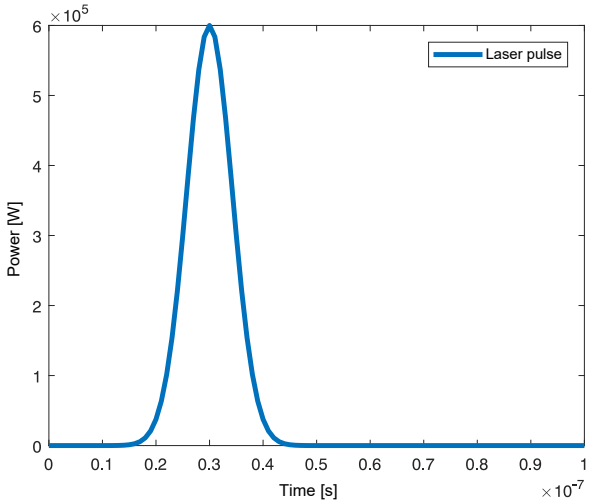


Fig. 4. Function diagram of laser pulse radiated as a point source on tissue, with power W_p = 8 mJ/cm^2, pulse duration τ_p = 10 ns, and τ_{center} = 30 ns.

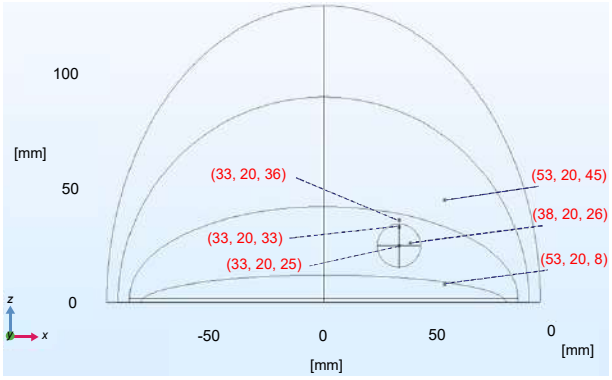


Fig. 5. Coordinate location of points mentioned in simulation studies.

The radiated laser fluence shown in Fig. 6, represents the laser fluence variation in time at four points

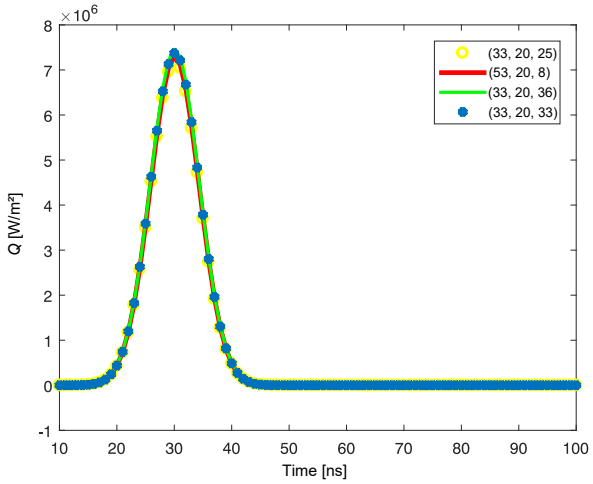


Fig. 6. Radiated laser fluence of four points: inside and outside the tumor close to the wall, center of the tumor and inside a muscle layer, in the L1 simulation model.

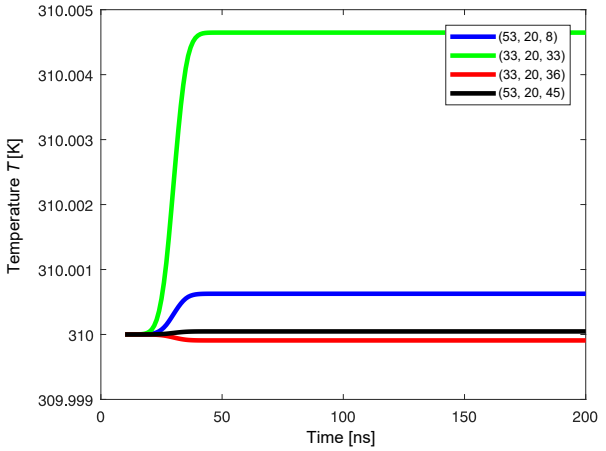


Fig. 7. Temperature changes in Kelvin, four points: inside and outside the tumor, one in glandular and the other in the adipose regions according to time variables.

of interest. The next step is converting the effect of laser light to heat and studying the behavior of tissues and tumors.

The temperature variations of four points are selected as representative of four places: inside and outside the tumor, near the wall of the tumor to determine the accuracy of detecting the edges of the tumor, a point in the muscular region, and a point in the glandular region. These are shown in Fig. 7. As is shown in Fig. 7, due to the amount of absorption of the laser energy, the temperature change of points is different, it is greater for the point inside the tumor and depends on the different optical and thermal characteristics of the tumor (all these specifications are listed in Tables 1 and 2) compared to breast tissue. The increase in temperature inside the tumor is approximately 0.0046 degrees in Kelvin, while the temperature of the study points slightly outside the tumor wall and in the adipose region has not changed much. Due to muscle tissue's thermal and optical characteristics, a slight increase in temperature is observed at the point located in this section, which can be ignored according to the increased value in the tumor temperature. In all calculations, the human body temperature, 310 degrees in Kelvin, is considered.

For further clarity on temperature changes in tissue and tumor, a cut surface was considered in the z–y coordinate plane passing through the center of the tumor in x coordinate (x = 33). Figure 8 indicates an increase in a temperature gradient in the tumor region compared to other parts of the tissue at time = 30 ns. High levels of the laser light absorption by the tumor tissue and its conversion to heat are the reason for the increase in heat of tumor tissue compared to other parts of breast tissue.

Increasing the temperature of the tumor compared to other parts of the tissue will cause the tumor tis-

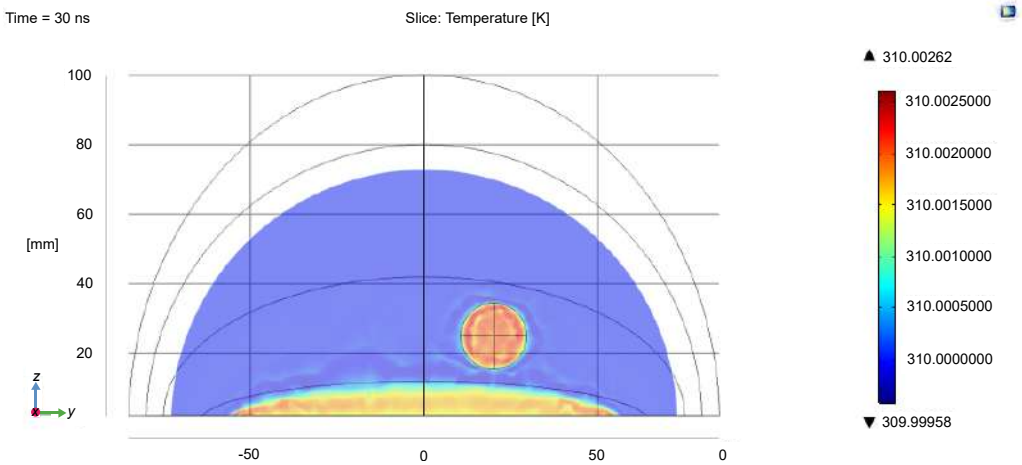


Fig. 8. Temperature gradient distribution of breast and tumor tissues on cut surfaces in the z–y plane at x = 33 mm and t = 30 ns.

sue to expand. Figure 9 shows the internal expansion changes of the three internal points of the tumor. The increase in the internal expansion of the tumor is about 3 N/m^2.

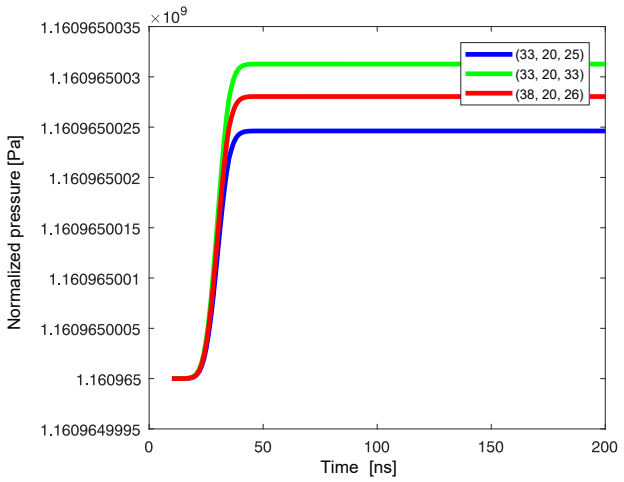


Fig. 9. Internal expansion changes of the three internal points of the tumor.

Expansion changes inside the tumor will cause an acoustic wave inside the tumor tissue and propagate it into the breast tissue. In the latest module of COMSOL, we have studied the propagation of acoustic pressure waves around tumor tissue. The images of the acoustic pressure wave propagation gradient in the surface z–y at x = 33 mm are shown in Fig. 10 in two steps and times as the first cycle of acoustic wave propagation.

The last step is to transfer information obtained from COMSOL simulation to MATLAB software and reconstruct the 2D tomographic image slices. The sensor's data that contain ultrasonic wave information are defined as "Domain Point Probe" in COMSOL, and 32 number of these "Domain Point Probes", according to previous statements, are located around the breast cup on the arc of 2D tomographic slices. Fifteen 2D tomographic slices of breast tissue have been considered according to Fig. 11, while these slices cutting the breast tissue Perpendicular to the x-axis, are located at a distance of about 1 cm from each other. Each "Domain Point Probe" information is exported from COMSOL as a matrix whose dimension depends on the simulation time and number of sensors.

The exported matrices are to be used for subsequent processing in MATLAB. The data of four sensors is simply drawn in Fig. 12.

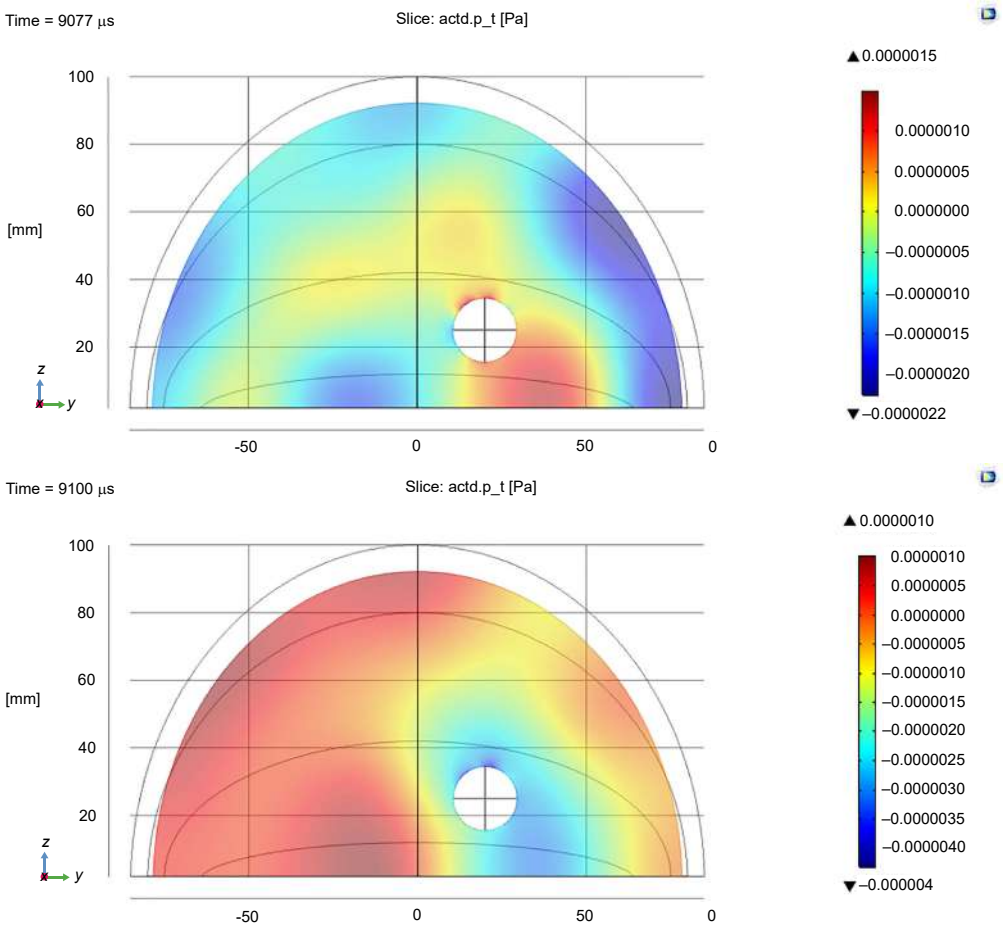


Fig. 10. Acoustic pressure wave gradient distribution at two steps and times on cut surfaces in the z–y plane at x = 33 mm.

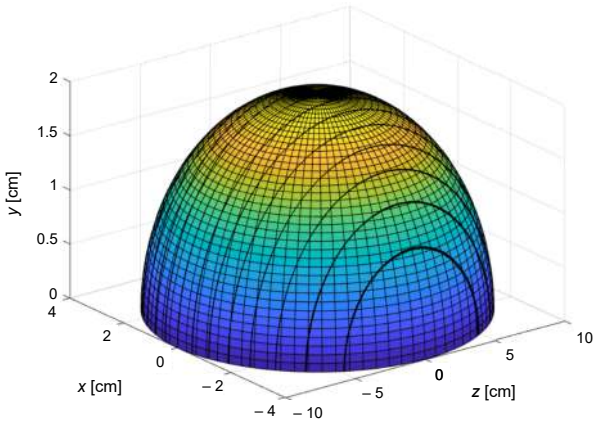


Fig. 11. Fifteen 2D tomographic slices of breast tissue used in this simulation.

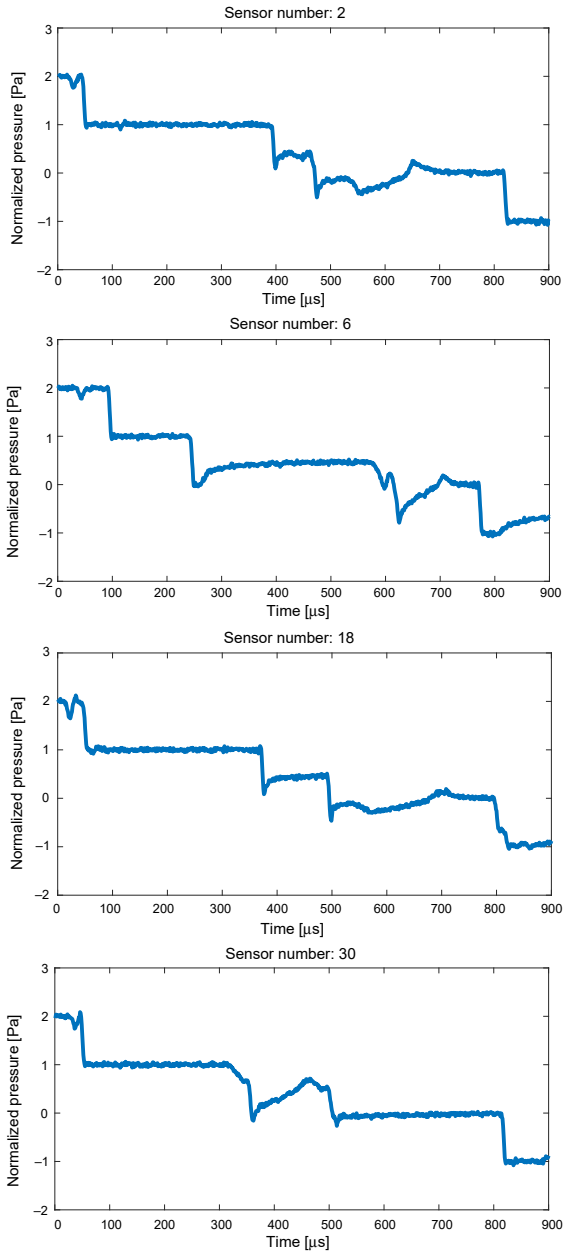


Fig. 12. Simple plotting of four ultrasonic sensors data.

2D tomographic reconstructed image slices for 3D image creation should be repeated in different coordinates; here we just showed two slices in Fig. 13: a cut surface, the y–z surface, at x = 33 mm and x = 37 mm. MATLAB code results for the reconstruction took a 1 minute processing. Finally, by Amira (version 5.5, TGS Template Graphics Software) software, we could visualize these 2D reconstructed slices to 3D, for operator of device, but here for our analyzing the 2D slices and sensor's data are adequate.

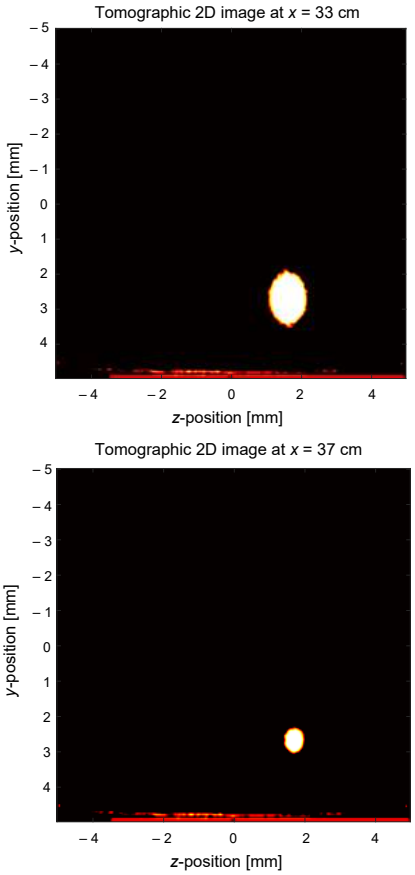


Fig. 13. 2D tomographic image in one slice of simulation, derived from MATLAB code, at x = 33 mm and x = 37 mm.

According to the anatomy of the breast tissue, the muscle layer is at the junction of the breast tissue with the chest wall, and due to the proximity of the characteristics of muscle and tumor, a certain error in image reconstruction is shown. This error could be easily removed based on the fixed location of the muscle and its specific shape and brightness.

3.2. Effect of tumor location

According to Table 1 and the different scenarios we have considered for the tumor location and size, in this section, we study the changes in temperature and intensity of the emitted pressure wave based on the change in the tumor location compared to the LOT1 mode.

According to the chart in Fig. 14 and the values of temperature changes for the tumor in different places, it was observed that for the case where the tumor was located in the adipose tissue, the absorption of radiant laser energy by the tumor would be higher than others; consequently, temperature changes were high.

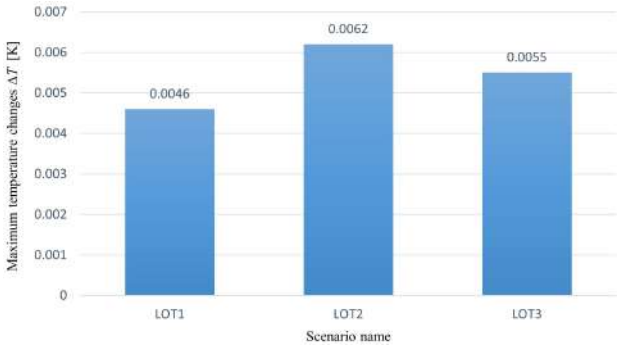


Fig. 14. Temperature changes based on different scenarios.

Based on the chart in Fig. 15, the amount of pressure change created in the tumor located in the adipose tissue is also high. This is because of the large difference in optical and thermal properties between the tumor tissue and the adipose tissue. By studying the tissue property tables (Tables 2 to 4), the cause can be observed.

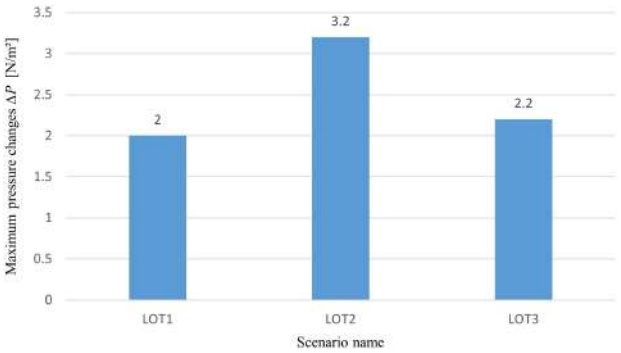


Fig. 15. Pressure changes based on different scenarios.

Increasing the temperature and pressure changes will cause the acoustic pressure wave to propagate more intensely, resulting in more accurate and easier tumor detection.

The final result of our study for the tumor location in different layers of breast tissue is completely consistent with the results of breast imaging by the thermoacoustic method (Soltani *et al.*, 2019). In both methods, the tumor placement in the adipose tissue results in a strong PA wave. In other words, it would be easier to diagnose a tumor that is located in the adipose section compared to other layers.

3.3. Effect of tumor size

To study the effect of increasing the diameter of the tumor, two scenarios have been considered: one with a diameter of 19 mm (LOT1); the other with 40 mm (DOT1). The choice of a tumor with a diameter of less than 2 cm is to evaluate the ability of PAI to detect the tumor in the early stages (Hammer *et al.*, 2008).

Based on the charts in Figs. 16 and 17, which show a comparison of the increase in temperature and pressure changes in the two scenarios respectively, it is obvious that the magnitude of the temperature change of a large-diameter tumor is greater than that in a small-diameter tumor.

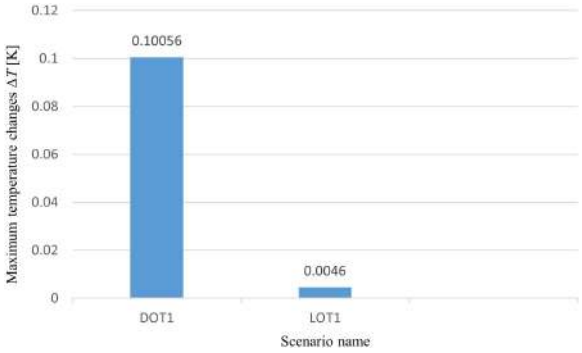


Fig. 16. Temperature changes based on different scenarios.

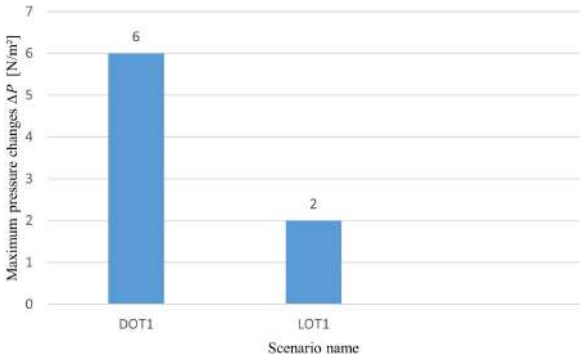


Fig. 17. Pressure changes based on different scenarios.

The reason for this is obvious. As the size of the tumor increases, the amount of energy absorbed by the radiant laser light will be higher. Compared to the thermoacoustic method (Soltani *et al.*, 2019), our results also indicate that the amplitude of the PA wave will increase with increasing tumor diameter.

4. Conclusions

In this study, breast tissue tomography was simulated using the PA phenomenon in real dimensions and different constituent layers in suitable and efficient software. To select the proper software, various software were examined to select the most suitable option. Attempts were also made to consider the different layers that make up breast tissue and enter all the required specifications in the simulation completely and accurately. In all the stages, the simulation process tried to be as close as possible to the real state to

cite the results of this simulation implementation of the PAI system and evaluate the strengths and weaknesses of the PA breast tissue imaging system. The simulations were performed for four different scenarios based on the size and location of the tumor to investigate the changes in the resulting PA signal in different states. It is possible to define different scenarios. In each scenario, the magnitude of temperature changes and consequently the acoustic pressure magnitude were obtained. Due to the optical, thermal, elastic, and acoustic characteristics of different layers of breast tissue and tumor which show a greater difference between the adipose tissue and the tumor, it was expected that the absorption of radiant laser light will be high if more tumor forms in the adipose layer of breast tissue.

Consequently, temperature changes and the resulting acoustic pressure will have a large amplitude. However, the acoustic pressure magnitude will be the lowest for tumors located in the glandular layer and make diagnosis more difficult. Also, by examining the simulation for a tumor with a larger diameter and the consequent increase in the acoustic pressure amplitude, it is quite clear that large tumors can easily be identified. But due to the capability of the PA phenomenon and incorporation of optical and ultrasound imaging advantages, it was able to detect tumors smaller than 2 cm in size so that cancerous tumor tissue could be detected in early stages.

Acknowledgments

We wish to thank Duckman Proofreading for the English language editing. The author acknowledges the usage of the COMSOL Multiphysics® software version 5.4 for Linux at a data processing center of the Sahand University of Technology, Tabriz, Iran.

Data availability

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

References

1. Ai M., Shu W., Salcudean T., Rohling R., Abolmaesumi P., Tang S. (2017), Design of high energy laser pulse delivery in a multimode fiber for photoacoustic tomography, *Optics Express*, **25**(15): 17713–17726, doi: 10.1364/OE.25.017713.

2. Akhlaghi N., Pfefer T.J., Wear K.A., Garra B.S., Vogt W.C. (2019), Multidomain computational modeling of photoacoustic imaging: Verification, validation, and image quality prediction, *Journal of Biomedical Optics*, **24**(12): 121910, doi: 10.1117/1.JBO.24.12.121910.

3. American Cancer Society (2019), Breast cancer facts & figures 2019–2020, *American Cancer Society, Inc.*, https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf.

4. Bengtson B.P., Glicksman C.A. (2015), The standardization of bra cup measurements: Redefining bra sizing language, *Clinics in Plastic Surgery*, **42**(4): 405–411, doi: 10.1016/j.cps.2015.06.002.

5. Bhatti S.N., Sridhar-Keralapura M. (2012), A novel breast software phantom for biomechanical modeling of elastography, *Medical Physics*, **39**(4): 1748–1768, doi: 10.1118/1.3690467.

6. Boppart S.A., Luo W., Marks D.L., Singletary K.W. (2004), Optical coherence tomography: feasibility for basic research and image-guided surgery of breast cancer, *Breast Cancer Research and Treatment*, **84**(2): 85–97, doi: 10.1023/B:BREA.0000018401.13609.54.

7. Cassidy J., Nouri A., Betz V., Lilge L. (2018), High-performance, robustly verified Monte Carlo simulation with FullMonte, *Journal of Biomedical Optics*, **23**(8): 085001, doi: 10.1117/1.JBO.23.8.085001.

8. Corlu A. *et al.* (2007), Three-dimensional in vivo fluorescence diffuse optical tomography of breast cancer in humans, *Optics Express*, **15**(11): 6696–6716, doi: 10.1364/OE.15.006696.

9. Dehghani H., Brooksby B.A., Pogue B.W., Paulsen K.D. (2005), Effects of refractive index on near-infrared tomography of the breast, *Applied Optics*, **44**(10): 1870–1878, doi: 10.1364/AO.44.001870.

10. Dehghani H. *et al.* (2009), Near infrared optical tomography using NIRFAST: Algorithm for numerical model and image reconstruction, *Communications in Numerical Methods in Engineering*, **25**(6): 711–732, doi: 10.1002/cnm.1162.

11. Dobrucki A.B., Opieliński K.J. (2000), Adaptation of image reconstruction algorithm for purposes of ultrasound transmission tomography (UTT), *Archives of Acoustics*, **25**(4): 395–422.

12. Downing J. (2008), Effects of light absorption and scattering in water samples on OBS measurements, *Campbell Scientific, Inc.*, pp. 1–4.

13. Gefen A., Dilmoney B. (2007), Mechanics of the normal woman's breast, *Technology and Health Care,* **15**(4): 259–271, doi: 10.3233/THC-2007-15404.

14. Grimal Q., Naïli S., Watzky A. (2005), A high-frequency lung injury mechanism in blunt thoracic impact, *Journal of Biomechanics*, **38**(6): 1247–1254, doi: 10.1016/j.jbiomech.2004.06.010.

15. Hale G.M., Querry M.R. (1973), Optical constants of water in the 200-nm to 200-μm wavelength region,

Applied Optics, **12**(3): 555–563, doi: 10.1364/AO.12.000555.

16. Hammer C., Fanning A., Crowe J. (2008), Overview of breast cancer staging and surgical treatment options, *Cleveland Clinic Journal of Medicine*, **75**: S10–S16, doi: 10.3949/ccjm.75.Suppl_1.S10.

17. Hasgall P.A. *et al.* (2018), IT'IS Database for thermal and electromagnetic parameters of biological tissues, Version 4.0, *IT'IS Foundation*, doi: 10.13099/VIP21000-04-0.

18. Jacques S.L. (2013), Optical properties of biological tissues: A review, *Physics in Medicine & Biology*, **58**(11): R37, doi: 10.1088/0031-9155/58/11/R37.

19. Jacques S.L., Wang L. (1995), Monte Carlo modeling of light transport in tissues, [in:] *Optical-Thermal Response of Laser-Irradiated Tissue*, Welch A.J., van Gemert M. [Eds.], pp. 73–100, Springer, Boston, MA, doi: 10.1007/978-1-4757-6092-7_4.

20. Lan H., Duan T., Jiang D., Zhong H., Zhou M., Gao F. (2019), Dual-contrast nonlinear photoacoustic sensing and imaging based on single high-repetition-rate pulsed laser, [in:] *IEEE Sensors Journal*, **19**(14): 5559–5565, doi: 10.1109/JSEN.2019.2902849.

21. Laufer J., Elwell C., Delpy D., Beard P. (2006), Absolute measurements of local chromophore concentrations using pulsed photoacoustic spectroscopy, [in:] *Photons Plus Ultrasound: Imaging and Sensing 2006: The Seventh Conference on Biomedical Thermoacoustics, Optoacoustics, and Acousto-optics*, Vol. 6086, pp. 398–405, doi: 10.1117/12.657372.

22. Li C., Duric N., Littrup P., Huang L. (2009), In vivo breast sound-speed imaging with ultrasound tomography, *Ultrasound in Medicine & Biology*, **35**(10): 1615–1628, doi: 10.1016/j.ultrasmedbio.2009.05.011.

23. Li C., Wang L.V. (2009), Photoacoustic tomography and sensing in biomedicine, *Physics in Medicine & Biology*, **54**(19): R59, doi: 10.1088/0031-9155/54/19/R01.

24. Lin L. *et al.* (2018), Single-breath-hold photoacoustic computed tomography of the breast, Nature communications, **9**(1): 2352, doi: 10.1038/s41467-018-04576-z.

25. Lynch P.J., Jaffe C.C. (1987), Generated for multimedia teaching projects by the Yale University School of Medicine, *Center for Advanced Instructional Media.*

26. Metwally M.K. *et al.* (2014), Influence of optical fluence distribution on photoacoustic imaging, *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering*, **8**(8): 1108–1112.

27. Oraevsky A.A., Karabutov A.A., Savateeva E.V. (2001), Enhancement of optoacoustic tissue contrast with absorbing nanoparticles, [in:] *Hybrid and Novel Imaging and New Optical Instrumentation for Biomedical Applications*, Vol. 4434, pp. 60–69, doi: 10.1117/12.446693.

28. Pogorzelski S.J., Szurkowski J., Śliwiński A. (1999), Detection of micellar structures in oil-water-surfactant systems with a photoacoustic method, *Archives of Acoustics*, **24**(3): 319–330.

29. Polyanskiy M.N. (2016), Refractive index database (access: 2020).

30. Ponikwicki N. *et al.* (2019), Photoacoustic method as a tool for analysis of concentration-dependent thermal effusivity in a mixture of methyl alcohol and water, *Archives of Acoustics*, **44**(1): 153–160, doi: 10.24425/aoa.2019.126361.

31. Prahl S. (2017), *Assorted Spectra* (assess: 2020).

32. Silverman R.H. *et al.* (2010), High-resolution photoacoustic imaging of ocular tissues, *Ultrasound in Medicine & Biology,* **36**(5): 733–742, doi: 10.1016/j.ultrasmedbio.2010.02.006.

33. Singh S., Repaka R. (2018), Numerical investigation of convective cooling in minimizing skin burns during radiofrequency ablation of breast tumor, *Sādhanā*, **43**(90): 1–8, doi: 10.1007/s12046-018-0872-4.

34. Soltani, M., Rahpeima R., Kashkooli F.M. (2019), Breast cancer diagnosis with a microwave thermoacoustic imaging technique – a numerical approach, *Medical & Biological Engineering & Computing*, **57**(7): 1497–1513, doi: 10.1007/s11517-019-01961-8.

35. Sowmiya C., Thittai A.K. (2017), Simulation of photoacoustic tomography (PAT) system in COMSOL and comparison of two popular reconstruction techniques, [in:] *Medical Imaging 2017: Biomedical Applications in Molecular, Structural, and Functional Imaging*, **10137**: 435–445, doi: 10.1117/12.2254450.

36. Tasinkevych Y., Lewandowski M., Klimonda Z., Walczak M. (2018), Synthetic aperture cardiac imaging with reduced number of acquisition channels. A feasibility study, *Archives of Acoustics*, **43**(3): 437–446, doi: 10.24425/123915.

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

38. Wang L., Jacques S.L. (1992), *Monte Carlo Modeling of Light Transport in Multi-Layered Tissues in Standard C.*, University of Texas, MD Anderson Cancer Center.

39. Wang L.V. [Ed.] (2017), *Photoacoustic Imaging and Spectroscopy,* CRC Press.

40. Wang L.V. (2004), Ultrasound-mediated biophotonic imaging: A review of acousto-optical tomography and photo-acoustic tomography, *Disease Markers*, **19**(2–3): 123–138, doi: 10.1155/2004/478079.

41. Wang L.V. (2008), Prospects of photoacoustic tomography, *Medical Physics*, **35**(12): 5758–5767, doi: 10.1118/1.3013698.

42. Wang Z., Ha S., Kim K. (2012), Evaluation of finite-element-based simulation model of photoacoustics in biological tissues, [in:] *Medical Imaging 2012: Ultrasonic Imaging, Tomography, and Therapy*, **8320**: 470–478, doi: 10.1117/12.912152.

43. Wilkie D.R. (1953), The coefficient of expansion of muscle, *The Journal of Physiology*, **119**(4): 369–375, doi: 10.1113/jphysiol.1953.sp004852.

44. Zho Q., Ji X., Xing D. (2011), Full-field 3D photoacoustic imaging based on plane transducer array and spatial phase-controlled algorithm, *Medical Physics*, **38**(3): 1561–1566, doi: 10.1118/1.3555036.

45. Zhutovsky S., Kovler K. (2015), Evaluation of the thermal expansion coefficient using non-destructive testing, [in:] *Proceedings of 110th International Conference on Mechanics and Physics of Creep, Shrinkage, and Durability of Concrete and Concrete Structures, CONCREEP*, **10**: 1137–1146, doi: 10.1061/9780784479346.136.

Archives of Acoustics Vol. 48, No. 1, pp. 39–48 (2023), doi: 10.24425/aoa.2023.144264

Research Paper

Acoustic Identification of Dolphin Whistle Types in Deep Waters of Arabian Sea Using Wavelet Threshold Denoising Approach

Madan M. MAHANTY*, Sanjana M. CHEENANKANDY, Ganesan LATHA, Govindan RAGURAMAN, Ramasamy VENKATESAN

National Institute of Ocean Technology, Ministry of Earth Sciences
Chennai, India

*Corresponding Author e-mail: mmmahanty@gmail.com

(*received February 15, 2022; accepted October 7, 2022*)

In situ time series measurements of ocean ambient noise, have been made in deep waters of the Arabian Sea, using an autonomous passive acoustic monitoring system deployed as part of the Ocean Moored buoy network in the Northern Indian Ocean (OMNI) buoy mooring operated by the National Institute of Ocean Technology (NIOT), in Chennai during November 2018 to November 2019. The analysis of ambient noise records during the spring (April–June) showed the presence of dolphin whistles but contaminated by unwanted impulsive shackle noise. The frequency contours of the dolphin whistles occur in narrow band in the range 4–16 kHz. However, the unwanted impulsive shackle noise occurs in broad band with the noise level higher by ~20 dB over the dolphin signals, and it reduces the quality of dolphin whistles. A wavelet based threshold denoising technique followed by a subtraction method is implemented. Reduction of unwanted shackle noise is effectively done and different dolphin whistle types are identified. This wavelet denoising approach is demonstrated for extraction of dolphin whistles in the presence of challenging impulsive shackle noise. Furthermore, this study should be useful for identifying other cetacean species when the signal of interest is interrupted by unwanted mechanical noise.

Keywords: deep water ambient noise; Arabian Sea; wavelet threshold denoising; impulsive shackle noise; dolphin whistle types.

1. Introduction

Sounds produced by marine species are often identified using time-frequency representations for extraction of salient and distinguishing features of their vocalization. The problem of extracting the sound from the spectrogram can be compounded by low signal to noise ratio and the obstruction of acoustic files with non-stationary anthropogenic sound sources. Wavelet transforms are preferred over a conventional method using the fast Fourier transform (FFT) in identifying a predominant biological noise source using the multiresolution denoising algorithm (Powell *et al.*, 1995; Learned, Willsky, 1995). The wavelet transform is used by Huynh *et al.* (1998) in classifying whale and porpoise sounds. A wavelet based threshold denoising technique followed by a subtraction method is implemented here that can be used for marine species identification. The algorithm is demonstrated for identification of dolphin whistles from noise data corrupted with mechanical noise.

Dolphins are mainly vocal mammalian family, and the vocal communication plays an important role in mediating social interactions (Slater, 1983). They are inhabited in all over the world's oceans and mostly distributed in warm equatorial to subpolar regions, and in coastal as well as offshore waters (Corkeron, Van Parijs, 2001). Dolphins produce mainly two primary types of sounds associated with specific behavioral contexts: non-pulse tonal frequency modulated whistles and rapid repetition of burst-pulsed click sounds (Au, 1993; Jones *et al.*, 2020). Whistles are non-pulse char-

acteristics, longer duration with a narrow frequency band (Boisseau, 2005; Akiyama, Ohta, 2007). These whistle sounds play a crucial role in maintaining contact between dispersed individuals (Caldwell *et al.*, 1990; Janik, Slater, 1998; Sayigh *et al.*, 1999; Janik *et al.*, 2006; 2013; Rachinas-Lopes *et al.*, 2017), group cohesion with male-male alliances, vocal communication between mother and calf pairs, and also promote as a greeting signal when the groups joining each other (Smolker *et al.*, 1993; Janik, 2009; Quick, Janik, 2012; King *et al.*, 2019). The amplitude and duration of the whistles may vary, however the stereotyped frequency contour over time of the whistle seems to comprise the information for recognition (Janik *et al.*, 2006). The discrete parameters to classify whistles include: start frequency, end frequency, minimum frequency, maximum frequency, bandwidth, duration, and number of inflection points (Janik *et al.*, 1994; Esch *et al.*, 2009). However, in some instances the whistle signals are often contaminated by the mechanical noises particular to the rubbing and mooring components such as shackles and cables. It is hard to distinguish the frequency contour of the whistles once the signal of interest is corrupted with unwanted sounds. Therefore, it is important to implement the denoising method which improves the quality of signal.

Denoising is the process of extracting an original signal from the noise (Bey, 2006). In general, the signal is corrupted by noise during its transmission, acquisition, reception, and processing (Isabona, Azi, 2012). Many researchers studied different denoising techniques, such as median filtering, mean filtering, the Fourier transform, and Wiener filtering which are a linear approach and suitable for stationary signals (Chen *et al.*, 2006; Lukac *et al.*, 2007; Zhang, Xiong, 2009). However, the ambient noise in the ocean is non-stationary because of the combination of many oceanic sources including unwanted impulsive mechanical mooring noises which are difficult to extract using a linear approach. Therefore, it is significant to implement the non-linear wavelet threshold denoising approach (Khan *et al.*, 2015). Various non-stationary signal extraction methods have emerged in recent years, and the algorithms are wavelet decomposition, empirical mode decomposition, the Hilbert-Huang transform, and variational mode decomposition (Dragomiretskiy, Zosso, 2013; Ukte *et al.*, 2014; Xiang, Wang, 2015). However, to work these approaches, certain conditions must be met such as decomposition levels, modal number, and termination thresholds. Among these, the wavelet decomposition method is designed for non-stationary signals, which combines both the time and frequency domain. Mallat (1989) describes the theoretical and mathematical approach for understanding wavelet decomposition on signal denoising. The key advantage of wavelet denoising is to split the data into different frequency components and study the noise spikes in each frequency component at different resolution (Chang *et al.*, 2000). The wavelet denoising is an emerging advance technique in signal processing that used in a various applications particularly image processing, data compression, impulsive events characterization, pattern recognition signal extraction and denoising (Yu *et al.*, 2007; Li, Zhou, 2008). This type of technique will be useful for removing impact noise produced in the mooring, when acoustic recorders are incorporated in multisuite ocean moorings. Also data acquisition during rough seas creates more platform noise which is unavoidable. Metal chains and shackles in mooring cause clonking noise in the same frequency range 100 Hz to a few kHz (Marley *et al.*, 2017). So the algorithm described in this paper should be useful for the above mentioned types of noise.

In this paper, the study area is located in the South of Lakshadweep Islands with the close proximity to the Maldives which presents itself a highly varying bathymetric and oceanographic environments so, the area affords a wide variety of cetaceans (Prakash *et al.*, 2015). It shows as different habitat types particularly the coral reefs, seagrass beds, rocky and sandy shores, deep water canyons and trenches that offer a vast marine biodiversity for cetaceans (Pillai, Jasmine, 1989; Mallik, 2017). To identify the cetacean species, visual observation was the commonly used method in which observers can identify the species in a limited sighting conditions (weather, seastate, and daylight). Basing on the visual surveys along with anecdotal evidence, 14 species of cetaceans have been documented in this area (Panicker *et al.*, 2020). Among them, the most commonly sited species are dolphins and studies on acoustic identification is very scarce.

The signal denoising methods would be adequate to eliminate the noise if the unwanted noise levels are lower as compared to the source signals. However, noise removal will be a challenge when the unwanted noise level is higher than the source signals. In this study, the unwanted impulsive shackle noise is higher as compared to that of dolphin whistle signals, and both are non-stationary with transient and vary quickly. However, the frequency contours of the dolphin whistle signals are different from that of unwanted impulsive shackle noise which enables the wavelet denoising technique effective for implementation.

This study is mainly on identification of whistle signals produced by dolphins from the passive acoustic measurements by implementing the wavelet threshold denoising technique along with the subtraction method to remove the unwanted impulsive shackle noise of the mooring system. This study addresses how to tackle the contaminated acoustic data due to impulsive mechanical noise and identify species in a marine based biological ecosystem which would provide the baseline information regarding cetaceans.

2. Materials and methods

2.1. Acoustic measurements

The deep ocean ambient noise measurement system combined with the Ocean Moored buoy network for the Northern Indian Ocean (OMNI) buoy was deployed at the south of Lakshadweep (AD9) in the South Eastern Arabian Sea at the ocean depth of ~2100 m during November 2018 to October 2019 (Fig. 1). The system firmly holds a glass sphere along with hydrophone (bandwidth 10 Hz to 100 kHz), data acquisition system, and power pack. The noise data were acquired at a sampling rate of 32 kHz for a duration of 12 min for every half an hour. The hydrophone-sensed acoustic pressure fluctuation caused by different sources of noise, which translates into electrical signals and converts to units of micropascal [μPa] by applying the receiving sensitivity (−165 dB re 1 V/μPa) of the hydrophone.

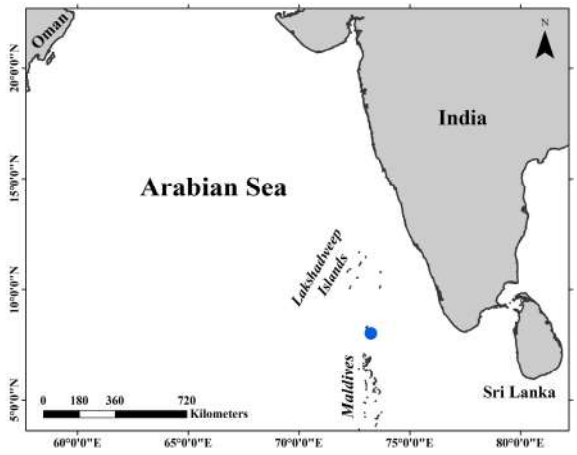


Fig. 1. Deployment location of OMNI buoy in the Arabian Sea (AD9) is indicated in blue filled dot.

The oscillogram, spectrograms, and Welch's averaged power spectral density (dB re 1 μPa^2/Hz) were analyzed using MATLAB (MATLAB R2021a). Multiple spectra were obtained by segmenting the data into smaller portions using a Hamming window and 2048 point FFT with 50% overlap. The resulting spectra were then averaged to obtain the final spectrum. The frequency resolution of each power spectrum is 15.6 Hz, which is determined by the sampling frequency and the number of points in the FFTs in each power spectrum.

2.2. Wavelet threshold denoising

The wavelet threshold denoising technique contains three steps: signal decomposition, thresholding, and signal reconstruction (Donoho, Johnstone, 1994). In this method, the signal is decomposed into approximation and detail coefficients at each level. The approximations are high-scale and low frequency components, and the details are low-scale and high frequency components of the signal (Tikkanen, 1999). The threshold value is very important parameter in the wavelet threshold denoising technique. There are four threshold selection methods particularly the universal threshold, rigorous Stein's Unbiased Risk Estimate (SURE), heuristic SURE, and minimax (Donoho, Johnstone, 1994). In this study, the rigorous SURE threshold estimation is adopted. The threshold (T) is defined by:

$$T = \sqrt{2\log_e\left(N\log_2\left(N\right)\right)}, \tag{1}$$

where N is the number of samples in the input signal. Once the value of threshold is estimated using this method, a hard or soft thresholding function is needed to filter the wavelet coefficients which contain unwanted noise (Donoho, Johnstone, 1995). For the hard threshold, the absolute values of wavelet coefficients below the threshold level are set to zero, and the values above the threshold are kept unchanged. In soft threshold, the wavelet coefficients whose values are lower than the zero threshold, and the coefficients above the threshold level are also modified (Donoho, Johnstone, 1995). In this study, the soft threshold method is considered because the wavelet coefficients become more stable and smoothening as compared to the hard threshold. Finally, the in situ signal is denoised and reconstructed using modified level coefficients.

In this study, the in situ time series data of ambient noise is the combination of unwanted impulsive shackle noise along with dolphins whistle signals. The noise level of the impulsive shackle dominates the whistles of dolphin which is difficult to identify from frequency contours. Therefore, the wavelet threshold denoising technique is implemented using the MATLAB function wden:

$$\text{signal}_{\text{denoised}} = \text{wden (input data, rigrsure, s, mln, level, wname)}, \tag{2}$$

where $\text{signal}_{\text{denoised}}$ is the denoised signal, input data is the original in situ data, rigrsure specifies the adaptive threshold selection using the principle of SURE. The term s denotes soft thresholding, mln indicates multi-threshold re-scaling at level coefficients, the level determines the decomposition of the signal using the syntax level = wmaxlev (N, wname), where N is the number of samples in the input signal. The wname is a wavelet family and the function wden performs wavelet denoise of the input signal. In this study, we chose the wavelet function Daubechies [db5] (Daubechies, 1992; Rowe, Abbott, 1995), and estimated the decomposition level at 20 using the above syntax.

After implementing the wavelet threshold denoising technique, the unwanted impulsive shackle noise

exists and suppresses the resulted wanted signal because the value of noise is higher as compared to the signal. Hence a subtraction method is implemented followed by the wavelet threshold denoising technique. A flowchart on the threshold denoising algorithm described in this work is shown in Fig. 2.

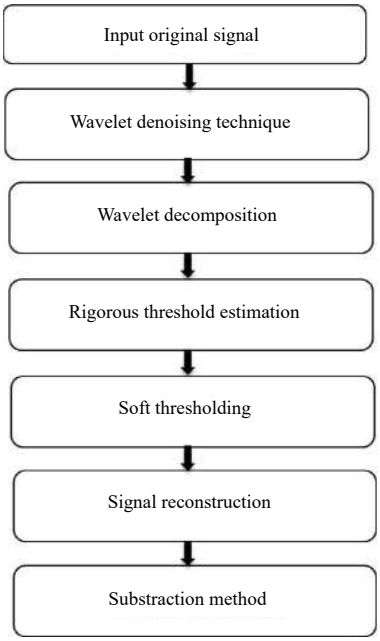


Fig. 2. Block diagram of wavelet based threshold denoising algorithm.

In this method, a residual signal is estimated by subtracting the denoised signal from the original in situ data (Math Works, n.d). The unwanted impulsive shackle noise is isolated by subtracting the residual signal from the denoised signal, and the wanted signal (dolphin whistle signals) is obtained by subtracting the unwanted impulsive shackle noise from the original in situ data.

3. Results

In situ time series measurements of deep water ambient noise in the spring period were considered for analysis, i.e., April–June, 2019. The in situ time series data have been recorded as audio files during the measurement, and analyzed using time-frequency spectrogram. During the study period, 12 data sets of recorded audio files resembled dolphin whistles, which are mostly contaminated by unwanted impulsive shackle noise. Among these, Fig. 3a represents the oscillogram of an in situ data recorded on 31/05/2019 at 11:58 hr. The spectrogram in Fig. 3b shows the mixing of unwanted background noise, impulsive shackle noise as well as the dolphin whistle sounds. The sounds produced by impulsive shackle noise in the frequency range of 0–16 kHz contaminates the dolphin whistle signals.

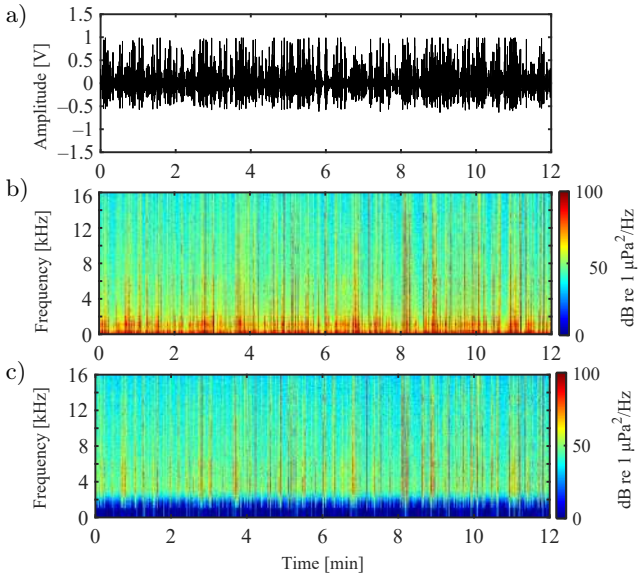


Fig. 3. a) Oscillogram; b) spectrogram of the original signal; c) spectrogram of the original signal using butterworth high pass filter up to 3 kHz. The vertical striplines in spectrogram are shown as impulsive shackle noise.

A butterworth high pass filter with 3 kHz cut-off is employed to the in situ data in order to suppress the unwanted low frequency background noise which is falling below the dolphin whistles (Fig. 3c). The spectrograms in Fig. 4 show the whistles of dolphin which are subsequently extracted from the original spectrogram of Fig. 3c. The results shown here are from in situ data recorded on 31/05/2019 at 11:58 hr with duration of 12 minutes. The whistles produced by dolphins are narrow band with the spectral peak in the frequency range of 4–16 kHz (Fig. 4). However, it is difficult to identify the frequency contour of different whistle types because the signals are indistinguishable due to the impact of impulsive shackle noise. It is analysed that the averaged noise level is about ~82 dB due to impulsive shackle noise whereas it is about ~62 dB produced by dolphins whistle signal (Fig. 4a). When compared the noise levels of impulsive shackle noise to that of the dolphin signal, it is observed that the unwanted impulsive shackle noise is higher by ~20 dB as compared to that of dolphin signals. It means that the noise level of dolphins whistle signals are significantly lower and buried under the impulsive shackle noise, which is difficult to retrieve.

Therefore, to retrieve the dolphin whistle signals, the wavelet threshold denoising approach has been implemented along with the commonly used subtraction method. The oscillogram and spectrogram of the proposed wavelet denoising approach is shown in Figs. 5a and 5b.

The denoised version of the spectrogram (Fig. 5b) shows the dominance of impulsive shackle noise, since the noise levels of impulsive shackle are higher than the noise level of dolphin whistle signals. The residual sig-

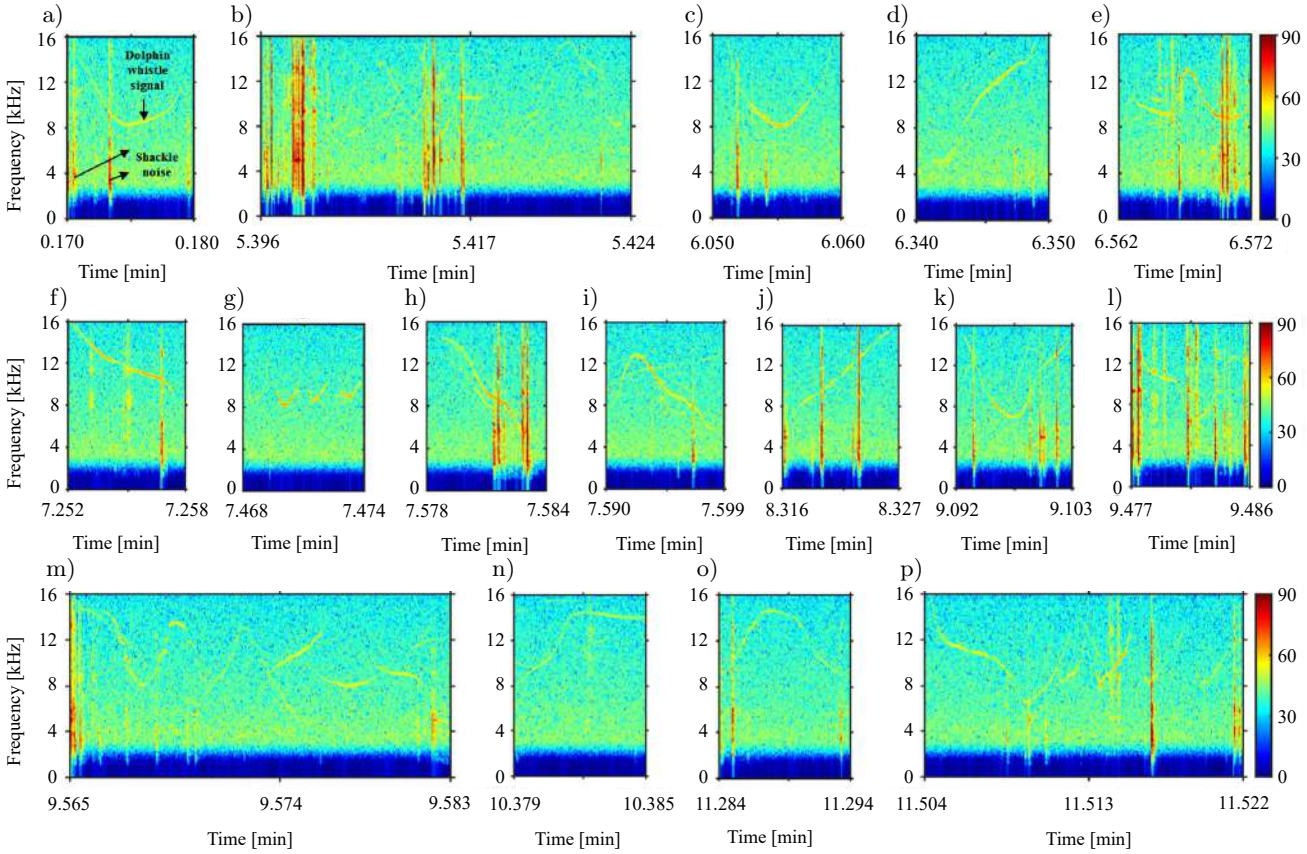


Fig. 4. Spectrograms of dolphin whistle signals (a–p) along with unwanted impulsive shackle noise which are extracted from Fig. 3c. The time axis represents Time in minutes from the start of the recording till the 12th minute reading from top panel.

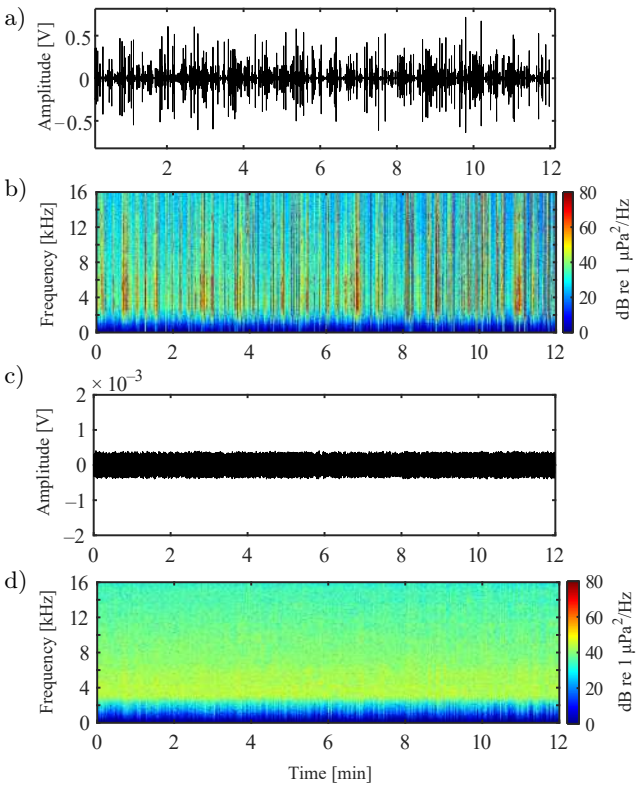


Fig. 5. a) Oscillogram and b) spectrogram of denoised signal; c) oscillogram and d) spectrogram of residual signal.

nal is obtained by taking the difference between the original signal and the denoised signal (Figs. 5c and 5d). In residual output, it contains the combination of whistle signals along with white Gaussian noise other than the impulsive shackle noise. However, denoised signal contains higher values of impulsive shackle noise and lower values of the whistle signals. Hence, only shackle noise can be estimated by subtracting the residual signal from the denoised signal, which is shown in Fig. 6. Finally, the wanted signal of dolphin whistles can be extracted by subtracting only shackle noise from the original signal (Fig. 7).

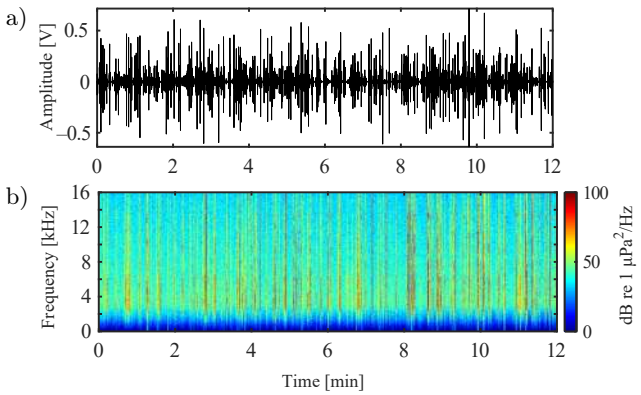


Fig. 6. a) Oscillogram and b) spectrogram of shackle noise.

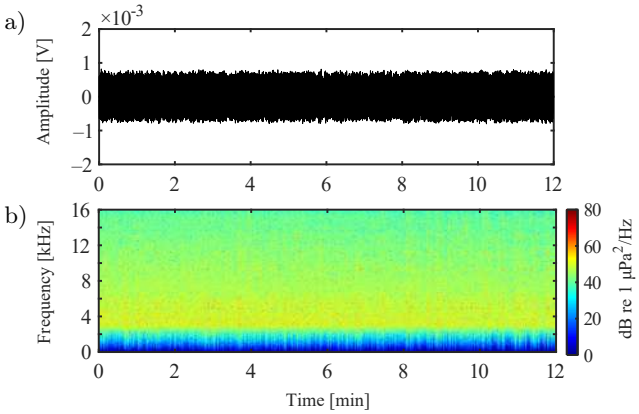


Fig. 7. a) Oscillogram and b) spectrogram of wanted dolphin whistle signals.

The segmented spectrograms of Fig. 8 are extracted from the Fig. 7b, which illustrates the removal of impulsive shackle noise and significant improvement of dolphin whistle signals. It is easy to detect the frequency contour and whistle types after the implementation of wavelet threshold denoising technique followed by the subtraction method. The frequency of dolphin whistles are ranged from ~4 to ~16 kHz with the duration ranges from ~0.4 to ~1.08 s. Based on the frequency contours, whistles are classified as different types such as concave, convex, upsweep, downsweep, sine, and multi-looped. Table 1 gives a detailed description of acoustic parameters for each of the dolphin whistle types. Majority of the whistle types are concave (Figs. 8a, 8c, 8k, and 8l) and convex (Figs. 8e, 8i,

Table 1. Acoustic parameters of the dolphin whistle types in deep water of Arabian Sea from in situ data recorded on 31/05/2019 at 11:58 hr. These metrics are calculated from time-frequency spectrograms of a single data considering 2048 FFT points.

Whistle types	Start frequency [kHz]	End frequency [kHz]	Maximum frequency [kHz]	Minimum frequency [kHz]	Duration [s]
Concave	15.90	15.25	15.90	8.21	0.6
Convex	8.25	14.25	14.62	8.25	0.6
Upsweep	4.98	15.39	15.39	4.93	0.6
Downsweep	15.60	8.92	15.60	8.92	0.4
Sine	15.25	4.15	15.25	4.15	0.9
Multi-looped	15.85	11.35	15.85	6.25	1.08

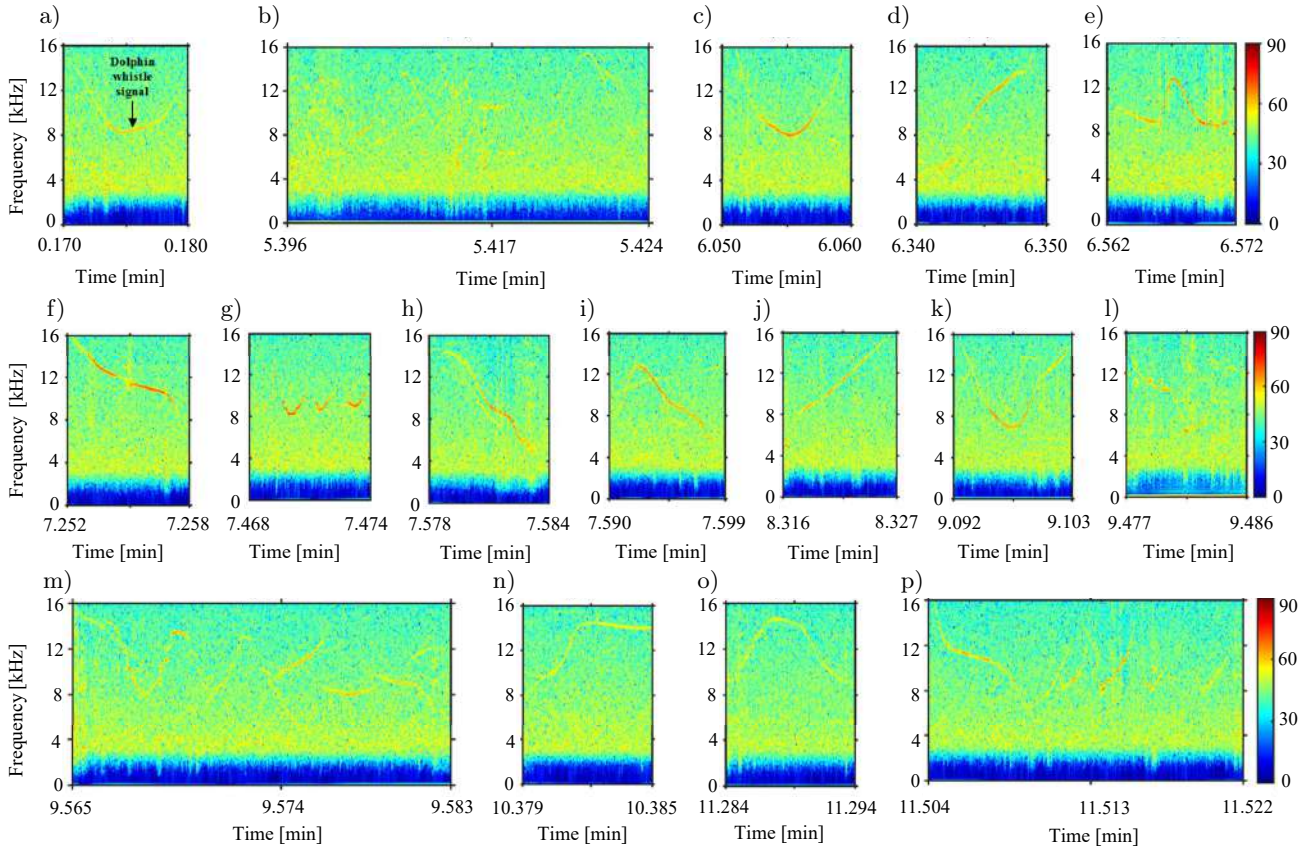


Fig. 8. Spectrograms of dolphin whistle signals (a–p) which are extracted from the Fig. 7b.

and 8o) followed by upsweep (Figs. 8d and 8j), downsweep (Figs. 8f and 8h), multi-looped (Figs. 8g and 8p), and sine (Fig. 8m).

4. Discussion

In this study, it is observed that the in situ time series data of acoustic noise have been contaminated by unwanted impulsive shackle noise, which conceals the dolphin whistle signals. The results show that the unwanted impulsive shackle noise is non-stationary, with the noise levels higher by ~20 dB as compared to that of dolphin whistle signals. The wavelet threshold denoising approach followed by the subtraction method is implemented successfully, and the impulsive shackle noise is removed which are overlapped on dolphin whistle contours. The study suggests that the optimum conditions of denoising are mainly considered by Daubechies [db5] along with 20 multilevel wavelet decomposition and the rigrsure soft threshold method. The satisfactory results of wavelet denoising for dolphin whistle types are obtained and shown in the spectrogram of Fig. 8.

Seramani *et al.* (2006) implemented the wavelet denoising along with the independent component analysis to separate dolphin whistles in the underwater noise environment. Lopez-Otero *et al.* (2018) used the discrete wavelet transform to model dolphin whistle contours for species classification. An earlier study has also demonstrated the extraction of time-frequency dolphin contours based on the automated denoising method (Mallawaarachchi *et al.*, 2008). However, there is no research paper on the wavelet denoising of dolphin whistle contours in the presence of impulsive shackle noise. In the present study, the use of the wavelet denoising threshold approach to identify the dolphin whistle signal has proved to be fruitful to remove impulsive shackle noise present in the in situ time series recording. This is because the impulsive shackle noise has the different temporal and frequency structure as compared to that of dolphin whistle contours.

Many previous studies on dolphin whistle types have been reported worldwide (Janik, Slater, 1998; Wang *et al.*, 1995; Acevedo-Gutiérrez, Stienessen, 2004; Azevedo *et al.*, 2007; Kriesell *et al.*, 2004; Heiler *et al.*, 2016). Some researchers have studied the acoustic detection along with visual observation of cetacean species in the offshore waters of the Maldives and Sri Lanka (Clark *et al.*, 2012; de Vos *et al.*, 2012). However, no detailed identification of dolphin whistle types have been studied in shallow and deep waters of the Arabian Sea. The recent study has described the acoustic identification of dolphin whistle types in deep water of Lakshadweep in the Arabian Sea, and analyzed their acoustic parameters which confirms the six major whistle types in the frequency range approximately from ~4 to ~16 kHz with the whistle duration ranges from ~0.4 to ~1.08 s (Fig. 8, Table 1).

As from the recent study, it has been revealed that the wavelet threshold denoising approach has been taken under consideration for removing the non-stationary impulsive shackle noise, and effectively identify the dolphin whistle contours. The ability to detect and characterize the different whistle contours that provides significant information on dolphin communication and behavioral signals. This method of a wavelet threshold denoising approach identifies the dolphin whistle sounds, and could be used for future studies on other cetacean whistle signals which would be effected by anthropogenic as well as unwanted mechanical noise sources.

