eISSN 2300-262X (online)

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

ARCHIVES of ACOUSTICS

QUARTERLY

Vol. 49, No. 1, 2024

WARSAW

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

ARCHIVES of ACOUSTICS

QUARTERLY, Vol. 49, No. 1, 2024

Research Papers

Technical Notes

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

Editorial Board Office

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

Indexed in BazTech, Science Citation Index-Expanded (Web of Science Core Collection),
ICI Journal Master List, Scopus, PBN – Polska Bibliografia Naukowa,
Directory of Open Access Journals (DOAJ)

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

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

PUBLISHED IN POLAND

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

Archives of Acoustics Vol. **49**, No. 1, pp. 3–9 (2024), doi: 10.24425/aoa.2023.146825

Research Paper

Enhancing Speech Recognition in Adverse Listening Environments: The Impact of Brief Musical Training on Older Adults

Akhila R. NANDAKUMAR, Haralakatta Shivananjappa SOMASHEKARA, Vibha KANAGOKAR, Arivudai Nambi PITCHAIMUTHU*

Department of Audiology and Speech-Language Pathology, Kasturba Medical College, Mangalore
Manipal Academy of Higher Education
Manipal, Karnataka, India

*Corresponding Author e-mail: arivudainambi11@gmail.com

(*received December 14, 2022; accepted July 23, 2023; published online December 11, 2023*)

The present research investigated the effects of short-term musical training on speech recognition in adverse listening conditions in older adults. A total of 30 Kannada-speaking participants with no history of gross otologic, neurologic, or cognitive problems were divided equally into experimental (M = 63 years) and control groups (M = 65 years). Baseline and follow-up assessments for speech in noise (SNR50) and reverberation was carried out for both groups. The participants in the experimental group were subjected to Carnatic classical music training, which lasted for seven days. The Bayesian likelihood estimates revealed no difference in SNR50 and speech recognition scores in reverberation between baseline and followed-up assessment for the control group. Whereas, in the experimental group, the SNR50 reduced, and speech recognition scores improved following musical training, suggesting the positive impact of music training. The improved performance on speech recognition suggests that short-term musical training using Carnatic music can be used as a potential tool to improve speech recognition abilities in adverse listening conditions in older adults.

Keywords: musical training; carnatic music; speech recognition in noise; speech recognition in reverberation.

1. Introduction

Several anatomical and physiological changes occur in the auditory system of older adults as part of the aging process (Chisolm *et al.*, 2003). Aging causes alterations in the metabolic activity of the cochlea, leading to a decrease in endo-cochlear potentials (EP) (Wangemann, 2002). This reduction in EP impairs the functioning of the cochlear amplifier and raises the neural threshold (Schmiedt *et al.*, 2002). Additionally, the aging process disrupts the precise timing of the neuronal firing, resulting in inaccuracies in the phase locking of auditory neurons (Moser *et al.*, 2006). Animal models of aging have demonstrated a decrease in the size of spiral ganglion cells in Rosenthal's canal and a reduction of approximately 15 to 25% of cells throughout the cochlear duct (Mills *et al.*, 2006). The progressive degeneration of cells within the auditory system leads to various auditory perceptual deficits (Tun *et al.*, 2012), including reduced audibility (Schuknecht, Gacek, 1993), deterioration in suprathreshold auditory spectral processing (Nambi *et al.*, 2016), temporal processing (He *et al.*, 2008) and cognitive abilities (Verhaeghen, Cerella, 2002). These deficits can contribute to difficulties in speech perception in noisy and reverberant environments (Helfer, Wilber, 1990; Nambi *et al.*, 2016). The most common complaint among older adults is the difficulty in comprehending speech in adverse listening conditions. This difficulty stems from their diminished auditory processing abilities, which hinder their ability to separate target speech from background noise, resulting in reduced speech perception in noisy environments (Schoof, Rosen, 2014). Due to the decline in auditory processing abilities in older adults, their passive and effortless speech processing in noisy environments is compromised (Rabbitt, 1990). As a compensatory mechanism, older adults rely on active and conscious signal processing, which relies on intact cognitive functioning. However, the aging process also impacts cognitive abilities, which can contribute to difficulties in speech perception in noisy situations (Tun *et al.*, 2002).

Methods to overcome communication difficulties in older adults have been the topic of interest among researchers. One preventive measure often recommended to counter the effects of aging is engaging in physical exercise (Alessio *et al.*, 2002; Curhan *et al.*, 2013). Physical exercise may help in preventing age-related auditory disorders, although it remains unclear whether it can reverse hearing changes that have already occurred due to aging. Other studies have demonstrated the benefits of auditory training using different stimuli, such as monosyllables (Burk *et al.*, 2006) and consonant-vowel transitions at the syllable, word, sentence, and context levels (Anderson *et al.*, 2013). These training methods have improved neural timing, processing speed, and speech perception in noisy environments.

In a broader sense, musical training can be considered a form of auditory training, and long-term musical training has been found to have positive effects on auditory and cognitive abilities (Parbery-Clark *et al.*, 2009a; 2009b; 2012; 2013; Kraus, Chandrasekaran, 2010; Patel, 2011). Older adult musicians with long-term expertise retain neuro-physiological advantages due to music which may improve their speech coding abilities. Anderson and Kraus (2010) found that these musicians outperformed their non-musician counterparts in tasks involving auditory, spectral, temporal, and cognitive processing. These promising findings suggest that musical training could serve as an effective strategy to mitigate speech perception deficits in older adults (Kraus, White-Schwoch, 2014). Therefore, it would be interesting to investigate whether musical training can be employed as an auditory training method to overcome speech recognition deficits in challenging listening conditions.

To the best of our knowledge, only Jain *et al.* (2015) investigated the effect of short-term musical training on speech recognition in noise among young adults, reporting enhancements in speech recognition. However, the impact of short-term musical training on speech recognition abilities in older adults remains unexplored. Hence, the present study aims to investigate the effects of short-term musical training on speech recognition in adverse listening conditions in older adults.

2. Method

The Institutional Ethics Committee (IEC) at Kasturba Medical College (KMC), Mangaluru, approved the research protocol. A total of 30 participants were selected using the convenient sampling method and were evenly divided into experimental and control groups. All participants were native Kannada speakers with no prior musical training experience or significant ear, neurological, or cognitive issues. Before conducting the study, informed consent was obtained from all individuals. Table 1 depicts the mean age of the groups with their average pure tone thresholds at 500 Hz, 1 kHz, and 2 kHz (PTA1), as well as 1, 2, and 4 kHz (PTA2). An independent t-test revealed no statistically significant difference ($p > 0.05$) in PTA1 ($t_{28} = 1.619$, $p = 0.117$) and PTA2 ($t_{28} = 1.337$, $p = 0.192$) between the two groups.

Table 1. Age and hearing thresholds of all the participants in the experimental and control group.

		PTA (M & SD)	
	Mean age	PTA1	PTA2
Experimental group	63 years	31.77 (6.50)	35.55 (6.53)
Control group	65 years	36.66 (9.71)	40.55 (12.92)

PTA1: average of pure tone thresholds at 500 Hz, 1 kHz, 2 kHz.
PTA2: average of pure tone thresholds at 1, 2, 4 kHz.

3. Procedure

The research was conducted in three distinct phases. During the initial phase, participants from both groups underwent testing to evaluate their speech recognition ability in noisy and reverberant conditions. In the subsequent phase, participants in the experimental groups received music training. Finally, in the last phase, the speech recognition ability in noise and reverberation was reassessed for all participants in both groups. Stimuli for speech recognition tests were presented from a personal laptop and routed through the Creative Soundblaster X-Fi USB sound card, while the Sennheiser HD 280 Pro headphones were used for stimulus presentation. All stimuli were digitized at a sampling rate of 44 100 Hz. The signal processing for speech recognition in noise and the music training paradigm was implemented in the MATLAB version 7.10.0 platform. Additionally, the signal processing for speech recognition in reverberation was performed using Adobe Audition Version 3 software.

4. Assessment of speech recognition in noise

The standard QuickSIN protocol was employed to estimate speech recognition in noise. Two lists of the standard QuickSIN Kannada (Methi *et al.*, 2009) sentences, spoken by the female speakers were used as the targets and a 4-talker speech babble was used as the background noise. Each list consisted of seven sentences, with the first sentence presented at a signal-to-noise ratio (SNR) of 20 dB. Subsequently, the SNR was gradually decreased in 5 dB increments until reaching −10 dB SNR for the final sentence. The sentences were presented at the most comfortable level (MCL) of the participant. For each sentence, the count of correctly identified keywords by each participant was determined and converted into the proportion of correct responses for each list. The SNR required to achieve a 50% correct recognition score (SNR50) was then estimated by fitting the cumulative Gaussian psychometric function to the proportion of correct responses at each SNR level. SNR50 was calculated as the midpoint of the psychometric function, separately for each list, and then averaged. In total, four sentence lists were employed to assess SNR50, with two sets used for pre-training evaluation and the remaining two sets for post-training assessment.

4.1. Assessment of speech recognition in reverberation

A single list of sentences from the QuickSIN test was convolved with binaural room impulse responses (BRIRs) to simulate speech recognition in a reverberant environment. This BRIR was generated to simulate a standard rectangular auditorium with an average reverberation time of 0.6 seconds. The reverberant material was presented to the participants at the MCL set by the participants. The total count of accurately identified keywords was tallied, with a maximum achievable score of 35. For assessment purposes, two sets of sentences were utilized, one for the pre-training evaluation and another for the post-training assessment.

4.2. Music training

The participants in the experimental group were subjected to short-term musical training spanning approximately seven days. The training initially consisted of ten Sampoorna ragas of Carnatic classical music. The ascending and descending pattern (Arohana and Avarohana) of all ten Sampoorna ragas were recorded using violin, veena, and flute instruments played by three professional artists with over ten years of experience. These ten ragas were divided into two lists, each containing five ragas. Subsequently, based on a pilot study, only one list comprising the ragas Mayamalavagowla, Kalyani, Thodi, Natabhairavi, and Charukeshi was selected for the musical training. A custom training module was developed in the graphical user interface (GUI) format, incorporating a training component and an assessment module for raga identification.

During the initial training session, the participants were familiarized with all five ragas by listening to violin samples. The unique characteristics of each raga were explained to them. Gradually, they were taught to identify and discriminate the ragas based on the ascends and descends. Throughout the training, multiple rehearsals and feedback were provided. At the end of each session, the participant's ability to identify each raga was assessed by randomly presenting each raga ten times. The training session continued until the participant achieved a 100% correct score.

Once the training with the violin samples was completed, a similar process was followed using veena samples. In the final phase of the training, the participant's ability to transfer the knowledge of ragas acquired from the violin and veena to the flute was ensured. In this stage, the participants underwent a raga identification test where each raga played on the flute was randomly presented ten times. The training was considered finished when the participants achieved a flawless score of 100%. If any participants failed to attain a perfect score of 100%, they were taken back to the previous stage, where they received further training with veena samples. Once they achieved a perfect score for the veena samples, they progressed to the next stage for the raga identification test with flute samples. This process continued until all participants obtained perfect scores of 100% for the flute samples.

5. Results

The statistical analyses were performed using the JASP version 1.17.1.0 software. JASP is a comprehensive and user-friendly statistical software that offers a wide range of tools for data analysis, including Bayesian and frequentist methods. With its intuitive interface and extensive statistical capabilities, JASP provides researchers with a powerful platform for conducting rigorous and transparent statistical analyses. In the current study, series of Bayesian paired sample t-tests were employed to investigate the main effect of music training on speech recognition outcomes in noise and reverberation. Series of Bayesian independent sample t-tests were performed to examine the disparity in speech recognition performance in noise and reverberation between the control and experimental groups.

5.1. Speech recognition in noise

The statistical analysis revealed that the SNR50 of participants in the experimental group was significantly different in the post-training session compared to the pre-training session ($BF_{10} = 9.20$). Music training had a positive influence by reducing the SNR50 in the experimental group. On the other hand, there was no significant difference in SNR50 between the baseline and follow-up sessions ($BF_{10} = 0.44$) in the control group.

The statistical analysis revealed that there was no significant difference in SNR50 between the control group and experimental group in the pre-training session ($BF_{10} = 0.35$). However, after subjecting the experimental group to musical training, the SNR50 was estimated in both the control and experimental groups.

The SNR50 in the experimental group was found to be better than the control group (BF_{10} = 141.5). The mean and standard deviation of SNR50 in baseline and follow-up sessions in both the control and experimental group is depicted in Fig. 1.

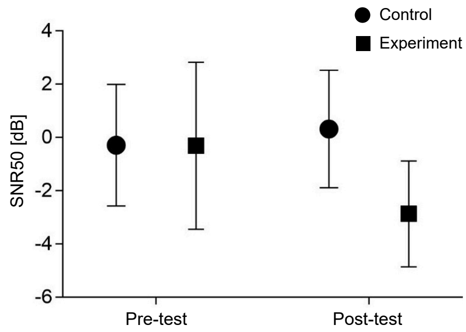

Fig. 1. Mean and standard deviation of SNR50 in baseline and follow-up sessions for both the control and experimental groups.

5.2. Speech recognition in reverberation

The main effect of music training on speech recognition scores in reverberation was evaluated by comparing the scores obtained in pre-training and post-training sessions. The total correct speech recognition scores of the participants in the experimental group were significantly larger in post-training sessions than in pre-training sessions (BF_{10} = 7.57). This result suggests that music training has improved speech recognition ability in reverberation. In contrast, there was no significant difference (BF_{10} = 0.45) in the baseline and follow-up performance of the control group on speech recognition scores.

Speech recognition scores measured at the pre-training session were not different between the control and experimental group (BF_{10} = 0.26). The speech recognition scores in reverberation measured following the music training in the experimental group were higher (BF_{10} = 11.68) than the speech recognition scores of the control group. The mean and standard deviations of the correct scores are depicted in Fig. 2.

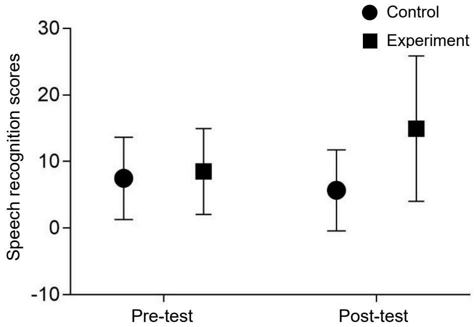

Fig. 2. Mean and standard deviations of speech recognition scores in baseline and follow-up sessions for both the control and experimental group.

6. Discussion

The present study suggests a positive impact of short-term musical training on speech recognition abilities in older adults in adverse listening conditions. This is evident from the improved SNR50 and speech recognition scores under reverberant conditions. Previous research has consistently shown that musicians tend to exhibit enhanced auditory abilities compared to non-musicians, as demonstrated in various studies (Kraus, Chandrasekaran, 2010; Kraus, White-Schwoch, 2014; Musacchia *et al.*, 2007; Parbery-Clark *et al.*, 2012; Rammsayer, Altenmüller, 2006; Slater *et al.*, 2015; Strait, Kraus, 2011). Furthermore, even older adults with musical experience performed better than their non-musician counterparts in speech-in-noise tasks (Anderson *et al.*, 2013; Kraus, White-Schwoch, 2014; White-Schwoch *et al.*, 2013).

Electrophysiological studies have indicated that long-term musical training can influence neural encoding by altering the responsiveness of sub-cortical and cortical neurons, thereby enhancing auditory processing ability. Recent electrophysiological studies focussing on short-term musical training lasting eight days, conducted on young non-musicians, observed changes primarily in cortical responses rather than subcortical responses (Devi *et al.*, 2015; Jain *et al.*, 2014). These findings provide evidence that even brief musical training can lead to improvements in the neural encoding process. Thus, it can be inferred that the improvements observed in the current study may also be attributed to enhanced neural encoding mechanisms associated with musical training.

Parbery-Clark *et al.* (2009) investigated subcortical speech coding in musicians and non-musicians using speech-evoked auditory brainstem responses. They found that the peaks of the waveform, carrying crucial temporal cues, were better preserved in musicians than in non-musicians, both in quiet and in the presence of background noise. Musicians also exhibited enhanced phase-locking abilities compared to non-musicians. The process of learning music enables the auditory system to adapt and extract essential cues from complex signals, resulting in improved neural representation within the auditory system. It permits better coding of the temporal and spectral aspects of the signal and also helps in concurrent stream segregation, which is essential for perceiving speech in adverse listening conditions (Zendel, Alain, 2009).

Exposure to music can strengthen the neural responses to stimuli and facilitate bottom-up processing. The auditory efferent system, known for suppressing irrelevant background noise, can enhance the perception of target speech (Luo *et al.*, 2008; Zhang *et al.*, 1997). Through prolonged musical training, top-down processing may modulate neural responses and magnify the cues that are important for stimulus identification. The formation of the auditory template plays a vital role in speech perception, and speaker identification (Best *et al.*, 2008). It is possible when good timber perception produces an excellent harmonic representation of the complex stimulus. A good perception of timbre, which generates a high-quality harmonic representation of complex stimuli, was observed in musicians compared to non-musicians. Additionally, musicians exhibited heightened sensitivity to subtle harmonic changes (Musacchia *et al.*, 2008; Zendel, Alain, 2009). These factors could have also influenced our study, potentially contributing to improved speech performance in older adults following musical training.

Various hypotheses have been proposed to explain the musical training-dependent changes in auditory processing abilities. Patel (2011) introduced the OPERA (overlap, precision, emotions, repetition, and attention) hypothesis, which offers potential explanations for the changes observed in auditory processing abilities resulting from musical training. The overlap hypothesis suggests anatomical overlap in the brain networks responsible for processing music and speech. According to the precision theory, the heightened precision required for music processing can also be beneficial for speech processing. The emotion theory proposes that the positive emotions evoked by music activate the brain's reward centres, leading to neural plasticity. Additionally, the brain's networks are frequently exposed to musical stimuli, leading to the repetition effect. Lastly, the attention theory states that the networks engaged in music processing are linked to focused attention, which is also crucial for recognizing speech in noisy environments. Therefore, the OPERA hypothesis provides a framework for understanding the improvements in auditory processing and speech recognition abilities in challenging listening conditions associated with music training.

Musical training also presents challenges to short-term memory and attention. Throughout the training, the participants were required to listen attentively to the ragas being played, placing a cognitive load on their memory as they aimed to recognize the raga based on the notes. Patel (2011) hypothesized that focused attention on the intricate details of the musical sounds promotes plasticity. Studies on animals have also demonstrated that training-induced plasticity is enhanced when active listening is involved (Fritz *et al.*, 2005). Anderson *et al.* (2013) expressed a similar viewpoint, emphasizing the importance of cognitive involvement in auditory training programs. The cognitive demand on memory leads to an increased reliance on perceptual cues mediated by the prefrontal cortex. Consequently, the perceptual demands and memory interacted during the training program to strengthen the neural representation of speech perception in the presence of background noise. Therefore, the heightened cognitive load experienced during the training is likely to positively impact the auditory processing and speech recognition abilities of older adults.

Kraus and White-Schwoch (2014) believed that music training enhances auditory processing regardless of duration and intensity. The findings of the present study align with their viewpoint, indicating that short-term musical training improves auditory processing and speech recognition abilities in older adults. Consequently, short-term music training holds potential as a way to alleviate auditory processing and speech recognition deficits in older adults. However, further investigation is necessary to determine the minimum duration of training required to maintain the generalized benefits. This aspect presents a promising avenue for future research in this field.

One notable finding in the present study is the extent of improvement observed in SNR50. Specifically, the magnitude of improvement observed in our study slightly exceeds that observed in long-term trained musicians who are native English speakers (Parbery-Clark *et al.*, 2009b). Conversely, Jain *et al.* (2015) reported a similar magnitude of improvement following short-term music training in young native Kannada language speakers. These observations lead to the speculation that Carnatic music may be more effective in enhancing speech understanding abilities compared to other genres of music. Additionally, the favourable phonetic characteristics of the Kannada language may contribute to the manifestation of the effects of music training on speech understanding. However, further exploration is necessary to investigate these speculations. Mishra and Panda (2014), also reported a positive effect of Carnatic music on auditory perceptual abilities, observing improved auditory perceptual abilities in Carnatic musicians compared to non-musicians. Each Carnatic music raga has unique ascending and descending musical patterns, distinguished by variations in the pitch of the notes. The ragas chosen for this study included all seven notes of music, known as Sampoorna ragas in Carnatic music. Indian classical music experts recognize the distinct properties of each raga, such as the tonic frequency, Swaras, Arohana (ascending notes), Avarohana (descending notes), Vaadi (primary note), Samvaadi (secondary note), and more. Each raga follows specific rules that define its characteristics and set it apart from others. While some ragas may share the same set of notes or Swaras, their combinations differ. The ascending pitch sequence is known as Arohana, while the descending sequence is called Avarohana. The fundamental basis of differentiation lies in the frequency and corresponding pitch. Unlike Western music, Indian musical notes do not adhere to standardized frequencies. Instead, artists choose a convenient frequency as a reference, which serves as the base for the entire raga. The ragas selected for this study differed from one another by one or two notes, with these differing notes falling in frequencies close to each other. Due to these unique qualities of Carnatic music ragas, they possess a higher potential as effective tools for auditory training.

7. Conclusion

The ability to perceive and distinguish important cues such as timber, pitch, and timing is critical in processing complex signals like speech and music. Developing precise auditory discrimination skills is vital for effectively extracting these cues. Musical training plays a crucial role in refining these skills and strengthening the neural representation of the auditory system, thereby enhancing speech perception. The present findings suggest that even a short period of musical training can significantly improve the speech perception abilities of older adults, especially in challenging listening conditions. Furthermore, the enjoyable nature of music further underscores its potential as a valuable tool for enhancing speech perception skills in adverse listening situations for older adults. However, the long-term sustainability of the training effect cannot be determined solely based on the current study, calling for further research on the long-term maintenance of short-term training outcomes.

References

1. Alessio H.M., Hutchinson K.M., Price A.L., Reinart L., Sautman M.J. (2002), Study finds higher cardiovascular fitness associated with greater hearing acuity, *The Hearing Journal*, **55**(8): 32–40, doi: 10.1097/01.HJ0000293292.87003.3d.
2. Anderson S., Kraus N. (2010), Sensory-cognitive interaction in the neural encoding of speech in noise: A review, *Journal of the American Academy of Audiology*, **21**(9): 575–585, doi: 10.3766/jaaa.21.9.3.
3. Anderson S., White-Schwoch T., Parbery-Clark A., Kraus N. (2013), Reversal of age-related neural timing delays with training, *Proceedings of the National Academy of Sciences of the United States of America*, **110**(11): 4357–4362, doi: 10.1073/pnas.1213555110.
4. Best V., Ozmeral E.J., Kopčo N., Shinn-Cunningham B.G. (2008), Object continuity enhances selective auditory attention, *Proceedings of the National Academy of Sciences of the United States of America*, **105**(35): 13174–13178, doi: 10.1073/pnas.0803718105.
5. Burk M.H., Humes L.E., Amos N.E., Strauser L.E. (2006), Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners, *Ear and Hearing*, **27**(3): 263–278, doi: 10.1097/01.aud.0000215980.21158.a2.
6. Chisolm T.H., Willott J.F., Lister J.J. (2003), The aging auditory system: Anatomic and physiologic changes and implications for rehabilitation, *International Journal of Audiology*, **42**(Suppl 2): 3–10, doi: 10.3109/14992020309074637.
7. Curhan S.G., Eavey R., Wang M., Stampfer M.J., Curhan G.C. (2013), Body mass index, waist circumference, physical activity, and risk of hearing loss in women, *The American Journal of Medicine*, **126**(12): 1142.e1–e8. 10.1016/j.amjmed.2013.04.026.
8. Devi N., Uppunda A.K., Mohamed H. (2015), Short-term perceptual training of music and its effect on neural encoding, *International Journal of Health Sciences & Research*, **5**(5): 347–356.
9. Fritz J., Elhilali M., Shamma S. (2005), Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex, *Hearing Research*, **206**(1–2): 159–176, doi: 10.1016/j.heares.2005.01.015.
10. He N., Mills J.H., Ahlstrom J.B., Dubno J.R. (2008), Age-related differences in the temporal modulation transfer function with pure-tone carriers, *The Journal of the Acoustical Society of America*, **124**(6): 3841–3849, doi: 10.1121/1.2998779.
11. Helfer K.S., Wilber L.A. (1990), Hearing loss, aging, and speech perception in reverberation and noise, *Journal of Speech & Hearing Research*, **33**(1): 149–155, doi: 10.1044/jshr.3301.149.
12. Jain C., Mohamed H., Kumar U.A. (2014), Short-term musical training and pyschoacoustical abilities, *Audiology Research*, **4**(102): 102, doi: 10.4081/audiores.2014.102.
13. Jain C., Mohamed H., Kumar U.A. (2015), The effect of short-term musical training on speech perception in noise, *Audiology Research*, **5**(5): 111, doi: 10.4081/audiores.2015.111.
14. Kraus N., Chandrasekaran B. (2010), Music training for the development of auditory skills, *Science and Society*, **11**(8): 599–606, doi: 10.1038/nrn2882.
15. Kraus N., White-Schwoch T. (2014), Music training: Lifelong investment to protect the brain from aging and hearing loss, *Acoustics Australia*, **42**(2): 117–123.
16. Luo F., Wang Q., Kashani A., Yan J. (2008), Corticofugal modulation of initial sound processing in the brain, *Journal of Neuroscience*, **28**(45): 11615–11621, doi: 10.1523/JNEUROSCI.3972-08.2008.
17. Methi R., Avinash, Kumar U. (2009), Development of sentence material for Quick Speech in Noise test (Quick SIN) in Kannada, *Journal of Indian Speech and Hearing Association*, **23**: 59–65.
18. Mills J.H., Schmiedt R.A., Schulte B.A., Dubno J.R. (2006), Age-related hearing loss: A loss of voltage, not hair cells, *Seminars in Hearing*, **27**(4): 228–236, doi: 10.1055/s-2006-954849.
19. Mishra S.K., Panda M.R. (2014), Experience-dependent learning of auditory temporal resolution, *NeuroReport*, **25**(2): 134–137, doi: 10.1097/WNR.0000000000000093.
20. Moser T., Neef A., Khimich D. (2006), Mechanisms underlying the temporal precision of sound coding at the inner hair cell ribbon synapse, *The Journal of Physiology*, **576**(1): 55–62, doi: 10.1113/jphysiol.2006.114835.

21. Musacchia G., Sams M., Skoe E., Kraus N. (2007), Musicians have enhanced subcortical auditory and audiovisual processing of speech and music, *Proceedings of the National Academy of Sciences of the United States of America*, **104**(40): 15894–15898, doi: 10.1073/pnas.0701498104.

22. Musacchia G., Strait D., Kraus N. (2008), Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians, *Hearing Research*, **241**(1–2): 34–42, doi: 10.1016/j.heares.2008.04.013.

23. Nambi P.M.A., Sangamanatha A.V., Vikas M.D., Bhat J.S., Shama K. (2016), Perception of spectral ripples and speech perception in noise by older adults, *Ageing International*, **41**(3): 283–297, doi: 10.1007/s12126-016-9248-4.

24. Parbery-Clark A., Anderson S., Kraus N. (2013), Musicians change their tune: How hearing loss alters the neural code, *Hearing Research*, **302**: 121–131, doi: 10.1016/j.heares.2013.03.009.

25. Parbery-Clark A., Skoe E., Kraus N. (2009a), Musical experience limits the degradative effects of background noise on the neural processing of sound, *Journal of Neuroscience*, **29**(45): 14100–14107, doi: 10.1523/JNEUROSCI.3256-09.2009.

26. Parbery-Clark A., Skoe E., Lam C., Kraus N. (2009b), Musician enhancement for speech-in-noise, *Ear and Hearing*, **30**(6): 653–661, doi: 10.1097/AUD.0b013e3181b412e9.

27. Parbery-Clark A., Tierney A., Strait D.L., Kraus N. (2012), Musicians have fine-tuned neural distinction of speech syllables, *Neuroscience*, **219**: 111–119, doi: 10.1016/j.neuroscience.2012.05.042.

28. Patel A.D. (2011), Why would musical training benefit the neural encoding of speech? The OPERA hypothesis, *Frontiers in Psychology*, **2**, doi: 10.3389/fpsyg.2011.00142.

29. Rabbitt P. (1990), Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ, *Acta Oto-Laryngologica*, **111**(476): 167–175, doi: 10.3109/00016489109127274.

30. Rammsayer T., Altenmüller E. (2006), Temporal information processing in musicians and nonmusicians, *Music Perception*, **24**(1): 37–48, doi: 10.1525/mp.2006.24.1.37.

31. Schmiedt R.A., Lang H., Okamura H., Schulte B.A. (2002), Effects of furosemide applied chronically to the round window: A model of metabolic presbyacusis, *Journal of Neuroscience*, **22**(21): 9643–9650, doi: 10.1523/jneurosci.22-21-09643.2002.

32. Schoof T., Rosen S. (2014), The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners, *Frontiers in Aging Neuroscience*, **6**, doi: 10.3389/fnagi.2014.00307.

33. Schuknecht H.F., Gacek M.R. (1993), Cochlear pathology in presbycusis, *The Annals of Otology, Rhinology, and Laryngology*, **102**(1): 1–16, doi: 10.1177/00034894931020S101.

34. Slater J., Skoe E., Strait D.L., O'Connell S., Thompson E., Kraus N. (2015), Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program, *Behavioural Brain Research*, **291**: 244–252, doi: 10.1016/j.bbr.2015.05.026.

35. Strait D.L., Kraus N. (2011), Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise, *Frontiers in Psychology*, **2**, doi: 10.3389/fpsyg.2011.00113.

36. Tun P.A., O'Kane G.,Wingfield A. (2002), Distraction by competing speech in young and older adult listeners, *Psychology and Aging*, **17**(3): 453–467, doi: 10.1037/0882-7974.17.3.453.

37. Tun P.A., Williams V.A., Small B.J., Hafter E.R. (2012), The effects of aging on auditory processing and cognition, *American Journal of Audiology*, **21**(2): 344–350, doi: 10.1044/1059-0889(2012/12-0030).