5. Conclusion

This study details a technique primarily based on the wavelet threshold approach followed by subtraction for denoising the dolphin whistle contours. The time series recordings of noise data are made in deep waters, where the dolphin whistle signals are contaminated by impulsive shackle noise. The results show that the optimal conditions for denoising are mainly based on Daubechies [db5] along with 20 multilevel wavelet decomposition and the rigrsure soft threshold method. Finally, the contaminated impulsive shackle noise is removed from the dolphin signals by using a wavelet approach. Based on the frequency contours, whistle types are identified as concave, convex, upsweep, downsweep, sine, and multi-looped. Hence, it is proven that, this method is found suitable to extract other species vocalization particularly the non-impulsive signals from passive acoustic recordings.

Acknowledgments

The authors express their sincere thanks to the Director of the National Institute of Ocean Technology, the Ministry of Earth Sciences, for encouragement to the team to do this work. The authors gratefully acknowledge the funding support from the Ministry of Earth Sciences, Government of India. The authors are very thankful to the team of the Ocean Observation System (OOS), NIOT for their support in this work. The authors sincerely thank A. Malarkodi, G. Raguraman, A. Thirunavukkarasu, P.S.S.R. Sridhar, V. Preyenga, and C. Dhanaraj of the Ocean Acoustics Group for their help in calibration of the hydrophones.

References

1. Acevedo-Gutiérrez A., Stienessen S.C. (2004), Bottlenose dolphins (Tursiops truncatus) increase number of whistles when feeding, *Aquatic Mammals*, **30**: 357–362, doi: 10.1578/AM.30.3.2004.357.

2. Akiyama J., Ohta M. (2007), Increased number of whistles of bottlenose dolphins, Tursiops truncatus, arising from interaction with people, *Journal of Veterinary Medical Science*, **69**(2): 165–170, doi: 10.1292/jvms.69.165.

3. Au W.W.L. (1993), Characteristics of dolphin sonar signals, [in:] *The Sonar of Dolphins*, pp. 115–139, Springer, New York, doi: 10.1007/978-1-4612-4356-4_7.

4. Azevedo F.A., Oliveira A.M., Dalla Rosa L., Lailson-Brito J. (2007), Characteristics of whistles from resident bottlenose dolphins (Tursiops truncatus) in southern Brazil, *The Journal of Acoustical Society of America*, **121**(5): 2978–2983, doi: 10.1121/1.2713726.

5. Beslin W.A., Whitehead H., Gero S. (2018), Automatic acoustic estimation of sperm whale size distributions achieved through machine recognition of on-axis clicks, *The Journal of Acoustical Society of America*, **144**(6): 3485–3495, doi: 10.1121/1.5082291.

6. Bey N.Y. (2006), Extraction of signals buried in noise. Part I: Fundamentals, *Signal Processing*, **86**(9): 2464–2478, doi: 10.1016/j.sigpro.2005.11.014.

7. Boisseau O. (2005), Quantifying the acoustic repertoire of a population: The vocalizations of free-ranging bottlenose dolphins in Fiordland, New Zealand, *The Journal of Acoustical Society of America*, **117**(4): 2318–2329, doi: 10.1121/1.1861692.

8. Caldwell M.C., Caldwell D.K., Tyack P.L. (1990), Review of the signature whistle hypothesis for the Atlantic bottlenose dolphin, [in:] Leatherwood S., Reeves R.R. [Eds.], *The Bottlenose Dolphin*, pp. 199–234, Amsterdam, Elsevier, doi: 10.1016/B978-0-12-440280-5.50014-7.

9. Chang S.G., Yu B., Vetterli M. (2000), Adaptive wavelet thresholding for image denoising and compression, *IEEE Transactions on Image Processing*, **9**(9): 1532–1546, doi: 10.1109/83.862633.

10. Chen J., Benesty J., Huang Y., Doclo S. (2006), New insights into the noise reduction Wiener filter, *IEEE Transactions on Audio, Speech, and Language Processing*, **14**(4): 1218–1234, doi: 10.1109/TSA.2005.860851.

11. Clark R.A. *et al.* (2012), Cetacean sightings and acoustic detections in the offshore waters of the Maldives during the northeast monsoon seasons of 2003 and 2004, *Journal of Cetacean Research and Management*, **12**(2): 227–234.

12. Corkeron P.J., Van Parijs S.M. (2001), Marine mammal migrations and movement patterns, [in:] Steele J.H., Turekian K.K., Thorpe S.A. [Eds.], *Encyclopedia of Ocean Sciences*, Academic Press, pp. 596–604, Oxford, doi: 10.1016/B978-012374473-9.00442-2.

13. Daubechies I. (1992), *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, USA, doi: 10.1137/1.9781611970104.

14. Donoho L.D., Johnstone I.M. (1995), Adatpting to unknow smoothness via wavelet shrinkage, *Journal of the American Statistical Association*, **90**(432): 1200–1224, doi: 10.1515/9781400827268.833.

15. Donoho L.D., Johnstone M.J. (1994), Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**: 425–455, doi: 10.1093/biomet/81.3.425.

16. Dragomiretskiy K., Zosso D. (2013), Variational mode decomposition, *IEEE Transactions on Signal Processing*, **62**(3): 531–544, doi: 10.1109/TSP.2013.2288675.

17. Esch H.C., Sayigh L.S., Blum J.E., Wells R.S. (2009), Whistles as potential indicators of stress in bottlenose dolphins (Tursiops truncatus), *Journal of Mammalogy*, **90**(3): 638–650, doi: 10.1644/08-MAMM-A-069R.1.

18. Esch H.C., Sayigh L.S., Wells R.S. (2009), Quantifying parameters of bottlenose dolphin signature whistles, *Marine Mammal Science*, **45**(4): 976–986, doi: 10.1111/j.1748-7692.2009.00289.x.

19. Heiler J., Elwen S.H., Kriesell H.J., Gridley T. (2016), Changes in bottlenose dolphin whistle parameters related to vessel presence, surface behaviour and group composition, *Animal Behavior*, **117**: 167–177, doi: 10.1016/j.anbehav.2016.04.014.

20. Hunynh Q.Q., Cooper L.N., Intrator N., Shouval H. (1998), Classification of underwater mammals using feature extraction based on time-frequency analysis and BCM theory, *IEEE Transactions on Signal Processing*, **46**(5): 1202–1207, doi: 10.1109/78.668783.

21. Isabona J., Azi S. (2012), Optimised Walficsh-Bertoni Model for pathloss prediction in urban propagation environment, *International Journal of Engineering and Innovative Technology*, **2**(5): 14–20.

22. Janik V.M. (2009), Acoustic communication in delphinids [in:] Naguib M., Janik V.M. [Eds.], *Advances in the Study of Behavior*, **40**: 123–157, doi: 10.1016/S0065-3454(09)40004-4.

23. Janik V.M., King S.L., Sayigh L.S., Wells R.S. (2013), Identifying signature whistles from recordings of groups of unrestrained bottlenose dolphins (Tursiops truncatus), *Marine Mammal Science*, **29**(1): 109–122, doi: 10.1111/j.1748-7692.2011.00549.x.

24. Janik V.M., Sayigh L.S., Wells R.S. (2006), Signature whistle shape conveys identity information to bottlenose dolphins, *Proceedings of the National Academy of Sciences of the United States of America*, **103**(21): 8293–8297, doi: 10.1073/pnas.0509918103.

25. Janik V.M., Slater P. (1998), Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls, *Animal Behaviour*, **56**(4): 829–838, doi: 10.1006/anbe.1998.0881.

26. Janik V.M., Todt D., Dehnhardt G. (1994), Signature whistle variations in a bottlenosed dolphin, Tur-

siops truncatus, *Behavioral Ecology and Sociobiology*, **35**(4): 243–248, doi: 10.1007/BF00170704.

27. Jones B., Zapetis M., Samuelson M.M., Ridgway S. (2020), Sounds produced by bottlenose dolphins (Tursiops): a review of the defining characteristics and acoustic criteria of the dolphin vocal repertoire, *Bioacoustics*, **29**(4): 399–440, doi: 10.1080/09524622.2019.1613265.

28. Khan M.M., Ashique R.H., Liya B.N., Sajjad M.M., Rahman M.A., Amin M.H. (2015), New wavelet thresholding algorithm in dropping ambient noise from underwater acoustic signals, *Journal of Electromagnetic Analysis and Applications*, **7**(3): 53–60, doi: 10.4236/jemaa.2015.73006.

29. King S.L., Allen S.J., Krützen M., Connor R.C. (2019), Vocal behaviour of allied male dolphins during cooperative mate guarding, *Animal Cognition*, **22**(6): 991–1000, doi: 10.1007/s10071-019-01290-1.

30. Kriesell H.J., Elwen S.H., Nastasi A., Gridley T. (2014), Identification and characteristics of signature whistles in wild bottlenose dolphins (Tursiops truncatus) from Namibia, *PLOS ONE*, **9**(9): e106317, doi: 10.1371/journal.pone.0106317.

31. Learned R.E., Willsky A.S. (1995), A wavelet packet approach to transient signal classification, *Applied and Computational Harmonic Analysis*, **2**(3): 265–278, doi: 10.1006/acha.1995.1019.

32. Li N., Zhou M. (2008), Audio denoising algorithm based on adaptive wavelet soft-threshold of gain factor and teager energy operator, [in:] *2008 International Conference on Computer Science and Software Engineering*, pp. 787–790, doi: 10.1109/CSSE.2008.1523.

33. Lopez-Otero P., Docio-Fernandez L., Cardenal-López A. (2018), Using discrete wavelet transform to model whistle contours for dolphin species classification, *Proceedings*, **2**(18): 1183, doi: 10.3390/proceedings2181183.

34. Lukac R., Smolka B., Plataniotis K.N. (2007), Sharpening vector median filters, *Signal Processing*, **87**(9): 2085–2099, doi: 10.1016/j.sigpro.2007.02.009.

35. Mallat S.G. (1989), A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**(7): 674–693, doi: 10.1109/34.192463.

36. Mallawaarachchi A., Ong S.H., Chitre M., Taylor E. (2008), Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles, *The Journal of Acoustical Society of America*, **124**(2): 1159–1170, doi: 10.1121/1.2945711.

37. Mallik T.K. (2017), Coral atolls of Lakshadweep, Arabian Sea, Indian Ocean, *MOJ Ecology Environmental Sciences*, **2**(2): 68–83, doi: 10.15406/mojes.2017.02.00021.

38. Marley S.A., Salgado-Kent C.P., Erbe C., Parnum I. (2017), Effects of vessel traffic and underwater noise on the movement, behaviour and vocalisations of bottlenose dolphins in an urbanised estuary, *Scientific Reports*, **7**: 13437, doi: 10.1038/s41598-017-13252-z.

39. Math Works (n.d.), *Wavelet Interval-Dependent Denoising*, https://in.mathworks.com/help/wavelet/ug/wavelet-interval-dependent-denoising.html.

40. Panicker D., Sutaria D., Kumar A., Stafford K.M. (2020), Cetacean distribution and diversity in Lakshadweep waters, India, using a platform of opportunity: October 2015 to April 2016, *Aquatic Mammals*, **46**(1): 80–92, doi: 10.1578/AM.46.1.2020.80.

41. Pillai C.S.G., Jasmine S. (1989), The coral fauna of Lakshadweep, [in:] James P.S.B.R., Suseelan C. [Eds.], *Marine Living Resources of the Union Territory of Lakshadweep: An Indicative Survey with Suggestions or Development*, Central Marine Fisheries Research Institute, **43**: 179–195.

42. Powell K.J., Sapatinas T., Bailey T.C., Krzanowski W.J. (1995), Application of wavelets to the pre-processing of underwater sounds, *Statistics and Computing*, **5**: 265–273.

43. Prakash T.N., Nair L.S., Hameed T.S.S. (2015), *Geomorphology and Physical Oceanography of the Lakshadweep Coral Islands in the Indian Ocean*, Springer Cham, doi: 10.1007/978-3-319-12367-7.

44. Quick N.J., Janik V.M. (2012), Bottlenose dolphins exchange signature whistles when meeting at sea, *Proceedings of the Royal Society B: Biological Sciences*, **279**: 2539–2545, doi: 10.1098/rspb.2011.2537.

45. Rachinas-Lopes P., Luís A.R., Borges A.S., Neto M., dos Santos M.E. (2017), Whistle stability and variation in captive bottlenose dolphins (Tursiops truncatus) recorded in isolation and social contexts, *Aquatic Mammal*, **43**(1): 1–13, doi: 10.1578/AM.43.1.2017.1.

46. Rowe A.C., Abbott P.C. (1995), Daubechies wavelets and mathematica, *Computers in Physics*, **9**(6): 635–648, doi: 10.1063/1.168556.

47. Sayigh L.S., Tyack P.L., Wells R.S., Solow A.R., Scott M.D., Irvine A.B. (1999), Individual recognition in wild bottlenose dolphins: A field test using playback experiments, *Animal Behaviour*, **57**(1): 41–50, doi: 10.1006/anbe.1998.0961.

48. Seramani S., Taylor E.A., Seekings P.J., Yeo K.P. (2006), Wavelet de-noising with independent component analysis for segmentation of dolphin whistles in a noisy underwater environment, [in:] *OCEANS 2006 – Asia Pacific*, pp. 1–7, doi: 10.1109/OCEANSAP.2006.4393920.

49. Slater P.J.B. (1983), The study of communication, [in:] Halliday T.R., Slater P.J.B [Eds.], *Animal Behaviour, Communication*, 2nd ed., pp. 9–42, Blackwell, Oxford.

50. Smolker R.A., Mann J., Smuts B.B. (1993), Use of signature whistles during separations and reunions by wild bottlenose dolphin mothers and infants, *Behavioral Ecology and Sociobiology*, **33**(6): 393–402, doi: 10.1007/BF00170254.

51. Tikkanen P.E. (1999), Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal, *Biological cybernetics*, **80**(4): 259–267, doi: 10.1007/s0042 20050523.

52. Ukte A., Kizilkaya A., Elbi M.D. (2014), Two empirical methods for improving the performance of statistical multirate high-resolution signal reconstruction, *Digital Signal Processing*, **26**: 36–49, doi: 10.1016/ j.dsp.2013.11.014.

53. de Vos A. *et al.* (2012), Cetacean sightings and acoustic detections in the offshore waters of Sri Lanka: March–June 2003, *Journal of Cetacean Research and Management*, **12**(2): 185–193.

54. Wang D., Würsig B., Evans W.E. (1995), Whistles of bottlenose dolphins: Comparisons among populations, *Aquatic Mammals*, **21**: 65–77.

55. Xiang L.W., Wang W.B. (2015), Harmonic signal extraction from noisy chaotic interference based on synchrosqueezed wavelet transform, *Chinese Physics B*, **24**(8): 080203, doi: 10.1088/1674-1056/24/8/080203.

56. Yu G., Bacry E., Mallat S. (2007), Audio signal denoising with complex wavelets and adaptive block attenuation, [in:] *IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP'07*, **3**: 869–872, doi: 10.1109/ICASSP.2007.366818.

57. Zhang X., Xiong Y. (2009), Impulse noise removal using directional difference based noise detector and adaptive weighted mean filter, *IEEE Signal Processing Letters*, **16**(4): 295–298, doi: 10.1109/LSP.2009.20 14293.

Archives of Acoustics Vol. 48, No. 1, pp. 49–61 (2023), doi: 10.24425/aoa.2022.142909

Research Paper

Using SVM Classifier and Micro-Doppler Signature for Automatic Recognition of Sonar Targets

Abbas SAFFARI[(1)], Seyed Hamid ZAHIRI[(1)*], Navid KHOZEIN GHANAD[(2)]

[(1)] *University of Birjand*
Birjand, Iran

[(2)] *Sajjad University of Mashhad*
Mashhad, Iran

*Corresponding Author e-mail: hzahiri@birjand.ac.ir

(*received September 20, 2021; accepted August 29, 2022*)

In this paper, we propose using a propeller modulation on the transmitted signal (called sonar micro-Doppler) and different support vector machine (SVM) kernels for automatic recognition of moving sonar targets. In general, the main challenge for researchers and craftsmen working in the field of sonar target recognition is the lack of access to a valid and comprehensive database. Therefore, using a comprehensive mathematical model to simulate the signal received from the target can respond to this challenge. The mathematical model used in this paper simulates the return signal of moving sonar targets well. The resulting signals have unique properties and are known as frequency signatures. However, to reduce the complexity of the model, the 128-point fast Fourier transform (FFT) is used. The selected SVM classification is the most popular machine learning algorithm with three main kernel functions: RBF kernel, linear kernel, and polynomial kernel tested. The accuracy of correctly recognizing targets for different signal-to-noise ratios (SNR) and different viewing angles was assessed. Accuracy detection of targets for different SNRs (−20, −15, −10, −5, 0, 5, 10, 15, 20) and different viewing angles (10, 20, 30, 40, 50, 60, 70, 80) is evaluated. For a more fair comparison, multilayer perceptron neural network with two back-propagation (MLP-BP) training methods and gray wolf optimization (MLP-GWO) algorithm were used. But unfortunately, considering the number of classes, its performance was not satisfactory. The results showed that the RBF kernel is more capable for high SNRs (SNR = 20, viewing angle = 10) with an accuracy of 98.528%.

Keywords: sonar micro-Doppler; automatic recognition; SVM; RBF kernel; linear kernel; polynomial kernel.

1. Introduction

Due to the increasing use of automatic target recognition systems in various military and commercial industries, the issue of classification and automatic target recognition has become one of the favorite interests of craftsmen and activists in this field (Ehrman, Lanterman, 2020; Smith *et al.*, 2007). The most important advantage of using these systems is eliminating the human role from the target identification and detection processes (Bhanu, 1986). One of the most important reasons for replacing these systems with systems controlled by human operators is a slow human reaction, low reliability, and high dependence on mental conditions (Zhou *et al.*, 2022). The main element of many defense and military missions in the sea is the automatic detection and identification of sonar targets (Liu *et al.*, 2019). The complex physical properties of sonar targets and the heterogeneous conditions of sound propagation at sea have led to many features being extracted to identify and detect sonar targets (Khishe, Safari, 2019). Obviously, with increasing the dimensions of the feature vector, the dimensions of the data also increase (Saffari *et al.*, 2022a).

The first challenge that sonar researchers face is obtaining reliable datasets (Dong *et al.*, 2021; Glover, Laguna, 2008). Creating a sonar dataset generates high costs (Chen *et al.*, 2022; Waite, 2002). According to the research, the sonar datasets are usually acquired by performing a sonar collection scenario

in a real environment (Kaveh *et al.*, 2019; Qiao *et al.*, 2021; Saffari *et al.*, 2022b; Wu *et al.*, 2021). Another solution for collecting sonar datasets is to use a cavitation tunnel (Khishe, Mosavi, 2019; Khishe, Mohammadi, 2019; Khishe, Safari, 2019). The most important disadvantage of using these methods is the presence of noise in the environment. In other words, the dataset is reliable and can be used for practical systems that perform well even in high-noise environments. One of the main motivations of this paper is to provide a practical model for simulating the emitted signals from sonar targets with the ability to adjust the SNR value and the target angle of view of the sonar receiver.

The next step after preparing the data is extracting the signal's attribute. All submarines have propellers for propulsion. When acoustic signals are propagated toward the target (floating propeller), the propeller's rotation reflects the transmitted signal (Clemente *et al.*, 2013). It can be said that this signal is unique; in other words, it is known as the corresponding floating frequency signature (Chen, Li, 2022; Tahmoush, 2015). The phenomenon of modulation of rotating parts (propeller) is known as sonar micro-Doppler. This phenomenon, called radar micro-Doppler, is widely used to classify aerial targets such as helicopters (Anderson, 2004; Mamgain *et al.*, 2018) and ground targets such as tanks (Foued *et al.*, 2017; Molchanov *et al.*, 2013). Unfortunately, we have not found any researches on the classification of sonar targets and in particular naval vessels using sonar micro-Doppler. Therefore, one of the main goals of this research is to investigate the effect of using this method in the automatic recognition of sonar targets for practical application.

The next step after feature extraction is the classifier design. There are two general approaches to classifying data (Koturwar *et al.*, 2015). The first method is to use definitive computational methods, which have very high reliability and definitely achieve the best results, but the disadvantage of these methods is clearly seen when the size of the data increases (such as sonar dataset). Then the spatial and temporal complexity increases (Lakhwani, 2020). Therefore, this method does not work well for sonar data (Cai *et al.*, 2021). The second approach is to use artificial intelligence (Jin *et al.*, 2020). The main subset of artificial intelligence is machine learning (Liu *et al.*, 2019; Sclavounos, Ma, 2018). One of the most popular supervised learning algorithms is support vector machine (SVM) (Uddin *et al.*, 2019). SVM has a relatively simple training phase and, unlike neural networks, does not fall into local maxima (Xu *et al.*, 2019). In addition, it works relatively well for high-dimensional data and, despite providing the optimal answer, has less time and space complexity than specific methods (Berthold *et al.*, 2018; Gaye *et al.*, 2021). According to the explanations given in this paper, sonar micro-Doppler and SVM will be used for the automatic recognition of sonar targets. In order to have a fair and comprehensive comparison, considering that in references (Saffari *et al.*, 2022c; 2022d) artificial neural network has been used to classify sonar targets, in this article (Kazimierski, Zaniewicz, 2021), two types of MLP-NNs with two types of BP and GWO training are used (Wawrzyniak, Stateczny, 2017; Zhang *et al.*, 2020).

The paper is organized in such a way that the second part explains SVM. In the third section, the micro-Doppler phenomenon is introduced. The fourth part introduces the automatic sonar detection system using sonar micro-Doppler. In the fifth part, the simulation results are presented. The sixth part is the conclusion.

2. Support vector machines

SVMs are supervised learning algorithms and a subset of heuristic algorithms (Kavzoglu, Colkesen, 2009). In SVM, hyperplanes usually separate the two classes and the training data set of a hyperplane are determined. The generalizability can then be verified using the test dataset. SVMs have been able to perform powerfully in many applications (Uddin *et al.*, 2019; Xu *et al.*, 2019).

To classify a data set with dimensional D, a *D-1* hyperplane is required. Figure 1 shows the different hyperplanes separating two different classes. However, only one optimal hyperplane divides data into two classes with the longest distance (Fig. 2). All points that limit the width of the margin are called support vectors. SVMs in binary class mode seek to find a hyperplane so that the distance between the members of each class to the optimal hyperplane is maximum. For example, it is assumed that a set of training data with a D sample is represented by $\{x_i, y_i\}$ and $(I = 1, 2, ..., D)$, where $x \in R^D$ is a D-dimensional space and $y \in \{-1, +1\}$ is a class label (Kavzoglu, Colkesen, 2009). The optimal hyperplane performance is to maximize margins. This hyperplane is defined as $w \cdot x_i + b = 0$, where w determines the orientation of the hyperplane in space, x is the points on the hyperplane, b is the bias of the distance of hyperplane

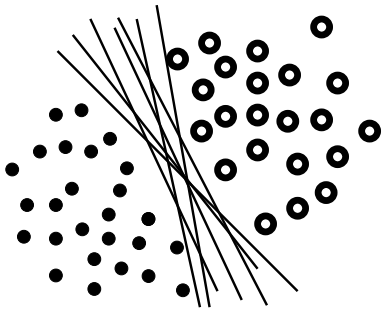


Fig. 1. Linear separation of data by different hyperplanes.

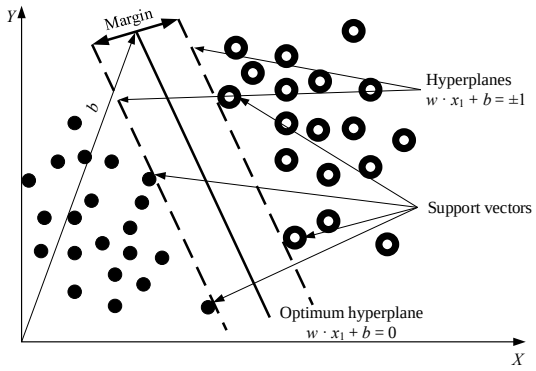


Fig. 2. Optimum hyperplane and support vectors for linearly separable data.

from the origin (Fig. 2). Equations (1) and (2) are the equations of a separating hyperplane for the separable state of two classes:

$$w \cdot x_i + b \geq +1 \qquad \text{for all} \quad y = +1, \tag{1}$$

$$w \cdot x_i + b \leq -1 \qquad \text{for all} \quad y = -1. \tag{2}$$

The aforementioned inequalities can be summed up in relation (3):

$$y_i(w \cdot x_i + b) - 1 \geq 0. \tag{3}$$

Support vectors are all points that exist in two parallel hyperplanes with the optimal hyperplane and are defined by the function $(w \cdot x_i + b) \pm 1 = 0$. If there is a hyperplane and Eq. (3) is established, the classes are separated linearly. Therefore, the margin between these aircrafts is $2/\|w\|$, distance to the nearest point. Minimize $\|w\|^2$ with the limit of Eq. (3) found the optimal hyperplane. Therefore, the following optimization problem must be solved to calculate the optimal aircraft hyperplane:

$$\min\left[\frac{1}{2}\|w\|^2\right]. \tag{4}$$

Of course subject to restrictions:

$$w \cdot x_i + b \geq 1 \qquad \text{and} \qquad y \in \{-1, +1\}. \tag{5}$$

Figure 3 shows the classification of separable nonlinear data. For such data, it is certainly not possible to

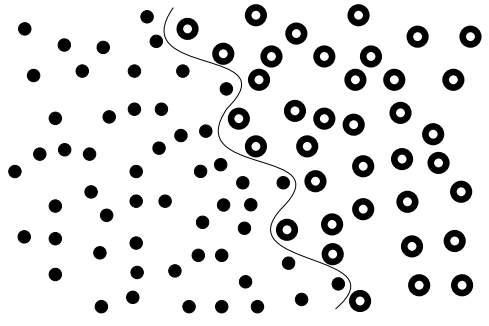


Fig. 3. Nonlinear separation of data.

classify data in two classes linearly. Therefore, in such cases where it is not possible to use a hyperplane with linear equations on the data, nonlinear decision levels should be used. Therefore, ξ slack variables replace the optimization problem (Fig. 4):

$$\min\left[\frac{\|w\|^2}{2} + C\sum_{i=1}^{r}\xi_i\right],$$

considering the following limitations:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, ..., N, \tag{6}$$

where C is a constant parameter and balances between two criteria, error minimization and margin maximization. Slack variables ξ_i show the distance of classified points from the optimal hyperplane cloud. If it is not possible to use a hyperplane with linear equations, it may be mapped into a high-dimensional space ($\circ D$) through some nonlinear mapping functions ($\varnothing$).

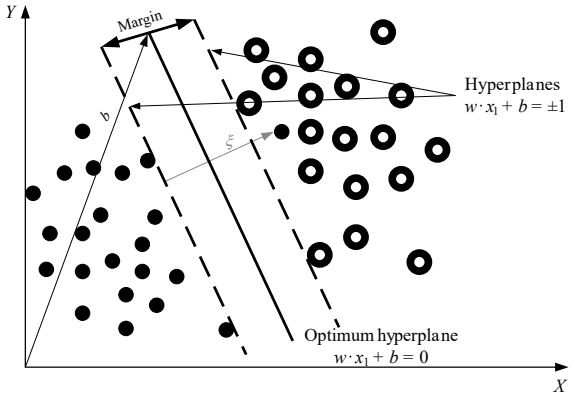


Fig. 4. Introducing slack variable for nonlinear data and generalization of the solution.

As shown in Fig. 5, the point x can be represented as $\varnothing(x)$ in the feature space. Complex computations $(\varnothing(x) \cdot \varnothing(x_i))$ are reduced using a kernel function. Therefore, the decision function for classification is as follows:

$$f(x) = \text{sign}\left(\sum_{i}^{z} a_i y_i \cdot K(X, X_i) + b\right), \tag{7}$$

where each z of the training case, there are X_i vectors that indicate the spectral response, $a_i y_i$ are Lagrange multipliers, and $K(X, X_i)$ is the kernel function, a_i depends on parameter C and its value is determined by it.

Kernel functions can be classified into four groups. In the following, 4 groups of SVM kernels are presented:

– linear kernel: the simplest kernel function is a linear kernel. Inner product of $X_i \cdot X_j$ and the constant coefficient C represent the linear kernel. Equation (8) represents the linear kernel function:

$$K(X_i, X_j) = X_i \cdot X_j; \tag{8}$$

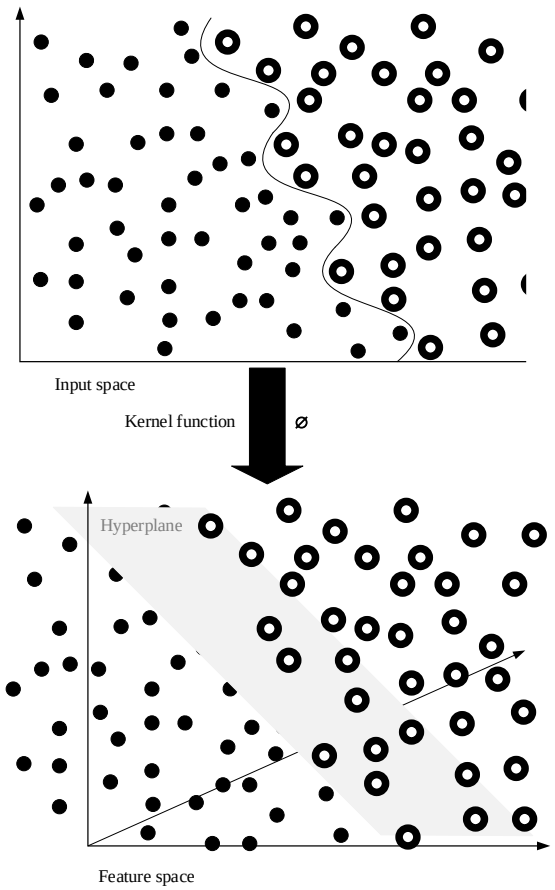


Fig. 5. Mapping of the data sets to the high-dimensional space with a kernel function.

– polynomial kernel: data separation is not possible linearly when useful features are not extracted or the amount of noise is high. To solve this problem, data must be mapped in different feature spaces so that they can be separated linearly. One of the kernel functions used in nonlinear separation is the polynomial kernel:

$$K\left(X_i, X_j\right) = \left(\gamma X_i \cdot X_j + C\right)^d, \tag{9}$$

where $C \geq 0$ represents a free parameter in the polynomial that trades off the impact of higher-order versus lower order terms;

– RBF kernel: the Gaussian kernel or RBF kernel is shown in Eq. (10):

$$K_{\mathrm{RBF}}\left(X_i, X_j\right) = \exp\left(-\gamma \left|X_i \cdot X_j\right|^2\right), \tag{10}$$

where γ is a parameter that used to set the spread of the kernel.

In this paper, the One-vs-All (OVA) method is used for classification.

3. Micro-Doppler

Micro-motions, such as vibrations or rotations of an object or structure on an object (Saffari *et al.*, 2022b; Yang *et al.*, 2006), cause changes in the extra frequencies on the signal, leading to sidebands on the object's Doppler frequency (Saffari *et al.*, 2022b; Tahmoush, 2015). This phenomenon is called sonar micro-Doppler. Recent research has shown that micro-Doppler techniques can identify or classify a target with its micro-Doppler properties. To explore the micro-Doppler properties of an object, time-frequency analysis is used to provide information about these local properties over time and frequency (Smith *et al.*, 2007). In most cases, the micro motion has a unique signature object. Micro-motion is created directly by the dynamic motion properties of an object, and the micro-Doppler properties are a direct reflection of micro-motion. Therefore, a micro-Doppler signature can be used to classify an object with unique motion characteristics (Yang *et al.*, 2006).

3.1. Theory

The analytic signal of a pure tone $s(t)$ is defined as the signal $\widehat{s}(t)$, such that $s(t)$ ¼ Real $\{\widehat{s}(t)\}$, and is generally expressed in a polar format as (Chen *et al.*, 2014):

$$\widehat{s}(t) = e^{j2\pi f_0 t}. \tag{11}$$

A target moving at a constant radial velocity u has the following Doppler shift relative to the sonar (or radar) system:

$$f_D = 2f_0 \frac{v}{C_s}, \tag{12}$$

where f_0 is the carrier frequency of the active sensor and C_s is the speed of propagation of sound in water (or air). If the target has a number M parts and each part moves at a velocity component $v_i(t)$, the Doppler shift is the sum of each single Doppler shift:

$$f_D(t) = \sum_{i=1}^{M} 2f_0 \frac{v_i(t)}{C_s}. \tag{13}$$

For such a target, the analytical signal of the echo return is as follows:

$$\widehat{s}_R(t) = e^{j2\pi f_0 t} \cdot e^{j2\pi f_D(t) t}. \tag{14}$$

Mixing the received signal $\hat{s}_R(t)$ with the conjugate of the transmitted signal $\hat{s}(t)$ is as follows:

$$\widehat{s}_R(t) \cdot \widehat{s}(t)^* = e^{j2\pi f_D(t) t}. \tag{15}$$

The aforementioned relation makes it possible to extract the Doppler signature from the data. This is the signal component that contains the micro-Doppler information on the target, which can be used for target recognition and classification.

4. Design of automatic sonar target recognition system using sonar micro-Doppler signature

Like any other pattern recognition system, designing an automatic sonar target recognition system has the following steps.

4.1. Data acquisition

One of the severe challenges for sonar research is the lack of reliable data. On the other hand, such things as the complex and heterogeneous environment of the sea, as well as the presence of unwanted signals in the sea (noise, clutter, and resonance) are the motivation for preparing a simulated data set using the mathematical model of the return signal of the rotation of the propeller. The targets tested are according to Table 1.

Table 1. Information on reference targets.

Number	Type	Model
1	Container ship	Emma Maersk
2	Container ship	MV Barzan
3	Container ship	MSC Oscar
4	Oil tanker	Front Century
5	Oil tanker	Seawise Giant
6	Passenger ship	Motorboat
7	Passenger ship	Oasis of the seas
8	Passenger ship	Leading Atlas
9	Cruise	Harmony of the Seas
10	Tugboat	ASD TUG 2913
11	Tugboat	Chinese oceanic tug boat
12	Research vessel	Nathaniel B. Palmer
13	Autonomous underwater vehicle	Tech 475 AUV
14	Military	Torpedo No. 1
15	Military	Torpedo No. 2
16	Military	Logistic Support
17	Military	Littoral Combat Ship No. 1
18	Military	Littoral Combat Ship No. 2
19	Military	Destroyer No. 1
20	Military	Destroyer No. 2
21	Military	Frigates
22	Military	Aircraft Carrier
23	Military	light submarine
24	Military	semi-heavy submarine
25	Military	heavy submarine

Different types of vessels were used in selecting the samples, including container vessels, tankers, passengers, cruises, autonomous underwater vehicles, tugboats, different classes of navy ships, submarines, and military torpedoes to evaluate the performance of the proposed model.

4.2. Extracting feature vectors

To generate a data set of return signals from the rotating part (propeller) of sonar targets, which is discussed in Subsec. 4.1, a suitable mathematical model (Eq. (16)) was used to simulate the return signal from the propeller:

$$s_r(t) = \sum_{n=0}^{N-1} A_r (L_2 - L_1) e^{j(\beta)} \cdot \operatorname{sinc}\left(\frac{4\pi}{\lambda}\frac{(L_2-L_1)}{2}\cos(\theta)\sin\left(\omega_r t + \frac{2\pi n}{N}\right)\right), \quad (16)$$

$$\beta = \omega_c t - \frac{4\pi}{\lambda} \cdot \left(R + vt + \frac{L_1+L_2}{2}\cos(\theta)\sin\left(\omega_r t + \frac{2\pi n}{N}\right)\right).$$

The parameters used in Eq. (16) are described in Table 2.

Table 2. Relationship parameters (16).

Parameters	Descriptions
$s_r(t)$	Return signal in time
N	Number of blades
A_r	A normalizing factor
L_1	The distance from the beginning of the blades to the center of rotation
L_2	The distance from the end of the blades to the center of rotation
ω_c	Radian frequency of the transmitted signal
λ	The wavelength of the transmitted signal
R	The distance from the center of rotation to the sonar receiver
v	Target speed relative to sonar receiver
θ	Target viewing angle
ω_r	Radian frequency of rotation

Figure 6 shows how to obtain a return signal using Eq. (16).

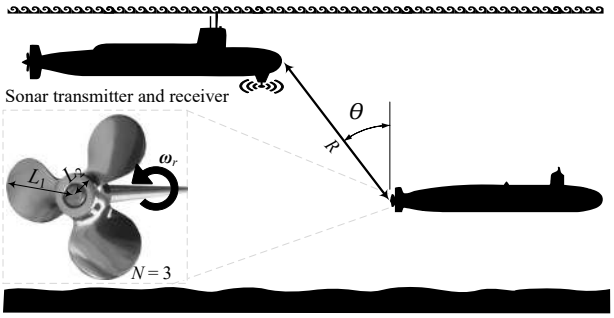


Fig. 6. How to obtain a return signal using Eq. (16).

The feature extracted from these signals is the Components 128-point from FFT. The structure of the feature vector expresses the target in the viewing angle (θ), and the SNR specified as follow:

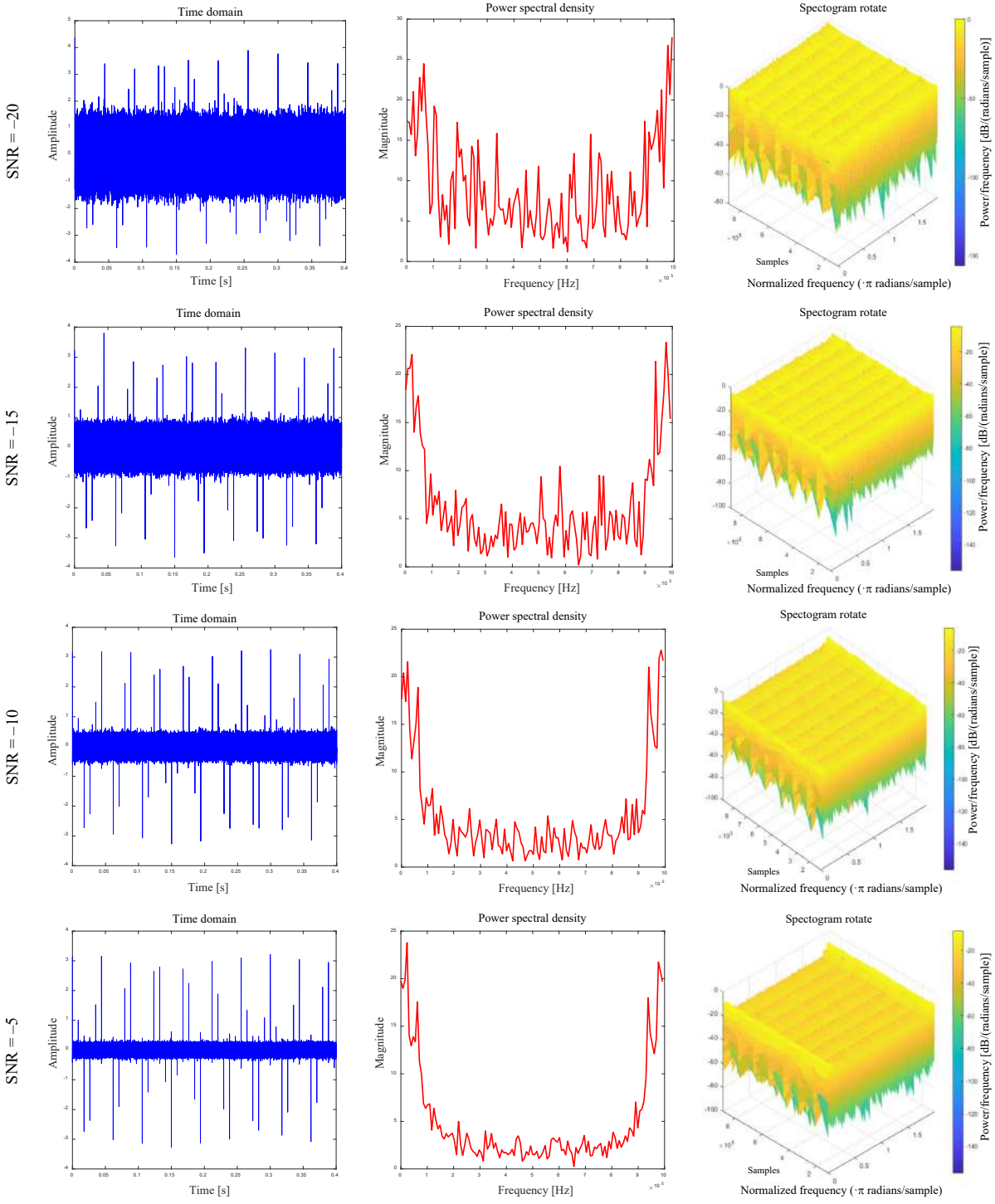
$$\text{feature vector} = [f_1, f_2, f_3, ..., f_{126}, f_{127}, f_{128}]_{(\text{SNR},\theta)}, \quad (17)$$

where each of its components corresponds to a point of 128-point FFT in the angle of view and the specified SNR.

The reference classes correspond to the twenty-five objectives of Table 1. Samples of each class include feature vectors in nine SNRs (20, 15, 10, 5, 0, –5, –10, –15, and –20 dB) and eight viewing angles (10, 20, 30, 40, 50, 60, 70, and 80 degrees). Each class contains 30 samples in SNR and specified viewing angles. Thus, there are 2160 samples in each class (corresponding to each target) for all SNRs and viewing angles. In total, the dataset contains 54,000 samples.

Figure 7 shows samples of simulated acoustic signals and frequency signatures from sonar micro-Doppler at different SNRs for target No. 8.

Figure 8 shows the effect of the viewing angles on the return signal at SNR = –20 for target No. 8.



[Fig. 7.]

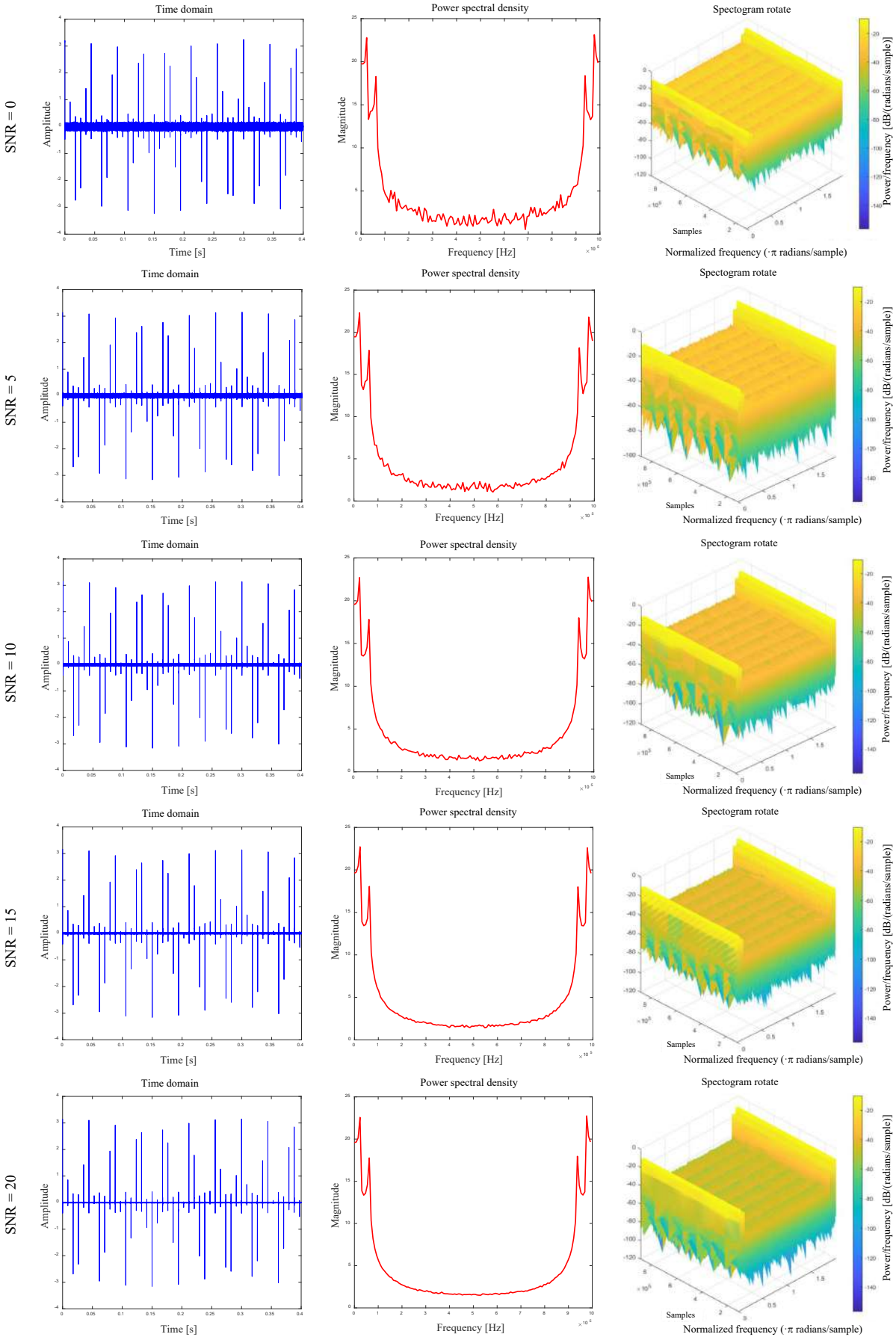
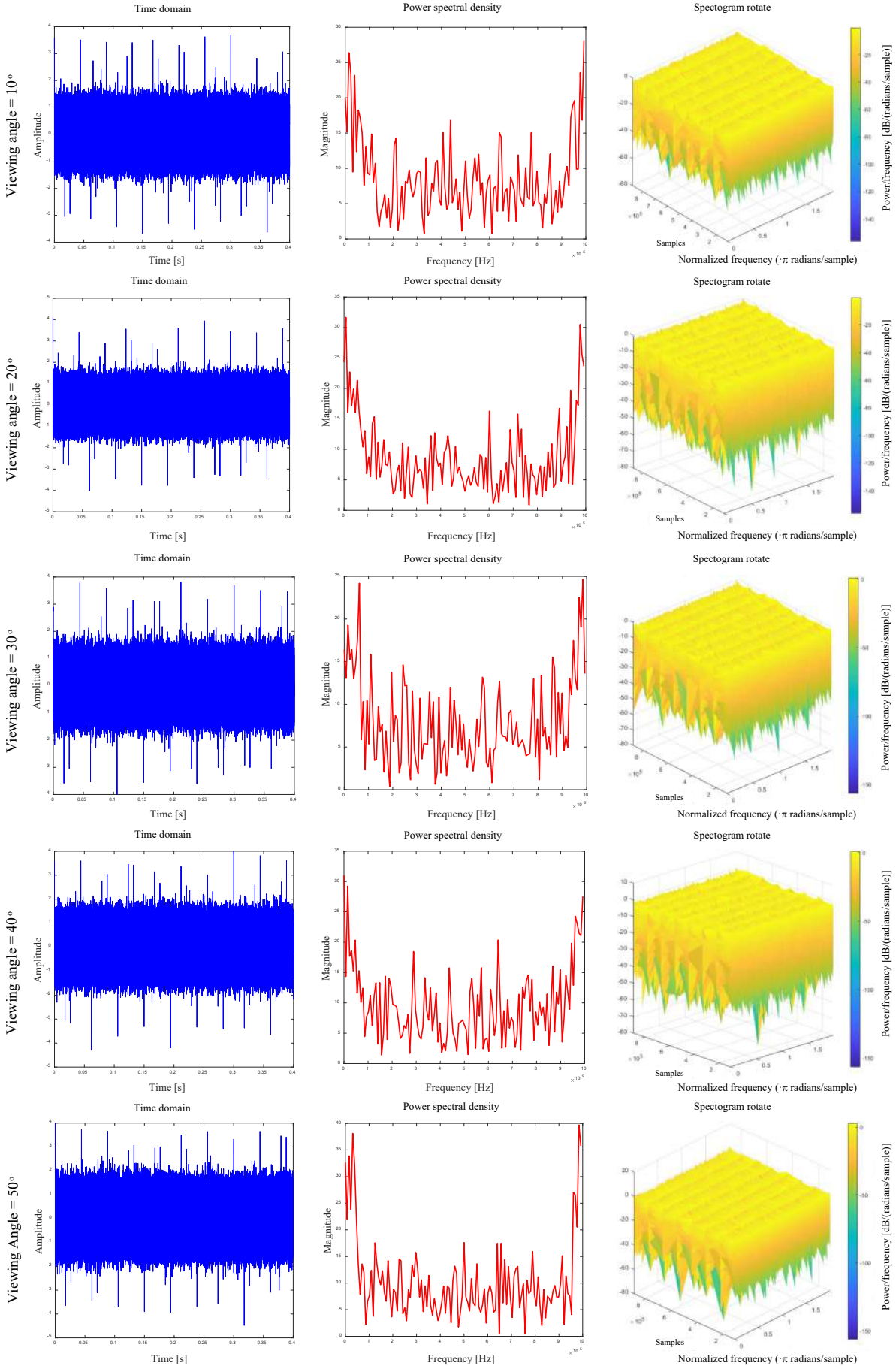


Fig. 7. Samples of simulated acoustic signals and frequency signatures of sonar micro-Doppler in different SNRs for target No. 8.


Time domain
Power spectral density
Spectogram rotate
Viewing angle = 10°
Amplitude
Time [s]
Magnitude
Frequency [Hz]
Samples
Normalized frequency (·π radians/sample)
Power/frequency [dB/(radians/sample)]
Viewing angle = 20°
Viewing angle = 30°
Viewing angle = 40°
Viewing Angle = 50°

[Fig. 8.]

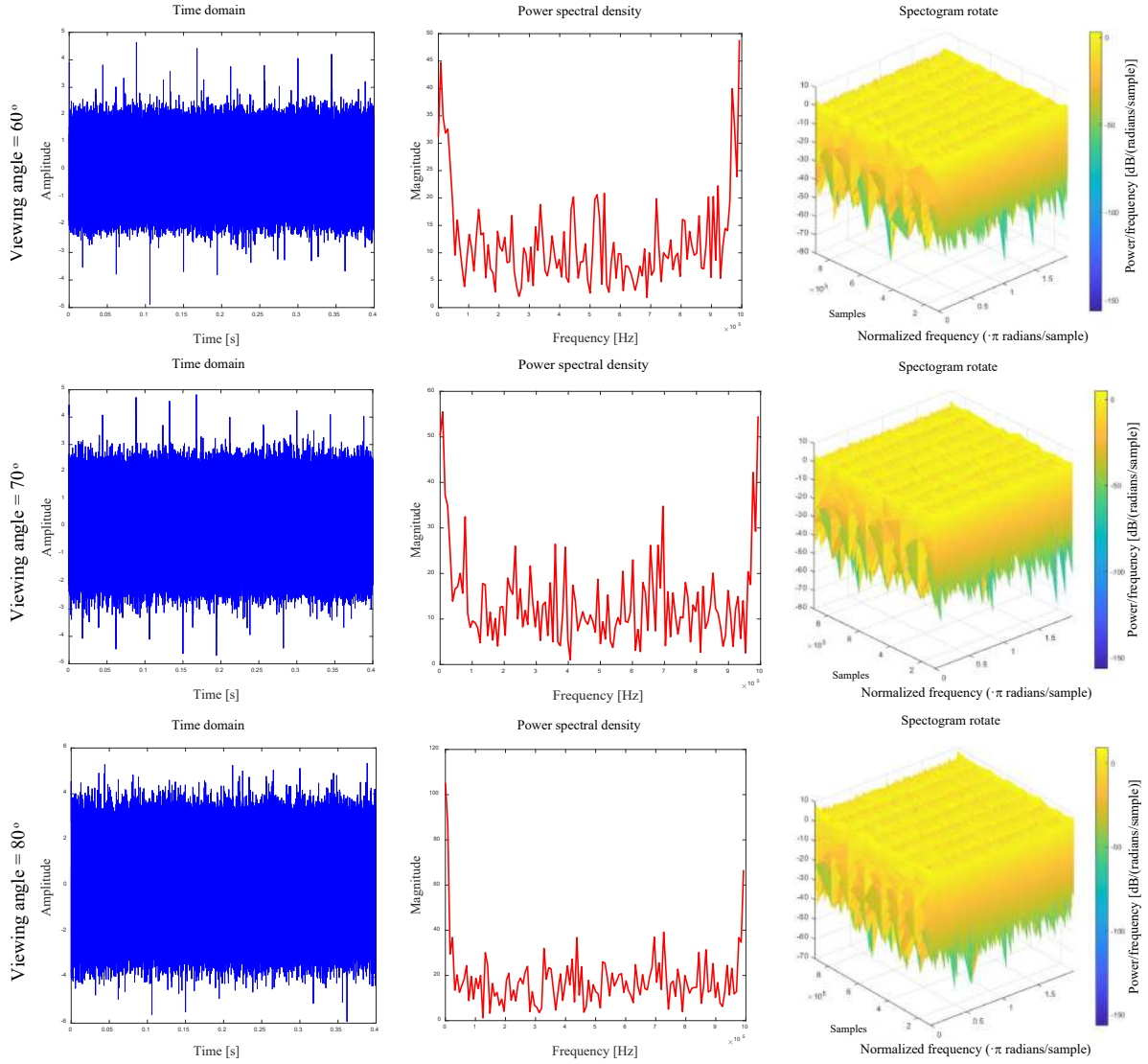


Fig. 8. Effect of the viewing angles on the return signal at SNR = −20 for target No. 8.

4.3. How to decide on a sonar target

This paper uses the three main kernel functions RBF, linear, and polynomial SVM classifier.

4.4. How to add noise

Noise mixed with the return signal from the target is assumed to be white Gaussian noise (Zhong *et al.*, 2022), and its random samples are simulated using uniform random variables. The power of the noise changes with the change of its variance. Different SNR ratios are performed by separately changing the noise power for each target. The signal strength for each target is its corresponding power at the same viewing angle.

5. Simulation results

In this section, the results obtained from the simulated system are analyzed. The results of the classification are the average of 15 program executions for each of the experiments. Each experiment assumes that the target angle of view and the SNR ratio is known. Due to the random nature of noise, in order to get the reference information in each SNR ratio as comprehensive as possible, the operation of generating random noise samples is performed 30 times. 70% of the samples are used to form a reference class related to a specific target and the other 30% are used as experimental data (unknown targets).

In general, according to the simulation results, it can be seen that in the SVM classifier, the RBF kernel works better for optimal conditions in which the noise level is lower. As the amount of noise increases, the linear kernel provides better results. Therefore, RBF, linear, and polynomial kernels performed better in terms of classification accuracy. Table 3 shows the results of the classification with the classifiers SVM-RBF, SVM-linear, and SVM-polynomial.

Table 3. The results of the classification with the classifiers SVM-RBF, SVM-linear, and SVM-polynomial.

SNR [dB]	Kernel	Angle 10°	Angle 20°	Angle 30°	Angle 40°	Angle 50°	Angle 60°	Angle 70°	Angle 80°
20	linear	93.008	92.016	91.696	90.160	90.192	88.672	86.960	81.520
	polynomial	94.432	94.800	94.096	92.736	92.480	91.440	84.592	77.872
	RBF	98.272	97.920	98.416	97.920	98.352	97.760	97.936	74.288
15	linear	92.688	92.032	91.088	90.272	91.136	87.776	84.864	79.056
	polynomial	92.016	92.096	91.840	90.672	90.848	88.864	80.480	70.656
	RBF	98.240	97.776	98.528	97.840	98.176	95.552	71.808	52.720
10	linear	92.144	91.600	89.088	89.088	88.640	86.720	81.968	75.024
	polynomial	87.824	87.600	86.384	85.536	84.048	82.416	74.432	64.144
	RBF	91.504	89.328	83.104	86.992	73.536	62.240	50.112	38.672
5	linear	90.032	89.760	87.136	88.832	86.992	82.688	75.664	75.488
	polynomial	81.904	82.176	79.136	81.264	78.448	71.328	63.200	58.096
	RBF	56.464	57.280	54.512	57.392	51.232	42.944	34.576	33.168
0	linear	82.912	83.936	83.984	80.464	81.056	77.072	74.208	64.256
	polynomial	70.240	71.216	71.248	69.488	70.912	64.640	59.488	53.136
	RBF	46.128	31.776	48.480	42.192	40.912	37.184	30.560	31.440
−5	linear	83.200	82.624	78.976	78.336	76.976	74.800	74.448	66.224
	polynomial	68.576	67.056	67.488	63.280	63.280	60.304	58.656	53.232
	RBF	38.720	34.720	39.312	38.720	35.424	43.904	38.960	44.992
−10	linear	70.864	69.520	67.584	65.680	69.488	70.944	66.816	68.560
	polynomial	53.776	52.384	50.080	51.040	50.816	50.784	51.136	48.656
	RBF	26.512	30.016	22.576	29.696	34.896	24.464	32.256	27.952
−15	linear	68.784	69.376	69.296	70.016	63.904	63.024	66.464	58.672
	polynomial	46.544	49.920	47.360	48.752	44.736	45.840	49.120	46.144
	RBF	25.808	24.544	28.496	30.992	22.208	24.832	15.104	12.688
−20	linear	70.112	73.568	69.312	67.840	67.440	66.752	64.576	62.304
	polynomial	48.160	45.056	48.656	46.800	45.072	48.128	52.208	49.568
	RBF	21.712	23.200	31.632	28.000	18.656	34.736	20.592	33.600

However, in terms of use in real environment with a lot of noise, linear, polynomial, and RBF kernels perform better.

Figure 9 shows the correct recognition rate for different SNR ratios at a 10° viewing angle for the three SVM classifier kernels, and Fig. 10 shows the correct recognition rate for different viewing angles at SNR = 20 dB for the three SVM classifier kernels.

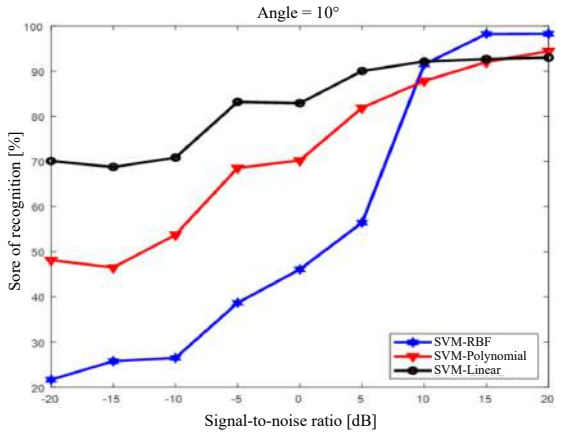


Fig. 9. Comparison of correct recognition rate for different SNR ratios at a 10° viewing angle for the three SVM classifier kernels.

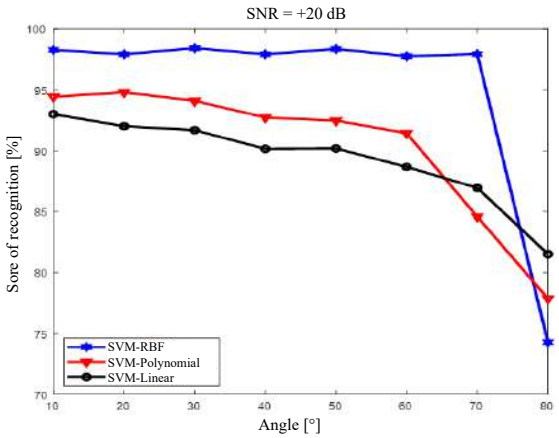


Fig. 10. Comparison of correct recognition rate for different viewing angles at SNR = 20 dB for the three SVM classifier kernels.

For a more comprehensive comparison, Table 4 presents the target recognition results using MLP-BP and MLP-GWO.

Table 4. The results of the classification with the classifiers MLP-BP and MLP-GWO.

SNR [dB]	Classifier	Angle 10°	Angle 20°	Angle 30°	Angle 40°	Angle 50°	Angle 60°	Angle 70°	Angle 80°
20	MLP-BP	13.60	19.60	20.00	15.10	16.80	18.39	14.43	11.22
	MLP-GWO	17.13	17.00	16.26	17.00	17.00	17.00	13.03	13.01
15	MLP-BP	14.39	11.66	13.60	8.79	11.61	12.40	7.19	5.6
	MLP-GWO	17.18	17.16	17.00	17.00	17.14	17.01	13.57	13.02
10	MLP-BP	11.51	11.62	10.41	9.94	8.42	7.12	8.82	7.83
	MLP-GWO	17.17	17.05	17.01	17.04	17.02	17.01	13.00	13.00
5	MLP-BP	8.41	9.34	4.00	8.46	10.81	9.63	6.47	7.19
	MLP-GWO	17.16	17.55	17.01	17.27	17.02	17.00	13.03	12.87
0	MLP-BP	7.19	7.87	5.20	7.19	6.80	4.39	4.39	4.80
	MLP-GWO	17.16	17.01	17.27	17.26	17.01	17.00	13.00	12.34
−5	MLP-BP	3.21	4.00	5.62	6.40	6.00	4.00	5.20	5.20
	MLP-GWO	17.14	17.40	17.40	17.01	17.40	17.02	13.01	13.00
−10	MLP-BP	3.60	2.80	4.00	7.60	6.00	4.00	2.41	4.82
	MLP-GWO	17.13	17.14	17.20	17.01	17.06	17.03	16.35	16.07
−15	MLP-BP	3.15	4.39	2.82	4.00	3.20	4.00	6.89	3.20
	MLP-GWO	16.07	16.80	15.62	16.71	14.80	14.75	14.68	13.80
−20	MLP-BP	3.20	4.12	5.69	3.20	5.61	5.54	3.11	3.59
	MLP-GWO	15.00	15.87	15.28	13.94	13.96	13.67	13.69	12.94

As shown in Table 4, MLP-BP has shown poor performance. The results for MLP-GWO are better than MLP-BP. However, in both classifiers, the results are not satisfactory. One of the reasons for the poor performance of neural networks is the number of classes. Obviously, increasing the number of classes causes an increase in the probability of errors and, as a result, a decrease in performance.

6. Conclusion

This paper uses a new method of using sonar micro-Doppler to automatically detect moving sonar targets. In other words, when the signal hits the propeller, each propeller has a unique effect on the signal according to its own metrics. Then, by transferring the signal to the frequency domain, its useful properties were extracted. Movable sonar targets were classified using linear kernels, RBF and polynomial SVM classifiers. The simulation results showed that the RBF kernel is significantly suitable for positive signal-to-noise ratios. For values with more noise, the linear kernel has different and significant performance compared to the other two kernels. Using GWO algorithm for neural network training improved the performance of the classifier compared to using BP for training, but overall, the result of using neural networks was not satisfactory. Therefore, the approach of using neural network is not suitable for this problem. However, due to the new idea of using sonar micro-Doppler to classify moving targets, the need to consider other machine learning methods and artificial intelligence techniques to improve the classifier performance in all SNRs and viewing angles is strongly noticeable.

Some of the topics that will be explored for future research are as follows:

- improving the performance of SVM classifiers using metaheuristic algorithms;
- use other machine learning algorithms to improve accuracy;
- use of hybrid classifiers to achieve more accurate accuracy for sensitive applications;
- use deep learning to improve classifier performance.

References

1. Anderson S.J. (2004), Target classification, recognition and identification with HF radar, [in:] *RTO SET Symposium on Target Identification and Recognition Using RF Systems*, pp. 1–20.
2. Berthold T., Leichter A., Rosenhahn B., Berkhahn V., Valerius J. (2018), Seabed sediment classification of side-scan sonar data using convolutional neural networks, [in:] *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–8, doi: 10.1109/SSCI.2017.8285220.
3. Bhanu B. (1986), Automatic target recognition: State of the art survey, *IEEE Transactions on Aerospace and Electronic Systems*, **22**(4): 364–379, doi: 10.1109/TAES.1986.310772.
4. Cai X. *et al.* (2021), Dynamically controlling terahertz wavefronts with cascaded metasurfaces, *Advanced Photonics*, **3**(3): 036003, doi: 10.1117/1.AP.3.3.036003.

5. Chen B. *et al.* (2022), DPM-LES investigation on flow field dynamic and acoustic characteristics of a twin-fluid nozzle by multi-field coupling method, *International Journal of Heat and Mass Transfer*, **192**: 122927, doi: 10.1016/j.ijheatmasstransfer.2022.122927.
6. Chen H., Li S. (2022), Collinear nonlinear mixed-frequency ultrasound with FEM and experimental method for structural health prognosis, *Processes*, **10**(4): 656, doi: 10.3390/pr10040656.
7. Chen V.C., Tahmoush D., Miceli W.J. (2014), *Radar Micro-Doppler Signature: Processing and Applications*, Institution of Engineering and Technology.
8. 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.
9. Dong J., Deng R., Quanying Z., Cai J., Ding Y., Li M. (2021), Research on recognition of gas saturation in sandstone reservoir based on capture mode, *Applied Radiation and Isotopes*, **178**: 109939, doi: 10.1016/j.apradiso.2021.109939.
10. Ehrman L.M., Lanterman A.D. (2020), Incorporation of aircraft orientation into automatic target recognition using passive radar, *IET Radar, Sonar & Navigation*, **14**(7): 1079–1087, doi: 10.1049/iet-rsn.2020.0010.
11. Foued C., Ammar M., Arezki Y. (2017), Detection and classification of ground targets using a Doppler RADAR, [in:] *2017 Seminar on Detection Systems Architectures and Technologies (DAT)*, pp. 1–8, doi: 10.1109/DAT.2017.7889168.
12. Gaye B., Zhang D., Wulamu A. (2021), Improvement of support vector machine algorithm in big data background, *Mathematical Problems in Engineering*, **2021**: 5594899, doi: 10.1155/2021/5594899.
13. Glover F., Laguna M. (2008), *Principles of Tabu Search*, Springer Science.
14. Jin C. *et al.* (2020), Development and evaluation of an artificial intelligence system for COVID-19 diagnosis, *Nature Communications*, **11**: 5088, doi: 10.1038/s41467-020-18685-1.
15. Kaveh M., Khishe M., Mosavi M.R. (2019), Design and implementation of a neighborhood search biogeography-based optimization trainer for classifying sonar dataset using multi-layer perceptron neural network, *Analog Integrated Circuits and Signal Processing*, **100**(2): 405–428, doi: 10.1007/s10470-018-1366-3.
16. Kavzoglu T., Colkesen I. (2009), A kernel functions analysis for support vector machines for land cover classification, *International Journal of Applied Earth Observation and Geoinformation*, **11**(5): 352–359, doi: 10.1016/j.jag.2009.06.002.
17. Kazimierski W., Zaniewicz G. (2021), Determination of process noise for underwater target tracking with forward looking sonar, *Remote Sensing*, **13**(5): 1014, doi: 10.3390/rs13051014.
18. Khishe M., Mohammadi H. (2019), Passive sonar target classification using multi-layer perceptron trained by salp swarm algorithm, *Ocean Engineering*, **181**: 98–108, doi: 10.1016/j.oceaneng.2019.04.013.
19. Khishe M., Mosavi M.R. (2019), Improved whale trainer for sonar datasets classification using neural network, *Applied Acoustics*, **154**: 176–192, doi: 10.1016/j.apacoust.2019.05.006.
20. Khishe M., Safari A. (2019), Classification of sonar targets using an MLP Neural Network trained by Dragonfly Algorithm, *Wireless Personal Communications*, **108**: 2241–2260, doi: 10.1007/s11277-019-06520-w.
21. Koturwar P., Girase S., Mukhopadhyay D. (2015), A survey of classification techniques in the area of big data, *Computer Science*, pp. 1–7.
22. Lakhwani K. (2020), Big data classification techniques: A systematic literature, *Journal of Natural Remedies*, **21**(2).
23. Liu J., Zhou Z., Zeng X. (2019), Multi-static active sonar target recognition method based on bionic signal, [in:] *2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP)*, pp. 285–289, doi: 10.1109/ICICSP48821.2019.8958532.
24. Mamgain R., Jain R., Deb D. (2018), Study and Simulation of Radar Targets' Micro-Doppler Signature, [in:] *2018 International Conference on Radar (RADAR)*, pp. 1–5, doi: 10.1109/RADAR.2018.8557264.
25. Molchanov P.O., Astola J.T., Egiazarian K.O., Totsky A.V. (2013), Classification of ground moving targets using bicepstrum-based features extracted from Micro-Doppler radar signatures, *EURASIP Journal on Advances in Signal Processing*, **2013**(1): 1–13, doi: 10.1186/1687-6180-2013-61.
26. Qiao W., Khishe M., Ravakhah S. (2021), Underwater targets classification using local wavelet acoustic pattern and Multi-Layer Perceptron neural network optimized by modified Whale Optimization Algorithm, *Ocean Engineering*, **219**: 108415, doi: 10.1016/j.oceaneng.2020.108415.
27. Saffari A., Khishe M., Zahiri S. (2022a), Fuzzy-ChOA: An improved chimp optimization algorithm for marine mammal classification using artificial neural network, *Analog Integrated Circuits and Signal Processing*, **111**: 403–417, doi: 10.1007/s10470-022-02014-1.
28. Saffari A., Zahiri S., Khishe M. (2022b), Automatic recognition of sonar targets using feature selection in micro-Doppler signature, *Defence Technology*, doi: 10.1016/j.dt.2022.05.007.
29. Saffari A., Zahiri S.H., Khishe M. (2022c), Fuzzy grasshopper optimization algorithm: A hybrid technique for tuning the control parameters of GOA using Fuzzy System for big data sonar classification, *Iranian Journal of Electrical and Electronic Engineering*, **18**(1): 1–13, doi: 10.22068/IJEEE.18.1.2131.