38. Verhaeghen P., Cerella J. (2002), Aging, executive control, and attention: a review of meta-analyses, *Neuroscience and Biobehavioral Reviews*, **26**(7): 849–857, doi: 10.1016/s0149-7634(02)00071-4.

39. Wangemann P. (2002), K^+ cycling and the endocochlear potential. *Hearing Research*, **165**(1–2): 1–9, doi: 10.1016/S0378-5955(02)00279-4.

40. White-Schwoch T., Carr K.W., Anderson S., Strait D.L., Kraus N. (2013), Older adults benefit from music training early in life: Biological evidence for long-term training-driven plasticity, *Journal of Neuroscience*, **33**(45): 17667–17674, doi: 10.1523/JNEUROSCI.2560-13.2013.

41. Zendel B.R., Alain C. (2009), Concurrent sound segregation is enhanced in musicians, *Cognitive Neutroscience*, **21**(8): 1488–1498, doi: 10.1162/jocn.2009.21140.

42. Zhang Y., Suga N., Yan J. (1997), Corticofugal modulation of frequency processing in bat auditory system, *Nature*, **387**(6636): 900–903, doi: 10.1038/43180.

Archives of Acoustics Vol. 49, No. 1, pp. 11–18 (2024), doi: 10.24425/aoa.2024.148774

Research Paper

The Influence of Violin Tailpiece Material on Acoustic Properties of a Violin

Adam ŁAPIŃSKI[(1)], Ewa SKRODZKA[(2)*], Andrzej WICHER[(2)]

[(1)] *The Ignacy Jan Paderewski Academy of Music in Poznań*
Poznań, Poland; e-mail: adam.lapinski113@gmail.com

[(2)] *Department of Acoustics, Faculty of Physics*
Adam Mickiewicz University
Poznań, Poland; e-mail: awaku@amu.edu.pl

*Corresponding Author e-mail: afa@amu.edu.pl

(*received July 25, 2023; accepted October 22, 2023; published online January 18, 2024*)

The different mechanical properties of the materials from which the tailpieces are made have a noticeable effect on the acoustic performance of the violin. These elements are made today from ebony, rosewood, boxwood, aluminium, or plastic. The aim of this study was to check the exact impact of tailpieces made of different materials on the frequency response function (FRF) of a violin's bridge and the timbre of the instrument's sound. For this purpose, the bridge FRF measurement was carried out, and a psychoacoustic test was conducted. The material from which the tailpiece is made to the greatest extent affects the modal frequencies in the range 530–610 Hz (mode B1+), which mainly manifested itself in a change in the instrument's timbre in terms of the brightness factor. The study showed that the lighter the tailpiece, the darker the sound of the violin. It was also revealed that the selection of accessories affects factors such as openness, thickness, and overall quality of the sound.

Keywords: tailpiece; violin; acoustical properties; exotic materials; violin-making; luthiery.

1. Introduction

Violin-making is an extraordinary art form that passionately combines artistic freedom with scientific objectivity. The mathematical complexity that prevents a comprehensive definition in terms of acoustics, the beauty of the harmonious form, and the appeal of each curve and line make the violin more than just a musical instrument. Even though for years, physicists, acousticians, and luthiers have been trying to explain the relationship between the physical parameters of a violin and the timbre of its sound (Hutchins, 1983; Skrodzka *et al.*, 2013; 2014), there is still a vast number of unexplored variables whose influence on the sound is unknown. One such variable is the tailpiece, a small plate used to attach the strings to the instrument. Its appearance has undergone constant changes throughout the evolution of the violin. Today, these pieces are made from the wood of dense hardwoods such as ebony, rosewood, and boxwood (Bucur, 2016). However, in the 17th century, due to the difficult availability of exotic materials, violin accessories were made from more common woods, such as sycamore (Pollens, 2009). Tailpieces made of this material were often decorated with numerous ornaments, sometimes with intricate marquetry or intarsia. Having different shapes and dimensions, they were an element that gave the instrument its individual and unique character (Houssay, 2014). Unfortunately, the 19th century industrial revolution standardised the appearance of violin accessories. In the manufactories and factories, there was no time for individual attention to each instrument and no room for woodcarving show-offs when making tailpieces. In the 20th century, with the development of new synthetic materials, plastic, aluminium, and graphite composite also began to be used.

There have been many attempts to explain the influence of a tailpiece on a violin sound (Fouilhé *et al.*, 2009; 2010; Leung, 2016). In order to objectively present the acoustic properties of tailpiece vibration, an experimental modal analysis was carried out by Stough (1996). He outlined five tailpiece modes, which were divided into two groups based on the nature of the vibration: three swing modes and two rotation modes (see Table 1). It was also noted that their frequency is influenced by the tailpiece's weight and the length of the tailgut. The topic was later revisited by Borman and Stoppani (nd), who made and published visual representations of the tailpiece's modal vibrations on his website (Stoppani *et al.*, nd). The available animations clearly explain the nature of the dynamics and movement of each violin element at a specific frequency.

Comparing Tables 1 and 2, it can be seen that, some frequency modes of the tailpiece overlap with some of the modal frequencies of the plates. This phenomenon has preoccupied scientists for years. Both Hutchins (1993) and Fouilhé *at al.* (2011), noted that matching the tailpiece's horizontal rotation mode (Rh) and vertical rotation mode (Rv) modal frequencies to the instrument's air modes can noticeably affect the violin's tone and timbre. Fouilhé also points out that matching the tailpiece's main resonance to a frequency he calls "the body's wolf resonance" allows wolf suppression. Wolves are a significant issue, particularly for cello players. Thus, many luthiers experiment with the parameters of the tailpiece to weaken them or change the frequency at which they occur (Zhang, Woodhouse, 2018; Gourc *et al.*, 2022).

Table 1. Modes of a tailpiece identified by Stough (1996). Frequency and quality factor (Q) are shown in ranges because their exact values depend on individual tailpiece parameters.

Mode	Full name	Frequency [Hz]	Q
Sb	Swing bass side mode	100–140	50–80
St	Swing treble side mode	120–160	60–80
Su	Swing under mode	180–230	35–70
Rh	Rotation mode, horizontal axis	300–800	34–110
Rv	Rotation mode, vertical axis	300–800	38–110

Table 2. Violin modes observed by Stoppani *et al.* (nd). Frequency and quality factor (Q) are shown as the average of two measurement sessions.

Mode	Full name	Frequency [Hz]	Q
A0	Fundamental air resonance	283	33.5
CBR	Centre bout romboid	416	50.3
A1	Second air resonance	470	54.0
B1−	First breathing mode	494	50.5
B1+	Second breathing mode	588	38.4

Nowadays, as shown in Fig. 1, there are three main styles that tailpieces are made in: English (Hill), which has a pointed shape similar to a house's roof; French, which has an elegant, rounded shape; and

Fig. 1. Different types of tailpiece shapes described by Folland (2010).

"tulip", which calls to mind a wine glass or a tulip. However, according to Folland (2010), the shape does not significantly affect the sound. On the other hand, it turns out that the tailpiece's position relative to the bridge can make a huge difference in sound timbre. Modern luthiers pay particular attention to the length of the string between the bridge and the tailpiece nut while installing a new tailpiece. This distance is called the after-length of the string. In a violin, it should be about 54.5 mm long ($^1/_6$ of the length of the vibrating string). By shortening or lengthening the tail gut, the after-length can be modified, and different sound effects can be achieved. According to Fouilhé and Houssay (2013), the shortening of the tail gut stiffened the cello tailpiece, resulting in a more powerful, harmonically richer, yet more demanding in emission sound. While the maximum extension of the tail gut resulted in a milder, less powerful, aggressive sound and diminution of the wolf note appearance. Adjusting the after-length by changing the length of the tailpiece did not noticeably affect the sound.

Another critical factor that can affect the overall sound of a violin is the material from which the tailpiece is made. In everyday use are those made of rosewood, boxwood, ebony, plastic, or metal. The physical properties of these materials differ considerably. It is difficult to imagine that the significant differences in density, the modulus of elasticity, and the damping coefficient of these materials do not affect the natural frequencies of the tailpiece and the sound of the entire instrument. Much research has been devoted to the wood used in the construction of violin plates (Mania *et al.*, 2015; 2017). Both scientists and luthiers pay a lot of attention to the properties of the material used in this process. It is difficult, however, to find information in the literature that does not deal with spruce and maple, the woods most commonly used to make instruments. The influence of ebony fingerboards or tailpieces made of exotic materials has not yet been well described. To the best of the authors' knowledge, no studies have yet been conducted to answer questions about the tailpiece material's effect on the violin's acoustic properties. The aim of this work is to find the answer to the question if there are any audible differences between violin timbre with different tailpieces attached and if the material of a tailpiece can affect the bridge FRF measurement and instrument sound.

2. Material and the method of bridge FRF measurements with different tailpieces attached

In order to objectively illustrate how the vibrations of tailpieces made of various materials affect the sound, the FRFs of the violin's bridge were measured. The bridge was hammered at a treble side in a direction perpendicular to the fingerboard, mimicking the excitation of the strings by the bow, while the response signal was measured with the accelerometer placed on the rear side of the bridge near the top left corner, with a measuring axis pointing towards the tailpiece (Fig. 3). As in (Fouilhé *et al.*, 2011; Fouilhé, Houssay, 2013) the measured function was accelerance (a/f). At this point, it should be noted that the acceleration function is not a standard in bridge dynamic measurements. The literature is richer in measurements of bridges mobility (v/f) as in (Jansson 1997; 2004), which can be executed by aligning the excitation axis parallel to the measurement axis. It is also worth noting that in several cases, the accelerometer or excitation point is placed in locations other than the side of the bridge, which leads to the measurement of different characteristics, for instance: (Boutin, Besnainou, 2008) or (Alonso Moral, Jansson, 1982).

According to (Minnaert, Vlam, 1937), the movements of the violin bridge during playing can be divided into: a) vibrations in the plane of the bridge, b) bending vibrations perpendicular to this plane, and c) torsional vibrations, also perpendicular to this plane (Fig. 2). Fouilhé *et al.* (2011) also indicate that modes 1, 2, and 4 of the tailpiece were found to be the most important. As could be seen in the visual representations from Stopanni's *et al.* website (nd) second and fourth modes of the tailpiece exhibit a pronounced vertical movement, which could affect the forward and backward displacement motion of the bridge. Let us assume that, the accelerometer was intentionally placed on the axis of this motion, perpendicular to the axis of excitation.

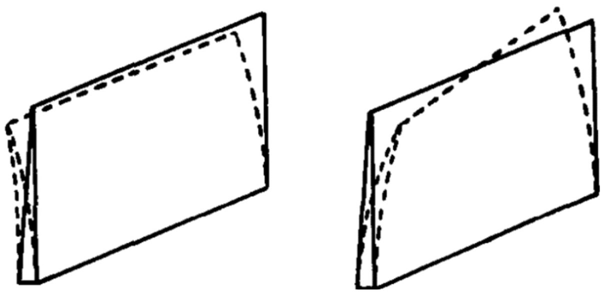

Fig. 2. Schematic representation of "flexural" (left) and "torsional" (right) deformation of the bridge (Minnaert, Vlam, 1937).

Five tailpieces, made of ebony, rosewood, boxwood, plastic (Wittner), and aluminium (Otto Infeld), were investigated (Fig. 3). Wooden ones were shaped in a "tulip" style. Their geometric parameters and weight are listed in Tables 3 and 4. Tailpieces made of wood had two attachable fine tuners on A and E string, although the plastic and aluminium ones had four built-in fine tuners. All of them were attached with a nylon tailgut to two instruments, which is further labelled as instrument A – based on the model of A. Stradivari – "Leonora Jackson" and instrument B – on the model of A. Stradivari "Dancla". Professional violin-makers made both instruments. In terms of the

Fig. 3. Tailpieces used in the study – from left: ebony, rosewood, boxwood, plastic, aluminium.

Table 3. Geometric parameters of the tested tailpieces.

Material	After-length [mm]		Length of a tailpiece [mm]
Instrument A and B	Instrument A	Instrument B	Instrument A and B
Ebony	55.0	54.0	111
Rosewood	54.5	54.5	114
Boxwood	54.0	54.0	112
Plastic	54.5	54.0	108
Aluminium	55.0	51.0	115

Table 4. Mass of the tested tailpieces.

Material	Tailpiece mass [g]	Fine tuner mass [g]	Tail gut mass [g]	Total [g]
Ebony	15.8	5.2 × 2	1.2	27.4
Rosewood	14.5	5.2 × 2	1.2	26.1
Boxwood	11.0	5.2 × 2	1.2	22.6
Plastic	19.3*		1.2	20.5
Aluminium	38.8*		1.2	40.0

* The tailpiece has four built-in fine tuners.

Table 5. Strings attached to the tested instruments.

	String G	String D	String A	String E
Instrument A	Evah Pirazzi Gold	Evah Pirazzi Gold	Chromcor	Evah Pirazzi Gold
Instrument B	Evah Pirazzi	Evah Pirazzi	Chromcor	Evah Pirazzi

material, model, arching, varnish, and strings, they were similar to each other. The violin plates, measuring 15.5 mm in height, were crafted using wood of the finest quality. Notably, material in both instruments exhibits exceptional properties, with spruce possessing a density lower than 0.35 g/cm^3 and a soundwave propagation velocity surpassing 5850 m/s, while maple demonstrates a density lower than 0.56 g/cm^3 and a soundwave propagation velocity greater than 5100 m/s, which according to (Bucur, 2006), is noticeably better than the usual wood used in violinmaking. The instruments were covered with spirit varnish and set up with Pirastro brand strings (Table 5). The only significant difference was the year of manufacture. Before the experiment, instrument B had been in use for two years, while instrument A had only been played for about a month.

The experimental tool used was an experimental modal analysis with a fixed response point and varied excitation point. The response signal was measured by the Ono Sokki accelerometer NP-2110, of 0.6 g in mass, attached with bee wax. The bridge was excited by an impact hammer with a piezoelectric force transducer (PCB Piezoelectronics Impact Hammer Model 086C05) (Fig. 4). The accelerometer and impact hammer were connected to the ONO SOKKI analyzer CF 5210. The modal parameters were calculated using the software packet SMS STAR Modal. Measurements were made for the frequencies 10–1600 Hz with a spectral resolution of 2 Hz. Ten spectral averages were used to improve the signal-to-noise ratio. Each measurement was controlled by the coherence function. The measured and analyzed function was the frequency response function module, similar to our previous works (Mania *et al.*, 2015; 2017; Mania, Skrodzka, 2020). After each tailpiece change, the violins were tuned.

Fig. 4. Position of the accelerometer and the impact hammer in the modal experiment.

Every effort has been made to ensure that the afterlength is 54.5 mm ±0.5 mm (Table 3). To prevent any influence from the vibrating strings on the experimental results, a cloth was carefully inserted between the strings and the fingerboard near the upper nut.

3. Measurement results and discussion

3.1. Instrument A

In Fig. 5, parts of FRFs registered for the instrument A are shown in frequency ranges of 240–300, 400–460, 460–520, and 540–600 Hz. The frequency ranges were selected so that each potentially contains one or two violin signature modes (Table 2).

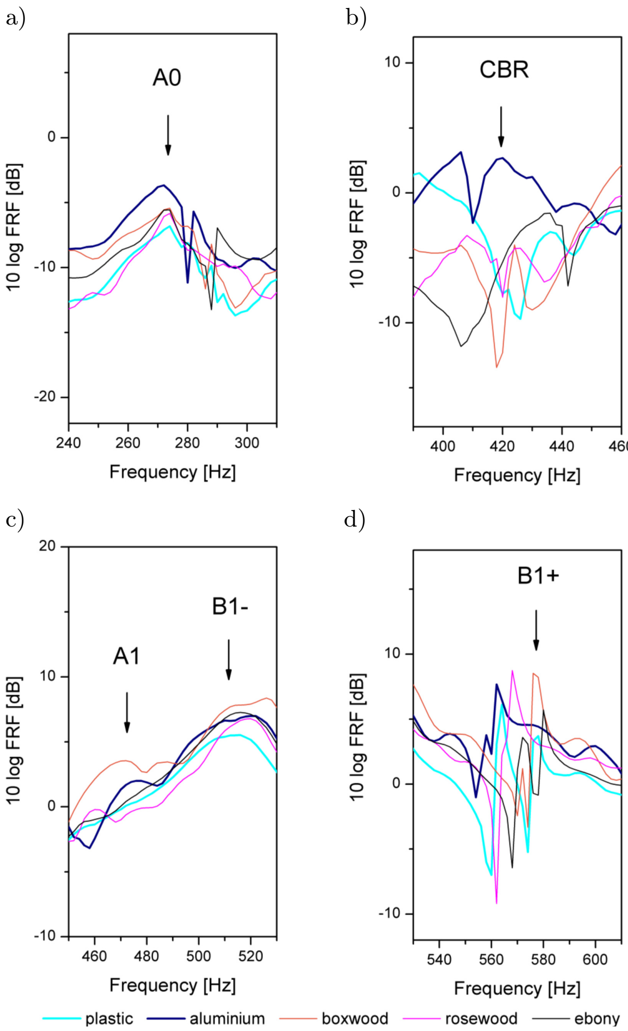

Fig. 5. FRFs for the instrument A in the range of:
a) 240–320 Hz mode A0; b) 380–460 Hz – mode CBR;
c) 450–530 Hz – modes A1 and B1−;
d) 530–610 Hz – mode B1+.

The most significant influence of the tailpiece material on the bridge FRF can be observed around modes A0 and B1+. In Fig. 5a, it can be seen that the highest FRF value for the A0 mode was observed for the aluminium tailpiece and the lowest for the plastic one. Although the frequency at which this mode occurs remained constant, the distinct peak around 285 Hz, which varied depending on the tailpiece used, is worth noting. A more significant effect of the tailpiece material can be observed in the frequency range of 530–610 Hz. From Fig. 5d, it can be deduced that the frequency of the maximum of the B1+ mode can be split, increased or decreased (Table 6). The lowest frequency of this mode was observed for the aluminium tailpiece and the highest for ebony one. It is difficult to interpret the effect of tailpiece material on the bridge FRF in the 400–530 Hz frequency range (Figs. 5b–c).

3.2. Instrument B

Figure 6 shows parts of the instrument B's bridge FRF divided into four frequency ranges, the same as

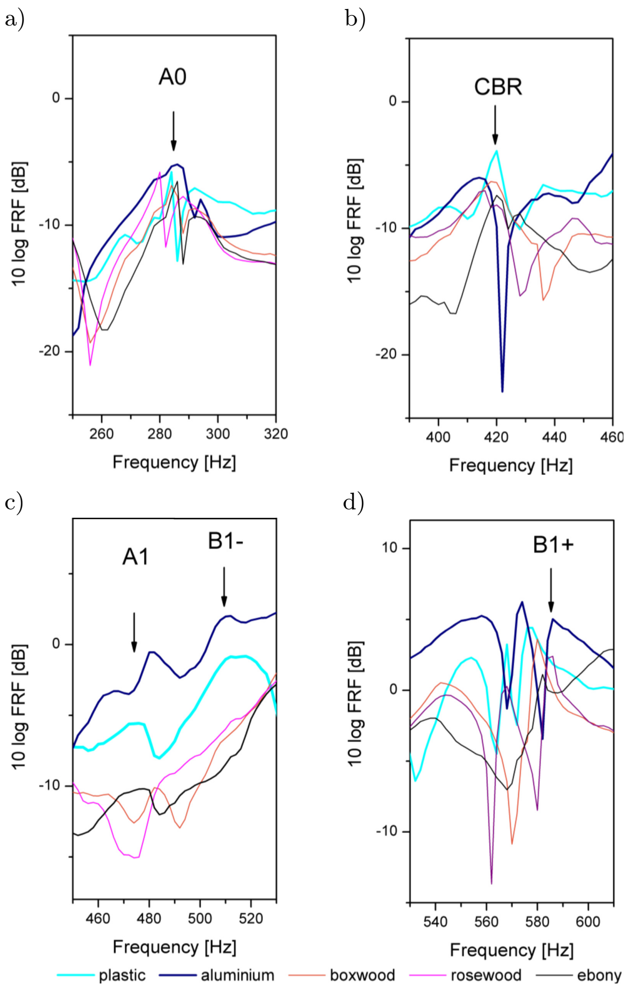

Fig. 6. FRFs for the instrument B in the range of: a) 240–320 Hz mode A0; b) 380–460 Hz – mode CBR; c) 450–530 Hz – modes A1 and B1−; d) 530–610 Hz – mode B1+.

for the instrument A. From Fig. 6a it can be derived that the frequency of A0 can be increased or decreased within a range of several Hertz depending on the tailpiece used. In the range of 440–530 Hz (Figs. 6b–c), it is noticeable that wooden tailpieces have a lower FRF value than those made of plastic or metal. Also worth highlighting is the characteristic minimum visible only for the aluminium tailpiece at 423 Hz in the direct vicinity of the CBR mode. From Fig. 6d, the B1+ mode is visible, which maximum was split for aluminium, plastic, and rosewood tailpieces. The frequency of this mode was lowest for the aluminium one and greatest for that made of rosewood. Depending on the tailpiece used, the frequency of the B1+ mode can be adjusted in the 12 Hz range.

3.3. Instruments A and B in a wide frequency range

Bridge FRFs registered for both instruments A and B are shown in Fig. 7 over a wide frequency range of 100–1000 Hz. It may be noted that the bridge FRF of instrument A slightly differs from that of instrument B. Despite these discrepancies, regularity can be observed in the form of similar effects of the individual tailpieces that appear in both instruments. The shapes of the bridge FRF graphs for ebony, rosewood and boxwood tailpieces are similar across the entire spectrum depicted in Fig. 7. Furthermore, tailpieces made of aluminium or plastic have higher FRF values up to 600 Hz than wooden ones. Also worth mentioning is the frequency range between the A0 and CBR modes – 320–400 Hz. It notes high FRF values for the plastic and metal tailpieces and a clear maximum around 380 Hz. It is also important to note the peak below 180 Hz for the plastic one, which repeats for both instruments A and B but is absent for the other tailpieces.

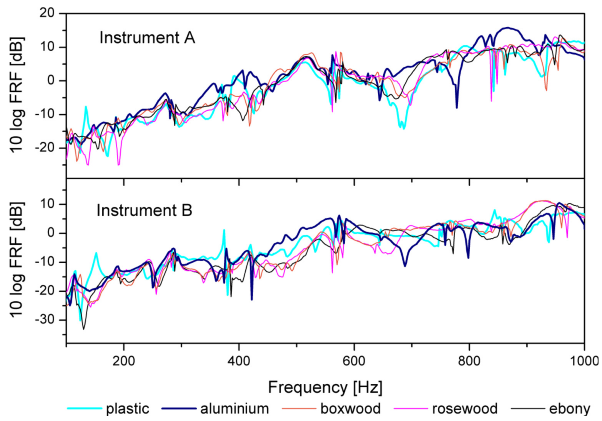

Fig. 7. FRFs showing amplitude changes for instrument A and B in a wide frequency range of 100–1000 Hz.

3.4. Interim summary

As seen from Figs. 5–7, the tailpiece material really has an impact on the frequencies of the bridge FRF.

Table 6. Modal frequency B1+ [Hz] for the material from which the tailpiece is built.

	Plastic	Aluminium	Boxwood	Rosewood	Ebony
Instrument A	564*	562	576	567	580*
Instrument B	576*	574*	580	586*	582

* Maximum clearly splits.

It affects frequencies in the 530–610 Hz range the most. In the case of the aluminium tailpiece, the frequency of B1+ mode has the smallest value of FRF in both instruments A and B. The modal frequency of the ebony tailpiece relative to the metal one was greater by an average of 13 Hz, while that of the rosewood was greater by 9 Hz, that made of the boxwood by 10 Hz, and that of the plastic by 2 Hz (Table 6). The frequency differences shown in Table 6 are generally greater than the measurement error, but they may not be solely due to the tailpiece replacement, and may be caused by some degree of tailpiece manipulation, which is unavoidable in this type of experiment. As reported by Torres *et al.* (2020) even after a very serious interference in the violin (removing the top plate and replacing it with another plate and after removing the top plate, drilling holes through the blocks of the soundbox, and then regluing the top) changes in mobility measured at both stages were of the same order of magnitude compared with the initial stage. In our case, the modification of the instrument was not so deep. It can therefore be assumed that our modifications consisting in replacing the tailpiece made a small, consistent, repeatable contribution to the results of the modal experiment. In addition, the replacement of some violin components, including the tailpiece, is a normal procedure for servicing the instrument. In both instruments, the plastic tailpiece caused the maximum of this mod to split. On the broadband spectrum graph, Fig. 7, it is hard not to notice the characteristic peak around 150 Hz for plastic, and around 380 Hz for plastic and metal tailpieces. It is also worth noting that non-wooden ones have generally high amplitude values of FRF below 600 Hz.

4. Results of subjective assessment of violin timbre with different tailpieces attached

Young musicians who attend music schools hone their aural skills in ear training classes. Development in this area is essential for the efficient and mindful performance of music. Scientists agree that sensitivity to differences in sound timbre differs between people who have never had much to do with music and professional musicians who do it professionally (Loebach *et al.*, 2010). From the point of view of neuroscience, along with learning how to play an instrument, numerous changes occur in the cerebral cortex, including areas related to motor coordination, memory, or feeling emotions (King, Nelken, 2009). The brain's ability to remodel neural connections, called neuroplasticity, is responsible for this phenomenon. According to researchers, auditory training leads to transformations in sound perception. Therefore, in order to describe as accurately and reliably as possible the changes in the timbre of instruments with attached tailpieces made of different materials, research has been conducted on a group of 40 qualified musicians and luthiers at the student and professional levels. The group included 13 violinists, 3 violists, 4 cellists, 1 double bass player, 12 luthiers, and 7 musicians who do not play any stringed instrument.

4.1. Method

The 10 recordings of an excerpt from Tchaikovsky's Violin Concerto in D Major, Op. 35 were recorded under concert conditions (in a large hall). Each of them was conducted under the same conditions. Only the tailpiece used was changed. The study used the same two instruments and the same five tailpieces used in the experiment described in Sec. 2 of the paper. The recordings were made using an Audio Technica AT2035 condenser microphone, which was at a distance of about 1.5 meters from a professional violinist. An anonymous test was then conducted on a 40-person study group. Respondents were asked to rate the recordings on a scale of 1 to 5 in terms of brightness, openness, thickness, and overall quality. For example, a "1" indicated that the excerpt was the brightest among the others, a "5" the darkest, and "3" that it was moderate. Respondents were required to point out extreme recordings and rate at least one tailpiece as a "1" and the other as a "5".

The study was conducted on one person at a time and under the same conditions. Silence and professional monitor headphones (Beyerdynamic DT990 PRO) were provided. Each respondent had unlimited time to respond and was free to compare portions of the recording, which were synchronized in a digital audio workstation software. A single survey lasted between 20 and 50 minutes.

4.2. Results and discussion of the survey

Statistical analysis of the survey responses was performed for each of the four rating categories, i.e., brightness, openness, thickness, and overall quality. The factors in the analysis were: instrument (A, B), subject (1–40), material (plastic, aluminium, boxwood, rosewood, ebony). A significance level of $p = 0.05$ was assumed.

The results of the statistical analysis showed that for each rating category, both subject and instrument were not significantly statistical factors. For the dark/bright evaluation, material was found to be the statistically significant factor $F(4{,}394) = 4.55$ for $p = 0.02$. A post-hoc test (Tukey) showed that there were statistically significant differences in evaluation between metal and boxwood, ebony and plastic, metal and plastic, and rosewood and plastic. The lowest mean score was obtained for metal, the highest for plastic (Fig. 8).

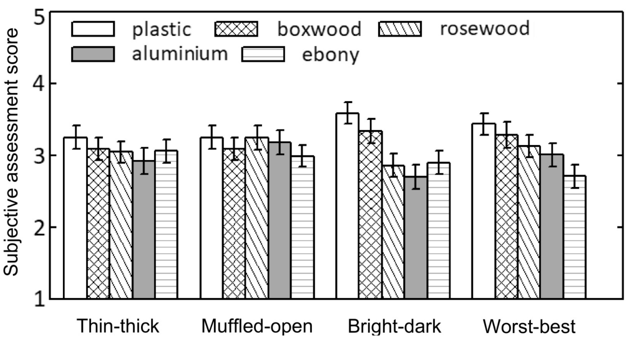

Fig. 8. Average survey scores, with standard errors indicated, for each material type in the four rating categories.

For the overall best/worst rating, the material was again found to be a significant factor $F(4{,}394) = 3.166$ for $p = 0.014$. Post hoc tests showed that a significantly statistical difference occurred between ebony and plastic. Ebony was found to be the worst material, while plastic was found to be the best. According to most participants, the survey was challenging and required a lot of concentration. The differences between the recordings were noticeable but very subtle. The complicated nature of violin performance was also an issue. It is difficult to perform the same one-minute piece of music identically 10 times. The sound of a violin depends enormously on how the bow is guided along the string and the pressure of the fingers of the left hand. An additional complication was the fact that the instrument's timbre varied from string to string. In addition, it is important to remember that abstract concepts, such as the thickness or darkness of sound, are interpreted differently by different people.

5. Conclusions

We conclude that:

- a tailpiece that is properly matched to a specific instrument can improve its sound;
- the lighter the tailpiece, the darker the sound of the violin and vice versa;
- sound of a violin with plastic or boxwood tailpieces over one with the ebony tailpiece was generally preferred.

Acknowledgments

The authors would like to thank the Ignacy Jan Paderewski Academy of Music in Poznań for providing the apparatus needed for the study. Many thanks also to prof. Andrzej Łapiński for his help editing this article, his good word, and his daily support.