30. Saffari A., Zahiri S.H., Khishe M. (2022d), Fuzzy whale optimisation algorithm: A new hybrid approach for automatic sonar target recognition, *Journal of Experimental & Theoretical Artificial Intelligence*, doi: 10.1080/0952813X.2021.1960639.

31. Sclavounos P.D., Ma Y. (2018), Artificial intelligence machine learning in marine hydrodynamics, [in:] *Proceedings of the ASME 2018 37th International Conference on Ocean, Offshore and Arctic Engineering. Volume 9: Offshore Geotechnics; Honoring Symposium for Professor Bernard Molin on Marine and Offshore Hydrodynamics*, doi: 10.1115/OMAE2018-77599.

32. Smith G.E., Woodbridge K., Baker C.J. (2007), Multiperspective micro-Doppler signature classification, [in:] *IET International Conference on Radar Systems 2007*, doi: 10.1049/cp:20070522.

33. Tahmoush D. (2015), Review of micro-Doppler signatures, *IET Radar, Sonar & Navigation*, **9**(9): 1140–1146, doi: 10.1049/iet-rsn.2015.0118.

34. Uddin S., Khan A., Hossain M.E., Moni M.A. (2019), Comparing different supervised machine learning algorithms for disease prediction, *BMC Medical Informatics and Decision Making*, **19**: 281, doi: 10.1186/s12911-019-1004-8.

35. Waite A.D. (2002), *Sonar for Practising Engineers*, Wiley.

36. Wawrzyniak N., Stateczny A. (2017), MSIS image postioning in port areas with the aid of comparative navigation methods, *Polish Maritime Research*, **24**(1): 32–41, doi: 10.1515/pomr-2017-0004.

37. Wu C., Khishe M., Mohammadi M., Taher Karim S.H., Rashid T.A. (2021), Evolving deep convolutional neutral network by hybrid sine–cosine and extreme learning machine for real-time COVID19 diagnosis from X-ray images, *Soft Computing*, doi: 10.1007/s00500-021-05839-6.

38. Xu H., Zhou J., Asteris P.G., Armaghani D.J., Tahir M.M. (2019), Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate, *Applied Sciences*, **9**(18): 3715, doi: 10.3390/app9183715.

39. Yang Y., Lei J., Zhang W., Lu C. (2006), Target classification and pattern recognition using micro-Doppler radar signatures, [in:] *Seventh ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD'06)*, pp. 213–217, doi: 10.1109/SNPD-SAWN.2006.68.

40. Zhang T., Liu S., He X., Huang H., Hao K. (2020), Underwater target tracking using forward-looking sonar for autonomous underwater vehicles, *Sensors*, **20**(1): 102, doi: 10.3390/s20010102.

41. Zhong T., Cheng M., Lu S., Dong X., Li Y. (2022), RCEN: A deep-learning-based background noise suppression method for DAS-VSP records, [in:] *IEEE Geoscience and Remote Sensing Letters,* **19**: 3004905, doi: 10.1109/LGRS.2021.3127637.

42. Zhou J., Bai J., Liu Y. (2022), Fabrication and modeling of matching system for air-coupled transducer, *Micromachines*, **13**(5): 781, doi: 10.3390/mi13050781.

Archives of Acoustics Vol. 48, No. 1, pp. 63–70 (2023), doi: 10.24425/aoa.2022.142907

Research Paper

Improvement of Sound Insulation Through Double-Panel Structure by Using Hybrid Local Resonator Array

Kyong-Su RI, Myong-Jin KIM*, Se-Hyon SON-U

Institute of Acoustics, Department of Physics, Kim Il Sung University
Pyongyang, Democratic People's Republic of Korea

*Corresponding Author e-mail: mj.kim0903@ryongnamsan.edu.kp

(*received March 5, 2021; accepted September 2, 2022*)

In this paper, we present one approach to improve the soundproofing performance of the double-panel structure (DPS) in the entire audible frequencies, in which two kinds of local resonances, the breathing-type resonance and the Helmholtz resonance, are combined. The thin ring resonator row and slit-type resonator (Helmholtz resonator) row are inserted between two panels of DPS together. Overlapping of the band gaps due to the individual resonances gives a wide and high band gap of sound transmission in the low frequency range. At the same time, the Bragg-type band gap is created by the structural periodicity of the scatterers in the high audible frequency range. In addition, the number of scatterer rows and the filling factor are investigated with regard to the sound insulation of DPS with sonic crystals (SCs). Consequently, the hybrid SC has the potential of increasing the soundproofing performance of DPS in the audible frequency range above 1 kHz by about 15 dB on average compared to DPS filled only with glass wool between two panels, while decreasing the total thickness and mass compared to the counterparts with the other type of local resonant sonic crystal.

Keywords: sound insulation; local resonance; insertion loss; sonic crystal.

1. Introduction

Brillouin (1946) described mathematically the propagation of waves in periodic structures in 1946. Subsequently, the transmission properties of the electromagnetic wave in periodic structures was studied in late 1980s. It was shown that infinite periodic structures do not support wave propagation in certain frequency ranges relating to the spacing between the scattering elements (Lattice constant).

The first experimental study on the sound attenuation by periodic structure was undertaken in 1995 (Martínez-Sala *et al.*, 1995). The sculpture consists of a periodic square symmetry arrangement of stainless steel pipe with a diameter of 29 mm, a distance between 2 pipes next to each other of 100 mm. Filling factor of 0.066 was used in their experiment. Sound attenuation was measured at various angles in outdoor conditions for sound wave incidences perpendicular to the cylinders' vertical axes, resulting in several maxima (sound attenuation) and minima (sound reinforcement) in the frequency spectrum. The first (lowest) band gap had a center frequency at 1.7 KHz, which could be attributed to the periodicity of the structure. Ever since, researches on the application of periodic arrays of cylinders for noise control have increased (Sanchez-Perez *et al.*, 2002; Romero-García *et al.*, 2011). Investigation of sound attenuation effects of regularly planted trees has also been conducted (Martínes-Sala *et al.*, 2006). There were numerical and theoretical works in respect to periodic arrays of elastic scatterers in gas involving hollow spheres and cylinders (Sainidou *et al.*, 2006).

Local sound absorption properties have been used in sonic crystal noise barriers. For example, both numerical and experimental studies have been made of the reflectance and transmittance spectra of the sonic crystals (SCs) consisted of rigid perforated cylindrical shells filled with recycled rubber crumb material (Sánchez-Dehesa *et al.*, 2011). Such design offers the additional mechanism of absorption, apart from the multiple scattering phenomenon in periodic structure, to further attenuate noise (Umnova *et al.*, 2006).

A conventional way for reducing low frequency noise transmission is to increase the thickness or mass per unit area of the sound insulation material. However, a drawback of this technique is that usually it results in large size and mass of the insulator. One possibility for targeting the band gap without increasing the size is to use the resonant scatterers. Besides Helmholtz resonator, there are several cavity-type resonators such as half- and quarter-wave tubes, which are used comonly on air intake systems to attenuate noise at specific frequencies (Sohn, Park, 2011).

The concept of split ring resonator (SRR) was initially introduced in 1999 in the electromagnetism field. There have been theoretical and experimental studies on SRR which gave the frequency range related to the resonant frequency in which waves could not propagate through the system (Movchan, Guenneau, 2004; Wu *et al.*, 2008). Another type of SC using concentrically placed Helmholtz resonators has been studied numerically (Elford *et al.*, 2011). It was found that the natural resonance properties of the six shell Matryoshka SC give rise to multiple independent resonance band gaps below the first Bragg band gap (i.e., due to the periodicity of the SC) between 400 and 1600 Hz. There have also been numerical studies of acoustical performance of a periodic array of resonant silicone rubber scatterers embedded in an epoxy resin matrix (Hirsekorn *et al.*, 2004). Experimental work in respect to periodic array of scatterers in air has investigated the use of pressurized gas-filled cylindrical balloons (Fuster-Garcia *et al.*, 2007). It was found that resonance attenuation peaks can be obtained at frequencies which do not depend on the periodicity of the SC but on the resonance frequency of the resonators. Predictions and measurements of sound transmission through a periodic array of elastic shells in air have been conducted (Kyrnkin *et al.*, 2010). Besides, several shapes of local resonant SCs have been proposed for enhancing the sound insulation through the multi-layer structures in the entire audible frequency range as well as in the low frequencies (Cavalieri *et al.*, 2019; Gulia, Gupta, 2018; 2019; Kim, 2019; Kim *et al.*, 2021; Qian, 2018).

Plane wave expansion (PWE) method is one of the most studied methods for research into the phononic or sonic crystal band gaps (Chen, Ye, 2001; Vasseur *et al.*, 2008). It can be applied to the infinite arrays of any scatterer shape. Multiple scattering approach was first published for the potential flow through a periodic rectangular array of identical cylinders in 1892. Subsequently, investigations on multiple scattering have been made for a variety of 2D or 3D problems (Linton, Evans, 1990).

Those two methods are not always feasible to solve for scatterers with non-geometrical shapes or resonant constructions by means of an exact analytical solution. In such a case, the finite element method (FEM) based on numerical solutions of partial differential equations (PDEs) offers a method for finding approximate numerical solutions of the scattered and total fields in a wide range of physical and engineering problems. The numerical results for the acoustical properties of SCs have been reported using the FEM commercial software such as COMSOL or ANSYS.

In this paper a DPS is proposed with a local resonant SC based upon the combination of the breathing resonance of a thin ring-type shell and Helmholtz resonance, and some issues for its application are also explored.

2. Numerical modelling

Figure 1 shows the schematic of DPS with SC as well as the Brillouin zone, which depicts the ΓX and ΓM directions of acoustic wave propagation. The SC consists of $7 \times n$ square arrangement of scatterers. Two panels of DPS have thickness of 4 mm and 6 mm, respectively, and are assumed to be made of plastic of which density, Young's modulus, and Poisson's ratio are 1380 kg/m^3, 3.2 GPa, and 0.37, respectively. The space between two panels changes with the number of scatterer rows. The domain between two panels is filled with air or porous material, glass wool. Density and sound speed of air are 1.2 kg/m^3 and 344 m/s, respectively.

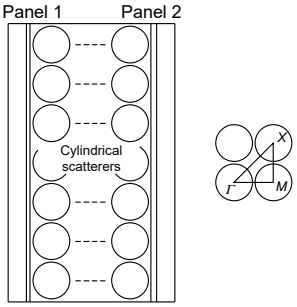


Fig. 1. DPS with $7 \times n$ square arrangement of scatterers and Brillouin Zone.

A plane wave source is applied to the entrance of DPS, i.e., the inward normal velocity is applied to the left boundary surface of the DPS in Fig. 1, while the infinite radiation conditions to the source and receiver boundaries. It can be assumed that there exists no back reflection of acoustic waves on such boundaries, and thus the following boundary conditions are satisfied (Gulia, Gupta, 2018b):

$$\left(-\frac{\nabla p}{\rho}\right) \cdot \mathbf{n} = \frac{i\omega}{\rho c_c} p - \frac{i\omega}{\rho c_c} p_0, \tag{1}$$

$$\left(-\frac{\nabla p}{\rho}\right) \cdot \mathbf{n} = \frac{i\omega}{\rho c_c} p, \tag{2}$$

where $p = p_0 e^{i\mathbf{kr}}$ and p is the acoustic pressure, $\mathbf{k}$ is the wave vector, ρ is the density, $\mathbf{n}$ is the normal vector,

ω is the angular frequency, and c_c is the sound speed. The surfaces of scatterers are assumed to be free to take account of the vibration of thin shells. The other boundaries of DPS are all considered to be solid wall. The normal particle velocity is equal to zero on the solid wall, where the Neumann boundary condition is satisfied as follows:

$$\left(-\frac{\nabla p}{\rho}\right)\cdot\mathbf{n}=0. \tag{3}$$

In this work, glass wool is taken as porous material and modelled by Delay-Bazley model (Delany, Bazley, 1970). The wave number k_g and the characteristic impedance z_g of glass wool are then expressed by the complex forms:

$$k_g = \frac{\omega}{c_0}\left[1+0.0978\chi^{-0.7}-j0.189\chi^{-0.595}\right], \tag{4}$$

$$z_g = \rho_0 c_0\left[1+0.0571\chi^{-0.754}-j0.087\chi^{-0.732}\right], \tag{5}$$

where $\chi=\rho_0 f/R_g$, ρ_0 and c_0 are the density of fluid and the sound velocity in fluid without porous material, respectively, f denotes the frequency, ω indicates the angular frequency, and R_g marks the flow resistance of glass wool.

The mean fiber diameter (d_g) and density (ρ_g) are 10 μm and 12 kg/m^3, respectively. The flow resistance can be calculated by the expression as following (Bies, Hansen, 1980):

$$R_g=\frac{3.18\cdot 10^{-9}\rho_g^{1.53}}{d_g^2}. \tag{6}$$

SC is a periodic array of the finite scatterers embedded in homogeneous medium and can be characterized by the geometry of its primitive cell as in semiconductor. The concept of wave reflection in Bragg's Law which refers mainly to light diffraction can be analogously used for sound waves. In the periodic structures comprised of rigid scatterers, the sound waves could interfere with each other constructively and destructively, resulting in a total reflection regime (complete band gap) in certain frequency range. The frequency range depends on the lattice constant of the array as given by Bragg's Law. In a square lattice, the fundamental Bragg resonance frequencies in the ΓX and ΓM directions are, therefore, as follows:

$$f_{\Gamma X} = \frac{c}{2L}, \tag{7}$$

$$f_{\Gamma M} = \frac{c}{\sqrt{2}\,(2L)}, \tag{8}$$

where c is the sound speed in host medium. If the two Bragg resonances are wide enough in frequency to overlap, then a complete band gap can be realized based solely on the Bragg resonance condition. The width and depth of Bragg resonance is dependent on the acoustic impedance mismatch between the host medium and scatterer as well as the filling fraction (function of lattice constant). Forming a sonic band gap requires the careful selection of materials with both the mass densities and modulus to yield the desired acoustic impedance, and velocity mismatch between the matrix and scatterer.

The resonant frequency of SRR is given by Chalmers *et al.* (2009):

$$f=\frac{c}{2\pi}\sqrt{\frac{\sigma}{S\left\{L+\frac{1}{2}\sqrt{\pi\sigma}\right\}}}, \tag{9}$$

where L is the neck length, σ is the slit width, and S is the inner cross-sectional area. It can be seen that the Helmholtz resonance frequency decreases with increasing the volume of the air cavity and the length of the neck while increases with increasing the slit width.

Calculations are conducted in the FE software ANSYS Multiphysics (v18.0) acoustic package and the soundproofing performances of DPSs are evaluated through the sound transmission loss (STL) and insertion loss (IL), assuming that the flanking sound transmission is negligibly small.

If the power incident on the DPS (left side in Fig. 1) is W_i and the power transmitted through the DPS (right side in Fig. 1) is W_t, STL is determined by:

$$\mathrm{STL}=10\log\left(\frac{W_i}{W_t}\right). \tag{10}$$

IL is calculated from the following equation:

$$\mathrm{IL}=10\log\left(\frac{W}{W'}\right), \tag{11}$$

where W and W' refer to the power transmitted through the DPS without or with the scatterer array.

Harmonic analysis is performed in the range of 20–8000 Hz with the frequency interval of 20 Hz. The model is meshed with an element size (about 4 mm) of one-tenth of the minimum wavelength.

3. Results and discussion

3.1. Number of scatterer rows and filling fraction in DPS

The filling fraction is defined as the ratio between the volume occupied by one scatterer and the volume occupied by a unit cell and thus can be determined once the lattice constant and the size of the scatterers are known. Properly specifying the number of scatterer rows and the filling fraction is very important for decreasing the geometrical size and cost of DPS. For this purpose, the Bragg band gap was investigated with varying such two parameters by the finite element method.

The shapes of scatterers are cylindrical and they are arranged on $7 \times n$ square lattice as in Fig. 1. The cylinders have an outer diameter of 40 mm and the lattice constant of SC is 42.8 mm. The number of rows, n, is varied from 1 to 3. The domains between two panels and scatterers are considered to be filled with air.

Figure 2 shows the IL spectra as a function of the number of scatterer rows. For all three cases, there is a peak of IL around 4020–4040 Hz, which matches well with the Bragg frequency calculated by Eq. (7) for the lattice constant of 42.8 mm. Bragg band gap cannot be created for one row of scatterers in free space. However, DPS with one row of scatterers gives a distinct band gap around the Bragg frequency due to the multiple reciprocating reflection between two panels. The ILs at the Bragg frequency are 49.7 dB, 56.9 dB, and 64.8 dB for the number of rows from 1 to 3, respectively. For the number of 2 and 3, several sharp peaks appear in the insertion loss spectrum, which might be attributed to the suppression of stand wave resonance by the inserted array of scatterers.

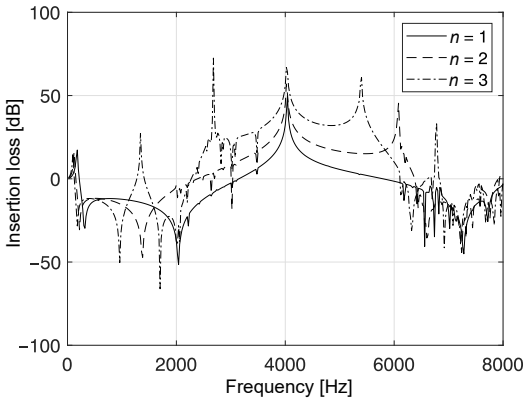


Fig. 2. IL as a function of number of scatterer rows in DPS.

The widths of frequency range where the IL is larger than zero are 2660 Hz, 3980 Hz, and 3820 Hz. And the widths for number of 2 and 3 are about the same.

Figure 3 shows the STL through the DPS for the number of scatterer rows of 2. The sound insulation, added the effect of two panels, exhibits a high transmission loss (TL) of larger than 80 dB in a broad frequency range from 2360 Hz to 6540 Hz, even though the number of scatterer rows is 2. It is therefore obvious that two rows of scatterers are enough to give a sufficiently wide and high band gap around the Bragg frequency in DPS.

Next, the maximum TL and width of the Bragg-type band gap are evaluated with varying the filling fraction from 10% to 70% through the outer radius of the 2 rows of cylindrical scatterers (Figs. 4 and 5).

For the filling fraction of larger than 10%, the peak of the IL increases slowly giving a relatively small

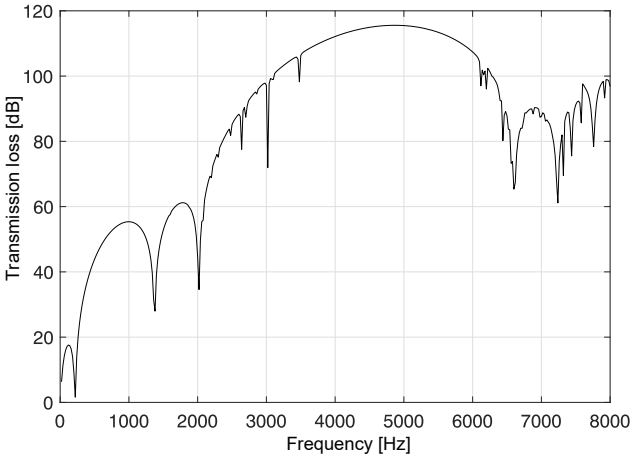


Fig. 3. STL of DPS with 7×2 cylindrical scatterers arranged on square lattice.

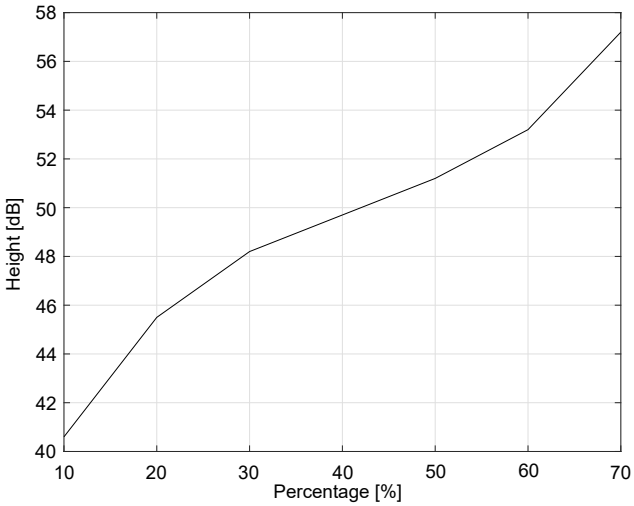


Fig. 4. Peak of IL as a function of filling fraction.

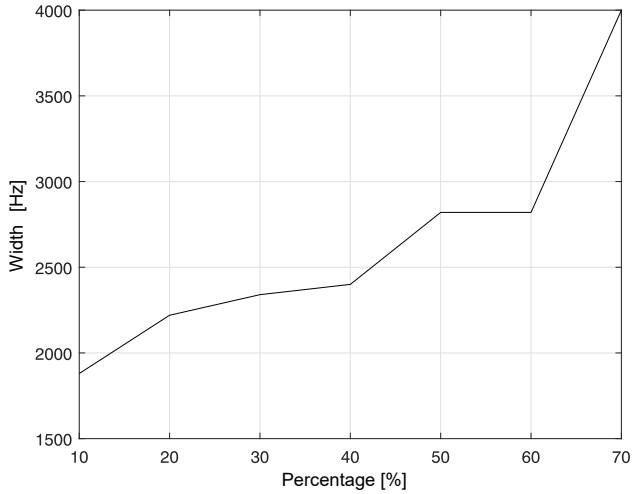


Fig. 5. Width of insertion loss peak as a function of filling fraction.

steepness. The variation of the peak is about 10–14 dB in the entire range of filling fraction from 10% to 70%.

Besides, it can be found from Fig. 5 that the width of frequency range where the IL is larger than zero does not change remarkably for the filling fraction from 10% to 60%. However, it increases rapidly for the filling fraction larger than 60%. As a result of this, it can be concluded that the effect of the filling fraction on the Bragg-type band gap is relatively weak.

3.2. Effects of local resonances

Since an array of elastic shells exposed to the environment would not be practical for an outdoor noise barrier, protecting the shells using concentric outer PVC cylinders offers a simple solution. However, for those inserted between two panels, there is no such problem.

The ring-type shells are considered to be made of commercially available non-vulcanized rubber, of which the material properties are the density of ρ = 1100 kg/m^3, the Young's modulus of E = 1.75 MPa, and the Poisson's ratio of ν = 0.4997. The outer diameter and thickness of the ring shells are 40 mm and 0.5 mm, respectively. Spacing between neighboring scatterers is 42.8 mm and they are arranged on 7×3 square lattice.

Figure 6 shows STL through the DPS with three rows of the thin ring shells.

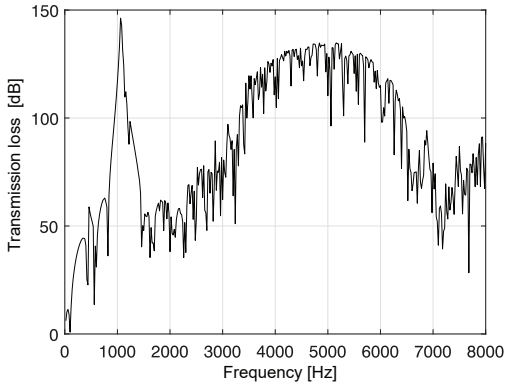


Fig. 6. STL through DPS with a ring-type scatterer array.

The curve has a very high band gap around the low frequency, 1060 Hz, as well as a Bragg-type band gap caused by the periodicity of the lattice, which is attributed to the local resonance of the ring-type scatterers. Meanwhile, TLs in the Bragg band gap fluctuate constantly. This might be due to the variation of the outer diameter caused by the breathing vibration of the scatterers. Fluctuating the outer diameter of the scatterers follows by one of the filling fraction, resulting in the following variation of the sound performance of the SCs.

Figure 7 shows the IL spectra for the solid cylinder array and the ring-type scatterers. As shown in the figure, unlike the case of cylindrical scatterers, there exists a band gap of very large ILs (78 dB) at low frequencies (center frequency of 1020 Hz, width of 600 Hz) attributed to the local breathing-type resonance for the ring scatterers.

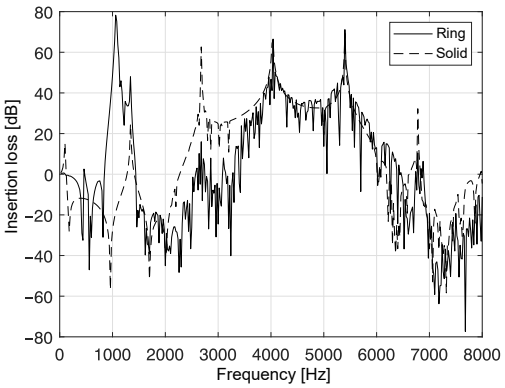


Fig. 7. IL spectra for solid cylinder arrays and ring-type scatterers.

Figure 8 shows the IL for SC consisted of the cylindrical scatterers compared to one for the Helmholtz resonant scatterers. Helmholtz resonators have the outer diameter of 35 mm, the thickness of 1.5 mm, and the slit width of 14 mm. They are assumed to be made of the same plastic as two panels.

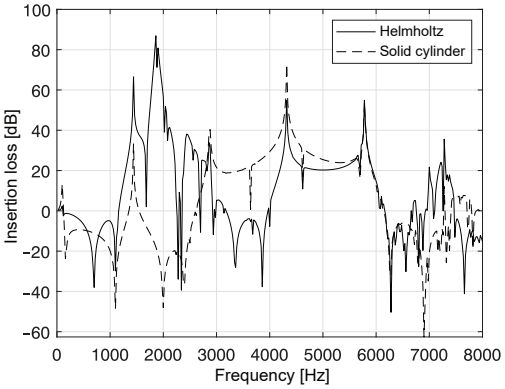


Fig. 8. ILs for solid cylindrical scatterers versus Helmholtz resonant scatterers.

Two ILs at the Bragg frequency are approximately comparable with each other. However, Helmholtz resonator array gives the IL higher than that for the cylindrical scatterer array at the stand wave resonance frequencies as in 1400 Hz. It is also the same at the local resonance frequencies.

3.3. Combination effect of breathing resonance and Helmholtz resonance

Coupling the local resonances in the individual scatterers enables the sound insulation in a certain frequency range to be enhanced. From Eq. (9), it is necessary to increase the thickness or the diameter of the scatterers in order to decrease the resonance frequency of SRRs. This results in the thicker and more weighted construction for soundproofing.

On the other hand, as shown in Figs. 6 and 7, the thin ring scatterer exhibits a very high local resonant band gap. And because the ring scatterers are very thin and made of the materials with a relatively small density, they are lighter than the Helmholtz resonators with the same outer diameter. For example, the mass per length of a Helmholtz resonator used above is about 250 g, while one of a ring scatterer is 68 g, lighter by a factor of about 3.6. If the Helmholtz resonators were made of metallic material rather than plastic, the mass difference would become larger. Thus, coupling the two kind of resonators would not only create the several local resonant band gaps at low frequencies, but also reduce markedly the total weight of the soundproofing structure.

Figure 9 shows the schematic and STL of a DPS with one row of Helmholtz resonators, of which the outer diameter is 35 mm, the thickness 1.5 mm, and the slit width 14 mm, and two rows of ring scatterers, of which the outer diameter is 35 mm and the thickness 0.5 mm. The STL for the DPS filled with air between two panels is also plotted in the figure.

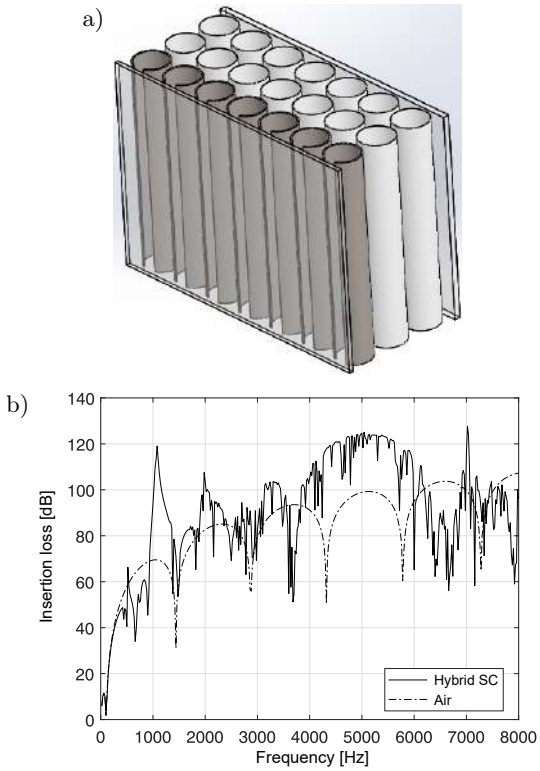


Fig. 9. Schematic (a) and STL (b) of DPS with 7×3 hybrid SC and filled with air.

In the TL curve, there exist very high peaks at the frequencies of 1080, 1440, 1980, and 2860 Hz below the Bragg-type band gap. The peaks at 1080 Hz and 1980 Hz are attributed to the breathing-type resonance of ring scatterers and Helmholtz resonance of SRR, while the other peaks are caused by the suppressions of stand wave resonances. These two types of resonances improve the sound insulation in the frequency range from 1000 Hz to 2300 Hz by an average of 13 dB compared to that without the hybrid SC, with a maximum of 50 dB.

Figure 10 shows the STLs for one row versus two rows of ring scatterers (solid line in Fig. 9b).

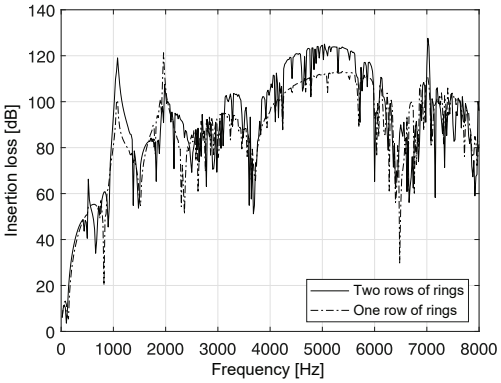


Fig. 10. STLs through DPSs with 7×2 and 7×3 hybrid SCs.

It can be seen from the figure that when the total number of scatterer rows decreases from three to two, the maximum TL of the Bragg band gap was reduced by 8 dB, and the maximum TL of the resonant band gap at the breathing resonance frequency was also reduced by a few dB due to a reduction in the number of ring resonator rows. However, the sound insulation performances in the other frequency ranges are almost similar, and the position and width of the Bragg band gap are also unchanged. In addition, even when the number of ring scatterer rows is 1, the maximum TLs of both the Bragg band gap and the resonant band gap at the breathing resonance frequency are higher than 100 dB. At the same time, the total thickness of DPS decreased by 40 mm from 130 to 90 mm as the scatterer decreased by one.

It can be concluded that combining the two resonances can improve the low-frequency sound insulation, reduce the mass and volume of the entire structure as well as SC.

3.4. Combination effect of porous material and hybrid local resonant SC

Figures 11 and 12 show the effect of porous material (glass wool) on the sound insulation of DPS with the hybrid SC as in Fig. 9 through the STL and IL. As shown in Fig. 11, inserting the porous material leads to the elimination of the deep dips in the STL spectrum due to the two kinds of local resonant scatterer arrays, a significant increase in the soundproofing performance in the overall frequency range, and a relatively smooth TL over 60 dB in the frequency range above 1 kHz. Figure 11 also shows TL through the DPS filled only with glass wool (without hybrid SC)

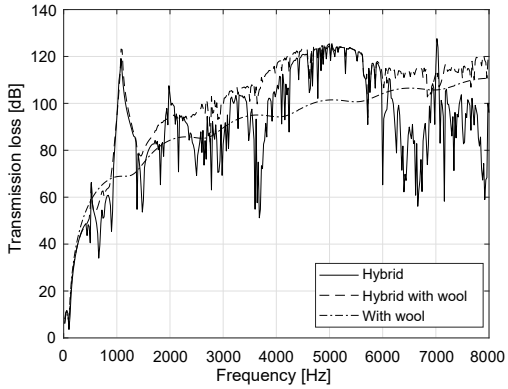


Fig. 11. STLs of DPSs with hybrid SC embedded in porous material or without it.

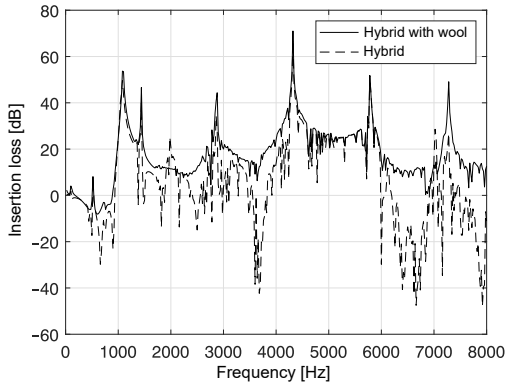


Fig. 12. ILs in DPSs with hybrid SC embedded in porous material and without it.

(dashed dotted line). It can be seen that the sound insulation is improved by more than 15 dB on average in the frequency range above 1 kHz compared with DPS only with glass wool.

These tendencies can be seen in Fig. 12, and by filling with the glass wool, the insertion loss at almost all frequencies is greater than zero. Therefore, in order to enhance sound insulation of DPS only in a particular low frequency range, a hybrid local resonator array with appropriate resonant frequencies should be inserted in two panels of the DPS, while in the case of high insulation performance in the whole frequency range, SCs and porous materials must be combined together.

4. Conclusions

In this paper, a new approach was proposed to reduce the size and mass of DPS while improving the acoustic insulation performance by using the SC coupled with two kinds of local resonance effects. It is found that the minimum number of scatterer rows to form a broad and high forbidden band at the Bragg frequency is two and the filling fraction has relatively little effect on the Bragg band gap in the DPS. In addition, combining the resonances of Helmholtz resonators with the breathing resonances of ring scatterers can not only reduce the thickness and mass of DPS, but also create several resonant band gaps at low frequencies. Also, when the local resonant SCs are applied to DPS, the use of porous materials together can lead to a relatively smooth and high forbidden band around the resonance frequencies. It also provides about 15 dB higher sound insulation than DPS only with glass wool in the frequency range above 1 kHz.

Acknowledgments

The authors wish to thank the anonymous reviewers for their valuable comments, constructive remarks, and suggestions to improve the quality of the paper.

References

1. Bies D.A., Hansen C.H. (1980), Flow resistance information for acoustical design, *Applied Acoustics*, **13**(5): 357–391, doi: 10.1016/0003-682X(80)90002-X.
2. Brillouin L. (1946), *Wave Propagation in Periodic Structures*, 2nd ed., Dover Publications Inc.
3. Cavalieri T., Cebrecos A., Groby J.-P., Chaufour C., Romero-García V. (2019), Three-dimensional multiresonant lossy sonic crystal for broadband acoustic attenuation: Application to train noise reduction, *Applied Acoustics*, **146**: 1–8, doi: 10.1016/j.apacoust.2018.10.020.
4. Chalmers L., Elford D.P., Kusmartsev F.V., Swallowe G.M. (2009), Acoustic band gap formation in two-dimensional locally resonant sonic crystals comprised of Helmholtz resonators, *International Journal of Modern Physics B*, **23**(20n21): 4234–4243, doi: 10.1142/S0217979209063390.
5. Chen Y.-Y., Ye Z. (2001), Theoretical analysis of acoustic bands in two-dimensional periodic arrays, *Physical Review E*, **64**: 036616, doi: 10.1103/PhysRevE.64.036616.
6. Delany M.E., Bazley E.N. (1970), Acoustical properties of fibrous absorbent materials, *Applied Acoustics*, **3**: 105–116, doi: 10.1016/0003-682X(70)90031-9.
7. Elford D.P., Chalmers L., Kusmartsev F.V., Swallowe G.M. (2011), Matryoshka locally resonant sonic crystal, *The Journal of the Acoustical Society of America*, **130**(5): 2746–2755, doi: 10.1121/1.3643818.
8. Fuster-Garcia E., Romero-García V., Sánchez-Pérez J.V., García-Raffi L.M. (2007), Targeted band gap creation using mixed sonic crystal arrays including resonators and rigid scatterers, *Applied Physics Letter*, **90**(24): 244104, doi: 10.1063/1.2748853.
9. Gulia P., Gupta A. (2018), Enhancing the sound transmission loss through acoustic double panel using sonic crystal and porous material, *The Journal of the Acoustical Society of America*, **144**(3): 1435–1442, doi: 10.1121/1.5054296.

10. Gulia P., Gupta A. (2019), Sound attenuation in triple panel using locally resonant sonic crystal and porous material, *Applied Acoustics*, **156**: 113–119, doi: 10.1016/j.apacoust.2019.07.012.

11. Hirsekorn M., Delsanto P.P., Batra N.K., Matic P. (2004), Modelling and simulation of acoustic wave propagation in locally resonant sonic materials, *Ultrasonics*, **42**(1–9): 231–235, doi: 10.1016/j.ultras.2004.01.014.

12. Kim M.-J. (2019), Numerical study for increasement of low frequency sound insulation of double-panel structure using sonic crystals with distributed Helmholtz resonators, *International Journal of Modern Physics B*, **33**(14): 1950138, doi: 10.1142/S0217979219501388.

13. Kim M.-J., Rim C.-G., Won K.-S. (2021), Broadening low-frequency band gap of double-panel structure using locally resonant sonic crystal comprised of slot-type Helmholtz resonators, *Archives of Acoustics*, **46**(2): 335–340, doi: 10.24425/aoa.2021.136587.

14. Kyrnkin A., Umnova O., Chong Y.B.A., Taherzadeh S., Attenborough K. (2010), Predictions and measurements of sound transmission through a periodic array of elastic shells in air, *The Journal of the Acoustical Society of America*, **128**(6): 3496–3506, doi: 10.1121/1.3506342.

15. Linton C.M., Evans D.V. (1990), The interaction of waves with arrays of vertical circular cylinders, *Journal of Fluid Mechanics*, **215**: 549–569, doi: 10.1017/S0022112090002750.

16. Martínez-Sala R., Rubio C., Garcia-Raffi L.M., Sánchez-Pérez J.V., Sánchez-Pérez E.A., Llinares J. (2006), Control of noise by trees arranged like sonic crystals, *Journal of Sound and Vibration*, **291**(1–2): 100–106, doi: 10.1016/j.jsv.2005.05.030.

17. Martínez-Sala R., Sancho J., Sánchez J.V., Gómez V., Llinares J., Meseguer F. (1995), Sound attenuation by sculpture, *Nature*, **378**: 241, doi: 10.1038/378241a0.

18. Movchan A.B., Guenneau S. (2004), Split-ring resonators and localized modes, *Physical Review B*, **70**: 125116, doi: 10.1103/PhysRevB.70.125116.

19. Qian D. (2018), Wave propagation in a LRPC composite double panel structure with periodically attached pillars and etched holes, *Archives of Acoustics*, **43**(4): 717–725, doi: 10.24425/aoa.2018.125165.

20. Romero-García V., Garcia-Raffi L.M., Sánchez-Perez J.V. (2011), Evanescent waves and deaf bands in sonic crystals, *American Institute of Physics Advances*, **1**: 041601, doi: 10.1063/1.3675801.

21. Sainidou R., Djafari-Rouhani B., Pennec Y., Vasseur J.O. (2006), Locally resonant phononic crystals made of hollow spheres or cylinders, *Physical Review B*, **73**(2): 024302, doi: 10.1103/PhysRevB.73.024302.

22. Sánchez-Dehesa J., Garcia-Chocano V.M., Torrent D., Cervera F., Cabrera S. (2011), Noise control by sonic crystal barriers made of recycled material, *The Journal of the Acoustical Society of America*, **129**(3): 1173–1183, doi: 10.1121/1.3531815.

23. Sanchez-Perez J.V., Rubio C., Martinez-Sala R., Sanchez-Grandia R., Gomez V. (2002), Acoustic barriers based on periodic arrays of scatterers, *Applied Physics Letters*, **81**: 5240–5242, doi: 10.1063/1.1533112.

24. Sohn C.H., Park J.H. (2011), A comparative study on acoustic damping induced by half-wave, quarter-wave and Helmholtz resonators, *Aerospace Science and Technology*, **15**(8): 606–614, doi: 10.1016/j.ast.2010.12.004.

25. Umnova O., Attenborough K., Linton C.M. (2006), Effects of porous covering on sound attenuation by periodic arrays of cylinders, *The Journal of the Acoustical Society of America*, **119**(1): 278–284, doi: 10.1121/1.2133715.

26. Vasseur J.O., Deymier P.A., Djafari-Rouhani B., Pennec Y., Hladky-Hennion A-C. (2008), Absolute forbidden bands and waveguiding in two-dimensional phononic crystal plates, *Physical Review B*, **77**(8): 085415, doi: 10.1103/PhysRevB.77.085415.

27. Wu L.-Y., Chen L.-W., Wu M.-L. (2008), The nondiffractive wave propagation in the sonic crystal consisting of rectangular rods with a slit, *Journal of Physics: Condensed Matter*, **20**: 295229, doi: 10.1088/0953-8984/20/29/295229.

Archives of Acoustics Vol. 48, No. 1, pp. 71–79 (2023), doi: 10.24425/aoa.2023.144262

Research Paper

Series Expanding of the Ultrasound Transmission Coefficient Through a Multilayered Structure

Ahmed CHITNALAH[(1)], Noureddine AOUZALE[(1)], Hicham JAKJOUD[(2)*]

[(1)] *Electrical Systems Energetic Efficiency and Telecommunications Laboratory, Cadi Ayyad University* Marrakesh, Morocco

[(2)] *Energy Engineering Materials and Systems Laboratory, Ibn Zohr University* Agadir, Morocco

*Corresponding Author e-mail: hicham.jakjoud@gmail.com

(*received March 1, 2022; accepted August 31, 2022*)

To calculate the transmission coefficient of ultrasonic waves through a multi-layered medium, a new approach is proposed by expanding it into Debye's series. Using this formalism, the transmission coefficient can be put in the form of resonance terms series. From this point of view, the relative amplitude of the transmitted wave can be considered as an infinite summation of terms taking into account all possible reflections and refractions on each interface. Our model is then used to investigate interaction between the ultrasonic plane wave and the N-plane-layer structure.

Obviously, the resulting infinite summation has to be reduced to a finite one, according to some level of accuracy. The numerical estimation of the transmission coefficient using the exact expression (Eq. (1)) is then compared to the one of our method in the case of two or three plane-layer structure. The effect of the order of the finite summation on the calculated value of the transmission coefficient is, as well, studied. Finally, our proposed method may be used, with the decomposition into Gaussian beams of a pressure field created by a circular source, to draw a 3D image of the pressure field transmitted through a multilayered structure.

Keywords: multilayered structure; Debye's series; resonance formalism; ultrasonic NDT.

Nomenclature

Γ_N – transmission coefficient of N-layer structure,
d_n – thickness of the n-th layer,
$\mathbf{k}_n$ – propagation wave vector in the n-th layer,
ρ_n – mass density of the n-th layer,
c_n – sound speed in the n-th layer,
α_n – absorption coefficient in the n-th layer,
t_{n-1} – transmission coefficient of the n-th interface,
r_n – reflection coefficient of the n-th interface,
n_i – summation degree in i-layer structure,
ε – precision needed to define the minimal value of n_i,
p_n – acoustic pressure in the n-th layer,
Z_n – acoustic impedance of the n-th layer.

1. Introduction

The measurement of ultrasound reflection and transmission coefficients, from and through a layered structure, is of great interest in many nondestructive testing and characterization applications. Biological tissues or rocks are some natural example of the layered structure. Experimental data with a theoretical model are exploited to extract acoustical properties as the attenuation coefficient, density, sound speed and other mechanical properties. Some of these applications are reviewed in (Hsu, 2009).

Several techniques can be used to solve the inverse problem. The inverse problem solution can be obtained by minimizing a cost functional formulated as the least square error between the waveform calculated using an equivalent model, and the measured waveform obtained from ultrasonic transmission tests (Messineo *et al.*, 2016). A particle swarm optimization (PSO) algorithm based least squares estimation and using the ultrasonic reflection spectrum has been

used (Yang *et al.*, 2019). Multilayer structures are involved in the design of piezoelectric transducers. Implementing two-layer matching structure improves the transmission of the acoustic power into the medium (Bakhtiari-Nejad *et al.*, 2020). In the case of several layers, genetic algorithms can help in an optimal selection of the materials used as adaptation layers (Gudra, Banasiak, 2020).

Ultrasound propagation through periodic structures is another area of interest of the layered structure. Indeed, periodic structures with a wavelength scale periodicity (Potel, Belleval, 1993; Shenand, Cao, 2000; Khaled *et al.*, 2013; Maréchal *et al.*, 2014) are known to exhibit acoustical band gaps, which is of great interest for many applications like wave filtering, guiding, focusing waves, silent blocks, and it can also help improving the efficiency of transducers (Maréchal *et al.*, 2008).

One way to study these structures is to calculate the transmission and reflection coefficients of the structure immersed in a fluid such as air or water. The problem can be resolved numerically. However, analytic solutions are still a great way to understand the physical mechanisms involved. Different approaches and techniques like the plane wave method (PWM) (Potel, Belleval, 1993; Deschamps, Chengwei, 1991) and the transfer matrix method (TMM) (Solyanik, 1977; Rokhlin, Wang, 1992; Haskell, 1953; Folds, Loggins, 1977; Lowe, 1995) or the global matrix method (Storheim *et al.*, 2015) are often adopted. The iterative method (Scott, Gordon, 1977) or the equivalent impedance (Messineo *et al.*, 2013) can also be used.

Debye's series decomposition method (Maréchal *et al.*, 2014; Gérard, 2022) allows developing the reflection and transmission coefficients into a sum of multiple reflection terms. Typically, some resonance terms are ignored. However, expressing the transmission coefficient into a sum of these terms is very useful for resonance analysis.

Fiorito and Überall (1979) showed that the acoustic transmission and the reflection coefficient of a fluid layer embedded in another fluid can be written in the form of a sum of resonance terms. The resonance theory of a fluid layer has been extended to include viscous effects (Fiorito *et al.*, 1981). Three layered elastic medium have been investigated (Ainslie, 1995) using ray path analyses. The solution of the reflection/refraction of a plane wave at a single solid layer has been expanded into Debye's series (Deschamps, Chengwei, 1991). Using a matrix notation and a generalized Debye theory, Gérard *et al.* (1979; 1980; 1982; 1987) derived an exact solution in an elastic multilayered sphere. The case of submerged cylinders (Derem, 1982) and plates (Conoir, 1991; Derible, Tinel, 2011) were studied too. Earlier we exploited this resonance formalism to study the interaction of a bounded ultrasonic beam with an immersed plate (Soucrati *et al.*, 2018). A global transfer matrix has been constructed to study the interaction of harmonic elastic waves with n-layered anisotropic medium (Nayfeh, 1991).

In the present work, firstly, we determined the transmission coefficient (Γ_N) through N-layered structure using the plane wave theory. The details of the calculus are given in Appendix A, the solution named an exact solution is then given in Eq. (1). This exact solution is broken down into series translating the individual contribution of each resonance. A novel way is applied to write this solution as a product of the Debye series. Then a new resonance model for the transmission (Γ_N) was developed. The model provides analytical expressions for the characteristics of each resonance. This facilitates the resonance decomposition of the transmitted wave and help understanding resonance phenomena. Further, the model provides a useful tool to solve the inverse problem. In the same manner, our model can be applied to the calculation of the reflection coefficient.

In this paper, we describe firstly the problem of the propagation of ultrasonic waves in multilayered media as well as the theoretical formalism that governs this propagation. Applying the pressure continuity and the particle velocity continuity at the two interfaces of each layer we derived the exact formulation of the transmission coefficient named Γ_N. Details are given in Appendix A. The result is given in Eq. (1). Then Γ_N is expanded into a sum of resonance terms like Debye's series. Details about the method used to expand the transmission coefficient into Debye's series are given in Appendix B. The expanded expressions are given in the Eqs. (10), (15), (18), and (22), respectively, for one, two, three, and N-layers. Finally, numerical evaluation of the transmission coefficient Γ_N given by the exact solution (Eq. (1)) and the expanded formulation are compared. The comparison shows good agreement subject to choosing the right number of resonance terms.

2. Theoretical formalism

2.1. Studied configuration

We consider a layered structure hit by an ultrasonic plane wave in normal incidence (Fig. 1). The structure to be analyzed is composed of N-layers indexed from 1 to N. Each layer is of the thickness d_n. Note

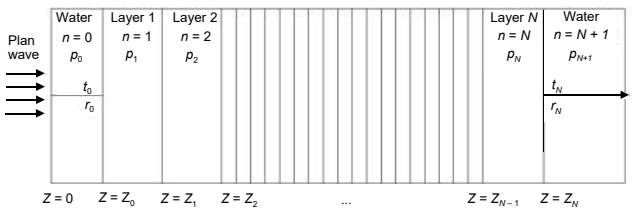


Fig. 1. Geometrical arrangement.

that the propagating vector is $\mathbf{k}_n$, the mass density is ρ_n, the layer's sound celerity is c_n, the acoustic impedance Z_n, and the attenuation coefficient α_n.

The structure is immersed in water characterized by its density ρ_0, velocity c_0, acoustic impedance Z_0, and wave number k_0. The surrounding medium is taken to be nonabsorbent, so $\alpha_0 = 0$. The second half medium surrounding the layers is also water and corresponds to the slice indexed $N+1$. Then $Z_{N+1} = Z_0$.

The N-layered structure is hit, in a normal incidence, by a plane harmonic wave. So, only a longitudinal wave is to be considered. The transverse waves are not considered.

Plexiglas is used as a layer with $\rho = 1200$ kg/m^3, $c = 2650$ m/s, and $\alpha = 1.13$ dB/[MHz$\cdot$cm]. For aluminum $\rho = 2800$ kg/m^3, $c = 6380$ m/s.

2.2. Transmission coefficient

We demonstrate (Appendix A) that the transmission coefficient Γ_N of N-layered structure can be expressed as:

$$\Gamma_N = T_N \frac{\varphi_N}{D_N} \tag{1}$$

with:

$$T_N = \prod_{n=0}^{N} t_n; \qquad t_n = \frac{2Z_{n+1}}{Z_n + Z_{n+1}}, \tag{2}$$

$$\varphi_N = \prod_{n=1}^{N} X_n; \qquad X_n = e^{-i\gamma_n d_n}; \qquad \gamma_n = \mathbf{k}_n - i\alpha_n, \tag{3}$$

where T_N is the transmission coefficient through the N-layer structure, while φ_N is the accumulation of the phase induced by the propagation into the different N-layers.

Moreover, t_n correspond to:

- t_0: transmission from water of the first half medium surrounding the structure, to the first layer (layer 1),
- t_N: transmission from the last layer (layer N) to the water of the second half medium surrounding the structure.

The transmission coefficient Γ_N consists of a fraction of two terms, namely the numerator $(T_N\varphi_N)$ that takes into account transmission attenuation and the denominator (D_N) that takes into account multiple reflections at each interface. This latter can be expressed as $D_N = 1 + \Phi_N$, where Φ_N is responsible for reflections/refractions at all the interfaces.

A similar formula has been already given in (Stovas, Arnsten, 2006). Indeed, starting from Eq. (40) in Appendix B, we can deduce:

$$D_N = D_1 + \sum_{m=0}^{N-1} r_{N-m}\, x_{N-m}\, D'_{N-m}. \tag{4}$$

So, one can see that D_N can be written in the form of:

$$D_N = 1 + \Phi_N \tag{5}$$

with:

$$\Phi_N = r_0\, r_1\, x_1 + \sum_{m=0}^{N-1} r_{N-m}\, x_{N-m}\, D'_{N-m}. \tag{6}$$

For $N > 1$, D_N is developed in a new manner to allow decomposition into series terms. Details are given in Appendix B. So D_N can be expressed in a simple way as:

$$\begin{aligned} D_N &= D_{N-1} + r_N x_N \widetilde{D}_{N-1} x_n = X_n^2, \\ r_n &= \frac{Z_{n+1} - Z_n}{Z_n + Z_{n+1}}, \end{aligned} \tag{7}$$

where

$$\widetilde{D}_N = r_N D_{N-1} + x_N \widetilde{D}_{N-1}, \qquad \widetilde{D}_1 = r_1 + r_0 x_1, \tag{8}$$

where r_N is the reflection coefficient at the last interface from water to the layer N. It is expressed as:

$$r_N = \frac{Z_{N+1} - Z_N}{Z_N + Z_{N+1}} = \frac{Z_0 - Z_N}{Z_N + Z_0}. \tag{9}$$

Now, the idea is to expand Γ_N into a sum of resonance terms as Debye's series.

For one layer ($N = 1$), we have:

$$\Gamma_1 = T_1 X_1 \sum_{n_1=0}^{\infty} \left(-r_0\, r_1\, x_1\right)^{n_1}. \tag{10}$$

For more than one layer ($N > 1$), we derive an expression of D_N (Appendix B) that allows expanding Γ_N into Debye's series.

Let us start with two layers $N = 2$:

$$\frac{1}{D_2} = \frac{1}{C_1} \sum_{n_2=0}^{\infty} \left(-\beta_1\, r_2\, x_2\right)^{n_2}. \tag{11}$$

We express β_1 as:

$$\beta_1 = \frac{1}{r_1}\left(1 - \frac{t_1 t'_1}{C_1}\right), \tag{12}$$

thus:

$$\frac{1}{D_2} = \frac{1}{C_1} \sum_{n_2=0}^{\infty} \left(-\frac{r_2}{r_1} x_2\right)^{n_2} \sum_{m=0}^{n_2} \frac{n_2!}{m!\,(n_2-m)!} \left(-\frac{t'_1 t'_1}{C_1}\right)^{m}. \tag{13}$$

The term $(1/C_1)^m$ can be expanded in series as:

$$\frac{1}{C_1^m} = \sum_{n_1=0}^{\infty} \left(-r_0\, r_1\, x_1\right)^{n_1} \frac{(m+n_1)!}{m! n_1!}. \tag{14}$$

Replacing Eq. (10) in (9), we get the final form of the transmission coefficient through two layers:

$$\Gamma_2 = T_2 \varphi_2 \sum_{n_2=0}^{\infty} \left(-r_2\, x_2\right)^{n_2} \sum_{n_1=0}^{\infty} W_{n_2 n_1} \left(-r'_0\, x'_1\right)^{n_1} \tag{15}$$

with:

$$W_{n_2 n_1} = r_1^{n_1-n_2}\frac{n_2!}{n_1!}\sum_{m=0}^{n_2}\frac{(m+n_1)!}{m!m!\,(n_2-m)!}\left(-t_1' t_1'\right)^m. \quad (16)$$

Let us rewrite the two first terms of Γ_2 with respect to n_2, in order to give a physical interpretation of each of them:

$$\Gamma_2 = T_2\varphi_2\left\{\sum_{n_1=0}^{\infty}\left(-r_0\, r_1\, x_1\right)^{n_1}\right.$$

$$\left.+\sum_{n_1=0}^{\infty}\left[1-(n_1+1)(-r_2\, x_2)\frac{t_1 t_1'}{r_1}\right]\left(-r_0\, r_1\, x_1\right)^{n_1}+\ldots\right\}. \quad (17)$$

The first summation in Eq. (13) counts for the resonance into the first layer and transmitted through the second (Fig. 2a). The second summation term is obtained for $n_2 = 1$ and composed of two components: the first sum is the same as for $n_2 = 0$, the second sum counts for the reflection from the interface between layer 2 and the ambient medium followed by resonances into layer 1 (Fig. 2b). The coefficient (n_1+1) in the second sum is due to the Fabry-Pérot like effect, which means that the response is composed of many echoes that arrive at the same time as shown in Fig. 2.

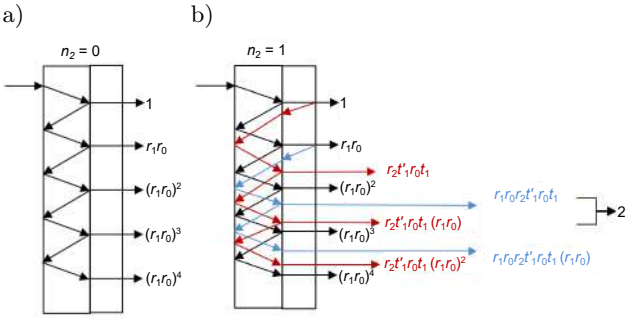


Fig. 2. Schematic of (a) the first layer resonance and transmission ($n_2 = 0$) and (b) the multiple reflections from the second layer and ambient medium interface ($n_2 = 1$).

In the same manner we derive the expression for three adjacent layers:

$$\Gamma_3 = T_3\varphi_3\sum_{n_2=0}^{\infty}\left(-x_2\right)^{n_2}\sum_{n_1=0}^{\infty}W_{n_2 n_1}\left(-r_0' x_1'\right)^{n_1}$$

$$\cdot\sum_{n_3=0}^{\infty}W_{n_2 n_3}\left(-r_3' x_3'\right)^{n_3}, \quad (18)$$

where

$$W_{n_2 n_3} = r_2^{n_3-n_2}\frac{n_2!}{n_3!}\sum_{m=0}^{n_2}\frac{(m+n_3)!}{m!m!\,(n_2-m)!}\left(-t_2' t_2'\right)^m. \quad (19)$$

In the case of four layers, we have:

$$\Gamma_4 = T_4\varphi_4\sum_{n_2=0}^{\infty}\left(-x_2\right)^{n_2}\sum_{n_1=0}^{\infty}W_{n_2 n_1}\left(-r_0' x_1'\right)^{n_1}$$

$$\cdot\sum_{n_3=0}^{\infty}W_{n_2 n_3}\left(-x_3'\right)^{n_3}\sum_{n_4=0}^{\infty}W_{n_3 n_4}\left(-r_4 x_4'\right)^{n_4} \quad (20)$$

with:

$$W_{n_3 n_4} = r_3^{n_4-n_3}\frac{n_3!}{n_4!}\sum_{m=0}^{n_3}\frac{(m+n_4)!}{m!m!\,(n_3-m)!}\left(-t_3' t_3'\right)^m. \quad (21)$$

In the general case of N-layers with $N > 3$:

$$\Gamma_N = T_N\varphi_N\sum_{n_2=0}^{\infty}\left(-x_2\right)^{n_2}\sum_{n_1=0}^{\infty}W_{n_2 n_1}\left(-r_0' x_1'\right)^{n_1}$$

$$\cdot\prod_{j=3}^{N-1}\left[\sum_{n_j=0}^{\infty}W_{n_{j-1} n_j}\left(-x_j'\right)^{n_j}\right]$$

$$\cdot\sum_{n_N=0}^{\infty}W_{n_{N-1} n_N}\left(-r_N x_N'\right)^{n_N}, \quad (22)$$

where

$$W_{n_j n_{j+1}} = r_j^{n_{j+1}-n_j}\frac{n_j!}{n_{j+1}!}\sum_{m=0}^{n_j}\frac{(m+n_{j+1})!}{m!m!\,(n_j-m)!}\left(-t_j' t_j'\right)^m. \quad (23)$$

Γ_N can be interpreted as a sum of waves of the amplitude W_{ij} reflected at each interface.

We derived here a more generalized formula than those based on the amplitude of echoes derived in the reflection mode (Chern, Nielsen, 1989) and through the transmission mode (Chern, Nielsen, 1990).

3. Simulation results

In this section, we show the validity of the proposed model by simulating the transmission coefficient of different configurations of multilayered structures. The cases of one, two, and three layers are studied for different numbers of resonances.

The exact expression of the transmission coefficient Γ_N through an N-layer structure is given by Eq. (1). The series expansion of the resonance term is given by Eqs. (10), (17), and (18), respectively for 1, 2, or 3 layers. In all these expressions we have infinite sums which are truncated during the simulation. The simulation results are then compared to the calculations of Γ_N using Eq. (1) to determine firstly the degree of truncation and secondly to validate our model.

3.1. One-layer case

Figure 3 gives the transmission coefficient versus frequency, for one layer made of 4 mm of plexiglass. The exact solution (red curve) is compared to our model (blue curve). The results in Fig. 3a are given for $n_1 = 1$ which means that the series in Eq. (10) is truncated to the first term. However in Fig. 3b the series is truncated to the 5 first terms. So the maximum number of resonance taken into account should be determined according to the acoustical parameters of the layer. The same results for aluminum plate are given in Fig. 4.

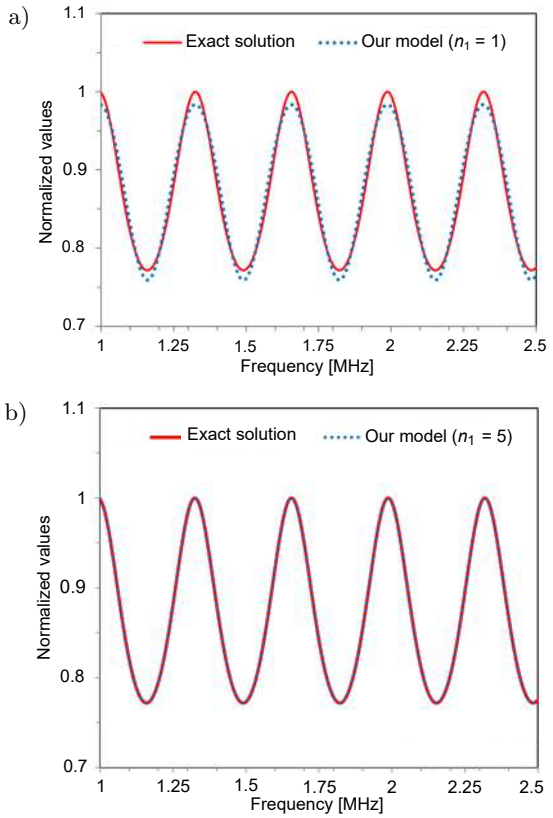


Fig. 3. $|\Gamma_1|$, versus frequency, for 4 mm plexiglas plate: exact solution (red), our model for $n_1 = 1$ (blue in a), and $n_1 = 5$ (blue in b).

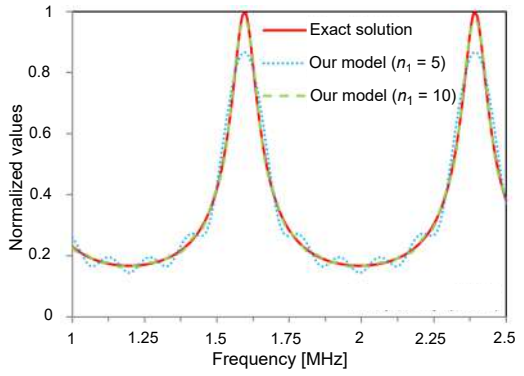


Fig. 4. $|\Gamma_1|$, versus frequency, for 4 mm aluminum plate: exact solution (red), our model for $n_1 = 5$ (blue), and $n_1 = 10$ (green).

As expected, Fig. 3 and Fig. 4 show the well-known resonance frequency due to different modes of propagation in the layer.

Discrepancies are noticed, especially for the minima and maxima of the curves, if the number of resonances taken into account is insufficient (blue curve Figs. 3a and 4). To match the exact solution one need to define a precision ε. Then the minimal value of n_i ($i = 1, 2, ..., N$) should be determined according to this accuracy ε.

The minimal value of n_1 is determined according to the expected accuracy using this expression:

$$|r_0\, r_1|^{n_1} < 10^{-\varepsilon} \quad \text{then} \quad n_1 \geq \frac{-\varepsilon}{\log|r_0\, r_1|}. \tag{24}$$

If we take $\varepsilon = 3$ for example, one should take the summation of the 20 terms for aluminum and only 4 terms for plexiglas.

3.2. Two layers

We studied a structure made of two layers aluminum/polyethylene (Al/PE) immersed in water. Each layer has 4 mm thickness. Transmission coefficient variation according to the frequency is presented in Fig. 5 for the Al/PE structure. This structure has been studied theoretically and experimentally using the reflection response (Lenoir, Maréchal, 2009).

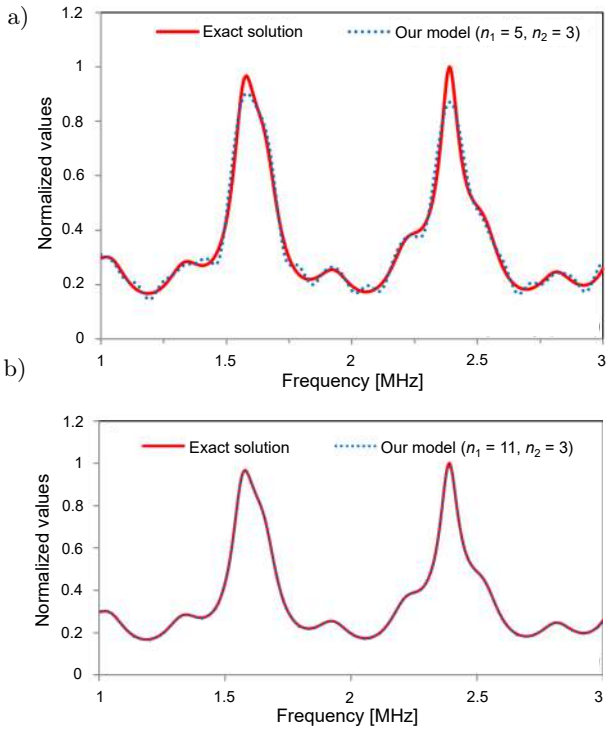


Fig. 5. Variation of $|\Gamma_2|$ versus frequency: exact solution (red), our model with $n_2 = 3$ (blue), $n_1 = 5$ in (a), and $n_1 = 11$ in (b).

The minimal value of n_1 is determined using Eq. (24). Taking $\varepsilon = 10^{-2}$ we deduce $n_1 = 11$. For n_2 we should use according to Eq. (7):

$$|\beta_1\, r_2|^{n_2} < 10^{-\varepsilon} \quad \text{then} \quad n_2 \geq \frac{-\varepsilon}{\log|r_2\, \beta_1|}. \tag{25}$$

If we take $\varepsilon = 2$ it gives $n_1 = 11$ and $n_2 = 5$. Using these minimal values, the approximated solution fits well with the exact one as it is seen in Fig. 5b.

3.3. Three layers

Let us now study the structure made of two plexiglas plates of 5 mm separated by 1 mm of water (Fig. 6).

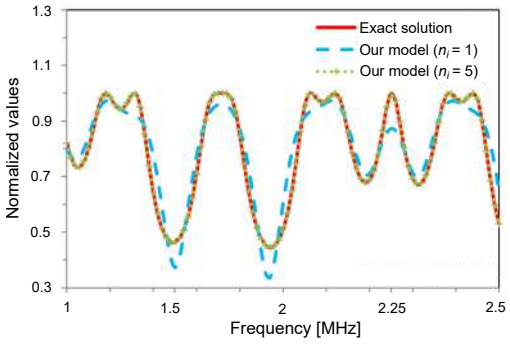


Fig. 6. $|\Gamma_3|$ versus frequency for plexiglas/water/plexiglas of 5/1/5 mm structure: exact solution (red), our model (blue $n_1 = n_2 = n_3 = 1$ and green $n_1 = n_2 = n_3 = 5$).

We can notice from simulation results, that our model based on Debye's series gives very good results for the calculation of the ultrasound transmission coefficient in a multilayered structure. Especially if the number of resonance terms used in calculation is well chosen.

4. Conclusion

In this work, we present a new analytical method for calculating the acoustic frequency response of a multilayered structure. The transmission coefficient calculated using this new method is put into a sum of resonance terms. This way, we can consider the relative amplitude of each wave as a summation of several terms taking into account all possible reflections/refractions.

The expanded solution is in good agreement with the exact solution, subject to take a suitable number of resonances. We have also proposed a method helping defining this resonances' number. Our method gives a more generalized formula than other methods based on reflection or transmission modes. Our work can also help extracting geometrical and acoustical parameters of each layer. The considered layers can be either liquid or solid, since the incidence is normal to avoid the shear waves that are not taken into account by the formula.

Instead of a plane wave, our model can be associated with superposition of a bounded beam to find the 3D diffracted field by a multilayered structure. In the same manner we derived an analytical expression of a 3D ultrasonic field transmitted through a single plane layer (Soucrati *et al.*, 2018).

Appendix A

For, $n = 1$ to N, we express the ultrasonic pressure p_n inside each layer as ($z_0 < z < z_N$):

$$p_n = U_n e^{-i\gamma_n(z-z_{n-1})} + V_n e^{i\gamma_n(z-z_{n-1})}. \tag{26}$$

The pressure wave in the first half of embedded medium (water $n = 0$) is then ($0 < z < z_0$):

$$p_0 = U_0 e^{-i\gamma_0 z} + V_0 e^{i\gamma_0 z}. \tag{27}$$

The transmitted wave in water after the layered structure ($z > z_N$) is written as:

$$p_{Nt} = U_t e^{-i\gamma_0(z-z_N)}. \tag{28}$$

We put η_n as the ration of the impedance of the slice before to that of the next slice:

$$X_n = e^{-i\gamma_n d_n}, \qquad \eta_n = \frac{Z_n}{Z_{n+1}}. \tag{29}$$

So from the water to the first layer we have:

$$X_0 = e^{-i\gamma_0 d_0}, \qquad \eta_0 = \frac{Z_0}{Z_1}, \tag{30}$$

and from the last layer to the water we write:

$$X_N = e^{-i\gamma_N d_N}, \qquad \eta_N = \frac{Z_N}{Z_{N+1}} = \frac{Z_N}{Z_0}. \tag{31}$$

The origin of the propagation axis is placed at the interface between layers n and $n-1$. By applying the pressure continuity (Ingard, Morse, 1968) and particle velocity continuity equations at each interface and solving them, we get the ratio of the sound pressure level of the transmitted wave (U_t) to the incident wave (U_0). This corresponds to the transmission coefficient (Γ_N). In the same way we can deduce the reflection coefficient.

For one layer, we have:

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{pmatrix} U_0 \\ V_0 \end{pmatrix} = \frac{1}{\eta_0}\begin{bmatrix} \eta_0 & \eta_0 \\ 1 & -1 \end{bmatrix}\begin{pmatrix} U_1 \\ V_1 \end{pmatrix}, \tag{32}$$

$$\frac{1}{X_1}\begin{bmatrix} x_1 & 1 \\ x_1 & -1 \end{bmatrix}\begin{pmatrix} U_1 \\ V_1 \end{pmatrix} = U_{1t}\begin{pmatrix} 1 \\ \eta_1 \end{pmatrix}. \tag{33}$$

Solving this system, we get the expression of Γ_1 as:

$$\Gamma_1 = \frac{U_{1t}}{U_0} = \frac{t_0\, t_1 X_1}{1 + r_0\, r_1\, x_1}. \tag{34}$$

For a structure of two layers $N = 2$, we have to add this equations system for $n = 1$:

$$\frac{1}{X_n}\begin{bmatrix} x_n & 1 \\ x_n & -1 \end{bmatrix}\begin{pmatrix} U_n \\ V_n \end{pmatrix} = \frac{1}{\eta_n}\begin{bmatrix} \eta_n & \eta_n \\ 1 & -1 \end{bmatrix}\begin{pmatrix} U_{n+1} \\ V_{n+1} \end{pmatrix}, \tag{35}$$

$$\frac{1}{X_N}\begin{bmatrix} x_N & 1 \\ x_N & -1 \end{bmatrix}\begin{pmatrix} U_N \\ V_N \end{pmatrix} = U_{Nt}\begin{pmatrix} 1 \\ \eta_N \end{pmatrix}. \tag{36}$$

Solving this system, we get Γ_2:

$$\Gamma_2 = \frac{U_{2t}}{U_0} = \frac{t_0\, t_1\, t_2\, X_1\, X_2}{1 + r_0\, r_1\, x_1 + r_1\, r_2\, x_2 + r_0\, r_2\, x_1\, x_2}. \tag{37}$$

For N layers, we can get the expression of the coefficient of transmission using the same technique:

$$\Gamma_N = \frac{U_{Nt}}{U_0} = \frac{T_N \varphi_N}{D_N}, \qquad T_N = t_0 t_1 \ldots t_N, \qquad \varphi_N = X_0 X_1 \ldots X_N. \tag{38}$$

We notice that D_N is expressed as a sum of combination of all the possible product of x_n.

For 3 layers for example, we have x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, and $x_1x_2x_3$. So D_3 is expressed as:

$$D_3 = 1 + c_1x_1 + c_2x_2 + c_3x_3 + c_{12}x_1x_2 + c_{13}x_1x_3 + c_{23}x_2x_3 + c_{123}x_1x_2x_3, \tag{39}$$

where the coefficients c_i are defined as the product of the reflection coefficient r_n at all the interfaces in which the considered layers are not adjacent.

If we take only one layer from N, we have $c_n = r_{n-1}r_n$. So, for example $c_1 = r_0r_1$.

If we take m layers from N, we have $\frac{N!(N-m)!}{m!}$ possibilities and we should consider the reflection coefficients at the interfaces where the layers considered are not adjacent.

If we consider two layers from three we have three possibilities x_1x_2, x_1x_3, and x_2x_3 (see Fig. 7).

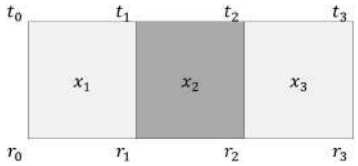


Fig. 7. Theoretical scene we consider three different layers.

For layer 1/layer 2 case, we have $c_{12} = r_0r_2$. There are only two free interfaces. For layer 2/layer 3 case, we have $c_{23} = r_1r_3$. There are, again, only two free interfaces. However, for the layer 1/layer 3 case, we got $c_{13} = r_0r_1r_2r_3$. There are four free interfaces.

The expression of every D_N can be deduced using this technique.

Appendix B

We are interested to determine the transmission coefficient Γ_N of the layered structure. The idea is to express the denominator D_N as a sum of D_{N-n} where n goes from 1 to N. We noticed that D_N can be put into the form:

$$D_N = D_{N-1} + r_N x_N \widetilde{D}_{N-1}, \tag{40}$$

where $\widetilde{D}_{N-1}$ can be written as:

$$\widetilde{D}_{N-1} = r_{N-1}D_{N-2} + x_{N-1}\widetilde{D}_{N-2}, \quad \widetilde{D}_1 = r_1 + r_0x_1, \tag{41}$$

which yields:

$$D_N = D_{N-1} + r_N r_{N-1} x_N (D_{N-2} + x_{N-1}D'_{N-2}), \tag{42}$$

we put:

$$\begin{aligned} C_n &= 1 + r_n r_{n-1} x_n, \\ C_1 &= D_1 = 1 + r_1 r_0 x_1 E_n = r_{n-1} + r_n x_n. \end{aligned} \tag{43}$$

For two layers:

$$D_2 = D_1 + r_2x_2\widetilde{D}_1 = C_1(1 + \beta_1 r_2 x_2) = C_1C_{21} \tag{44}$$

with:

$$C_{21} = 1 + \beta_1 r_2 x_2, \qquad \beta_1 = \frac{\widetilde{D}_1}{C_1} = \frac{1}{r_1}\left(1 - \frac{t_1t'_1}{C_1}\right). \tag{45}$$

For three layers we arrange D_3 in the form:

$$\begin{aligned} D_3 &= D_2 + r_3x_3\widetilde{D}_2 \\ &= D_1 + r_2x_2\widetilde{D}_1 + r_3x_3\left(r_2D_1 + x_2\widetilde{D}_1\right) \\ &= C_1C_3 + x_2\widetilde{D}_1E_3 = C_1C_3C_{31}, \end{aligned} \tag{46}$$

where

$$C_{23} = 1 + \beta_1\beta_3 x_2, \qquad \beta_3 = \frac{E_3}{C_3} = \frac{1}{r_2}\left(1 - \frac{t_2t'_2}{C_3}\right). \tag{47}$$

For four layers:

$$\begin{aligned} D_4 &= D_3 + r_4x_4\widetilde{D}_3 \\ &= C_4\left(D_2 + \beta_4x_3\widetilde{D}_2\right) \\ &= C_1C_4\left(C_{34} + \beta_1x_2\left(r_2 + \beta_4x_3\right)\right), \end{aligned} \tag{48}$$

$$D_4 = C_1C_4C_{24}C_{341} \tag{49}$$

with:

$$\begin{cases} C_{34} = 1 + \beta_4 r_2 x_3, \\ C_{24} = 1 + \beta_1\beta_{34} x_2, \end{cases}$$

$$\begin{cases} \beta_4 = \dfrac{E_4}{C_4} = \dfrac{1}{r_3}\left(1 - \dfrac{t_3t'_3}{C_4}\right), \\ \beta_{34} = \dfrac{r_2 + \beta_4x_3}{C_{34}}. \end{cases} \tag{50}$$

Generalizing for N layers D_N is put in the form:

$$D_{N\geq 4} = C_NC_1\prod_{m=2}^{N-1} C_{mN}, \tag{51}$$

$$\begin{cases} C_n = 1 + r_{n-1}r_nx_n, \\ C_1 = 1 + r_0r_1x_1, \\ C_{2N} = 1 + \beta_1\beta_{3N}x_2, \\ C_{mN} = 1 + \beta_{m+1N}r_{m-1}x_m, \end{cases}$$

$$\begin{cases} \beta_n = \dfrac{E_n}{C_n} = \dfrac{1}{r_{n-1}}\left(1 - \dfrac{t_{n-1}t'_{n-1}}{C_n}\right), \\ \beta_{mN} = \dfrac{r_{m-1} + \beta_{m+1N}x_m}{C_{mN}}. \end{cases} \tag{52}$$

References

1. Ainslie M.A. (1995), Plane-wave reflection and transmission coefficients for a three layered elastic medium, *The Journal of the Acoustical Society of America*, **97**(2): 954–961, doi: 10.1121/1.412074.
2. Bakhtiari-Nejad M., Hajj M.R., Shahab S. (2020), Dynamics of acoustic impedance matching layers in contactless ultrasonic power transfer systems, *Smart Materials and Structure*, **29**: 035037, doi: 10.1088/1361-665X/ab6fe5.
3. Chern E.J., Nielsen H.T.C. (1989), Generalized formulas for reflected pulse response of multilayered structures, *Journal of Applied Physics*, **66**(7): 2833–2837, doi: 10.1063/1.344212.
4. Chern E.J., Nielsen H.T.C. (1990), Generalized pulse equations for through-transmission evaluation of arbitrary multilayered structures, *Research in Nondestructive Evaluation*, **2**: 1–9.
5. Conoir J.M., (1991), Interferences and periodic distribution of resonances in an elastic plate [in French: Interférences et périodicités sur les résonances dans une plaque élastique], *Journal d'acoustique*, **4**: 377–412.
6. Derem A. (1982), Series of transmitted waves for a fluid and hollow cylinder: an exact solution [in French: Série des ondes transmises pour un cylindre fluide et creux: une solution exacte], *Revue du CETHEDEC*, **19**(70): 1–27.
7. Derible S., Tinel A. (2011), Resonances of two elastic plates separated by a thickness of water. Study by means of transition terms [in French: Résonances de deux plaques élastiques séparées par une épaisseur d'eau. Etude au moyen des termes de transition], [in:] *20ème Congrès Français de Mécanique Besançon.*
8. Deschamps M., Chengwei C. (1991), Reflection/refraction of a solid layer by Debye's series expansion, *Ultrasonics*, **29**(4): 288–293, doi: 10.1016/0041-624X(91)90024-3.
9. Fiorito R., Madigoskay W., Überall H. (1981), Acoustic resonances and the determination of the material parameters of a viscous fluid layer, *The Journal of the Acoustical Society of America*, **69**(4): 897–903, doi: 10.1121/1.385610.
10. Fiorito R., Überall H. (1979), Resonance theory of acoustic reflection and transmission through a fluid layer, *The Journal of the Acoustical Society of America*, **65**(1): 9–14, doi: 10.1121/1.382275.
11. 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.
12. Gérard A. (1979), Field resulting from the incidence of P or SV waves on an elastic sphere [in French: Champ résultant de l'incidence d'ondes P ou SV sur une sphère élastique], *Comptes rendus de l'Académie des Sciences. Série IIb, Mécanique, Elsevier*, **289**: 237–240.
13. Gérard A. (1980), Field resulting from the incidence of P and SV waves on a stratified medium with spherical symmetry [in French: Champ résultant de l'incidence d'ondes P et SV sur un milieu stratifié à symétrie sphérique], *Comptes rendus de l'Académie des Sciences. Série IIb, Mécanique, Elsevier*, **290**(3): 43–46.
14. Gérard A. (1982), Factorization of the characteristic equation of a multi-layered elastic sphere: interpretation of resonances [in French: Factorisation de l'équation caractéristique d'une sphère élastique multicouches: interprétation des résonances], *Comptes rendus de l'Académie des Sciences. Série IIb, Mécanique, Elsevier*, **297**: 17–19.
15. Gérard A. (1987), Modal formalism: interpretation [in French: Formalisme modal: interprétation], [in:] *Diffusion Acoustique* [in French: *La Diffusion Acoustique*], Gespa N., Poirée B. [Eds.], Revue du CETHEDEC, pp. 165–287.
16. Gérard A. (2022), Generalized Debye series theory for acoustic scattering: Some applications, [in:] *Generalized Models and Non-classical Approaches in Complex Materials 1. Advanced Structured Materials*, Visakh P.M. [Ed.], Vol. 89, Springer Singapore, pp. 349–374.
17. Gudra T., Banasiak D. (2020), Optimal selection of multicomponent matching layers for piezoelectric transducers using genetic algorithm, *Archives Acoustics*, **45**(4): 699–707, doi: 10.24425/aoa.2020.135276.
18. Haskell N.A. (1953), The dispersion of surface waves on multilayered media, *Bulletin of the Seismological Society of America*, **43**(1): 17–34, doi: 10.1785/BSSA0430010017.
19. Hsu D.K. (2009), Nondestructive evaluation of sandwich structures: A review of some inspection techniques, *Journal of Sandwich Structures and Materials*, **11**(4): 275–291, doi: 10.1177/1099636209105377.
20. Ingard K.U., Morse P.M. (1968), *Theoretical Acoustics*, Princeton University Press, Princeton, New Jersey.
21. Khaled A., Maréchal P., Lenoir O., Ech-Cherif El-Kettani M., Chenouni D. (2013), Study of the resonances of periodic plane media immersed in water: Theory and experiment, *Ultrasonics*, **53**(3): 642–647, doi: 10.1016/j.ultras.2012.11.011.
22. Lenoir O., Maréchal P. (2009), Study of plane periodic multilayered viscoelastic media: Experiment and simulation, [in:] *2009 IEEE International Ultrasonics Symposium Proceedings*, pp. 1028–1011, doi: 10.1109/ULTSYM.2009.5441518.
23. Lowe M.J.S. (1995), Matrix techniques for modeling ultrasonic waves in multilayered media, [in:] *IEEE Transaction of Ultrasonic, Ferroelectric, and Frequency Control*, **42**(4): 525–542, doi: 10.1109/58.393096.
24. Maréchal P. *et al.* (2008), Modeling of a high frequency ultrasonic transducer using periodic structures, *Ultrasonics*, **48**(2): 141–149, doi: 10.1016/j.ultras.2007.11.007.
25. Maréchal P., Lenoir O., Khaled A., Ech-Cherif El-Kettani M., Chenouni D. (2014), Viscoelasticity effect on a periodic plane medium immersed in water,

Acta Acustica united with Acustica, **100**(6): 1036–1043, doi: 10.3813/AAA.918783.

26. Messineo M.G., Frontini G.L., Eliçabe G.E., Gaete-Garretón L. (2013), Equivalent ultrasonic impedance in multilayer media. A parameter estimation problem, *Inverse Problems in Science and Engineering*, **21**(8): 1268–1287, doi: 10.1080/17415977.2012.757312.
27. Messineo M.G., Rus G., Eliçabe G.E., Frontini G.L. (2016), Layered material characterization using ultrasonic transmission: An inverse estimation methodology, *Ultrasonics*, **65:** 315–328, doi: 10.1016/j.ultras.2015.09.010.
28. Nayfeh A.H. (1991), The general problem of elastic wave propagation in multilayered anisotropic media, *The Journal of the Acoustical Society of America*, **89**(4): 1521–1531, doi: 10.1121/1.400988.
29. Potel C., de Belleval J.-F. (1993), Propagation in an anisotropic periodically multilayered medium, *The Journal of the Acoustical Society of America*, **93**(5): 2669–2677, doi: 10.1121/1.405842.
30. Rokhlin S.I., Wang Y.J. (1992), Equivalent boundary conditions for thin orthotropic layer between two solids: Reflection, refraction, and interface waves, *The Journal of the Acoustical Society of America*, **91**(4): 1875–1887, doi: 10.1121/1.403717.
31. Scott W.R., Gordon P.F. (1977), Ultrasonic spectrum analysis for nondestructive testing of layered composite materials, *The Journal of the Acoustical Society of America*, **62**(1): 108–116, doi: 10.1121/1.381491.
32. Shenand M., Cao W. (2000), Acoustic band gap formation in a periodic structure with multilayer unit cells, *Journal of Physics D: Applied Physics*, **33**(10): 1150–1154, doi: 10.1088/0022-3727/33/10/303.
33. Solyanik F.I. (1977), Transmission of plane waves through a layered medium of anisotropic materials, *Soviet Physics Acoustics-USSR*, **23**: 533–536.
34. Soucrati H., Chitnalah A., Aouzale N., Jakjoud H. (2018), Analytical model of three-dimensional ultrasonic beam interaction with an immersed plate, *Archives of Acoustics*, **43**(4): 669–679, doi: 10.24425/aoa.2018.125160.
35. Storheim E., Lohne K.D., Hergum T. (2015), Transmission and reflection from a layered medium in water. Simulations and measurements, [in:] *Proceedings of the 38th Scandinavian Symposium on Physical Acoustics*, Norway.
36. Stovas A., Arntsen B. (2006), Vertical propagation of low-frequency waves in finely layered media, *Geophysics*, **71**(3): T87–T94, doi: 10.1190/1.2197488.
37. Yang X., Zhang C., Wang C., Sun A., Ju B.-F., Shen Q. (2019), Simultaneous ultrasonic parameter estimation of a multi-layered material by the PSO-based least squares algorithm using the reflection spectrum, *Ultrasonics*, **91**: 231–236, doi: 1016/j.ultras.2018.08.003.

Archives of Acoustics Vol. 48, No. 1, pp. 81–91 (2023), doi: 10.24425/aoa.2023.144266

Research Paper

Evaluation of the Effect of Uncertainties on the Acoustic Behavior of a Porous Material Located in a Duct Element Using the Monte Carlo Method

Hanen HANNACHI[(1),(2)], Hassen TRABELSI[(1)], Marwa KANI[(1),(2)], Mohamed TAKTAK[(1),(2)*], Mabrouk CHAABANE[(2)], Mohamed HADDAR[(2)]

[(1)] *Laboratory of Mechanics, Modeling and Productivity (LA2MP)*
National School of Engineers of Sfax, University of Sfax
Sfax, Tunisia

[(2)] *Faculty of Science of Sfax, University of Sfax*
Sfax, Tunisia

[*]Corresponding Author e-mail: mohamed.taktak@fss.rnu.tn

(*received January 26, 2022; accepted September 6, 2022*)

When studying porous materials, most acoustical and geometrical parameters can be affected by the presence of uncertainties, which can reduce the robustness of models and techniques using these parameters. Hence, there is a need to evaluate the effect of these uncertainties in the case of modeling acoustic problems. Among these evaluation methods, the Monte Carlo simulation is considered a benchmark for studying the propagation of uncertainties in theoretical models. In the present study, this method is applied to a theoretical model predicting the acoustic behavior of a porous material located in a duct element to evaluate the impact of each input error on the computation of the acoustic proprieties such as the reflection and transmission coefficients as well as the acoustic power attenuation and the transmission loss of the studied element. Two analyses are conducted; the first one leads to the evaluation of the impacts of error propagation of each acoustic parameter (resistivity, porosity, tortuosity, and viscous and thermal length) through the model using a Monte Carlo simulation. The second analysis presents the effect of propagating the uncertainties of all parameters together. After the simulation of the uncertainties, the 95% confidence intervals and the maximum and minimum errors of each parameter are computed. The obtained results showed that the resistivity and length of the porous material have a great influence on the acoustic outputs of the studied model (transmission and reflection coefficients, transmission loss, and acoustic power attenuation). At the same time, the other physical parameters have a small impact. In addition, the acoustic power attenuation is the acoustic quantity least impacted by the input uncertainties.

Keywords: porous material; physical parameters; transmission loss; acoustic power attenuation; Monte Carlo method.

1. Introduction

Studying acoustic propagation in duct systems containing a porous material is still today important in acoustic research and industrial communities. In fact, this kind of duct element is used in many industrial applications such as transport and building domains. The objective of these studies is to predict and better understand the involved physical phenomena related to the acoustic propagation of these duct elements (reflection, transmission, attenuation, absorption, convection, diffraction, refraction, etc.). The use of porous materials is justified by the fact that they possess good absorption properties, and are easy to manufacture and install. The theoretical description of acoustic propagation in porous media has been con-

stantly progressing since the 1940s. It is now relatively well-known thanks to the contributions of many models which start with Zwikker and Kosten (1949), who are the first to model sound propagation in porous media. Then Delany and Bazley (1970) established an empirical model according to which the acoustic characteristics (the characteristic impedance and the propagation coefficient) depend only on the ratio of the frequency f to the air flow resistivity. Then, the porosity φ and the tortuosity α_∞ were introduced as presented in the models (Attenborough, 1982; 1983) to take into account the complexity of the pore geometry in high frequencies. Johnson *et al.* (1987) introduced the physical concept of viscous characteristic length Λ. This model was next completed by Champoux and Allard (1991) by adding the description of thermal characteristic length Λ' effects. Later on, Lafarge *et al.* (1997) refined the Champoux and Allard model by introducing a new parameter called the thermal permeability k_0', which describes the damping of sound waves due to the thermal exchanges between the fluid and the structure at the pore surface.

To study the acoustic behavior of duct elements containing porous materials, some matrices can be used coupled with the previous porous acoustic models. Among these matrices, two present a great interest: the first is the transfer matrix as presented in (Peat, 1988; Tanaka *et al.*, 1985; Othmani *et al.*, 2016; 2017; Kani *et al.*, 2019; 2021) from which the acoustic transmission loss (TL) can be computed. The second is the scattering matrix (Bi *et al.*, 2006; Sitel *et al.*, 2006; Taktak *et al.*, 2010; 2013; Jdidia *et al.*, 2014; Othmani *et al.*, 2015; Kessentini *et al.*, 2016; Masmoudi *et al.*, 2017; Ben Souf *et al.*, 2017; Dhief *et al.*, 2020; Tounsi *et al.*, 2022) that contains important information about the transmission and reflection phenomena and is used for the acoustic power attenuation computation. Consequently, these two matrices give complete information about the acoustic behavior of a duct element.

In general, most of the parameters used in the theoretical modeling are characterized by the presence of some uncertainties that affect the robustness of such modeling. In order to avoid any errors, it is necessary to evaluate the effect of these errors in the final results and determine the more influential parameters. To achieve this objective, uncertainty analyses are used. A widely used stochastic technique, called the Monte Carlo method, can be integrated into the theoretical models in order to evaluate the propagation of errors and their degree of influence (Taktak *et al.*, 2009; Trabelsi *et al.*, 2017; Bouazizi *et al.*, 2019).

The aim of the work presented in this paper is to evaluate the impact of uncertainties affecting the physical and geometrical parameters of a duct element containing a porous material on its acoustic behavior. To reach this goal, the Monte Carlo technique is coupled with theoretical modeling to compute the transfer and scattering matrices of the studied duct element. The used acoustic porous model is the Johnson–Champoux–Allard–Lafarge (JCAL) model (Johnson *et al.*, 1987; Champoux, Allard, 1991; Lafarge *et al.*, 1997), which incorporates the maximum of parameters.

In the present study, the effects of uncertainties of model parameters on the porous material acoustic properties such as reflection and transmission coefficients, as well as the transmission loss and the acoustic power attenuation, are evaluated and investigated. This is obtained using the Monte Carlo method allowing the computation of the 95% confidence intervals of the model outputs as well as the corresponding errors.

The outline of the paper is as follows: in Sec. 2, the theoretical basis of modeling the acoustic behavior of porous element using the JCAL model and the computation of transfer and scattering matrices of the studied duct element are presented in detail. These matrices are then used to calculate the transmission loss and the acoustic power attenuation of the porous material. Section 3 presents the details of the uncertainty analysis based on the Monte Carlo method. Finally, the numerical results are presented and discussed in Sec. 4.

2. Theoretical basis

2.1. Transfer matrix and transmission loss computation

The studied duct element is located between the two axial coordinates z_1 and z_2, as shown in Fig. 1. It contains a porous material with a length equal to L. According to the duct element dimensions used in the present study, only the propagation of the acoustic plane wave through the duct element in the z-direction is assumed. This propagation is modeled using the transfer matrix $[T]$, which provides the relationship between the acoustic pressure P_1 and the particle velocity U_1 in duct I at $z_1 = 0$ and the acoustic pressure P_2 and the particle velocity U_2 in duct II at $z_2 = L$ (Fig. 1).

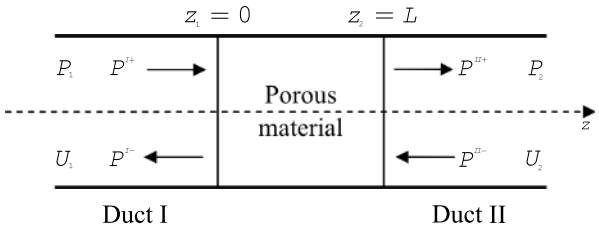


Fig. 1. Description of the studied duct.

The general formulation of the transfer matrix is given as follows (Atalla, Allard, 2009):

$$\begin{pmatrix} P_1 \\ U_1 \end{pmatrix} = [T] \begin{pmatrix} P_2 \\ U_2 \end{pmatrix}. \tag{1}$$

For a single fluid layer, the transfer matrix is constructed as:

$$[T] = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} \cos(k(w)L) & jZ(w)\sin(k(w)L) \\ \dfrac{j}{Z(w)}\sin(k(w)L) & \cos(k(w)L) \end{bmatrix}, \quad (2)$$

where $j = \sqrt{-1}$, w is the angular frequency, $Z(w)$ is the characteristic impedance, and $k(w)$ is the acoustic wavenumber of the porous material. These latter two intrinsic quantities are linked to the dynamic mass density $\rho(w)$ and the bulk modulus $K(w)$ of the porous material by the following relations:

$$Z(w) = \sqrt{\rho(w)K(w)}, \quad (3)$$

$$k(w) = w\sqrt{\frac{\rho(w)}{K(w)}}. \quad (4)$$

According to the JCAL model (Johnson *et al.*, 1987; Champoux, Allard, 1991; Lafarge *et al.*, 1997), the expressions of $\rho(w)$ and $K(w)$ are given as follows:

$$\rho(w) = \frac{\alpha_\infty \rho_0}{\phi}\left(1 - j\frac{\sigma\phi}{w\rho_0\alpha_\infty}\sqrt{1 + j\frac{4\alpha_\infty^2\eta\rho_0 w}{\sigma^2\Lambda^2\phi^2}}\right), \quad (5)$$

$$K(w) = \frac{\gamma P_0}{\phi}\left(\gamma - \frac{\gamma - 1}{1 - j\frac{8\eta}{\Lambda' N_{\mathrm{Pr}}\sqrt{1 + j\frac{\Lambda'^2 N_{\mathrm{Pr}}}{16\eta}}}}\right)^{-1}, \quad (6)$$

where ϕ is the open porosity, σ is the static air-flow resistivity, α_∞ is the high-frequency limit of the dynamic tortuosity, Λ is the characteristic viscous length, Λ' is the characteristic thermal length, ρ_0 is the density at rest of the fluid saturating the pores, η is its dynamic viscosity, N_{Pr} is its Prandtl number, γ is its specific heat ratio, and P_0 is the atmospheric pressure.

The power transmission factor τ of the porous material is defined as the ratio of the transmitted power W_t and the power incident on the porous material W_i:

$$\tau = \frac{W_t}{W_i}. \quad (7)$$

The transmission loss (TL) is defined in dB as:

$$\mathrm{TL} = 10\log\left(\frac{1}{\tau}\right). \quad (8)$$

Using the four elements T_{11}, T_{12}, T_{21}, and T_{22} of the transfer matrix $[T]$, the transmission loss can be calculated as follows:

$$\mathrm{TL} = 20\log\left(\frac{1}{2}\left|T_{11} + \frac{T_{12}}{Z_0} + Z_0 T_{21} + T_{22}\right|\right), \quad (9)$$

where $Z_0 = \rho_0 c_0$ is the characteristic impedance of the surrounding medium, and c_0 is its speed of sound.

2.2. Scattering matrix and acoustic power attenuation computation

The scattering matrix $[S]$ of the studied duct element located between z_1 and z_2 (Fig. 1) is a linear relationship between the incoming pressures vector $\{P^{\mathrm{in}}\}$ and the outgoing pressures vector $\{P^{\mathrm{out}}\}$ and can be expressed as:

$$\{P^{\mathrm{out}}\} = [S]_{2\times 2}\{P^{\mathrm{in}}\}, \quad (10)$$

where

$$\{P^{\mathrm{in}}\} = \begin{Bmatrix} P^{I+}(z_1) \\ P^{II-}(z_2) \end{Bmatrix} \quad \text{and} \quad \{P^{\mathrm{out}}\} = \begin{Bmatrix} P^{I-}(z_1) \\ P^{II+}(z_2) \end{Bmatrix}.$$

$P^{I+}(z_1)$, $P^{I-}(z_1)$, $P^{II+}(z_2)$, and $P^{II-}(z_2)$ represent the incident, the reflected, the transmitted, and the retrograde pressures, respectively (Fig. 1).

This matrix

$$[S]_{2\times 2} = \begin{bmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{bmatrix}$$

depicts only the studied duct element and is independent of the upstream and downstream acoustic conditions. The physical meaning of each coefficient is as follows:

- S^{11} is the reflection coefficient of the wave coming into the element from the left side,
- S^{22} is the reflection coefficient of the wave coming into the element from the right side,
- S^{12} is the transmission coefficient of the wave coming into the element from the right side,
- S^{21} is the transmission coefficient of the wave coming into the element from the left side.

For a symmetric studied duct element and the same mediums on both sides of the element, the following can be written:

$$\begin{Bmatrix} S^{11} = S^{22} \\ S^{12} = S^{21} \end{Bmatrix}. \quad (11)$$

The four scattering matrix coefficients can be expressed in terms of the transfer matrix coefficients as follows (Hu, 2010):

$$S^{11} = \frac{X^+ - W^+}{X^+ + W^+}, \quad (12)$$

$$S^{22} = -\frac{X^- + W^-}{X^+ + W^+}, \quad (13)$$

$$S^{12} = \frac{X^+W^- - W^+X^-}{X^+ + W^+}, \quad (14)$$

$$S^{21} = \frac{2}{X^+ + W^+}, \quad (15)$$

where

$$\left\{ \begin{array}{l} X^{\pm} = T_{11} \pm \dfrac{T_{12}}{Z_0} \\ W^{\pm} = Z_0 T_{21} \pm T_{22} \end{array} \right\}. \tag{16}$$

The acoustical power attenuation W_{att} of the studied duct element is defined in decibel [dB] as follows:

$$W_{\text{att}} = 10 \log \left(\frac{W^{\text{in}}}{W^{\text{out}}} \right), \tag{17}$$

where W^{in} is the total acoustic power of incoming waves and W^{out} is the total acoustic power of outgoing waves. This acoustic power attenuation can be computed using the scattering matrix $[S]$ (Taktak *et al.*, 2010; Kani *et al.*, 2021):

$$W_{\text{att}} = 10 \log \left(\frac{|d_1|^2 + |d_2|^2}{\lambda_1 |d_1|^2 + \lambda_2 |d_2|^2} \right) \tag{18}$$

with:

– λ_1 and λ_2 are the eigenvalues of the matrix $[H]$ defined as follows:

$$[H] = [S]^H \cdot [S] = [V] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} [V]^H, \tag{19}$$

where the subscript H denotes the conjugate transpose, and $[V]$ is the matrix of the eigenvectors of the matrix $[H]$.

– d_1 and d_2 are the components of the vector $\{d\}$ calculated by the following equation:

$$\{d\} = \sqrt{\frac{1}{2Z_0}} \cdot [V]^H \cdot \left\{ \begin{array}{l} P^{I+} \\ P^{II-} \end{array} \right\}. \tag{20}$$

3. Uncertainty analysis: Monte Carlo simulation

The JCAL model (this model was detailed in Sec. 2) adopted for the sound propagation in the porous material study uses the five material physical parameters: porosity, static air-flow resistivity, high-frequency limit of the dynamic tortuosity, characteristic viscous and thermal lengths.

In order to study the propagation of uncertainties through the used model, uncertainty analysis is conducted (Taktak *et al.*, 2009; Trabelsi *et al.*, 2017; Bouazizi *et al.*, 2019). A probability distribution is first applied to each model's physical parameters input, and then an investigation of their effect on model outputs (transmission and reflection coefficients, transmission loss, and acoustic power attenuation) is performed.

The computing steps for the Monte Carlo simulation used for this analysis are regrouped in the algorithm illustrated in Fig. 2. In the first step, a Gaussian distribution for each input is defined. Then, M-samples are created for each studied model input. After that, the algorithm is executed N-times to generate a set of outputs. Finally, the obtained outputs are regrouped, and the 95% confidence interval with the corresponding maximum and minimum errors are estimated.

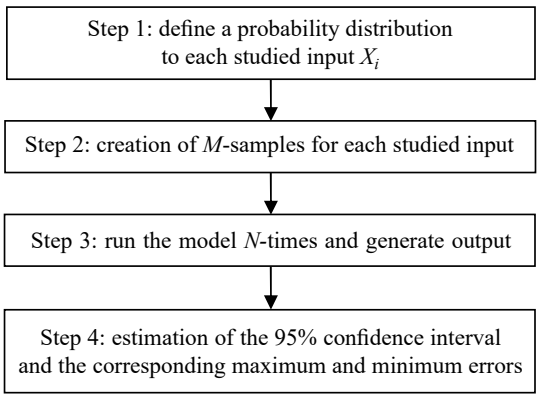


Fig. 2. Monte Carlo algorithm for uncertainty analysis.

4. Results and discussion

The presented uncertainty analysis is applied to a porous material located in a duct element in the frequency band (0–4000 Hz). The properties of this studied porous material are presented in Table 1. Only the plane wave is propagating in the duct element. This table also presents the mean value and the standard derivation used for each parameter to obtain the 10 000 values of studied acoustic parameters, which are then used in the proposed Monte Carlo method.

Table 1. Properties of the studied porous material.

Parameter	Mean value	Standard deviation
Flow resistivity σ [N·s·m^{-4}]	25000	306
Porosity ϕ	0.95	0.012
Tortuosity α_∞	1	0.013
Viscous length Λ [μm]	170	5.10^{-6}
Thermal length Λ' [μm]	510	5.10^{-6}
Length of the porous material L [m]	0.05	7.10^{-4}
Density [kg·m^{-3}]	60	0.9

This section presents the results of the developed uncertainty analysis when considering the variation of one of the used parameters separately. The uncertainty analysis through the Monte Carlo method is performed as described in Fig. 2 through MATLAB software.

Initially, a ±5% variation is applied to each mean input value (shown in Table 1) according to the Gaussian distribution and by performing N = 10 000 calculations. After randomly selecting a set of each input (resistivity, porosity, tortuosity, viscous and thermal lengths), the corresponding model outputs (transmission and reflection coefficients, transmission loss, and acoustic power attenuation) are computed. Finally, these steps are repeated N-times to obtain, for each input parameter, the 95% confidence intervals limited by the minimum and the maximum values of each model output of the studied porous material and the corresponding mean value.

Uncertainty analysis results considering the flow resistivity uncertainty are presented in Fig. 3. This figure illustrates that the flow resistivity has an important effect on the studied acoustic outputs, with errors varying from 0.5% to 6%. This observation is confirmed by the 95% interval of each output parameter: these intervals present a thickness, which means that the influence of this parameter is significant. It is important

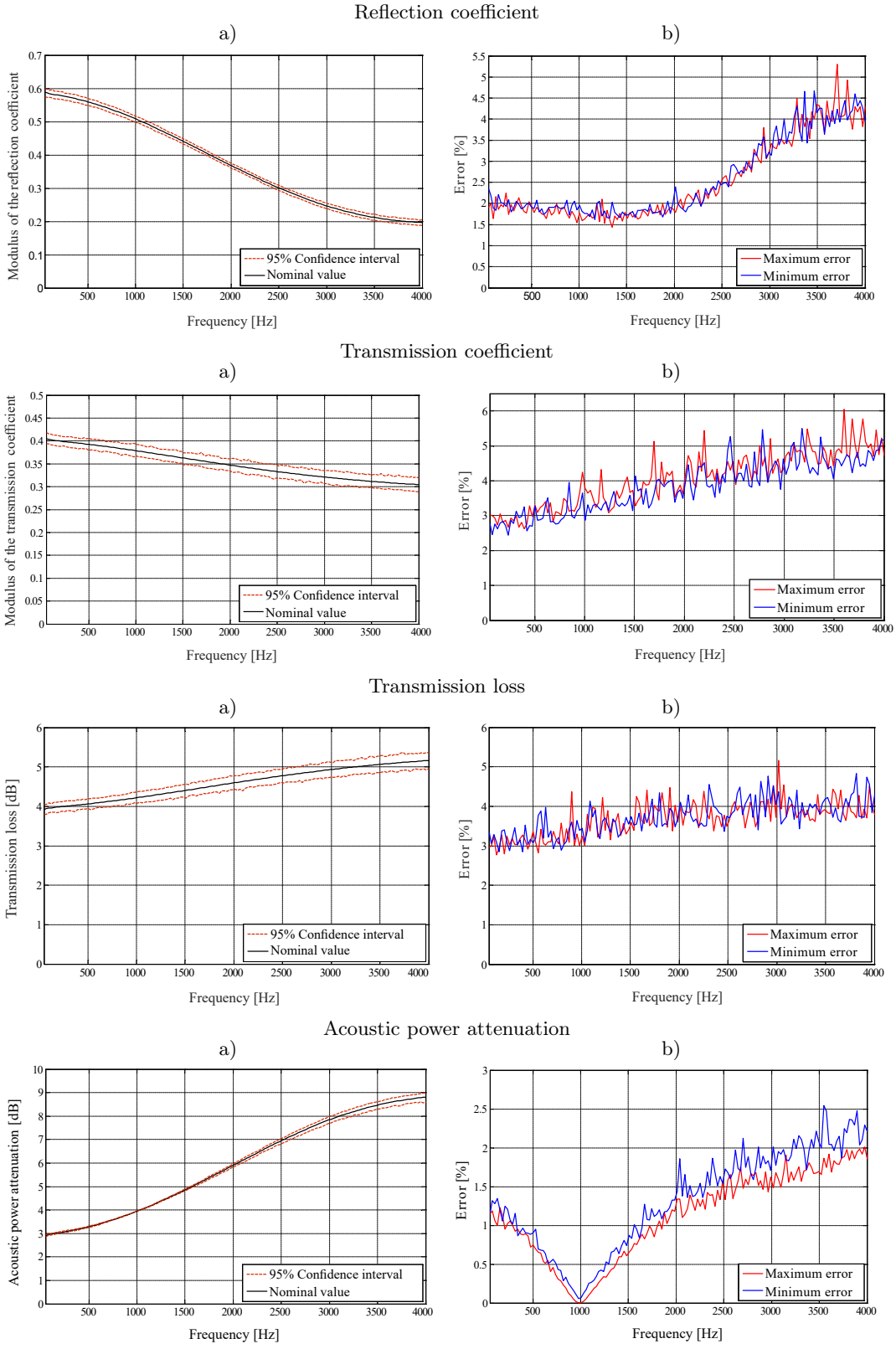


Fig. 3. Flow resistivity uncertainties' effect on the acoustic outputs: 95% confidence interval (a) and the corresponding error (b).

to indicate also that the uncertainty of the flow resistivity has a minimum influence on the acoustic power attenuation, as indicated in Fig. 3 (with an error not exceeding 2.5%). It is also observed that errors on all the output parameters dues to uncertainties increase when the frequency increases.

Figure 4 shows that the porosity has the same effect as the flow resistivity with a significant influence on the

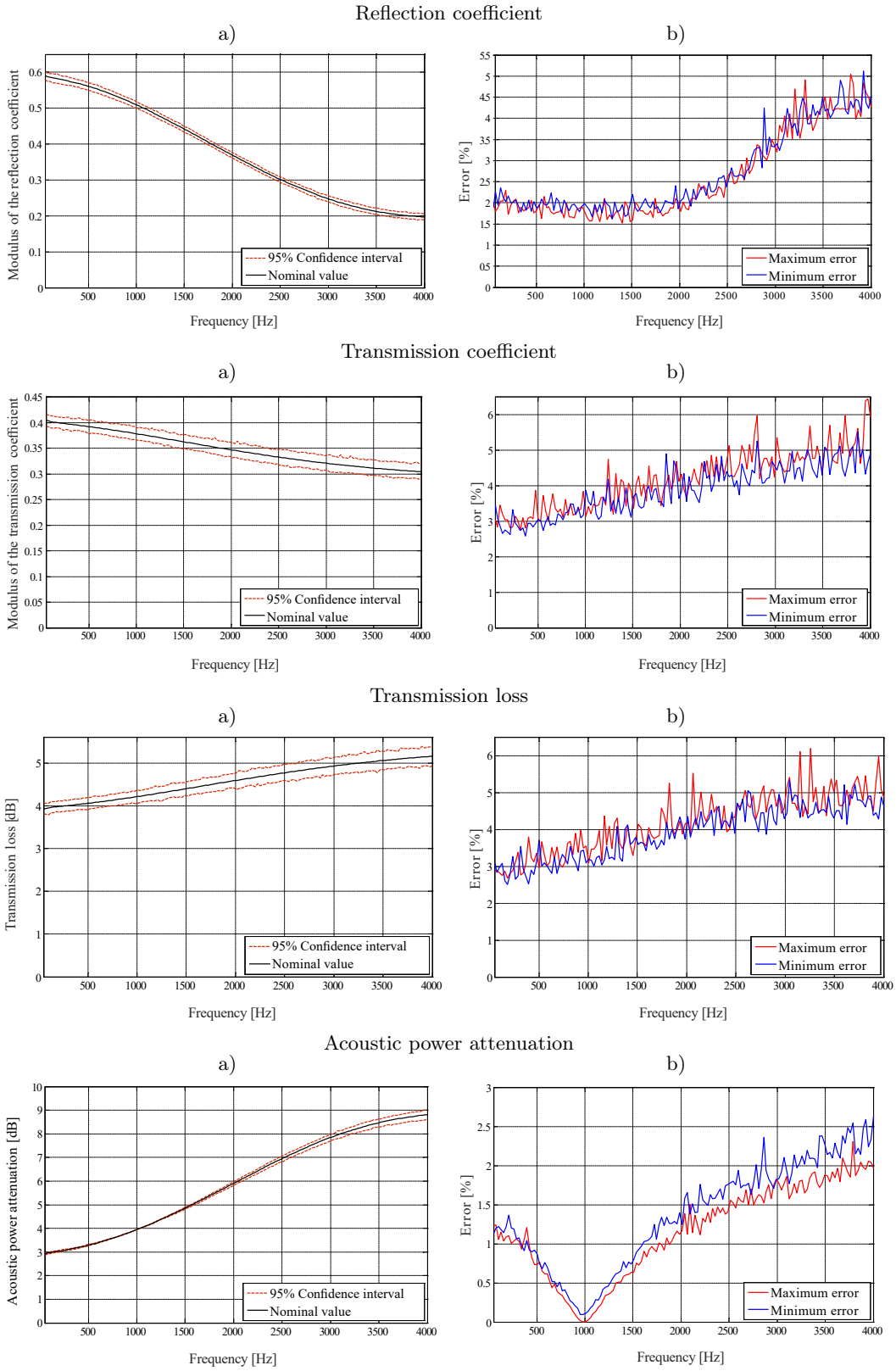


Fig. 4. Porosity uncertainties' effect on the acoustic behaviors: 95% confidence interval (a) and the corresponding error (b).

transmission and reflection coefficients as well as the transmission loss (TL) with errors reaching 6% and a small effect on the acoustic attenuation of the porous material with a maximum error equal to 3%. It is observed that the thickness of the confidence interval in Fig. 4a is small, which indicates that the effect of this

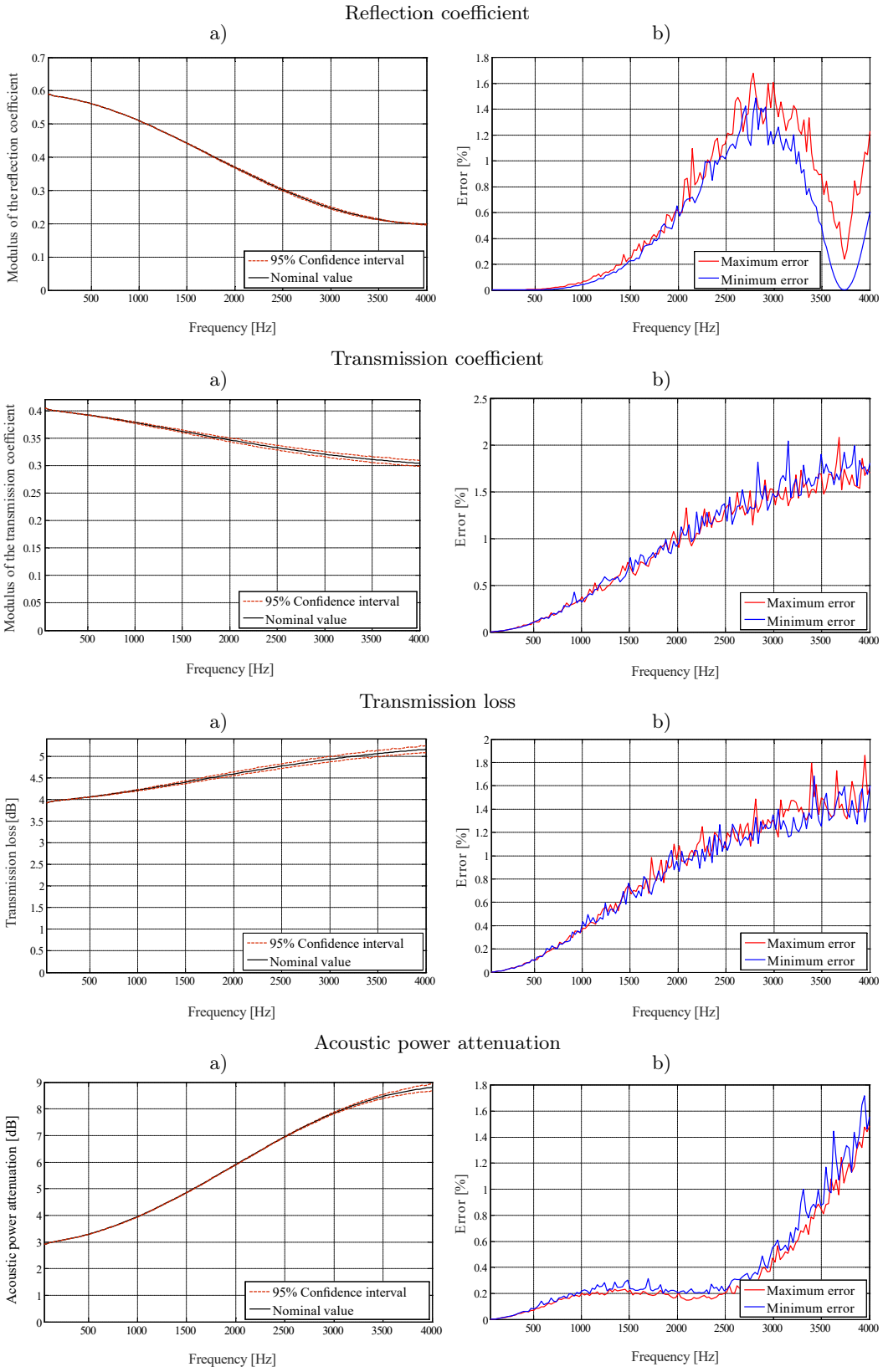


Fig. 5. Tortuosity uncertainties' effect on the acoustic behaviors: 95% confidence interval (a) and the corresponding error (b).

parameter is minor. This result is confirmed by the results presented in Fig. 4b showing the variation of the maximum and minimum errors due to the normal distribution of the porosity variation. Thus, a variation of ±5% of the nominal value of the porosity generates a minimum error equal to 2% (at high frequency) in

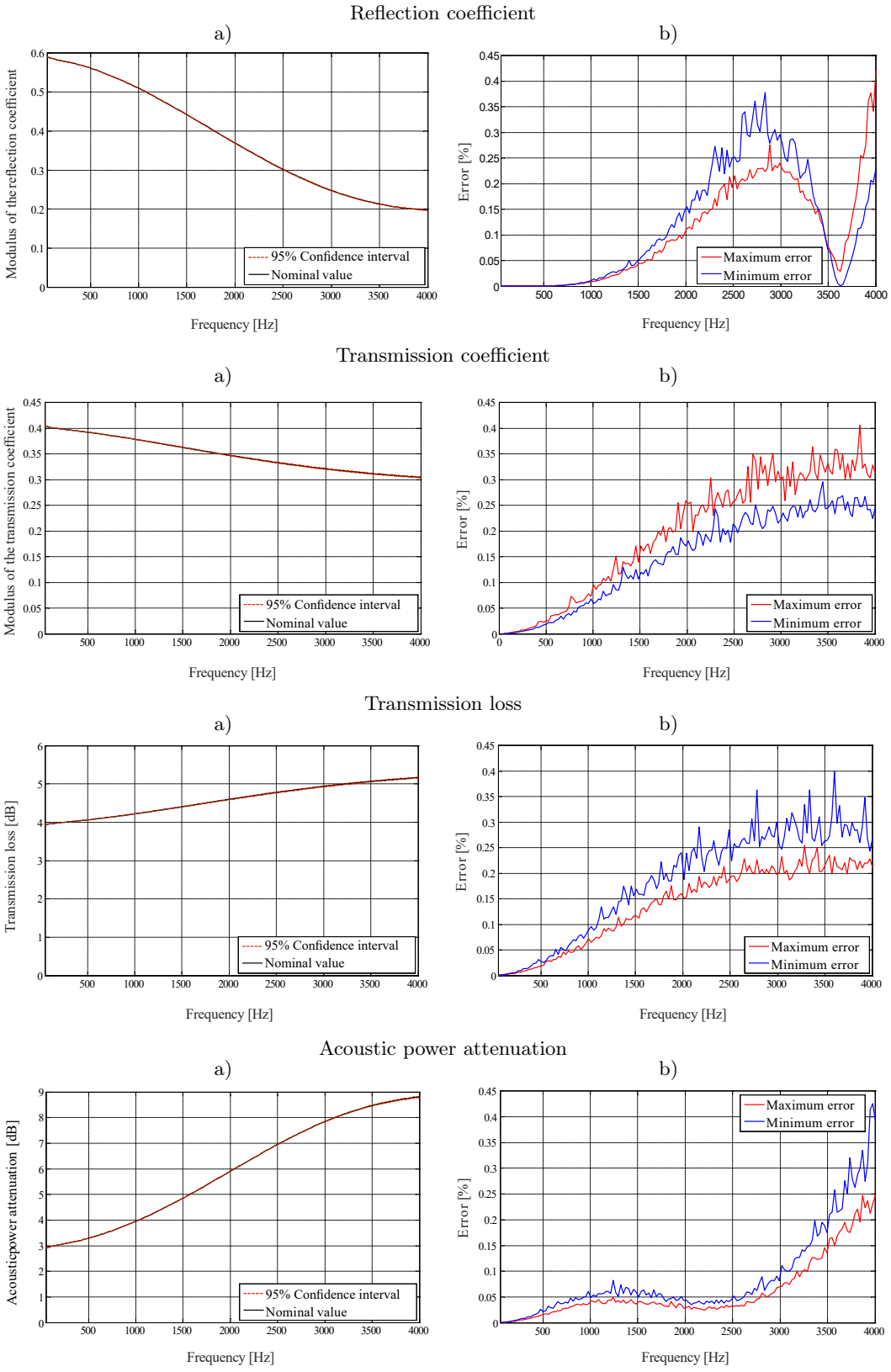


Fig. 6. Viscous characteristic length uncertainties' effect on the acoustic behaviors: 95% confidence interval (a) and the corresponding error (b).

the frequency band between 2000–4000 Hz and a maximum error equal to 2.5% in this band.

Figure 5 shows that the tortuosity has a small effect on the acoustic output parameters with small errors

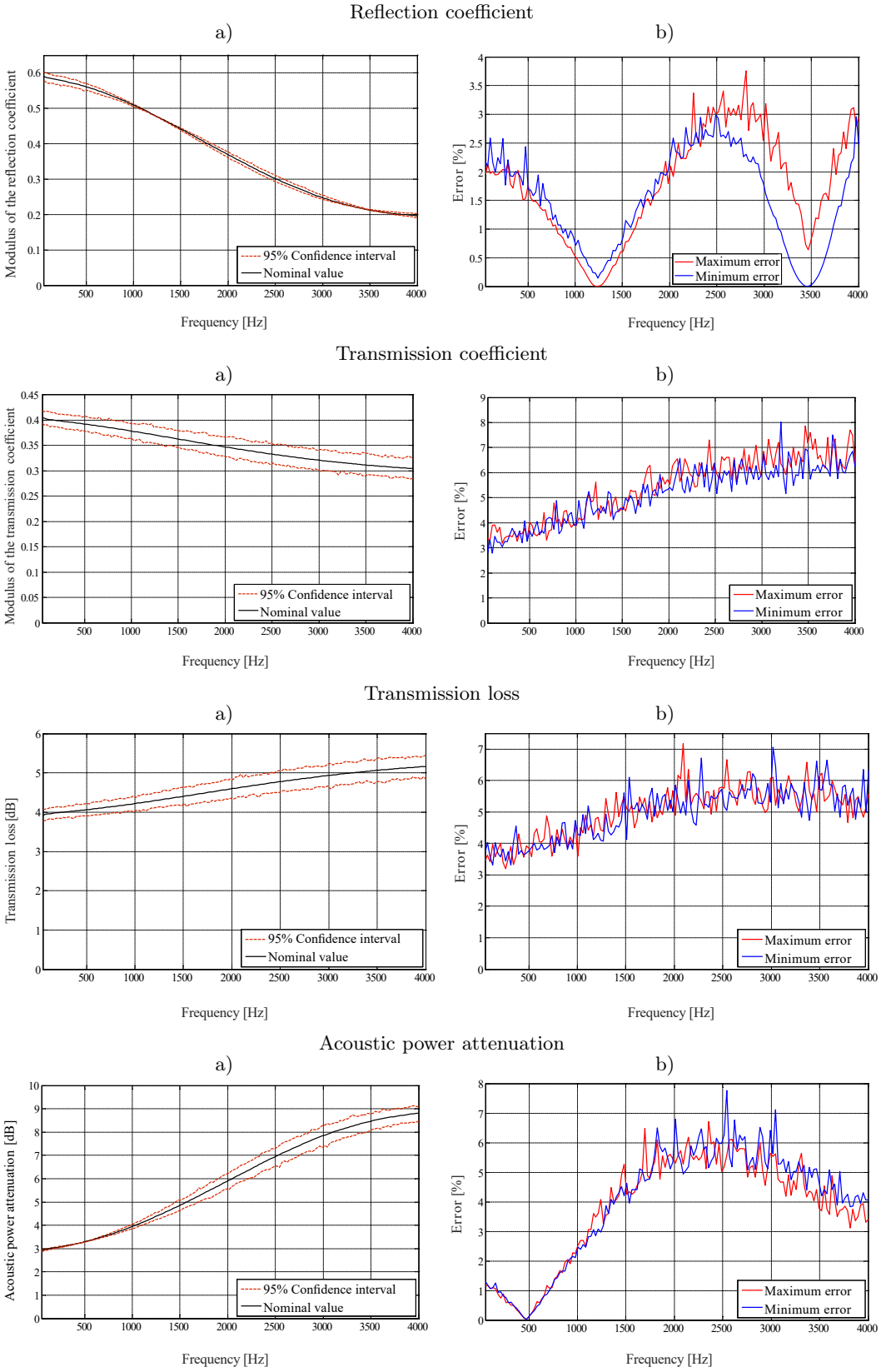


Fig. 7. Porous material length uncertainties' effect on the acoustic behaviors: 95% confidence interval (a) and the corresponding error (b).

and 95% confidence. This result is confirmed by the results presented in Fig. 5b, which shows the variation of the maximum and minimum errors due to the normal distribution of the variation of tortuosity; the maximum of these errors reaches the value of 2.5%.

Thus, a variation of ±5% of the nominal value of the tortuosity generates a minimum error equal to 0.3% (at low frequency) in the frequency band between (0–2500 Hz) and then an increase in value to 1.4% in the frequency band (2500–4000 Hz) and a maximum error equal to 1.5% in the frequency band between (2500–4000 Hz).

Figure 6 demonstrates that the viscous characteristic length Λ (as well as the thermal length Λ') has a negligible effect on the acoustic outputs the studied porous material. It is noticed that the thickness of the confidence interval is very small, which shows that the effect of this parameter is small, as shown in Fig. 6a, and with minimum errors. Thus, a variation of ±5% of the nominal value generates a minimum error equal to 0.23% and a maximum error equal to 0.33% at high frequency.

The effect of a variation of ±5% in the nominal value of the viscous and thermal characteristic lengths are, respectively, presented in Figs. 6 and 7. It is observed that the effect of these two lengths is similar.

When 5% errors are added to the nominal value of the length of porous materials, we observe that the thickness of the confidence interval is high in all the acoustic outputs. The curves show that this parameter has an important influence. A variation of ±5% of the nominal value of the length of porous materials on the attenuation generates a minimal error equal to 7.5% and a maximal variation error equal to 7% at a frequency equal to 2000 Hz.

5. Conclusion

This paper presented the results of the simulation of errors obtained by the Monte Carlo method, which allowed to determine the confidence interval of the coefficients of the scattering matrix transmission loss and acoustic attenuation for each input parameter and then for all parameters together. The Monte Carlo method is interesting for studying the degree of dependence of the model output on the inputs.

The results concluded that the resistivity and length of the porous material have a great influence on the acoustic outputs of the studied model (transmission and reflection coefficients, transmission loss, and acoustic power attenuation). At the same time, the effect of the uncertainty on the other parameters is negligible. It is important to indicate that acoustic power attenuation is the parameter less affected by input errors. The present study justifies the choice of this parameter in a previous work (Kani *et al.*, 2021) in the cost function of an inverse technique to determine porous parameters of a porous material in a duct element. By using this parameter, it is guaranteed to have results less sensitive to parameter uncertainties.

References

1. Atalla N., Allard J.F. (2009), *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials*, 2nd ed., John Wiley & Sons.
2. Attenborough K. (1982), Acoustical characteristics of porous materials, *Physics Reports*, **82**(3): 179–227, doi: 10.1016/0370-1573(82)90131-4.
3. Attenborough K. (1983), Acoustical characteristics of rigid fibrous absorbents and granular materials, *The Journal of the Acoustical Society of America*, **73**(3): 785–799, doi: 10.1121/1.389045.
4. Ben Souf M.A., Kessentini A., Bareille O., Taktak M., Ichchou M.N., Haddar M. (2017), Acoustical scattering identification with local impedance through a spectral approach, *Comptes Rendus Mécanique*, **345**(5): 301–316, doi: 10.1016/j.crme.2017.03.006.
5. Bi W.P., Pagneux V., Lafarge D., Aurégan Y. (2006), Modelling of sound propagation in a non-uniform lined duct using a multi-modal propagation method, *Journal of Sound and Vibration*, **289**(4–5): 1091–1111, doi: 10.1016/j.jsv.2005.03.021.
6. Bouazizi L., Trabelsi H., Othmani C., Taktak M., Chaabane M., Haddar M. (2019), Uncertainty and sensitivity analysis of porous materials acoustic behavior, *Applied Acoustics*, **144**: 64–70, doi: 10.1016/j.apacoust.2018.01.025.
7. Champoux Y., Allard J.-F. (1991), Dynamic tortuosity and bulk modulus in air-saturated porous media, *Journal of Applied Physics*, **70**(4): 1975–1979, doi: 10.1063/1.349482.
8. Delany M.E., Bazley E.N. (1970), Acoustical properties of fibrous absorbent materials, *Applied Acoustics*, **3**(2): 105–116, doi: 10.1016/0003-682X(70)90031-9.
9. Dhief R., Makni A., Taktak M., Chaabane M., Haddar M. (2020), Investigation on the effects of acoustic liner variation and geometry discontinuities on the acoustic performance of lined ducts, *Archives of Acoustics*, **45**(1): 49–66, doi: 10.24425/aoa.2020.132481.
10. Hu Y. (2010), *Development of passive/active hybrid panels for acoustics* [in French: *Développement de panneaux hybrides passifs/actifs pour l'acoustique*], Ph.D. Thesis, Ecole Centrale de Lyon, France.
11. Jdidia M.B., Akrout A., Taktak M., Hammami L., Haddar M. (2014), Thermal effect on the acoustic behavior of an axisymmetric lined duct, *Applied Acoustics*, **86**: 138–145, doi: 1016/j.apacoust.2014.03.004.
12. Johnson D.L., Koplik J., Dashen R. (1987), Theory of dynamic permeability and tortuosity in fluid-

saturated porous media, *Journal of Fluid Mechanics*, **176**: 379–402, doi: 10.1017/S0022112087000727.

13. Kani M. *et al.* (2019), Acoustic performance evaluation for ducts containing porous materials, *Applied Acoustics*, **147**: 15–22, doi: 10.1016/j.apacoust.2018.08.002.

14. Kani M., Makni A., Taktak M., Chaabane M., Haddar M. (2021), Identification of physical parameters of a porous material located in a duct by inverse methods, *Archives of Acoustics*, **46**(4): 657–665, doi: 10.24425/aoa.2021.139642.

15. Kessentini A., Taktak M., Ben Souf M.A., Bareille O., Ichchou M.N., Haddar M. (2016), Computation of the scattering matrix of guided acoustical propagation by the wave finite elements approach, *Applied Acoustics*, **108**: 92–100, doi: 10.1016/j.apacoust.2015.09.004.

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

17. Masmoudi A., Makni A., Takatak M., Haddar M. (2017), Effect of geometry and impedance variation on the acoustic performance of a porous material lined duct, *Journal of Theoretical and Applied Mechanics*, **55**(2): 679–694, doi: 10.15632/jtam-pl.55.2.679.

18. Othmani C. *et al.* (2016), Experimental and theoretical investigation of the acoustic performance of sugarcane wastes based material, *Applied Acoustics*, **109**: 90–96, doi: 10.1016/j.apacoust.2016.02.005.

19. Othmani C. *et al.* (2017), Acoustic characterization of a porous absorber based on recycled sugarcane wastes, *Applied Acoustics*, **120**: 90–97, 2017, doi: 10.1016/j.apacoust.2017.01.010.

20. Othmani C., Hentati T., Taktak M., Elnady T., Fakhfakh T., Haddar M. (2015), Effect of liner characteristics on the acoustic performance of duct systems, *Archives of Acoustics*, **40**(1): 117–127, doi: 10.1515/aoa-2015-0014.

21. Peat K.S. (1988), The transfer matrix of a uniform duct with a linear temperature gradient, *Journal of Sound and Vibration*, **123**(1): 43–53, doi: 10.1016/S0022-460X(88)80076-2.

22. Sitel A., Ville J-M., Foucart F. (2006), Multiload procedure to measure the acoustic scattering matrix of a duct discontinuity for higher order mode propagation conditions, *The Journal of the Acoustical Society of America*, **120**(5): 2478–2490, doi: 10.1121/1.2354040.

23. Taktak M., Ville J.-M., Haddar G., Gabard M., Foucart F. (2009), A 3D multiport scattering matrix based-method for educing wall impedance of cylindrical lined duct section: Simulation and error evaluation, *Advances in Acoustics and Vibration*, pp. 1–17, doi: 10.1155/2009/928367.

24. Taktak M., Ville J.-M., Haddar M., Gabard G., Foucart F. (2010), An indirect method for the characterization of locally reacting liners, *The Journal of the Acoustical Society of America*, **127**(6): 3548–3559, doi: 10.1121/1.3365250.

25. Tanaka T., Fujikawa T., Abe T., Utsuno H. (1985), A method for the analytical prediction of insertion loss of a two-dimensional muffler model based on the transfer matrix derived from the boundary element method, *Journal of Vibration Acoustics Stress and Reliability in Design*, **107**(1): 86–91, doi: 10.1115/1.3274721.

26. Tounsi D., Taktak W., Dhief R., Taktak M., Chaabane M., Haddar M. (2022), Evaluation of the acoustic performance of porous materials lined ducts with geometric discontinuities, *Archives of Acoustics*, **47**(2): 223–240, doi: 10.24425/aoa.2022.141652.

27. Trabelsi H., Abid M., Taktak M., Fakhfakh T., Haddar M. (2017), Effect of the aerodynamic force modeling on the tonal noise prediction model for axial fan: Sensitivity and uncertainty analysis, *Applied Acoustics*, **117**(Part A): 61–65, doi: 10.1016/j.apacoust.2016.10.017.

28. Zwikker C.W., Kosten C. (1949), *Sound Absorbing Materials*, Elsevier Publishing Company, New York.

Archives of Acoustics Vol. 48, No. 1, pp. 93–101 (2023), doi: 10.24425/aoa.2022.142908

Research Paper

Airborne and Structure-Borne Noise Control in the MB Truck Cabin Interior by the Noise Reduction in the Transmission Path

Nader MOHAMMADI

Department of Mechanical Engineering, Islamic Azad University, Parand Branch
Tehran, Iran; e-mail: nader.mohammadi@gmail.com; nmohamady@ut.ac.ir

(*received May 20, 2022; accepted September 9, 2022*)

In the current study, investigations are made to control the MB truck cabin interior noise by reducing noise in the transmission path. The main sources of cabin noise include the engine, exhaust system, air inlet system, driveline system, and tyres (especially at higher speeds). Furthermore, vibrations of the body and interior parts of the truck may significantly impact the overall in-cabin sound level. Noise is transmitted into the cabin via air (airborne noise) and cabin structure (structure-borne noise). In the noise treatment phase, noise transmission paths are considered. A viscoelastic layer damping material is used to reduce the vibration amplitude of the cabin back wall. The overall loss factor and vibration amplitude reduction ratio for the structure treated is calculated. Computational results are then compared with the values obtained by the experimental modal analysis results. Choosing the suitable material and thickness can significantly reduce the vibration amplitude. A sound barrier, silicon adhesive, and foam are also utilised for noise control in the transmission path. The effectiveness of the mentioned acoustic materials on cabin noise reduction is evaluated experimentally. The experimental SPL values are reported in the frequency range of 20 Hz–20 kHz based on a 1/3 octave filter. The experimental results show that using acoustics materials reduces the overall in-cabin sound level for a wide range of frequencies.