References

1. Alonso Moral J., Jansson E.V. (1982), *Input admittance, eigenmodes and quality of violins*, Report STL-QPSR, www.speech.kth.se/prod/publications/files/qpsr/1982/1982_23_2-3_060-075.pdf (access: 3.08.2023).
2. Borman T., Stoppani G. (nd), *Modal Animations. Borman Violins*, www.bormanviolins.com/modalanalysis.html (access: 8.01.2023).
3. Boutin H., Besnainou C. (2008), Physical parameters of the violin bridge changed by active control, *Acoustics'08*, Paris.
4. Bucur V. (2006), *Acoustics of Wood*, 2nd ed., Springer Berlin, Heidelberg.
5. Bucur V. (2016), *Handbook of Materials for String Musical Instruments*, Springer Cham.
6. Folland D. (2010), *How Tailpiece Can Affect Sound and Playablility. David Folland Violins*, www.follandviolins.com/articles/tailpiece/ (access: 8.01.2023).
7. Fouilhé E., Goli G., Houssay A., Stoppani G. (2009), *Preliminary Study on the Vibrational Behaviour of Tailpieces in Stringed Instruments – COSTIE0601 STSM Results*, www.researchgate.net/publication/267560695_Preliminary_study_on_the_vibrational_behaviour_of_tailpieces_in_stringed_instruments_-_COSTIE0601_STSM_results (access: 8.01.2023).
8. Fouilhé E., Goli G., Houssay A., Stoppani G. (2010), The cello tailpiece: How it affects the sound and response of the instrument, *Proceedings of the Second Vienna Talk*, pp. 63–67.
9. Fouilhé E., Goli G., Houssay A., Stoppani G. (2011), Vibration modes of the cello tailpiece, *Archives of Acoustics*, **36**(4): 713–726, doi: 10.2478/v10168-011-0048-2.
10. Fouilhé E., Houssay A. (2013), String "after-length" and the cello tailpiece: Acoustics and perception, [in:] *Proceedings of the Stockholm Music Acoustics Conference 2013, SMAC 2013*, pp. 60–65.
11. Gourc E., Vergez C., Mattei P.-O., Missoum S. (2022), Nonlinear dynamics of the wolf tone production, *Journal of Sound and Vibration*, **516**: 116463, doi: 10.1016/j.jsv.2021.116463.
12. Houssay A. (2014), The string "after-length" of the cello tailpiece: History, acoustics and performance techniques, [in:] *Proceedings of the International Symposium on Musical Acoustics*, pp. 207–213.
13. Hutchins C.M. (1983), A history of violin research, *The Journal of the Acoustical Society of America*, **73**(5): 1421–1440, doi: 10.1121/1.389430.

14. Hutchins C.M. (1993), The effect of relating the tailpiece frequency to that of other violin modes, *Catgut Acoustical Society Journal*, **2**(3): 5–8.

15. Jansson E.V. (1997), Admittance measurements of 25 high quality violins, *Acustica – Acta Acustica*, **83**: 337–341.

16. Jansson E.V. (2004), Violin frequency response – bridge mobility and bridge feet distance, *Applied Acoustics*, **65**: 1197–1205, doi: 10.1016/j.apacoust.2004.04.007.

17. King A., Nelken I. (2009), Unraveling the principles of auditory cortical processing: can we learn from the visual system?, *Nature Neuroscience*, **12**: 698–701, doi: 10.1038/nn.2308.

18. Leung J. (2016), *Resonant effects of the violin tailpiece*, Msc. Thesis, McGill University.

19. Loebach J.L., Conway C.M., Pisoni D.B. (2010), Audition: cognitive influences, [in:] *Encyclopedia of Perception*, Goldstein B. [Ed.], pp. 138–141, SAGE Publications, Inc.

20. Mania P., Fabisiak E., Skrodzka E. (2015), Differences in the modal and structural parameters of resonance and non-resonance wood of spruce (Picea abies L.), *Acta Physica Polonica A*, **127**(1): 110–113, doi: 10.12693/APhysPolA.127.110.

21. Mania P., Fabisiak E., Skrodzka E. (2017), Investigation of modal behaviour of resonance spruce wood samples (Picea abies L.), *Archives of Acoustics*, **42**(1): 23–28, doi: 10.1515/aoa-2017-0003.

22. Mania P., Skrodzka E. (2020), Modal parameters of resonant spruce wood (Picea abies L.) after thermal treatment, *Journal of King Saud University – Science*, **32**(1): 1152–1156, doi: 10.1016/j.jksus.2019.11.007.

23. Minnaert M.G.J., Vlam C.C. (1937), The vibrations of the violin bridge, Physica, **4**(5): 361–372, doi: 10.1016/S0031-8914(37)80138-X.

24. Pollens S. (2009), Some misconceptions about the Baroque violin, *Performance Practice Review*, **14**(1): 6, doi: 10.5642/perfpr.200914.01.06.

25. Skrodzka E., Linde B.B.J, Krupa A. (2014), Effect of bass bar tension on modal parameters of violin's top plate, *Archives of Acoustics*, **39**(1): 145–149, doi: 10.2478/aoa-2014-0015.

26. Skrodzka E.B., Linde B.B.J., Krupa A. (2013), Modal parameters of two violins with different varnish layers and subjective evaluation of their sound quality, *Archives of Acoustics*, **38**(1): 75–81, doi: 10.2478/aoa-2013-0009.

27. Stoppani G., Zygmuntowicz S., Bissinger G. (nd), *The Signatures Modes*, www.strad3d.org/st_4.html (access: 8.01.2023).

28. Stough B. (1996), The lower violin tailpiece resonances, *Catgut Acoustical Society Journal*, **3**(1): 17–24.

29. Torres J.A., Soto C.A., Torres-Torres D. (2020), Exploring design variations of the Titian Stradivari violin using a finite element model, *The Journal of the Acoustical Society of America*, **148**(3): 1496–1506, doi: 10.1121/10.0001952.

30. Zhang A., Woodhouse J. (2018), Playability of the wolf note of bowed string instruments, *The Journal of the Acoustical Society of America*, **144**(5): 2852–2858, doi: 10.1121/1.5079317.

Archives of Acoustics Vol. 49, No. 1, pp. 19–26 (2024), doi: 10.24425/aoa.2024.148765

Research Paper

Auditory Spatial Illusion – A Psychoacoustic Study in Young Adults

Mehri MALEKI[(1)], Mohsen AHADI[(1)*], Amirsalar JAFARPISHEH[(2)], Amin ASGHARZADEH[(3)], Alireza Akbarzadeh BAGHBAN[(4)], Reyhane TOUFAN[(1)]

[(1)] *Rehabilitation Research Center, Department of Audiology*
School of Rehabilitation Sciences
Iran University of Medical Sciences
Tehran, Iran

[(2)] *University of Social Welfare and Rehabilitation Sciences*
Tehran, Iran

[(3)] *Department of Medical Physics, School of Medicine*
Iran University of Medical Sciences
Tehran, Iran

[(4)] *Proteomics Research Center, Department of Biostatistics*
School of Allied Medical Sciences
Shahid Beheshti University of Medical Sciences
Tehran, Iran

*Corresponding Author e-mail: ahadi.m@iums.ac.ir

(*received April 16, 2023; accepted August 17, 2023; published online December 13, 2023*)

The Franssen illusion, or Franssen effect (FE), is one of the auditory spatial illusions. Few studies have explored the FE, and the mechanisms underlying it remain unknown. The present study was conducted to clarify the FE occurrence with different tasks and presentation modes in young adults. It also sought to investigate possible neurophysiological similarities between interaural time difference (ITD) cue processing and FE perception. FE perception was evaluated using two different tasks and two presentation modes (i.e., insert phones and loudspeakers). Sound reflections (reverberation) were presented in the diffuse field (loudspeaker mode). ITD performance was investigated using different stimuli delivered via insert phones. No significant difference between the two FE perception tasks was found ($F_{1,25} = 0.138$, $p = 0.713$). However, the FE perception showed a significant difference between the two presentation modes ($F_{1,25} = 434.03$, $p < 0.001$). Spearman's correlation did not reveal a significant relationship between FE perception and ITD scores ($p > 0.05$).

The current findings show the importance of reverberation in the FE occurrence. Also, the non-significant correlation between the results of the behavioral binaural temporal resolution test and FE perception in young people with normal temporal resolution may indicate that room reflections (reverberation) complicate the ability to process ITDs (rather than poor ITD processing for the "steady state" portion of signal).

Keywords: auditory spatial illusions; the Franssen effect; interaural time difference; binaural temporal resolution.

1. Introduction

Illusions are valuable tools for studying perception mechanisms since one can relate neural responses to both the physical aspects of the stimuli and subject's answers (Rajala *et al.*, 2013). In contrast to visual illusions (e.g., change-blindness illusion (O'Regan *et al.*, 1999)) and multisensory illusions (such as kappa and tau effects (Schroeger *et al.*, 2022)), which are well-known and extensively studied, there is limited understanding of spatial auditory illusions and the mechanisms underlying them (Rajala *et al.*, 2013). As the spatial representation of auditory signals does not occur at the receptor level (Grothe *et al.*, 2010), this may make designing experiments a challenge. The Franssen illusion, or Franssen effect (FE), is one of the auditory spatial illusions named after Nico Valentinus Franssen, who found and presented this phenomenon in 1960 in his Ph.D. dissertation. This illusion occurs when a sound is divided into two components: a transient onset – (first few milliseconds) characterized by a smooth rising, and a sustained tone. These components are simultaneously presented from different loudspeakers (Franssen, 1960). Indeed, in the Franssen illusion perception, the onset predominates over other cues (Franssen, 1960; Higgins *et al.*, 2017), and this domination of the onset cue continues throughout the entire sound perception process. Therefore, the entire sound is perceived as originating from the onset loudspeaker, even though it is the other loudspeaker that delivers nearly all of the sound energy (Franssen, 1960). The perception of this illusion may last for tens of seconds (Rajala *et al.*, 2013).

A previous study proposed a possible relationship between the FE and a more widely known phenomenon, i.e., the precedence effect (Yagcioglu, Ungan, 2006). Due to the precedence effect, similar sounds are localized by directional cues carried in the first-arriving sound, resulting in their perception as a single auditory event (fusion) (Wallach *et al.*, 1949). Litovsky and Godar (2010) introduced several different psychophysical tasks to investigate different aspects of the precedence effect, including discrimination-based task, fusion, and localization. Notably, studies on the impacts of these tasks on precedence effect assessments yielded inconsistent results. Previous studies (Donovan *et al.*, 2012; Saberi, 1996; Seeber, Hafter, 2011) showed that the temporal (fusion) and spatial (localization/discrimination) aspects of the precedence effect probably have different mechanisms. On the other hand, Seeber and Hafter (2011) investigated fusion and both lead and lag localization using virtual auditory space stimuli across different conditions. They concluded that the fusion and localization aspects of the precedence effect share similar time courses. However, until now, the effect of various tasks on FE perception has not been investigated.

In the Franssen illusion perception, stream separation based on spatial cues does not occur as intended (Van Deun *et al.*, 2009), which may be related to incorrect processing of ITD and interaural intensity difference (IID) cues, required for sound localization, particularly due to sound reflections and reverberation (Hartmann, Rakerd, 1989; Stevens, Newman, 1936). In sound localization, ITD holds greater significance (Babkoff *et al.*, 2002). ITD cues can be divided into ongoing envelope, fine structure, and transient onset (Haqqee *et al.*, 2021; Verschooten *et al.*, 2019). In pure tone signals, ITD is processed up to 1500 Hz, referred to as fine-structure ITD ($\mathrm{ITD_{FS}}$) (Brughera *et al.*, 2013; Delphi *et al.*, 2017; Verschooten *et al.*, 2019). In complex sounds, ITD at higher frequencies (beyond 1500 Hz) is perceived through the difference in the sound's envelope timing, known as envelope ITD ($\mathrm{ITD_{ENV}}$) (Yost, 2017); and transient onset ITD cues, which refer to the difference in the arrival time of the sound to the perceiver's different ears ($\mathrm{ITD_{ONSET}}$) (Scharf *et al.*, 1976).

Yost and Zhong (2014) investigated sound localization by stimuli with a bandwidth ranging from 1/20 to two octaves. Their findings revealed that sound localization accuracy in noise burst stimuli is higher than that in pure tone. Soeta and Nakagawa (2007) conducted a study on binaural hearing filters. Their results showed that N1m amplitudes increase with the increasing frequency separation to above 200 Hz. Hafter *et al.* (1979) indicated that, for tones with extended durations, onsets and offsets were unnecessary for detecting ITD or IID with earphones. On the other hand, Hartmann and Rakerd (1989) showed that the sustained component of a stimulus does not provide valuable localization information in a room. Thus, they introduced the plausibility hypothesis. In this hypothesis, reflections make ITDs unreliable as cues for localizing steady-state stimuli (Hartmann, Rakerd, 1989). Previous studies showed that the Franssen illusion exists in live (reverberant) rooms (Hartmann, Rakerd, 1989; Yost *et al.*, 1997), and it is particularly strong for midfrequency tones (near 1500 Hz) (Yost *et al.*, 1997). With regard to previous studies on onset dominance, questions arise whether the presentation of FE stimuli with earphones causes FE perception, and whether there is any similarity in the underlying neurophysiological mechanism between FE and ITD cues processing.

The present study was designed to clarify the Franssen illusion occurrence in young adults. For this purpose, the perception of the FE was evaluated through two different tasks conducted under the insert phones mode and the diffuse field mode within a typical room (in which reverberation times ranged from ~0.2 to roughly 1.3 seconds (Gelfand, 2016)). Since the abnormal ITD cues are implicated as the cause of the FE occurrence, in line with the plausibility hypothesis, the correlation of the illusion with ITD scores (around frequency 1500 Hz where all ITD cues are weak and do not overcome each other) was investigated using tone burst and noise bursts.

2. Material and method

2.1. Participants

A total of 26 young adults aged between 19 and 32 years old (mean = 23.30 ± 3.18) participated in the research. They had no history of ear and hearing problems, head trauma, or neurologic disease. Audiometric thresholds were equal to or less than 10 dB hearing level (HL) at octave frequencies ranging between 250 to 8000 Hz. Interaural audiometric threshold asymmetry was usually less or equal to 5 dB at all frequencies tested. The average gap in noise threshold was obtained at 4.41 ms in the right ear and 4.54 ms in the left ear. The experiment was administered in accordance with the ethical guidelines of the Iran University of Medical Sciences (ethics code: IR.IUMS.REC.1400.1099).

2.2. Stimuli, experimental design, and procedure

Experiment 1: The Franssen illusion investigation with identification/discrimination tasks at 1500 Hz tone burst stimuli using insert phones and diffuse field modes.

Two types of stimuli (generated digitally in MATLAB) were used. The Franssen stimuli (including a transient component (total duration was 50 ms), a sustained (steady state) component with a 50 ms linear onset, 100 ms linear offset, and a 350 ms plateau), and a single sound non-Franssen stimulus. Both stimuli had a total duration of 500 ms. The stimuli were presented with two loudspeakers (Pejvak Ava Corporation, Tehran, Iran) and ER-2 insert earphones (Etymotic Research, Inc., El Grove Village, IL, United States) to subjects seating in a typical room (reverberation time (RT) = 0.827 s, estimated through Sabine's reverberation equation (Fig. 1)). The ER-2 insert earphones were calibrated at 1000 Hz without any frequency-dependent correction. The level of stimuli presentation was at 70 dB sound pressure level.

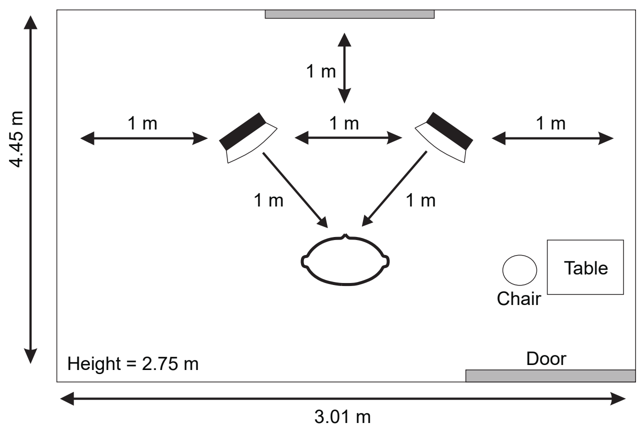

Fig. 1. Sketch of the test room and the relative positions and orientations of the speakers and the listener. Dimensions of the room: length – 4.45 m, width – 3.01 m, height – 2.75 m, positions of the speakers: height from the floor – 1 m, distance from the side and back walls – 1 m, facing of loudspeakers – medially 45°; height of the listener's ears from the floor – 1 m, and reverberation time with Sabine's equation (RT) = 0.827 s.

Procedure for the Franssen effect identification task: stimuli were presented alternately, and in each trial, the participants were asked "which speaker/insert phones presented the stimuli? right, left, or two sounds from both speakers" and the participant identified the direction of presentation (Fig. 2). A block of 50 trials was used to estimate the Franssen illusion for each mode (under insert phones and loudspeakers). The number of Franssen and non-Franssen stimuli were equal, and their distribution was random. Finally, the number of illusions was calculated and reported in terms of total Franssen stimuli numbers (i.e., 25).

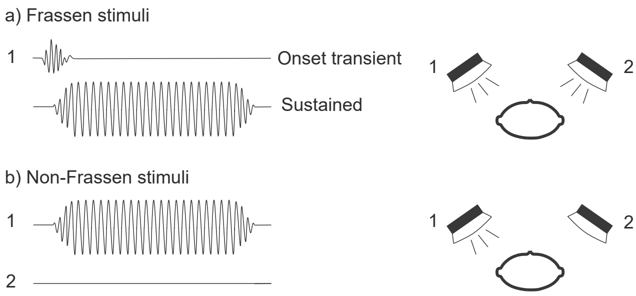

Fig. 2. Identification task procedure in diffuse field: a) in the Franssen stimuli, the onset transient was presented from one loudspeaker (1) and quickly cross-faded to a second loudspeaker (2); b) in the non-Franssen stimuli, the whole tone was presented from one loudspeaker (1), and the other loudspeaker was off (2).

Procedure for the Franssen-effect discrimination task: a block of 48 trials was used to estimate the Franssen illusion perception for each mode (under insert phones and loudspeakers). In each block, experimental and catch trials were presented randomly. In each trial, pairs of stimuli were presented. For example, in the Franssen right and non-Franssen right experiment, the onset of the Franssen stimuli was presented to the right earphone/speaker, and the sustained part of the Franssen stimuli was presented to the left earphone/speaker. Then, the second stimuli (non-Franssen right) were presented to the right earphone/speaker. The participant was asked to determine whether the stimuli were the same or different (Fig. 3). A "same" response indicated that both stimuli presented in each trial were similar in terms of perceived spatial location. The distribution of stimuli was random. To ensure that participants attended to the requested task, 16 catch trials (~30% of all stimuli) were used in which pairs of non-Franssen stimuli were presented in the same or different directions. Finally, the number of illusions (a condition in which a person does not differentiate between the Franssen and non-Franssen stimulus pairs) was calculated and reported in total real trials (i.e., 32).

Experiment 2: ITD assessment at 1500 Hz tone burst and as a function of bandwidth (1/200 to 1/2 octaves wide).

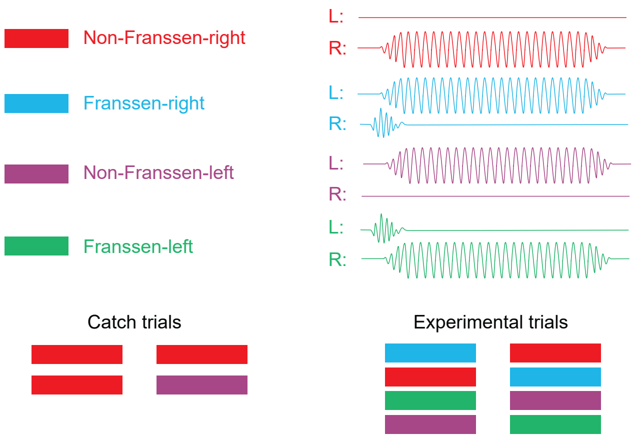

Fig. 3. Discrimination task procedure: in catch trials, pairs of the same (two non-Franssen stimuli from the right loudspeaker) (red) or different (two non-Franssen stimuli from the right and left loudspeakers) (red, purple) stimuli were presented randomly; in experimental trials, pairs of Franssen and non-Franssen stimuli were presented.

Stimuli were generated digitally in MATLAB (including 1500 Hz tone burst and filtered noises of various bandwidths (1/200, 1/100, 1/50, 1/20, and 1/2 octaves wide), filtered with a six-pole elliptic filter with passband ripple 3 dB and stopband ripple 20 dB, geometrically centered at 1500 Hz). Two durations were used: 250 ms and 20 ms rise and fall time to investigate ITD. In ITD tests, stimuli were presented binaurally with ER-2 insert earphones (Etymotic Research, Inc., El Grove Village, IL, United States), with delay times of 700, 300, 100, 0, −100, −300, and 2700 μs between the two ears at an intensity level of 70 dB SPL. Therefore, the sounds could be perceived to be coming from seven different positions in a semicircle. To understand each position, a pair of stimuli was presented with an interval of 500 ms. In stimulus pairs, the first signal (standard signal) always shows the midline position, and the second signal (test signal) indicates the position that the person must understand and point to its position. The participant was asked not to respond to the first stimuli and determine the position of the second stimuli orally after each stimulus, which was presented randomly. Each position was evaluated twice, and errors (i.e., wrong answers) were calculated. Hence, the maximum number of true or wrong answers was 14. All participants were trained before the main test, and after becoming familiar with the response method, they proceeded with the main test.

3. Results

Table 1 shows the minimum, maximum, mean, and standard deviation of the number of errors in experiment 1 for each of the tasks and modes. In the present study, a two-way repeated measures ANOVA did not reveal a significant difference in the Franssen illusion perception between the two different tasks ($F_{1,25}$ = 0.138, p = 0.713). The total number of Franssen stimuli presented in each procedure/task were converted into a ratio and then compared due to different stimuli numbers. However, the Franssen illusion perception showed a significant difference with two different presentation modes, under insert phones and diffuse field ($F_{1,25}$ = 434.03, $p < 0.001$). Also, the interaction of the two factors was not significant ($F_{1,25}$ = 2.609, p = 0.119). The minimum, maximum, mean, and standard deviation of the number of errors for each of the stimuli used in experiment two are given separately in Table 2.

Statistical analysis of the ITD score was performed between stimuli. Results of the pairwise comparison with the Bonferroni post hoc test between tone burst stimuli (1500 Hz) and noise burst stimuli with

Table 1. Statistics of the Franssen illusion perception test results with insert phones and diffuse field modes.

Franssen illusion perception tasks	Presentation modes	Participants [n]	Number of errors			
			Minimum	Maximum	Mean	Standard deviation
Identification task	Diffuse field	26	11.00	25.00	21.11	4.79
	Insert phones	26	00.00	13.00	4.34	4.70
Discrimination task	Diffuse field	26	13.00	32.00	28.03	6.15
	Insert phones	26	00.00	20.00	4.00	5.34

Table 2. Statistics of the ITDs test results at 1500 Hz and as a function of bandwidth (1/200 to 1/2 octaves wide).

Types of stimuli	Participants [n]	Number of errors			
		Minimum	Maximum	Mean	Standard deviation
Tone burst 1500 Hz	26	4.00	13.00	9.61	2.26
Noise burst with 1/200 octave wide	26	00.00	8.00	2.73	2.30
Noise burst with 1/100 octave wide	26	00.00	7.00	3.11	2.02
Noise burst with 1/50 octave wide	26	00.00	9.00	3.23	2.15
Noise burst with 1/20 octave wide	26	00.00	6.00	3.19	1.67
Noise burst with 1/2 octave wide	26	00.00	3.00	1.07	0.93

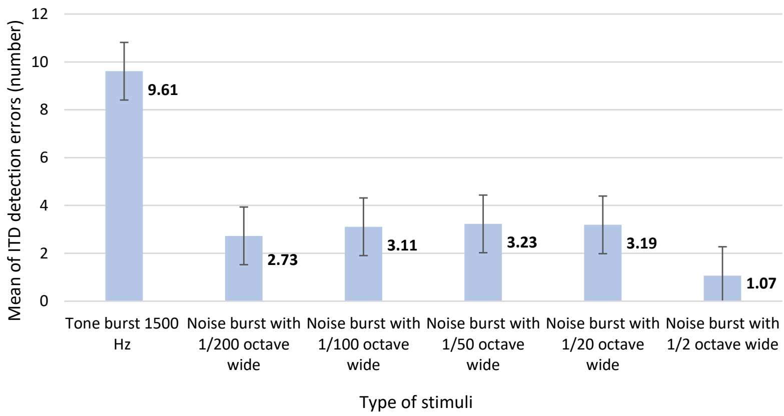

Fig. 4. ITD scores for the six different stimulus waveforms using a stimulus level of 70 dB SPL. The rectangular bars and the error bars indicate the mean and standard deviations of the data across the twenty-six participants.

bandwidths of one-half to 1/200 octaves showed a significant difference. The one-half octave wide noise burst differed from other noise bursts as well (Fig. 4). Additionally, Spearman's correlation did not show a significant relationship between the Franssen illusion perception and ITD scores ($p > 0.05$).

4. Discussion

In the present study, Franssen illusion perception was evaluated using two different tasks (the sound source identification task/the Franssen discrimination task) to investigate the impact of different task types on Franssen illusion outcomes. Statistical analysis did not show any significant differences between the results of these two task types, which is in accordance with the study of Seeber and Hafter (2011). The average error number for the identification task in diffuse field mode was 84.44%, and for the discrimination task with a similar presentation method, it was 87.59%. This shows the similar performance of these two procedures in the Franssen illusion investigation. On the other hand, prior studies (Donovan *et al.*, 2012; Saberi, 1996; Seeber, Hafter, 2011) indicated that temporal (fusion) and spatial (localization/discrimination) aspects of the precedence effect probably have different mechanisms. This discrepancy may be due to binaural fusion dependence on the task. Also, interpreting the percentage of fused responses is complicated (Suneel *et al.*, 2017). In general, both tasks appear to have the same ability in investigating the occurrence rate of the Franssen illusion, and the probability of illusion occurrence in a typical room is almost the same for both types of procedures. Fusion and spatial mechanisms probably have the same contribution to the occurrence of FE. This seems reasonable, considering that fusion stimuli perception correlates with localization performance (Suneel *et al.*, 2017).

The perception of the FE with both tasks in diffuse field and insert phones mode exhibited a significant difference. In other words, the average Franssen illusion (i.e., the average percentage of errors (number) in both procedures/tasks) decreased from 86% in the diffuse field mode to 14.93% with insert phones. These errors arise because individuals tend to localize the stimulus using the onset cue and have difficulty in identifying the location of the sustained part of stimuli. This observed result is consistent with previous studies (Hafter *et al.*, 1979; Hartmann, Rakerd, 1989; Yost *et al.*, 1997). Hartmann and Rakerd (1989), and Yost *et al.* (1997) investigated the FE in the diffuse field mode. They showed the occurrence of the Franssen illusion in a reverberant room, and observed a reduction in the Franssen illusion when it was performed in a dead room. It seems that in typical listening environments, the ITD cue in identifying the sustained part of the Franssen stimulus becomes abnormal due to reverberation.

Hafter *et al.* (1979), in a study on stimulus onset, demonstrated that people could lateralize stimuli in the absence of abrupt onset, which can be a justification for reducing the occurrence of illusions and the non-dominancy of onset in the insert phones condition (a state in which there were no environment reflections) in the present study. It seems that the onset predominance or, in other words, interaural onset difference leads to the Franssen illusion perception in the diffuse field mode. However, this onset dominance is lost in the insert phones mode, and people can distinguish the onset and sustained parts individually. Studies on the role of onset dominance in the precedence effect occurrence showed that if the delay between onset and post-onset pulses is less than 5 ms, participant's spatial judgments are dominated by onset cues (lead). However, if the inter-pulse interval is 12 ms, it causes the same sensitivity to the onset "leading" and post-inset pulses "lagging" in spatial judgment (i.e., failure in precedence effect) (Brown *et al.*, 2015; Saberi, 1996). In the present study, despite the sustained part of the stimulus reaching its maximum value after 50 ms and continuing up to 500 ms, we still have the predominance of onset in diffuse field presentation, which may be due to the gradual and progressive onset of the Franssen stimuli or the different origins of these two effects (i.e., precedence effect and the Franssen illusion).

The statistical analysis showed a significant difference between the ITD score with tone burst 1500 Hz stimulus and noise burst stimuli (including bandwidths of 1/200 to 1/2 octave) so that the average number of errors decreased from 9.61 to 1.07, which is in accordance with previous studies (Pierce, 1901; Stevens, Newman, 1936; Soeta, Nakagawa, 2007; Yost, Zhong, 2014). Yost and Zhong (2014) investigated the accuracy of sound localization using stimuli with a bandwidth of 1/20 to two octaves and center frequency of 250, 2000, and 4000 Hz. The study's findings showed that sound localization accuracy in noise burst stimuli is higher than in pure tone. This suggests that modulations or oscillations around the center frequency over time help participants to use the difference in arrival time of the stimulus envelope (ITD_{ENV}) and ITDFS for sound lateralization.

A comparison of the ITD scores between the noise burst stimulus with a 1/2 octave bandwidth and other noise burst stimuli showed a significant difference. However, it did not show any significant difference between 1/200 to 1/20 octave bands, probably due to the stimulation of the same auditory filters. Soeta and Nakagawa (2007) performed a study on binaural hearing filters with auditory-induced magnetic fields. Their results showed that N1m amplitudes remain roughly constant if frequency separation is below 100 Hz, but they increase with increasing frequency separation to above 200 Hz. Given that only the frequency bandwidth changes in stimuli with different octave bands, Soeta and Nakagawa's findings (2007) can justify the present study's results. Based on their study results, ITD discrimination improves with an increase in bandwidth, even as small as 1/200 octaves. However, when the difference in frequency separation between two stimuli is such that they are placed in the same physiological filter, it will not create a significant difference in the listener's performance in ITD discrimination.