Keywords: truck noise sources; airborne noise; structure-borne noise; acoustic materials; viscoelastic layer damping; modal analysis.

1. Introduction

Unwanted sound or noise has some undesired impacts on human conversations. They can be as strong as to cause a person hearing loss or hearing impairments. Today, noise pollution is among those problems on which widespread efforts are focused to attenuate its effects. A low noise level is an essential feature of a product indicating its quality. Many companies, including car manufacturers, implement widespread attempts to improve their products' quality by reducing the levels of noise. Among other noise sources, vehicle noise (especially those generated by heavy trucks) significantly contributes to noise pollution. As car manufacturing industries developed and advanced, creating a competitive environment among them, the importance of vehicle quality and comfort has become more notable. A vehicle's interior and exterior noise and vibration level are essential factors determining its quality. In a vehicle, unwanted sounds and vibrations hurt vehicle parts and can cause additional dynamic loads, fatigue, and loss of power, thus reducing the vehicles efficiency. Also, vehicle sounds and vibrations may significantly affect the passengers' comfort. Sound is a part of a structure vibrational energy transmitted to its surrounding environment. Therefore, there is a direct relationship between sound and vibration.

Mohanty *et al.* (2000) used a CAE method for noise reduction in a truck cabin interior. The finite element (FE) and the boundary element method (BEM) were used to characterize the acoustic field of a truck cabin interior in terms of the natural frequencies and the mode shapes. Structural vibration responses of the cabin were computed for excitations at the cabin mounts in the frequency range from 50 to 250 Hz. Interior noise levels at the driver's right ear were determined using the boundary element method for excitations at the cabin mounts. A panel acous-

tic contribution analysis (PACA) was done to determine the structural areas of the cabin contributing most to the noise levels at the driver's right ear. Structure-borne noise was reduced in the cabin's interior by selecting and placing sound absorbing material at the appropriate locations in the cabin, as determined by PACA.

Li *et al.* (2008) employed active noise control methods to control the noise of the heavy truck's cabin interior. An interior noise field test for the heavy truck was performed, and frequencies of interior noise of this vehicle were analysed. Then the least squares lattice (LSL) algorithm was used as the signal processing algorithm of the controller, and a closed-loop control DSP system was developed. The residual signal at the driver's ear was used as a feedback signal. Lastly, the developed active noise control (ANC) system was loaded into the heavy truck cabin, and controlling the noise at the driver's ear for that truck at different driving speeds was attempted.

Antila *et al.* (2008) simulated an ANC system using noise data measured in a truck cabin. The data were treated in the simulation process with a control system model. The result was evaluated both numerically and by listening tests. The possible benefits of the proposed ANC system included less fatigue for driver and co-driver, no need for excessive noise insulation in the truck, and more comfortable driving conditions. The challenges in designing the system were its complexity, reliability, and potentially high price. These pros and cons were discussed in the paper, and a concept of the system realisation was given.

Bealko (2009) examined noise exposure inside haul truck cabins experienced during a typical workday with normal operator practices, the effect of noise-reduction features inside the cabin, and the consequence of disabling noise controls (unnecessary open doors/windows), and the significance of haul truck and cabin maintenance factors.

Lu *et al.* (2013) studied an adaptive active noise control (AANC) system of the interior truck cabin to reduce low-frequency noise. A normalisation Fractional Least Mean Square (FLMS) algorithm Simulink model was established in MATLAB/Simulink. Then taking it as the core, a feedforward adaptive active control system and a feedback adaptive active control system of the interior tuck cabin were established in MATLAB/Simulink. Considering the actual channel error effects on systems, the noise reduction effects of two adaptive active control systems were verified from the simulation results. Comparing the two adaptive active control systems showed that the feedforward adaptive active control system was more stable than the other one.

Ang *et al.* (2016) provided an overview of the existing industrial practices used for cabin noise control in various industries such as automotive, marine, aerospace, and defense. The current industrial practices pertaining to cabin noise control were discussed. Also, the potential of acoustic metamaterials was highlighted.

Saxena and Jadhav (2021) measured the interior noise and vibration on one of the light trucks, and a few dominant low-frequency noise booms were observed in the operation range. Modal analysis was done for the cabin at virtual and experimental levels, and a few modes were found close to these noise booms. Vibrations were measured across the cabin mounts, and it was found that the isolation of front mounts is not effective at lower frequencies. The mount design was modified to shift the natural frequency and improve the isolation behaviour at the lowest dominant frequency. Also, the interior noise and vibration measurement was carried out on the truck fitted with selected mounts, and substantial vibration, overall noise reduction, and drastic boom noise reduction were achieved.

Herein, the main cabin noise sources are investigated for an MB truck, including both vibrations and acoustic noises. Viscoelastic damping layers are employed to control vibration, while sound barriers, silicon adhesive, and foam are used to control noise. The effect of viscoelastic damping layers on the vibration amplitude of the cabin back wall is studied both theoretically and experimentally. Also, the impact of acoustic materials on cabin noise attenuation is experimentally investigated.

2. Main sources of cabin noise in an MB truck

Preliminary investigations indicate that the main sources of noise in an MB truck (including acoustic noise as well as vibration) can be categorised as follows:

- power system including engine, air inlet system, and exhaust system,
- driveline system including gearbox, driveshaft, and differential,
- tyre/road interaction.

The main sources of interior noise (the cabin noise) are the power system, driveline system, and tyre/road interactions. However, the noise generated by tyre/road interaction dominates the driveline system noise at speeds exceeding 80 km/h (Johnson, 1996).

2.1. Engine

The engine is the main source of interior noise in trucks. The engine vibration and engine noise are studied in the following sections (Taylor, 1982; SAE, 1992).

2.1.1. Engine vibration

The inertia of the engine's moving parts and the cylinder pressure changes create forces that cause en-

gine components' movement. As a result, vibration with variable frequency and amplitude covers the motor's whole structure, and it is called internal vibration. Internal engine vibration includes the flexural and torsional vibration of the crankshaft, the piston's torsional vibration, and the auxiliary systems such as the oil pump, water pump, and turbocharger. Continuous systems such as crankcase and crankshaft housing, pipes and tubes such as oil pipelines and exhaust have an important role in the internal vibration of the engine. The high-frequency vibration (noise) from combustion and gears involvement is also important (Taylor, 1982; SAE, 1992).

2.1.2. Engine noise

Engine noise includes mechanical, combustion, fuel injection, air inlet, and exhaust system noise (Taylor, 1982; Harris, 1991; Beranek, 1992).

Diesel engines have higher pressure rise rates than spark ignition engines, indicating higher importance of combustion noise in diesel engines than that in spark ignitions engines. The acoustic and vibrational properties of the parts related to the combustion phenomenon play an important role in the amount of combustion noise. Today, research shows that the most important noise in diesel engines is combustion noise. Experimental research shows that 106 BTU/hr of heat released by combustion can generate about 29 W of acoustic power (equivalent to a sound power level of 135 dB) (Taylor, 1982; Harris, 1991).

Exhaust in internal combustion engines is one of the main sources of noise. Noise is caused by the periodic release of gases from the exhaust manifold (Harris, 1991).

2.2. Driveline system

The driveline system in trucks has a more intense effect on the overall noise level than in passenger cars and should be considered. Driveline system noise includes gearbox, driveshaft, and differential noise. The differential noise has minor effects on the overall noise level (SAE, 1992).

2.3. Tyre/road noise

Since the speed of the truck is generally less than 80 km/h, the effect of the tyre/road on the cabin noise is less important (Johnson, 1996).

3. Noise and vibration control

The noise (including acoustic noise and vibration) control methods have been categorised into three categories: noise control at the source, noise control in the transmission path, and noise control at the receiver. This study considers the transmission path for both noise and vibration. In other words, the cabin noise is controlled in its transmission paths. The sound from different sources is transmitted into the cabin via air (airborne sound transmission) and structure (structure-borne sound transmission). Figure 1 shows principal noise sources and their transmission paths to the cabin. Also, Fig. 2 shows structure-borne and airborne paths of noise transmission to the cabin.

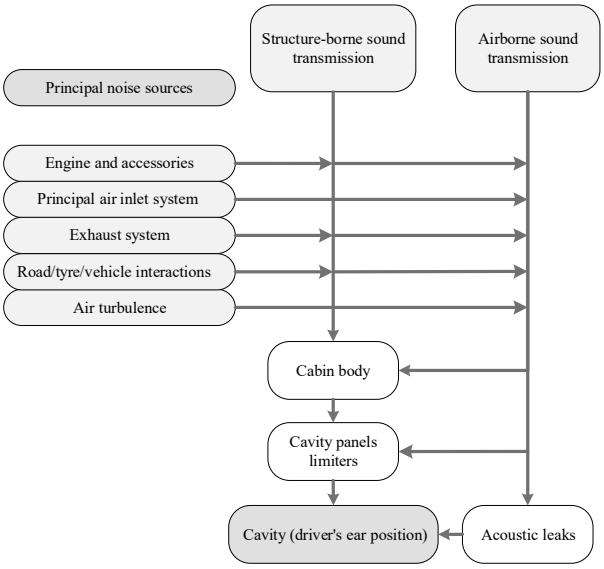


Fig. 1. Principal noise sources and their transmission paths to the cabin.

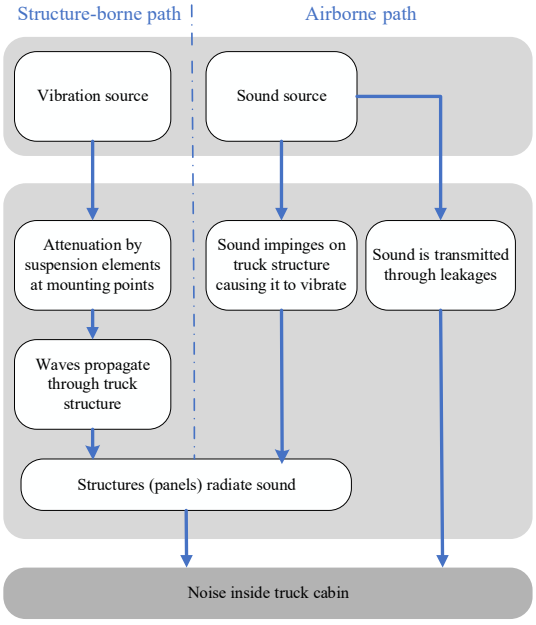


Fig. 2. Structure-borne and airborne paths of noise transmission to the cabin.

3.1. Vibration

In many machines, excited vibrations cover a wide range of frequencies. In this case, the conventional method of vibration control, i.e., separation of the

natural frequencies of the system from the excitation frequencies, is not possible. While dynamic vibration absorbers are not used in such a situation because it is possible to excite several natural modes, viscoelastic layer damping treatment is applied to control the vibration of a structure. Such layers reduce the structural vibration amplitude by energy dissipation, causing the overall noise level to be lower (Pujara, 1992; Barber, 1992). A layer damping can be used to control the vibration of the truck cabin, engine housing, cylinder head cover, and crankcase.

Researchers have studied two types of viscoelastic damping treatments: unconstrained layer damping (UCLD) treatment and constrained layer damping (CLD) treatment. Unconstrained layer (free layer) treatments are widely used in the automotive industry as additional layers on large sheet metal panels. The elongation between the supporting metallic and the viscoelastic due to the bending of supporting plates in the low-frequency range introduces the material damping. In the constrained layer treatments, a viscoelastic layer is placed between the vibrating structure and a solid plate (usually metal). In this method, most of the vibrational energy is lost due to the shear deformation of the viscoelastic layer (Jones, 1985; Nashif *et al.*, 1985; Malik, 1990). In this study, the unconstrained layer damping treatment is used.

The overall loss factor η_s of the structure treated can be obtained by Eq. (1) (Jones, 1985):

$$\eta_s = \frac{\eta_D}{1 + A/(Be)}, \tag{1}$$

where

$$A = \frac{(1-n^2e)^3 + [1+(2n+n^2)e]^3}{(1+ne)^3},$$

$$B = \frac{(1+2n+n^2e)^3 - (1-n^2e)^3}{(1+ne)^3},$$

$$n = h_D/h, \qquad e = E_D/E,$$

and where h is the thickness of the structure, h_D is the thickness of the layer damping, E is the Young modulus of the structure, E_D is the real part of the complex modulus of the layer damping, η_D is the loss factor of the layer damping, and η_s is the overall loss factor of the structure treated.

3.2. Noise

Sound barriers are used to block the transmission of airborne sound by providing mass to existing structures or hung as limp mass partitions. The performance of a sound barrier is measured in terms of its transmission loss (TL). In practical applications, the value of TL for a sound barrier is often expressed as the mass law: the more the surface density of a sound barrier, the higher its TL. Herein, a sound barrier, silicon adhesives, and foam are used to control noise.

4. Treatment and results

The engine, the exhaust system, and the air inlet system are the main sources of cabin noise in the MB truck. Figure 3 shows the MB truck cabin, the main sources of interior noise, and their transmission paths into the cabin. Engine noise is transmitted into the cabin by the air (airborne noise: A) through acoustic leaks and the cabin body (floor and firewall), and by the cabin structure (structure-borne noise: S) through the floor and the firewall. The exhaust system noise is transmitted into the cabin by the air through acoustic leaks and the floor, and by the cabin structure through the floor and back wall. The air inlet system noise is transmitted into the cabin by the air through acoustic leaks and firewall, and by the cabin structure through the firewall.

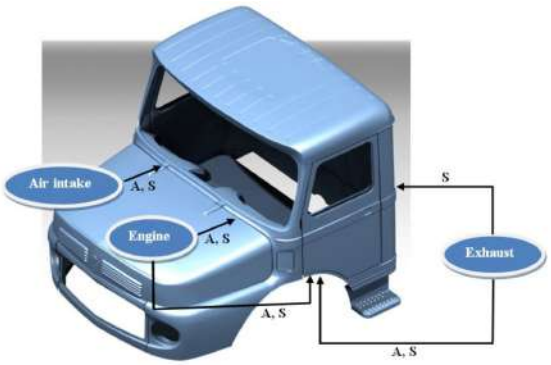


Fig. 3. MB truck cabin, its main sources of interior noise, and their transmission paths into the cabin.

Here, a viscoelastic layer damping is used on the back wall to control vibration, and acoustic materials are used on the floor and the firewall to control noise.

4.1. Viscoelastic damping layer

4.1.1. Theoretical model

A viscoelastic layer damping is used here to control vibration. The back wall vibration is considered and treated. A suitable viscoelastic material is selected, and the overall loss factor and the ratio of vibration amplitude reduction are calculated for the structure treated (back wall with viscoelastic layer installed).

The behaviour of viscoelastic materials is a function of temperature and frequency. First, the structure's operating temperature range and vibration frequency range (back wall) must be determined. Then, the material with the maximum value $E_D\eta_D$ is selected in that temperature and frequency range. The temperature of the back wall varies in the range of 35–45°C, and the

structure's natural frequencies are calculated to determine the frequency range of the back wall vibration.

The truck's cabin is designed such that its back wall is nearly a flat surface. So, the back wall can be appropriately modelled as a rectangular sheet. The natural frequencies of a rectangular sheet are calculated as follows (Pujara, 1992):

$$\omega_{pq} = \pi^2 \sqrt{\frac{gD}{\rho h}\left[\left(\frac{p}{a}\right)^2 + \left(\frac{q}{b}\right)^2\right]}, \tag{2}$$

where $D = Eh^3/[12(1-\nu^2)]$, a, b, and h are length, width, and thickness of the sheet, respectively, ρ, E, and ν are density, Young modulus, and Poisson's ratio of the plate, respectively, g is the gravitational acceleration, and ω_{pq} are the natural frequencies of a rectangular sheet.

The back wall plate in the MB truck is made of USt 37-2, and its dimensions and properties are listed in Table 1. From Eq. (2), natural frequencies corresponding to the back wall are presented in Table 2.

Table 1. Dimensions and physical and mechanical properties of the back wall plate in the MB truck.

a [m]	b [m]	h [m]	ρ [kg/m^3]	E [GPa]	ν	D
1.96	1.45	1.5×10^{-3}	7850	210	0.3	64.9

Table 2. Natural frequencies of the back wall.

f_{11} [Hz]	f_{12} [Hz]	f_{21} [Hz]	f_{22} [Hz]	f_{23} [Hz]	f_{32} [Hz]
10	17	14	19.8	26.7	23.8

From Table 2, the frequency range of the back wall vibration 10–25 Hz is considered. According to the temperature and frequency range, LD-400 viscoelastic material is selected as the most suitable material, as it has the highest $E_D\eta_D$ value over the mentioned temperature and frequency ranges. In Table 3, the E_D, η_D, and $E_D\eta_D$ values are tabulated for frequencies of 10, 15, 20, and 25 Hz and temperatures of 35, 40, and 45°C. The density of the viscoelastic layers is $\rho_D = 1500$ kg/m^2.

Table 3. Properties of the LD-400 viscoelastic layer damping (AFML data).

Frequency [Hz]	T [°C]	η_D	E_D [GPa]	$E_D\eta_D$
10	35	0.512	1.2986	0.6649
	40	0.443	1.0517	0.4660
	45	0.413	0.8648	0.3572
15	35	0.521	1.4213	0.7405
	40	0.463	1.1691	0.5413
	45	0.412	0.9607	0.3958
20	35	0.529	1.4828	0.7844
	40	0.471	1.2041	0.5671
	45	0.418	0.9903	0.4139

The overall loss factor can be calculated by Eq. (1). The E_D and η_D values are extracted from Table 3. Here, the temperature and frequency are equal to 40°C and 10 Hz (the first natural frequency). Table 4 presents the overall loss factor of the structure treated for $n = 2$, 4, 6, and 10.

Table 4. Overall loss factor of the structure treated.

n	2	4	6	8	10
η_s	0.1034	0.2812	0.3653	0.3921	0.4062

Assuming that the first mode of vibration is important, the structure treated is modelled as a simple one degree of freedom system. Values of the vibration amplitude reduction ratio of the structure are computed as follows (Malik, 1990):

$$\frac{X}{Y} = \left[\frac{1+(2\xi r)^2}{(1-r^2)^2+(2\xi r)^2}\right]^{1/2}, \tag{3}$$

where X/Y represents the ratio of vibration amplitude reduction, r is the frequency ratio, and ξ is the damping ratio equal to $\eta_s/2$ (Malik, 1990). Figure 4 shows the values of X/Y for $n = 2$, 4, 6, and 10.

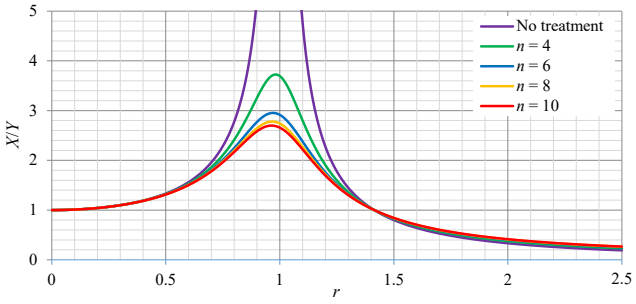


Fig. 4. Ratio of the vibration amplitude reduction for the structure treated.

4.1.2. Experimental work

In order to calculate the vibration amplitude, the structure is excited, and the acceleration response is obtained. A stringer that has been installed on the head of a shaker (B&K Type 4808) does the excitation; a force transducer (Endevco Model 2311-100) on the head of the shaker, and an accelerometer (Endevco Model 65-100) are used to sense the input and output of the system, respectively. A signal analyser (B&K Type 3560-B), a power amplifier (B&K Type 2719), and a PC equipped with PULSE 8 software were utilised for data acquisition and signal processing. Figure 5 shows a schematic sketch of the experimental setup used for the measurements.

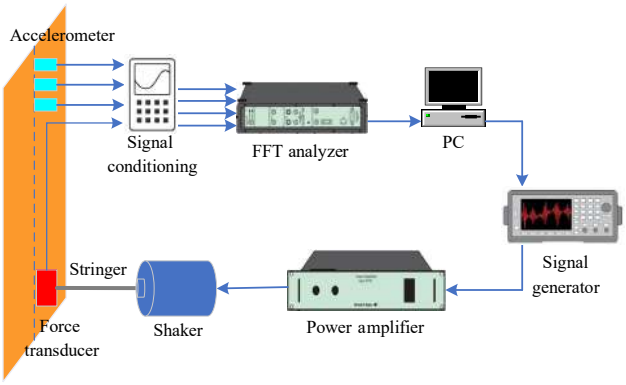


Fig. 5. Schematic sketch of the experimental setup used for the measurements.

First, natural frequencies of the structure are measured. In Table 5, the experimental and theoretical values of the natural frequency are shown. Then, the vibration amplitude of the structure is determined before and after the treatment. The excitation frequency is normalised with respect to the first natural frequency. Figure 6 shows experimental and theoretical values of the ratio of the vibration amplitude reduction for the structure treated for $n = 4$.

Table 5. Natural frequencies of the back wall.

Mode number	f_{11} [Hz]	f_{12} [Hz]	f_{21} [Hz]	f_{22} [Hz]	f_{23} [Hz]	f_{32} [Hz]
Theoretical	10	17	14	19.8	26.7	23.8
Experimental	9.5	16.1	13.2	18.7	25.1	22.8

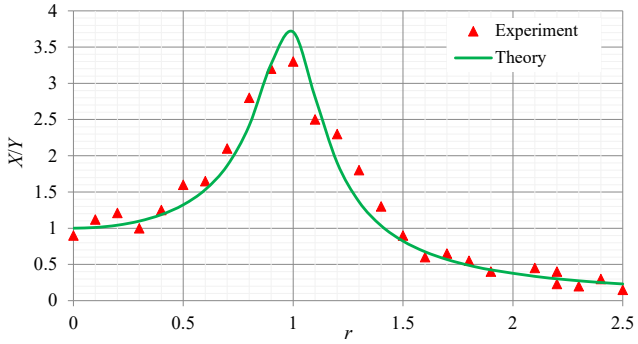


Fig. 6. Experimental and theoretical values of the ratio of the vibration amplitude reduction for the structure treated, $n = 4$.

4.2. Acoustic materials

Herein, acoustic materials, including sound barrier, Silicon adhesives, and polyurethane foam are used. Rubber sheets, namely "Genaral seal & Panchari", made in Iran, are used as a sound barrier on the floor and firewall. This material has a density of 5 kg/m^2, with its TL_E values presented in Table 6. These values are obtained by testing a squared specimen with an area of 900 cm^2. Also, the TL_M values calculated by mass law are shown in Table 6.

Table 6. TL values corresponding to mass law (TL_M) and testing (TL_E) of the sound barrier.

Frequency [Hz]	TL [dB]	
	Experimental	Mass law
125	14	9
250	16	15
500	22	21
1000	26	27
2000	27	33
4000	30	39
8000	36	45

Silicone adhesives are here used in tape and liquid forms. The silicone tape is used on the floor and firewall, and the liquid silicone is used to treat leakages of the structure. Also, polyurethane foam is injected into the side member and other noise transmission paths.

4.2.1. Experimental work and results

The experiments are performed according to standards ISO 5128 (1980) and SAE-J336 (2011). Figure 7 shows a schematic sketch of the experimental setup used for the measurements. In addition to the shown equipment, a loudspeaker is used as an external noise source.

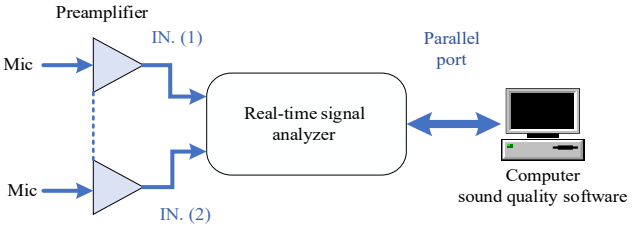


Fig. 7. Schematic sketch of the experimental setup used for the measurements.

Herein, a B&K 4155 1/2 free-field microphone is used. The noise level measured inside the cabin changes drastically with the microphone's position, and the microphone must be able to accurately describe the noise sensed by the driver and his/her assistant. The microphone's position is determined via ISO 5128 (see Fig. 8). A Norsonic microphone calibrator type 1251 is used for calibration. Also, a Norsonic preamplifier type 1201 is used here. A four-channel signal analyser (B&K Type 3560-C) and a PC equipped with PULSE 8 software are utilised for data acquisition and signal processing.

The test placement must be a place with minimum background noise and distance from direct noise sources and reflective surfaces exceeding 15 m. The

Fig. 8. Position of the microphone inside the cabin.

background noise level will be measured separately and subtracted from the overall noise level in places where the background noise level is high.

Three tests, including the loudspeaker, engine, and road, are performed in 6 steps. In the loudspeaker test, the engine is turned off, and a loudspeaker is used as the noise source. The loudspeaker is a Norsonic reference sound source with a weighted sound power output: 94 dB re 1 pW (50 Hz line frequency). The loudspeaker is located 1.1 m from the front axle on the centreline of the two axles. In the engine test, the engine is operating at a speed of 1500 rpm for 5–10 seconds (the truck is stationary). In the road test, the truck runs 200 m at a constant speed of 30 km/h in 3rd gear. The 6 test steps are:

Step 1: Performing loudspeaker, engine, and road tests before any treatment.

Step 2: Performing loudspeaker and engine tests after removing the carpet from the cabin floor.

Step 3: Performing engine and loudspeaker tests once the sound barrier and silicon adhesives cover the firewall leaks.

Step 4: Performing loudspeaker and engine tests once the sound barrier covers the floor.

Step 5: Performing the road test after all leaks and the floor are covered by the sound barrier and silicon adhesives.

Step 6: Performing the road test after complete treatment, including the coverage of all leaks in the floor and firewall by the sound barrier and silicon adhesive, and foam injection into the side member.

The test results are shown in Figs. 9 to 15. The experimental SPL values are reported in the frequency range of 20 Hz–20 kHz based on a 1/3 octave filter. Figure 9 shows the measured sound pressure level (SPL) inside the cabin for the loudspeaker test (steps 2 and 3). Knowing that the structure rather than the air transmits low-frequency sound waves (compared to high-frequency sound waves), one may see that coverage of leaks does not affect the noise level reduction for frequencies lower than 1000 Hz. The maximum noise

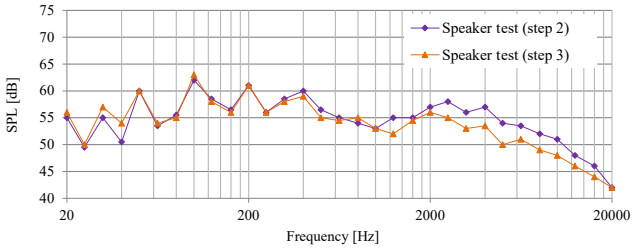


Fig. 9. Measured SPL inside the cabin; the loudspeaker test (steps 2 and 3).

reduction of 4 dB is seen at 5000 Hz. In addition, the carpet has little impact on cabin noise reduction.

Figure 10 shows the measured SPL inside the cabin for the loudspeaker test (steps 2 and 4). Coverage of the cabin floor with the sound barrier reduces the cabin noise level for frequencies higher than 1000 Hz (a maximum of 6 dB at 5000 Hz).

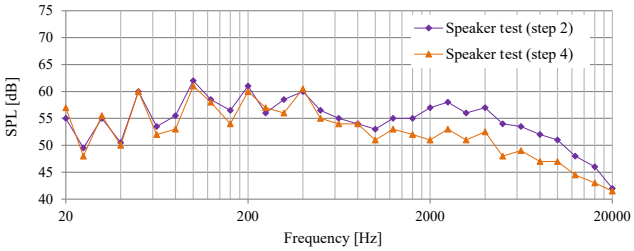


Fig. 10. Measured SPL inside the cabin; the loudspeaker test (steps 2 and 4).

According to Fig. 11, the cabin carpet has very small contribution to the sound pressure level inside the cabin (a maximum of 2 dB at 2500 Hz).

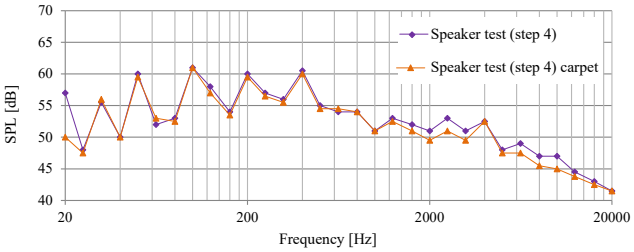


Fig. 11. Impact of the carpet on the SPL inside the cabin.

Figure 12 shows the measured SPL inside the cabin for the engine test (steps 1 and 3). The results indicate

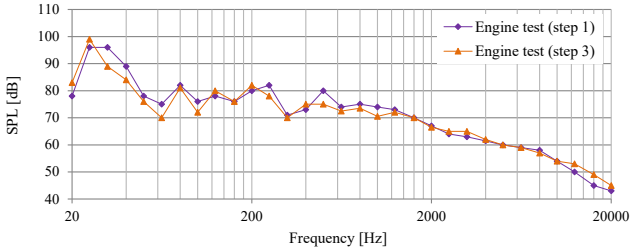


Fig. 12. Measured SPL inside the cabin; the engine test (steps 1 and 3).

that even when all the leaks are entirely covered, the engine noise is transmitted into the cabin by the structure itself (structure-borne noise).

Figure 13 shows the measured SPL inside the cabin for the engine test (steps 1 and 2). The impact of the cabin carpet on the cabin noise attenuation is significant for the frequencies higher than 1000 Hz. An average reduction of 13 dB is observed in the SPL inside the cabin. However, the carpet does not affect the cabin noise reduction for the frequencies lower than 1000 Hz. In this frequency range, the engine noise is transmitted into the cabin by the structure itself.

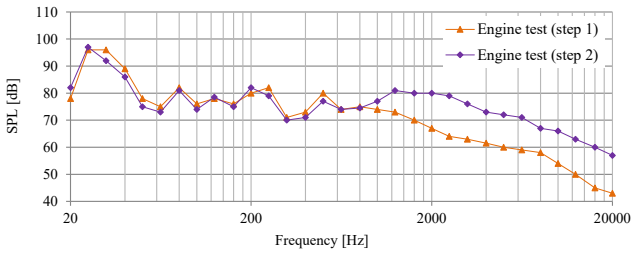


Fig. 13. Measured SPL inside the cabin; the engine test (steps 1 and 2).

Figure 14 shows the SPL inside the cabin for the engine test (steps 2 and 4). For the frequencies above 1000 Hz, coverage of the floor with the sound barrier effectively reduces the SPL inside the cabin. An average reduction of 17 dB is observed in the SPL inside the cabin. However, coverage of the floor with acoustic materials does not affect the cabin noise reduction for the frequencies lower than 1000 Hz. The engine noise is transmitted into the cabin by the structure itself in this frequency range.

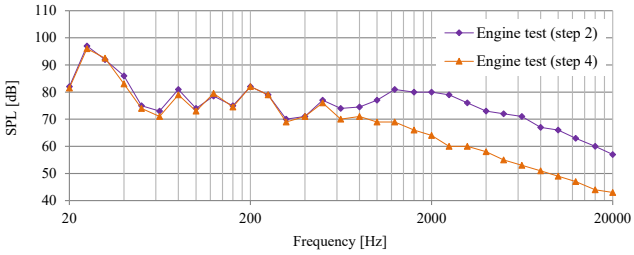


Fig. 14. Measured SPL inside the cabin; the engine test (steps 2 and 4).

Figure 15 shows the measured SPL inside the cabin for the road test (steps 1, 5, and 6) in dBA. The overall noise level at the microphone position decreases from 80 to 70 dBA. The final value is mainly influenced by the peak of acoustic energy in the frequency band of 250 Hz, a frequency component related to the engine rotation. In order to obtain a further reduction of the overall dBA level, it will be necessary to investigate this peak first, probably due to some structure-born vibration. Foam injection has a relatively good effect at low frequencies.

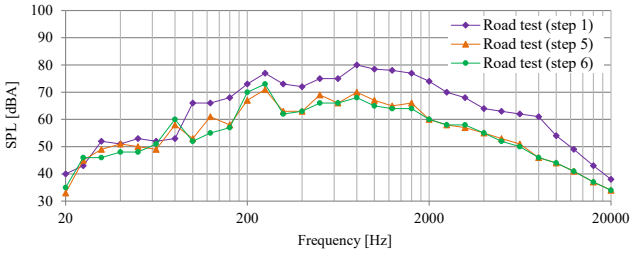


Fig. 15. Measured SPL inside the cabin; the road test (steps 1, 5, and 6) in dBA.

5. Conclusions

Investigations were made to control the MB truck cabin interior noise by reducing noise in the transmission path. A viscoelastic layer damping material was used to reduce the vibration amplitude of the cabin back wall. Computational results were then compared with the values obtained by the experimental modal analysis results. Good agreement was found between the theoretical and experimental results. A sound barrier, silicon adhesive, and foam were utilised for noise control in the transmission path. The effectiveness of the mentioned acoustic materials on cabin noise reduction was experimentally evaluated. The measurements focused on carpet impact, coverage of leaks in the firewall (by sound barriers and silicon adhesive), and floor coverage (by sound barriers). By looking at the graphs, it is possible to understand the influence of the different trim parts and the modifications made, but it is also interesting to see the peaks in the spectra. Those peaks are probably related to the presence of excited acoustic modes, and this phenomenon should be investigated to verify this. The excitation of acoustic modes in acoustic cavities of the vehicle by the engine running is a common problem in noise reduction. However, it requires specific investigations and acoustic modelling to understand what can be done.

The loudspeaker test results show that the carpet impact, coverage of leaks in the firewall by the sound barrier and silicone adhesive, and coverage of the floor by the sound barrier have a significant effect on the cabin noise reduction for the frequencies above 1000 Hz. For the frequencies lower than 1000 Hz, a significant portion of the noise is transmitted into the cabin by the structure itself; therefore, those treatments do not reduce cabin noise.

The engine test results show that the coverage of the firewall leaks does not affect the cabin noise reduction. It seems that the engine noise is still transmitted into the cabin by the structure itself (structure-borne noise). However, for the frequencies above 1000 Hz, the carpet and coverage of the floor have a significant effect on the cabin noise reduction. For the frequencies lower than 1000 Hz, those treatments do not affect the cabin noise reduction.

The road test results show that a complete treatment, including the coverage of the floor and firewall leaks by acoustic materials and silicone adhesive and foam injection into the side member, has a significant effect on the cabin noise reduction. The overall noise level at the microphone position decreases from 80 to 70 dBA. The final value is mainly influenced by the peak of acoustic energy in the frequency band of 250 Hz, a frequency component related to the engine rotation. In order to obtain a further reduction of the overall dBA level, it will be necessary to investigate this peak first, probably due to some structure-born vibration. Foam injection has a relatively good effect at low frequencies.

References

1. Ang L.Y.L., Koh Y.K., Lee H.P. (2016), Acoustic metamaterials: a potential for cabin noise control in automobiles and armored vehicles, *International Journal of Applied Mechanics*, **8**(5), doi: 10.1142/S1758825116500721.
2. Antila M., Hao Y., Lankila A., Yu J. (2008), Possibilities and benefits of active noise control (ANC) in truck cabins, [in:] *Proceedings: 37th International Congress and Exhibition on Noise Control Engineering, INTER-NOISE 2008*, **1**: 444–449, Shanghai, China.
3. Barber A. (1992), *Handbook of Noise and Vibration Control*, 6th ed., Elsevier Science, Oxford.
4. Bealko S.B. (2009), Mining publication: Mining haul truck cab noise: An evaluation of three acoustical environments, *The National Institute for Occupational Safety and Health (NIOSH)*, **61**(10): 36–42.
5. Beranek L.L. (1992), *Noise and Vibration Control*, John Wiley.
6. Harris C.M. (1991), *Handbook of Acoustical Measurements and Noise Control*, 3rd ed., McGraw-Hill, New York.
7. ISO 5128 Standard (1980), *Acoustics-Measurement of Noise Inside Motor Vehicles*.
8. Johnson E. (1996), *Tires and Handling*, Society of Automotive Engineers.
9. Jones D.I.G. (1985), *Damping Materials for Vibration & Sound Control*, John Wiley, New York.
10. Li H.B., Liu H.G., Liu J.C., Shangguan Y.F. (2008), Active noise control of the heavy truck interior cab, *Journal of Beijing Institute of Technology*, **17**(4): 400–404.
11. Lu Q.C., Li H.B., Huang H. (2013), Simulation of active noise control system on the truck interior cab, *Advanced Materials Research*, **798–799**: 443–447, doi: 10.4028/www.scientific.net/AMR.798-799.443.
12. Malik A.K. (1990), *Principle of Vibration Control*, Affiliated East-West Press, India.
13. Mohanty A.R., Pierre B.D., Suruli-Narayanasami P. (2000), Structure-borne noise reduction in a truck cab interior using numerical techniques, *Applied Acoustics*, **59**(1): 1–17, doi: 10.1016/S0003-682X(99)00018-3.
14. Nashif A.D., Jones D.I.G., Henderson J.P. (1985), *Vibration Damping*, John Wiley, New York.
15. Pujara K. (1992), *Vibration & Noise for Engineers*, 4th ed., Dhanpat Rai & Sons, Delhi.
16. SAE-J336 Standard (2011), *Sound Level for Truck Cab Interior*, SAE International.
17. Saxena S., Jadhav S. (2021), Truck front cabin mount tuning for cabin noise boom, overall interior noise and vibration reduction, [in:] *Symposium on International Automotive Technology*, Pune, India, doi: 10.4271/2021-26-0286.
18. Society of Automotive Engineers (SAE) (1992), *Truck Systems Design Handbook*, Society of Automotive Engineers.
19. Taylor C.F. (1982), *The Internal Combustion Engine in Theory and Practice*, Vol. 2, 5th ed., Massachusetts Institute of Technology.

Archives of Acoustics Vol. 48, No. 1, pp. 103–126 (2023), doi: 10.24425/aoa.2023.144265

Research Paper

Evaluation and Analysis of Long-term Environmental Noise Levels in 7 Major Cities of India

Naveen GARG[(1),(2)*], Saurabh KUMAR[(1),(2)], Chitra GAUTAM[(1),(2)], Vishal GANDHI[(3)], Nalin Kumar GUPTA[(3)]

[(1)] *CSIR-National Physical Laboratory*
New Delhi 110 012, India

[(2)] *Academy of Scientific and Innovative Research*
Ghaziabad 201 002, India

[(3)] *Central Pollution Control Board*
Parivesh Bhawan, Delhi, India

[*]Corresponding Author e-mail: ngarg@nplindia.org

(*received January 1, 2022; accepted October 12, 2022*)

This paper describes the long-term noise monitoring data for ten consecutive years (2011–2020) acquired from the diversified National Ambient Noise Monitoring Network (NANMN) set up across 7 major cities of India and consisting of 70 stations for continuous noise monitoring throughout the year. The annual average ambient noise levels observed throughout these ten years in 70 locations under study, including 25 locations in commercial zones, 12 in industrial, 16 in residential, and 17 in silence zones, are described. The noise scenario in these ten years is compared and analyzed. It is observed that no site in residential or silence zones meets the ambient noise limits for the past ten years. The study suggests guidelines for a policy framework for environmental noise management and control to regulate noise pollution in Indian cities.

Keywords: National Ambient Noise Monitoring Network (NANMN); day equivalent sound level (L_{day}); night equivalent sound level (L_{night}).

1. Introduction

Environmental noise exposure and associated health effects have become a serious concern globally. The Guideline Development Group (World Health Organization [WHO], 2018) has enunciated to use the term "environmental noise" and defined it as "noise emitted from all sources except sources of occupational noise exposure in workplaces" (WHO, 2018). The recommended limits for noise exposure from road traffic noise are 53 dB L_{den} and 45 dB L_{night}, and for noise exposure from aircraft noise the recommended limits are 45 dB L_{den} and 40 dB L_{night} (WHO, 2018). Every nation is concerned about the health effects of noise emitted from the increasing number of vehicles on the roads and aircraft noise in residential areas near the airports (Babisch *et al.*, 2005; WHO, 2011; van Kempen *et al.*, 2018; Hansell *et al.*, 2013; Schmidt *et al.*, 2015). It is thus imperative to adopt long-term noise monitoring strategies to assess and control the accentuated environmental noise levels and plan for suitable noise abatement measures. The European Environmental Noise Directive 2002/49/EC requires that the values of acoustic parameters L_{den} and L_{night} are representative of a year period (The European Parliament and the Council of the European Union, 2002). The Directive articulates the assessment methods for the purpose of strategic noise mapping and the corresponding action plans, which imply the use of harmonized indicators and evaluation methods as well as criteria for noise mapping. The noise maps should present noise levels expressed in the harmonized indicators L_{den} and L_{night}. The WHO estimates that, in Western Europe alone, 1 million healthy life years are lost annually

to environmental noise (WHO, 2011). Other estimates put the external cost of noise-related health issues in the European Union between 0.3–0.4% of GDP (European Conference of Ministers of Transport [EMCT], 1998) and 0.2% in Japan (Mizutani *et al.*, 2011). Some studies have correlated the relationship between environmental noise and real estate markets, with housing prices falling with the increase in noise levels (Nelson, 1982; Theebe, 2004). Thus, it is imperative from an Indian perspective to monitor the environmental noise levels in citiesand devise Noise Action Plans (NAPs) for abatement and control of noise pollution. The Central Pollution Control Board (CPCB) in India initiated the process of developing the National Ambient Noise Monitoring Network (NANMN), through which it was decided to include ambient noise as a regular parameter for monitoring in specified urban areas (CPCB, 2011; 2012). The real-time noise monitoring network, the NANMN program, was established with the objective of getting real-time continuous noise monitoring data. The present work discusses the noise monitoring data from 70 sites located in 7 major cities for the past ten years (2011–2020) with the following objectives:

- ascertain the noise scenario and the annual average ambient noise levels of 70 sites for the ten years under study and compare them with ambient noise standards of India (Table 1 – Appendix);
- inculcate the awareness of the general public towards the status of noise pollution and dissemination of information publicly through the website (CPCB, n.d.);
- ascertain the Most Exposed Urban Sites (MEUS) among the 70 sites and suggest the need for Noise Action Plans required, if any;
- ascertain to what extent the residential and silence zone sites meet the current ambient noise standards;
- analyze the difference of L_{day} and L_{night} levels to ascertain the severity of night noise levels as compared to the day levels;
- annual increment or decrement in the ambient noise levels for each of these sites in the decade under study for forecasting the future noise scenario;
- recommendations on policy framework for reducing the noise pollution levels in Indian cities based on the long-term evaluation and analysis of noise monitoring data for the 70 sites of 7 metropolitan cities.

It may be noted that although the installation of 10 noise monitoring stations for each city is insufficient to represent the noise environment of the concerned cities, yet the present study is focused on the evaluation and analysis of continuous long-term noise levels obtained from these 70 stations to ascertain and analyze the status of ambient noise levels and planning for suitable measures to control them. The present study shall be very helpful for understanding the noise scenario, analyzing the status of compliance of sites in each zone with the ambient noise standards, and planning for suitable measures and action plans for noise abatement and control in metropolitan cities of India. The study shall be helpful for pollution control bodies and planning and development authorities in managing and controlling environmental noise levels in the metropolitan cities of India.

2. Materials and methods

The diversified NANMN project was established in 2011 covering 70 stations in 7 major cities of the country, namely, Bengaluru, Chennai, Delhi, Hyderabad, Kolkata, Lucknow, and Mumbai. The details of 70 locations under study established in 7 cities of India, with each city having 10 noise monitoring stations, are shown in Fig. 1. The 70 locations cover 25 commercial sites, 16 residential sites, 17 sites in silence zones, and 12 sites in industrial zones. The Noise Monitoring Terminals (NMTs) manufactured and installed by Geónica Earth Sciences, Spain, have been discussed in detail earlier (Garg *et al.*, 2016; 2017b; 2017c; 2017d). In addition, a website application (CPCB, n.d.) has been developed to disseminate the data in real time to the public to generate awareness. The Noise Monitoring Network so established is unique and one of the largest noise monitoring networks of its kind across the globe.

In the present study, day equivalent levels L_{day}, and night equivalent levels L_{night} were acquired through the CPCB website and analyzed for each year from 2011 to 2020. In order to compare the ambient noise scenario of the noise monitoring locations with international guidelines and several studies reported so far (WHO, 2009; Babisch, 2002), the study also reports day-night average sound levels L_{dn}, and 24-hour equivalent sound levels $L_{\mathrm{Aeq,24h}}$. The day equivalent levels are the average equivalent sound levels of 16 hours duration of the day from 06:00 AM to 10:00 PM and the night equivalent levels are the average equivalent sound levels of 8 hours duration from 10:00 PM to 06:00 AM (Ministry of Environment & Forests, 2000).

The average day and night equivalent sound levels for each year are calculated as (Garg *et al.* 2016):

$$L_{\mathrm{day},n} = 10\log_{10}\left[\frac{1}{n}\sum_{i=1}^{n}10^{0.1(L_{\mathrm{day},i})}\right], \tag{1}$$

$$L_{\mathrm{night},n} = 10\log_{10}\left[\frac{1}{n}\sum_{i=1}^{n}10^{0.1(L_{\mathrm{night},i})}\right], \tag{2}$$

where n is the number of days in the year, and $L_{\mathrm{day},i}$ and $L_{\mathrm{night},i}$ are the day and night equivalent sound pressure levels of the i-th day of the year, respectively.

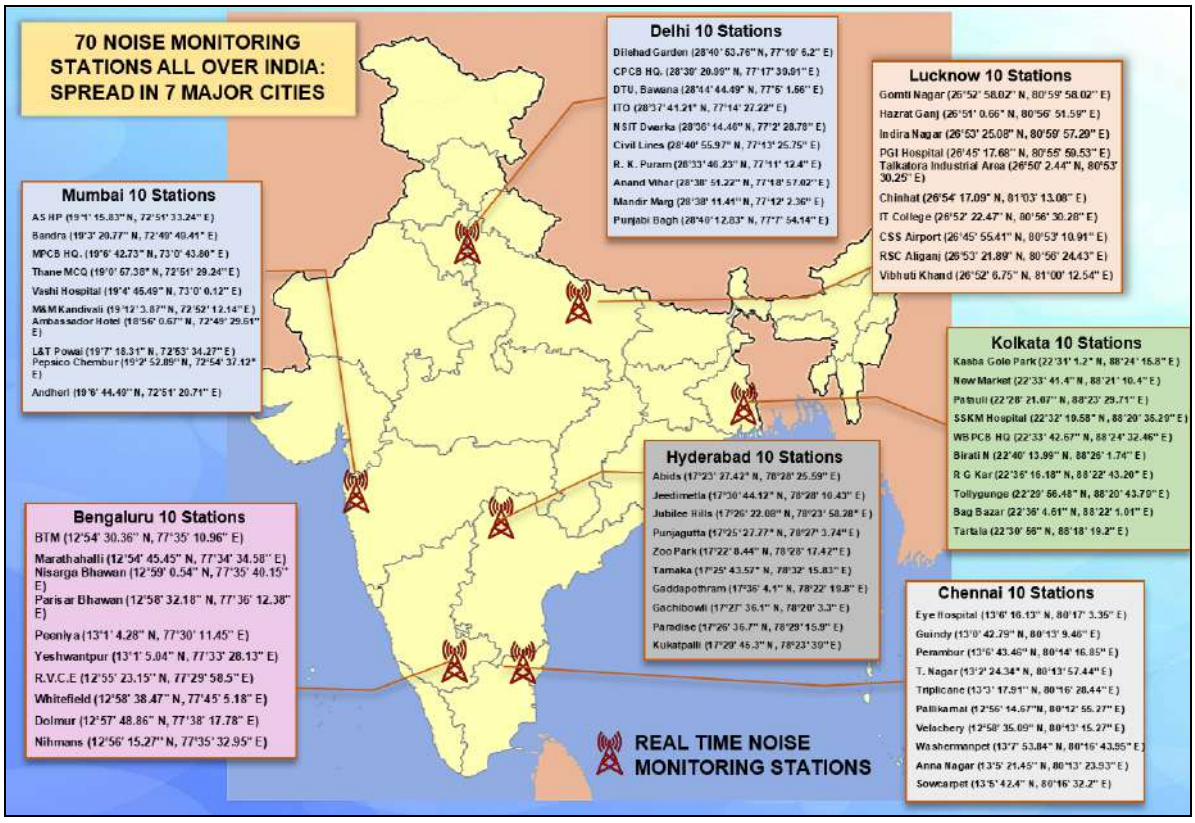


Fig. 1. Noise monitoring stations at 70 sites in 7 major cities of India installed by the CPCB, India for continuous noise monitoring throughout the year.

The 24-hour equivalent sound pressure levels are the average equivalent sound levels of 24 hours of a single day. Also, the annually-averaged day-night average sound pressure levels of the 70 noise monitoring sites are calculated (Garg *et al.*, 2016):

$$L_{\mathrm{dn}} = 10\log_{10}\left[\frac{16\cdot 10^{L_{\mathrm{day}}/10} + 8\cdot 10^{(L_{\mathrm{night}}+10)/10}}{24}\right], \quad (3)$$

where L_{day} and L_{night} are the day and night equivalent noise levels per year, respectively.

The standard deviation associated with the day equivalent sound levels of the noise monitoring sites is calculated as:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(L_{\mathrm{day},i} - \overline{L_{\mathrm{day}}})}{n-1}}, \quad (4)$$

where $\overline{L_{\mathrm{day}}}$ is the average of the day equivalent sound levels per year. The standard deviation for the average night equivalent sound levels for each noise monitoring site is also calculated in the same way.

The study also analyzes the noise exceedance factor (EF) calculated in each zone, as shown in Eq. (5) (Chowdhury *et al.*, 2016), in order to assess the environmental noise pollution scenario of different zones. The noise EF of a site is defined as the ratio of the ambient noise level of the site and the permissible noise level limit of the zone in which the site is located. The average noise limit exceedance factor (NEF) for all the sites lying in one zone is called the average exceedance factor (AEF):

$$\mathrm{EF} = \frac{L_0}{L_p}, \quad (5)$$

where L_0 is the observed ambient noise level, and L_p is the legally permissible limit recommended by the CPCB, India (Table 1 – Appendix).

The study undertakes a zone-wise and site-wise analysis in order to analyze the noise scenario at various sites and a specific trend of noise scenario in the past ten years.

3. Results and discussion

Tables 2 and 3 (Appendix) show the details of the annual average ambient day and night equivalent levels monitored for the 35 stations installed in 7 cities from 2011 to 2020 (CPCB, 2015a; 2015b; 2016; 2018; Garg, 2022). Table 4 (Appendix) shows the details of the annual average day and night equivalent sound levels for the past five years (2015–2020) for the additional 35 noise monitoring stations in 7 cities, which were installed in November 2014 (CPCB, 2015a; 2015b; 2016; 2018; Garg, 2022). Figures 2 and 3 show the

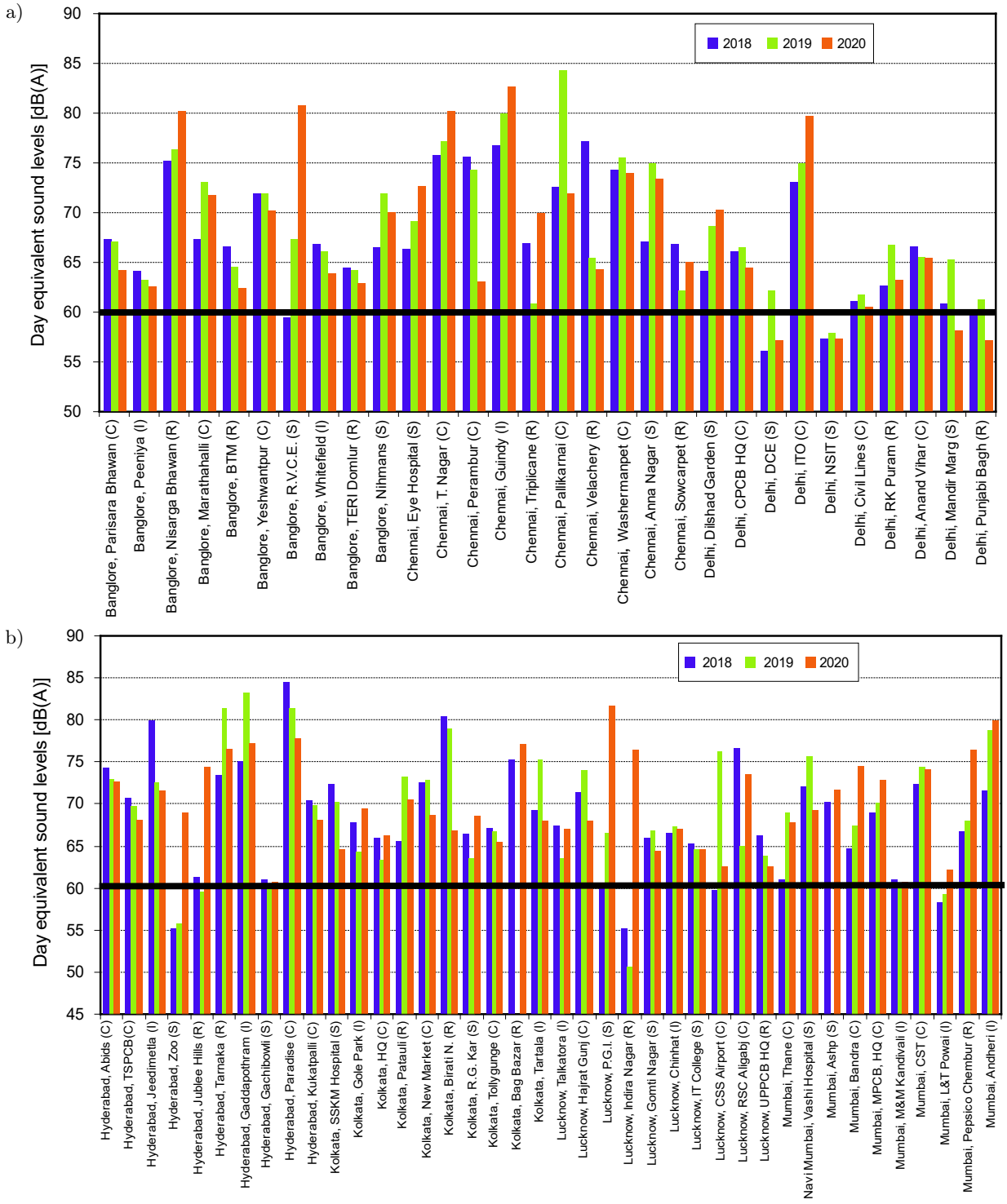


Fig. 2. Annual night equivalent average sound levels L_{night} [dB(A)] observed for the three years (2018–2020) for 7 major cities (Garg, 2022).

annual average day and night equivalent noise levels for the past three years (2018-2020) for the 70, sites in the 7 major cities of India (Garg, 2022). The 60 dB(A) L_{day} limit can be seen as NOAEL (no observed adverse effect level) for the correlation between road traffic noise and myocardial infarction (MI); the risk of MI increases incessantly for noise levels higher than 60 dB(A) (WHO, 2009; Babisch, 2002). It can be observed that the 60 dB L_{day} limit was met by only 4 sites in 2020, while 7 sites met the limit in 2019, and 9 sites in 2018. The Interim Target (IT) of 55 dB L_{night}, as recommended in the EU Night Noise Guidelines report (WHO, 2009), was met by only 4 sites in 2020, 7 sites in 2019, and 8 sites in 2018. The major-

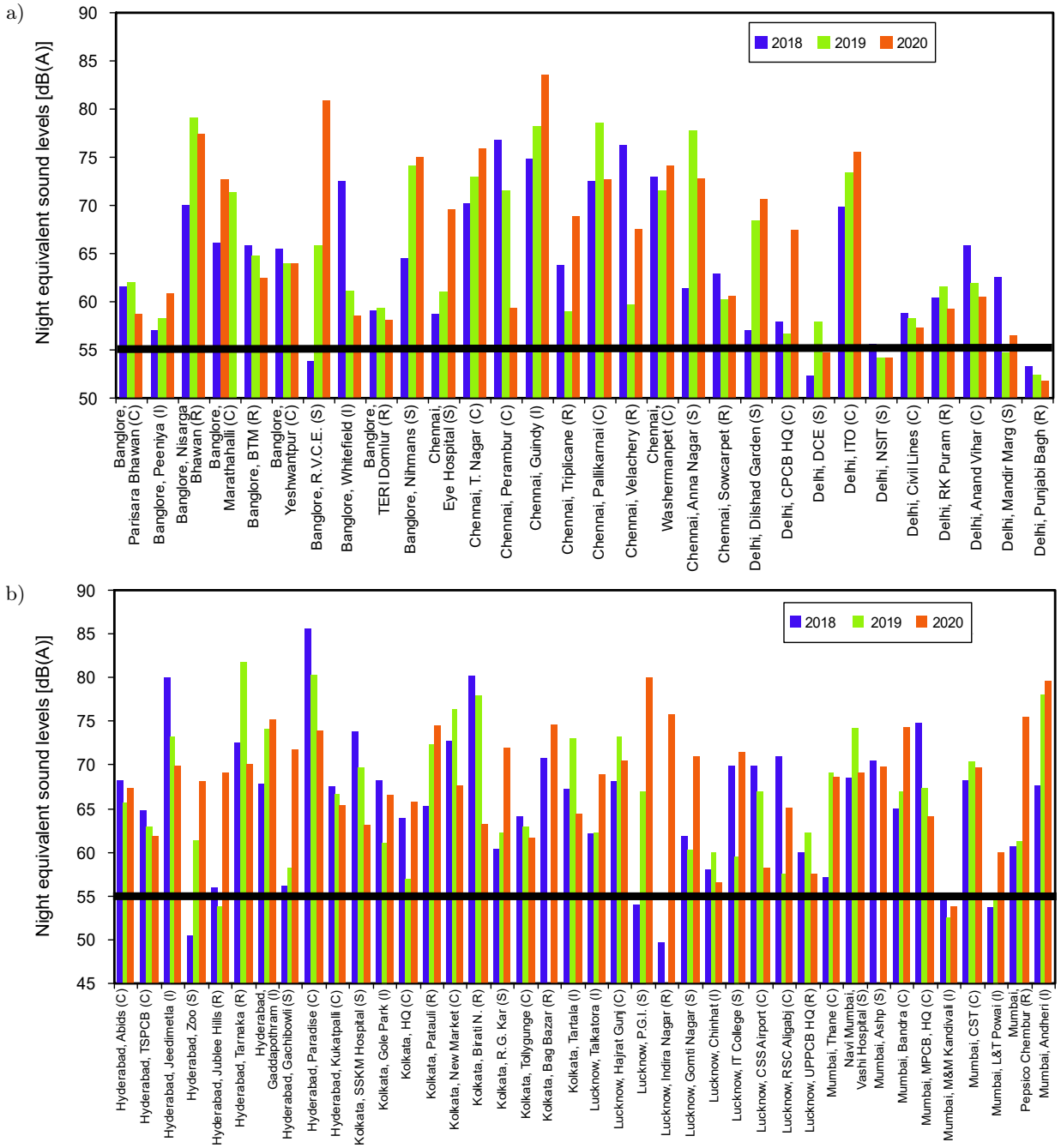


Fig. 3. Annual night equivalent average sound levels L_{night} [dB(A)] observed for the three years (2018–2020) for 7 major cities (Garg, 2022).

ity of sites (7 out of 10) in Chennai city showed day sound levels above 70 dB(A) and 6 sites showed night equivalent sound levels above 70 dB(A). However, in Delhi, only 2 sites showed day and night equivalent levels ≥70 dB(A). Also, in Mumbai and Hyderabad, 6 sites showed day equivalent levels ≥70 dB(A) and 5 sites showed night equivalent levels ≥70 dB(A). Overall, 32 sites (45.7%) comprising of 12 commercial, 5 industrial, 7 in silence zone, and 8 residential were observed to be the MEUS with day equivalent sound levels ≥70 dB(A). Thirty sites (42.9%) comprising of 9 commercial, 4 industrial, 11 in silence zone, and 6 residential were observed to be the MEUS with night equivalent sound levels ≥70 dB(A). Table 5 (Appendix) shows the frequency distribution of the noise descriptors: annual average day equivalent levels L_{day}, annual average night equivalent levels L_{night}, annual average 24-hour equivalent sound level $L_{\text{Aeq,24h}}$, and annual day-night average sound level L_{dn} for the 70 sites for the past four years (2017–2020). It was observed that the majority of the sites (75.7%) registered day equivalent noise levels between 60 to 75 dB(A), while

64.3% of the sites registered night equivalent noise levels between 60 to 75 dB(A). Also, the majority of sites (69.5%) showed L_{dn} values between 65 to 80 dB(A), and 77.6% of the sites showed $L_{\mathrm{Aeq,24h}}$ values between 60 to 75 dB(A). Figures 4 and 5 show the 24-hour equivalent average annual sound levels $L_{\mathrm{Aeq,24h}}$ and annual day-night average sound levels L_{dn} observed for the past three years (2018–2020) in the 7 major cities of India (Garg, 2022). It can be observed that the majority of $L_{\mathrm{Aeq,24h}}$ values range from 60 to 75 dB(A) for commercial (83%) and industrial zone sites (65.9%), while for the residential (75%) and silence zone sites (76.9%), the majority of $L_{\mathrm{Aeq,24h}}$ values range from 55 to 70 dB(A). The majority of L_{dn} values range from 65 to 80 dB(A) for commercial (66%), residential (69%), and industrial zone sites (67%), while for the silence zone sites, the majority (80.5%) of L_{dn} values range from 60 to 75 dB(A).

The zone-wise analysis was also conducted as shown in Table 6 (Appendix) in order to ascertain the range of day and night equivalent noise levels in all 4 zones (Garg, 2022). The analysis of the 2020 noise monitoring data shows that 88% of the commercial sites, 62.6% of the residential sites, 70.5% of the silence zone sites, and 74.9% of the industrial zone sites registered day equivalent sound levels between 60 to 75 dB(A). Furthermore, 76% of the sites registered night equivalent sound levels between 60 to 75 dB(A), while 56.4% of the residential sites, 70.6% of the silence zone sites, and 66.7% of the industrial sites registered

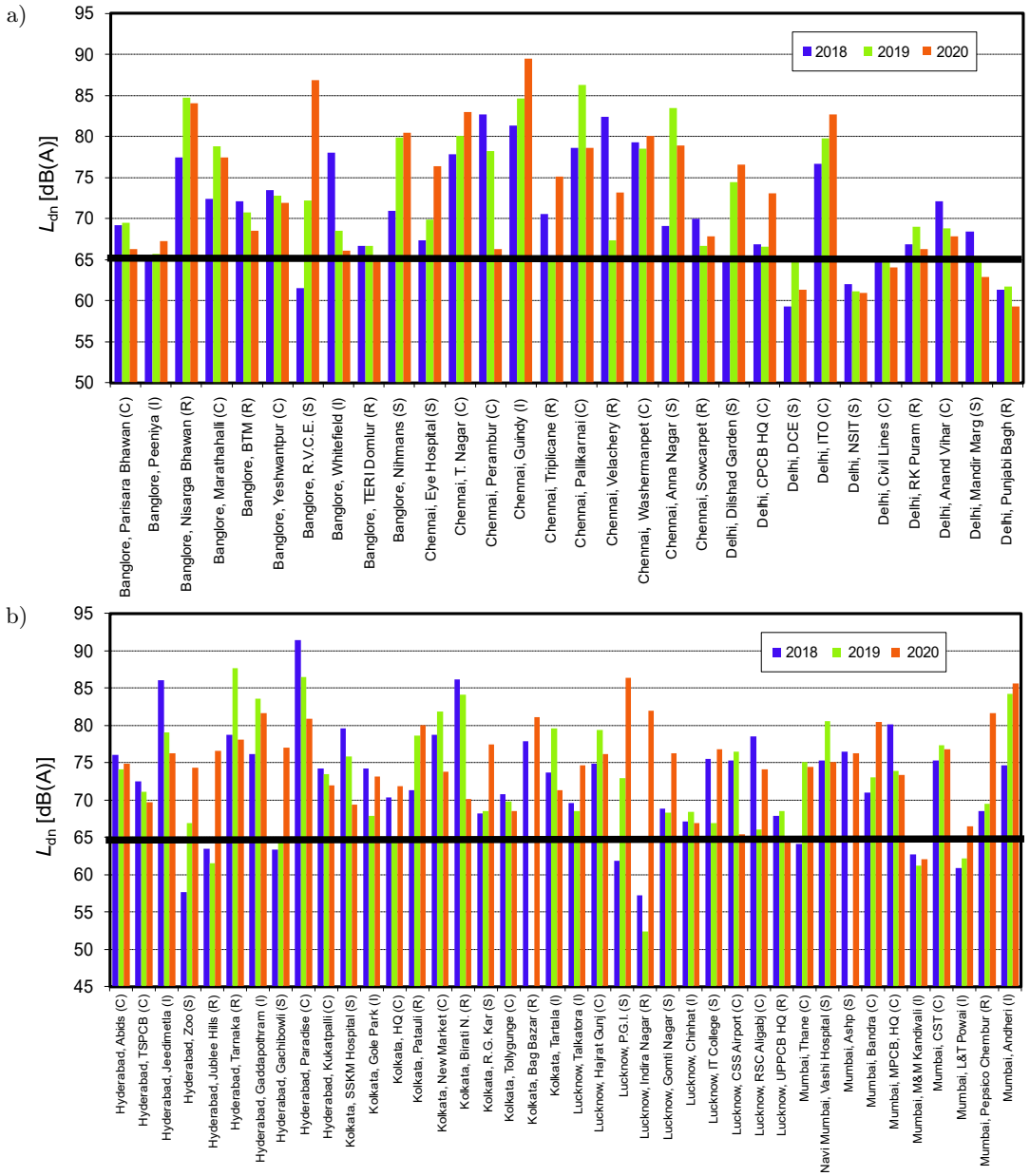


Fig. 4. Annual day-night average sound levels L_{dn} in dB(A) observed for the past three years (2018–2020) for 7 major cities (Garg, 2022).