This study showed no significant correlation between ITD scores and the Franssen illusion occurrence. This finding can be explained in several ways. First, the probable site of the mechanism leading to the Franssen illusion remains ambiguous and controversial. Higgins *et al.* (2017) concluded that the perceptual representation of auditory space occurs at the higher level of the auditory cortex, which gives the basis for the formation of the FE. Yagcioglu and Ungan (2006) introduced the primary auditory cortex as a possible site for the FE mechanism. On the other hand, Rajala *et al.* (2013) showed a considerable correlation between the neural activity of the inferior colliculus in rhesus monkeys and behavioral responses to Franssen stimuli, indicating a possible subcortical origin of the Franssen illusion. Likewise, Haqqee *et al.* (2021) recently indicated the high sensitivity of the inferior colliculus to interaural onset difference in bats, which is remarkable considering the dominance of onset in perceiving the Franssen illusion. Therefore, if the FE is assumed to be a perceptual illusion, the absence of a significant correlation between the illusion perception and ITD, which has a subcortical origin and is considered a bottom-up process, is justifiable. Secondly, the different sensitivity of mammal's auditory system to onset and envelope cues (Haqqee *et al.*, 2021) may lead to the illusion occurrence in diffuse field conditions. The participants in this study were young adults with normal monoaural and binaural temporal resolution and normal ITD processing. It seems that the possible role of abnormal ITD cues in Franssen illusion perception is not due to a deficiency in neural processing of ITD but rather to the reverberation and its interference with ITD perception.

The current study demonstrated that Franssen illusion perception occurs in a typical room and is reduced in the insert phones condition. Moreover, the task variety does not affect the results, and it seems that the illusion occurrence does not depend on behavioral ITD discrimination. It is important to note that this research was conducted on listeners with a normal temporal resolution, fixed ITD (i.e., 100, 300, and 700 μs), and a constant center frequency. Performing a study on people with abnormal monaural and binaural temporal resolution, such as the elderly, across the entire human frequency range, and employing different ITDs in diffuse field conditions, could provide more information about the importance and role of ITD cues in the Franssen illusion perception. On the other hand, correlation is a type of statistical analysis that depends on individual participants' data rather than overall average, and thus a significant correlation may be observed with a large sample size. Based on the present study results and considering the onset dominance in the diffuse field mode, it may be possible to use the FE in examining the tone onset time (TOT), which is vital in understanding plosive (stop) consonants (Pisoni, 1977). Furthermore, the study suggests that the effect of age, hearing loss, and training on the Franssen illusion may be investigated with behavioral and nonbehavioral tools (e.g., electrophysiological tests) in order to obtain more information about the mechanism of the Franssen illusion perception, the precedence effect, and its potential applications in clinical settings.

5. Conclusion

This study investigated the impacts of different tasks and presentation modes on the Franssen auditory-spatial illusion perceptual aspect. Subsequently, the relationship between illusion perception and the listener's ability in binaural temporal resolution was studied. The findings of the present study, in accordance with previous works, showed the importance of reverberation in the Franssen illusion occurrence and onset's non-dominancy under insert phone conditions. Furthermore, there was no significant correlation between the results of the behavioral binaural temporal resolution test (with a fixed interaural difference and center frequency) and the Franssen illusion perception among young people with normal temporal resolution, which may suggest that room reflections (reverberation) complicate the ability to process ITDs (rather than poor ITD processing for the "steady state" part of signal).

Funding

This study is part of Mehri Maleki's Ph.D. dissertation at the Iran University of Medical Sciences (code: 1401-1-6-23114).

References

1. Babkoff H., Muchnik C., Ben-David N., Furst M., Even-Zohar S., Hildesheimer M. (2002), Mapping lateralization of click trains in younger and older populations, *Hearing research*, **165**(1–2): 117–127, doi: 10.1016/S0378-5955(02)00292-7.

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

3. Brughera A., Dunai L., Hartmann W.M. (2013), Human interaural time difference thresholds for sine tones: The high-frequency limit, *The Journal of the Acoustical Society of America*, **133**(5): 2839–2855, doi: 10.1121/1.4795778.

4. Delphi M., Lotfi M.-Y., Moossavi A., Bakhshi E., Banimostafa M. (2017), Reliability of interaural time difference-based localization training in elderly individuals with speech-in-noise perception disorder, *Iranian Journal of Medical Sciences*, **42**(5): 437–442.

5. Donovan J.M., Nelson B.S., Takahashi T.T. (2012), The contributions of onset and offset echo delays to auditory spatial perception in human listeners, *The Journal of the Acoustical Society of America*, **132**(6): 3912–3924, doi: 10.1121/1.4764877.

6. Franssen N.V. (1960), *Some considerations on the mechanism of directional hearing*, Ph.D. Thessis.

7. Gelfand S.A. (2016), *Essentials of Audiology*, 4th ed., Thieme, New York.

8. Grothe B., Pecka M., McAlpine D. (2010), Mechanisms of sound localization in mammals, *Physiological Reviews*, **90**(3): 983–1012, doi: 10.1152/physrev.00026.2009.

9. Hafter E.R., Dye R.H., Gilkey R.H. (1979), Lateralization of tonal signals which have neither onsets nor offsets, *The Journal of the Acoustical Society of America*, **65**(2): 471–477, doi: 10.1121/1.382346.

10. Haqqee Z., Valdizón-Rodríguez R., Faure P.A. (2021), High frequency sensitivity to interaural onset time differences in the bat inferior colliculus, *Hearing Research*, **400**: 108133, doi: 10.1016/j.heares.2020.108133.

11. Hartmann W.M., Rakerd B. (1989), Localization of sound in rooms IV: The Franssen effect, *The Journal of the Acoustical Society of America*, **86**(4): 1366–1373, doi: 10.1121/1.398696.

12. Higgins N.C., McLaughlin S.A., Da Costa S., Stecker G.C. (2017), Sensitivity to an illusion of sound location in human auditory cortex, *Frontiers in Systems Neuroscience*, **11**, doi: 10.3389/fnsys.2017.00035.

13. Litovsky R.Y., Godar S.P. (2010), Difference in precedence effect between children and adults signifies development of sound localization abilities in complex listening tasks, *The Journal of the Acoustical Society of America*, **128**(4): 1979–1991, doi: 10.1121/1.3478849.

14. O'Regan J.K., Rensink R.A., Clark J.J. (1999), Change-blindness as a result of 'mudsplashes', *Nature*, **398**: 34, doi: 10.1038/17953.

15. Pierce A.H. (1901), *Studies in Auditory and Visual Space Perception*, Longmans, Green, and Company.

16. Pisoni D.B. (1977), Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops, *The Journal of the Acoustical Society of America*, **61**(5): 1352–1361, doi: 10.1121/1.381409.

17. Rajala A.Z., Yan Y., Dent M.L., Populin L.C. (2013), The inferior colliculus encodes the Franssen auditory spatial illusion, *European Journal of Neuroscience*, **38**(7): 3056–3070, doi: 10.1111/ejn.12325.

18. Saberi K. (1996), Observer weighting of interaural delays in filtered impulses, *Perception & Psychophysics*, **58**(7): 1037–1046, doi: 10.3758/BF03206831.

19. Scharf B., Florentine M., Meiselman C. (1976), Critical band in auditory lateralization, *Sensory Processes*, **1**(2): 109–126.

20. Schroeger A., Raab M., Cañal-Bruland R. (2022), Tau and kappa in interception – how perceptual spatiotemporal interrelations affect movements, *Attention, Perception, & Psychophysics*, **84**: 1925–1943, doi: 10.3758/s13414-022-02516-0.

21. Seeber B.U., Hafter E.R. (2011), Failure of the precedence effect with a noise-band vocoder, *The Journal of the Acoustical Society of America*, **129**(3): 1509–1521, doi: 10.1121/1.3531836.

22. Soeta Y., Nakagawa S. (2007), Effects of the binaural auditory filter in the human brain, *NeuroReport*, **18**(18): 1939–1943, doi: 10.1097/WNR.0b013e3282f1d4fc.

23. Stevens S.S., Newman E.B. (1936), The localization of actual sources of sound, *The American Journal of Psychology*, **48**(2): 297–306, doi: 10.2307/1415748.

24. Suneel D., Staisloff H., Shayman C.S., Stelmach J., Aronoff J.M. (2017), Localization performance correlates with binaural fusion for interaurally mismatched vocoded speech, *The Journal of the Acoustical Society of America*, **142**(3): 276–280, doi: 10.1121/1.5001903.

25. Van Deun L. *et al.* (2009), Sound localization, sound lateralization, and binaural masking level differences in young children with normal hearing, *Ear and Hearing*, **30**(2): 178–190, doi: 10.1097/AUD.0b013e318194256b.

26. Verschooten E. *et al.* (2019), The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints, *Hearing research*, **377**: 109–121, doi: 10.1016/j.heares.2019.03.011.

27. Wallach H., Newman E.B., Rosenzweig M.R. (1949), The precedence effect in sound localization, *The American Journal of Psychology*, **62**(3): 315–336, doi: 10.1121/1.1917119.

28. Yagcioglu S., Ungan P. (2006), The 'Franssen' illusion for short duration tones is preattentive: A study using mismatch negativity, *Brain Research*, **1106**(1): 164–176, doi: 10.1016/j.brainres.2006.05.075.

29. Yost W.A. (2017), Sound source localization identification accuracy: Envelope dependencies, *The Journal of the Acoustical Society of America*, **142**(1): 173–185, doi: 10.1121/1.4990656.

30. Yost W.A., Mapes-Riordan D., Guzman S.J. (1997), The relationship between localization and the Franssen effect, *The Journal of the Acoustical Society of America*, **101**(5): 2994–2997, doi: 10.1121/1.418528.

31. Yost W.A., Zhong X. (2014), Sound source localization identification accuracy: Bandwidth dependencies, *The Journal of the Acoustical Society of America*, **136**(5): 2737–2746, doi: 10.1121/1.4898045.

Archives of Acoustics Vol. 49, No. 1, pp. 27–35 (2024), doi: 10.24425/aoa.2024.148772

Research Paper

Effect of Curvature Shape of Transparent COVID-19 Protective Face Shields on the Speech Signal

Miloš BJELIĆ*, Miomir MIJIĆ, Tatjana MILJKOVIĆ, Dragana ŠUMARAC PAVLOVIĆ

School of Electrical Engineering, University of Belgrade
Belgrade, Serbia; e-mails: emijic@etf.rs; tm@etf.rs; dsumarac@etf.rs

*Corresponding Author e-mail: bjelic@etf.rs

(*received June 21, 2023; accepted October 21, 2023; published online January 16, 2024*)

Recent papers and studies over the course of last three years have shown that COVID-19 has a negative impact on the speech communication quality between people. This paper presents an influence analysis of the curvature shape of protective transparent shields on the speech signal. Five shields made of the same material and dimensions but with different curvatures were analyzed, from a completely flat to a very curved shield which has the same shape of curvature at its top and bottom and covers the entire face. The influence of the shield is analyzed with two types of experiments – one using dummy head with integrated artificial voice device, and the other using real speakers (female and male actors). It has been shown that usage of protective shields results in a relative increase in the speech signal level, in the frequency range of around 1000 Hz, compared to the situation when protective shields are not used. The relative increase in speech signal levels for large-curvature shields can be up to 8 dB. The possible causes of this phenomenon have been analyzed and examined.

Keywords: acoustic effects; COVID-19; curvature; face shield; pandemic; protection; speech.

1. Introduction

The COVID-19 pandemic has affected lives of people worldwide by introducing social distance, hand sanitizers, and the mandatory wearing of face protectors such as protective masks and shields. Protective equipment was used to prevent the spread of the virus but also negatively impacted communication between people. Masks as a visual barrier affected verbal communication, readability of emotional facial expressions, and lip-reading. In everyday communication, lip-reading is extremely useful for people (Dalka *et al.*, 2006).

Transmission of speech signal through the physical channel leads to reduced speech intelligibility, especially in the situation when parasitic signals overlay parts of smaller speech signal amplitudes. Overlaying quieter parts of the speech signal can occur as time-uniform noise coverage or time-limited coverage by reflections in the room (Nábělek *et al.*, 1989). It has been shown that speech intelligibility can be reduced by more than 15% in rooms with a long reverberation time (Liu *et al.*, 2020). Also, in the literature has been shown that at low signal-to-noise ratios in rooms, typically less than 5 dB, speech intelligibility can be below 75% (Liu *et al.*, 2020; Kociński, Sęk, 2005; Choi, 2020; 2021).

Face protective equipment degrades speech intelligibility and many researchers in acoustics have focused on examining the impact of protective masks on speech communication (Nobrega *et al.*, 2020; Caniato *et al*, 2021; Pörschmann *et al*, 2020; Corey *et al.*, 2020; Atcherson *et al.*, 2017; 2020; Magee *et al.*, 2020; Bottalico *et al.*, 2020; Wolfe *et al.*, 2020; Rudge *et al.*, 2020; Kopechek, 2020). Several groups of authors examined speech intelligibility with the use of face protective equipment in rooms, such as school classrooms (Choi, 2021; Caniato *et al.*, 2021; Bottalico, 2020; Wolfe *et al.*, 2020; Rudge *et al.*, 2020). The results of these studies have shown that speech intelligibility with protective masks is less than 25%. Some research analyzed the effects of masks through the forms of long-term speech spectrum as an objective measure of the impact of protective equipment (Choi, 2021; Corey *et al.*, 2020; Atcherson *et al.*, 2021).

These studies show that the attenuation introduced by protective masks can be up to 20 dB at high frequencies. In addition to objective methods for testing some of the hypotheses related to protective masks, subjective tests were also used (Magee *et al.*, 2020; Rudge *et al.*, 2020). The influence of the material from which the protective equipment was made on the speech signal has been investigated in several studies (Corey *et al.*, 2020; Atcherson *et al.*, 2021). Choosing the design and material of the face mask, the boundary conditions of acoustic radiation at the mouth opening of the speaker are changing, so it is necessary to make articulatory adjustments to maintain speech quality (Vojnović *et al.*, 2018). Research has dealt with the direct impact of poor intelligibility due to the usage of protective equipment. Studies have shown that a protective mask affects listener fatigue during prolonged listening due to reduced intelligibility (Choi, 2020; Nobrega *et al.*, 2020).

Research conducted by two groups of scientists has shown that the usage of protective masks reduces speech intelligibility in people with hearing impairments because they do not have lip reading possibility (Dalka *et al.*, 2006; Atcherson *et al.*, 2017). It has been shown that speech intelligibility improves when transparent protective equipment is used in their case. Therefore, some research focuses on examining the impact of transparent protective shields on speech communication (Caniato *et al.*, 2021; Corey *et al.*, 2020; Atcherson *et al.*, 2021). The influences of protective shields of various sizes, primarily on the shape of the speech signal spectrum, were considered. In the first approximation, the shield effect on the speech signal can be described as low-pass filtering. By comparing the difference in the speech signal spectrum in cases with and without the use of a protective shield, it was noticed that in certain areas, typically around 1000 Hz, there is an increase in these differences (Corey *et al.*, 2020; Atcherson *et al.*, 2021). When varying recording microphone positions, it was concluded that the highest levels of the speech signal occur when the microphone is in the space between the mouth and the transparent shield. Explanations for the causes of these increases, can be found related to resonant phenomena in the literature (Atcherson *et al.*, 2021). This indicates that protective shields have some other effects on the speech signal, apart from the weakening that occurs with protective masks.

The value of the energy, between the shields and the face increase of the speech signal, in case of the use of different transparent shields varies (Corey *et al.*, 2020; Atcherson *et al.*, 2021). Therefore, it was hypothesized that the shape of the protective shield, primarily its curvature, may affect the energy increase of the speech signal in a certain frequency range. The stated hypothesis was the motivation for the research presented in this paper. The studies aimed primarily at examining and describing possible reasons for the increase in the speech signal level in determining spectrum parts when protective shields were used. In the literature, different shields were compared. However, they were of different dimensions and made of different materials, so the influence of the shield curvature on the speech signal could not be clearly seen (Atcherson *et al.*, 2021). This article examines the change in the shape of the long term spectrum of the speech signal when protective shields of the same shape were made of the same material but with different curvatures. The realized experiments aimed at quantifying the curvature influence of the transparent shield on the long-term speech spectrum. The comparing approach of the long-term speech spectrum in the case without the shield's usage and in cases where shields of different curvature sizes were used, from a completely flat shield to a shield of large curvature, was used. Experiments were conducted with real speakers reading text in Serbian and also by another method, using a dummy head to enable the repeatability of the experiment.

2. Methods

2.1. Shields of different curvatures

In order to examine and quantify the curvature influence of protective shields on the speech signal, several shields made of the same transparent material with the same dimensions were analyzed. Figure 1a shows the shape of the transparent plastic foil used for protective shields in this paper. The width and height of the foil are 24 cm each, and the thickness is 0.5 mm. The used foil was not purposely made but used plastic foil from shields commercially available in stores and pharmacies. For achieving different curvatures, appropriate plastic supports were used which were produced with a 3D printer. The plastic support was always present on the upper side of the plastic foil so that the transparent foil could be attached to the head. In a situation where it was necessary, the plastic support was placed on the lower end of the plastic foil in order to obtain the appropriate shield curvature.

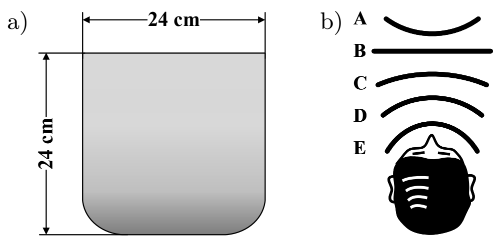

Fig. 1. a) Transparent plastic foil for shields; b) curvature shapes of protection shields.

Figure 1b shows a sketch with the curvature shapes of the protective shields used in this paper. The shields are arranged according to the degree of curvature and marked with letters from A to E. The distances of the middle of all types of shields from the surface of the face are approximately the same, but due to the need to compare the curvature in one place, the sketch is given with shields placed in front of each other. In order to better perceive the shape of the used shields, Fig. 2 shows all the used shields and their position on the human head, frontally and from the profile. Figure 2a shows the protective shield case A, which is not a protective shield in the sense that it is used as a protective device. This shield is the reverse version of the standard shield, i.e., the curve is turned away from the face and not towards it. To analyse the shield curvature influence on the speaking signal, this shield was chosen as the antipode to the shield that has the greatest curvature.

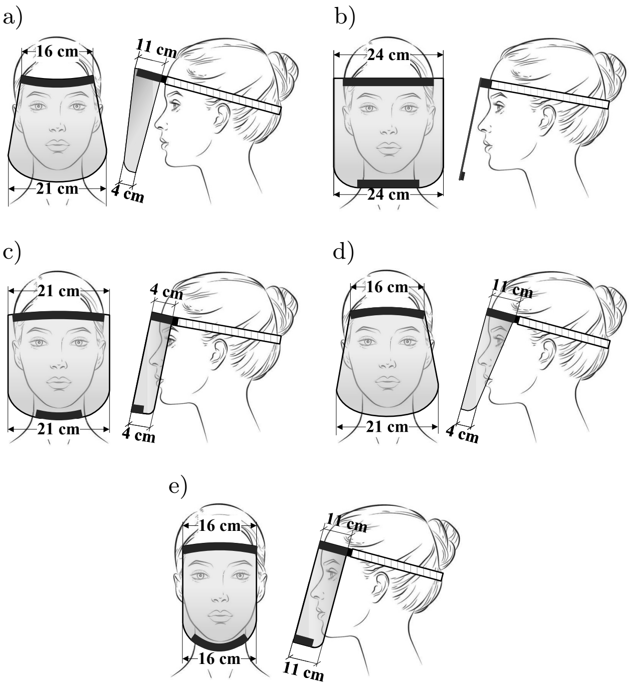

Fig. 2. Shield dimensions and position on the head: a) case A; b) case B; c) case C; d) case D; e) case E.

Figure 2b shows a flat shield (case B), i.e., a shield with no curvature. Special plastic supports were constructed on the upper and lower sides of the shield, which ensured that the transparent foil from Fig. 1a was flat over its entire surface. The case B shield represents a case where the shield has no curvature. In Fig. 2c, a shield with a slight curvature (case C) is shown. Its width protects the face from the outer environment and such shields can be found on the market as protective equipment. Plastic supports on the lower and upper sides were used to form this shield. The case C shield in this paper represents the mean curviness. The shield showed in Fig. 2d is a shield of greater curvature compared to the shields shown in the previous figures. It has been formed by placing the plastic support on the upper side of the transparent foil, which enables the shield placement on the head. The formed shield is 5 cm wider on the lower side compared to the upper side. This type of shield can also be found on the market as a protective device during the COVID-19 pandemic. In the experiments, the case D shield was used to obtain the case A shield so that it is turned away from the face and attached to the head using a special bracket. The last type of shield presented in this paper is shown in Fig. 1e. This shield has the greatest curvature compared to all other used types of shields. Its width on the lower and upper sides is the same and it is 16 cm. This was achieved by adding a plastic bracket on the underside of the case D shield. Observed in relation to the human head, this type of shield leaves the smallest air space to the face, seen from the profile, i.e., it protects the face from outer influences the most. When all the types of shields shown are compared, it can be said that the case A represents the shield with the lowest curviness (inversely curved). In contrast, the type E shield represents the shield with the highest positive curviness.

2.2. Experimental setup

The experiments were performed in anechoic chamber with a volume of 50 m^3. The level of ambient noise in the room was around 20 dB(A), and the reverberation time of the anechoic chamber is around 0.1 s. Two groups of experiments have been performed:

- the first group – with real speakers reading a text (without the shield, and with five shields of different curvatures);
- the second group – with a dummy head reproducing speech signals (without the shield, and with five shields of different curvatures).

In Fig. 3a one of the speakers is presented with the case D shield. As a source of the speech signal in

a) b)


Fig. 3. Experimental setting in anechoic conditions: a) one of the speakers with protective shield (case D); b) artificial head with protective shield (case D).

Table 1. Statistical distribution of letters in the used text.

Type	Vowels	Semivowels	Plosives	Fricatives	Affricates	Nasals
Part of text [%]	45.3	10.6	17.6	13.1	2.6	10.8

the second group of experiments, a dummy head with an integrated artificial voice device is used (*Technical documentations of the manufacturer*, 1971). In Fig. 3b the dummy head with the case D shield is presented. At the distance of 20 cm from the middle of the dummy head (or speaker head), a measurement microphone is placed (*Technical documentations of the manufacturer*, 2010). The recording of the signal emitted by the dummy head is performed using an audio interface (*Technical documentations of the manufacturer*, 2012) with a sampling rate of 48 kHz.

In the first group of experiments, real speakers were reading a literary text in Serbian. The literary text is composed of 436 words (1021 letters) which corresponded to average reading time of 2 min. Statistical distribution of letters of this text per articulation is given in Table 1.

The letter distribution of this text is in accordance with general distribution for the Serbian, obtained on a large sample of literary texts (Jovičić, 1999). Ten speakers, five male and five female, participated in the first group of experiments, with the age range between 20 to 27 years old. In this paper, the speakers were professional actors, final year students of the Faculty of Dramatic Arts, without speech defects or vocal tract diseases. To conduct the experiments, reading the same text as much repeatability as possible was required, in order to eliminate the influence of the speaker on the results of the shield tests. Each speaker had to read the same text 6 times, once without a shield and 5 times with shields of different curvature. This is why actors were chosen because they have good control over intonation, rhythm and dynamics of speech. Each of the speakers was recorded reading the text not wearing the shield, and then the recordings were made of speakers reading the text when wearing each type of shield (five types with different curvatures). Therefore, 60 recordings were made with real speakers.

In the second group of experiments speech signals were reproduced using a dummy head. Experiments were performed with a dummy head in order to examine whether, in addition to the influence of the shield, there is also an influence of the speaker's articulation. The reproduced speech signals (ten signals) were taken from the first group of experiments and correspond to the case when real speakers are not wearing a protective shield. The microphone records a speech signal emitted by the dummy head (ten recordings). Afterwards, shields of different curvatures are placed on the dummy head and recordings are made (50 recordings). In such a way, 60 recordings of speech emitted by the dummy head were made.

2.3. Long term speech spectrum

For speech analysis in this paper, a spectrum in 1/3 octave bands is used, obtained with 1/3 octave filter bank (ANSI, 2004). In Fig. 4, a block diagram is shown for obtaining the long term spectrum of the speech signal, used in this paper. The speech signal (of real speakers or from dummy head) is at the input of the 1/3 octave filter bank, which is composed of filters numerated from 1 to m. In this paper, the frequency range of interest is 125 Hz to 16 kHz. The filtering is performed for all speech signals of the same category (e.g., for speech signals of all ten speakers without the protective shield).

Next, for all filtered signals, a square of RMS values is calculated which is a value proportional to the power of the speech signal in the observed frequency range. For each frequency range (defined by the corresponding filter), averaging is performed of the RMS values obtained for all speakers. This procedure is repeated for all frequency ranges of interest (1 to m). After averaging, m values are obtained which represent the long-term spectrum in 1/3 frequency bands.

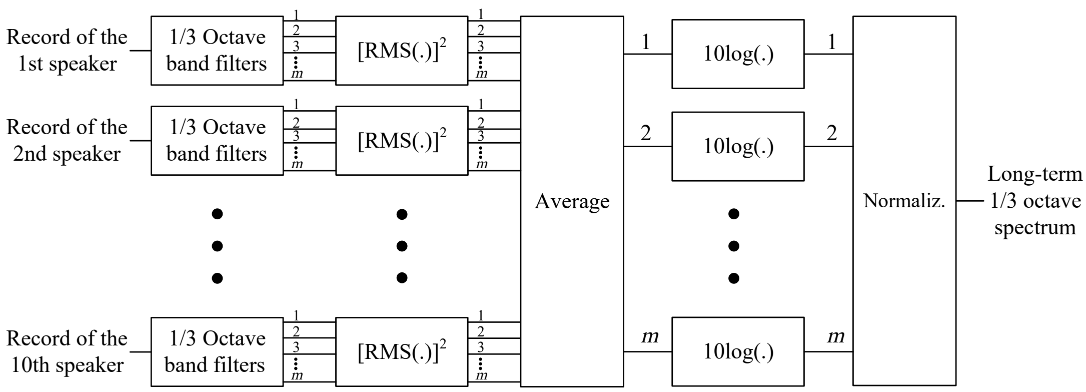

Fig. 4. Diagram for obtaining long term speech spectrum.

The values are then converted to dB and normalized so that the total level in the entire frequency range of interest equals 0 dB.

The presented procedure obtains the 1/3 octave long term spectrum for one group of recordings (e.g., speech signals of ten speakers without shield). This procedure is performed for all groups of recordings, which include all types of shields. In such a way, six long term spectrums are obtained (one without shield, and five with shields of different curvatures), in the case of real speakers. In the same way, six long term spectrums are obtained for the dummy head experiments as well.

3. Results and discussion

3.1. Speech signal without the shield

Figure 5 shows the 1/3 octave band long term spectrum of speech in the case of real speakers reading a text, not wearing the protective shield. The maximum level value of the normalized long-term speech spectrum corresponds to the 1/3 octave range with a center frequency of 500 Hz. In the range from 200 to 500 Hz, the spectrum decreases by 3 dB per octave, in the area below 200 Hz by 8 dB per octave, and in the range up to 3.1 kHz, the speech spectrum decreases 6 dB per octave. In the frequency range between 3.1 and 10 kHz, the spectrum is approximately flat, and after 10 kHz, it can be considered that there are no significant components in the long-term speech spectrum. The obtained spectrum coincides with the Serbian speech spectrum data from the literature (Vojnović, Mijić, 1997; Byrne *et al.* 1994). This shows that the selected group of speakers is relevant. The standard deviation of the averaged speech spectrum for the frequency range of interest belongs to the interval (2.4 ± 1.5 dB). The highest standard deviation value of 3.9 dB was obtained for the frequency range of 1250 Hz.

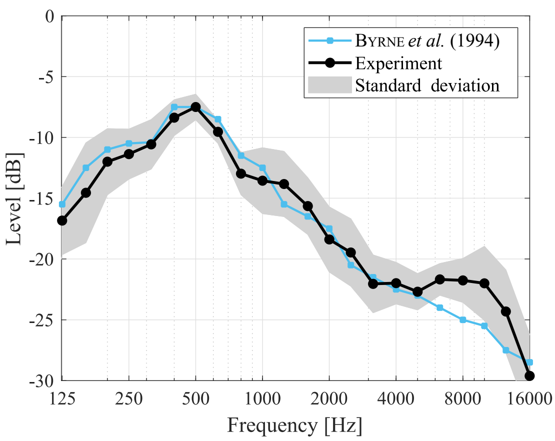

Fig. 5. Average speech spectrum of 10 speakers without a shield, standard deviation and speech spectrum from the literature.

3.2. Dummy head results

In order to see the influence of the shield curvature on the speech signal, differences in the spectrum of broadcasted speech were calculated for cases with and without protective shields. Long-term spectrum differences over 1/3 octave frequency bands averaged for 10 recorded signals for each of the shields. The calculated differences represent the insertion losses introduced by the shields in frequency bands. In this way, five frequency dependent insertion loss curves were calculated (five protective shields, case A to case E), which are shown in Fig. 6. The figure also depicts a symbolic representation of the different shield curvatures and their label for easier tracking of the results.

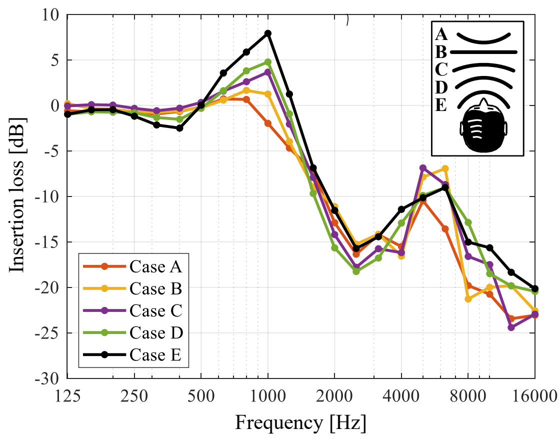

Fig. 6. Insertion loss introduced by the shields of different curvature – dummy head experiments.

In the frequency range up to 250 Hz, the differences shown in Fig. 6 are less than 0.5 dB, i.e., there are no significant differences between the speech signal spectrum with and without protective shields. In this frequency range, the sound wavelengths are up to several meters. Hence, the protective shield with dimensions of several tens of centimeters and a millimeter thickness is not a significant obstacle (Pierce, 2019). In the range from 250 to 500 Hz, the differences are slightly larger for all types of shields (for example in the case E – 2 dB). For all shields except the case A shield, the maximum difference with regard to the case without the usage of a protective shield is occurring for the 1/3 octave range with a center frequency of 1000 Hz. In the case of the shield with the largest curvature (case E), the difference is even 8 dB. The differences are 5, 4, and 2 dB (cases D, C, and B, respectively). It can be concluded that the level difference increases too at 1000 Hz with an increasing shield curvature. In case when the case A shield was used (reverse shield), there is no local maximum as with other shield types.

From Fig. 6, it could be seen that only negative differences occur in the range above 1250 Hz, i.e., that all shield types bring attenuation into the speech signal.

In addition to the local maximum another local maximum can be observed. The position of this maximum for different types of shields corresponds to the 1/3 frequency bands with central frequencies of 5000 and 6300 Hz. From Fig. 6, it can be seen that there is no dependence of the numerical values of the differences on the shield curvature. The connection between curvature and this local maximum does not exist because the obtained local maximum is a consequence of the resonance in the space (chamber) in front of the face on the shortest side, i.e., the resonance corresponding to the distance between the face and the shield. For all shield types, the distance between the speaker's mouth and the protective shield is approximately the same, about 3 cm, which can be seen in Fig. 2. Therefore, the differences in this frequency range for different shields are relatively small. In the band above 10 kHz, there are no significant components of the speech signal.

3.3. Real speaker results

As in the dummy head case, the differences of spectrums (insertion loss) were calculated for the cases of real speakers wearing shields of different curvature and the case when real speakers were without any protective shield. Insertion losses were averaged over all 10 speakers. The results are shown in Fig. 7. For all shields except the case A shield, the maximum difference with regard to the case without the usage of a protective shield is occurring for the 1/3 octave range with a center frequency of 1000 Hz. In the case of the shield with the largest curvature (case E), the difference is even 6.6 dB. Table 2 shows the averaged insertion loss values and standard deviations for the 1/3 octave range with a central frequency of 1000 Hz, for all speakers who participated in the experiment.

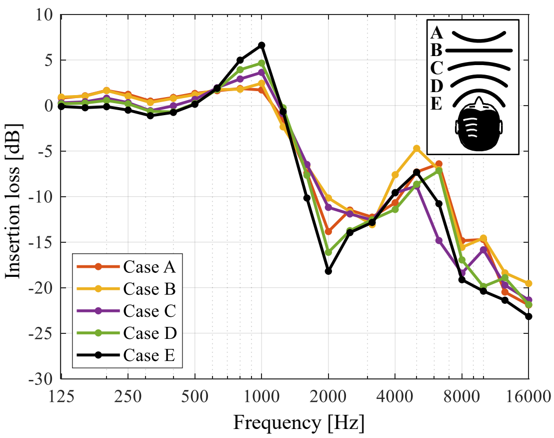

Fig. 7. Insertion losses introduced by the shields of different curvature – real speaker experiments.

As in the dummy head experiment, it can be concluded that the level difference increases at 1000 Hz with increasing shield curvature. In case when the case A shield was used (reverse shield), there is no local maximum as with other shield types. The values of differences for the frequency band at 1000 Hz in the case of real speaker experiments are somewhat lower compared to the dummy head experiments. Certain differences in the results are due to the different structure of surfaces of the face. In the real speaker experiment case it is human skin, whereas in the case of the dummy head experiment the surface is hard rubber. The absorption characteristics of these two surfaces are different, predominantly in the mid and high frequency range, where the result differences occur in the presented two groups of experiments. In studies involved in examining the influence of protective devices which use the dummy head and human head as signal sources, it was shown that certain differences occur (Pörschmann *et al.*, 2020; Corey, 2020; Atcherson *et al.*, 2021) which is in accordance with the results presented in this paper. Furthermore, the cause of differences in results obtained by two experiment methods can partially be due to differences in reading of the text when wearing protective mask. The dummy head is used in order to enable the repeatability of results, i.e., to eliminate the possible human influence because of different manner of text reading. In the case of experiments with real loudspeakers, another local maximum is present, in the frequency range of 5000 and 6300 Hz. The numerical values of the differences in these bands are within 4 dB, and are somewhat larger than in the case of the dummy head experiment. The connection between the curvatures and these local maximums is not present, which was also the case in the dummy head experiments.

Table 2. Insertion loss and standard deviation of insertion loss for all speakers for 1/3 octave range 1000 Hz.

	Insertion loss [dB]	Standard deviation [dB]
Case A	1.7	1.2
Case B	2.5	1.1
Case C	3.7	1.3
Case D	4.6	1.8
Case E	6.6	1.5

Based on the results presented in Figs. 6 and 7, it is concluded that the phenomena of pronounced maximum in the 1/3 frequency band at 1000 Hz is present regardless of the experiment type, i.e., whether the speech signal is reproduced with a dummy head or real speakers. The numerical values are different in two cases of experiments due to differences in surface characteristics of the human head and dummy head.

3.4. Analysis of the pronounced maximum in the spectrum

The obtained form of differences in the speech spectrum with and without a protective shield also appears in experiments conducted in the literature (Corey, 2020; Atcherson *et al.*, 2021). However, the causes of the relative increase in the long-term spectrum in the region around 1000 Hz have not been considered. The spectrum shape is similar to the situation when experiments were carried through on both human and artificial heads (Corey, 2020). There are claims that the appearance of the maximum difference in speech signal spectrum with and without the usage of a protective shield is a consequence of resonance in the protective shield foil (Atcherson *et al.*, 2020). In order to examine this claim, the vibrations on the shield were analyzed. For this experiment, the dummy head is used. An accelerometer recorded a transparent shield (case D) response to the impulse excitation (by the impact of a stick on the shield). In this way, the system impulse response was recorded.

Figure 8a shows the position of the accelerometer on the shield. An accelerometer weighing 0.635 grams was used (*Technical documentations of the manufacturer*, 2018) so that its mass would not affect the response of shields whose mass is not large. Several measurements were performed for different accelerometer positions and different excitation positions. The results are averaged, and Fig. 8b shows the 1/3 octave spectrum of the recorded vibrations (the frequency response of the shield). The biggest 1/3 octave value was set to 0 dB. Based on Fig. 8b, it could be seen that the maximum frequency response of the shield is positioned in the frequency range around 250 Hz and that it decreases towards higher frequencies. The maximum positive value of the differences in the speech signal spectrum with and without using a protective shield occurs in the range of about 1000 Hz. Since the positions of the maximums of the vibration spectrum on the shield (Fig. 8) and the positions of the maximum differences of the speech spectrum (Figs. 6 and 7) do not coincide, this means that vibrations generated in the transparent foil of the shield are not responsible for the local maximum in 1000 Hz.

a)

b)

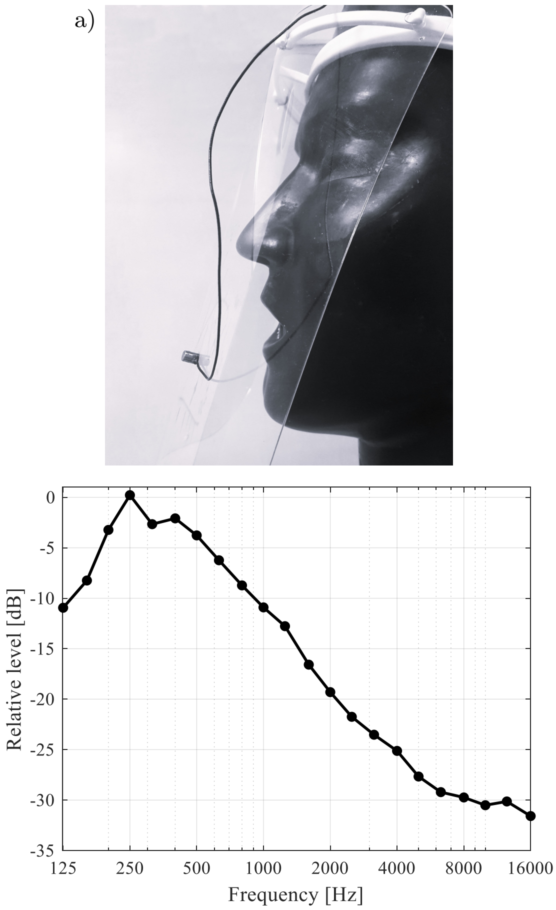

Fig. 8. a) Accelerometer on protective shield; b) shield's response.

A new hypothesis was introduced, given the proof that the shield material has no influence that could be perceived in the long-term speech spectrum. The maximum positive differences are due to the resonance that occurs in the space between the face and the shield. This narrow space (chamber) is determined by the surface of the face on the back, the shield surface on the front, the plastic shield holder on the upper side, while there is air on the other three sides of the chamber. Part of the speech signal energy (at lower frequencies, in the range up to 500 Hz) passes through the protective shield, so the protective shield does not affect this frequency range, as seen in Figs. 6 and 7. For the frequency range in which the transparent material represents an obstacle (medium and high frequencies), the speech signal excites the space (chamber) in front of the face in which the sound field is established. The chamber width is 24 cm, height 24 cm, and thickness is about 3 cm (a distance of the face from the shield). Due to the small thickness, for the frequency range of the speech signal above 500 Hz, this space could be considered a sound pipeline that is open at the ends. Based on the given shield dimensions (width and height), it is possible to determine the frequency corresponding to the surface resonance in this sound pipeline, approximately 1000 Hz (that resonance covers a band around 1000 Hz, not just a discrete frequency). For this frequency range in the space between the face and the shield, there is an increase in sound energy, which is also sustained in the increase in the value of sound pressure recorded by the microphone placed in front of the shield. Consequently, the differences shown in Figs. 6 and 7 have a maximum positive value for the 1/3 octave frequency range with a center frequency of 1000 Hz. The increase in the difference value with increasing shield curvature is a consequence of the fact that in the case of greater curvature, the air space at the ends of the chamber becomes smaller, i.e., the chamber behaves more like a sound pipe. The losses from the sides top and bottom are less allowing for stronger resonances. The smallest opening exists in the case of the case E shield, and in that case the space between the face and the shield most closely resembles a sound pipe, so the value of the difference shown in Fig. 5 is the largest and amounts to 8 dB. In contrast, for the case A shield, the largest opening is at the end of the chamber (practically a funnel), so in this case we cannot talk about the space in front of the face as a sound pipeline. Therefore, there is no increase in the difference for this type of shield at 1000 Hz. Other shield types due to the shape of the space they define, are between these two extreme cases, as shown by the results obtained.

4. Conclusion

This paper examines the influence of protective shields' curvature on the speech signal. In literature, results show that using transparent protective shields increases the level of speech signal in certain frequency ranges. However, detailed explanations for these phenomena are not found. This paper analyses and explains the reasons for the level increase in the long-term speech spectrum. The experiments were performed using an artificial head emitting recorded speech signals of real-life speakers. Furthermore, another set of experiments was performed with real speakers to test whether the speaker's articulation affects the end result. Five shields were examined, having the same overall dimensions, made of the same material, but with different curvature. The curvature of the shields is varied in order to examine whether there exists an influence of the level of curvature on the speech signal. It was shown that the resonant processes do not occur within the transparent material of the shield. Rather, they occur in the air space between the face of the speaker and the transparent material. Protective shields form a sound pipeline between the face of the speaker and the transparent material of the shield. Therefore, in this space, resonant processes occur which result in the level increase of the speech spectrum in certain frequency ranges. It has been concluded that increasing the curvature of the protective shield increases the relative difference in long-term speech spectrums, when comparing the long-term spectrum of the speech signal recorded with and without shield. This phenomenon is explained by the fact that increasing curvature reduces the volume of the air space between the face of the speaker and the shield, which means that the air space has the characteristics of a sound pipeline. This is the reason for the level increase in the long-term speech spectrum, when compared to the shields with lower curvature. The dummy head experiments show that in the frequency range around 1000 Hz, the level of speech signal when using protective shield with large curvature can be higher up to 8 dB compared to the speech signal when no shield is used. In the case of real speaker experiments this difference is 6.6 dB. Since the phenomenon is observed in both types of experiments, it is concluded that articulation has no influence and that the increase in the level of speech signal in this frequency range is a consequence of the curvature of the shield. Increase in speech level was also observed in the frequency range around 5000 Hz, and it is concluded that this phenomenon is a consequence of the resonance which corresponds to the distance between the shield and the face of the speaker. The increase in level in this frequency range is not dependent on the shield curvature, because the distance from the mouth of the speaker to the shield is approximately the same for all tested shields.

Acknowledgments

This work was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia under contract numbers 451-03-47/2023-01/200125.

References

1. ANSI (2004), *Specification for octave, half-octave, and third octave band filter sets*, No. S1.11, Acoustical Society of America.
2. Atcherson S.R. *et al.* (2017), The effect of conventional and transparent surgical masks on speech understanding in individuals with and without hearing loss, *Journal of the American Academy of Audiology*, **28**(1): 58–67, doi: 10.3766/jaaa.15151.
3. Atcherson S.R., Finley E.T., McDowell B.R., Watson C. (2020), More speech degradations and considerations in the search for transparent face coverings during the COVID-19 pandemic, *Audiology Today*, **32**(6): 20–27.
4. Atcherson S.R., McDowell B.R., Howard M.P. (2021), Acoustic effects of non-transparent and transparent face coverings, *The Journal of the Acoustical Society of America*, **149**: 2249–2254, doi: 10.1121/10.0003962.
5. Bottalico P., Murgia S., Puglisi G.E., Astolfi A., Kirk K.I. (2020), Effect of masks on speech intelligibility in auralized classrooms, *The Journal of the Acoustical Society of America*, **148**(5): 2878–2884, doi: 10.1121/10.0002450.
6. Byrne D. *et al.* (1994), An international comparison of long-term average speech spectra, *The Journal of the Acoustical Society of America*, **96**: 2108–2120, doi: 10.1121/1.410152.
7. Caniato M., Marzi A., Gasparella A. (2021), How much COVID-19 face protections influence speech intelligibility in classrooms?, *Applied Acoustics*, **178**: 1–14, doi: 10.1016/j.apacoust.2021.108051.
8. Choi Y. (2020), The intelligibility of speech in university classrooms during lectures, *Applied Acoustics*, **162**: 1–8, doi: 10.1016/j.apacoust.2020.107211.
9. Choi Y. (2021), Acoustical measurements of masks and the effects on the speech intelligibility in uni-

versity classrooms, *Applied Acoustics*, **180**: 1–8, doi: 10.1016/j.apacoust.2021.108145.

10. Corey R.M., Jones U., Singer A.C. (2020), Acoustic effects of medical, cloth, and transparent face masks on speech signals, *The Journal of the Acoustical Society of America*, **148**: 2371–2375, doi: 10.1121/10.0002279.
11. Dalka P., Kostek B., Czyżewski A. (2006), Vowel recognition based on acoustic and visual features, *Archives of Acoustics*, **31**(3): 275–288.
12. Jovičić S. (1999), *Fundamental of Speech Communication*, p. 81, Beograd.
13. Kociński J., Sęk A. (2005), Speech intelligibility in various spatial configurations of background noise, *Archives of Acoustics*, **30**(2): 173–191.
14. Kopechek J.A. (2020), Increased ambient noise and elevated vocal effort contribute to airborne transmission of COVID-19, *The Journal of the Acoustical Society of America*, **148**(5): 3255–3257, doi: 10.1121/10.0002640.
15. Liu H., Ma H., Kang J., Wang C. (2020), The speech intelligibility and applicability of the speech transmission index in large spaces, *Applied Acoustics*, **167**: 1–12, doi: 10.1016/j.apacoust.2020.107400.
16. Magee M. *et al.* (2020), Effects of face masks on acoustic analysis and speech perception: Implications for peri-pandemic protocols, *The Journal of the Acoustical Society of America*, **148**(5): 3562–3568, doi: 10.1121/10.0002873.
17. Nábělek A.K., Letowski T.R., Tucker F.M. (1989), Reverberant overlap- and self-masking in consonant identification, *The Journal of the Acoustical Society of America*, **86**(4): 1259–1265, doi: 10.1121/1.398740.
18. Nobrega M., Opice R., Lauletta M.M., Nobrega C.A. (2020), How face masks can affect school performance, *International Journal of Pediatric Otorhinolaryngology*, **138**: 1–2, doi: 10.1016/j.ijporl.2020.110328.
19. Pierce A. (2019), *Acoustics an Introduction to Its Physical Principles and Applications*, 3rd ed., Springer Nature, Cham Switzerland.
20. Pörschmann C., Lübeck T., Arend J.M. (2020), Impact of face masks on voice radiation, *The Journal of the Acoustical Society of America*, **148**: 3663–3670, doi: 10.1121/10.0002853.
21. Rudge A.M., Sonneveldt V., Brooks B.M. (2020), The effects of face coverings and remote microphone technology on speech perception in the classroom, *The Moog Center for Deaf Education*, pp. 1–8.
22. *Technical documentations of the manufacturer* (1971), https://www.opweb.de/english/company/Brüel_and_Kjær/4219 (access: 16.06.2023).
23. *Technical documentations of the manufacturer* (2010), http://www.nti-audio.com/Portals/0/data/en/MiniSPL-Measurement-Microphone-Product-Data.pdf (access: 16.06.2023).
24. *Technical documentations of the manufacturer* (2012), http://download.steinberg.net/downloads_hardware/UR22/UR22_documentation/UR22_OperationManual_en.pdf (access: 16.06.2023).
25. *Technical documentations of the manufacturer* (2018), https://www.bksv.com/-/media/literature/Product-Data/bp2038.ashx (access: 16.06.2023).
26. Vojnović M., Mijić M. (1997), The influence of the oxygen mask on longtime spectra of continuous speech, *The Journal of the Acoustical Society of America*, **102**(4): 2456–2458, doi: 10.1121/1.421021.
27. Vojnović M., Mijić M., Šumarac Pavlovic D. (2018), Transfer characteristics of vocal tract closed by mask cavity, *Archives of Acoustics*, **43**(2): 307–311, doi: 10.24425/122378.
28. Wolfe J. *et al.* (2020), Optimizing communication in schools and other settings during COVID-19, *The Hearing Journal*, **73**(9): 40–45, doi: 10.1097/01.HJ.0000717184.65906.b9.

Archives of Acoustics Vol. 49, No. 1, pp. 37–48 (2024), doi: 10.24425/aoa.2024.148768

Research Paper

Short Utterance Speaker Recognition Based on Speech High Frequency Information Compensation and Dynamic Feature Enhancement Methods

Yunfei ZI*, Shengwu XIONG

School of Computer and Artificial Intelligence, Wuhan University of Technology
Wuhan, China

*Corresponding Author e-mail: yfzi@whut.edu.cn

(*received April 5, 2022; accepted September 20, 2023; published online January 8, 2024*)

This work aims to further compensate for the weaknesses of feature sparsity and insufficient discriminative acoustic features in existing short-duration speaker recognition. To address this issue, we propose the Bark-scaled Gauss and the linear filter bank superposition cepstral coefficients (BGLCC), and the multi-dimensional central difference (MDCD) acoustic feature extracted method. The Bark-scaled Gauss filter bank focuses on low-frequency information, while linear filtering is uniformly distributed, therefore, the filter superposition can obtain more discriminative and richer acoustic features of short-duration audio signals. In addition, the multi-dimensional central difference method captures better dynamics features of speakers for improving the performance of short utterance speaker verification. Extensive experiments are conducted on short-duration text-independent speaker verification datasets generated from the VoxCeleb, SITW, and NIST SRE corpora, respectively, which contain speech samples of diverse lengths, and different scenarios. The results demonstrate that the proposed method outperforms the existing acoustic feature extraction approach by at least 10% in the test set. The ablation experiments further illustrate that our proposed approaches can achieve substantial improvement over prior methods.

Keywords: Bark-scaled Gauss; linear filter; filter bank superposition; multi-dimensional central difference; speaker recognition.

1. Introduction

Speaker recognition, as one of the most popular biometric technologies (Wu *et al.*, 2016) today has been widely used in many fields such as access control, forensic evidence provision, security, and telephone banking user authentication (Vogt *et al.*, 2010). The purpose of speaker recognition is to recognize the claimed identity of the speaker, which includes speaker verification and speaker identification (Campbell, 1997). One of its main purposes is to determine whether the test sound from the speaker is acceptable. After decades of development, the technology of speaker verification has been extensively studied, and the recognition system has achieved relatively satisfactory performance, provided that the enrollment and test voices are long enough and the signal-to-noise ratio (SNR) is large enough (Zinchenko *et al.*, 2017; Greenberg *et al.*, 2013; Kinnunen, Li, 2010).

However, in some application scenarios, it is not easy to collect a suitable speech. The current speaker verification system has a significant decrease of the recognition rate in a short utterance environment (Nosratighods *et al.*, 2010). A short-duration speech means that the speech contains insufficient acoustic characteristics. Obtaining enough speech data is difficult for many real-world applications and users are reluctant to provide sufficient voice data, especially during the testing phase asking the user to speak for a long time, for instance in phone banking. In other cases, it is very difficult to collect enough data, e.g., in forensic applications, in the security field. The performance degradation caused by insufficient data is called the short-duration issue.

Current speaker recognition systems have achieved great success and performed well when the enrollment and test data are sufficiently long; hence, the traditional acoustic feature extraction methods are designed based on long-duration speech, and the long-duration speech feature extraction filter arrangement method mainly focuses on the low-frequency domain, this makes high-frequency domain features more sparse in the short duration speech, and high-frequency domain information best represents timbre and detail (Huang, Pun, 2020). At the same time, the traditional acoustics features include fewer dynamic features of speakers, as a result, fewer acoustic features are extracted that can be discriminated for speaker recognition. Research on the more challenging short-duration text-independent speaker recognition of discriminative feature compensation has been more in demand lately, which is also our focus in this work.

Although the traditional speaker model has obvious feature specificity, because the number of features is too few, it is still susceptible to noise interference, and awful recognition performance. The acoustic feature extraction design should address how to extract the high discriminative embeddings more effectively in short-duration audio speaker recognition. Therefore, how improving the effectiveness of discriminative acoustic feature extraction, in short utterance speaker environment, is an urgent problem to be solved.

To address the problems, the solution is proposed in this paper. In the Bark-scaled Gauss filter bank acoustic feature extraction method the filter bank distribution puts more emphasis on the low-frequency bands, which portray the low-frequency spectrum of speech in great detail. In comparison, the Bark-scaled Gauss filter distribution less emphasizes the high-frequency bands, so some helpful information is easily lost from the high-frequency domain. However, the details of the high frequency can enhance the information of one's timbre. To enhance the valuable information on the high-frequency, the Bark-scaled Gauss and linear filter bank superposition cepstrum coefficients (BGLCC) are proposed to portray more precise high-frequency details. The filter bank of the conventional acoustic feature extraction method puts more emphasis on the low-frequency band. In contrast, the linear triangle filter is uniformly distributed, which can remedy the weakness of the sparse high-frequency information and insufficient acoustic feature extraction brought by the uneven distribution of a single filter, thus, integrating the advantages of both and constructing new hybrid feature parameters is a way to enhance the feature sparsity problem.

Moreover, aiming to capture better dynamics features of speakers, we propose multi-dimensional central difference (MDCD) features based on the BGLCC features matrix, simultaneously, to improve the performance of short utterance speaker recognition. The MDCD are multi-dimensional central difference features in the time-frequency plane. Different speakers speak the same word or sentence in different ways. The proposed MDCD feature concatenate information about the speaker from four different dimensions, this can explain why it performs significantly better than traditionally used speech features in speaker recognition tasks under various conditions. Therefore, the MDCD features can further compensate for the limited and sparse dynamic acoustic characteristics of short-duration audio signals based on extracting dynamic speaker features.

1.1. Related works

To enhance the efficiency of performance of short-duration audio speaker recognition algorithms, some approaches have been presented by previous research studies. In terms of front-end acoustic feature extraction, the vast majority of existing acoustic feature extraction is based on some form of the short-term frequency spectrum to implement short utterance speaker recognition algorithms like Mel-frequency cepstral coefficients (MFCCs) (Herrera-Camacho *et al.*, 2019; Paseddula, Gangashetty, 2018) linear prediction cepstral coefficients (LPCCs) (Yang *et al.*, 2019; Atal, 1974) and constant Q cepstral coefficients (CQCC) (Todisco *et al.*, 2017), acoustic features. For instance, by judiciously combining MFCC and LPCC for short-duration audio signal speaker recognition (Chowdhury, Ross, 2020), the hypothesis is that MFCC and LPC capture two different aspects of speech, namely, speech perception and speech production. By using the model method, there is speaker recognition based on GMM-UBM from MFCC features in the limited enrollment and test data (Omar, Pelecanos, 2010). Another work is the I-vector approach and factor analysis subspace estimation introduced by (Kenny *et al.*, 2005; Dehak *et al.*, 2010) to reduce the number of redundant model parameters, resulting in more accurate speaker models. Some approaches attempt to increase performance by selecting segments with better discriminability based on speaker features (Nosratighods *et al.*, 2010) GMM and the CNN hybrid method (Liu *et al.*, 2018), the method is an initial alignment method for short utterance feature, which can improve the effect of short utterance speaker recognition. In their work, front-end feature extraction methods are based on Fourier transform Mel-triangle filtering and linear prediction cepstral coefficients for model training and testing as well as model inference.

With further developments in deep learning, various methods for speaker recognition or short utterance speaker recognition have been proposed, by Povey *et al.* (2018), the factorized time delay neural network (F-TDNN) has been proposed which divides the parameter matrix of TDNN into smaller matrices to increase the training effectiveness and the extended time delay neural networks (E-TDNN) was proposed in (Snyder *et al.*, 2019), E-TDNN is based on its broader and deeper network structure, thus allowing more information to be learned, they both improve speaker recognition performance significantly. In (Villalba *et al.*, 2020), based on F-TDNN and E-TDNN, the best results were obtained for speaker evaluation in SRE18 and in the field. In addition, a focus on aggregation information, channel attention, and propagation method were proposed (Desplanques *et al.*, 2020), called TDNN-based speaker verification (ECAPA-TDNN), which further improves the robustness of speaker recognition. After years of development, the performance of short utterance speaker recognition has improved considerably, but it is still unsatisfactory in some complex scenarios.

Most of the aforementioned methods would benefit from the optimization model, enhance data characteristics and extract more discriminative features for speaker recognition. With 5~10 seconds of speech duration, they all improve speaker recognition performance when audio speech becomes shorter, but they still face significant challenges.

Generally speaking, there are two types of speech recognition features, namely linear prediction cepstral coefficients (LPCCs) and Mel-frequency cepstral coefficients (MFCCs), but when used in a short-duration environment, they suffer from a drop in performance. As we know, there is no reasonably good short-duration speaker verification model. Unfortunately, there is no better feature extraction method to obtain sufficient and discriminative speaker information models from short-duration speech signals, there are no better training methods.

1.2. Contribution

To compensate for the problems of difficult short-utterance discriminative feature capture and insufficient discriminative acoustic features, we propose a filter superposition-based multi-dimensional central difference discriminative acoustic feature extraction method for feature compensation and enhancement of short-duration speech speaker recognition. The proposed method can significantly improve the performance and accuracy of the the short-duration speech speaker recognition system.

The contributions of this paper:

– we propose the Bark-scaled Gauss and linear filter bank superposition acoustic feature extraction method, which compensates for the weakness of the sparse filter and the sparse feature in the high-frequency information for a short utterance feature, this can improve the performance of short utterance speaker recognition by providing rich timbre information;

– we propose the multi-dimensional central difference method for capturing the dynamic features of speakers, which is used to simulate real speech and enhance the diversity of acoustic features with limited speech data.

1.3. Organization

This paper is organized as follows. Section 2 details the proposed filter superposition-based multi-dimensional central difference discriminative acoustic feature extraction method. Then we analyze the experiments and results of the proposed method in Sec. 3. Finally, the conclusion is given in Sec. 4.

2. Proposed method

In this section, which mainly includes the discriminative acoustic feature extraction algorithm, we elaborate on the proposed feature extraction technique, which the design based on the Bark-scaled Gauss and linear filter banks superposition algorithm and then the multi-dimensional central difference dynamic features extraction method based on the BGLCC features matrix. In addition, the effect of the introduced feature extraction of BGLCC and MDCD was achieved through mathematical analysis.

2.1. BGLCC feature extraction method

The speech signal is performed by a high-pass filter as pre-emphasis, this filter is equivalent to:

$$H(z) = 1 - az^{-1}, \tag{1}$$

where a is a pre-emphasis coefficient, the value is chosen in the interval [0.95, 0.97] and it can increase the energy of higher frequencies.

The following Hamming window w is used for smoothing the edge of framed speech signals:

$$w(k) = \left[0.54 - 0.46\cos\left(\frac{2\pi k}{K-1}\right)\right] R_K(k), \tag{2}$$

where $K-1$ is the window length, $K-1$ equals 256, $0 \le k \le K-1$, $R_K(k)$ equals rectangular window.

In speech processing, the Bark-frequency cepstrum (BFC) affects the speech short-term power spectrum, which is transformed on the Bark-scale of frequency. The BFC can be obtained as:

$$F_{\text{Bark}}(f) = 13\tan^{-1}\left(\frac{0.76f}{1000}\right) + 3.5\tan^{-1}\left(\frac{f}{7500}\right)^2. \tag{3}$$

In contrast to the well-known Mel-scaled triangular filter, the proposed Bark-scaled Gauss filter structure has a smoother response and enhances the correlation between adjacent sub-bands. The coefficients are derived from a type of cepstral representation of the speech clip. The frequency response of the Bark-scaled Gauss filter bank can be obtained as:

$$H_{\text{Bark}_b}(k) = \frac{1}{\sqrt{2\pi}\sigma_b} e^{\frac{-[k-f(b)]^2}{2\sigma_b^2}}, \tag{4}$$

where σ_b is the standard deviation, and $f(b)$ is the b-th filter boundary point (Bark-scaled center frequency), as defined:

$$\sigma_b = \frac{f(b+1) - f(b)}{\alpha}, \tag{5}$$

where α is equal to 2.0.

The signal presents 24 critical bands in the band, which is also the Bark center frequency, and this is the Bark domain.