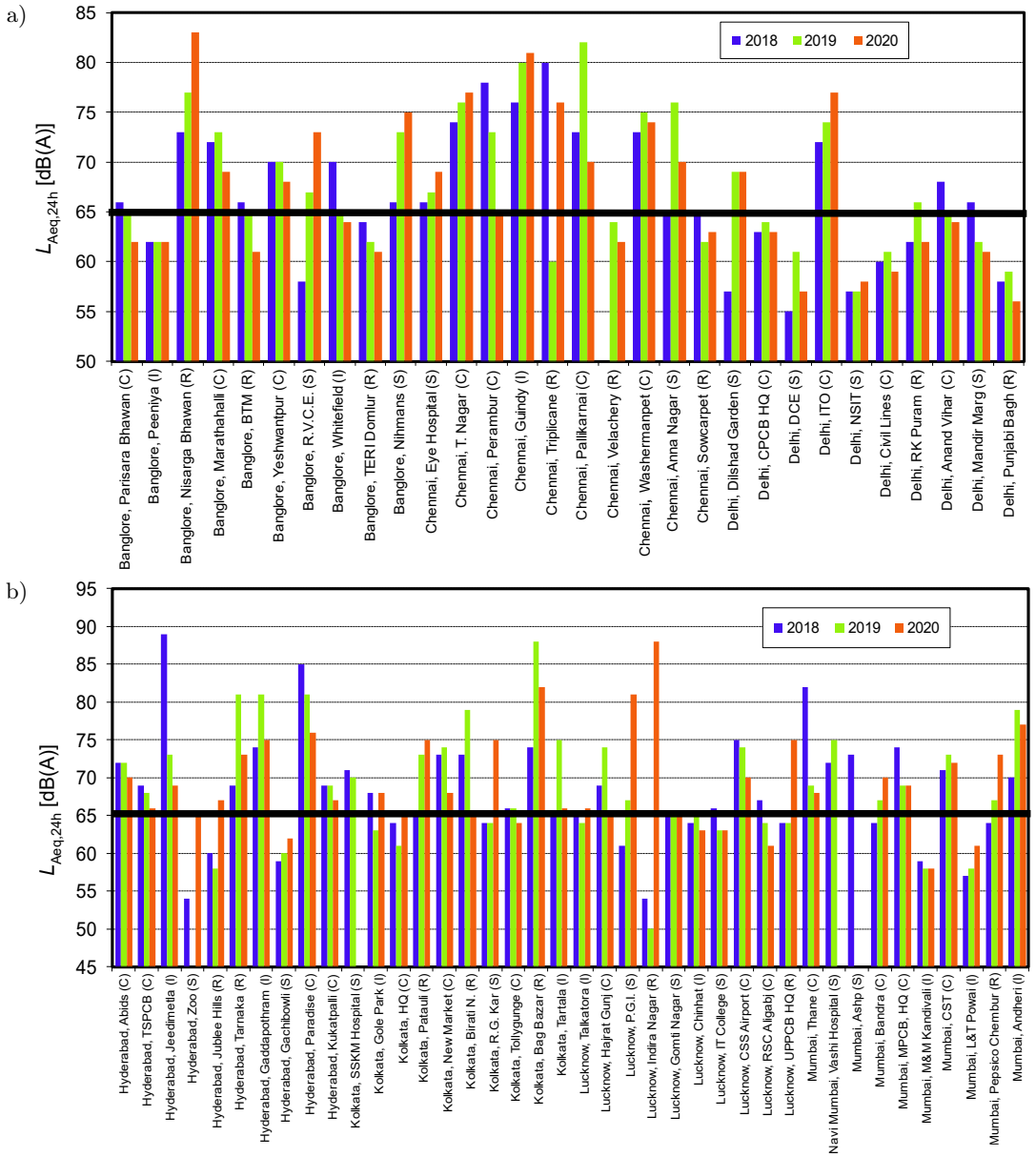


Fig. 5. Annual 24-hour equivalent average sound levels $L_{\mathrm{Aeq,24h}}$ in dB(A) observed for the past three years (2018–2020) for 7 major cities (Garg, 2022).

night equivalent noise levels between 55 to 70 dB(A). Thus, it is evident that the majority of the sites in all 4 zones registered day equivalent sound levels in the range of 60 to 75 dB(A), while the majority of sites in all zones except the commercial zone showed night equivalent sound levels between 55 to 70 dB(A). The analysis of (L_{day}–L_{night}) for the 70 sites was conducted to analyze the severity of day equivalent noise levels compared to night equivalent noise levels. Table 7 (Appendix) shows the frequency distribution (in %) of the difference of annual average (L_{day}–L_{night}) values observed in dB for the 70 sites spread across the 7 major cities of India (Garg, 2022). It is revealed that the majority of observations (75.7% for 2020 to 88.6% for 2014) showed a difference between 0 to 10 dB(A) and less than or equal to 5 dB(A) in the past ten years from 2011 to 2020 (31.4% in 2012 to 58.6% in 2020). These observations thus suggest that the night equivalent noise levels are comparable to the day equivalent levels for the majority of sites. The 10 dB night-time adjustment in day-night average sound level is not appropriate in such a scenario. On a similar analogy, these observations also suggest that the 5 dB evening time correction in the day-evening-night average sound level descriptor is not justified as the evening noise levels are similar to the day equivalent noise levels. Thus, the 24-hour equivalent continuous sound level $L_{\mathrm{Aeq,24h}}$ would be more suitable as it is a common way of expressing day-night average sound level without the 10 dB night-time adjustment (Garg, 2019). It may be noted that

these results are observed for the case of metropolitan cities only, and the environmental noise scenario for other cities, especially tier 2 cities, shall be helpful to conclude a generic trend about the suitability of day-night average sound levels with the 10 dB night-time adjustment in India.

The comparison of the environmental noise levels in the 70 sites in comparison to the previous years was ascertained to understand the change in noise scenario in the past decade. Figure 6a shows the difference in day and night equivalent sound levels in 2020 sound levels for the 35 sites in which the NMTs were installed in 2011. It can be observed that 19 sites (54.3%) showed an increment in day and night equivalent sound levels of more than 5 dB(A) in these ten years, 9 sites (25.7%) showed an increment in day equivalent sound levels up to 10 dB(A), and 19 sites (54.3%) showed an increment in night equivalent sound levels up to 10 dB(A). Only 3 sites showed a decrement in day equivalent sound levels up to 5 dB(A) and 1 site showed a decrement in night equivalent sound levels by more than 5 dB(A). Figure 6b shows the difference in day and night equivalent sound levels in 2020 sound levels for the 35 sites in which the NMTs were

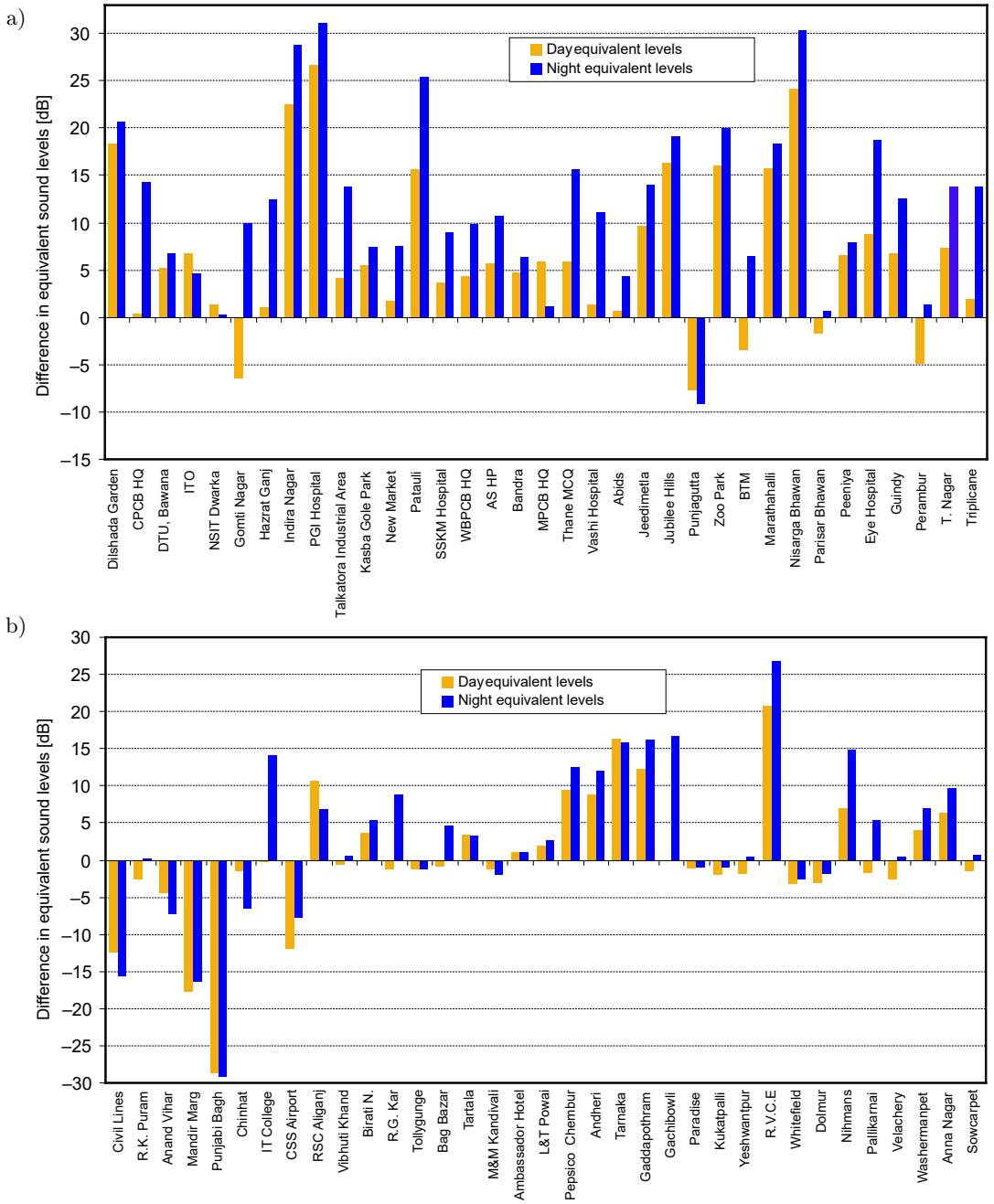


Fig. 6. Difference in equivalent sound levels: a) difference in day and night equivalent noise levels w.r.t. 2020 sound levels for 35 sites in which the NMTs were installed in 2011; b) difference in day and night equivalent sound levels w.r.t. 2020 sound levels for 35 sites in which the NMTs were installed in 2015.

installed in 2015. It can be observed that 8 sites (22.9%) showed an increment in day and night equivalent sound levels of more than 5 dB(A) in these five years, 5 sites (14.3%) showed an increment in day equivalent sound levels ≥10 dB(A) and 9 sites (25.7%) showed an increment in night equivalent sound levels ≥10 dB(A). Also, 4 sites (11.4%) showed a decrement in day equivalent sound levels ≥5 dB(A), while 6 sites (17.1%) showed a decrement in night equivalent sound levels ≥5 dB(A).

The comparison of $L_{\mathrm{Aeq,24h}}$ and L_{dn} levels for the 35 sites for the ten years under study, as shown in Fig. 7a, also revealed that 8 sites (22.9%) showed an increment in $L_{\mathrm{Aeq,24h}}$ levels ≥ 10 dB(A) and 14 sites (40%) showed an increment in L_{dn} levels ≥ 10 dB(A). Only 1 site showed a decrement in $L_{\mathrm{Aeq,24h}}$ and L_{dn} levels of more than 5 dB(A). Overall, it was observed that the commercial and silence zone sites exhibited higher increment in the day and night equivalent sound levels in the considered years.

The comparison of $L_{\mathrm{Aeq,24h}}$ and L_{dn} levels for the 35 sites for the past five years, as shown in Fig. 7b, also revealed that 5 sites (14.3%) showed an increment in $L_{\mathrm{Aeq,24h}}$ levels ≥10 dB(A) and 8 sites (22.9%) showed an increment in L_{dn} levels ≥10 dB(A). Additionally, 5 sites (14.3%) showed a decrement in $L_{\mathrm{Aeq,24h}}$ ≥5 dB(A), while 5 sites (14.3%) showed a decrement in L_{dn} levels ≥5 dB(A). These observations also revealed that commercial and silence zone sites exhibited higher increment in the 24-hour equivalent sound levels and day-night average sound levels in the considered years.

A piecewise linear regression analysis was also performed to assess the approximate rate of variation of day and night equivalent levels in the past ten and six years, for the new addition in 2014, on an annual basis. The slope of variation (in dB(A)/year) of day and night equivalent levels and the corresponding Pearson's correlation coefficients for all the 70 sites in the 7 cities of India were evaluated, as shown in Tables 8 and 9

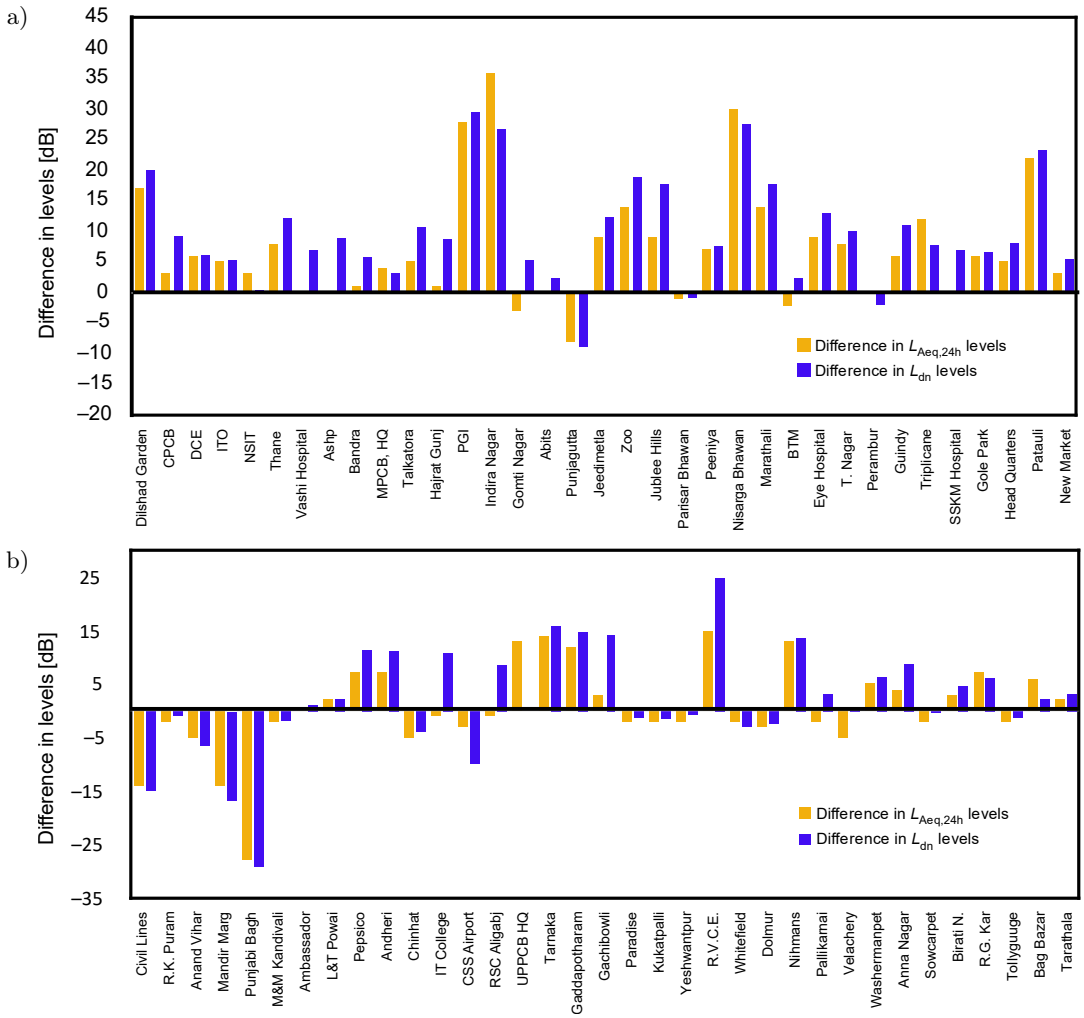


Fig. 7. Difference in equivalent sound levels: a) difference in 24-hour equivalent sound levels and day-night average sound levels w.r.t. 2020 sound levels for 35 sites in which the NMTs were installed in 2011; b) difference in 24-hour equivalent sound levels and day-night average sound levels w.r.t. 2020 sound levels for 35 sites in which the NMTs were installed in 2015.

(Appendix). Table 8 shows the slope of variation of ambient noise levels in the past ten years, 2011–2020, and the corresponding correlation coefficient for the 35 sites where the noise monitoring stations were established in 2011.

Table 9 presents the slope of variation of ambient noise levels in the past six years, 2015–2020, and the corresponding correlation coefficient for the 35 sites where the noise monitoring stations were established in 2014. It is revealed in Table 8 that except for 1 commercial, 2 residential, and 1 silence zone site, all other sites showed positive slopes for day equivalent levels and except for 1 commercial zone site, all other sites reported positive slopes for night equivalent levels for the past ten years (2011–2020), which indicates the increment in the ambient noise levels in these ten years. Also, for the variation of day equivalent levels, more than 70% of the 35 sites exhibited a slope within the range of 0 to 1.5 dB(A)/year, and almost 60% of the sites exhibited a slope within 0 to 1 dB(A)/year, while for the variation of night equivalent levels, nearly 77% of the 35 sites reported the slope of variation within 0 to 2 dB(A)/year. It is observed in Table 9 that for the past six years (2015–2020), 5 commercial, 3 industrial, 6 residential, and 4 silence zone sites exhibited a negative slope of variation of day equivalent levels and 4 commercial, 3 industrial, 4 residential, and 1 silence zone site reported negative slope of variation of night equivalent levels in the past six years (2015–2020), indicating a decrement in the day and night equivalent levels in these six years. Overall, 18 and 12 sites out of 35 sites reported a negative slope of variation for the day and night equivalent levels, respectively. Thus, it can be deduced that more commercial, industrial, and residential zone sites exhibited decrement in the day and night equivalent sound levels than the silence zone sites in the six years under study. Also, 21 out of 35 sites reported the slope of variation of day equivalent levels within the range −1 to 1 dB(A)/year, and 17 out of 35 sites exhibited a slope of variation of night equivalent levels within the range 0 to 2.5 dB(A)/year for the past six years from 2015 to 2020.

Population growth, rising transportation needs, an increase in vehicular density particularly heavy vehicles and cars, etc., and road congestion are all factors that are primarily attributed to the increased noise levels in the ten years under study (Jamir *et al.*, 2014). Also, the increased industrial activities account for the elevated ambient noise levels of the noise monitoring sites in ten years span of time. The increased encroachment of vehicles alongside the main roads in silence, residential, and commercial zones, congestion due to heavy vehicles, and unnecessary honking events are also associated with the elevated noise levels in these zones. Furthermore, loud music from the various night events (including marriage ceremonies and other concerts), noise from dog-barking at night at some sites, and movement of heavy trucks at night-time are primarily attributed to the higher noise levels at some sites.

The present study is the first study to comprehensively analyze the extensive noise monitoring data for 7 major cities of India for the past ten years from 2011 to 2020. Although the non-compliance of the silence and residential zone sites with the ambient noise standards is supported by several studies for other cities in India as well (Datta *et al.*, 2006; Aggarwal, Swami, 2011; Bhosale *et al.*, 2010; Garg *et al.*, 2016; 2017b; 2017c), a comprehensive study focussed in tier 2 and tier 3 cities shall give a broader picture of noise scenario in the country for the various zones.

4. Overall noise scenario and compliance with ambient standards

Long-term noise monitoring for the past decade revealed that the environmental noise levels were high compared to the recommended limits for some of the sites and thus noise control measures (or noise action plans) are essentially required for controlling the noise levels. It was observed that only 4 sites (5.7%) met the target of 55 dB L_{night}. Table 10 (Appendix) shows the status of compliance of the day and night equivalent levels explicitly for the various sites with the ambient noise limits. It can be observed that day-time compliance is observed in more sites than night-time compliance. A minimum of 9 to a maximum of 14 sites out of 70 sites have shown day-time compliance in the studied ten years, while a minimum of 5 to a maximum of 12 sites showed night-time compliance. No silence zone site ever met the ambient noise limits in the span of ten years under study, while only 2 residential sites showed compliance with the day ambient noise limits for 2011–2013, and 1 residential zone site showed day-time compliance in 2014, 2018, and 2019. Table 11 (Appendix) enlists the status of overall compliance of all the sites in 7 cities in the considered decade (Garg, 2022). It is evident that primarily the industrial sites met the ambient noise limits in these ten years. Also, in accordance with the U.S Department of Housing and Urban Development criteria (1984) that recommends the $L_{\text{Aeq}} \leq 49$ dB(A) as clearly acceptable and $49 < L_{\text{Aeq}} \leq 62$ dB(A) as normally acceptable. In 2020, 15 sites (21.4%), including 3 industrial, 5 commercial, 4 residential, and 3 silence zones, met these criteria. The level of significance for assessing noise impacts has been identified as an L_{dn} of 65 dB(A), whereby an L_{dn} value of 65 dB(A) is described as the onset of a normally unacceptable zone. In accordance with these criteria, 9 sites (12.9%) in 2020, including 1 industrial, 3 commercial, 3 residential, and 2 silence zones, met these criteria. These observations thus suggest a retrospective and prospective

view of ambient noise standards, particularly for the residential areas and sites lying in the silence zones and mixed zone prevalent at some sites. It is rightly pointed out in the European Night Noise Guidelines (WHO, 2009) report that limits could be reasonably high but firmly imposed or very minimal with no legal obligation whatsoever.

Also, in order to ascertain the most severely affected zone, the NEF for each zone in the past four years was analyzed, as shown in Table 12 (Appendix). The NEF was calculated as the ratio of the ambient noise level (day/night) observed at the site to the noise limit recommended by the ambient noise standards (Chowdhury *et al.*, 2016). The analysis of NEF and AEF revealed that the silence zone has the maximum AEF of 1.4 for day equivalent sound levels and 1.7 for night equivalent sound levels, followed by residential zones (AEF 1.3–1.5) and commercial zone (AEF 1.1–1.2) in the 2020 noise monitoring data. The analysis of the four years AEF values suggested that the silence zone sites were the most affected sites, followed by residential zone sites. A planned land-use pattern of commercial, residential, silence, and industrial zones in a city shall be thus pivotal in reducing environmental noise levels. The numerical meta-analyses ascertaining the exposure-response relationship between community noise and cardiovascular risk recommends an empirical formulation as (WHO, 2011):

$$\mathrm{OR} = 1.63 - 6.13 \cdot 10^{-4} \cdot L^2_{\mathrm{day,16h}} + 7.36 \cdot 10^{-6} \cdot L^3_{\mathrm{day,16h}}, \quad (6)$$

where $L_{\mathrm{day,16h}}$ is 16 hours ambient day level and OR is the odds ratio that is used to compare the relative odds of the occurrence of the outcome of disease, given exposure to the variable of interest (noise exposure level). Thus, for the Gunidy site in Chennai that experienced the highest day equivalent sound levels of 82.8 dB(A) in 2020 and the Bag Bazar site in Kolata that experienced the highest day equivalent sound levels of 87.2 dB(A) in 2019, the ORs of 1.6 and 1.8 were evaluated, respectively. The variation of OR was evaluated to be in the range of 1 to 1.6 for 2019 and 1 to 1.8 in 2020. Thus, epidemiological meta-analysis and noise annoyance studies are essentially required from an Indian perspectives to correlate noise exposure with the health effects. The empirically described conversion rules between the environmental noise exposure metrics as that presented by Brink *et al.* (2018) in Switzerland shall facilitate the estimation of the value of one (unknown) noise metric from the value of another (known) metric, e.g., in the scope of epidemiological meta-analyses or systematic reviews, when results from different studies are pooled and need to be related to one common exposure metric. Table 13 (Appendix) shows the inter-conversion of the various noise descriptors for the four zones based on the analysis of noise monitoring data of the past three years. The interconversion rule shall facilitate the prediction of noise descriptors eventually when short-term noise monitoring studies are carried out (Garg *et al.*, 2015a; Tiwari *et al.*, 2022).

5. Noise action plans and noise control policy

The undue violation of ambient noise limits for the residential and silence zone sites is primarily due to the mixed category zone prevalent for some of the locations. Thus, these observations suggest a need for the national policy framework for the management and control of noise pollution in India (Garg, Maji, 2016; Garg *et al.*, 2022). Figure 8 shows the proposed plan for management and control of noise pollution and the noise action plans for controlling noise pollution in the urban cities of India. The major elements of such a policy framework shall be: noise mapping, monitoring and certification, enforcement of ambient noise standards, selection and execution of appropriate noise action plans, traffic management policy, legal measures and noise screening policy, and noise awareness campaigns for inculcating awareness amongst the community towards reducing the noise pollution. The proposal in the master plan with major impetus on traffic decongestion includes the following: Unified Metro Transport Authority, synergy between land-use and transport integrated multi-modal public transport system for reducing the dependence on personalized vehicles, road and rail-based mass transport system recommended to be a major mode of public transport, and optimal use of existing road network. National Transport Oriented Development (TOD) policy can serve as guidelines and play a catalytic role in formulating state/city-level policies to promote transit-oriented development (CPCB, 2017). Decongestion plans for busy road junctions, special drives for "no honking", promotion of carpooling policy, and "work from home" culture for some of the offices (Kumar *et al.*, 2022) for reducing road traffic, launching of odd-even traffic measures for some time (Garg *et al.*, 2017a), launching car-free day initiatives, installing synchronized traffic signaling, constructing flyovers in metropolitan cities for deflecting the traffic, easing the traffic density, and restricting entry of heavy vehicles are some of the major administrative issues that should be implemented for reducing the environmental noise levels. Installation of noise barriers at hotspots in an adequate manner is also a feasible option for noise abatement in various zones of metropolitan cities of India (Lokhande *et al.*, 2021a). Demarcation of all the silence zone and residential zone sites in cities, demarcation of no-honking zones, proper land-use planning, and inculcating awareness amongst the community on associated health hazards due to noise exposure shall be very effective steps to control noise pollution. Periodic noise monitoring of sites in conjunction with noise mapping of cities and periodic review

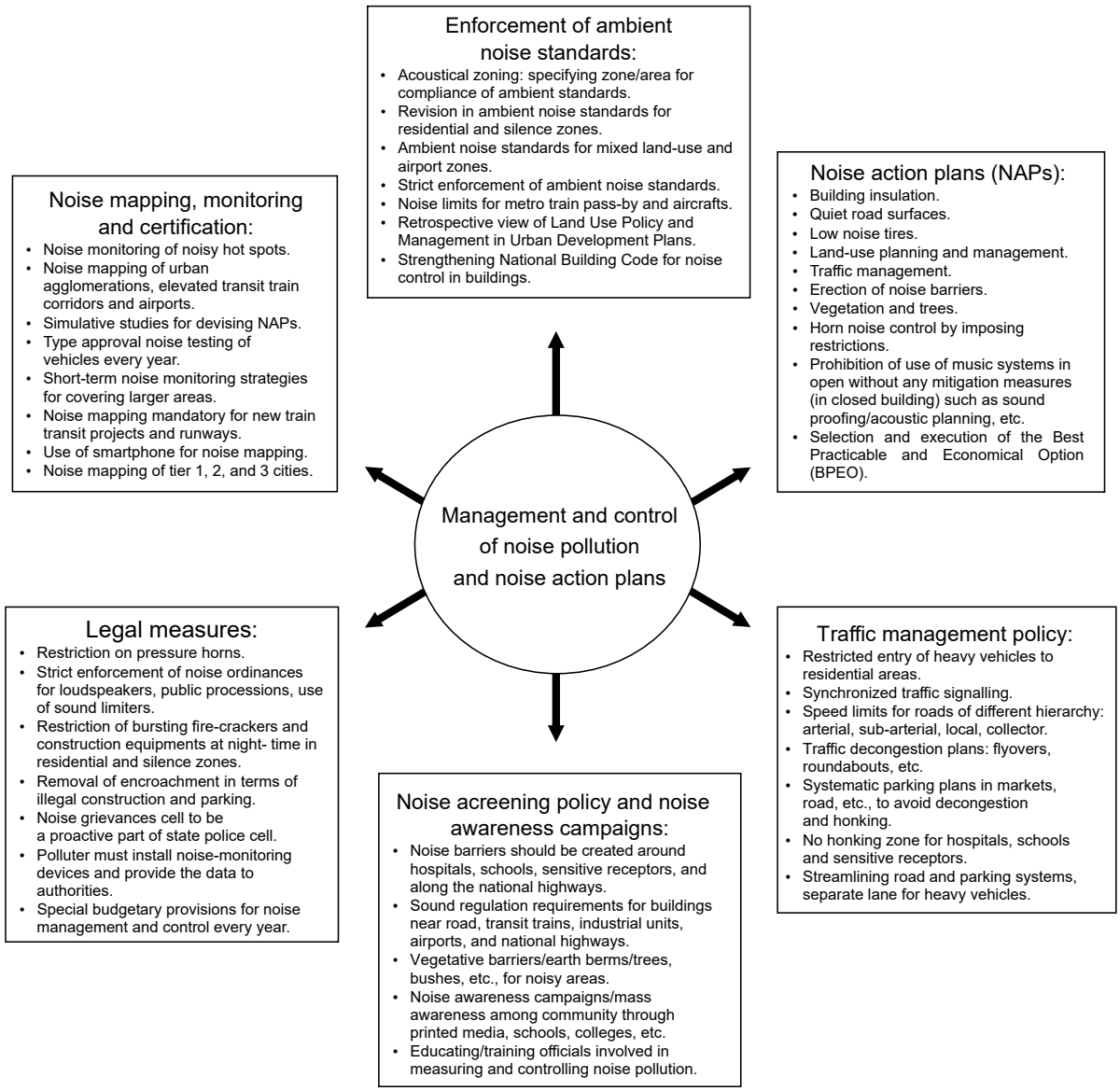


Fig. 8. Proposed plan for management and control of noise pollution and noise action plans in Indian scenario to effectively control noise pollution in urban cities.

of noise scenarios after fixed intervals shall be indispensable to analyze, understand, and devise suitable noise action plans (Garg *et al.*, 2015a; 2020; 2021; Kumar *et al.*, 2023; Lokhande *et al.*, 2019). In addition, noise measurements with the help of smartphone applications are also an effective, novel and economical way of disseminating awareness among the community, especially the young generations (Lokhande *et al.*, 2021b). It is recommended that a mixed zone category should be specifically considered for the noise abatement guidelines/ambient standards in the future from an Indian perspective. Reduction at the source (travel demand reduction, setting restrictive speed limits in residential areas and sensitive sites, prohibition of heavy vehicle traffic in residential areas and sensitive sites, minimization of slopes in urban roads, control of acoustic emissions generated by vehicles with sirens, strict enforcement of regulations governing the emission limits and conditions of use of vehicles, etc.) is the effective action plan for reducing the environmental noise levels (Torija *et al.*, 2021). The European Union has recently reported the noise action plans for major roads as the measures on the propagation path (40%) followed by the source-orientated measures (38%). Noise barriers and traffic management measures were the most commonly reported, followed by improving the road surface. However, the actions related to urban planning only account for a small percentage (13%) and the ones related to education and communication account for 6% (Blanes *et al.*, 2019; Conference of European Directors of Roads [CEDR], 2013; European Environment Agency, 2020). The cost-

effectiveness of the noise action plans is very crucial for selecting an optimal strategy (Münzel *et al.*, 2018; European Union, 2017; Lokhande *et al.*, 2022; Garg *et al.*, 2012; 2022; Tiwari *et al.*, 2021). Thus, an effective noise policy for a sustainable environment, including noise action plans in urban planning, is essential from an Indian perspective for accomplishing the desired goals towards noise pollution control in metropolitan cities. Thus, these measures have to be implemented for metropolitan cities of India while considering the recent WHO (2018) guidelines such as reducing noise exposure while conserving silence areas, promoting the interventions to reduce noise exposure and improve health, coordinating the various approaches to control noise source and other environmental health risks and information and involvement of the communities (WHO, 2018).

6. Conclusions and recommendations

This paper analyzed and reported the long-term noise-monitoring data for 2011–2020 obtained from the diversified NANMN set up across the 7 major cities of India and covering 70 stations for continuous noise monitoring throughout the year. The annual average ambient noise levels observed in these years for these 70 locations under study, in which 25 locations were in commercial zones, 12 in industrial, 16 in residential, and 17 in silence zones, were described. These observations were instrumental in ascertaining the noise scenario, the status of compliance with the ambient noise limits and planning a national policy framework for reducing noise pollution in metropolitan cities of India. Such a study shall be helpful for predicting and forecasting future noise scenarios in the decade. The following conclusions can be drawn from this study:

- the comparison of the ambient noise levels in the studied decade revealed that the majority of sites (19 out of 35) registered an increment in day and night equivalent sound levels of more than 5 dB(A) in this decade. Only 3 sites showed a decrement in day equivalent sound levels by 5 dB(A) and 1 site showed a decrement in night equivalent sound levels by 5 dB(A). Overall, it was observed that the commercial and silence zone sites exhibited a higher increment in the day and night equivalent sound levels in the past few years;
- in the past three years, from 2018 to 2020, 7 out of 10 sites in Chennai and Mumbai exhibited day and night equivalent levels greater than 70 dB(A). For Kolkata, 6 sites reported day and night equivalent sound levels greater than 70 dB(A). Also, 4 out of 10 sites in Lucknow, Hyderabad, and Bengaluru exhibited day and night equivalent levels greater than 70 dB(A). For Delhi, only 2 out of 10 sites showed day and night equivalent sound levels greater than 70 dB(A). Forty-two sites (60%) comprising of 18 commercial, 5 industrial, 9 in silence zone, and 10 residential were observed to be the most exposed urban sites with day equivalent sound levels ≥70 dB(A);
- the NOAEL limit of 60 dB(A) L_{day} was met by 4 sites only for 2020, while 7 sites met this limit in 2019 and 9 sites in 2018. The interim target (IT) of 55 dB L_{night}, as recommended in the EU Night Noise Guidelines report, was met by 4 sites in 2020, 7 sites in 2019, and 8 sites in 2018. Overall, 94.3% of the observations exceeded the interim target recommended by the Night Noise Guidelines (NNG) report, which indicates that 55 dB L_{night} is an ambitious target for all these sites under consideration;
- the zone-wise analysis of ambient noise levels showed that the majority of sites registered day equivalent sound levels in the range of 60 to 75 dB(A) in all 4 zones. Also, the majority of sites registered night equivalent sound levels in the range of 55 to 70 dB(A) for residential, silence, and industrial zones, while the majority of the sites (76%) in the commercial zones registered night equivalent sound levels in the range of 60 to 75 dB(A);
- the analysis of the noise monitoring data for the 70 sites for the year 2020 showed that the majority of $L_{\mathrm{Aeq,24h}}$ values range from 60 to 75 dB(A) for commercial (83%) and industrial zone sites (65.9%), while for the residential (75%) and silence zone sites (76.9%), the $L_{\mathrm{Aeq,24h}}$ values range from 55 to 70 dB(A). Also, the analysis of L_{dn} values revealed that the majority of these values range from 65 to 80 dB(A) for commercial (66%), residential (69%), and industrial zone sites (67%), while for the silence zone sites, the majority (80.5%) of L_{dn} values range from 60 to 75 dB(A);
- the analysis of compliance with the ambient standards showed that day-time compliance is shown by more sites than the night-time. A minimum of 9 sites and a maximum of 14 sites out of 70 sites showed day-time compliance from 2011 to 2020, while only a minimum of 5 sites and a maximum of 12 sites showed night-time compliance during the past ten years. Primarily, only the commercial and industrial zone sites complied with the ambient noise standards from 2011 to 2020. No silence zone site ever met the ambient noise limits, while only 2 residential sites showed compliance with the day ambient noise limits. No site met the limits of 53 dB L_{den} and 45 dB L_{night} recommended by the Guidelines Development Group of the WHO for road traffic noise. These observations suggest the need to reconsider ambient noise standards, especially for residential and silence zones. Also, the mixed category zone was available for some of these sites, however it was very

difficult to exclusively classify them as either residential or silence zones or commercial or industrial zones. It is recommended that a mixed zone category should be specifically considered for the noise abatement guidelines/ambient standards in the future from an Indian perspective;

– the analysis of (L_{day}–L_{night}) for 70 sites revealed that the majority of observations show a difference between 0 to 10 dB(A) and less than or equal to 5 dB(A) in the decade under study. Therefore, these observations suggest that the night equivalent noise levels are comparable to the day equivalent levels for the majority of sites. Thus, the 10 dB night time adjustment in the day-night average sound level is not appropriate in such a scenario. These observations also suggest that a 5 dB evening time correction in the day-evening-night average sound level descriptor is not justified in the Indian context as the evening noise levels are the same as the day equivalent noise levels. Thus, the 24-hour equivalent continuous sound level $L_{\mathrm{Aeq,24h}}$ would be a more suitable descriptor as it is a common way of expressing the day-night average sound level without a 10 dB night-time adjustment. It is thus recommended that for developing exposure-effect relationships and correlating noise annoyance, and correlating noise exposure with the health aspects, the single noise descriptor, 24-hour equivalent continuous sound level $L_{\mathrm{Aeq,24h}}$, would be more suitable compared to L_{dn} or L_{den} descriptors;

– it was observed that the approximate annual rate of variation of ambient noise levels was positive for 31 out of 35 sites for the ten years, indicating an increment in the ambient noise levels in the these ten years. For the rest of the 35 sites where the noise monitoring stations were installed in 2014, the approximate annual rate of variation of day equivalent levels of 18 out of 35 sites was negative and for the night equivalent levels, the approximate annual rate of variation of 12 out of 35 sites was negative, which indicates the decrement in day and night equivalent levels of 18 and 12 sites, respectively, in the past six years. Also, it was revealed that relatively more commercial, industrial, and residential zone sites exhibited a decrement in the day and night equivalent sound levels in comparison to silence zones in the same six years;

– the analysis of the NEF and AEF showed that the silence zone has the maximum AEF of 1.4 for day equivalent noise levels and 1.7 for night equivalent noise levels, followed by residential zones (AEF 1.3–1.5) and commercial zones (AEF 1.1–1.2) in 2020 noise monitoring data. Thus, it suggests that silence zone sites are the most affected sites, followed by residential zone sites.

The present study considered the analysis of noise monitoring data from 70 sites in 7 metropolitan cities. Future studies focusing on long-term evaluation and analysis of noise monitoring data of tier 2 and tier 3 cities shall be helpful in understanding the generic noise scenario of the country and developing a national, coherent noise policy from an Indian perspective. The undue violation of ambient noise limits for the residential and silence zone sites is primarily due to the mixed category zone prevalent for some of the locations. Figure 8 recommends the road map of national policy on noise management and control for controlling the noise pollution in metropolitan cities of India. Thus, the execution and implementation of the NAPs and the administrative measures such as traffic management policy, noise screening policy, legal measures, and strict enforcement of ambient noise standards can be very instrumental in reducing ambient noise levels. Thus, the future developments and establishment of "Smart Cities" should consider these aspects at the designing stages for controlling noise pollution and developing sustainable cities promoting good health and quality of life.

Acknowledgments

The authors are very thankful to the CPCB, India and web portal CPCB (n.d.) for sharing and disseminating the noise monitoring data of 70 sites. The present work is an extension of the previous work reported on the pilot project of establishment of the NANMN project by CPCB, India. The paper analyzes and reports the noise monitoring data and the views and opinions expressed in this article are those of the authors' own and do not necessarily reflect the official policy or position of any agency of the Government of India. The content of the papers is solely to present a retrospective and prospective view of noise levels in the studied ten years and may not be used or considered for dispute redressal in the legal framework.

Appendix

Table 1. Ambient noise standards of India (Ministry of Environment & Forests, 2000).

Area code	Category of area/zone	Limits in dB(A) L_{eq}*	
		Day-time	Night-time
A	Industrial area	75	70
B	Commercial area	65	55
C	Residential area	55	45
D	Silence zone	50	40

* L_{eq} denotes the time-weighted average of the sound level in decibels in A-weighting.

Table 2. Annual average ambient levels L_{day} and L_{night} for 35 noise monitoring stations installed across 7 major cities in India from 2011 to 2014 (CPCB, 2015).

Name of location	City	Area characteristics	2011		2012		2013		2014	
			L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}
Dilshad Garden	Delhi	Silence	52.0 ± 0.9	50.0 ± 1.4	52.0 ± 1.1	49.0 ± 2.1	51.0 ± 1.1	49.0 ± 2.4	52.0 ± 0.9	48.0 ± 1.6
CPCB HQ		Commercial	64.0 ± 2.0	53.0 ± 1.4	62.0 ± 1.0	53.0 ± 1.3	63.0 ± 0.8	53.0 ± 1.0	65.0 ± 1.7	54.0 ± 1.6
DTU, Bawana		Silence	52.0 ± 1.3	48.0 ± 2.1	51.0 ± 0.9	49.0 ± 3.2	52.0 ± 1.7	49.0 ± 3.0	52.0 ± 1.1	48.0 ± 2.5
ITO		Commercial	73.0 ± 0.6	71.0 ± 1.0	72.0 ± 4.0	69.0 ± 5.3	74.0 ± 0.7	73.0 ± 0.4	74.0 ± 1.0	73.0 ± 1.4
NSIT Dwarka		Silence	56.0 ± 1.3	54.0 ± 0.8	56.0 ± 0.7	54.0 ± 1.1	56.0 ± 0.5	53.0 ± 0.9	56.0 ± 1.4	53.0 ± 1.8
Gomti Nagar	Lucknow	Silence	71.0 ± 0.8	61.0 ± 1.5	63.0 ± 0.9	55.0 ± 1.1	66.0 ± 2.2	57.0 ± 1.6	69.0 ± 1.4	61.0 ± 2.0
Hazrat Ganj		Commercial	67.0 ± 0.9	58.0 ± 1.0	72.0 ± 0.5	61.0 ± 1.0	72.0 ± 0.5	62.0 ± 1.3	72.0 ± 0.5	61.0 ± 1.5
Indira Nagar		Residential	54.0 ± 1.2	47.0 ± 2.9	53.0 ± 1.1	47.0 ± 3.0	54.0 ± 1.4	48.0 ± 3.6	57.0 ± 0.9	49.0 ± 4.7
PGI Hospital		Silence	55.0 ± 2.5	49.0 ± 2.8	58.0 ± 1.2	52.0 ± 3.6	60.0 ± 1.4	53.0 ± 3.0	62.0 ± 1.3	55.0 ± 3.5
Talkatora Industrial Area		Industrial	63.0 ± 0.4	55.0 ± 1.6	64.0 ± 0.7	56.0 ± 1.6	63.0 ± 0.5	56.0 ± 1.9	64.0 ± 1.2	57.0 ± 2.0
Kasba Gole Park	Kolkata	Industrial	64.0 ± 1.2	59.0 ± 1.3	65.0 ± 1.6	61.0 ± 2.6	68.0 ± 3.5	64.0 ± 4.7	70.0 ± 2.5	67.0 ± 2.9
New Market		Commercial	67.0 ± 0.5	60.0 ± 1.4	67.0 ± 0.7	59.0 ± 1.4	68.0 ± 0.5	60.0 ± 1.6	70.0 ± 2.3	65.0 ± 5.2
Patauli		Residential	55.0 ± 1.0	49.0 ± 2.0	55.0 ± 1.0	49.0 ± 3.2	55.0 ± 1.6	48.0 ± 6.2	55.0 ± 1.3	49.0 ± 4.5
SSKM Hospital		Silence	61.0 ± 0.4	54.0 ± 0.9	62.0 ± 0.8	56.0 ± 1.8	62.0 ± 1.2	57.0 ± 1.9	62.0 ± 1.1	56.0 ± 1.7
WBPCB HQ		Commercial	62.0 ± 0.6	56.0 ± 1.3	61.0 ± 0.7	54.0 ± 1.1	62.0 ± 1.4	55.0 ± 1.4	64.0 ± 0.6	58.0 ± 0.9
AS HP	Mumbai	Silence	66.0 ± 1.2	59.0 ± 1.5	65.0 ± 1.0	59.0 ± 0.3	65.0 ± 0.8	60.0 ± 1.5	66.0 ± 2.0	60.0 ± 0.9
Bandra		Commercial	70.0 ± 0.5	68.0 ± 0.8	69.0 ± 0.7	67.0 ± 1.9	69.0 ± 0.4	67.0 ± 0.5	70.0 ± 0.5	67.0 ± 0.7
MPCB HQ		Commercial	67.0 ± 0.6	63.0 ± 0.5	66.0 ± 0.5	63.0 ± 0.7	68.0 ± 1.6	65.0 ± 2.0	71.0 ± 0.6	68.0 ± 1.0
Thane MCQ		Commercial	62.0 ± 1.8	53.0 ± 2.3	62.0 ± 0.7	55.0 ± 1.9	62.0 ± 1.2	55.0 ± 1.4	64.0 ± 1.0	56.0 ± 1.3
Vashi Hospital		Silence	68.0 ± 1.7	58.0 ± 1.4	69.0 ± 0.9	57.0 ± 2.7	69.0 ± 0.8	57.0 ± 0.8	69.0 ± 1.5	59.0 ± 3.6
Abids	Hyderabad	Commercial	72.0 ± 0.5	63.0 ± 0.9	72.0 ± 0.9	63.0 ± 1.9	72.0 ± 0.8	64.0 ± 2.1	74.0 ± 1.9	65.0 ± 2.5
Jeedimetla		Industrial	62.0 ± 0.5	56.0 ± 1.4	63.0 ± 1.2	56.0 ± 2.1	63.0 ± 1.3	56.0 ± 1.6	65.0 ± 0.6	58.0 ± 0.8
Jubilee Hills		Residential	58.0 ± 1.0	50.0 ± 1.7	56.0 ± 0.7	49.0 ± 0.5	56.0 ± 0.6	49.0 ± 1.2	57.0 ± 1.6	49.0 ± 1.2
Punjagutta		Commercial	76.0 ± 0.6	71.0 ± 1.0	75.0 ± 0.5	70.0 ± 0.5	76.0 ± 1.7	71.0 ± 1.3	79.0 ± 0.7	73.0 ± 0.5
Zoo Park		Silence	53.0 ± 1.5	48.0 ± 2.8	54.0 ± 1.8	48.0 ± 2.0	54.0 ± 1.4	49.0 ± 1.1	56.0 ± 1.2	50.0 ± 2.2
BTM	Bengaluru	Residential	66.0 ± 0.4	56.0 ± 0.4	66.0 ± 0.5	56.0 ± 1.0	66.0 ± 0.8	56.0 ± 0.8	66.0 ± 0.7	57.0 ± 1.1
Marathahalli		Commercial	56.0 ± 1.9	53.0 ± 1.8	54.0 ± 0.7	52.0 ± 0.6	57.0 ± 2.1	54.0 ± 2.8	59.0 ± 0.7	56.0 ± 0.8
Nisarga Bhawan		Residential	56.0 ± 3.0	47.0 ± 1.8	56.0 ± 2.0	47.0 ± 1.9	56.0 ± 1.9	48.0 ± 1.6	56.0 ± 1.5	49.0 ± 1.4
Parisar Bhawan		Commercial	66.0 ± 1.1	58.0 ± 0.7	65.0 ± 0.3	57.0 ± 1.6	65.0 ± 0.7	57.0 ± 0.8	65.0 ± 0.8	57.0 ± 0.5
Peeniya		Industrial	56.0 ± 1.6	53.0 ± 2.6	56.0 ± 1.2	49.0 ± 1.2	58.0 ± 1.1	53.0 ± 2.3	58.0 ± 0.8	55.0 ± 2.0
Eye Hospital	Chennai	Silence	64.0 ± 0.6	51.0 ± 1.2	62.0 ± 1.5	52.0 ± 3.1	64.0 ± 1.5	53.0 ± 2.2	61.0 ± 3.9	53.0 ± 1.2
Guindy		Industrial	76.0 ± 0.6	71.0 ± 1.1	75.0 ± 1.1	71.0 ± 1.3	75.0 ± 1.0	71.0 ± 1.5	76.0 ± 1.9	72.0 ± 1.2
Perambur		Commercial	68.0 ± 0.9	58.0 ± 0.8	69.0 ± 1.2	58.0 ± 1.2	68.0 ± 0.5	57.0 ± 0.7	69.0 ± 1.3	58.0 ± 0.8
T. Nagar		Commercial	73.0 ± 0.5	62.0 ± 1.1	73.0 ± 0.3	62.0 ± 1.0	74.0 ± 1.0	64.0 ± 2.0	75.0 ± 0.9	67.0 ± 1.6
Triplicane		Residential	68.0 ± 0.4	55.0 ± 1.0	68.0 ± 0.5	56.0 ± 0.8	68.0 ± 0.5	56.0 ± 0.7	68.0 ± 1.4	57.0 ± 2.0

Table 3. Annual average ambient levels L_{day} and L_{night} for 35 noise monitoring stations installed across 7 major cities in India from 2015 to 2020 (CPCB, 2015a; 2015b; 2016; 2018; Garg, 2022).

Name of location	City	Area characteristics	2015		2016		2017		2018		2019		2020	
			L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}
Dilshad Garden	Delhi	Silence	57.3 ± 3.3	54.0 ± 5.6	53.0 ± 1.4	51.0 ± 2.7	54.0 ± 1.5	53.0 ± 0.4	64.1 ± 1.1	57.0 ± 0.6	68.7 ± 2.6	68.4 ± 2.8	70.4 ± 3.5	70.6 ± 3.5
CPCB HQ		Commercial	69.2 ± 1.8	59.1 ± 1.5	66.0 ± 0.9	57.0 ± 0.8	66.0 ± 0.5	57.0 ± 1.5	66.1 ± 0.4	58.0 ± 0.7	66.5 ± 0.5	56.7 ± 0.7	64.5 ± 2.3	67.4 ± 2.3
DTU, Bawana		Silence	63.0 ± 6.9	62.0 ± 11	57.0 ± 3.5	57.0 ± 7.6	56.0 ± 2.0	52.0 ± 3.1	56.2 ± 1.0	52.3 ± 2.0	62.2 ± 2.9	57.9 ± 2.3	57.2 ± 3.7	54.8 ± 3.3
ITO		Commercial	74.0 ± 0.9	70.0 ± 1.2	75.0 ± 4.2	71.0 ± 4.3	72.0 ± 0.9	67.0 ± 0.7	73.1 ± 0.4	69.9 ± 0.6	75.1 ± 3.1	73.5 ± 3.7	79.8 ± 4.4	75.6 ± 2.7
NSIT Dwarka		Silence	60.0 ± 2.0	56.4 ± 2.1	58.0 ± 1.0	55.0 ± 1.9	57.0 ± 0.6	54.0 ± 1.0	57.4 ± 0.5	55.6 ± 1.0	57.9 ± 0.9	54.2 ± 4.0	57.4 ± 1.9	54.3 ± 1.8
Gomti Nagar	Lucknow	Silence	–	–	72.0 ± 6.2	64.0 ± 6.4	67.0 ± 1.4	59.0 ± 1.3	66.1 ± 0.6	61.8 ± 2.8	66.9 ± 1.0	60.2 ± 1.1	64.5 ± 2.1	70.9 ± 4.1
Hazrat Gunj		Commercial	73.3 ± 0.8	64.1 ± 1.5	74.0 ± 0.9	66.0 ± 2.3	73.0 ± 5.2	69.0 ± 4.8	71.5 ± 2.6	68.0 ± 2.2	74.0 ± 4.1	73.2 ± 4.5	68.0 ± 3.3	70.5 ± 4.3
Indira Nagar		Residential	61.9 ± 2.8	56.0 ± 4.7	79.0 ± 6.2	77.0 ± 8.4	57.0 ± 1.1	51.0 ± 2.2	55.3 ± 4.0	49.6 ± 4.0	50.6 ± 1.9	44.7 ± 3.0	76.5 ± 4.5	75.8 ± 4.8
PGI Hospital		Silence	63.9 ± 1.0	58.0 ± 2.3	65.0 ± 1.7	59.0 ± 2.9	63.0 ± 3.7	59.0 ± 4.4	60.4 ± 3.1	53.9 ± 3.0	66.7 ± 2.1	67.0 ± 3.7	81.7 ± 3.7	80.1 ± 5.2
Talkatora Industrial Area		Industrial	67.3 ± 0.5	61.0 ± 3.1	68.0 ± 0.9	61.0 ± 2.5	67.0 ± 2.0	61.0 ± 2.7	67.4 ± 2.4	62.1 ± 3.3	63.7 ± 3.4	62.3 ± 4.4	67.1 ± 4.6	68.9 ± 8.7
Kasba Gole Park	Kolkata	Industrial	81.0 ± 6.7	81.0 ± 7.1	81.0 ± 9.7	83.0 ± 4.2	73.0 ± 4.4	73.0 ± 4.2	68.0 ± 2.6	68.3 ± 2.6	64.4 ± 5.2	61.1 ± 3.8	69.5 ± 3.2	66.5 ± 4.0
New Market		Commercial	80.0 ± 1.1	79.0 ± 1.0	87.0 ± 0.5	87.0 ± 0.8	78.0 ± 0.7	78.0 ± 0.8	72.7 ± 2.1	72.7 ± 1.9	72.9 ± 2.8	76.3 ± 4.0	68.8 ± 5.6	67.6 ± 5.1
Patauli		Residential	70.0 ± 1.3	69.0 ± 2.2	76.0 ± 1.4	76.0 ± 4.3	67.0 ± 2.4	67.0 ± 4.3	65.6 ± 2.9	65.2 ± 3.0	73.4 ± 5.2	72.4 ± 5.7	70.6 ± 4.9	74.4 ± 6.0
SSKM Hospital		Silence	63.9 ± 7.3	58.0 ± 8.2	64.0 ± 4.9	59.0 ± 4.2	67.0 ± 4.3	60.0 ± 4.2	72.4 ± 3.4	73.8 ± 2.7	70.4 ± 3.2	69.6 ± 3.9	64.7 ± 6.1	63.0 ± 5.3
WBPCB HQ		Commercial	64.0 ± 9.9	58.1 ± 1.2	62.0 ± 4.4	56.0 ± 4.6	63.0 ± 4.2	57.0 ± 4.6	66.1 ± 3.2	63.9 ± 3.5	63.4 ± 0.9	57.0 ± 1.1	66.4 ± 4.9	65.8 ± 6.3
AS HP	Mumbai	Silence	77.0 ± 1.6	75.0 ± 1.3	74.0 ± 1.1	74.0 ± 1.0	74.0 ± 1.2	74.0 ± 3.8	70.4 ± 4.4	70.5 ± 4.8	–	–	71.7 ± 11.5	69.8 ± 10.8
Bandra		Commercial	68.0 ± 1.3	64.0 ± 1.7	70.0 ± 7.4	69.0 ± 9.0	70.0 ± 4.6	70.0 ± 4.7	64.8 ± 3.2	65.0 ± 3.4	67.5 ± 4.6	66.9 ± 4.7	74.7 ± 5.3	74.4 ± 5.0
MPCB HQ		Commercial	71.0 ± 0.9	70.0 ± 0.7	69.0 ± 2.6	67.0 ± 5.9	66.0 ± 6.3	61.0 ± 12.9	69.0 ± 2.3	74.7 ± 7.5	70.2 ± 0.5	67.3 ± 3.3	72.9 ± 4.6	64.1 ± 5.1
Thane MCQ		Commercial	64.0 ± 7.6	57.0 ± 9.4	73.0 ± 1.9	75.0 ± 2.1	61.0 ± 1.9	58.0 ± 2.3	61.0 ± 4.0	57.1 ± 2.7	69.1 ± 0.5	69.1 ± 1.1	67.9 ± 0.9	68.6 ± 1.3
Vashi Hospital		Silence	69.0 ± 1.2	59.0 ± 1.5	70.0 ± 2.2	65.0 ± 4.2	78.0 ± 1.9	83.0 ± 4.7	72.1 ± 2.3	68.5 ± 3.9	75.8 ± 7.2	74.2 ± 7.7	69.4 ± 4.7	69.1 ± 5.1
Abids	Hyderabad	Commercial	77.6 ± 0.8	69.7 ± 1.8	53.0 ± 2.0	51.0 ± 2.6	74.0 ± 0.8	66.0 ± 1.5	74.4 ± 0.8	68.2 ± 1.3	73.1 ± 0.7	65.7 ± 1.4	72.7 ± 4.7	67.4 ± 4.7
Jeedimetla		Industrial	68.0 ± 0.6	63.0 ± 0.9	57.0 ± 1.3	57.0 ± 3.6	74.0 ± 1.5	85.0 ± 1.6	80.0 ± 1.0	80.1 ± 4.6	72.6 ± 5.3	73.2 ± 6.9	71.7 ± 4.0	70.0 ± 4.6
Jubilee Hills		Residential	60.0 ± 0.5	53.2 ± 0.8	58.0 ± 3.3	55.0 ± 5.2	61.0 ± 0.7	55.0 ± 1.3	61.4 ± 1.2	55.9 ± 1.2	59.7 ± 1.2	53.8 ± 1.1	74.4 ± 4.0	69.1 ± 4.6
Punjagutta		Commercial	80.0 ± 1.8	75.0 ± 3.0	66.0 ± 2.6	57.0 ± 3.5	72.0 ± 0.6	65.0 ± 1.0	70.7 ± 1.4	64.8 ± 1.4	69.8 ± 0.9	62.9 ± 0.8	68.2 ± 2.3	61.8 ± 2.5
Zoo Park		Silence	56.0 ± 0.7	51.0 ± 1.5	75.0 ± 2.0	71.0 ± 1.7	56.0 ± 1.9	50.0 ± 1.5	55.2 ± 0.8	50.4 ± 1.0	55.8 ± 3.4	61.4 ± 3.3	69.0 ± 8.9	68.1 ± 8.8
BTM	Bengaluru	Residential	65.8 ± 3.7	58.0 ± 5.0	66.0 ± 2.9	62.0 ± 2.3	66.0 ± 3.4	66.0 ± 4.0	66.6 ± 3.6	65.9 ± 3.6	64.7 ± 3.0	64.7 ± 3.3	62.5 ± 5.3	62.5 ± 4.6
Marathahalli		Commercial	59.0 ± 0.5	57.2 ± 0.9	56.0 ± 1.1	56.0 ± 1.2	64.0 ± 0.5	63.0 ± 1.1	67.4 ± 4.5	66.2 ± 5.2	73.1 ± 3.7	72.7 ± 7.2	71.8 ± 6.1	71.4 ± 6.4
Nisarga Bhawan		Residential	58.0 ± 2.6	52.0 ± 4.3	58.0 ± 1.5	53.0 ± 4.6	57.0 ± 1.0	50.0 ± 1.0	75.2 ± 3.7	70.0 ± 4.1	76.4 ± 4.0	79.1 ± 8.0	80.2 ± 8.8	77.4 ± 8.6
Parisar Bhawan		Commercial	66.7 ± 0.8	59.0 ± 1.1	67.0 ± 1.9	62.0 ± 3.1	67.0 ± 2.3	61.0 ± 2.4	67.4 ± 1.0	61.6 ± 1.0	67.2 ± 0.5	62.1 ± 0.6	64.3 ± 2.6	58.7 ± 3.4
Peeniya		Industrial	61.0 ± 1.2	59.0 ± 1.2	64.0 ± 0.5	57.0 ± 2.4	66.0 ± 9.3	60.0 ± 7.9	64.1 ± 1.1	57.0 ± 0.6	63.4 ± 0.8	58.4 ± 1.4	62.6 ± 2.5	60.9 ± 2.7
Eye Hospital	Chennai	Silence	67.0 ± 3.1	60.0 ± 4.3	67.0 ± 1.8	59.0 ± 2.3	67.0 ± 0.0	62.0 ± 3.3	66.4 ± 6.3	58.7 ± 3.3	69.2 ± 3.8	61.0 ± 4.3	72.8 ± 6.9	69.7 ± 5.3
Guindy		Industrial	79.8 ± 1.0	75.7 ± 2.5	79.0 ± 0.6	76.0 ± 1.9	76.0 ± 0.4	72.0 ± 2.0	76.8 ± 0.4	74.9 ± 1.4	80.1 ± 3.5	78.2 ± 4.8	82.8 ± 6.8	83.6 ± 7.1
Perambur		Commercial	72.0 ± 1.5	65.0 ± 4.7	73.0 ± 1.2	71.0 ± 4.5	75.0 ± 3.7	76.0 ± 4.5	75.8 ± 1.3	76.8 ± 5.7	74.4 ± 2.3	71.6 ± 2.7	63.1 ± 3.9	59.4 ± 5.4
T. Nagar		Commercial	76.6 ± 0.4	69.3 ± 0.7	76.0 ± 1.6	68.0 ± 5.2	75.0 ± 3.0	67.0 ± 1.0	75.8 ± 1.3	70.2 ± 2.0	77.3 ± 0.9	73.0 ± 1.9	80.3 ± 6.5	75.9 ± 4.0
Triplicane		Residential	73.0 ± 3.1	68.0 ± 5.6	73.0 ± 5.8	72.0 ± 7.3	69.0 ± 10.4	58.0 ± 4.6	67.1 ± 3.5	63.8 ± 3.5	60.9 ± 3.1	59.0 ± 4.1	70.0 ± 4.6	68.9 ± 5.0

Table 4. Annual average ambient levels L_{day} and L_{night} for additional 35 noise monitoring stations installed in November 2014 across 7 major cities from 2015 to 2020 (CPCB, 2015a; 2015b; 2016; 2018; Garg, 2022).

Name of location	City	Area characteristics	2015		2016		2017		2018		2019		2020	
			L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}
Civil Lines	Delhi	Commercial	73.0 ± 5.3	73.0 ± 5.3	69.0 ± 1.0	70.0 ± 1.9	66.0 ± 4.7	65.0 ± 5.7	61.2 ± 0.5	58.8 ± 0.6	61.7 ± 0.6	58.3 ± 0.5	60.6 ± 1.0	57.3 ± 1.1
R.K. Puram		Residential	66.0 ± 6.6	59.0 ± 7.6	65.0 ± 5.0	62.0 ± 7.8	62.0 ± 2.9	56.0 ± 4.9	62.8 ± 2.5	60.3 ± 4.4	66.8 ± 2.9	61.6 ± 1.0	63.3 ± 3.1	59.2 ± 3.3
Anand Vihar		Commercial	70.0 ± 3.3	68.0 ± 5.6	68.0 ± 0.8	67.0 ± 4.1	68.0 ± 1.4	65.0 ± 1.4	66.6 ± 1.3	65.9 ± 2.9	65.6 ± 0.4	61.9 ± 0.6	65.5 ± 5.0	60.6 ± 4.2
Mandir Marg		Silence	76.0 ± 5.8	73.0 ± 4.8	76.0 ± 12.8	61.0 ± 9.4	66.0 ± 7.6	49.0 ± 1.8	61.0 ± 3.9	62.6 ± 2.3	65.4 ± 2.6	54.8 ± 3.0	58.3 ± 3.5	56.6 ± 2.5
Punjabi Bagh		Residential	86.0 ± 1.1	81.0 ± 2.3	59.0 ± 1.0	53.0 ± 1.7	64.0 ± 5.7	62.0 ± 8.5	59.8 ± 1	53.4 ± 1.7	61.3 ± 1.1	52.4 ± 1.3	57.3 ± 1.2	51.7 ± 1.2
Chinhat	Lucknow	Industrial	68.6 ± 5.5	63.0 ± 9.1	68.0 ± 3.4	62.0 ± 5.2	65.0 ± 2.1	56.0 ± 1.8	66.7 ± 1.7	57.9 ± 2.2	67.3 ± 2.4	60.0 ± 4.0	67.1 ± 3.5	56.5 ± 4.8
IT College		Silence	65.0 ± 4.1	57.4 ± 1.3	65.0 ± 2.0	59.0 ± 1.9	68.0 ± 3.5	62.0 ± 4.6	65.3 ± 1.0	70.0 ± 3.8	64.7 ± 1.0	59.5 ± 1.3	64.7 ± 2.3	71.5 ± 2.8
CSS Airport		Commercial	74.7 ± 8.6	66.1 ± 6.0	76.0 ± 0.9	67.0 ± 0.8	76.0 ± 0.8	67.0 ± 4.7	59.8 ± 1.2	70.0 ± 3.8	76.3 ± 0.9	67.0 ± 1.1	62.7 ± 2.1	58.3 ± 2.7
RSC Aliganj		Commercial	63.0 ± 0.8	58.1 ± 0.7	66.0 ± 0.7	61.0 ± 0.7	66.0 ± 0.5	60.0 ± 0.6	76.7 ± 0.5	70.9 ± 2.6	65.1 ± 0.6	57.6 ± 0.6	73.6 ± 4.2	65.0 ± 4.6
Vibhuti Khand		Residential	63.4 ± 1.3	57.0 ± 2.1	64.0 ± 3.3	60.0 ± 4.8	64.0 ± 0.9	60.0 ± 2.2	66.4 ± 2.8	59.9 ± 1.7	63.9 ± 0.7	62.2 ± 1.3	62.7 ± 1.7	57.5 ± 1.4
Birati N.	Kolkata	Residential	63.2 ± 2.6	58.0 ± 4.2	67.0 ± 5.9	70.0 ± 3.6	64.0 ± 2.8	58.0 ± 3.6	80.4 ± 1.8	80.1 ± 3.5	79.1 ± 8.9	77.9 ± 10.6	66.9 ± 3.5	63.3 ± 3.8
R.G. Kar		Silence	70.0 ± 8.6	63.0 ± 6.0	64.0 ± 0.9	60.0 ± 1.7	64.0 ± 0.8	60.0 ± 1.7	66.5 ± 1.9	60.4 ± 1.3	63.7 ± 3.3	62.2 ± 3.0	68.7 ± 3.1	71.8 ± 3.0
Tollygunge		Commercial	66.8 ± 0.5	62.9 ± 0.7	67.0 ± 0.4	63.0 ± 0.6	67.0 ± 2.4	63.0 ± 0.6	67.2 ± 0.4	64.1 ± 0.4	66.8 ± 0.4	62.9 ± 0.6	65.5 ± 1.5	61.6 ± 1.4
Bag Bazar		Residential	78.0 ± 4.4	70.0 ± 2.5	78.0 ± 3.5	69.0 ± 1.0	75.0 ± 0.9	68.0 ± 1.0	75.3 ± 1.1	70.7 ± 1.8	87.2 ± 4.6	90.1 ± 3.9	77.1 ± 9.5	74.6 ± 11.2
Tartala		Industrial	64.7 ± 1.1	61.1 ± 1.7	65.0 ± 1.2	62.0 ± 1.3	65.0 ± 1.6	62.0 ± 1.3	69.3 ± 1.2	67.2 ± 1.2	75.3 ± 1.9	73.1 ± 1.6	68.1 ± 2.0	64.4 ± 2.0
M&M Kandivali	Mumbai	Industrial	62.1 ± 1.1	55.8 ± 1.6	66.0 ± 4.5	69.0 ± 11.5	62.0 ± 0.7	57.0 ± 2.7	61.1 ± 0.7	54.8 ± 1.2	60.4 ± 1.0	52.6 ± 1.5	60.8 ± 2.9	53.8 ± 2.4
Ambassador Hotel		Commercial	73.1 ± 0.9	68.5 ± 1.3	73.0 ± 0.8	68.0 ± 0.8	72.0 ± 0.9	67.0 ± 1.1	72.5 ± 0.6	68.1 ± 0.8	74.5 ± 2.9	70.3 ± 3.2	74.2 ± 3.8	69.6 ± 5.1
L&T Powai		Industrial	60.3 ± 1.2	57.4 ± 1.5	60.0 ± 1.2	57.0 ± 1.9	59.0 ± 1.7	55.0 ± 2.7	58.4 ± 1.6	53.6 ± 2.3	59.4 ± 2.7	55.1 ± 3.2	62.2 ± 2.8	60.0 ± 4.8
Pepsico Chembur		Residential	67.0 ± 4.3	63.0 ± 7.8	64.0 ± 1.0	58.0 ± 2.1	64.0 ± 1.6	60.0 ± 4.8	66.8 ± 2.6	60.6 ± 2.9	68.1 ± 1.9	61.3 ± 1.4	76.5 ± 4.5	75.4 ± 4.6
Andheri		Industrial	71.3 ± 1.4	67.6 ± 0.8	72.0 ± 2.7	67.0 ± 0.8	71.0 ± 1.6	68.0 ± 1.6	71.7 ± 2.1	67.7 ± 1.5	78.9 ± 3.1	78.0 ± 4.0	80.1 ± 5.3	79.6 ± 5.7
Tarnaka	Hyderabad	Residential	60.2 ± 1.1	54.2 ± 3.2	69.0 ± 2.7	70.0 ± 5.6	67.0 ± 5.6	62.0 ± 7.6	73.5 ± 2.8	72.6 ± 3.4	81.4 ± 10.7	81.7 ± 11.2	76.6 ± 6.4	70.0 ± 5.6
Gaddapothram		Industrial	65.0 ± 2.0	59.0 ± 3.5	65.0 ± 3.8	62.0 ± 5.1	78.0 ± 4.4	75.0 ± 3.9	75.1 ± 4.5	67.8 ± 5.3	83.3 ± 5.0	74.0 ± 6.9	77.3 ± 4.8	75.2 ± 5.2
Gachibowli		Silence	60.9 ± 2.2	55.0 ± 3.4	68.0 ± 4.9	67.0 ± 4.8	63.0 ± 5.6	60.0 ± 3.4	61.0 ± 2.3	56.1 ± 3.1	60.0 ± 2.5	58.2 ± 2.7	60.9 ± 2.4	71.7 ± 5.5
Paradise		Commercial	79.1 ± 0.9	74.9 ± 1.0	76.0 ± 3.1	61.0 ± 1.8	76.0 ± 4.6	75.0 ± 7.7	84.6 ± 3.5	85.6 ± 3.5	81.4 ± 2.5	80.3 ± 2.9	78.0 ± 2.7	73.9 ± 3.3
Kukatpalli		Commercial	70.3 ± 1.1	66.4 ± 1.0	59.0 ± 2.2	53.0 ± 2.7	71.0 ± 1.1	66.0 ± 1.7	70.6 ± 1.1	67.5 ± 0.8	69.9 ± 0.7	66.7 ± 0.8	68.2 ± 2.4	65.4 ± 3.1
Yeshwantpur	Bengaluru	Commercial	72.1 ± 0.5	63.4 ± 0.7	72.0 ± 0.5	64.0 ± 0.8	72.0 ± 0.5	64.0 ± 0.9	72.0 ± 0.6	65.5 ± 0.8	72.1 ± 0.4	64.0 ± 0.6	70.2 ± 2.1	64.0 ± 2.2
R.V.C.E		Silence	60.1 ± 0.7	54.1 ± 0.7	60.0 ± 0.7	55.0 ± 0.8	60.0 ± 1.2	55.0 ± 0.7	59.6 ± 0.9	53.9 ± 0.6	67.4 ± 0.9	65.9 ± 2.6	80.9 ± 3.0	80.9 ± 4.2
Whitefield		Industrial	67.1 ± 0.7	61.2 ± 0.6	68.0 ± 1.7	62.0 ± 0.2	72.0 ± 5.6	66.0 ± 4.8	66.9 ± 1.1	72.5 ± 4.2	66.2 ± 0.8	61.1 ± 1.1	63.9 ± 1.9	58.6 ± 2.1
Dolmur		Residential	66.0 ± 3.3	60.0 ± 4.5	64.0 ± 1.8	59.0 ± 1.9	68.0 ± 5.7	61.0 ± 5.8	64.5 ± 1.8	59.2 ± 2.5	64.3 ± 1.8	59.3 ± 2.6	62.9 ± 1.5	58.1 ± 2.7
Nihmans		Silence	63.0 ± 3.4	60.0 ± 5.1	63.0 ± 2.9	65.0 ± 8.1	67.0 ± 6.8	66.0 ± 9.2	66.6 ± 5.6	64.5 ± 3.8	72.0 ± 3.9	74.1 ± 3.0	70.1 ± 5.4	74.9 ± 4.0
Pallikarnai	Chennai	Commercial	73.8 ± 0.6	67.2 ± 0.6	77.0 ± 4.3	71.0 ± 0.5	78.0 ± 4.2	77.0 ± 6.7	72.6 ± 1.9	72.5 ± 2.2	84.4 ± 4.9	78.6 ± 5.0	72.0 ± 5.4	72.6 ± 9.9
Velachery		Residential	67.0 ± 3.2	67.0 ± 9.4	68.0 ± 3.8	70.0 ± 10.4	74.0 ± 6.0	69.0 ± 7.7	77.3 ± 4.9	76.2 ± 8.2	65.5 ± 1.8	59.7 ± 4.1	64.4 ± 2.6	67.6 ± 2.0
Washermanpet		Commercial	70.0 ± 1.4	67.0 ± 5.7	72.0 ± 1.9	65.0 ± 2.6	72.0 ± 2.0	66.0 ± 5.7	74.4 ± 2.8	72.9 ± 4.2	75.6 ± 4.4	71.5 ± 4.3	74.0 ± 4.4	74.1 ± 5.3
Anna Nagar		Silence	67.0 ± 1.7	63.0 ± 4.6	68.0 ± 2.5	66.0 ± 7.9	66.0 ± 1.4	60.0 ± 3.0	67.1 ± 1.7	61.5 ± 3.4	75.0 ± 4.3	77.8 ± 4.0	73.4 ± 5.7	72.8 ± 7.8
Sowcarpet		Residential	66.6 ± 1.2	60.0 ± 2.7	70.0 ± 3.2	69.0 ± 6.8	68.0 ± 3.4	65.0 ± 5.7	66.9 ± 1.4	62.9 ± 3.1	62.3 ± 3.5	60.2 ± 3.0	65.1 ± 4.7	60.7 ± 3.6

Table 5. Frequency distribution of L_{day}, L_{night}, $L_{\text{Aeq,24h}}$, and L_{dn} [dB(A)] for 7 major cities (70 sites) for 2017–2020.