Next, the linear triangle filter bank processing details. The power spectrum is then processed, on the frequency, by a linear uniform filter bank. In these linear filter banks, each filter is a triangle filter. The filter can be defined as:

$$H_{\mathrm{Linear}_l}(k) = \begin{cases} 0 & k < f(l-1), \\ \dfrac{k - f(l-1)}{f(l) - f(l-1)} & f(l-1) \le k < f(l), \\ 1 & k = f(l), \\ \dfrac{f(l+1) - k}{f(l+1) - f(l)} & f(l) < k \le f(l+1), \\ 0 & k > f(l+1), \end{cases} \tag{6}$$

where $f(l)$ is the center frequency, $0 \le l < L$, and L is the number of filter banks, and the value of L is 24. We use more filter bands than usual on account that the resolution of high-frequency domains is essential for the timbre. Finally, we get the linear filter features.

The raw speech signal $x(n)$ is preprocessed to obtain $x_w(n)$. Subsequently, the fast Fourier transform of the framed speech signal to transform the speech data from the time domain to the frequency domain, the mathematical calculation can be written as:

$$X(i,n) = \mathrm{FFT}[x_w(i,n)], \tag{7}$$

where $x_w(n)$ indicates that after adding the window function i is the number of speech frames.

The power spectrum is calculated as:

$$E(i,n) = |X(i,n)|^2. \tag{8}$$

Therefore, the Bark-scaled Gauss and linear filter banks superposition feature extraction is made based on the power spectrum of the output from the fast Fourier transform. Thus, the BGLCC power calculation procedure can be given by:

$$S(i,t) = \sum_{k=0}^{N-1} E(i,n)[H_{\mathrm{Bark}_b}(k) + H_{\mathrm{Linear}_l}(k)], \qquad 0 \le b \le u, \qquad 0 \le l \le v, \tag{9}$$

where t denotes the t-th superposition filter, b denotes the b-th Bark-scaled Gauss filter, and l denotes the l-th linear triangle filter, respectively, u is the number of the Bark-scaled Gauss filter, v is the number of the linear triangle filter, t, u, v all are 48-channel filter banks; $S(i,t)$ is equivalent to multiplying the power spectrum $E(i,n)$ and the superposition of $H_{\mathrm{Bark}_b}(k)$, the Bark-scaled Gauss filter and $H_{\mathrm{Linear}_l}(k)$ the linear triangle filter on the frequency domain.

$$\mathrm{BGLCC}(i,t) = \sum_{t=0}^{T-1} \log[S(i,t)] \cos\left[\frac{\pi r(2t-1)}{2T}\right], \tag{10}$$

where $S(i,t)$ is the BGLCC power, i denotes the i-th frame, r is the spectral line after discrete cosine transformation, t denotes the t-th superposition filter, T is the number of superposition filters, and the value of T is 48.

The Bark-scaled Gauss and linear filter bank superposition features (BGLCC) are processed as shown in Fig. 1.

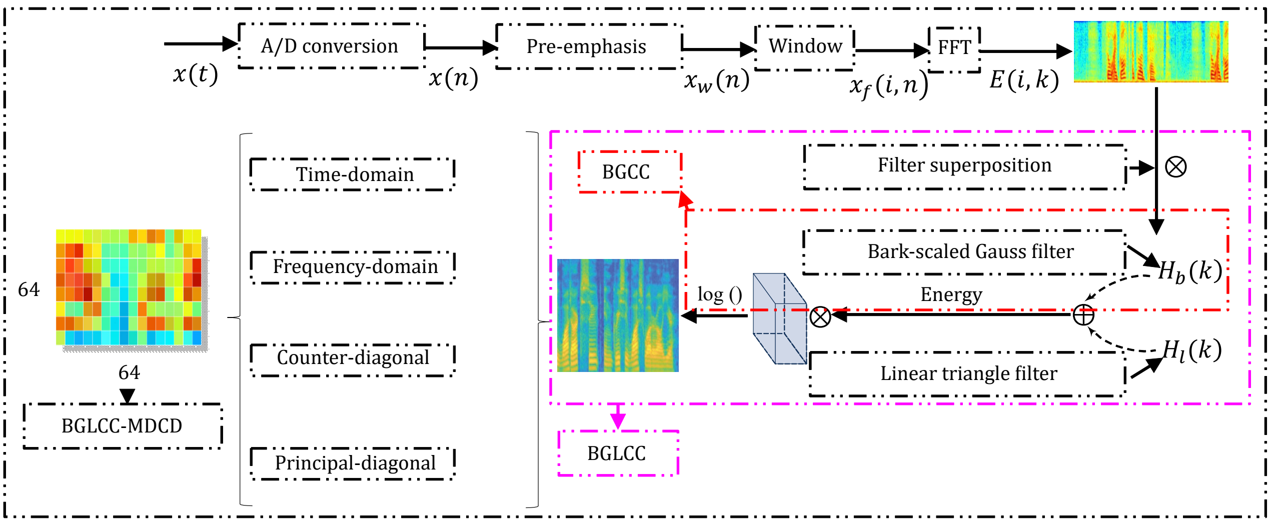

Fig. 1. Structure of the proposed acoustic features extraction method.

2.2. MDCD dynamic feature extraction method

The proposed multi-dimensional central difference dynamic feature extraction method was applied to the different dimensions of the BGLCC time-frequency matrix, where the horizontal dimension is the time domain axis and the vertical is the frequency domain axis dimension and it captures speech time-domain relevance and speech high-low-frequency correlation of the speaker. Similarly, the central difference of linear regression is applied to the time-frequency matrix principal diagonal and counter diagonal, therefore it can capture the voiceprint of the speaker.

The process of the proposed method is shown in Fig. 1; MDCD dynamic feature extraction of different dimensions on the BGLCC time-frequency matrix. First, a series of pre-processing is performed on a frame of the speech signal, which converts the input signal from a time-domain speech signal to a frequency-domain speech signal. Next, the proposed Bark-scaled Gauss and linear filter bank features superposition is applied to divide the spectrum into certain frequency bands, and the log compression is applied. Then, multi-dimensional central difference obtains four different types of features based on the BGLCC time-frequency matrix, which are calculated as in Eqs. (11)–(14):
time-domain:

$$T_h = M^t_{t,f} = \frac{M_{t+1,f} - 2M_{t,f} + M_{t-1,f}}{h^2}, \tag{11}$$

frequency-domain:

$$F_h = M^f_{t,f} = \frac{M_{t,f+1} - 2M_{t,f} + M_{t,f-1}}{h^2}, \tag{12}$$

counter-diagonal domain:

$$P_h = M^P_{t,f} = \frac{M_{t+1,f+1} - 2M_{t,f} + M_{t-1,f-1}}{h^2}, \tag{13}$$

principal-diagonal domain:

$$C_h = M^C_{t,f} = \frac{M_{t+1,f-1} - 2M_{t,f} + M_{t-1,f+1}}{h^2}. \tag{14}$$

In these equations, the value of h is 2, as the central difference of linear regression has been applied. Here, t stands for the time domain axis and f stands for the frequency domain axis. M is the point along which different dimensions of the axis have been applied.

The time domain's central difference and the frequency domain's central difference can better capture the contour of the speaker formants. By doing the matrix principal diagonal's central difference and matrix counter diagonal's central difference, speaker information about the uttering text phoneme of each speaker can be captured. Thus, the different dimensions of the time-frequency spectrum central difference can be regarded as multi-dimensional dynamic speaker information of each speaker and this explains the excellent results of the proposed MDCD features. To reduce the computationally derived high-dimensional MDCD features, we compress and decorrelate them by DCT.

It was our goal to perform speaker verification through the proposed BGLCC-MDCD as acoustic features, and use 34-layer ResNet as the backbone model, to perform the short-duration speaker verification. The detailed configuration is listed in Table 1.

Table 1. Detailed configuration of the backbone model of 34-layer ResNet. The input size is $T \times 64$.

Layer	Structure	Output shape
Conv0	CNN (7 × 7, 32), stride 2	$T \times 64 \times 32$
Conv1	$\begin{pmatrix}(3\times3,32)\\(3\times3,32)\end{pmatrix}$ × 3, stride 2	$T/2 \times 32 \times 32$
Conv2	$\begin{pmatrix}(3\times3,64)\\(3\times3,64)\end{pmatrix}$ × 4, stride 2	$T/2 \times 16 \times 64$
Conv3	$\begin{pmatrix}(3\times3,128)\\(3\times3,128)\end{pmatrix}$ × 6, stride 2	$T/2 \times 8 \times 128$
Conv4	$\begin{pmatrix}(3\times3,256)\\(3\times3,256)\end{pmatrix}$ × 3, stride 2	$T/2 \times 4 \times 256$

3. Experiments and analysis

3.1. Experiments

The short-duration speaker verification experiments presented in this paper are conducted using the three well-known speaker recognition datasets with different scenarios: VoxCeleb (Nagrani *et al.*, 2017; Chung *et al.*, 2018), Speaker in the Wild (SITW) (McLaren *et al.*, 2016), and the NIST SRE 2010 (Martin, Greenberg, 2010) to evaluate our proposed algorithm.

The short-duration text-independent dataset is generated from the VoxCeleb, SITW, and NIST SRE corpus, respectively. After removing silence frames using an energy-based VAD, the speech utterances are chopped into short segments (ranging from 0.25 to 10 seconds). This is to illustrate the efficiency of the work of our proposed method under short-duration audio conditions.

The three different scenarios of speech datasets: VoxCeleb, SITW, and NIST SRE corpus aim to evaluate the generalizability of the methods across a range of different audio lengths of scenarios. We focus on conducting speaker verification trials on voice samples of different speech lengths, which are used to investigate the effect of testing speech sample length changes and to validate the efficiency of the presented method on the performance of the speaker verification method. One thing to keep in mind is that in all of our tests, we assume that there is only one speaker in each voice sample and that there is no overlapping voice from several speakers in any of the training or testing speeches.

3.2. Corpus description

3.2.1. VoxCeleb and SITW corpus

VoxCeleb is a large open-source speaker recognition dataset with over a million utterances, 7000 speakers, and 2000 hours of audio. The average duration of utterances in the VoxCeleb dataset is 8 seconds, and the majority of utterances have a duration of fewer than 10 seconds. The audio sampling rate is 16kHz. VoxCeleb includes two sub-datasets, VoxCeleb-1 and VoxCeleb-2. The SITW dataset contains open-source media recordings of 299 public celebrities. The SITW dataset is used to generate the short-duration text-independent dataset. SITW speech segments range in length from 6 seconds to 180 seconds, where the majority are long utterances. As a result, the two datasets can be used to assess the performance of our proposed architectures on utterances of varying lengths as well as the model's generalizability.

Each of the three datasets, VoxCeleb-1, VoxCeleb-2, and SITW, is divided into two parts: development and testing (evaluation). The training set consists of 1 092 009 utterances and 5994 speakers from the VoxCeleb-2 development part (VoxCeleb2-Dev). The remaining datasets were treated as test sets, with two parts: the VoxCeleb-1 dataset and the SITW evaluation (SITW-Eval) set. There are 4706 utterances and 37 611 trials in the VoxCeleb-1. There are 1202 utterances and 721 788 trials in the SITW evaluation (SITW-Eval).

3.2.2. NIST SRE corpus

The NIST SRE corpus was used to generate the short-duration text-independent dataset. The SRE04-08, Switchboard II phase 2, 3, and Switchboard Cellular Part 1, Part 2 comprise the training set. The final training set includes 4000 speakers with 40 short utterances each. Similarly, the enrollment and test sets are derived from NIST SRE 2010. The enrollment speech includes 150 male and 150 female speakers, each of whom is enrolled by five utterances. The 4500 utterances in the enrollment speech data are used to test from the same 300 speakers. The trial list that was generated contains 392 660 trials. The website GitHub provides access to the trial list and the comprehensive segmentation files.

3.3. Feature extraction

All experiments use a 64-dimensional input feature from a 25 ms window with a 10 ms frameshift. The experiments evaluate using features: LPCC, MFCC, MFCC-LPCC, the proposed BGCC, BGLCC, and BGLCC-MDCD. The 64-dimensional features were extracted for LPCCs, with 32 for linear regression along the time axis and 32 along the frequency axis. The MFCCs used 64-dimensional features, and the 64-dimensional MFCC-LPCC features contain 32-dimensional MFCC and LPCC features, respectively. The use of delta 1/2 inputs is also a 64-dimensional feature. For the proposed acoustic feature, BGCC, BGLCC, the 64-dimensional feature vector has been extracted, BGCC-MDCD, BGLCC-MDCD, which contain 16 time-domain features, 16 frequency-domain features, 16 counter-diagonal domain features, 16 principal-diagonal domain features, respectively.

3.4. Loss function

In (Schroff *et al.*, 2015), the triplet loss was initially proposed to learn discriminatory image embedding. The embeddings need to satisfy the following relationship for model training to be successful. The cosine triplet embedded *Loss* (Zhang *et al.*, 2018) for training the model is:

$$\left\|f(s_i^a) - f(s_i^p)\right\|_2^2 + \alpha_{\text{margin}} < \left\|f(s_i^a) - f(s_i^n)\right\|_2^2, \qquad \forall\left(f(s_i^a), f(s_i^p), f(s_i^n)\right) \in \tau, \tag{15}$$

$$L = \sum_i^N \left[\left\|f(s_i^a) - f(s_i^p)\right\|_2^2 - \left\|f(s_i^a) - f(s_i^n)\right\|_2^2 + \alpha_{\text{margin}}\right]. \tag{16}$$

The cosine triplet embedding the loss function L is used here, where τ is the batch of triplet, with (s_i^a, s_i^p, s_i^n) is a triplet. N is the batch size. Samples of speech from a specific "a" are s_i^a, the anchor sample, and s_i^p, the positive sample with the same person. The negative sample, s_i^n, is a sample of speech from another person "b", so that $a \neq b$. The α_{margin} is a user-tunable hyper-parameter at the value of 0.25 that determines the minimum distance between negative and positive speech samples.

3.5. Implementation and reproducibility

The proposed discriminative acoustic feature method uses the PyTorch (Paszke *et al.*, 2017) toolkit to conduct the experiment, and training using the Triplet-loss (Schroff *et al.*, 2015). The initial learning rate is 0.001 and lasts for 200 epochs. The experiment embeds the cosine triplet loss, and the value of the α_{margin} hyper-parameter is 0.25, which is the best trade-off. The network is optimized using the Adam optimizer with a minibatch size of 32 and softmax as a classifier. The fully connected layers after the statistic pooling layer have 512 nodes. The training was done on a single Nvidia A100 GPU.

3.6. Evaluation metrics

We use the following metrics to evaluate the model performance: the Equal Error Rate (EER, in %), and the minimum detection cost function at the prior probability of specifying the targeted speaker of (Min-DCF*100), which is a standard-setting (Nagrani *et al.*, 2017), and partial AUC (pAUC) with $\alpha = 0$ and $\beta = 0.05$, the pAUC represents the partial area under the ROC curve, it meets the evaluation requirement of real-world applications that work on different parts of ROC curves. It is a supplement evaluation metric to the existing metrics. The pAUC is defined by two false positive rate (FPR) parameters: α and β, which is a detailed calculation (Bai *et al.*, 2020). The pAUCMetric evaluates the similarity between two speaker features by the squared Mahalanobis distance.

3.7. Results and analysis

3.7.1. Overall performance

Performance comparison of different acoustic features. Table 2 and Fig. 2 show the performance of our proposed acoustic features and the compared acoustic features on VoxCeleb-1, SITW, and NIST SRE 2010 datasets, respectively. Table 2 lists the results in terms of EER, Min-DCF, and pAUC, Fig. 2 plots the detection error trade-off (DET) curves of different acoustic features under 10 s speech length that include no dynamic features, using delta 1/2 dynamic features and using MDCD dynamic features. The acoustic feature extraction level for the short-duration audio signal, contains three conventional baseline features, which are MFCC, LPCC, MFCC-LPCC, and our proposed BGCC and BGLCC acoustic features. The speech length ranges from 0.25 to 10 seconds, including 3 segments.

From Table 2, on VoxCeleb-1, SITW, and NIST SRE 2010 datasets, it can be observed that BGLCC-MDCD acoustic feature significantly outperforms MFCC, LPCC, and MFCC-LPCC in terms of EER, Min-DCF, and pAUC, and BGLCC-MDCD acoustic feature achieves better performance in short-duration speaker verification.

Across the LPCC experiment in Table 2, on the VoxCeleb-1 dataset, compared to LPCC features, the proposed BGLCC features improve by 15.0%, compared to LPCC-delta1/2 features, BGLCC-MDCD features improve 19.0%, under 2 s duration speech length in terms of EER.

Across the MFCC experiment in Table 2, on the VoxCeleb-1 dataset, compared to MFCC features, the proposed BGLCC features improve by 10.6%, compared to MFCC-delta1/2 features, BGLCC-MDCD features improve 15.0%, under 2 s duration speech length in terms of EER.

Across the MFCC-LPCC experiment in Table 2, on the VoxCeleb-1 dataset, compared to MFCC-LPCC features, the proposed BGLCC features improve by 9.1%, compared to MFCC-LPCC-delta1/2 features, BGLCC-MDCD features improve 13.3%, under 2 s duration speech length in terms of EER.

At the same time, on the other speech with different lengths from VoxCeleb-1, SITW, and NIST SRE 2010 datasets, the proposed BGLCC-MDCD acoustic features for short-duration speaker verification achieve better performance, compared with conventional MFCC, LPCC, and MFCC-LPCC fusion acoustic features. The comparison of the performance of the baseline is shown in Table 2.

In order to visualize the effectiveness of our proposed acoustic features on the different length speech, we plot detection error trade-off (DET) curves for all comparable features, as illustrated in Fig. 2. The performance advantage of proposed BGLCC and MDCD can also be seen from the DET curves in Fig. 2. For example, the results of experiment 1 present the DET curves of the LPCC acoustic feature under three conditions: no dynamic features, using delta 1/2 dynamic features, and using our MDCD dynamic features, under 10 s speech length on the VoxCeleb-1 dataset; the results of experiment 2 present the DET curves of the LPCC acoustic feature under three conditions: no dynamic features, using delta 1/2 dynamic features, and using our MDCD dynamic features, under 10 s speech length on the SITW dataset; the results of experiment 3 present the DET curves of the LPCC acoustic feature under three conditions: no dynamic features, using delta 1/2 dynamic features, and using our MDCD dynamic features, under 10 s speech length on the NIST SRE 2010 dataset.

Similarly, experiments 4–6 represent the DET curves of the MFCC acoustic feature under three conditions, on VoxCeleb-1, SITW, and NIST SRE 2010 datasets, respectively; experiments 7–9 represent the DET curves of the MFCC-LPCC acoustic feature under three conditions, on VoxCeleb-1, SITW, and NIST SRE 2010 datasets, respectively; experiments 10–12 represent the DET curves of the BGCC acoustic feature under three conditions, on VoxCeleb-1, SITW, and NIST SRE 2010 datasets, respectively; and experiments 13–15 represent the DET curves of the BGLCC acoustic feature under three conditions, on VoxCeleb-1, SITW, and NIST SRE 2010 datasets, respectively.

The experimental results also show the lower DET curves achieved using our proposed MDCD dynamic features, compared to no dynamic features, and using delta 1/2 dynamic features on VoxCeleb-1, SITW, and NIST SRE 2010 datasets.

The proposed MDCD dynamic acoustic feature achieves lower EER, Min-DCF, and highest pAUC than delta 1/2, thus demonstrating that the proposed multi-dimensional central difference dynamic features perform better and are more effective than single-dimensional dynamic features. The results of that comparison are listed in Table 2.

Table 2. Comparison results of different acoustic features and proposed acoustic features under varying audio lengths using the ResNet-34 network on VoxCeleb-1, SITW, and NIST SRE 2010 datasets.

Features	Delta2	MDCD	Duration [s]	VoxCeleb-1			SITW			NIST SRE 2010		
				EER [%]	MinDCF	pAUC [%]	EER [%]	MinDCF	pAUC [%]	EER [%]	MinDCF	pAUC [%]
LPCC	–	–	0.25	11.19	32.94	75.38	13.22	36.31	70.52	12.01	34.17	74.43
	√	–		11.18	32.92	75.39	13.20	36.29	70.54	12.00	34.15	74.44
	–	√		**11.13**	**32.83**	**75.46**	**13.15**	**36.24**	**70.66**	**11.95**	**34.09**	**74.54**
	–	–	2	3.17	17.99	95.38	5.53	23.41	92.37	4.48	23.03	93.45
	√	–		3.16	17.98	95.39	5.52	23.40	92.40	4.46	23.01	93.47
	–	√		**3.11**	**17.92**	**95.46**	**5.47**	**23.35**	**92.46**	**4.40**	**22.95**	**93.58**
	–	–	10	1.61	10.33	98.01	3.60	19.17	94.96	2.54	12.10	96.73
	√	–		1.60	10.32	98.03	3.58	19.16	94.98	2.52	12.09	96.75
	–	√		**1.54**	**10.27**	**98.10**	**3.53**	**19.10**	**95.04**	**2.46**	**12.02**	**96.84**
MFCC	–	–	0.25	11.04	32.52	75.74	12.52	35.83	72.23	11.50	32.93	75.01
	√	–		11.02	32.51	75.75	12.51	35.82	72.26	11.48	32.92	75.03
	–	√		**10.98**	**32.47**	**75.79**	**12.45**	**35.74**	**72.44**	**11.44**	**32.86**	**75.12**
	–	–	2	3.01	17.90	95.40	4.46	23.01	93.53	3.33	18.07	95.27
	√	–		3.00	17.88	95.41	4.45	23.00	93.54	3.32	18.05	95.28
	–	√		**2.95**	**17.84**	**95.60**	**4.40**	**22.95**	**93.58**	**3.27**	**17.99**	**95.31**
	–	–	10	1.37	10.12	98.34	3.24	17.96	95.32	2.11	10.83	97.62
	√	–		1.37	10.11	98.35	3.23	17.95	95.33	2.10	10.81	97.64
	–	√		**1.36**	**10.04**	**98.40**	**3.19**	**17.66**	**95.51**	**2.05**	**10.75**	**97.74**
MFCC-LPCC	–	–	0.25	10.97	32.47	75.82	12.42	35.73	72.47	11.41	32.82	75.17
	√	–		10.96	32.46	75.83	12.41	35.72	72.49	11.40	32.81	75.19
	–	√		**10.90**	**32.42**	**75.86**	**12.34**	**35.70**	**72.58**	**11.35**	**32.75**	**75.30**
	–	–	2	2.96	17.79	95.44	4.37	21.98	93.73	3.28	18.01	95.30
	√	–		2.94	17.78	95.45	4.35	21.97	93.74	3.27	17.99	95.31
	–	√		**2.88**	**17.71**	**95.66**	**4.30**	**21.92**	**93.76**	**3.21**	**17.92**	**95.35**
	–	–	10	1.36	9.92	98.36	3.17	17.99	95.38	2.05	10.75	97.72
	√	–		1.35	9.91	98.38	3.16	17.98	95.39	2.04	10.73	97.74
	–	√		**1.34**	**9.82**	**98.42**	**3.11**	**17.92**	**95.46**	**1.99**	**10.70**	**97.89**
BGCC	–	–	0.25	10.98	32.47	75.79	12.43	35.74	72.46	11.42	32.84	75.14
	√	–		10.97	32.46	75.80	12.42	35.73	72.48	11.40	32.83	75.15
	–	√		**10.91**	**32.42**	**75.84**	**12.35**	**35.71**	**72.57**	**11.36**	**32.77**	**75.28**
	–	–	2	2.96	17.79	95.42	4.38	22.00	93.64	3.28	18.01	95.30
	√	–		2.95	17.78	95.44	4.37	21.98	93.65	3.27	17.99	95.31
	–	√		**2.90**	**17.72**	**95.64**	**4.31**	**21.92**	**93.69**	**3.23**	**17.94**	**95.34**
	–	–	10	1.36	9.93	98.35	3.18	18.01	94.36	2.06	10.77	97.72
	√	–		1.36	9.92	98.36	3.17	17.99	94.38	2.05	10.75	97.74
	–	√		**1.35**	**9.84**	**98.41**	**3.13**	**17.94**	**95.44**	**2.00**	**10.72**	**97.84**
BGLCC	–	–	0.25	10.71	31.95	75.84	12.26	34.91	72.57	11.11	32.64	75.34
	√	–		10.70	31.94	75.85	12.25	34.90	72.58	11.10	32.62	75.36
	–	√		**10.58**	**31.42**	**75.89**	**12.05**	**34.62**	**72.64**	**10.95**	**32.22**	**75.84**
	–	–	2	2.69	17.03	95.63	4.16	21.88	93.73	3.06	17.91	95.35
	√	–		2.67	17.02	95.64	4.15	21.87	93.74	3.05	17.91	95.36
	–	√		**2.55**	**16.95**	**95.72**	**3.99**	**21.02**	**93.78**	**2.86**	**17.57**	**95.72**
	–	–	10	1.34	9.82	98.42	2.96	17.79	95.42	1.87	10.66	97.85
	√	–		1.34	9.82	98.43	2.95	17.78	95.44	1.85	10.64	97.87
	–	√		**1.32**	**9.37**	**98.48**	**2.66**	**16.82**	**95.94**	**1.63**	**10.34**	**98.00**

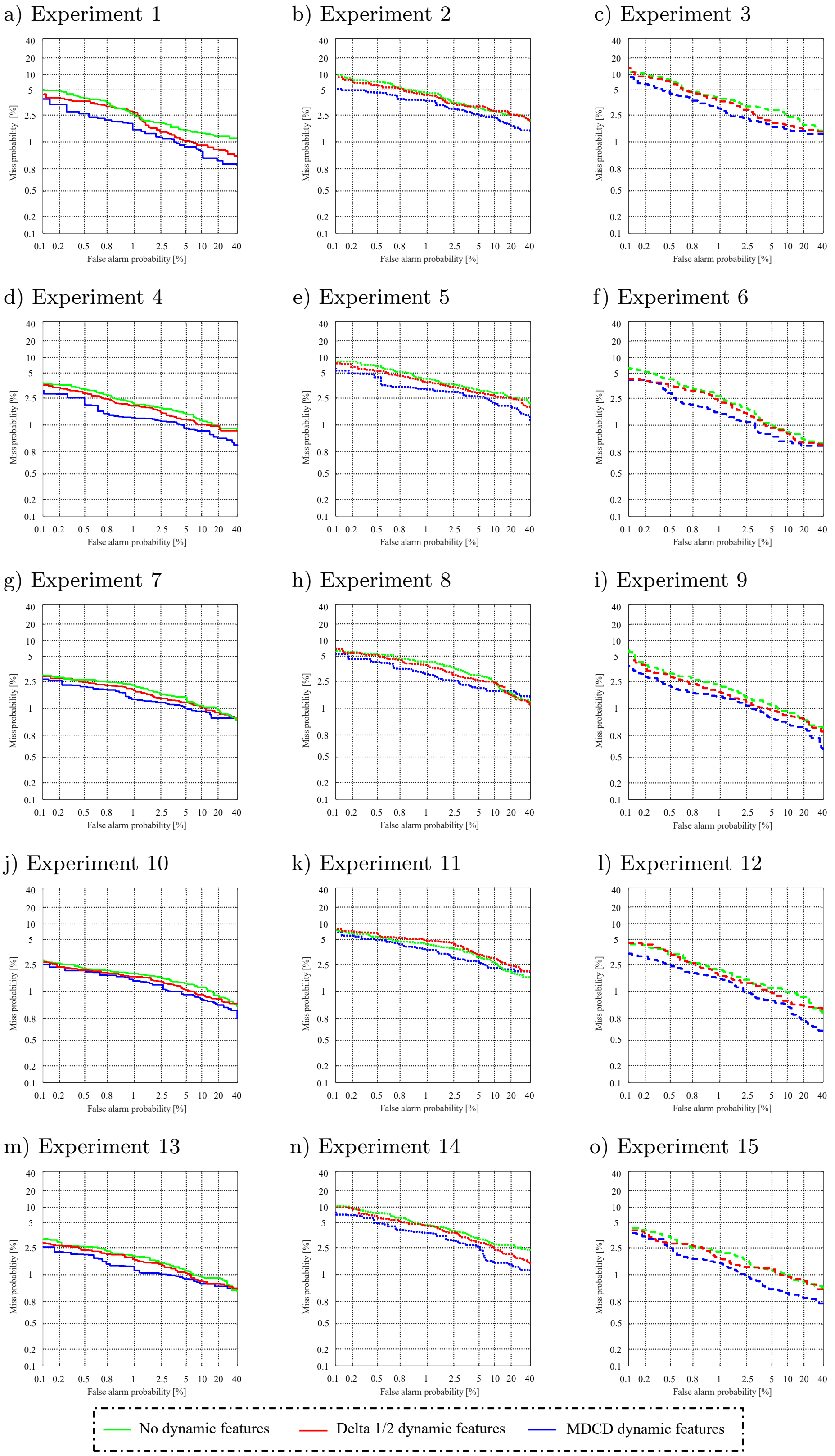

Fig. 2. The DET curves of different acoustic features and different dynamic features for speaker verification under varying audio lengths using the ResNet-34 model on VoxCeleb-1, SITW, and NIST SRE 2010 datasets. The experiments 1 to 3, the DET curves indicate, that on VoxCeleb-1, SITW, and NIST SRE 2010, under 10 s speech length, the LPCC uses no dynamic features, delta 1/2 dynamic features, and MDCD dynamic features, respectively. Similarly, experiments: 4–6 represent the MFCC method, 7–9 represent the MFCC-LPCC method, 10–12 represent the BGCC method, and 13–15 represent the BGLCC method.

At the same time, in the experiments comparing the different attributes of source information combination for short-duration speaker recognition (Das *et al.*, 2016), the proposed multi-source discriminative acoustic feature achieves consistent performance benefits across short-duration speech dataset experiments.

3.7.2. Ablation experiments

To evaluate each component of the BGLCC-MDCD feature, we conducted several ablation experiments on VoxCeleb-1, SITW, and NIST SRE 2010 datasets, where the results are shown in Tables 2 and 3, and Figs. 2 and 3.