Variation of noise descriptors	L_{day}				L_{night}				L_{dn}				$L_{\text{Aeq,24h}}$			
	2017	2018	2019	2020	2017	2018	2019	2020	2017	2018	2019	2020	2017	2018	2019	2020
$40 < L_{\text{eq}} \leq 45$ dB(A)	0	0	0	0	0	0	1.4	0	0	0	0	0	0	0	0	0
$45 < L_{\text{eq}} \leq 50$ dB(A)	0	0	0	0	4.3	1.4	0	0	0	0	0	0	0	0	1.5	0
$50 < L_{\text{eq}} \leq 55$ dB(A)	1.4	0	1.4	0	10.0	10.0	7.2	5.7	0	0	1.4	0	5.7	4.3	0.0	0
$55 < L_{\text{eq}} \leq 60$ dB(A)	10	11.4	5.8	5.7	25.7	17.1	18.8	15.7	7.1	4.3	0.0	1.4	8.6	12.9	10.3	7.5
$60 < L_{\text{eq}} \leq 65$ dB(A)	21.4	17.1	23.2	25.7	22.9	21.4	24.6	15.7	12.9	12.9	10.1	8.6	34.3	24.3	30.9	34.3
$65 < L_{\text{eq}} \leq 70$ dB(A)	32.9	34.3	29.0	25.7	22.9	25.7	14.5	25.7	35.7	24.3	36.2	21.4	25.7	24.3	20.6	28.4
$70 < L_{\text{eq}} \leq 75$ dB(A)	24.3	21.4	18.8	24.3	7.1	17.1	18.8	22.9	21.4	22.9	14.5	21.4	18.6	24.3	20.6	14.9
$75 < L_{\text{eq}} \leq 80$ dB(A)	10	12.9	13.0	10.0	4.3	2.9	10.1	10.0	11.4	25.7	18.8	25.7	7.1	4.3	8.8	7.5
$80 < L_{\text{eq}} \leq 85$ dB(A)	0	2.9	7.2	8.6	2.9	2.9	2.9	4.3	10.0	5.7	13.0	15.7	0	2.9	5.9	6.0
$85 < L_{\text{eq}} \leq 90$ dB(A)	0	0	1.4	0	0	1.4	0	0	1.4	2.9	4.3	5.7	0	1.4	1.5	1.5
$90 < L_{\text{eq}} \leq 95$ dB(A)	0	0	0	0	0	0	1.4	0	0	1.4	1.4	0	0	1.4	0	0

Table 6. Frequency distribution of annual average L_{day} and L_{night} values observed [dB] for the 70 sites spread across the 7 major cities of India for 2018–2020 (Garg, 2022).

Variation of L_{day} and L_{night} [dB]	Percentage of noise monitoring locations in various zones																							
	Commercial						Residential						Silence						Industrial					
	2018		2019		2020		2018		2019		2020		2018		2019		2020		2018		2019		2020	
	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}	L_{day}	L_{night}
$40 < L_{\text{eq}} \leq 45$	0	0	0	0	0	0	0	0	0	6.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$45 < L_{\text{eq}} \leq 50$	0	0	0	0	0	0	0	6.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$50 < L_{\text{eq}} \leq 55$	0	0	0	0	0	0	0	6.3	6.3	12.5	0	6.3	0	23.5	0	12.5	0	11.8	0	16.7	0	8.3	0	8.3
$55 < L_{\text{eq}} \leq 60$	4.0	12.0	0	16.0	0	16.0	12.5	18.8	6.3	18.8	6.3	18.8	23.5	23.5	12.5	18.8	17.6	5.9	8.3	16.7	8.3	25.0	0	25.0
$60 < L_{\text{eq}} \leq 65$	12.0	20.0	8.0	20.0	20.0	20.0	18.8	25.0	37.5	31.3	31.3	18.8	23.5	29.4	25.0	25.0	23.5	41.2	16.7	8.3	33.3	25.0	33.3	16.7
$65 < L_{\text{eq}} \leq 70$	28.0	36.0	36.0	24.0	32.0	32.0	37.5	18.8	18.8	0	18.8	18.8	35.3	11.8	37.5	25.0	17.6	23.5	41.7	33.3	16.7	0	33.3	25.0
$70 < L_{\text{eq}} \leq 75$	40.0	24.0	32.0	28.0	36.0	24.0	6.3	12.5	6.3	6.3	12.5	18.8	17.6	11.8	18.8	12.5	29.4	5.9	8.3	16.7	8.3	25.0	8.3	0
$75 < L_{\text{eq}} \leq 80$	12.0	4.0	16.0	8.0	8.0	8.0	18.8	6.3	12.5	12.5	25.0	18.8	0	0	6.3	6.3	0	0	25.0	0	16.7	16.7	8.3	16.7
$80 < L_{\text{eq}} \leq 85$	4.0	0	8.0	4.0	4.0	0	6.3	6.3	6.3	6.3	6.3	0	0	0	0	0	11.8	11.8	0	8.3	16.7	0	16.7	8.3
$85 < L_{\text{eq}} \leq 90$	0	4.0	0	0	0	0	0	0	6.3	6.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$90 < L_{\text{eq}} \leq 95$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 7. Frequency distribution [%] of difference of annual average (L_{day}–L_{night}) values observed [dB] for the 70 sites spread across the 7 major cities of India from 2011 to 2020 (Garg, 2022).

Variation of difference (L_{day}–L_{night}) values [dB]	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
$-15 < (L_{day}–L_{night}) \leq -10$	0	0	0	0	0	0	1.4	1.4	0	1.4
$-10 < (L_{day}–L_{night}) \leq -5$	0	0	0	0	0	0	1.4	2.9	1.4	2.9
$-5 < (L_{day}–L_{night}) \leq 0$	0	0	0	0	0	20.0	10.0	14.3	13.0	18.6
$0 < (L_{day}–L_{night}) \leq 5$	34.3	28.6	40.0	31.4	50.7	47.1	47.1	45.7	52.2	58.6
$5 < (L_{day}–L_{night}) \leq 10$	54.3	57.1	48.6	57.1	35.0	30.0	37.1	35.7	31.9	17.1
$10 < (L_{day}–L_{night}) \leq 15$	11.4	14.3	11.4	11.4	14.3	2.9	1.4	0	1.4	1.4
$15 < (L_{day}–L_{night}) \leq 20$	0	0	0	0	0	0	1.4	0	0	0

Table 8. Slope and correlation coefficient of L_{day} and L_{night} levels with respect to the transition period from 2011 to 2020 for the 35 sites in 7 major cities of India.

Area characteristics	Name of location	City	L_{day}		L_{night}	
			Slope	R	Slope	R
Commercial	CPCB HQ	Delhi	0.31	0.4654	1.14	0.7904
	ITO	Delhi	0.44	0.6006	0.24	0.2953
	Hazrat Ganj	Lucknow	0.15	0.1867	1.54	0.9561
	New Market	Kolkata	0.68	0.3108	1.82	0.5730
	WBPCB HQ	Kolkata	0.44	0.7374	0.90	0.7158
	Bandra	Mumbai	0.08	0.0938	0.37	0.3874
	MPCB HQ	Mumbai	0.43	0.5622	0.39	0.2948
	Thane MCQ	Mumbai	0.59	0.4416	1.66	0.6626
	Abids	Hyderabad	0.01	0.0038	0.39	0.2269
	Punjagutta	Hyderabad	–1.02	0.6629	–1.25	0.6604
	Marathahalli	Bengaluru	2.06	0.9016	2.37	0.9380
	Parisar Bhawan	Bengaluru	0.11	0.2956	0.48	0.6852
	Perambur	Chennai	0.31	0.2378	1.62	0.6087
	T. Nagar	Chennai	0.63	0.8749	1.40	0.9415
Industrial	Talkatora Industrial Area	Lucknow	0.40	0.5929	1.28	0.9274
	Kasba Gole Park	Kolkata	0.33	0.1598	0.66	0.2436
	Jeedimetla	Hyderabad	1.55	0.6787	2.68	0.7408
	Peeniya	Bengaluru	1.02	0.8483	1.03	0.8391
	Guindy	Chennai	0.64	0.7353	1.11	0.8264
Residential	Indira Nagar	Lucknow	1.27	0.3871	1.69	0.4280
	Patauli	Kolkata	2.21	0.7883	3.27	0.8415
	Jubilee Hills	Hyderabad	1.28	0.7198	1.57	0.7931
	BTM	Bengaluru	–0.23	0.5846	1.21	0.8633
	Nisarga Bhawan	Bengaluru	2.79	0.8405	3.71	0.8680
	Triplicane	Chennai	–0.20	0.1795	1.16	0.5621
Silence	DTU, Bawana	Delhi	0.92	0.6568	0.89	0.5614
	NSIT Dwarka	Delhi	0.21	0.4844	0.11	0.3219
	Gomti Nagar	Lucknow	–0.19	0.2132	0.89	0.6300
	PGI Hospital	Lucknow	1.86	0.7797	2.44	0.8167
	SSKM Hospital	Kolkata	0.96	0.7594	1.66	0.7824
	AS HP	Mumbai	0.97	0.6247	1.78	0.7320
	Vashi Hospital	Mumbai	0.63	0.5659	2.16	0.7474
	Zoo Park	Hyderabad	1.10	0.4516	1.83	0.6328
	Eye Hospital	Chennai	0.97	0.8430	1.73	0.9048
	Dilshad Garden	Delhi	2.12	0.8601	2.26	0.8430

Table 9. Slope and correlation coefficient of L_{day} and L_{night} levels with respect to the transition period within the years 2015–2020 for the 35 sites in 7 major cities of India.

Area characteristics	Name of location	City	L_{day}		L_{night}	
			Slope	R	Slope	R
Commercial	Civil Lines	Delhi	−2.53	0.9471	−3.42	0.9610
	Anand Vihar	Delhi	−0.89	0.9634	−1.47	0.9440
	CSS Airport	Lucknow	−2.15	0.5321	−1.03	0.4865
	RSC Aliganj	Lucknow	1.74	0.6003	1.01	0.3723
	Tollygunge	Kolkata	−0.20	0.6002	−0.16	0.3840
	Ambassador Hotel	Mumbai	0.30	0.5808	0.39	0.6069
	Paradise	Hyderabad	0.55	0.3085	1.81	0.4131
	Kukatpalli	Hyderabad	0.62	0.2536	1.07	0.3644
	Yeshwantpur	Bengaluru	−0.26	0.6533	0.13	0.3419
	Pallikarnai	Chennai	0.22	0.0900	1.29	0.5863
	Washermanpet	Chennai	0.95	0.8718	1.77	0.8519
Industrial	Chinhat	Lucknow	−0.23	0.3410	−1.05	0.6739
	Tartala	Kolkata	1.49	0.6815	1.57	0.6446
	M&M Kandivali	Mumbai	−0.69	0.6343	−1.75	0.5474
	L&T Powai	Mumbai	0.20	0.2866	0.17	0.1388
	Andheri	Mumbai	1.87	0.8398	2.65	0.8503
	Gaddapothram	Hyderabad	3.24	0.8157	3.14	0.8305
	Whitefield	Bengaluru	−0.76	0.5313	−0.26	0.0985
Residential	R.K. Puram	Delhi	−0.21	0.2047	0.12	0.1007
	Punjabi Bagh	Delhi	−4.02	0.7007	−4.48	0.7315
	Vibhuti Khand	Lucknow	−0.04	0.0599	0.26	0.2523
	Birati N.	Kolkata	2.03	0.4982	2.07	0.3984
	Bag Bazar	Kolkata	0.67	0.2788	2.54	0.5710
	Pepsico Chembur	Mumbai	1.79	0.7258	2.07	0.6183
	Tarnaka	Hyderabad	3.59	0.8948	3.56	0.7087
	Dolmur	Bengaluru	−0.52	0.5386	−0.30	0.5666
	Velachery	Chennai	−0.49	0.1793	−0.59	0.2077
	Sowcarpet	Chennai	−0.91	0.6472	−0.71	0.3791
Silence	Mandir Marg	Delhi	−3.58	0.8996	−2.49	0.5682
	IT College	Lucknow	−0.15	0.2148	2.29	0.7097
	R.G. Kar	Kolkata	−0.14	0.0967	1.20	0.5099
	Gachibowli	Hyderabad	−0.74	0.4691	1.52	0.4300
	R.V.C.E.	Bengaluru	3.59	0.7913	4.73	0.8150
	Nihmans	Bengaluru	1.77	0.9084	2.87	0.9140
	Anna Nagar	Chennai	1.55	0.7624	2.45	0.6541

Table 10. Status of compliance of day and night equivalent levels explicitly for the various sites with ambient noise standards of India.

Category	Number of compliant stations: day and night time exclusively																			
	2011		2012		2013		2014		2015		2016		2017		2018		2019		2020	
	a*	b*	a*	b*	a*	b*	a*	b*	a*	b*	a*	b*	a*	b*	a*	b*	a*	b*	a*	b*
Silence	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Residential	2	0	2	0	2	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0
Commercial	4	3	5	4	5	4	5	1	4	0	4	2	3	0	4	0	3	0	5	0
Industrial	4	4	5	4	5	4	4	4	10	10	10	10	10	8	10	9	9	9	9	9
Total	10	7	12	8	12	8	10	5	14	10	14	12	13	8	13	9	13	8	14	9

a* – Day-time, b* – Night-time.

Table 11. Status of compliance of various sites in the NANMN project with ambient noise standards of India (Garg, 2022).

Year	Number of sites meeting the ambient noise criteria	Name of the sites
2011	7 sites (3 commercial and 4 industrial)	CPCB Headquarters, Thane, Marathali, Peeniya, Jeedimetla, Talkatora, and Kasba Gole Park
2012	8 sites (4 commercial and 4 industrial)	CPCB Headquarters, West Bengal Headquarters, Thane, Marathali, Peeniya, Jeedimetla, Talkatora, and Kasba Gole Park
2013	8 sites (4 commercial and 4 industrial)	CPCB Headquarters, West Bengal Headquarters, Thane, Marathali, Peeniya, Jeedimetla, Talkatora, and Kasba Gole Park
2014	5 sites (1 commercial and 4 industrial)	CPCB Headquarters, Peeniya, Jeedimetla, Talkatora, and Kasba Gole Park
2015	10 industrial sites	Talkatora, Jeedimetla, Peeniya, Chinhat, Tartala, M&M Kandivali, L&T Powai, Andheri, Whitefield, and Gaddapothram
2016	12 sites (10 industrial and 2 commercial)	Talkatora, Jeedimetla, Peeniya, Chinhat, Tartala, M&M Kandivali, L&T Powai, Andheri, Whitefield, Gaddapothram, Kukatpalli, and Abids
2017	8 industrial sites	Talkatora, Peeniya, Chinhat, Tartala, M&M Kandivali, L&T Powai, Andheri, and Whitefield
2018	9 industrial sites	Peeniya, Gole Park, Tartala, Talkatora, Chinhat, M&M Kandivali, L&T Powai, Andheri, and Gaddapothram
2019	8 sites (7 industrial and 1 residential)	Whitefield, Gole Park, Talkatora, Peeniya, Indira Nagar, Chinhat, Kandivali, and L&T Powai
2020	9 industrial sites	Peeniya, Whitefield, Jeedimetla, Gole Park, Tartala, Talkatora, Chinhat, Kandivali, and L&T Powai

Table 12. AEF for different zones in 2017–2020.

Category of area/zone	L_{day}			L_{night}		
	Maximum value	Minimum value	AEF	Maximum value	Minimum value	AEF
2017 annual average values						
Industrial area (n = 12)	1.0	0.8	0.9	1.2	0.8	0.9
Commercial area (n = 25)	1.2	0.9	1.1	1.4	1.0	1.2
Residential area (n = 16)	1.4	1.0	1.2	1.5	1.1	1.3
Silence zone (n = 17)	1.6	1.1	1.3	2.1	1.2	1.5
2018 annual average values						
Industrial area (n = 12)	1.1	0.8	0.9	1.1	0.8	0.9
Commercial area (n = 25)	1.3	0.9	1.1	1.6	1.0	1.2
Residential area (n = 16)	1.5	1.0	1.2	1.8	1.1	1.4
Silence zone (n = 17)	1.4	1.1	1.3	1.8	1.3	1.5
2019 annual average values						
Industrial area (n = 12)	1.1	0.8	0.9	1.1	0.8	0.9
Commercial area (n = 25)	1.3	0.9	1.1	1.5	1.0	1.2
Residential area (n = 16)	1.6	0.9	1.2	2.0	1.0	1.4
Silence zone (n = 17)	1.5	1.1	1.3	1.9	1.4	1.6
2020 annual average values						
Industrial area (n = 12)	1.1	0.8	0.9	1.2	0.8	0.9
Commercial area (n = 25)	1.2	0.9	1.1	1.4	1.0	1.2
Residential area (n = 16)	1.5	1.0	1.3	1.7	1.1	1.5
Silence zone (n = 17)	1.6	1.1	1.4	2.0	1.4	1.7

Table 13. Inter-conversion of the various noise descriptors for the 4 zones.

Type of zone		L_{day}	L_{night}	L_{dn}	$L_{\mathrm{eq,24h}}$
Commercial	L_{day}	–	+4.0	–3.3	+0.7
	L_{night}	–4.0	–	–7.4	–3.4
	L_{dn}	+3.3	+7.4	–	+4.0
	$L_{\mathrm{eq,24h}}$	–0.7	+3.4	–4.0	–
Industrial	L_{day}	–	+3.2	–3.9	+0.5
	L_{night}	–3.2	–	–7.2	–2.7
	L_{dn}	+3.9	+7.2	–	+4.4
	$L_{\mathrm{eq,24h}}$	–0.5	+2.7	–4.4	–
Residential	L_{day}	–	+3.9	–3.5	+0.5
	L_{night}	–3.9	–	–7.8	–3.9
	L_{dn}	+3.5	+7.8	–	+3.8
	$L_{\mathrm{eq,24h}}$	–0.5	+3.9	–3.8	–
Silence	L_{day}	–	+2.3	–3.5	+0.5
	L_{night}	–3.9	–	–7.8	–3.9
	L_{dn}	+3.5	+7.8	–	+3.8
	$L_{\mathrm{eq,24h}}$	–0.5	+3.9	–3.8	–

References

1. Agarwal S., Swami B.L. (2011), Road traffic noise, annoyance and community health survey – A case study for an Indian city, *Noise and Health*, **13**(53): 272–276, doi: 10.4103/1463-1741.82959.

2. Babisch W. (2002), The noise/stress concept, risk assessment and research needs, *Noise and Health*, **4**(16): 1–11.

3. Babisch W., Beule B., Schust M., Kersten N., Ising H. (2005), Traffic noise and risk of myocardial infarction, *Epidemiology*, **16**(1): 33–40, doi: 10.1097/01.ede.0000147104.84424.24.

4. Bhosale B.J., Late A., Nalawade P.M., Chavan S.P., Mule M.B. (2010), Studies on assessment of traffic noise level in Aurangabad city, India, *Noise and Health*, **12**(48): 195–198, doi: 10.4103/1463-1741.64971.

5. Blanes N., Domingues D., José Ramos M., Sáinz de la Maza M., Fons-Esteve J., Peris E. (2019), *ETC/ATNI Report 8/2019: Noise Action Plans. Managing Exposure to Noise in Europe*, European Environment Agency, https://www.eionet.europa.eu/etcs/etc-atni/products/etc-atni-reports/etc-atni-report-8-2019-noise-action-plans-managing-exposure-to-noise-in-europe (access: 11.06.2021).

6. Brink M., Schäffer B., Pieren R., Wunderli J.M. (2018), Conversion between noise exposure indicators Leq_{24h}, L_{Day}, $L_{Evening}$, L_{Night}, L_{dn} and L_{den}: Principles and practical guidance, *International Journal of Hygiene and Environmental Health*, **221**(1): 54–63, doi: 10.1016/j.ijheh.2017.10.003.

7. Central Pollution Control Board (CPCB) (2011), *National Ambient Noise Monitoring Network. Information Brochure*, https://cpcb.nic.in/.

8. Central Pollution Control Board (CPCB) (2012), *Annual Report 2011-12*, pp. 94–96.

9. Central Pollution Control Board (CPCB) (2015a), *Status of Ambient Noise Level in India in 2011–2014*, NANMN/02/2015-16.

10. Central Pollution Control Board (CPCB) (2015b), *Status of Ambient Noise Level in India in 2015*, NANMN/03/2015-16.

11. Central Pollution Control Board (CPCB) (2016), *Status of Ambient Noise Level in India in 2016*, NANMN/04/2016.

12. Central Pollution Control Board (CPCB) (2017), *Delhi's Ambient Noise Levels Influenced by Traffic Flow – Case studies*, CUPS/87/2017-18.

13. Central Pollution Control Board (CPCB) (2018), *Status of Ambient Noise Level in India in 2017*, NANMN/05/2017.

14. Central Pollution Control Board (CPCB) (n.d.), https://cpcb.nic.in/.

15. Chowdhury A.K., Debsarkar A., Chakrabarty S. (2016), Assessment of seasonal variations of average traffic pollution levels in curbside open-air microenvironments in Kolkata, India, *Health Scope*, **5**(2): e33081, doi: 10.17795/ jhealthscope-33081.

16. Conference of European Directors of Roads (CEDR) (2013), *Value for Money in Road Traffic Noise Abatement*, CEDR Project Group Road Noise, http://www.carreteros.org/explotacion/cedr/3_CEDR.pdf (access: 11.06.2021).

17. Datta J.K., Sadhu S., Gupta S., Saha R., Mondal N.K., Mukhopadhyay B. (2006), Assessment of noise level in Burdwan town, West Bengal, *Journal of Environmental Biology*, **27**(3): 609–612.

18. European Conference of Ministers of Transport (ECMT) (1998), *Efficient Transport for Europe. Policies of Internalisation of External Costs.*

19. European Environment Agency (2020), *Environmental Noise in Europe – 2020*, No. 22/2019, Publications Office.

20. European Union (2017), *Future Brief: Noise Abatement Approaches*, Publication Office, doi: 10.2779/016648.

21. Garg N. (2019), On suitability of day-night average sound level descriptor in Indian scenario, *Archives of Acoustics*, **44**(2): 385–392, doi: 10.24425/aoa.2019.128502.

22. Garg N. (2022), *Environmental Noise Control: The Indian perspective in an International Context*, Springer Nature Switzerland, doi: 10.1007/978-3-030-87828-3.

23. Garg N. *et al.* (2017a), Study on the establishment of a diversified National Ambient Noise Monitoring Network in seven major cities of India, *Current Science*, **113**(7): 1367–1383, doi: 10.18520/cs/v113/i07/1367-1383.

24. Garg N., Chauhan B.S., Singh M. (2020), Realization and dissemination of unit Watt in airborne sound: Measurement methodology, sound emission regulations and implications, *Sound Emission Regulations and Implications (MAPAN)*, **35**(4): 601–612, doi: 10.1007/s12647-020-00417-x.

25. Garg N., Chauhan B.S., Singh M. (2021), Normative framework of noise mapping in India: Strategies, Implications and Challenges ahead, *Acoustics Australia*, **49**(1): 23–41, doi: 10.1007/s40857-020-00214-1.

26. Garg N., Gandhi V., Gupta N.K. (2022), Impact of COVID-19 lockdown on ambient noise levels in seven metropolitan cities of India, *Applied Acoustics*, **188**: 108582, doi: 10.1016/j.apacoust.2021.108582.

27. Garg N., Maji S. (2016), A retrospective view of noise pollution control policy in India: Status, proposed revisions and control measures, *Current Science*, **111**(1): 29–38, doi: 10.18520/CS/V111/I1/29-38.

28. Garg N., Mangal S.K., Saini P.K., Dhiman P., Maji S. (2015a), Comparison of ANN and analytical models in traffic noise modeling and predictions, *Acoustics Australia*, **43**(2): 179–189, doi: 10.1007/s40857-015-0018-3.

29. Garg N., Sharma O., Mohanan V., Maji S. (2012), Passive noise control measures for traffic noise abatement in Delhi, India, *Journal of Scientific and Industrial Research*, **71**(3): 226–234.

30. Garg N., Sinha A.K., Dahiya M., Gandhi V., Bhardwaj R.M., Akolkar A.B. (2017b), Evaluation and analysis of environmental noise pollution in seven major cities of India, *Archives of Acoustics*, **42**(2): 175–188, doi: 10.1515/aoa-2017-0020.

31. Garg N., Sinha A.K., Gandhi V., Bhardwaj R.M., Akolkar A.B. (2016), A pilot study on the establishment of national ambient noise monitoring network across the major cities of India, *Applied Acoustics*, **103**(Part A): 20–29, doi: 10.1016/j.apacoust.2015.09.010.

32. Garg N., Sinha A.K., Gandhi V., Bhardwaj R.M., Akolkar A.B. (2017c), Impact of Diwali celebrations on environmental noise pollution in India, *Acoustics Australia*, **45**(1): 101–117, doi: 10.1007/s40857-017-0081-z.

33. Garg N., Sinha A.K., Gandhi V., Bhardwaj R.M., Akolkar A.B. (2017d), Effect of odd-even vehicular restrictions on ambient noise levels at ten sites in Delhi city, *Indian Journal of Pure and Applied Physics*, **55**(9): 687–692.

34. Garg N., Vishesh, Maji S. (2015b), Fuzzy TOPSIS approach in selection of optimal noise barrier for traffic noise abatement, *Archives of Acoustics*, **40**(4): 453–467, doi: 10.1515/aoa-2015-0045.

35. Hansell A.L. *et al.* (2013), Aircraft noise and cardiovascular disease near Heathrow airport in London: small area study, *BMJ*, **347**, doi: 10.1136/bmj.f5432.

36. Jamir L., Nongkynrih B., Gupta S.K. (2014), Community noise pollution in urban India: Need for public health action, *Indian Journal of Community Medicine*, **39**(1): 8–12, doi: 10.4103/0970-0218.126342.

37. van Kempen E., Casas M., Pershagen G., Foraster M. (2018), WHO environmental noise guidelines for the European region: A systematic review on environmental noise and cardiovascular and metabolic effects: A summary, *International Journal of Environmental Research and Public Health*, **15**(2): 379, doi: 10.3390/ijerph15020379.

38. Kumar S. *et al.* (2022), Effect of lockdown amid second wave of COVID-19 on environmental noise scenario of the megacity Delhi, India, *The Journal of the Acoustical Society of America*, **152**(3): 1317–1336, doi: 10.1121/10.0013827.

39. Kumar S., Chauhan B.S., Garg N. (2023), Significance and implications of noise mapping for noise pollution control, [in:] *Recent Advances in Metrology*, *Lecture Notes in Electrical Engineering*, Yadav S., Chaudhary K., Gahlot A., Arya Y., Dahiya A., Garg N. [Eds], Vol. 906, pp. 335–341, Springer, Singapore, doi: 10.1007/978-981-19-2468-2_36.

40. Lokhande S.K., Chopkar P.F., Jain M.C., Hirani A. (2021a), Environmental noise assessment of Bhandara City in the adversity of COVID-19 pandemic: A crowdsourcing approach, *Noise Mapping*, **8**(1): 249–259, doi: 10.1515/noise-2021-0020.

41. Lokhande S.K., Garg N., Jain M.C., Rayalu S. (2022), Evaluation and analysis of firecrackers noise: Measurement Uncertainty, legal noise regulations and noise induced hearing loss, *Applied Acoustics*, **186**: 108462, doi: 10.1016/j.apacoust.2021.108462.

42. Lokhande S.K., Gautam R., Dhawale S.A., Jain M.C., Bodhe G.L. (2019), Evaluation of practical framework for industrial noise mapping: A case study, *Noise and Health*, **21**(102): 194–199, doi: 10.4103/nah.NAH_7_17.

43. Lokhande S.K., Sakhareb D.S., Dange S.S., Jain M.C. (2021b), A short review of road noise barriers focusing on ecological approaches, *International Journal of Engineering & Technology Sciences*, **2021**: IJETS-2106092112346.

44. Ministry of Environment & Forests (2000), *The Noise Pollution (Regulation and Control) Rules*, 2000.

45. Mizutani F., Suzuki Y., Sakai H. (2011), Estimation of social costs of transport in Japan, *Urban Studies*, **48**(16): 3537–3559, doi: 10.1177/0042098011399597.

46. Münzel T., Schmidt F.P., Steven S., Herzog J., Daiber A., Sørensen M. (2018), Environmental noise and the cardiovascular system, *Journal of the American College of Cardiology*, **71**(6): 688–697, doi: 10.1016/j.jacc.2017.12.015.

47. Nelson J.P. (1982), Highway noise and property values: A survey of recent evidence, *Journal of Transport Economics and Policy*, **16**(2): 117–138.

48. Schmidt F. *et al.* (2015), Nighttime aircraft noise impairs endothelial function and increases blood pressure in patients with or at high risk for coronary artery disease, *Clinical Research in Cardiology*, **104**(1): 23–30, doi: 10.1007/s00392-014-0751-x.

49. The European Parliament and the Council of the European Union (2002), *Directive 2002/49/EC of the European Parliament and of the Council of 25th June 2002, Relating to the Assessment and Management of Environmental Noise*, Official Journal of the European Communities.

50. Theebe M.A.J. (2004), Planes, trains, and automobiles: The impact of traffic noise on house prices, *The Journal of Real Estate Finance and Economics*, **28**: 209–234, doi: 10.1023/B:REAL.0000011154.92682.4b.

51. Tiwari S., Kumaraswamidhas L.A., Garg N. (2021), Accuracy of short-term noise monitoring strategy in

comparison to long-term noise monitoring strategy, *Indian Journal of Pure & Applied Physics (IJPAP)*, **59**(8): 569–576, doi: 10.56042/ijpap.v59i8.48569.

52. Tiwari S.K., Kumaraswamidhas L.A., Garg N. (2022), Time-series prediction and forecasting of ambient noise levels using deep learning and machine learning techniques, *Noise Control Engineering Journal*, **70**(5): 456–471, doi: 10.3397/1/377039.

53. Torija A.J., Ruiz D.P., Alba-Fernandez V., Ramos-Ridao Á. (2012), Noticed sound events management as a tool for inclusion in the action plans against noise in medium-sized cities, *Landscape and Urban Planning*, **104**(1): 148–156, doi: 10.1016/j.landurbplan.2011.10.008.

54. U.S. Department of Housing and Urban Development (1984), *Environmental Criteria and Standards*, 24 CFR Part 51.

55. World Health Organization (WHO) (2009), *Night noise guidelines for Europe*, World Health Organization, Regional Office for Europe, Denmark.

56. World Health Organization (WHO) (2011), *Burden of disease from environmental noise: Quantification of healthy life years lost in Europe*, World Health Organization, Regional Office for Europe, Denmark.

57. World Health Organization (WHO) (2018), *Environmental noise guidelines for the European region*, World Health Organization, Regional Office for Europe, Denmark.

Archives of Acoustics Vol. 48, No. 1, pp. 127–139 (2023), doi: 10.24425/aoa.2023.144267

Research Paper

Continuous Head-related Transfer Function Representation Based on Hyperspherical Harmonics

Adam SZWAJCOWSKI

Department of Robotics and Mechatronics
AGH University of Science and Technology
Kraków, Poland; e-mail: szwajcowski@agh.edu.pl

(*received June 4, 2022; accepted October 12, 2022*)

Expressing head-related transfer functions (HRTFs) in the spherical harmonic (SH) domain has been thoroughly studied as a method of obtaining continuity over space. However, HRTFs are functions not only of direction but also of frequency. This paper presents an extension of the SH-based method, utilizing hyperspherical harmonics (HSHs) to obtain an HRTF representation that is continuous over both space and frequency. The application of the HSH approximation results in a relatively small set of coefficients which can be decoded into HRTF values at any direction and frequency. The paper discusses results obtained by applying the method to magnitude spectra extracted from exemplary HRTF measurements. The HRTF representations based on SHs and HSHs exhibit similar reproduction accuracy, with the latter one featuring continuity over both space and frequency and requiring much lower number of coefficients. The developed HSH-based continuous functional model can serve multiple purposes, such as interpolation, compression or parametrization for machine-learning applications.

Keywords: hyperspherical harmonics; HRTF; spherical harmonics; continuous functional model; directivity.

1. Introduction

Continuous development of virtual and augmented reality applications rises the need for efficient binaural audio processing algorithms. Especially an important role in authentic recreation of an auditory scene is played by the directivity of human ears, which is commonly called head-related transfer functions (HRTFs). HRTFs are different for each individual as they depend on the shape of torso, head, and pinna. Even though there are some similarities among sets of HRTFs obtained for different people, application of individually measured or matched HRTFs has been proven to improve the localization abilities (Wenzel *et al.*, 1993; Begault *et al.*, 2001). The importance of HRTFs is especially prominent in recognizing the position of a sound source placed in sagittal planes, where simple binaural cues such as interaural level difference (ILD) and interaural time difference (ITD) are mostly invariant (Macpherson, Middlebrooks, 2002; Agterberg *et al.*, 2012).

HRTFs can be physically measured in two ways. Most measurements are performed by putting a pair of microphones inside the subject's ears and recording the response to the sound coming from different directions (e.g., (Andreopoulou, 2015a; Zhang, 2012)). Alternatively, one can make use of the Helmholtz reciprocity principle and place microspeakers inside the ears while the microphones are set around the subject (Zotkin, 2006). Either way, the results of the measurements are sets of HRTFs for a finite number of directions. They are usually stored in the form of discrete head-related impulse responses (HRIRs), for example using the Spatially Oriented Format for Acoustics (SOFA) (AES, 2015; Majdak *et al.*, 2013).

Since HRTFs are relatively large data sets, some attempts were made to develop an efficient model that would reduce the amount of data without a significant loss of accuracy. Initially, the research concerned approximating HRIRs, e.g., by expressing them as filters of either finite or infinite impulse response (e.g., (Kulkarni, Colburn, 1995; 2004)), but soon

the focus was moved to the spatial properties. Evans *et al.* (1998) proposed expressing HRTFs by means of spherical harmonics (SHs). Beside reducing the data size, this representation introduced an even more significant feature – continuity over space. Over the years, the SH-based method has been widely investigated, regarding, i.e., efficiency of different sampling schemes (Zhang *et al.*, 2012), preprocessing techniques (Brinkmann, Weinzierl, 2018) or mixed-order approximations (Ben-Hur *et al.*, 2019). Some alternative continuous representations were suggested, e.g., based on spherical wavelets (Hu *et al.*, 2019; Liu *et al.*, 2019) or Slepian functions (Bates *et al.*, 2015); however, SHs still prevail as the most popular basis functions to approximate not only HRTFs but also other directivity functions such as the sound source directivity (Szwajcowski *et al.*, 2021; Shabtai *et al.*, 2017) or the microphone directivity (Ziegler *et al.*, 2017).

Beside focusing on either only spatial or only time/frequency variations, several attempts were made to develop a model including both these dependencies. Kistler and Wightman (1992) employed the principal component analysis to an HRTF database. While resulting model indeed provided a good accuracy, it was discrete and required a priori knowledge of the database to determine optimal basis vectors. Chen *et al.* (1995) developed a representation based on thin-plate splines that was continuous in space but still discrete in frequency. In 2009, Zhang *et al.* described a model based on Fourier series and Bessel functions which was continuous over both frequency and horizontal angles (Zhang *et al.*, 2009) and which was later extended to cover the entire sphere and include distance dependence as well (Zhang *et al.*, 2010). Another proposition comes from authors who based their model on the fusion of infinite impulse response filters and Legendre polynomials (Shekarchi *et al.*, 2013). However, the authors focused purely on the compression of measured data and thus their model cannot be used to interpolate missing values. Zhang *et al.* (2015) created a functional model utilizing SHs for spatial dependency and complex exponentials for the frequency representation. Both mentioned fully continuous models (Zhang *et al.*, 2010; 2015) determined coefficients for basis functions partially by numerical integrating, which is computationally efficient, but provides worse accuracy than that obtained by fitting in the least-squares sense. Furthermore, these researches focused on retrieving high accuracy complex representations, even though it is known that humans are insensitive to fine spectral or spatial details of HRTFs (Kulkarni, Colburn, 1998; Romigh *et al.*, 2015; Breebaart, Kohlrausch, 2001; Xie, Zhang, 2010).

In this paper, a new approach to the HRTF dimensionality is presented: frequency is imagined as another spatial dimension. This way, the character of these functions becomes purely spatial with two truly spatial variables (horizontal and vertical angles) and one extra variable that represents frequency, but is also treated as a spatial dependence[1]. This approach emphasizes coupling between frequency and space in directivity functions and is inspired by the mathematical structure known as Minkowski space, where time and three-dimensional (3D) space are combined together to conveniently express some of the physical phenomena, most notably regarding the theory of special relativity. Assuming that SHs are a good choice for expressing directivity functions in the 3D space, including the fourth dimension would require an extention of the basis to a four-dimensional (4D) space. A similar problem was tackled in computer graphics, where standard 3D shape descriptors were insufficient for certain cases. Bonvallet *et al.* (2007) extended the popular approach to another dimension by replacing SHs with hyperspherical harmonics (HSHs). This method was later applied in medical imaging by Pasha Hosseinbor *et al.* (20015), providing further a proof that HSHs, previously reserved for theoretical chemistry and nuclear physics, can be successfully employed in engineering. However, to the best of our knowledge, HSHs have not yet been utilized in acoustics or any related field. Thus, the application of this basis to represent acoustical data is the main novel element of this paper.

The primary advantage of the proposed HSH representation over the state-of-the-art SH one is continuity over both space and frequency. Such a representation allows to extract the HRTF magnitude not only at a given direction but also at a given frequency without any additional operations (e.g., interpolation or resampling), thus being computationally attractive. Furthermore, varying the approximation parameters enables easy control of the balance between accuracy and amount of data; acknowledging psychoacoustical aspects in the process of deriving the functional model can lead to a significant reduction of data size by ignoring high-order HSHs responsible for imperceptible spectral details. Last but not least, a more holistic HRTF representation can be of great value in research requiring a thorough directivity parametrization method, e.g., for machine learning applications.

Section 2 provides the necessary theoretical background on 4D coordinate systems and HSHs. Section 3 describes conversion from raw measurement data to hyperspherical data and then to the HSH domain. Section 4 presents exemplary results of the conversion performed on a typical set of HRTFs and suggests further improvements of the process. Section 5 consists of general comments on the HSH representation, its accuracy and potential applications. Finally, Sec. 6 summarizes all the content of this paper.

[1]This paper concerns far-field HRTFs, where the radius dependence is neglected. This is why three- and four-dimensional spaces are described by only two and three variables, respectively.

2. Theoretical background

HRTFs are functions of direction, distance, and frequency. Under the assumption that the distance is greater than 1 m, the radial dependence can be dropped. Such simplified functions are called far-field HRTFs, but since they are measured and applied more commonly than full, distance-dependent HRTFs, they are often called just HRTFs. This is also the case in this paper; wherever HRTFs are mentioned, they mean far-field HRTFs, independent of distance.

2.1. Hyperspherical coordinate system

Assuming that the spatial dependence of HRTFs is to be described in the spherical coordinate system, an extension of this system is needed to capture variability over another dimension. This can be achieved by adding another linear or angular dimension, resulting in either spherindrical or 4D hyperspherical coordinate system (HCS), respectively. The HCS has already proven to be successful in similar research problems in other fields (Bonvallet *et al.*, 2007; Pasha Hosseinbor *et al.*, 2015) and its geometrical properties enhance natural physical properties of acoustic directivity characteristics (see Subsec. 2.2.1 and 3.1.1 for explanation). For these reasons, HCS was chosen over the spherindrical coordinate system for this research.

HCS consists of hyperspherical radius ρ and three angles: $\varphi \in [0, 2\pi)$, $\theta \in [0, \pi)$, and $\psi \in [0, \pi)$. The angles φ and θ correspond to azimuth and inclination defined in the spherical coordinate system, while ψ is an extra angle representing another dimension. Since there are no unified names for angles in 4D space, within this paper they are referred to just by their symbols or colloquially as the spatial angles (meaning angles φ and θ) and the frequency angle (meaning angle ψ) (the details of utilizing an angle to describe the frequency are provided in Subsec. 3.1.1). The relation between HCS and 4D Cartesian system (x, y, z, w) is following:

$$\begin{aligned} x &= \rho \sin\psi \sin\theta \sin\varphi, \\ y &= \rho \sin\psi \sin\theta \cos\varphi, \\ z &= \rho \sin\psi \cos\theta, \\ w &= \rho \cos\psi. \end{aligned} \tag{1}$$

2.2. Hyperspherical harmonics definition

HSHs can be defined for any multidimensional space, but within this paper only 4D HSHs are considered and simply referred to the HSHs. They can be defined as following (Domokos, 1967):

$$Z_{nl}^{m}(\varphi, \theta, \psi) \equiv N(n,l) \sin^{l}\psi \, C_{n-l}^{l+1}(\cos\psi) \, Y_{l}^{m}(\varphi, \theta), \tag{2}$$

where $N(n,l)$ is the normalization factor, $C_{\nu}^{\alpha}(x)$ are the Gegenbauer polynomials, and $Y_{l}^{m}(\varphi, \theta)$ are the SHs, while n, l, and m are the integer parameters limited as following:

$$\begin{aligned} n &\geq 0, \\ 0 \leq l &\leq n, \\ -l \leq m &\leq l. \end{aligned} \tag{3}$$

$N(n,l)$ is the normalization factor, making the HSH basis not only orthogonal but orthonormal. It is given by the formula:

$$N(n,l) \equiv 2^{l+\frac{1}{2}}(l+1)! \sqrt{\frac{2(n+1)(n-l+1)!}{\pi(n+l+2)!}}. \tag{4}$$

It is worth noting, that the SHs in Eq. (2) are also normalized, which is why $N(n,l)$ does not depend on the parameter m – this dependency is entirely included in the SH normalization factor.

$C_{\nu}^{\alpha}(x)$ are the Gegenbauer polynomials, also known as ultraspherical polynomials. They can be defined in multiple equivalent ways; however, for the computations, the most useful one is that given by the following recurrence relation (Ultraspherical polynomials, n.d.):

$$\begin{aligned} C_0^{\alpha}(x) &= 1, \\ C_1^{\alpha}(x) &= 2\alpha x, \\ C_{\nu}^{\alpha}(x) &= \frac{1}{\nu}\big(2x(\nu+\alpha-1)C_{\nu-1}^{\alpha}(x) \\ &\quad - (\nu+2\alpha-2)C_{\nu-2}^{\alpha}(x)\big). \end{aligned} \tag{5}$$

In the HSH definition (Eq. (2)), the Gegenbauer polynomials, together with the factor of $\sin^{l}\psi$, are responsible for variation among ψ angle (the frequency angle).

$Y_{l}^{m}(\varphi, \theta)$ are SHs. SHs can be defined either as complex- or real-valued functions. This paper focuses on their application to express magnitude spectra of HRTFs (see Subsec. 3.1 for explanation) and thus the real form is used, making the entire HSH basis real as well. Real SHs are defined as following[2] (Varshalovich *et al.*, 1998):

$$Y_{l}^{m}(\phi, \theta) \equiv \begin{cases} N_Y(l,m) P_{l}^{m}(\cos\theta) \cos(m\varphi), & \text{if} \quad m \geq 0, \\ N_Y(l,m) P_{l}^{|m|}(\cos\theta) \sin(|m|\varphi), & \text{if} \quad m < 0, \end{cases} \tag{6}$$

where P_{l}^{m} are the associated Legendre functions and $N_Y(l,m)$ is the normalization factor for SHs defined as:

$$N_Y(l,m) \equiv \sqrt{(2-\delta_{m0})\frac{2l+1}{4\pi}\frac{(l-|m|)!}{(l+|m|)!}}, \tag{7}$$

where δ is the Kronecker delta.

[2]Sometimes in real SH definitions there is also $(-1)^m$ factor included, but it is irrelevant for approximation purposes as it merely changes the sign of some coefficients.

2.2.1. Hyperspherical poles

Hyperspheres, by being extensions of 2-spheres (regular 3D spheres), display some analogous features. On a 2-sphere, there are two poles lying at inclinations $\theta \in \{0, \pi\}$, at which all meridians converge to the same point. In other words, at these poles, the actual direction is the same for every azimuthal angle φ. In HSC, such poles exist for every ψ, but there are also two major poles (hyperpoles) at $\psi \in \{0, \pi\}$, where the direction is dependent on neither φ nor θ. Now, applying the interpretation that ψ is the frequency angle, this means that values for the limit frequencies have to be constant for every physical direction. This is inline with how the directivity (including HRTFs) behaves at low frequency – the lower the frequency, the less variance along the directions can be observed, as the size and shape of objects become less and less relevant when compared to the length of acoustic waves corresponding to the analyzed frequency. This coincidence is used in frequency-to-angle mapping, so that the hyperpole convergence can be perceived as an advantage rather than a limitation of the proposed model (see Subsec. 3.1.1).

The presence of hyperpoles can also be seen as a part of general property of HSHs, which are meant to represent the hypersphere boundary evenly. Since the angular distance between two points on the hypersphere gets lower as the hyperpoles are approached (in the same vein as meridians get closer near the poles on a sphere), the ability to capture fine HRTF details decreases. For SHs, the same effect is often achieved by decreasing the maximum order for lower frequencies (e.g., (Zhang *et al.*, 2015; Li *et al.*, 2021)). This is not feasible when continuous representations are used; however, HSHs provide the same effect in a different way, which further shows that they fit well for handling directivity data.

3. Hyperspherical harmonic approximation

In order to express HRTFs in the HSH domain, two steps have to be taken: first, the measurement data have to be converted to HCS and then the HSH coefficients need to be determined so that the weighted sum of HSHs approximates the data in HCS as accurately as possible. Both these steps are described in details in the following subsections.

3.1. HRIRs to data in HCS

As stated in Introduction, HRTFs are commonly stored in the form of HRIRs. It might seem natural to try to find a way to efficiently represent the data in the time domain rather than in the frequency domain. However, there are several arguments in favour of the latter. Firstly, in previous research, both approaches were studied and models based on the frequency representation resulted in lower approximation errors using objective measures (Evans *et al.*, 1998; Hartung *et al.*, 1999). Secondly, the human auditory system, analyzes sounds mostly in the frequency domain and thus accurate reproduction of the impulse response shape is irrelevant to our ears, as opposed to the shape of the frequency spectrum. Furthermore, HRIRs usually have much sharper shapes than the corresponding HRTFs, which makes representing them as a sum of basis functions much harder, especially when the basis is truncated. Last but not least, HRIRs include information on phase, which is of little importance; it is widely acknowledged that phase spectra can be ignored as long as the interaural time difference is preserved (Kistler, Wightman, 1992; Romigh *et al.*, 2015; Kulkarni *et al.*, 1999)[3], although some contrary results have also been presented (Rasumow *et al.*, 2014). Basing a model on the frequency-domain representation thus allows the information on phase to be dropped, reducing the efficient amount of data twice without a significant loss as far as the applicatory aspect is concerned. However, HSHs can be defined as complex functions as well, so it is possible to include the phase information as well, if desired.

The magnitude spectra typically are represented either in linear or in logarithmic scale (in dBs). Referring again to the psychoacoustic aspects of sound perception, the logarithmic scale is relative and thus it is more reflective of how the sound pressure is perceived by the human auditory system. Since the approximation method of the model proposed within this paper is based on the least-squares fitting (see Subsec. 3.2.2 for more details), it seems reasonable to use a logarithmic scale, so that the solver minimizes relative errors (Kulkarni, Colburn, 2004; Romigh *et al.*, 2015; Hartung *et al.*, 1999; Blommer, Wakefield, 1997).

The reference value for the logarithmic scale does not matter as long as it is uniform within a given HRTF set; this value is linked only with the coefficient for Z_{00}^0 (the first HSH invariant along all the angles), but otherwise has no impact whatsoever. Therefore, to make it simple, the reference value was chosen to be 1.

3.1.1. Frequency mapping

As signaled earlier, in order to express HRTFs in HCS, frequency has to be mapped to the ψ angle. The simplest way of doing it would be to set the frequency $f = 0$ at $\psi = 0$ and the frequency $f = f_s/2$ to $\psi = \pi$, where f_s denotes the sampling frequency of HRIRs in the HRTF set. The remaining frequencies would be then mapped to ψ linearly, following the formula:

[3]Phase spectra are still important for low frequencies, but they can be replaced by a linear phase derived from interaural time differences, without any noticeable damage to the localization abilities.

$$\psi_k = \frac{2\pi f_k}{f_s}, \tag{8}$$

where f_k is the center frequency of the k-th frequency bin and ψ_k is the corresponding value of the frequency angle. However, such mapping implies convergence to uniform values at hyperpoles, i.e., at $f \in \{0, f_s/2\}$. When approaching 0 Hz, HRTFs tend to be omnidirectional anyway, but this is not the case for the highest frequencies, especially in the logarithmic scale (see Fig. 1). While this might seem to be of little importance, as the convergence to the hyperspherical pole at the Nyquist frequency would likely affect mostly the highest frequencies which are inaudible anyway, it is possible to avoid such distortions; $f = f_s/2$ can be set to $\psi = \pi/2$ instead of $\psi = \pi$, so that all the HRTFs can be mapped to effectively only half a hypersphere. This requires twice as large resolution of HSH along ψ since the spectra are squeezed to fit on only half of the full range of the frequency angle; however, such mapping allows to ignore all HSHs that are not symmetric about the hyperplane at $\psi = \pi/2$. The application of only ψ-symmetric HSHs (the HSHs that exhibit the aforementioned symmetry about $\psi = \pi/2$) removes the convergence at $f = f_s/2$ and makes the spectra for the entire range of $\psi \in [0, \pi)$ symmetric about the Nyquist frequency in the same manner as typical magnitude spectra obtained by performing the discrete Fourier transform on a real signal. With $f = f_s/2$ set to $\psi = \pi/2$, Eq. (8) takes the form of:

$$\psi_k = \frac{\pi f_k}{f_s}, \tag{9}$$

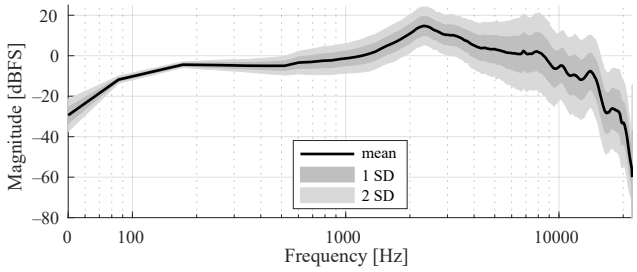


Fig. 1. Statistical distribution of an exemplary set of HRTFs (described in details in Sec. 4) for all directions. Solid black line is the mean spectrum, while the dark and light grey areas cover regions of 1 and 2 standard deviations, respectively. First frequency bin was shifted from 0 to 50 Hz, to be able to fit on the logarithmic axis. At the last frequency bin, standard deviation is especially large due to the occurrence of zero points that were converted to floating-point accuracy in MATLAB to avoid $-\infty$ values. The plot was shifted to fluctuate around 0 dB.

Alternatively, different frequency mappings could be used. Both limit values and the character of scale can be changed, e.g., to a logarithmic frequency scale, which is quite popular in audio engineering, or the mel scale, which is based on a psychoacoustic model of frequency perception. This paper concerns only the linear frequency mapping and leaves the employment of other scales for potential future research.

3.2. Data in HCS to HSH coefficients

Once the 4D data are prepared in the form of a set of triplets of angles (φ, θ, ψ) and the corresponding magnitudes, these data need to be approximated by a finite number of HSHs. The approximation comes down to determining the coefficients $\alpha_{n,l}^m$ for the weighted sum so that approximated function $\widehat{H}(\varphi, \theta, \psi)$ of the form[4]:

$$\widehat{H}(\varphi, \theta, \psi) = \sum_{n=0}^{n_{\max}} \sum_{l=0}^{n} \sum_{m=-l}^{l} \alpha_{nl}^m Z_{nl}^m(\varphi, \theta, \psi) \tag{10}$$

is as close to the original function $H(\varphi, \theta, \psi)$ as possible. In the analytical approach, when an infinite number of basis functions can be used ($n_{\max} = \infty$), consecutive coefficients can be determined by calculating the dot product of the approximated function and complex conjugate of the given HSH:

$$\alpha_{n,l}^m = \int_{\mathbb{S}^3} H(\varphi, \theta, \psi) \overline{Z_{nl}^m}(\varphi, \theta, \psi) \, \mathrm{d}\chi, \tag{11}$$

where $\int_{\mathbb{S}^3} \mathrm{d}\chi$ means integrating over the surface of the unit 3-sphere:

$$\int_{\mathbb{S}^3} \mathrm{d}\chi = \int_0^{2\pi} \int_0^{\pi} \int_0^{\pi} \sin^2\psi \, \mathrm{d}\psi \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi. \tag{12}$$

However, when the number of basis functions is truncated, the accuracy of the approximation decreases. For this reason, least-squares fitting of the coefficient values is preferred over the numerical integration of Eq. (11) to find the optimal set of coefficients.

3.2.1. Determination of the number of HSHs used

The choice of the number of basis functions is one of the most important decisions affecting the quality of the approximation. In general, increasing this number improves the overall accuracy of the approximation. However, using too many basis functions requires more computational resources and can lead to overfitting. It is thus important to determine what is the optimal number of basis functions for a given problem.

HSHs are described by three parameters: n, l, and m. Analyzing the definition given in Eq. (2), one can notice that these parameters are responsible for variations along the angles ψ, θ, and φ, respectively. Limiting just n results in the same spatial frequency

[4]Equation (10) assumes approximation using all the HSHs up to $n = n_{\max}$. However, all of the parameters n, l and m can have independent limits (see Subsec. 3.2.1).

along all the angles. However, some extra limitations for l and m can be imposed to obtain different resolution along angles θ and φ. Equation (10) then takes the form of:

$$\widehat{H}(\varphi,\theta,\psi) = \sum_{n=0}^{n_{\max}} \sum_{l=0}^{\min(n,l_{\max})} \sum_{m=-\min(l,m_{\max})}^{\min(l,m_{\max})} \cdot \alpha_{nl}^{m} Z_{nl}^{m}(\varphi,\theta,\psi), \tag{13}$$

where the limits have to satisfy the following relation[5]:

$$n_{\max} \geq l_{\max} \geq m_{\max}. \tag{14}$$

These maximum values of $n_{\max}$, $l_{\max}$, and $m_{\max}$ should be determined with respect to the sampling theorem and depend on both spatial and frequency sampling. However, HRTFs do not need to be reconstructed with the highest available accuracy, since fine details are psychoacoustically irrelevant. Thus, it is important to find the minimum values of the limits for which the approximation error is negligible.

The spatial dependence is described by SHs and thus the limits for spatial frequency ($l_{\max}$ and $m_{\max}$) can be determined basing on SH-related research. Romigh *et al.* (2015) noticed that some of the localization abilities occur for HRTFs approximated by SHs of the maximum order as low as 2. Approximations using SHs of the maximum order 6 are already statistically indistinguishable from the unprocessed HRTFs and the localization does not improve with further increasing the maximum order. It is important to notice that Romigh *et al.* (2015) used the same preprocessing as proposed within this paper, i.e., the least-squares fitting based on magnitudes in a logarithmic scale. These results suggest that setting $l_{\max}$ and $m_{\max}$ to 6 should be sufficient to prevail the localization properties of a set of HRTFs.

Similar research was conducted regarding sensitivity to spectral details of HRTFs. Kulkarni and Colburn (1998) checked how much smoothing in frequency can be applied to HRTFs without a significant impact on the localization abilities. For this reason, they prepared stimuli by truncating the Fourier series representing the logarithmic spectra of empirical HRTFs. The listeners that took part in the experiment were unable to discriminate between the real and smoothed virtual sound sources in most of the setups for HRTFs reconstructed from only 16 coefficients. For 32 coefficients and above, all of the listeners performed at chance for all the tested directions. Even though the frequency angle in HSHs is not represented by the Fourier series, the same spatial frequency along that angle can be chosen by setting appropriate $n_{\max}$. Since for chosen frequency mapping the spectra are supposed to fit on only half the hypersphere, $n_{\max}$ must be doubled to exhibit a desired resolution over the effectively used range of ψ. Thus, to match the spatial frequency of 32 spectrum coefficients from the described experiment, $n_{\max}$ should be set to 64.

Furthermore, as stated in Subsec. 3.1.1, effectively utilizing only half of the hypersphere allows to ignore non-ψ-symmetric HSHs. Following the HSH and Gegenbauer polynomials definitions given in Eqs. (2) and (5), it can be noted that for Z_{nl}^{m} to be ψ-symmetric, the difference between n and l has to be even. All the parameter configurations where this difference is odd should be thus disregarded when summing over l in Eq. (13).

3.2.2. Computations

Assuming sampling at K 4D directions $\Omega_k \equiv (\varphi_k, \theta_k, \psi_k)$, the HSH coefficients can be determined by solving the following matrix equation:

$$\begin{bmatrix} Z_{00}^{0}(\Omega_1) & \dots & Z_{n_{\max}l_{\max}}^{m_{\max}}(\Omega_1) \\ \vdots & \ddots & \vdots \\ Z_{00}^{0}(\Omega_K) & \dots & Z_{n_{\max}l_{\max}}^{m_{\max}}(\Omega_K) \end{bmatrix} \begin{bmatrix} \alpha_{00}^{0} \\ \vdots \\ \alpha_{n_{\max}l_{\max}}^{m_{\max}} \end{bmatrix} = \begin{bmatrix} H(\Omega_1) \\ \vdots \\ H(\Omega_K) \end{bmatrix}, \tag{15}$$

where $H(\Omega)$ are given in dBs to minimize errors in logarithmic scale.

Since the system is overdetermined, it is usually impossible to find coefficients which would perfectly satisfy this equation. However, it can be solved in the least-squares sense, minimizing the error of approximation. This approach is commonly embraced in research concerning the SH approximation (Zhang *et al.*, 2012; Romigh *et al.*, 2015; Pasqual *e al.*, 2014; Alon *et al.*, 2018). A few existing 4D directivity models featured separate computation of spatial and frequency dependencies, using least-squares fitting for space and then direct integration for frequency (Zhang *et al.*, 2010; 2015). Including both frequency and spatial dependent functions in a single least-squares minimization to the best of my knowledge has not been yet applied for these kind of data, making it another novel element of the paper. While such an approach demands more computational resources, it provides better fitting and acknowledges coupling of space and frequency in directivity functions. It is also worth noting that the increased computational complexity mainly concerns determining the HSH coefficients and is caused by replacing multiple smaller matrix equations by one large. In real-time, with proper optimization, rendering binaural sound should be comparably fast for both SH and HSH HRTF representations.

HRTF sets usually lack data for low elevation angles ($\theta \to \pi$) because of the measurement setup restrictions. This can lead to some irrational values in the

[5]It is possible to obtain a higher spatial frequency, e.g., along θ than along ψ by using only a selection of HSHs for higher values of the corresponding parameters. However, this case is not applicable within this paper and thus it is not further discussed.

unsampled region, since the least-squares solver minimizes an error only at the points where data is available. One way of dealing with this effect is to apply a proper regularization (Zhang *et al.*, 2015; Zotkin *et al.*, 2009). Ahrens *et al.* (2012) proposed another, even more efficient solution based on filling the missing region by means of the low-order SH approximation. However, since in the HSH definition the direction dependence is described by SHs, it is logical to assume that these issues can be handled for the HSH approximation in the same manner. For the comparison between HSHs and SHs, applying any of these methods should not impact the analysis results (both approximations are expected to be impacted in the same way). Thus, simple, non-regularized least-squares fitting was used for both bases.

The computations and analysis were performed in MATLAB using Objective-Oriented Directivity (Szwajcowski, 2021). Both the classes used in this research as well as precomputed objects containing raw and approximated data can be found in its database.

4. Exemplary approximations

To showcase efficiency of the HSH approximation, the method was tested on exemplary data. The chosen set of HRTFs comes from the original measurements of Knowles Electronics Manikin for Acoustic Research (KEMAR) with large pinnae performed at Massachusetts Institute of Technology (Gardner, Martin, 1995) and is a typical HRTF set for evaluating different models (Zhang *et al.*, 2009; 2010; 2015; Shekarchi *et al.*, 2013). The set contains 710 HRIRs for each ear, measured at different directions and each consisting of 512 samples recorded with the sampling frequency of 44.1 kHz. The HRIRs were converted to HCS as described in Subsec. 3.1. Only data for the left ear was used.

Although, according to Subsec. 3.2.1, the parameter limits of $\{l_{\max}, m_{\max}\} = 6$ and $n_{\max} = 64$ should be sufficient, it is important to remember that the SH-focused research used exact frequency spectra, and the spectral-focused one used exact spatial representation. Smoothing in both space and frequency can thus have more impact on the localization abilities. For this reason, the limits used in the exemplary approximation were set to slightly higher values of $\{l_{\max}, m_{\max}\} = 8$ and $n_{\max} = 80$.

For the comparison, the SH approximation was also performed on the same exemplary data. The procedure was exactly the same as for the HSH approximation, but the computations were carried out separately for each frequency bin. Then, approximated values for given directions were put together to retrieve discrete spectra. The maximum order and degree of SHs used for approximation was set to 8, to match the limiting parameters of SHs embedded in the HSHs. Exemplary raw spectra and their SH and HSH approximations are shown in Fig. 2.

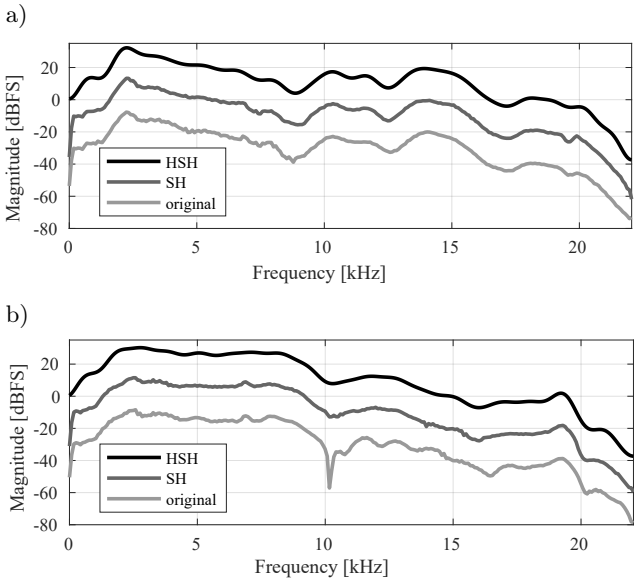


Fig. 2. Raw and approximated spectra of KEMAR HRTFs for two exemplary directions: a) azimuth: 0°, elevation: 0°, far-field; b) azimuth: 30°, elevation: 30°, far-field. The spectra are offset by 20 dB for clarity. See text for more details.

The HSH approximation retrieves the general shape very well, but the resulting spectra are smoother than both the ones computed from the raw data and the ones obtained by means of the SH approximation. The most problematic details to represent in the HSH domain are deep notches, e.g., around 10 kHz in Fig. 2b. However, this is not exclusive to HSH approximation but comes from the spatial dependence represented by SHs, as the SH approximation features the same issues. The notches are caused by rapid changes in phase. Zagala and Zotter (2019) developed a method to deal with such a problem for the low-order SHs. Their algorithm could be extended and applied to HSHs if necessary.

By the properties of ψ-symmetric HSHs, their first derivative with respect to ψ has to be 0 at $\pi/2$, which causes minor discrepancies at the highest frequencies. However, since they lie beyond the human hearing range, these discrepancies are irrelevant from the perceptual point of view. Another problematic frequency band is the very lowest one. In this case, the SH approximation is much more precise than the HSH one; for low frequencies, the directivity takes very simple shapes (mostly omnidirectional), which are easy to express in the SH domain. One of the reasons for the inaccuracy of the HSH approximation in this band are rapid changes of magnitude levels within first few frequency bins, which come from the audio chain used in the HRTF measurement rather than the physical properties of HRTF themselves. This issue is investigated in depth in the following subsection.

4.1. Frequency weighting

Since some frequency bands lie beyond human hearing range, their impact on the approximation results should not be as high as the impact of the audible part of the spectra. This variation in relevance can be achieved by applying weighting to the least-squares minimization. Equation (15) then takes the form of:

$$(\mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{Z})\boldsymbol{\alpha} = \mathbf{Z}^{\mathrm{T}}\mathbf{W}\mathbf{H}, \tag{16}$$

where $\mathbf{Z}$, $\boldsymbol{\alpha}$, and $\mathbf{H}$ denote the respective matrices from Eq. (15), $^{\mathrm{T}}$ denotes matrix transposition, and $\mathbf{W}$ is a diagonal matrix with weights for consecutive angle triplets:

$$\mathbf{W} = \begin{bmatrix} w(\Omega_1) & 0 & \dots & 0 \\ 0 & w(\Omega_2) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & w(\Omega_K) \end{bmatrix}. \tag{17}$$

One of the regions where fitting could be improved by applying proper weights is the frequency band represented by the first few frequency bins. For all the directions, a drop-off can be noticed as the frequency approaches 0, which could be caused by a low-frequency limit of sound sources used in the HRTF measurement or a high-pass filter embedded in the audio chain. For the lowest frequencies, where the corresponding sound waves are much longer than the size of a human head, the HRTFs can be considered omnidirectional (see Fig. 1). This omnidirectionality mirrors the convergence of HCS for the frequency angle $\psi \to 0$. Since the problematic first bins do not hold any important information, they can be completely removed at the stage of computing the HSH approximation (i.e., have weight equal to 0). However, one needs to be wary that, under some circumstances, removing data from certain regions can lead to overfitting.

The other perceptually irrelevant part of the spectra are the highest frequencies. In this case, there are more than one frequency bins lying beyond the human hearing range. Their weights were arbitrarily chosen to start decreasing above 20 kHz and reach 0 at the Nyquist frequency following the shape of cosine function.

The HSH approximation was performed once again, applying the weights as described above. For most parts of the frequency spectrum, no noticeable changes occurred in regard to the previously performed the HSH computation on the complete data. Since there is on average more weight on the audible frequency band, the resulting approximation should be closer to the original data, but the improvements were minuscule in most of the frequency range. However, the fitting for the lowest frequencies has been notably improved, except for the region corresponding to the first bins, which were effectively removed from the least-squares minimization (Fig. 3). Even though the resulting approximation is less accurate with regard to the measured data (large errors at first two bins), it is likely more accurate with regard to the factual physical HRTFs and improves fitting at several following bins.

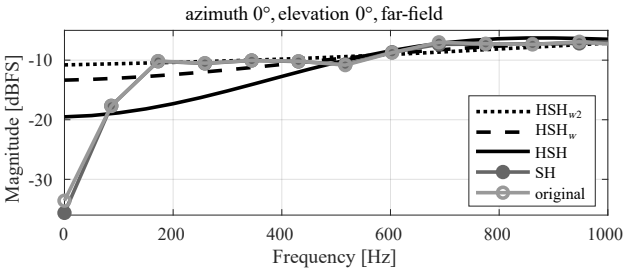


Fig. 3. Raw and approximated spectra of KEMAR HRTF for the direction straight ahead zoomed on the lowest frequencies. HSH$_w$ and HSH$_{w2}$ denote the HSH approximation with weights applied ignoring first or first two frequency bins, respectively (see text for more details).

From the perceptual point of view, the fitting in regions outside of the hearing range is irrelevant, and so the reduced impact of extreme frequency bins should improve the overall accuracy, even if only slightly. Thus, the following sections consider only this new approximation.

4.2. Approximation error analysis

Since 4D data is difficult to display efficiently as a 2D image, it is required to introduce numerical measures to compare accuracy of the SH and HSH approximations over the whole sphere. The key difference between these two approaches is the way of handling the frequency dependence and thus, for the comparison, the approximation error should be presented in the frequency domain. In similar research, various measures of approximation error were used. Currently, one of the most popular approaches is to evaluate it as the root-mean-square (RMS) error of the differences between logarithmic spectra (e.g., (Liu *et al.*, 2019; Romigh *et al.*, 2015; Nishino *et al.*, 1999; Li *et al.*, 2021)). A common alternative is to evaluate the error based on the linear magnitude (e.g., (Hu *et al.*, 2019; Zhang *et al.*, 2015)). Since this article has been written, there is no clear indication which of these approaches surpasses others, including the authors' speculations. However, the measure based on dB is more intuitive to interpret and the computations were performed on the logarithmic data, so it seems appropriate to use the logarithmic error as well. RMS errors for the HSH approximation[6] were computed by averaging

[6]From this section on, the HSH approximation is the one with first two frequency bins removed and weights applied for the high frequencies (earlier denoted as HSH$_{w2}$).

over all 710 directions available in the measurement data using the following formula:

$$\mathrm{RMS}_{\widehat{H}}(f) = \sqrt{\frac{a^*}{\sum_{k=1}^{K} \delta_{\psi_k \psi_f}}}, \tag{18}$$

where

$$a^* = \sum_{k=1}^{K} \left(\widehat{H}(\varphi_k, \theta_k, \psi_k) - H(\varphi_k, \theta_k, \psi_k)\right)^2 \delta_{\psi_k \psi_f}$$

and ψ_f is the frequency angle corresponding to the frequency f, derived from Eq. (9). The RMS error for the SH approximation was computed in an analogous way. In addition, for both methods and for each frequency bin, the 95th percentile (P_{95}) was determined for absolute values of $\widehat{H}(\Omega) - H(\Omega)$. Errors in the form of RMS an P_{95} values for both the SH and HSH approximations are plotted in Fig. 4, while Fig. 5 shows difference between the RMS plots from Fig. 4 as well as 5th and 95th percentile determined on the sets of differences between absolute errors in dB for these two methods:

$$\mathrm{P}_{x\,\mathrm{HSH\text{-}SH}}(f) = \mathrm{P}_x(\{|\widehat{\mathbf{H}}^f_{\mathrm{HSH}} - \mathbf{H}^f| - |\widehat{\mathbf{H}}^f_{\mathrm{SH}} - \mathbf{H}^f|\}), \tag{19}$$

where $\mathbf{H}^f$, $\widehat{\mathbf{H}}^f_{\mathrm{SH}}$, and $\widehat{\mathbf{H}}^f_{\mathrm{HSH}}$ are sub-vectors of the original data, SH and HSH approximations, respectively, which contain only values for a given frequency f.

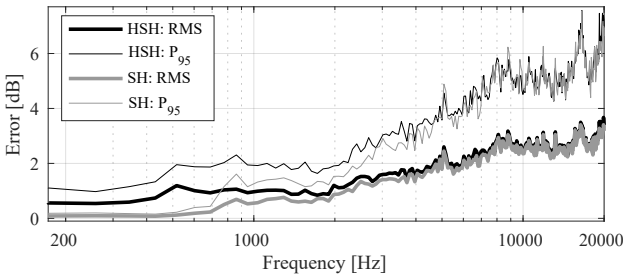


Fig. 4. RMS and P_{95} of the HSH and SH approximation errors in reference to the discrete data.

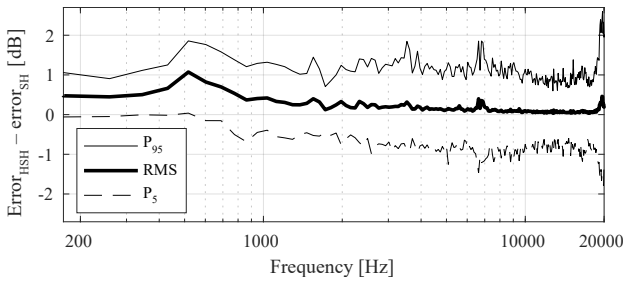


Fig. 5. Difference in RMS errors and 5th, and 95th percentile of differences between the HSH and SH approximation errors.

For the vast majority of the hearing range, the difference in accuracy between the HSH and SH approximations is small, reaching about 1 dB around 500 Hz and becoming less and less significant as the frequency rises. The largest discrepancies occur for the lower frequency regions (below 1 kHz), where the SH approximation provides near perfect reconstruction of the original data, but even there the absolute error is still overall low (around 0.5–1 dB). The values of P_5 indicate that in some cases the HSH approximation yields lower errors than the SH one (especially for higher frequencies). The RMS error, however, is lower for the SH approximation for each frequency bin, even if only to a small extent.

4.2.1. Single-term error

To improve the comparability of the HSH approximation accuracy, the error needs to be provided as a single value averaged over all directions and frequencies. For logarithmic scale, an error measure called spectral distortion (SD) is usually employed. Its definition resembles the RMS definition in Eq. (18) but without the Kronecker deltas, so that the averaging is performed over both space and frequency:

$$\mathrm{SD}_{\widehat{H}} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\widehat{H}(\Omega_k) - H(\Omega_k)\right)^2}. \tag{20}$$

SD values were computed for both SH and HSH approximations in the range 100 Hz to 20 kHz (the lower limit was imposed to exclude the first two frequency bins). The results were $\mathrm{SD}_{\mathrm{SH}} = 2.32$ dB and $\mathrm{SD}_{\mathrm{HSH}} = 2.44$ dB. Since SD requires a ground truth data, it cannot accurately capture the magnitude of differences between two approximations and so the difference between SD values is more representative than the SD of the differences between approximated data (just like the RMS difference is more accurate than the RMS differences in Fig. 5).

5. Discussion

In general, the HSH and SH approximations provide very similar accuracy within the hearing range. However, it is important to note that the minuscule difference in the approximations' errors for middle and audible high frequencies does not mean that there are no errors within these bands. In fact, the errors in both cases are lower for the lowest frequencies, but the SH approximation is very accurate in this region and thus the difference is bigger than for the middle and high frequencies. Such characteristics suggest that the lower accuracy of HSH approximation for higher frequencies comes from inability of low-order SHs to properly express the spatial variability of HRTFs in these bands. However, the maximum order of SHs was chosen based on perceptual studies and so it indicates that these magnitudes of errors are imperceptible. It is notable, that the approximations are not identical even when the RMS error is close to 0; the differences are small

(mostly below 1 dB) and are distributed relatively evenly between being in favor of either SH or HSH approximation. Such small differences in middle and high frequencies have been proven to be perceptually insignificant by many researchers (e.g., (Breebaart, Kohlrausch, 2001; Xie, Zhang, 2010; Huopaniemi *et al.*, 1999)).

At low frequencies, the accuracy of the HSH approximation improved significantly after removing the first two frequency bins from the least-squares minimization. However, the SH approximation still outperforms the HSH one, providing near perfect reconstruction of the original spectrum at low frequencies. The largest differences occur around 500 Hz, which might be specific for the dataset used for evaluation. The way accuracy depends on frequency can be impacted by frequency mapping. Linear mapping treats all frequencies as equally important in the least-squares fitting, while it may be beneficial to employ a frequency scale more aligned with how the human hearing system works, e.g., the logarithmic or mel scale. On the other hand, it is known that spectral localization cues are dependent mostly on the shape of pinna, which impacts only higher frequencies (Algazi *et al.*, 2001; Langendijk, Bronkhorst, 2002). Alternatively, the importance of accuracy in certain frequency bands can be manipulated by the frequency weighing if needed.

To further put the magnitude of errors in perspective, the approximation errors can be put against the measurement errors. Some research has been presented on the impact of different variables on the spectral variations between HRTF measurements (Andreopoulou *et al.*, 2013; 2015); the HSH approximation error is on par with (or lower than) the human-related variations in repeated individual HRTF measurements or the differences between HRTF sets of the same artificial head but obtained in different laboratories. What is more, the difference between SH and HSH approximations is lower than the differences between the measurements performed in the same laboratory within a week (0.12 dB versus 0.17 dB). The frequency smoothing introduced by the HSH approximation, although changing the original values, might also be perceived as a positive effect, since some of the distortion in the raw data might be caused by noise and imperfections of the measurement setup.

Generally, the application of HSHs instead of SHs to approximate a magnitude of HRTFs yields almost the same accuracy while exhibiting some additional advantages. First of all, the resulting representation is fully continuous, not only in space but also in frequency. This continuity enables retrieving discrete HRTFs of any frequency resolution without any additional processing. The most prominent advantage of application of SHs to directivity functions was thus extended to another dimension. An exemplary benefit of using the HRTF model of infinite frequency resolution is the control of balance between precision and latency of binaural rendering by reading discrete HRTFs of any desired resolution.

What is more, even less amount of data is required to store a HRTF set in the HSH domain; in the analyzed example, the HSH approximation was described by 3081 coefficients, while the SH one required 81 for each of 257 frequency bins (255 excluding the first two bins that were ignored in the HSH approximation either way). The total number of SH coefficients was thus 20817 (20655), which is almost seven times more than in the HSH approximation, while retaining similar amount of perceptually relevant information. Even if SH approximation was determined at only 40 frequency bands to match the reduced frequency resolution of HSHs for $n_{\max} = 80$, the total number of SH coefficients would be 3240 which is slightly more than the number of HSH coefficients (3081). It is also worth noting that 40-bin HRTFs would be computationally inefficient and would need to be resampled while no such operation is needed in the HSH representation, where spectra of any resolution can be read without any extra processing. Furthermore, comparing the number of HSH coefficients (3081) to the number of raw data samples (182470), the reduction of data size is almost 60-fold with SD below 2.5 dB, while the best configuration of the HRTF compression method proposed by Shekarchi *et al.* (2013) reached the compression ratio of only 40 for the same accuracy (to get the compression ratio of 60, SD increases to about 2.8 dB). Furthermore, their method is discrete in both space and frequency while the HSH representation is fully continuous. HSHs can be thus useful in designing a compression format that would include all the variability of HRTFs within a single set of coefficients, allowing them to be expressed holistically and compactly.

The downside of the HSH representation is that it requires more computational resources to determine the HSH coefficients via least-squares fitting. However, with the continuous development of processing units, this issue becomes less and less relevant.

The mathematical structure of the HSH representation allows it to be utilized not only for HRTFs but for any type of directivity function, e.g., directivity of electroacoustic devices, whether they are sound sources or receivers. These types of directivity functions are easier to measure and so applying HSHs as approximation tools might be of lesser value, but they can still be useful for such objects in any machine learning problems requiring proper directivity parametrization. HSHs, capturing holistic nature of the far-field directivity, seem to be better fitted for such tasks than SHs.

One of the interesting questions regarding the presented method is the impact of the choice of limiting parameters $n_{\max}$, $l_{\max}$, and $m_{\max}$ on the approximation accuracy. Increasing these limits will improve the

accuracy, but the quality and quantity of the improvement remains to be determined. This is intrinsically linked with the question of what is the minimum required accuracy from the perceptual standpoint and what is the best way to achieve it. However, the presented subject is very broad and requires a thorough analysis to provide reliable answers. For this reason, it is decided to be out of scope of the present paper, but will be considered in future studies.

6. Conclusions

Within this paper, the theory of HCS, real HSHs, and their application to express frequency-dependent directivity data such as HRTFs was presented. The entire computational process was described, starting from discrete HRIRs to determining HSH coefficients. The special focus was put on efficient mapping of frequency scale to an angle, which was required to express the directivity data in HCS. The location of the hyperspherical poles was leveraged to match the omnidirectionality of directivity functions such as HRTFs for low frequencies, making the HSH basis surprisingly well-suited to express such functions, without physical motivation.

The exemplary HSH representation was determined for KEMAR HRTFs using perceptually-motivated number of HSHs. The HSH approximation yields similar levels of accuracy to the corresponding SH one, while providing continuity in frequency and a significant reduction of required amount of data. Thanks to the continuity over both space and frequency, the HSH representation captures the holistic nature of far-field HRTF characteristics. Thus, the HSH representation can be considered an upgrade over the currently popular SH-based approach in practical applications such as binaural rendering, parametrization or data storage. However, there are still many aspects to investigate, such as, e.g., impact of the limiting parameters on the accuracy, exploring different frequency mapping or weighing, perceptual tests, etc.

The method described within this paper not only is a value by itself, but also presents wider possibilities of perceiving directivity functions by modeling frequency as an extra dimension, introducing coupling between their spectral and spatial properties. The HSH model can serve as a base for development of similar representations applying different 4D functions, e.g., created by merging basis functions of lower dimensions. Furthermore, this paper focuses only on the magnitude of far-field HRTFs, which can be in future complemented by the distance and phase dependencies if needed.

Acknowledgments

This research was supported by the National Science Centre, project No. 2020/37/N/ST2/00122. The author would like to thank Anna Snakowska for advice and helpful discussions as well as anonymous reviewers for their insightful suggestions.

References

1. AES (2015), *AES69-2015. AES standard for file exchange – Spatial acoustic data file format*, Audio Engineering Society.
2. Agterberg M.J.H, Snik A.F.M, Hol M.K.S, Van Wanrooij, M.M., Van Opstal A.J. (2012), Contribution of monaural and binaural cues to sound localization in listeners with acquired unilateral conductive hearing loss: Improved directional hearing with a bone-conduction device, *Hearing Research*, **286**(1–2): 9–18, doi: 10.1016/j.heares.2012.02.012.
3. Ahrens J., Thomas M.R.P., Tashev I. (2012), HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data, [in:] *Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (IEEE)*.
4. Algazi V.R., Avendano C., Duda R.O. (2001), Elevation localization and head-related transfer function analysis at low frequencies, *The Journal of the Acoustical Society of America*, **109**(3): 1110–1122, doi: 10.1121/1.1349185.
5. Alon D.L., Ben-Hur Z., Rafaely B., Mehra R. (2018), Sparse head-related transfer function representation with spatial aliasing cancellation, [in:] *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6792–6796, doi: 10.1109/ICASSP.2018.8462101.
6. Andreopoulou A., Begault D.R., Katz B.F. (2015), Inter-Laboratory round robin HRTF measurement comparison, *IEEE Journal on Selected Topics in Signal Processing*, **9**(5): 895–906, doi: 10.1109/JSTSP.2015.2400417.
7. Andreopoulou A., Rogińska A., Mohanraj H. (2013), Analysis of the spectral variations in repeated head-related transfer function measurements, [in:] *International Conference on Auditory Display*, pp. 213–218.
8. Bates A.P., Khalid Z., Kennedy R.A. (2015), On the use of Slepian functions for the reconstruction of the head-related transfer function on the sphere, [in:] *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*.
9. Begault D.R., Wenzel E.M., Anderson M.R. (2001), Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source, *Journal of the Audio Engineering Society*, **49**(10): 904–916.
10. Ben-Hur Z., Alon D.L., Rafaely B., Mehra R. (2019), Loudness stability of binaural sound with spherical harmonic representation of sparse head-related transfer functions, *EURASIP Journal on Audio and Music Processing*, **2019**: 5, doi: 10.1186/s13636-019-0148-x.

11. Blommer M.A., Wakefield G.H. (1997), Pole-zero approximations for head-related transfer functions using a logarithmic error criterion, *IEEE Transactions on Speech and Audio Processing*, **5**(3): 278–287, doi: 10.1109/89.568734.

12. Bonvallet B., Griffin N., Li J. (2007), A 3D shape descriptor: 4D hyperspherical harmonics an exploration into the fourth dimension, [in:] *Proceedings of IASTED International Conference on Graphics and Visualization in Engineering*, pp. 113–116.

13. Breebaart J., Kohlrausch A. (2001), Perceptual (ir)relevance of HRTF magnitude and phase spectra, [in:] *Proceedings of 110th AES Convention*.

14. Brinkmann F., Weinzierl S. (2018), Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition, [in:] *2018 AES International Conference on Audio for Virtual and Augmented Reality*.

15. Chen J., Van Veen B.D., Hecox K.E. (1995), A spatial feature extraction and regularization model for the head-related transfer function, *The Journal of the Acoustical Society of America*, **97**(1): 439–452, doi: 10.1121/1.413110.

16. Domokos G. (1967), Four-dimensional symmetry, *Physical Review*, **159**(5): 1387–1403, doi: 10.1103/Phys Rev.159.1387.

17. Evans M.J., Angus J.A.S., Tew A.I. (1998), Analyzing head-related transfer function measurements using surface spherical harmonics, *The Journal of the Acoustical Society of Americ*a, **104**(4): 2400–2411, doi: 10.1121/1.423749.

18. Gardner W.G., Martin K.D. (1995), HRTF measurements of a KEMAR, *The Journal of the Acoustical Society of America*, **97**(6): 3907–3908, doi: 10.1121/1.412407.

19. Hartung K., Braasch J., Sterbing S.J. (1999), Comparison of different methods for the interpolation of head-related transfer functions, [in:] *16th International Conference: Spatial Sound Reproduction*, pp. 319–3129.

20. Hu S., Trevino J., Salvador C., Sakamoto S., Suzuki Y. (2019), Modeling head-related transfer functions with spherical wavelets, *Applied Acoustics*, **146**: 81–88, doi: 10.1016/j.apacoust.2018.10.026.

21. Huopaniemi J., Zacharov N., Karjalainen M. (1999), Objective and subjective evaluation of head-related transfer function filter design, *Journal of the Audio Engineering Society*, **47**(4): 218–239.

22. Kistler D.J., Wightman F.L. (1992), A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction, *The Journal of the Acoustical Society of America*, **91**(3): 1637–1647, doi: 10.1121/1.402444.

23. Kulkarni A., Colburn H.S. (1995), Efficient finite-impulse-response filter models of the head-related transfer function, *The Journal of the Acoustical Society of America*, **97**(5): 3278–3278, doi: 10.1121/1.411579.

24. Kulkarni A., Colburn H.S. (1998), Role of spectral detail in sound-source localization, *Nature*, **396**(6713): 747–749, doi: 10.1038/25526.

25. Kulkarni A., Colburn H.S. (2004), Infinite-impulse-response models of the head-related transfer function, *The Journal of the Acoustical Society of America*, **115**(4): 1714–1728, doi: 10.1121/1.1650332.

26. Kulkarni A., Isabelle S.K., Colburn H.S. (1999), Sensitivity of human subjects to head-related transfer-function phase spectra, *The Journal of the Acoustical Society of America*, **105**(5): 2821–2840, doi: 10.1121/1.426898.

27. Langendijk E.H.A., Bronkhorst A.W. (2002), Contribution of spectral cues to human sound localization, *The Journal of the Acoustical Society of America*, **112**(4): 1583–1596, doi: 10.1121/1.1501901.

28. Li J., Wu B., Yao D., Yan Y. (2021), A mixed-order modeling approach for head-related transfer function in the spherical harmonic domain, *Applied Acoustics*, **176**: 107828, doi: 10.1016/j.apacoust.2020.107828.

29. Liu H., Fang Y., Huang Q. (2019), Efficient representation of head-related transfer functions with combination of spherical harmonics and spherical wavelets, *IEEE Access*, **7**: 78214–78222, doi: 10.1109/ACCESS.2019.2921388.

30. Macpherson E.A., Middlebrooks J.C. (2002), Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited, *The Journal of the Acoustical Society of America*, **111**(5): 2219–2236, doi: 10.1121/1.1471898.

31. Majdak P. *et al.* (2013), Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions, [in:] *Proceedings of 134th AES Convention*.

32. Nishino T., Kajita S., Takeda K., Itakura F. (1999), Interpolating head related transfer functions in the median plane, [in:] *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 167–170.

33. Pasha Hosseinbor A. *et al.* (2015), 4D hyperspherical harmonic (HyperSPHARM) representation of surface anatomy: A holistic treatment of multiple disconnected anatomical structures, *Medical Image Analysis*, **22**(1): 89–101, doi: 10.1016/j.media.2015.02.004.

34. Pasqual A.M. (2014), Spherical harmonic analysis of the sound radiation from omnidirectional loudspeaker arrays, *Journal of Sound and Vibration*, **333**(20): 4930–4941, doi: 10.1016/j.jsv.2014.05.006.

35. Rasumow E. et al. (2014), Smoothing individual head-related transfer functions in the frequency and spatial domains, *The Journal of the Acoustical Society of Americ*a, **135**(4): 2012–2025, doi: 10.1121/1.4867372.

36. Romigh G.D., Brungart D.S., Stern R.M., Simpson B.D. (2015), Efficient real spherical harmonic representation of head-related transfer functions, [in:] *IEEE Journal on Selected Topics in Signal Processing*, **9**(5): 921–930, doi: 10.1109/JSTSP.2015.2421876.

37. Shabtai N.R., Behler G., Vorländer M., Weinzierl S. (2017), Generation and analysis of an acoustic radiation pattern database for forty-one musical instruments, *The Journal of the Acoustical Society of America*, **141**(2): 1246–1256, doi: 10.1121/ 1.4976071.

38. Shekarchi S., Hallam J., Christensen-Dalsgaard J. (2013), Compression of head-related transfer function using autoregressive-moving-average models and Legendre polynomials, *The Journal of the Acoustical Society of America*, **134**(5): 3686–3696, doi: 10.1121/1.4822477.

39. *Ultraspherical polynomials* (n.d.), Encyclopedia of Mathematics, http://encyclopediaofmath.org/index.php?title=Ultraspherical_polynomials&oldid=52128.

40. Szwajcowski A. (2021), *Objective-Oriented Directivity. MATLAB toolbox for processing directivity models* (access: 3.06.2022).

41. Szwajcowski A., Krause D., Snakowska A. (2021), Error analysis of sound source directivity interpolation based on spherical harmonics, *Archives of Acoustics*, **46**(1): 95–104, doi: 10.24425/aoa.2021.136564.

42. Varshalovich D.A., Moskalev A.N., Khersonskii V.K. (1988), *Quantum Theory of Angular Momentum*, World Scientific.

43. Wenzel E.M., Arruda M., Kistler D.J., Wightman F.L. (1993), Localization using nonindividualized head-related transfer functions, *The Journal of the Acoustical Society of America*, **94**(1): 111–123, doi: 10.1121/1.407089.

44. Xie B., Zhang T. (2010), The audibility of spectral detail of head-related transfer functions at high frequency, *Acta Acustica United with Acustica*, **96**(2): 328–339.

45. Zagala F., Zotter F. (2019), Idea for sign-change retrieval in magnitude directivity patterns, [in:] *Conference: Fortschritte der Akustik – DAGA*.

46. Zhang M., Kennedy R.A., Abhayapala T.D. (2015), Empirical determination of frequency representation in spherical harmonics-based HRTF functional modeling, [in:] *IEEE/ACM Transactions on Audio Speech and Language Processing*, **23**(2): 351–360, doi: 10.1109/TASLP.2014.2381881.

47. Zhang W., Abhayapala T.D., Kennedy R.A., Duraiswami R. (2010), Insights into head-related transfer function: Spatial dimensionality and continuous representation, *The Journal of the Acoustical Society of America*, **127**(4): 2347–2357, doi: 10.1121/1.3336399.

48. Zhang W., Kennedy R.A., Abhayapala T.D. (2009), Efficient continuous HRTF model using data independent basis functions: Experimentally guided approach, *IEEE Transactions on Audio, Speech and Language Processing*, **17**(4): 819–829, doi: 10.1109/TASL.2009.2014265.

49. Zhang W., Zhang M., Kennedy R.A., Abhayapala T.D. (2012), On high-resolution head-related transfer function measurements: An efficient sampling scheme, *IEEE Transactions on Audio, Speech and Language Processing*, **20**(2), 575–584, doi: 10.1109/TASL.2011.2162404.

50. Ziegler J.D., Rau M., Schilling A., Koch A. (2017), Interpolation and display of microphone directivity measurements using higher order spherical harmonics, [in:] *Proceedings of 143rd AES Convention*.

51. Zotkin D.N., Duraiswami R., Grassi E., Gumerov N.A. (2006), Fast head-related transfer function measurement via reciprocity, *The Journal of the Acoustical Society of America*, **120**(4): 2202–2215, doi: 10.1121/1.2207578.

52. Zotkin D.N., Duraiswami R., Gumerov N.A. (2009), Regularized HRTF fitting using spherical harmonics, [in:] *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 257–260.

Archives of Acoustics Vol. 48, No. 1, pp. 141–148 (2023), doi: 10.24425/aoa.2023.144268

Research Paper

Non-invasive Ultrasound Doppler Effect Based Method of Liquid Flow Velocity Estimation in Pipe

Pawel BIERNACKI*, Stanislaw GMYREK, Wladyslaw MAGIERA

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

*Corresponding Author e-mail: pawel.biernacki@pwr.edu.pl

(*received June 8, 2022; accepted November 7, 2022*)

This paper discusses the estimation of flow velocity from a multi-sensor scenario. Different estimation methods were used, which allow the effective measurement of the actual Doppler shift in a noisy environment, such as water with air bubbles, and on this basis the estimation of the flow velocity in the pipe was calculated. Information fusion is proposed for the estimates collected. The proposed approach focuses on the density of the fluid. The proposed method is capable of determining the flow velocity with high accuracy and small variations. Simulation results for plastic and steel (both galvanized and non-galvanized) pipes show the possibility of accurate fluid flow measurements without the need for sensors inside the pipe.

Keywords: information fusion; flow velocity; Doppler effect.

1. Introduction

Calm-on flowmeters are becoming increasingly popular. Their main advantage is fast and non-invasive installation. They can be used as portable devices in applications such as channel infiltration capacity testing, delivery control, consumption verification, measurement campaigns. Direct and indirect approaches can be distinguished (Raffel *et al.*, 1998). Direct methods typically involve probes that are placed directly in the fluid drift. On the other hand, indirect methods do not disturb fluid flow and consist of particle image velocimetry (PIV) (Atkins, 2016; Kaipio *et al.*, 2017) and laser Doppler anemometry (LDA) (Solero, Beghi, 1995; Doran, 2013).

Methods based on digital signal processing are becoming more and more popular. The cross-correlation method is to look for the time delay of the flow structure passing from upstream sensors to downstream sensors (Beck, Plaskowski, 1987; Lucas *et al.*, 1999). The pixel-based cross-correlation solution can reconstruct the radial velocity distribution (Cui *et al.*, 2016; Xu *et al.*, 2009). The ability of acoustic transducers to characterize underwater dynamic phenomena may be used in non-invasive measurement scenarios (Cochran, 2001; Jones, 1995; Buermans *et al.*, 2009). In medical echography, the Doppler signal analysis is one of the most vital diagnostic techniques (Matani *et al.*, 1996).

There are four basic measurement methods based on:

1) estimation of the time of a sine wave propagation in the tube (transit-time flow) (Takeda, 2012),
2) using the Doppler effect for the emitted sinusoidal wave (pulsed Doppler flow) (Mori *et al.*, 2004; Wu, 2018),
3) measuring of ultrasonic wave lift (Kang *et al.*, 2019),
4) correlation method (Avilán *et al.*, 2013).

The operation of the transit flow involves sending and receiving ultrasound pulses through a pair of probes and measuring the difference in signal transit time. The probes used are mounted outside the pipe, generating pulses that pass through the wall of the pipe. The liquid flowing in the pipeline creates a difference in the signal beam transit time. This time is

measured by a flowmeter, and the flow rate is then calculated. The key principle of the method is that sound waves traveling in the direction of the liquid have a higher speed than those traveling in the opposite direction. The difference in the signal transit time is proportional to the liquid flow rate. It is shown in Fig. 1 and can be written as:

$$V_{\text{liquid}} = \frac{\Delta t c^2_{\text{liquid}}}{2L\cos\theta} = \frac{\Delta t c^2_{\text{liquid}}}{2L\sin\alpha}. \tag{1}$$

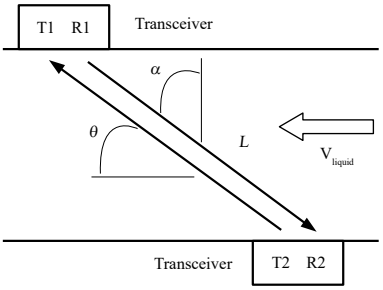


Fig. 1. Time transit measurement method of liquid flow velocity.

The time-of-flight flow measurement technology can provide reliable performance over a wide range of fluid flow conditions, but there are cases where a high percentage of undissolved gases or solids can scatter the acoustic beam and prevent the appropriate signal amplitude from reaching the receiving transducer (Takeda, 1995). Under such conditions, a Doppler measurement may be required to meet the customer's flow measurement needs.

The method of ultrasonic wave lift is based on shifting the point of incidence of the wave along the pipeline, in proportion to the average speed of the fluid), with:

$$V_{\text{liquid}} \propto (A_1 - A_2), \tag{2}$$

where A_1 and A_2 are the amplitudes of the ultrasonic waves received by the receivers R_1 and R_2. As the velocity of the fluid increases, the signal in the receiver R_1 decreases and in the receiver R_2 increases, and the difference in signal amplitudes carries information about the fluid velocity. This approach is shown in Fig. 2.

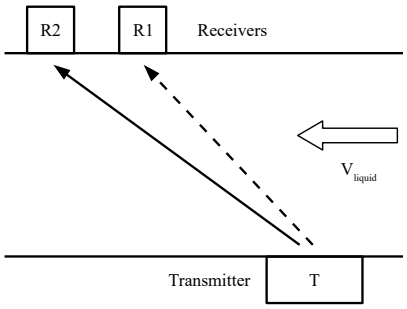


Fig. 2. Wave lift measurement method of liquid flow velocity.

The correlation method (Fig. 3) of measurement based on the time shift of the receiving signals for which there is a maximum of cross-correlation. Such a sensor can also be used for vortex frequency and thermal disturbance detections.

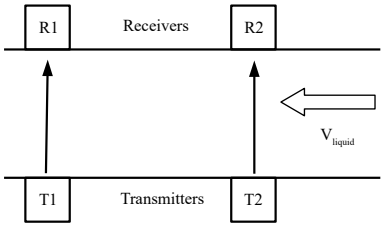


Fig. 3. Correlation measurement method of the flow velocity of liquid.

The pulse Doppler method appears to be the latest and most accurate way to measure the flow of liquids containing large amounts of undissolved gases or suspended solids (Lucas *et al.*, 1999). This method uses the Doppler phenomenon to measure the frequency difference of a continuous signal transmitted and received in a liquid with reflective elements (e.g., air bubbles). Its method of operation is described in the next section. It is possible to develop the pulsed Doppler method that automatically switches from time-of-flight to the Doppler measurement without changing the transducer position, suggesting the use of the transceiver method. In the minimum hardware version, this solution requires only one transceiver that acts as a transmitter and a receiver interchangeably. It is shown in Fig. 4. Ultrasonic liquid flow velocity methods have the following advantages:

- No moving parts: traditional mechanical flowmeters measure pressure through the use of moving parts that serve as mechanical sensors. Because there are no moving parts on ultrasonic flowmeters, one does not have to worry about them degrading or creating a blockage.
- Low maintenance: because ultrasonic flowmeters do not involve moving parts, they last a long time and need very little maintenance. They also have low power consumption, so they often last several years before the batteries are to be replaced.
- High accuracy: as long as the meter is properly mounted and installed, these meters are highly accurate. However, inline and insertion flowmeters

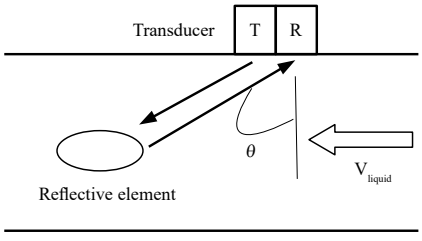


Fig. 4. Doppler measurement method of liquid flow velocity.

are generally more accurate than clamp-on ultrasonic flowmeters.
- Bi-directional measurements: although many traditional flowmeters only measure in a single direction, transit-time ultrasonic flowmeters measure flow in both directions, forward, and reverse.
- High stability: transit-time meters are unaffected by the temperature, density, or concentration of the liquids they measure, making them a more stable measuring device. Ability to measure liquids and gases. Ultrasonic flowmeters can be used to measure a wide variety of liquids.

The disadvantages include:
- Sensitivity to temperature changes: Doppler-type ultrasonic transducers are sensitive to changes in temperature, density, and concentration, meaning that any changes to the contents of the pipe may negatively affect the accuracy of the Doppler transducer results.
- Substance limitations: working with slurries where the flow is not linear may produce measurement errors.

The organization of the paper is as follows: in Secs. 2 and 3 we discuss the methodology for estimating the liquid flow velocity. Sections 4, 5, and 6 include mathematical tools used for measuring system. Section 7 describes some simulations performed in the real-life working system. The paper ends with some conclusions and suggestions for the next steps in increasing the accuracy of the measurement.

2. Metodology of measurements

In single-sensor detection, the signal is subjected to individual preprocessing and then different Doppler shift estimation methods are used. The environment is characterized by a high level of noise (especially in steel pipes), which leads to false results that are indistinguishable from the true values. When using multiple sensors, the actual fluid velocity is obtained by comparing the results from individual sensors using different methods. The decision is obtained from a joint density of the results taking into account the results from all sensors. A joint density for as few as two sensors gives significantly better speed detection than using only one sensor. We compare the use of individual parameterisation methods to the case of multisensor data fusion at the decision level. The results acquired from the individual sensors are used to estimate the Doppler shift; we estimate the probability density function of the fluid flow velocity.

3. Doppler shift and velocity flow equation

We have the continuous-wave Doppler system. In this case, two transducers are used: the first transducer transmits an acoustic signal into the fluid, while the second transducer is used to receive the reflected signal. Reflections come from the scattering particles within the fluid. The Doppler shift frequency depends on the fluid velocity. This affects the reflected signal, which is an expanded (or compressed) version of the transmitted signal. The difference between them gives us the Doppler shift frequency. In the standard approach, the Doppler flow equation also takes into account the fluid sound speed as well as the beam angle in the fluid. In our case, we do not need these two parameters. It would be redundant because using the refractive clamp-on transducer implies that the sine of the beam angle and the sound speed are constant.

We have the Doppler shift frequency:

$$\Delta f = 2f\sin(\theta)\frac{v}{c}. \tag{3}$$

From

$$Tr_{\text{phase}} = \frac{c}{\sin(\theta)} \tag{4}$$

we have

$$\Delta f = 2f\frac{v}{Tr_{\text{phase}}}. \tag{5}$$

It results in a form:

$$v = Tr_{\text{phase}}\frac{\Delta f}{2f}, \tag{6}$$

where V_{phase} is the transducer phase parameter that depends on the angle of the transmitted wave, f is the transmit frequency, Δf is the Doppler shift frequency, v is the flow velocity, c is the sound velocity in the liquid.

4. Extended Kalman filter

To adapt the Kalman filter to nonlinear optimal filtering problems, the Extended Kalman Filter (EKF) (Jazwinski, 1970; Maybeck, 1982; Bar-Shalom *et al.*, 2004; Grewal, Andrews, 2001; Särkkä, 2006) is used. It requires determining a Gaussian approximation to the joint distribution of state x and measurements y (with Taylor series-based transformation involved):

$$x_k = f(x_{k-1}, k-1) + q_{k-1}, \tag{7}$$

$$y_k = h(x_k, k) + r_k, \tag{8}$$

where x_k is the state, y_k is the measurement, q_{k-1} is the process noise, r_k is the measurement noise, f is the dynamic model function, h is the measurement model function. The state vector (in the sine wave case) may be written as:

$$x_k = (\theta_k, \omega_k, a_k), \tag{9}$$

where θ_k is the parameter for the sine function on the time step k, $\frac{\mathrm{d}\theta}{\mathrm{d}t} = \omega$, ω_k is the angular velocities in time

step k, a_k is the amplitude in the time step k. The dynamic equation in the discretized form is:

$$x_k = \begin{pmatrix} 1 & \Delta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{10}$$

The measurement function $h(x_k, k)$, given by

$$h(x_k, k) = a_k \sin(\theta_k), \tag{11}$$

is a sine function. The measurement model as follows:

$$y_k = h(x_k, k) + r_k, \tag{12}$$

where r_k is the white Gaussian noise with zero mean and variance 1.

5. Power spectral density estimation

Power spectral density (PSD) estimation techniques can be divided into parametric and non-parametric methods. Non-parametric methods estimate PSD explicitly from signal samples, without making any assumptions about the particular process structure. Parametric approaches assume that signal can be described as the stationary process (MA, AR, or ARMA) of order p. The power spetral density is then calculated using estimated model parameters. This paper presents PSD estimated with parametric approaches (Burg and Prony's method) and non-parametric methods (Welch's and Thomson multitaper method). On the basis of the signal spectrum, it is possible to determine the dominant frequency (weighted average), which estimates the Doppler shift, Δf.

5.1. The Burg algorithm

The Burg algorithm assumes that the signal can be described as the AR process of order p:

$$\widehat{x} = -\sum_{k=1}^{m} a_m(k) x(n-k). \tag{13}$$

There are many techniques for estimating the a_m parameters such as the Yule-Walker algorithm, or least squares estimator (Kaipio *et al.*, 2015). The Burg algorithm solves the ordinary least-squares problem. The AR parameters a_m are estimated by minimizing the prediction forward and backward errors, which are referred to as the error between the actual value signal and the corresponding estimators forward and backward (Atkins, 2016):

$$\mathrm{PSD}x(f) = \frac{E_m}{\left|1 + \sum_{k=1}^{m} a(k) e^{-j2\pi fk}\right|^2}. \tag{14}$$

The results obtained by the Burg algorithm have a high frequency resolution (Atkins, 2016), and are more objective and stable than the other algorithms for estimating the power spectral density using the AR model.

5.2. The Prony algorithm

Prony proposed that the N data samples can be approximated using the sum of complex, damped exponentials:

$$\widehat{x}(k) = \sum_{i=0}^{p} a_i \cdot z_i^k, \qquad k = 0, ..., N-1, \tag{15}$$

with $a_i = A_i \cdot e^{j\theta_i}$, $z_i = e^{\alpha_i + 2\pi j f T}$, where T is the period, A_i, α_i, f_i, θ_i are the amplitude, damping factor, frequency and initial phase of the complex, dumped exponentials. The fitness problem leads to a minimization error between the data $x(n)$ and the fitted value $\widehat{x}(n)$. The complicated nonlinear problem, with the Prony method, can be converted to the linear prediction problem, and $x(n)$ can be regraded as the output of the p-th order of AR process. The PSD is given by:

$$\mathrm{PSD}x(f) = \left| \sum_{i=1}^{p} A_i \cdot e^{j\theta_i} \frac{2\alpha_i}{(\alpha_i^2 + (2\pi(f - fi))^2)} \right|^2. \tag{16}$$

The Prony method provides greater accuracy and does not have a problem with spectral leakage.

5.3. The Welch algorithm

To estimate PSD with the Welch method, the signal should be divided into overlapping segments and multiplied by a window function. Then, for each part of the signal, the modified periodogram is computed. The power spectral density is estimated by averaging the periodograms. The estimate is given by:

$$\mathrm{PSD}x(f) = \frac{1}{M} \sum_{m=1}^{M} \left\{ \frac{2}{N_f} \left| \sum_{n=0}^{N-1} x(n+mD) e^{-j2\pi nk} \right|^2 \right\}, \tag{17}$$

where M denotes the number of signal fragments of length N_f and D denotes delay. The Welch method reduces the variance of the classic periodogram (Lyons, 2004).

5.4. Multitaper PSD estimate

Power spectral density estimate computed using the multitaper method utilizes mutually orthogonal windows – discrete prolate spheroidal (Slepian) sequences:

$$\mathrm{PSD}x^{MT}(f) = \frac{1}{L} \sum_{l=0}^{L-1} \mathrm{PSD}x_l(f), \tag{18}$$

where

$$\mathrm{PSD}x_l(f) = \Delta t \left| \sum_{n=0}^{N-1} s_l(n) x(n) e^{-j2\pi f_n \Delta t} \right| \tag{19}$$

can be considered as the modified periodogram computed with the l-th Slepian sequence $s_l(n)$. The multitaper PSD estimate averages the L periodograms.

The multitaper method (MTM) is similar to the Welch estimators, but in this approach, the periodograms are decorelated due to the orthogonality of Slepian sequences. The Welch approach computes the modified periodograms using the overlapping segments of the signal, whereas the MTM method uses the entire signal to compute the modified periodogram. The examples of the first five Slepian sequences are presented in Fig. 5.

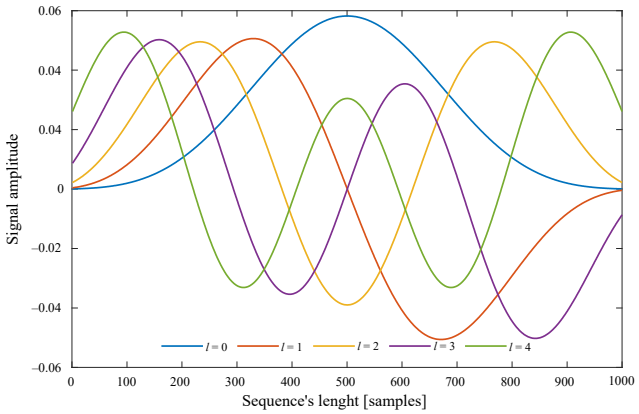


Fig. 5. Examples of five Slepian sequences used in MTM method.

6. Kernel density estimation

In order to estimate a probability density function $\widehat{f}(x)$ we need to use a statistical method using a set of samples $x_1, ..., x_n$. For this we use kernel density estimation (KDE). In each step, the i-th sample x_i is assigned to a kernel function $K(x,t)$:

$$\widehat{f} = \frac{1}{n}\sum_{i=1}^{n} K(x_i, t). \tag{20}$$

In the special case we have a form:

$$K_{\text{sym}}(x,t) = \frac{1}{h}K\left(\frac{x-t}{h}\right), \tag{21}$$

which is valid when our kernel is symmetric. In Eq. (21) h represents the smoothing parameter or bandwidth. Control the smoothing factor for each sample. It is very important to choose the right value for h. Taking a wrong value, too small or too large, will affect the estimator. There are insignificant details shown when h is too small. On the other hand, the estimated probability density function will be too smooth and the information from the sample may be lost.

In our proposal, we use the bivariate extension:

$$\widehat{f}(x,y) = \frac{1}{nh_xh_y}\sum_{i=1}^{n} K\left(\frac{x_i-x}{h_x}, \frac{y_i-y}{h_y}\right), \tag{22}$$

where x_i, y_i, for $i = 1, 2, ..., n$ are the samples, h_x, h_y stand for smoothing coefficients. It can be easily deduced from the univariate case.

There are many different multivariate kernels which can be found in applications, e.g., the Epanechnikov kernel:

$$K(u) = \frac{3}{4}(1-u^2), \tag{23}$$

or the Gaussian kernel:

$$K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}. \tag{24}$$

From Eq. (22) following estimators are available – the product kernel estimator and the radial kernel estimator.

7. Estimation of Doppler frequency signal scheme

The scheme of the measurement system is shown in Fig. 6.

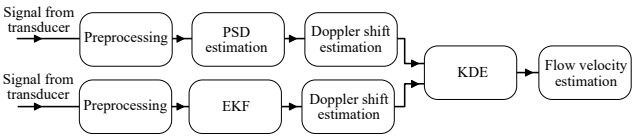


Fig. 6. Signal processing scheme.

The signals from the transducers are first preprocessed for noise cancelation. The PSD estimation is then calculated. When the results obtained from both transducers (oriented in opposite directions) are obtained, the Doppler shift can be determined. Using the KDE approach allows to estimate the fluid flow velocity.

8. Results

The schematic of the measurement system is visible in Figs. 7 and 8. Siemens 191N1S transducers were used with a sampling frequency of 10 MHz and with separate transmit and receive transducers. Such a transducer selection provides the high amount of reflected sound energy with the least synchronous noise from the pipe wall or transducer cross coupling, so this was considered to be the easiest case to prove the principle without any major barriers. The transmitting sine-wave signal was stored in a flash memory and sent to the pipe as a burst every 10 ms. The received signal was amplified by a fixed factor, sampled (16 bits) and sent to the signal processing unit

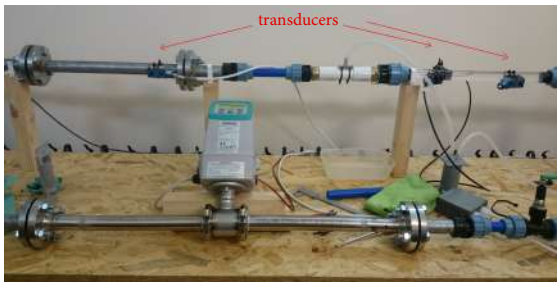


Fig. 7. Measurement equipment.

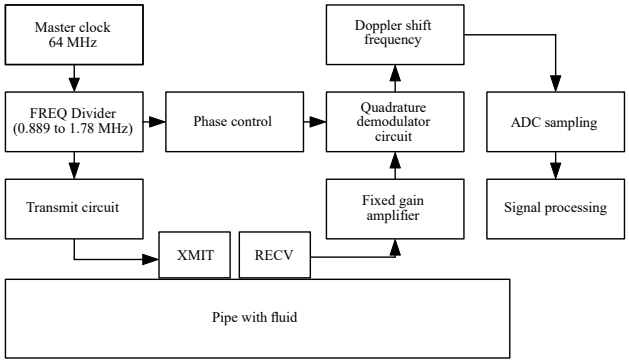


Fig. 8. Measurement system details.

(STM32 processor). Test studies were carried out on an actual system. The test environment consisted of a set of pipes made of PVC, galvanized, and ungalvanized steel. Examples of recorded signals are presented in Fig. 9.

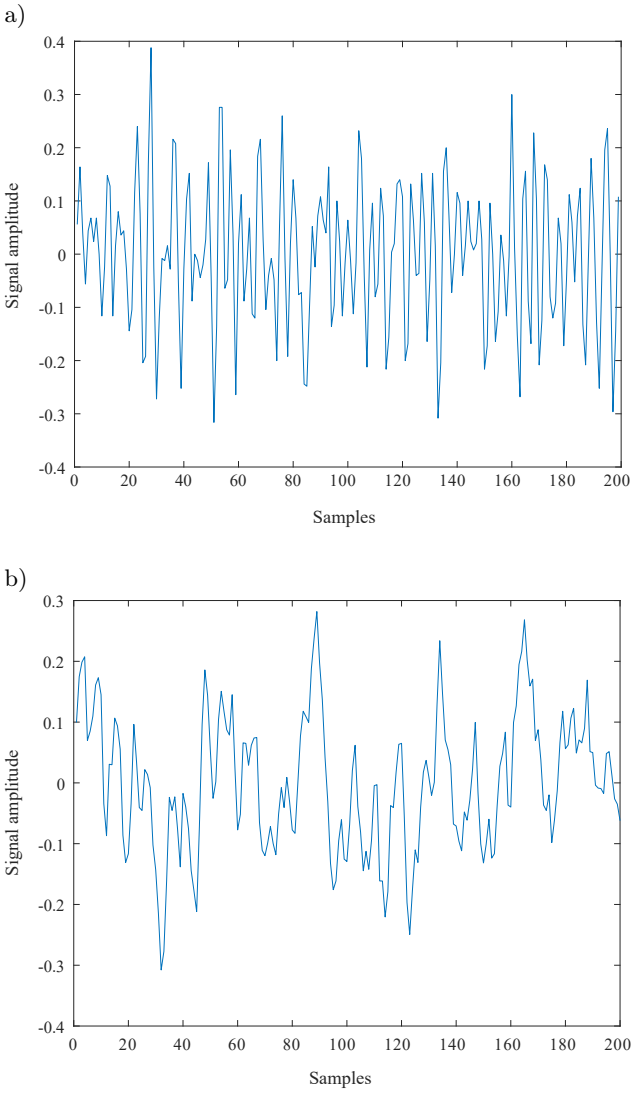


Fig. 9. Examples of the recorded signals:
a) PCV pipe; b) steel galvanized pipe.

Pipes with diameters ranging from 50 mm to 150 mm with closed circulation were used. Measurements were made using air bubbles with a diameter of 1–2 mm. This made it possible to use the pulsed Doppler method. The pump used a guaranteed constant and the same speed of air bubbles and water inside the pipe. Transducers acting as sensors provided measurements for our tests. These were fixed on the horizontal plane. Two transducers measure the Doppler shift – the first measures the negative shift and the second measures the positive shift. It is a result of the fluid velocity. The obtained liquid velocities are presented in Fig. 10.

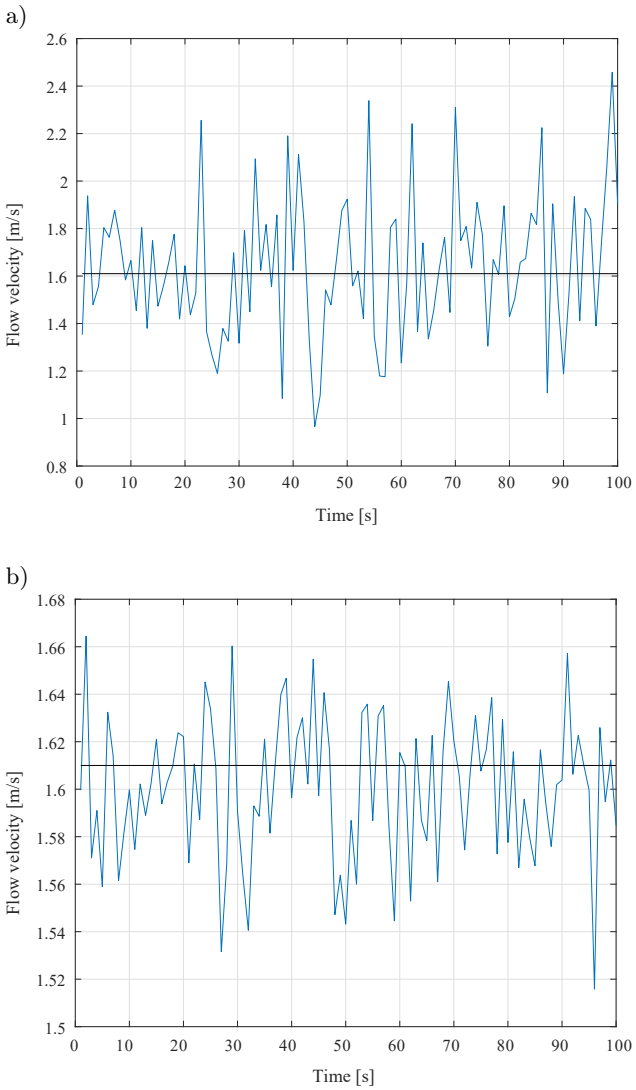


Fig. 10. Obtained liquid velocities:
a) PCV pipe; b) steel galvanized pipe.

The transducers operated at a sampling frequency rate of 10 MHz. The operating signal was a single sine wave with a frequency of 2 MHz and a burst duration of 20 µ. The transmit frequency was equal to 10 kHz. The flow velocity estimated by two methods separately

is presented in Fig. 11 and the accuracy speed liquid for the PCV pipe in Table 1.

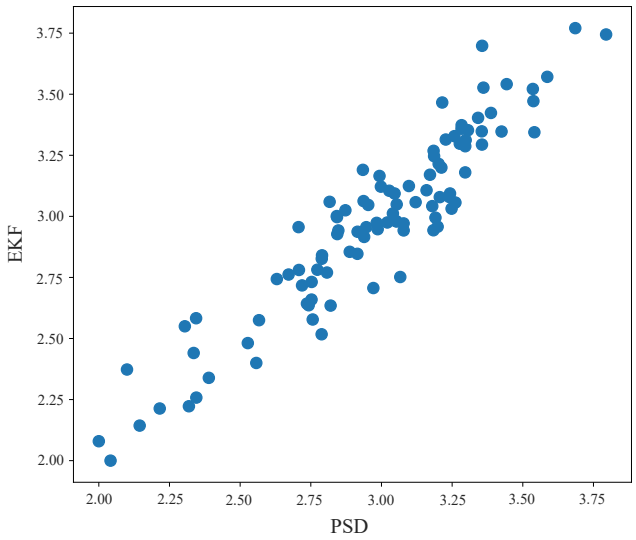


Fig. 11. Flow velocity estimated by two methods separately.

Table 1. Accuracy speed liquid for PCV pipe.

Real flow [m^3/h]	Single method	KDE
0.5	0.6	0.51
1.0	0.94	1.02
1.5	1.40	1.49
2.0	1.92	1.98
2.5	2.53	2.51
3.0	3.10	2.98

8.1. Comparison of fluid velocity estimation results for the analysed PSD estimators

PSD were estimated with two non-parametric methods: MTM and Welsh algorithms and two parametric: Burg and Prony methods. Different cases were analysed: positive and negative flow, different liquid flow velocity, different pipe diameters, and different pipe material. The mean flow velocity of the liquid estimation results for each of the four estimators of the PSD is shown in Table 2.

Table 2. Accuracy speed liquid for PSD pipe.

Pipe number	Real flow	Burg	Prony	MTM	Welch
1	0.20	0.2120	0.51	0.2202	0.2374
2	0.21	0.2099	0.4277	0.2126	0.1968
3	0.66	0.6639	1.2993	0.5740	0.6321
4	0.65	0.6776	1.3036	0.5898	0.6126
5	0.21	0.2162	1.0013	0.2189	0.2009

For each power spectral density estimator, the relative error between the exact value of the liquid flow velocity and the measured value was calculated. Errors are expressed as a percentage and results are shown in Fig. 12. It follows from Fig. 12 and Table 2 that the estimation of the power spectral density using the parametric Burg method had the smallest errors. To maintain the legibility of the above figure, the relative values of errors using the parametric Prony method were not included, as the errors reached several hundred percent.

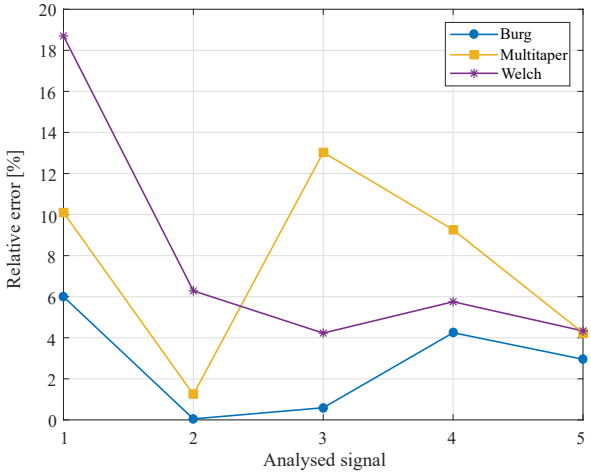


Fig. 12. Relative error plot of liquid flow velocity estimation for four power spectral density estimators.

After averaging the results, the global error was determined for each of the estimators. The results are presented in Table 3.

Table 3. Global error [%].

Burg	Prony	MTM	Welch
2.77	166.58	7.58	7.86

Figure 13 shows the variance of the power spectral density estimators used, for each of the analysed pipes in which fluid flow was studied. It is clear from

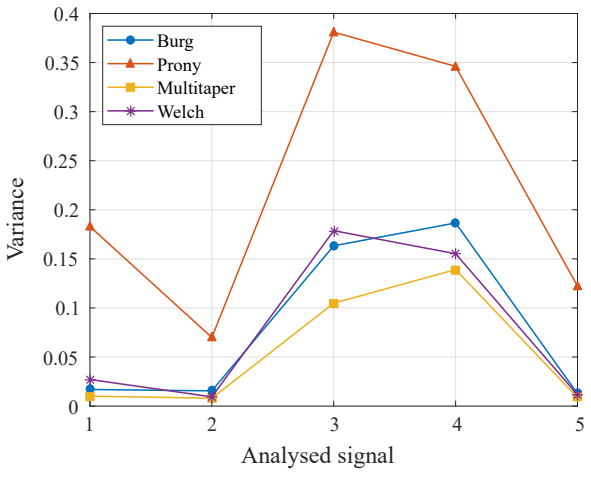


Fig. 13. The variance of the PSD estimators.

Fig. 13 above that the non-parametric MTM estimator has the smallest variance. Similar to the error analysis, the parametric Prony estimator has the highest variance.

9. Conclusions

The methods used to combine measurements for non-invasive measurement of the fluid flow velocity have proven to be accurate and return correct values. This has the advantage of using algorithms with low computational complexity. In combination with the KDE approach, a convenient and accurate tool was obtained. Another advantage of this solution is the ease of mounting the sensors in the pipe. A measurement accuracy of >95% was obtained.

References

1. Atkins M.D. (2016), Velocity field measurement using particle image velocimetry (PIV), [in:] *Application of Thermo-Fluidic Measurement Techniques*, Kim T., Lu T.J., Song S.J. [Eds.], pp. 125–166, doi: 10.1016/B978-0-12-809731-1.00005-8.
2. Avilán E.J., Reis V., Barreira L.E., Salgado C.M. (2013), Evaluation of cross correlation technique to measure flow in pipes of the oil industry, [in:] *2013 International Nuclear Atlantic Conference – INAC 2013*.
3. Bar-Shalom Y., Li X.-Rong, Kirubarajan T. (2004), *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*, John Wiley & Sons.
4. Beck M.S., Plaskowski A. (1987), *Cross Correlation Flowmeters, Their Design and Application*, Taylor & Francis.
5. Buermans J., Lampa J., Lemon D. (2009), Turbine flow measurement in low-head plants – Acoustic scintillation flow meter: Why? How? Where?, *Medicine*, Corpus ID: 113046881.
6. Cochran S. (2001), Ultrasonic instruments & devices – reference for modern instrumentation, techniques and technology, *Ultrasound in Medicine and Biology*, **27**(10): 1439, doi: 10.1016/S0301-5629(01)00386-6.
7. Cui Z. *et al.* (2016), A review on image reconstruction algorithms for electrical capacitance/resistance tomography, *Sensor Review*, **36**(4): 429–445, doi: 10.1108/SR-01-2016-0027.
8. Doran P.M. (2013), *Bioprocess engineering principles*, 2n ed., Academic Press Waltham.
9. Grewal M.S., Andrews A.P. (2001), *Kalman Filtering: Theory and Practice Using MATLAB*, 2nd ed., John Wiley & Sons.
10. Jazwinski A.H. (1970), *Stochastic Processes and Filtering Theory*, Elsevier Science.
11. Jones F.E (1995), *Techniques and topics in flow measurement*, CRC Press.
12. Kaipio J.P. *et al.* (2015), Process tomography and estimation of velocity fields, [in:] *Industrial Tomography*, Wang M. [Ed.], pp. 551–590, Woodhead Publishing, doi: 10.1016/B978-1-78242-118-4.00021-6.
13. Kang L. *et al.* (2019), Flow velocity measurement using a spatial averaging method with two-dimensional flexural ultrasonic array technology, *Sensors*, **19**(21): 4786, doi: 10.3390/s19214786.
14. Lucas G.P., Cory J., Waterfall R.C., Loh W.W., Dickin F.J. (1999), Measurement of the solids volume fraction and velocity distributions in solids-liquid flows using dual-plane electrical resistance tomography, *Flow Measurement and Instrumentation*, **10**(4): 249–258, doi: 10.1016/S0955-5986(99)00010-2.
15. Lyons R. (2004), *Understanding Digital Signal Processing*, 2nd ed., Pearson Education Incorporated.
16. Maybeck P.S. (1982), *Stochastic Models, Estimation and Control: Volume 2*, Academic Press.
17. Matani A., Oshiro O., Chihara K. (1996), Doppler signal processing of blood flow using a wavelet transform, *Japanese Journal of Applied Physics*, **35**(5S), doi: 10.1143/jjap.35.3131.
18. Mori M., Tezuka K., Aritomi M., Kukura H., Takeda Y. (2004), Industrial application experiences of new type flow-metering system based on ultrasonic-Doppler flow velocity-profile measurement, [in:] *Fourth International Symposium on Ultrasonic Doppler Methods for Fluid Mechanics and Fluid Engineering*, Japan.
19. Raffel M., Willert C.E., Kompenhans J. (1998), *Particle Image Velocimetry: A Practical Guide*, Springer Berlin, doi: 10.1007/978-3-662-03637-2.
20. Särkkä S. (2006), *Recursive Bayesian Inference on Stochastic Differential Equations*, Helsinki University of Technology, Laboratory of Computational Engineering Publications, Report B54, ISBN: 951-22-8127-9.
21. Solero G., Beghi M. (1995), Experimental fluid dynamic characterization of a premixed natural gas burner for domestic and semi-industrial applications, [in:] *The Institute of Energy's Second International Conference on Combustion & Emissions Control*, pp. 39–48, doi: 10.1016/B978-0-902597-49-5.50007-X.
22. Takeda Y. (1995), Instantaneous velocity profile measurement by ultrasonic Doppler method, *JSME International Journal Series B Fluids and Thermal Engineering*, **38**(1): 8–16, doi: 10.1299/jsmeb.38.8.
23. Takeda Y. (2012), *Ultrasonic Doppler Velocity Profiler for Fluid Flow, Fluid Mechanics and its Applications*, Springer, ISBN: 978-4-431-54026-7.
24. Xu Y., Wang H., Cui Z., Dong F. (2009), Application of electrical resistance tomography for slug flow measurement in gas/liquid flow of horizontal pipe, [in:] *2009 IEEE International Workshop on Imaging Systems and Techniques*, pp. 319–323, doi: 10.1109/IST.2009.5071657.
25. Wu J. (2018), Acoustic streaming and its applications, *Fluids*, **3**(4): 108, doi: 10.3390/fluids3040108.

Archives of Acoustics is the peer-reviewed journal of the Polish Academy of Sciences
and the Institute of Fundamental Technological Research PAS

AIMS AND SCOPE

Archives of Acoustics was founded in 1976 and since then publishes original research papers from all areas of acoustics including:

- acoustical measurements and instrumentation,
- musics acoustics,
- acousto-optics,
- architectural, building, and environmental acoustics,
- bioacoustics,
- electroacoustics,
- linear and nonlinear acoustics,
- noise and vibration,
- psychoacoustics,
- physical and chemical effects of sound,
- physiological acoustics,
- speech processing and communication systems,
- speech production and perception,
- sonochemistry,
- transducers,
- ultrasonics,
- underwater acoustics,
- quantum acoustics.

INFORMATION FOR AUTHORS

- Manuscripts intended for publication in *Archives of Acoustics* should be submitted in pdf format by an on-line procedure using **New Submission** link in **User Home** menu.
- Manuscript should be original and should not be submitted either previously or simultaneously elsewhere, neither in whole, nor in part.
- Submitted papers must be written in good English and proofread by a native speaker.
- In principle the papers should not exceed 40 000 typographic signs (about 20 typewritten pages).
- A brief abstract must be included. The keywords and subject classification for the article are required.
- Postal addresses, affiliations and email addresses for each author are required.
- The submitted manuscript should be accompanied by a cover letter containing the following information:
  - why the paper is submitted to *Archives of Acoustics*,
  - suggestion on the field of acoustics related to the topic of the submitted paper,
  - the statement that the manuscript is original, the submission has not been previously published, nor was it sent to another journal for consideration,
  - 3–5 names of suggested reviewers together with their affiliations, full postal and e-mail addresses; at least 3 suggested reviewers should be affiliated with other scientific institutions than the affiliations of the authors,
  - author's suggestion to classification of the paper as the research paper, review paper or technical note.
- Information regarding the preparation of manuscripts is provided on the journal's website **https://acoustics.ippt.pan.pl**.
- Reviews of manuscript is available online.
- Full information about the journal are available at **https://acoustics.ippt.pan.pl**.

Each manuscript is initially evaluated by the Editor and then allocated to an Section Editor according to specific subject area. Upon receiving the Section Editor sends it to reviewers. The Editor and the Section Editors make the final recommendation. Their opinions are crucial for the Editorial Committee to decide whether the paper can be accepted for publication or not.

The Word or LaTeX source and pdf files of the final version of articles will be requested after acceptance for publication. High-quality illustrations (format bitmap (gif, jpg) or PostScript, with resolution no lower than 300 dpi) should be sent along with the manuscript.

Editorial Board
Archives of Acoustics