First, we evaluate the effectiveness of our proposed enhancement of discriminative acoustic features. Table 2 lists the EER, Min-DCF, and pAUC results of different features on VoxCeleb-1, SITW, and NIST SRE 2010 datasets. From Table 2, it can be observed that the proposed acoustic feature vastly outperforms the baseline feature, and it is seen from Fig. 2 that the DET curve of using MDCD dynamic features is lower than that without dynamic features, and using delta 1/2 dynamic features. The main reason for the performance improvement is our proposed BGLCC feature which employs the Bark-scaled Gauss and the linear filter bank superposition methods, it can remedy the weakness of the sparse high-frequency information and insufficient acoustic feature extraction by enhancing more high-frequency domain information. Similarly, MDCD through four different dimension differences captures better dynamics features of voiceprints, and it can further compensate for the limited and sparse dynamic acoustic features of short-duration audio signals. The experimental results also prove this.

To verify that different multi-dimensional central differences can capture dynamic features of the voiceprint, we conducted several ablation experiments,

Table 3. Ablation study for different multi-dimensional dynamic features based on BGLCC under varying audio lengths using the ResNet-34 network on VoxCeleb-1, SITW, and NIST SRE 2010 datasets.

Methods	Duration [s]	VoxCeleb-1			SITW			NIST SRE 2010		
		EER [%]	MinDCF	pAUC [%]	EER [%]	MinDCF	pAUC [%]	EER [%]	MinDCF	pAUC [%]
MDCD-T_h	0.25	10.67	31.90	75.88	12.19	34.80	72.62	11.04	32.29	75.43
	2	2.65	16.99	95.68	4.08	21.72	93.81	3.01	17.89	95.42
	10	1.32	9.71	98.46	2.91	15.69	95.96	1.83	10.62	97.89
MDCD-F_h	0.25	10.68	31.91	75.87	12.21	34.82	72.61	11.05	32.31	75.42
	2	2.66	17.00	95.67	4.10	21.81	93.79	3.02	17.88	95.40
	10	1.33	9.72	98.44	2.92	17.70	95.94	1.84	10.63	97.88
MDCD-P_h	0.25	10.70	31.94	75.85	12.25	34.85	72.59	11.09	32.34	75.38
	2	2.68	17.02	95.65	4.14	21.84	93.76	3.05	17.90	95.37
	10	1.35	9.82	98.42	2.95	17.73	95.37	1.86	10.65	97.86
MDCD-C_h	0.25	10.69	31.93	75.86	12.23	34.84	72.60	11.07	32.32	75.39
	2	2.67	17.01	95.66	4.12	21.83	93.77	3.04	17.89	95.38
	10	1.34	9.81	98.43	2.94	17.72	95.38	1.85	10.64	97.87
MDCD	0.25	**10.58**	**31.42**	**75.89**	**12.05**	**34.62**	**72.64**	**10.95**	**32.22**	**75.84**
	2	**2.55**	**16.95**	**95.72**	**3.99**	**21.02**	**93.78**	**2.86**	**17.57**	**95.72**
	10	**1.32**	**9.37**	**98.48**	**2.66**	**16.82**	**95.94**	**1.63**	**10.34**	**98.00**

a) Experiment (EER, VoxCeleb-1) b) Experiment (EER, SITW) c) Experiment (EER, NIST SRE 2010)

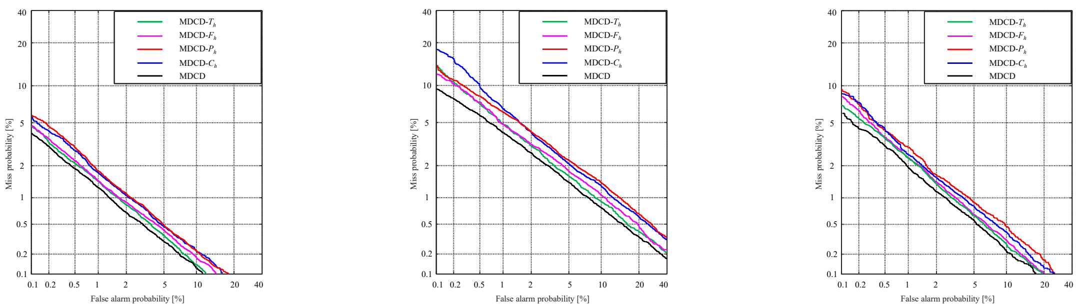

Fig. 3. DET curves of different dimensional dynamic features on VoxCeleb-1 (a), SITW (b), and NIST SRE 2010 (c) datasets under 10 s duration speech using the ResNet-34 model.

where the results are shown in Table 3 and Fig. 3. Compared to the diagonal domain, the time-frequency domain central difference captures better dynamic features, and the MDCD achieves the lower EER and Min-DCF. Figure 3 visualizes the DET curve of each dimension branch under the 10 s length utterance. The time-frequency domain performs better than the diagonal domain which is since the signal is mainly analyzed in the time-frequency domain.

Hence, the proposed BGLCC-MDCD discriminative acoustic features are the key reasons for the performance improvement in short utterance speaker verification, which: (a) extracts speaker-reliant characteristics successfully, from the BGLCC features to remedy the weakness of insufficient acoustic features to solve the problem of less emphasizes high-frequency information from the conventional acoustic feature extraction filter design; (b) then, the MDCD method can capture better dynamics features of voiceprints from short-duration audio signals.

4. Conclusion

In this paper, we propose the Bark-scaled Gauss and the linear filter bank superposition acoustic features extraction methods to enhance high-frequency domain information of short-duration audio, to deal with the problem of the high-frequency band feature sparsity. Compared with traditional acoustic features such as MFCC, LPCC, etc., our proposed BGLCC feature extraction method emphasizes a focus on both the low-high frequency band of speech, which is more helpful in extracting more discriminative acoustic features to compensate the sparsity of the effective information. Furthermore, a multi-dimensional central difference dynamic acoustic feature is proposed following the BGLCC spectrum characteristics, aiming to capture more diverse dynamic information. The MDCD feature concatenate information of the speaker from four different dimensions, this can explain why it performs significantly better than traditionally used speech features in short utterance speaker verification tasks under various conditions.

The proposed methods are evaluated on well-known datasets, VoxCeleb-1, SITW, and NIST SRE 2010 corpus. From the experimental results, the proposed method achieves continuous improvement over traditional acoustic features in all test sets. The ablation experiments further indicate that the proposed approaches substantially improve the enhanced discriminant features for speaker verification tasks. Future work involves the combination of acoustic feature-based and model-based compensations for short-duration speech speaker verification, and to improve the performance, accuracy, and richness of acoustic feature extraction in short-duration audio signals.

Acknowledgments

This work was in part supported by NSFC (grants nos. 62176194, 62101393), the Major project of IoV (grant no. 2020AAA001), Sanya Science and Education Innovation Park of the Wuhan University of Technology (grant no. 2021KF0031), CSTC (grant no. cstc2021jcyj-msxmX1148), and the Open Project of the Wuhan University of Technology Chongqing Research Institute (ZL2021-6).

References

1. Atal B.S. (1974), Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification, *The Journal of the Acoustical Society of America*, **55**(6): 1304–1312, doi: 10.1121/1.1914702.

2. Bai Z., Zhang X.-L., Chen J. (2020), Speaker verification by partial AUC optimization with Mahalanobis distance metric learning, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **28**: 1533–1548, doi: 10.1109/TASLP.2020.2990275.

3. Campbell J.P. (1997), Speaker recognition: A tutorial, *Proceedings of the IEEE*, **85**(9): 1437–1462, doi: 10.1109/5.628714.

4. Chowdhury A., Ross A. (2020), Fusing MFCC and LPC features using 1D triplet CNN for speaker recognition in severely degraded audio signals, *IEEE Transactions on Information Forensics and Security*, **15**: 1616–1629, doi: 10.1109/TIFS.2019.2941773.

5. Chung J.S., Nagrani A., Zisserman A. (2018), Voxceleb2: Deep speaker recognition, [in:] *Proceedings of Interspeech 2018*, pp. 1086–1090, doi: 10.21437/Interspeech.2018-1929.

6. Das R.K., Mahadeva Prasanna S.R. (2016), Exploring different attributes of source information for speaker verification with limited test data, *The Journal of the Acoustical Society of America*, **140**(1): 184, doi: 10.1121/1.4954653.

7. Dehak N., Dehak R., Glass J., Reynolds D., Kenny P. (2010), Cosine similarity scoring without score normalization techniques, [in:] *Proceedings of Odyssey 2010 – The Speaker and Language Recognition Workshop*.

8. Desplanques B., Thienpondt J., Demuynck K. (2020), ECAPA-TDNN: Emphasized channel attention, propagation and aggregation in TDNN based speaker verification, [in:] *Proceedings of Annual conference of the International Speech Communication Association 2020*, pp. 3830–3834, doi: 10.21437/Interspeech.2020-2650.

9. Greenberg C.S. *et al.* (2013), The 2012 NIST speaker recognition evaluation, [in:] *Proceedings of Interspeech 2013*, pp. 1971–1975, doi: 10.21437/Interspeech.2013-469.

10. Herrera-Camacho A., Zúñiga-Sainos A., Sierra-Martínez G., Tramgol-Curipe J., Mota-Montoya M., Jarquín-Casas A. (2019), Design and testing of a corpus for forensic speaker recognition using MFCC, GMM and MLE, [in:] *Proceedings of International Conference on Video, Signal and Image Processing 2019*, pp. 105–110, doi: 10.1145/3369318.3369330.

11. Huang L., Pun C.-M. (2020), Audio replay spoof attack detection by joint segment-based linear filter bank feature extraction and attention-enhanced DenseNet-BiLSTM network, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **28**: 1813–1825, doi: 10.1109/TASLP.2020.2998870.

12. Kenny P., Boulianne G., Dumouchel P. (2005), Eigenvoice modeling with sparse training data, *IEEE Transactions on Speech and Audio Processing*, **13**(3): 345–354, doi: 10.1109/TSA.2004.840940.

13. Kinnunen T., Li H. (2010), An overview of text-independent speaker recognition: From features to supervectors, *Speech Communication*, **52**(1): 12–40, doi: 10.1016/j.specom.2009.08.009.

14. Liu Z., Wu Z., Li T., Li J., Shen C. (2018), GMM and CNN hybrid method for short utterance speaker recognition, *IEEE Transactions on Industrial Informatics*, **14**(7): 3244–3252, doi: 10.1109/TII.2018.2799928.

15. Martin A.F., Greenberg C.S. (2010), The NIST 2010 speaker recognition evaluation, [in:] *Proceedings of Interspeech 2010*, pp. 2726–2729, doi: 10.21437/Interspeech.2010-722.

16. McLaren M., Ferrer L., Castan D., Lawson A. (2016), The speakers in the wild (SITW) speaker recognition database, [in:] *Proceedings of Interspeech 2016*, pp. 818–822, doi: 10.21437/Interspeech.2016-1129.

17. Nagrani A., Chung J.S., Zisserman A. (2017), VoxCeleb: A large-scale speaker identification dataset, [in:] *Proceedings of Interspeech 2017*, pp. 2616–2620, doi: 10.21437/Interspeech.2017-950.

18. Nosratighods M., Ambikairajah E., Epps J., Carey M.J. (2010), A segment selection technique for speaker verification, *Speech Communication*, **52**(9): 753–761, doi: 10.1016/j.specom.2010.04.007.

19. Omar M.K., Pelecanos J.W. (2010), Training universal background models for speaker recognition, [in:] *Proceedings of Odyssey 2010 – The Speaker and Language Recognition Workshop*, pp. 52–57.

20. Paseddula C., Gangashetty S.V. (2018), DNN based acoustic scene classification using score fusion of MFCC and inverse MFCC, [in:] *Proceedings of International Conference on Industrial and Information Systems 2018*, pp. 18–21, doi: 10.1109/ICIINFS.2018.8721379.

21. Paszke A. *et al.* (2017), Automatic differentiation in PyTorch, [in:] *Proceedings of NIPS 2017 Workshop*, pp. 1–4.

22. Povey D. *et al.* (2018), Semi-orthogonal low-rank matrix factorization for deep neural networks, [in:] *Proceedings of Interspeech 2018*, pp. 3743–3747, doi: 10.21437/Interspeech.2018-1417.

23. Schroff F., Kalenichenko D., Philbin J. (2015), FaceNet: A unified embedding for face recognition and clustering, [in:] *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2015*, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.

24. Snyder D., Garcia-Romero D., Sell G., McCree A., Povey D., Khudanpur S. (2019), Speaker recognition for multi-speaker conversations using x-vectors, [in:] *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing 2019*, pp. 5796–5800, doi: 10.1109/ICASSP.2019.8683760.

25. Todisco M., Delgado H., Evans N. (2017), Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification, *Computer Speech & Language*, **45**: 516–535, doi: 10.1016/j.csl.2017.01.001.

26. Villalba J. *et al.* (2020), State-of-the-art speaker recognition with neural network embeddings in NIST SRE18 and speakers in the wild evaluations, *Computer Speech & Language*, **60**: 101026, doi: 10.1016/j.csl.2019.101026.

27. Vogt R., Sridharan S., Mason M. (2010), Making confident speaker verification decisions with minimal speech, *IEEE Transactions on Audio, Speech, and Language Processing*, **18**(6): 1182–1192, doi: 10.1109/TASL.2009.2031505.

28. Wu Z., Yu Z., Yuan J., Zhang J. (2016), A twice face recognition algorithm, *Soft Computing – A Fusion of Foundations, Methodologies and Applications*, **20**(3): 1007–1019, doi: 10.1007/s00500-014-1561-9.

29. Yang H., Deng Y., Zhao H.-A. (2019), A comparison of MFCC and LPCC with deep learning for speaker recognition, [in:] *Proceedings of International Conference on Big Data and Computing 2019*, pp. 160–164, doi: 10.1145/3335484.3335528.

30. Zhang C., Koishida K., Hansen J.H.L. (2018), Text-independent speaker verification based on triplet convolutional neural network embeddings, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **26**(9): 1633–1644, doi: 10.1109/TASLP.2018.2831456.

31. Zinchenko K., Wu C.-Y., Song K.-T. (2017), A study on speech recognition control for a surgical robot, *IEEE Transactions on Industrial Informatics*, **13**(2): 607–615, doi: 10.1109/TII.2016.2625818.

Archives of Acoustics Vol. 49, No. 1, pp. 49–60 (2024), doi: 10.24425/aoa.2023.146824

Research Paper

Study on Mechanism and Suppression Method of Flow-Induced Noise in High-Speed Gear Pump

Peng ZHAN, Yan QIANG, Zhiyuan JIANG, Runxue YANG, Liejiang WEI*

Energy and Power Engineering College, Lanzhou University of Technology
Qilihe District, Lanzhou, Gansu Province, P.R. China

*Corresponding Author e-mail: weiliejiang@126.com

(*received December 15, 2022; accepted July 14, 2023; published online December 18, 2023*)

The flow-induced noise mechanism of a 5000 rpm high-speed gear pump is explored. On the basis of the CFD technology and the Lighthill acoustic analogy theory, a numerical model of the flow-induced noise of a high-speed gear pump is constructed, and the effect of oil suction pressure (0.1–0.2 MPa) on the internal flow field and flow-induced noise characteristics of the high-speed gear pump is investigated. To evaluate the accuracy of the numerical simulation, a noise testing platform for high-speed gear pumps was developed. Adding an oil replenishment groove to the high-speed gear pump suppresses its flow-induced noise. The results indicate that the discrete noise at the fundamental frequency and its harmonic frequency is the primary component of the flow-induced noise of the pump and that the oil-trapped area is the principal source of vibration. The overall sound pressure level of flow-induced noise in the inlet and outlet areas decreases with distance from the oil-trapped area, and the sound pressure level in the outlet area is greater than that in the inlet area. The oil replenishment groove may considerably minimize cavitation noise, enhance the oil absorption capacity, and reduce the outer field's overall sound pressure level by 4–5 dB.

Keywords: external gear pumps; flow-induced noise; the oil replenishment groove; flow pulsation rate.

1. Introduction

In volumetric pumps, gear pumps are widely used as power units in hydraulic systems due to their simple structure, strong anti-fouling ability, and low price. With the development of high-speed and high-pressure gear pumps, the problem of high noise induced by working is becoming more and more obvious (Woo, Vacca, 2020). The noise of the gear pump is composed of mechanical noise and flow-induced noise. The research shows that flow-induced noise is the major source of noise in high-speed gear pumps, which is mainly composed of cavitation noise, flow pulsation noise, oil-trapped noise, and turbulent noise (Fiebig, Wróbel, 2022). The mechanical vibration noise can be suppressed or eliminated by improving the processing and manufacturing accuracy, and there is research on the noise decomposition analysis (Wang *et al.*, 2016a). The flow-induced noise is due to its complex sound generation mechanism, the wide range of frequency involved, and the profound degree of harm, that we must pay attention to. The working speed of the gear pump is an important factor affecting its working stability. Excessive speed will aggravate the occurrence of cavitation and the phenomenon of trapped oil, causing serious vibration noise (Woo, Vacca, 2022). However, too low a speed will cause a low volumetric efficiency of the gear pump. Therefore, the working speed is usually set in the range of 600–3000 r/min. The high-speed gear pump studied in this topic has a high speed (3000–5000 r/min) and low oil suction pressure under actual working conditions. On the one hand, it is difficult for the oil to fill the tooth groove in time due to the centrifugal force generated by the high speed, resulting in serious gas accumulation at the tooth root and reducing the volumetric efficiency of the pump. Furthermore, it will also aggravate the cavitation of the pump and produce serious cavitation noise (Zhang *et al.*, 2021). Therefore, flow-induced noise is the main component of high-speed gear pump noise.

In this paper, the mechanism of flow-induced noise of a high-speed gear pump with a speed of 5000 r/min is analyzed. Based on the CFD technology and the Lighthill acoustic analogy theory, a numerical model of the flow-induced noise of a high-speed gear pump is established. The variation characteristics of internal and external sound fields of flow-induced noise of gear pumps at different speeds are studied. The noise test system of the high-speed gear pump is built to verify the accuracy of numerical simulation. Finally, the flow-induced noise suppression effect of the high-speed gear pump is achieved by adding an oil replenishment groove.

2. Mechanism analysis of flow-induced noise in high-speed gear pump

Flow-induced noise and mechanical noise are important sources of gear pump noise. With the development of a high-speed gear pump and the improvement of machining accuracy, flow-induced noise will be much greater than mechanical noise. The specific causes are summarized as follows.

2.1. Cavitation noise

For the operation of rotating machinery with fluid as the working medium, the occurrence of cavitation is often accompanied. When the fluid flows through the variable cross-section channel, the flow rate of the fluid will increase and the pressure will decrease. When the local fluid pressure is lower than the saturated vapor pressure of the oil, the gas incorporated in the oil will be vaporized and precipitated. The precipitated gas is brought to the high-pressure area by the oil. When the bubble is shattered while it is being subjected to pressure, it will generate a strong pressure in the core of the bubble, which will lead to a huge impact and a lot of noise, specifically cavitation noise (Liu *et al.*, 2015; 2016).

2.2. Oil trapped noise

In order to ensure the normal operation of involute gears, the contact degree of gear meshing is required to be greater than 1, so there is an inevitable trapped oil volume. When the oil in the trapped oil volume is squeezed by the rotation of the gear, the compressibility of the oil is very small, and the oil in the trapped oil area will produce strong hydraulic impact and extrusion leakage, which will cause vibration and noise of the gear and the pump. At the same time, when the volume of the trapped oil area changes from small to large, it will induce cavitation and aggravate noise pollution (Guo, Guan, 2021).

2.3. Flow pulsation noise

As a kind of volumetric pump, the working principle of the gear pump inevitably produces periodic flow changes due to the structural characteristics, which makes the discharged oil to exhibit periodic flow pulsation. The flow pulsation at the outlet will cause pressure pulsation under the coupling with the load, which can cause vibration and noise of gear pumps, pipelines, and other components of the system. If the frequency of the flow pulsation coincides with a certain frequency of the system, it will also cause resonance and produce stronger noise (Marinaro *et al.*, 2021).

2.4. Turbulent noise

When the gear pump works, the internal fluid flow law is complex, and the fluid flow is more disordered under high-speed working conditions, thus inducing strong noise, that is, turbulent noise, which is usually composed of a turbulent boundary layer and its wake, vortex shedding from the solid surface, turbulence impacting the solid surface and other factors, and the spectrum presents a broadband characteristic.

The flow-induced noise of a high-speed gear pump is primarily caused by the unsteady flow of its internal fluid. From the perspective of the generation mechanism of flow-induced noise, the types of fluid sound sources can generally be divided into monopole sources, dipole sources, and quadrupole sources. Figure 1 shows the sound source models and characteristics of three kinds of fluid sound sources (Wang *et al.*, 2016b). Among them, Ma represents the Mach number, Wa refers to sound radiation velocity, and U_0 represents sound radiation velocity without disturbance.

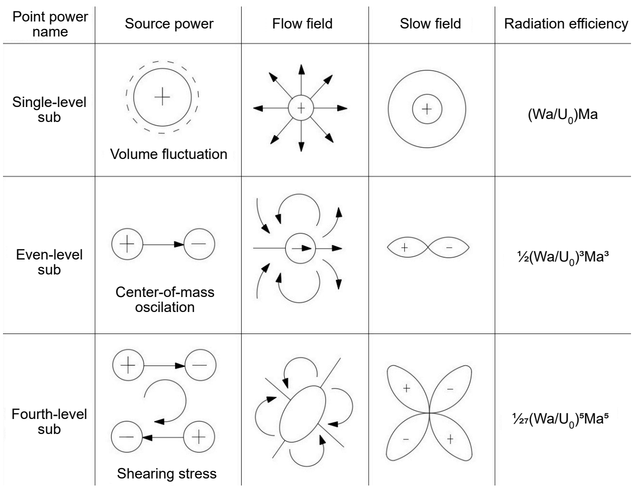

Fig. 1. Flow-induced noise source model and its characteristics.

The monopole source can be considered a point source of pulsating mass flow, which is mainly caused by the uneven mass or heat inflow in the medium.

The intensity of the sound source is proportional to the square of the Mach number. The dipole source can be regarded as the composition of two anti-phase monopole sources. The main reason is the interaction between the fluid and the solid contact with it. The intensity of the sound source is proportional to the third power of the Mach number, and the directivity of the sound radiation presents the "8" shape. The spectrum presents discrete characteristics. The common dipole source noise includes the boundary layer noise and the propeller rotation noise. Quadrupole sources can be considered to be composed of two anti-phase dipole sources, which are mainly produced in the stress change of turbulent fluid. The intensity of the sound source is proportional to the fifth power of the Mach number, and the directivity of acoustic radiation is a "four lobes" shape. The spectrum shows broadband characteristics. The common four-level sub-source noise includes jet noise and turbulent noise. From the perspective of noise radiation efficiency, the highest to lowest radiation intensity are: the monopole source, the dipole source, and the quadrupole source.

The flow-induced noise of the high-speed gear pump is composed of the above three fluid sound sources, in which the cavitation noise belongs to the monopole source noise. Since the high-speed rotation of the gear intensifies the occurrence of cavitation, and the radiation efficiency of the monopole source noise is higher than that of the other two sound sources, the monopole source noise is often a key factor in the noise reduction of the high-speed gear pump. The vibration radiation noise of the pump body contacted by the pressure fluctuation of the fluid belongs to the dipole source noise. Although the noise radiation efficiency is not as good as that of the monopole source noise, it often forms a strong noise level. Turbulent noise belongs to the fourth-order sub-source noise. Because its noise radiation efficiency is the weakest and is often considered at a high Mach number, the fluid flow in a high-speed gear pump belongs to a low Mach number, so it is ignored.

3. Numerical simulation

3.1. Calculation model

Since the involute gear is easy to manufacture and has a strong bearing capacity, the object of this study is designed as the involute external gear pump. The theoretical displacement of the pump is 72 mL/r, the working pressure is 2 MPa, and the working speed is 3000–5000 r/min. The simplified geometric modeling of the gear pump is carried out in Creo software, which is mainly composed of an inlet, an upper pump cover, a lower pump cover, a gear, and a floating side plate, as shown in Fig. 2. The main geometric parameters of the gear part of the external gear pump calculated according to the design parameters are shown in Table 1. The diameter of the oil suction port of the gear pump is 58 mm, and the diameter of the oil discharge port is 30 mm.

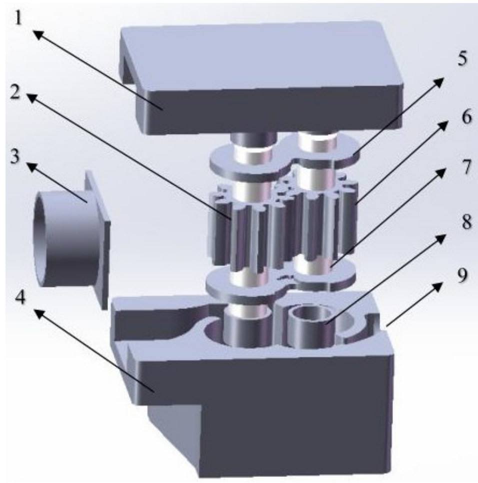

Fig. 2. Structure diagram of high-speed gear pump: 1) upper pump cover; 2) driving gear; 3) inlet; 4) lower pump cover; 5) floating side plate; 6) driven gear; 7) pump shaft; 8) needle roller bearing; 9) outlet.

Table 1. Gear design parameters.

Parameter	Parameter value
Module of gear	5
Tooth number	10
Gear indexing circle diameter [mm]	50
Addendum [mm]	4
Tooth dedendum [mm]	7.25
Root diameter [mm]	35.5
Addendum circle diameter [mm]	58
Pressure angle [°]	20
Displacement factor	–0.2
Operating center distance [mm]	47.5
Tooth width [mm]	45

3.2. Flow field calculation

Based on the Pumplinx flow field analysis software, the internal flow field of the gear pump was simulated and analyzed. The SIMPLEC algorithm was used, and the RNG k-ε turbulence model was used for the turbulence model. This model was an important correction based on the standard k-ε turbulence model. By reflecting the influence of small-scale vortex in the corrected viscosity term and large-scale vortex motion, this model can calculate the low Reynolds number turbulence and take into account the vortex effect, which improves the calculation accuracy of strong vortex flow (Chen *et al.*, 2003). The grid division of the computational model adopts the special grid generator template provided by software. The final number of grid cells is 462761, and the number of grid nodes is 1486231. The overall grid division is shown in Fig. 3.

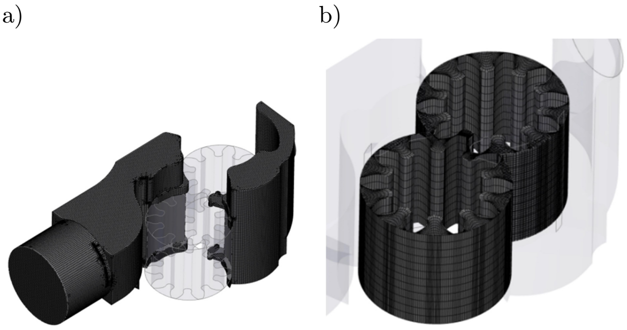

Fig. 3. Grid model of flow field: a) import and export areas; b) gear region.

In this paper, the residual convergence accuracy of the simulated physical quantity is set to 0.1, the time step of each tooth is set to 40 steps, the number of teeth is 10, and the gear is set to rotate for 10 rotations. Then the total time step is 4000. At the same time, to ensure the accuracy of the subsequent acoustic simulation, the flow field data showing periodic changes in the next five rotations are exported as the sound source term of the acoustic calculation with the EnSight format file. Table 2 shows the specific settings of simulation boundary conditions.

Table 2. Boundary conditions and parameter settings.

Dynamic viscosity [Pa·s]	0.2532516
Fluid density [kg/m^3]	860
Elasticity of modulus of liquid bulk [Pa]	2.15e + 09
Saturated vapor pressure [Pa]	37100
Initial gas content	0.00009
Temperature [°]	20
Inlet pressure [MPa]	0.1, 0.15, 0.2
Outlet pressure [MPa]	2
Revolution speed [r/min]	3000, 4000, 5000

3.3. Sound field calculation

In this paper, the acoustic simulation of the flow-induced noise of the external gear pump is carried out based on the acoustic finite element method and the infinite element method. The body sound source and surface sound source of the fluid are extracted by the CAA method and interpolated into the corresponding acoustic body grid and the surface grid. Then, the noises generated by the flow field are simulated and their sound pressure spectra were obtained (Carletti *et al.*, 2016). To improve the accuracy of acoustic calculation, acoustic modeling only retains the important structural characteristics of the high-speed gear pump, the structural mesh of the pump body only retains the upper and lower pump covers, does not include gear and bearing and other parts, and the bearing holes are closed. The acoustic calculation model is mainly composed of body sound source, surface sound source, sound propagation area, and acoustic infinite element (Paszkowski, 2020). The grid model is established in HyperMesh software and the material properties of different components are defined, as shown in Fig. 4.

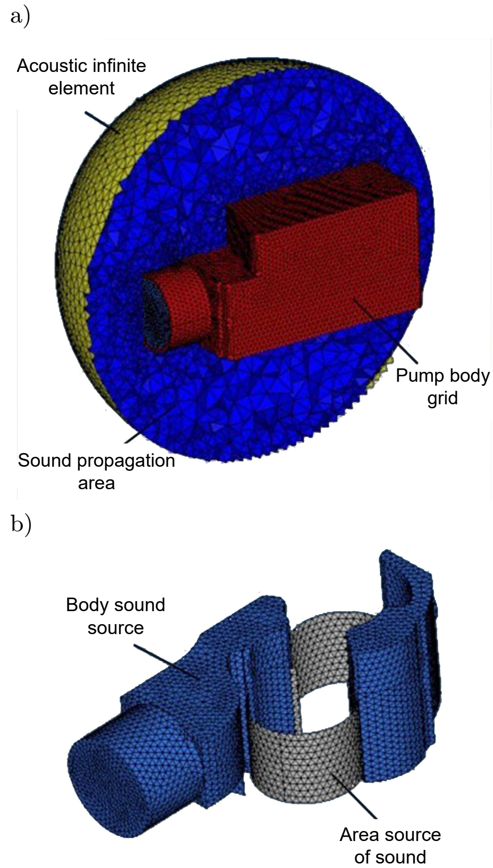

Fig. 4. Acoustic grid model: a) pump body and acoustic finite element infinite element mesh; b) acoustic grid in fluid domain.

Due to the consideration of the external radiation of the fluid sound source, the acoustic grid must include the grid of the external air domain, and the interface of each part of the grid is set as a common node, so that the transmission of acoustic information between various components can be realized. The size of the acoustic mesh is determined by the analysis frequency of the fluid. The maximum size $L_{\max}$ of the acoustic mesh should be calculated:

$$L_{\max} < c/6f_{\max}. \tag{1}$$

In the aforementioned equation, the sound speed c is about 1300 m/s, and the time step Δt of flow field calculation under each speed condition is 5×10^{-5} s, 3.75×10^{-5} s, and 3×10^{-5} s, respectively. Then the maximum frequency $f_{\max}$ at three speeds can be calculated by the sampling law, and the maximum frequency $f_{\max}$ is 16 666 Hz. Therefore, the maximum size $L_{\max}$ of the acoustic grid should be less than 13 mm, and the maximum size of the acoustic grid of each part of the high-speed gear pump should not exceed 0.2 mm, which can fully meet the requirements of the acoustic calculation. When converting time-domain acoustic information to frequency-domain acoustic information with a fast Fourier transform, the window function needs to be set as a Hanning window to reduce signal leakage. To explore the frequency response characteristics of the fluid sound source inside the high-speed gear pump, three monitoring points are set in the inlet and outlet areas of the pump, as shown in Fig. 5.

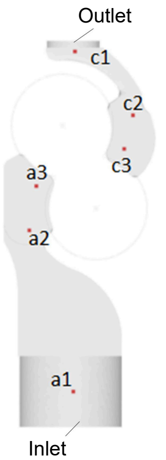

Fig. 5. Internal acoustic monitoring points.

4. Analysis of numerical simulation results

4.1. Analysis of flow field calculation results

Figure 6 depicts the flow pulsation curve of the high-speed gear pump outlet under each oil suction pressure after periodic stable fluctuation. From the information in the graph, it can be seen that the instantaneous flow curve of each graph shows periodic change, which conforms to the essence that the periodic change of the meshing point position leads to the periodic change of the output flow. With the increase of the rotational speed, the amplitude of outlet flow pulsation is stronger. When the oil suction pressure is 0.1 MPa, the fluctuation amplitude of the outlet flow pulsation of the high-speed gear pump is large under various rotational speeds. With the increase of the oil suction pressure, the amplitude of the outlet flow pulsation decreases gradually. The main reason is that the increase of the oil suction pressure inhibits the cavitation level in the gear rotor area, resulting in the decrease of the flow pulsation at the outlet. Therefore, the way of pressurizing the oil tank pressure can be used to reduce the occurrence of internal cavitation in high-speed gear pumps and to reduce the cavitation noise level.

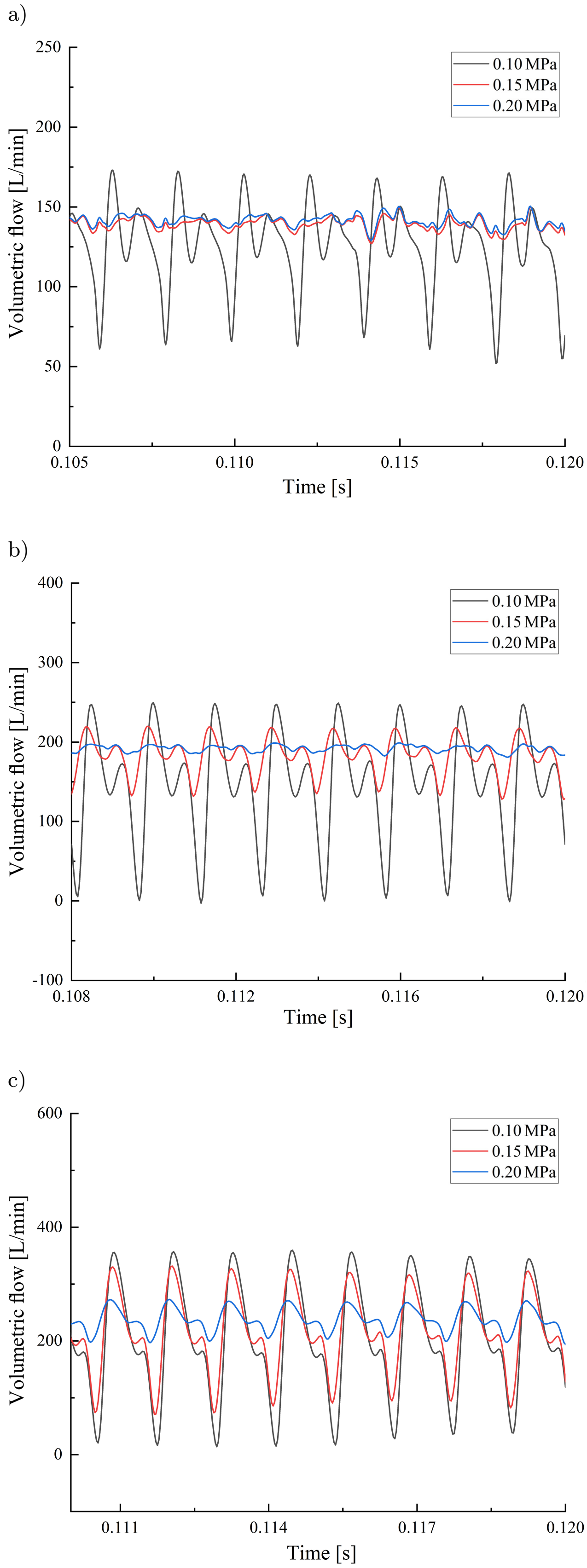

Fig. 6. Fluctuation curve of outlet flow under each suction pressure: a) 3000 r/min; b) 4000 r/min; c) 5000 r/min.

4.2. Analysis of sound field calculation result

After the acoustic calculation is completed, the distribution cloud diagram of surface and body sound sources at the fundamental frequency and its harmonics are selected. From the formula $f = n \cdot Z/60$, it can be seen that when the rotational speed is 4000 r/min, the passing frequency (fundamental frequency) of the gear to the fluid is 667 Hz, and then 1333, 2002, and 2665 Hz are all harmonics of the fundamental frequency (Huang *et al.*, 2019). From Fig. 7 the distribution of surface sound sources and sound pressure levels at different locations can be seen. In the distribution of the surface sound source on the outer wall of the gear rotation area, the alternating occurrence of the maximum and minimum sound pressure levels is the discrete sound source when the high-speed gear pump works. Compared with other frequencies, the sound pressure level at the fundamental frequency is the highest, and with the increase of frequency, the sound pressure level shows a significant decreasing trend.

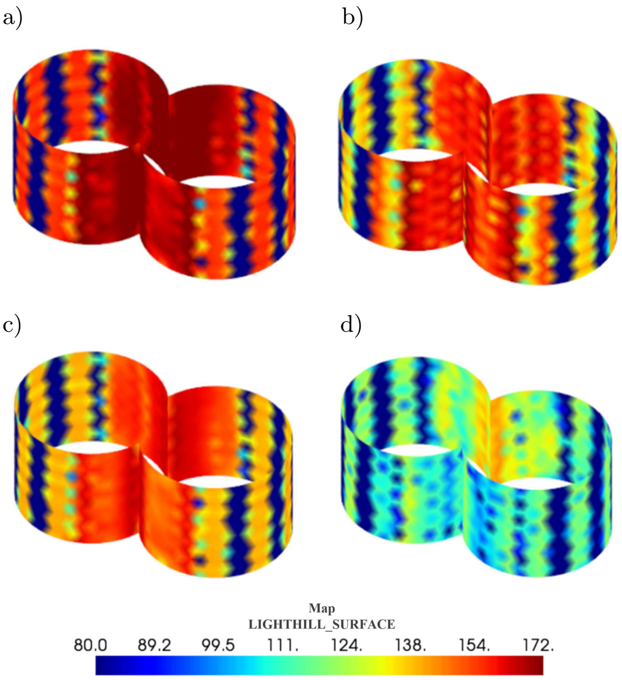

Fig. 7. Surface sound source images at fundamental frequency and harmonics frequency [dB]: a) 667 Hz; b) 1333 Hz; c) 2002 Hz; d) 2665 Hz.

It can be seen from the distribution cloud diagram of the body sound source in Fig. 8 that the sound pressure level of the body sound source is much higher than that of the surface sound source. The sound pressure level of the body sound source at the fundamental frequency is the highest, and with the increase of frequency, the sound pressure level also shows a significant decreasing trend. The sound pressure level in the outlet area is higher than that in the inlet area because the oil in the outlet area is subjected to a high-pressure load and has a strong impact on the pump body, which makes the sound pressure level relatively high. In addition, the sound pressure level at the junction of the gear and inlet and outlet fluid domain is also high. The unstable flow will produce obvious turbulence and vortex phenomena, and the fluctuation of oil pressure directly affects the level of the sound pressure.

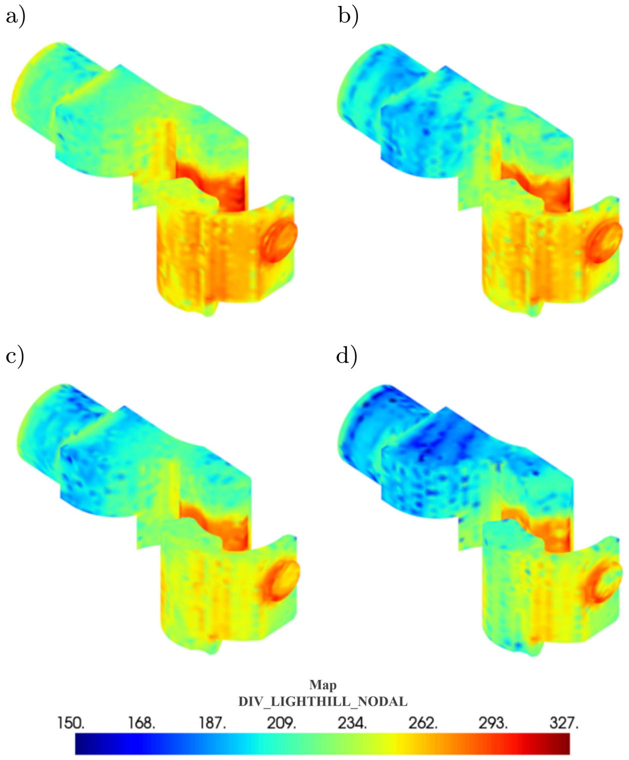

Fig. 8. Cloud images of body sound source at fundamental frequency and its harmonics frequency [dB]: a) 667 Hz; b) 1333 Hz; c) 2002 Hz; d) 2665 Hz.

Figure 9 shows the flow-induced noise spectrum curves of each acoustic monitoring point at the inlet and outlet under the working condition of 4000 r/min. It can be found that the sound pressure level spectrum is mainly composed of wideband noise and discrete noise. The peak value of the sound pressure level appears at the fundamental frequency (667 Hz) and harmonics frequency (1333, 2002, 2665 Hz). The sound pressure levels at the fundamental frequency and harmonics frequency are about 10–20 dB higher than those at other frequency bands, indicating that the flow pulsation formed by the periodic mutual force between the gear and the fluid is the key factor to induce the flow-induced noise.

The total sound pressure level of the inlet region is 182–199 dB, while the total sound pressure level of the outlet region is 195–212 dB. The total sound pressure level of the outlet region is higher than that of the inlet region, and the total sound pressure level of the import and export monitoring points decreases with the gear rotor area to the import and export sides.

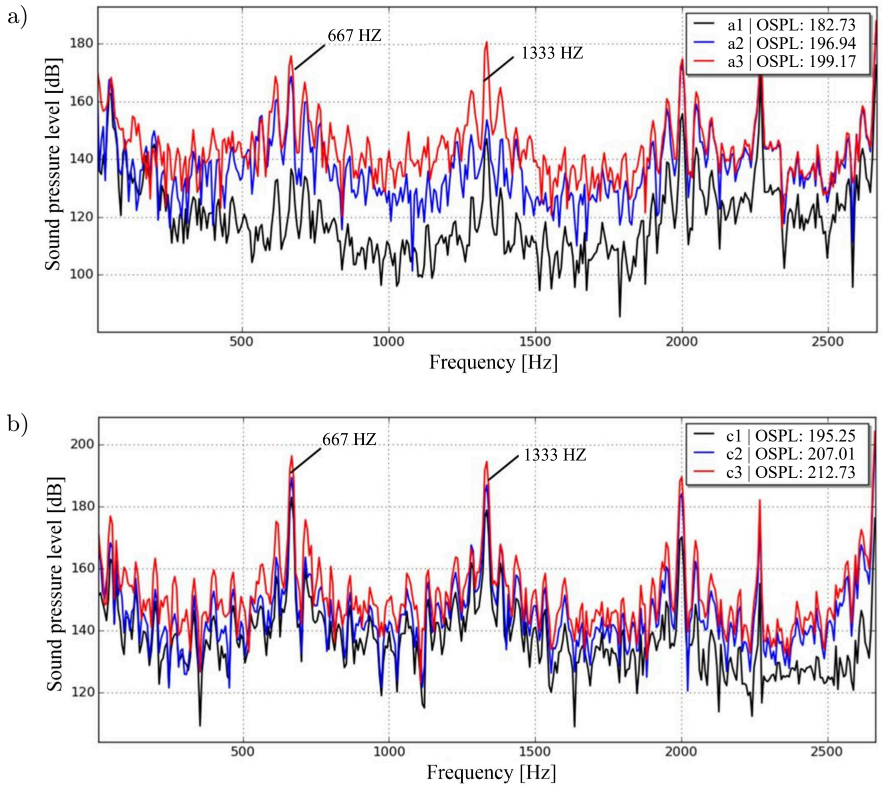

Fig. 9. Spectrum of sound pressure level in pump: a) imported spectrum curve; b) export spectrum curve.

Table 3 shows the summary of the total sound pressure levels at the inlet and outlet of the high-speed gear pump under different oil suction pressures. It can be found that the total sound pressure level in the internal fluid domain increases with the increase of the rotational speed. The total sound pressure level in the outlet area is much higher than that in the inlet area, and the total sound pressure level in the inlet and outlet area decreases with the increase of the oil suction pressure. The main reason is that the increase of the oil suction pressure can reduce the gas content of the oil, improve the effective volume elastic modulus of the oil, and reduce the compressibility of the oil, thus greatly inhibiting the occurrence of cavitation under high-speed conditions. Comparing the change of the total sound pressure level at the monitoring point a2 in the import area with that at the point c2 in the export area, it can be found that for every 0.05 MPa increase in the oil absorption pressure, the total sound pressure level at the monitoring point a2 in the import area decreases by 3–8 dB, while the total sound pressure level at the monitoring point c2 in the export area decreases by 10–14 dB, indicating that the noise reduction effect of increasing the oil absorption pressure in the export area is better than that in the import area.

Table 3. Total sound pressure level table of import and export monitoring points under oil suction pressure.

Revolution speed [r/min]	Suction pressure [MPa]	a2 total sound pressure level [dB]	c2 total sound pressure level [dB]
3000	0.1	191	200
	0.15	182	189
	0.2	179	185
4000	0.1	197	207
	0.15	189	195
	0.2	187	187
5000	0.1	205	217
	0.15	196	203
	0.2	193	192

5. Experimental verification

5.1. Experimental system

Due to the complex generation mechanism of flow-induced noise of a high-speed gear pump, it is often difficult to convincingly study the results only by numerical simulation. Therefore, to verify the accuracy of the above acoustic simulation, this section carried out the noise test experiment of the high-speed gear pump. The noise test experiment of the high-speed gear pump is carried out on the hydraulic system test bench of the oil pump production workshop in Dongguan. The experimental system mainly includes a gear pump operation system and a data acquisition system. The gear pump operation system includes a gear pump, drive motor, pressure regulating valve, pipeline, sound insulation cover, and oil tank. The data acquisition system collects various physical quantities in the operation of a high-speed gear pump with the help of an acquisition instrument and various sensors, including a sound pressure sensor, turbine flowmeter, photoelectric speed meter, pressure gauge, and thermometer. The experimental system is shown in Fig. 10.

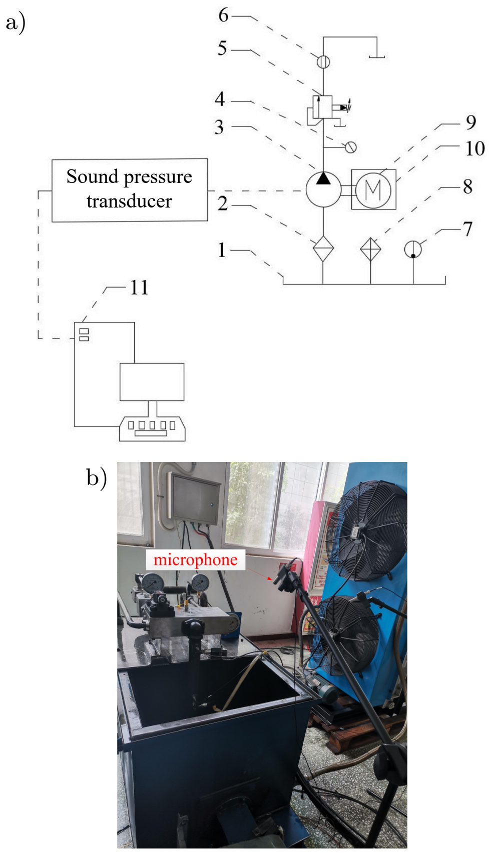

Fig. 10. Experimental system diagram: a) system schematic diagram (1 – tank; 2 – filter; 3 – gear pump; 4 – pressure gauge; 5 – pressure regulating valve; 6 – flowmeter; 7 – thermometer; 8 – heater; 9 – sound insulation cover; 10 – motor; 11 – data acquisition system); b) physical figure.

The acquisition instrument used in this experiment is the INV3062SC series 24-bit network distributed acquisition instrument of the Beijing Oriental Institute of Vibration and Noise. The sound pressure sensor adopts the IEPE microphone preamplifier of the Beijing Oriental Institute, which belongs to capacitive testing sensors. The frequency response range is 16 Hz–100 kHz, the measurement accuracy is ±0.5 dB, the maximum output voltage can reach 5.0 V_{rms}, and the working condition is −40 to 85°C. The dynamic measurement of the sound pressure sensor can ensure that the total distortion of the sound pressure level below 146 dB is not more than 3%. In order to record the noise generated by the high-speed gear pump, the sampling frequency of the experimental test is 8 kHz, and the sampling time is 90 s. After the sampling is completed, the time-frequency conversion window function of the acoustic signal is set as the Hanning window.

5.2. Comparison between experimental results and simulation

In this noise sampling, to prevent the noise signal generated during the operation of the driving motor from affecting the experimental results, the motor is wrapped by sound absorption materials. The arrangement of acoustic pressure sensors refers to the relevant standards (Tang *et al.*, 2014). The acoustic pressure sensors are arranged according to the hemispherical method and make appropriate adjustments according to the actual situation of the site. Due to the small volume of the gear pump in this experiment, the length, width, and height of the pump body are all less than 0.5 m, and the radius of the hemisphere should not be less than twice the length of the experimental object and not less than 1 m. Therefore, the sound pressure sensor arranged at a radius of 1 m from the pump source. In this experiment, three acoustic measuring points were placed around the outlet of the radial section of the gear pump and measured synchronously. Each measuring point is 1 m away from the gear pump, and the three measuring points are sandwiched by 40°. The actual measurement point arrangement is shown in Fig. 11.

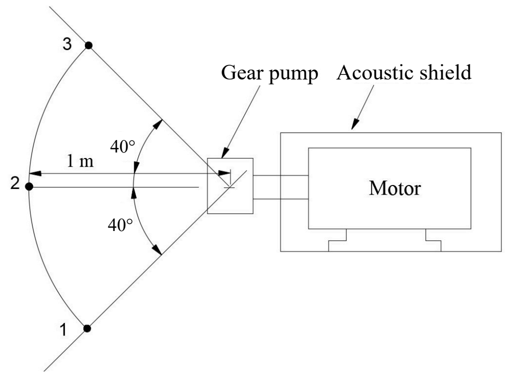

Fig. 11. Arrangement of sound pressure sensor measurement points.

The fast Fourier transform (FFT) is performed on the time domain information of the noise of the sampled high-speed gear pump to obtain the sound pressure level spectrum of the noise. Figure 12 is the comparison between the experimental values and the simulation values of three acoustic measuring points. Due to the error of the practical test and the numerical simulation itself, the experimental values are higher and lower than the simulation values in each frequency band, but the overall trend is the same. The discrete noise of the practical and simulation values is pronounced at the fundamental frequency and harmonics frequency. The experimental values of the three acoustic measuring points are slightly larger than the simula-

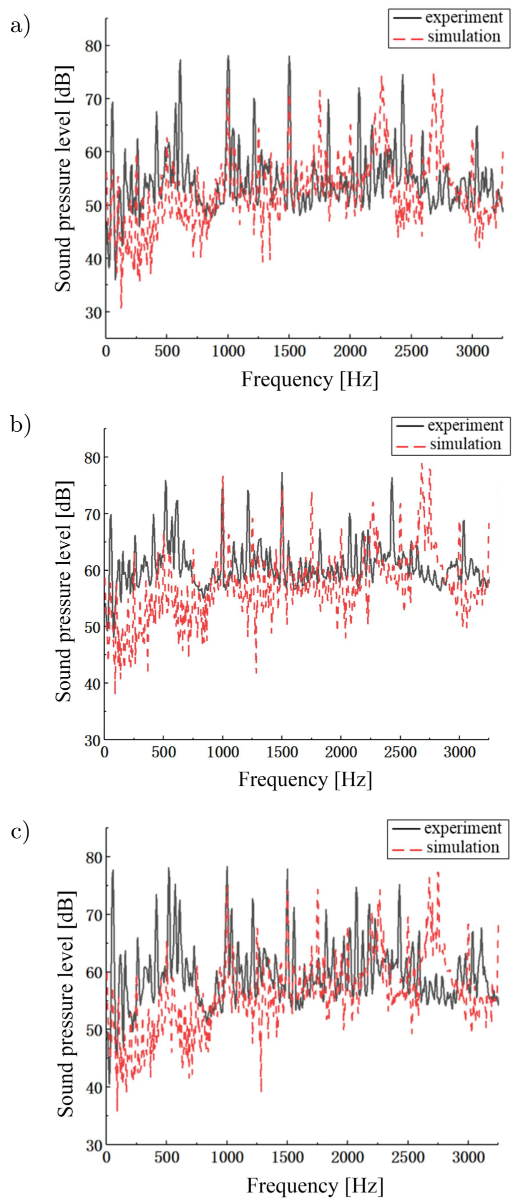

Fig. 12. Spectrum comparison of sound pressure level: a) measurement point 1; b) measurement point 2; c) measurement point 3.

tion values in the range of 0–1000 Hz. The experimental values are in good agreement with the simulation values at the second frequency (1000 Hz) and the third frequency (1500 Hz). Since the influence of the mechanical noise of the pump, actual measured noise is more than that of the simulated noise spectrum.

At 3000, 4000, and 5000 r/min measured experimentally, the average sound pressure levels considering A-weighting are 87, 90, and 94 dB, respectively. Table 4 compares the experimental and simulation noise values at each measuring point noise characteristic of 3000 r/min. According to the table, the noise value measured by the experiment is slightly higher than the simulation value because the noise measurement of the high-speed gear pump cannot exclude mechanical noise interference. However, the maximum error of the noise value obtained by experiment and simulation is less than 7 dB at each frequency. The feasibility of the algorithm used in this paper to simulate the flow-induced noise of a high-speed gear pump is demonstrated within an acceptable range.

Table 4. Experimental simulation comparison of noise values at characteristic frequencies [dB].

Frequency [Hz]		500	1000	1500
Measuring point 1	Experiment	76	75	78
	Simulation	69	76	75
	Error	7	1	3
Measuring point 2	Experiment	79	79	79
	Simulation	73	76	76
	Error	6	3	3
Measuring point 3	Experiment	78	79	78
	Simulation	72	75	74
	Error	6	4	4

6. Simulation analysis of noise reduction optimization

6.1. Structure design of oil replenishment tank

Under actual operating conditions, the high-speed gear pump in this paper has a low oil suction pressure. On the one hand, due to the centrifugal force, it is difficult for the oil to fill the grooves in time, resulting in the accumulation of gas in the oil at the root of the tooth, which severely reduces the volumetric efficiency of the pump. On the other hand, it will cause cavitation to become more intense. The precipitated gas flows to the high-pressure area of the outlet and collapses with the rotation of the gear, producing a loud cavitation noise. As a result, Gianluca Marinaro proposed a side plate structure that controls the reverse flow to achieve a consistent reduction in the amplitude of flow unevenness (Zhou *et al.*, 2018).

A scheme for opening an oil replenishment groove near the inlet side of the floating side plate is proposed in this paper. In time, the oil can be added to the tooth groove. The inner diameter (d = 35.5 mm) and outer diameter (D = 46.75 mm) of the oil replenishment groove are designed as the diameter of the tooth root circle and the diameter of the dividing circle, respectively, to improve the filling effect of the oil. Figure 13 depicts the structure of the floating side plate.

To find the best angle and depth of the oil replenishment tank, the angle of the oil replenishment tank in this simulation is interpolated once every 5° from 120° to 250°, and the angle of the two teeth in the high-pressure area is retained for oil sealing. The depth of the oil replenishment tank is set as a group from 0.4 to 4 mm every 0.4 mm. To analyze the influence of oil re-

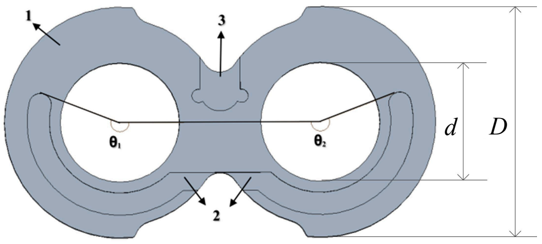

Fig. 13. Structure of oil tank: 1) side plate; 2) oil filling groove; 3) unloading groove.

plenishment tank structure on the flow-induced noise of the high-speed gear pump, the outlet flow pulsation rate under each oil suction pressure is calculated according:

$$\delta = \frac{Q_{\max} - Q_{\min}}{Q_{\text{avg}}}, \tag{2}$$

where δ is the flow pulsation rate; $Q_{\max}$ is the maximum value of instantaneous flow; $Q_{\min}$ is the minimum value of instantaneous flow; Q_{avg} is the average flow value, L/min. Figure 14 depicts the variation curve of

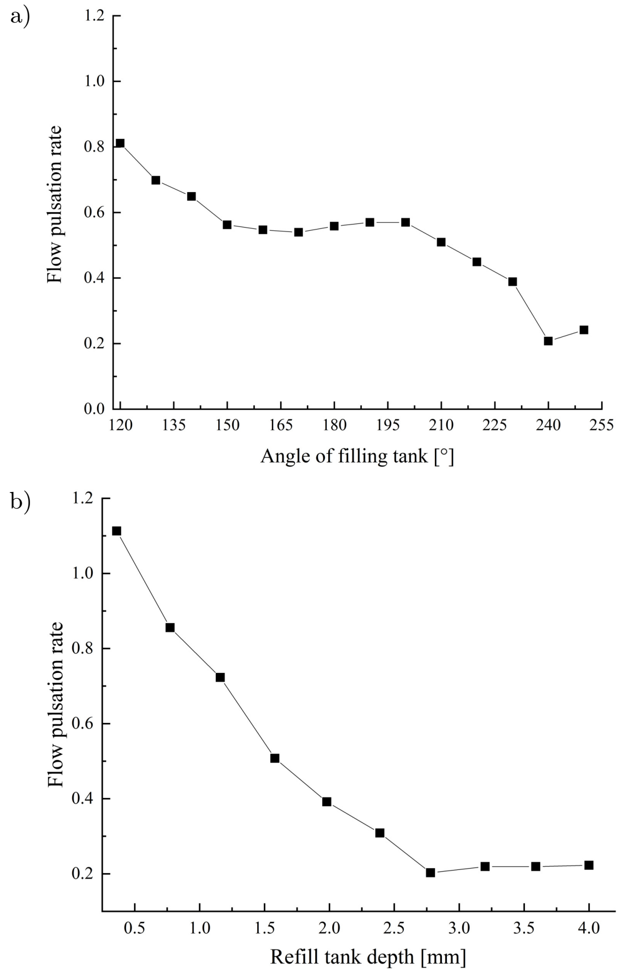

Fig. 14. Variation curve of flow pulsation rate – the relationship between flow pulsation rate and: a) the angle of oil feeding tank; b) tank depth.

the gear pump's outlet flow pulsation rate with the angle and depth of the oil replenishment tank.

Through the analysis of Fig. 14a, it can be seen that the fluctuation rate of the outlet flow of the high-speed gear pump changes gently before 240°. At 240°, the fluctuation rate of the outlet flow decreases sharply. Therefore, in order to avoid the decrease of the volumetric efficiency caused by too large an angle in the design of the oil replenishment tank, 240° is the best choice. Figure 14b shows that the flow pulsation rate at the outlet shows a decreasing trend before the oil replenishment tank depth is 2.8 mm. When the tank depth is greater than 2.8 mm, the flow pulsation rate tends to be gentle. Considering the structural strength problem caused by the deep oil tank, the selection of the oil replenishment tank with 2.8 mm depth in the design has the best effect on reducing the noise caused by the flow pulsation.

6.2. Simulation comparison after optimization

The flow field calculation results with the oil replenishment tank structure are loaded into acoustic software to calculate flow-induced noise. After the calculation, a section is set in the axial and radial directions of the original pump and the improved pump, respectively, to show the change in the sound pressure level of the external field before and after the improvement. The sound radiation slice nephograms before and after the improvement at the fundamental frequency are selected, as shown in Fig. 15. Table 5 summarizes the total sound pressure level summary of the external acoustic monitoring points after the modification.

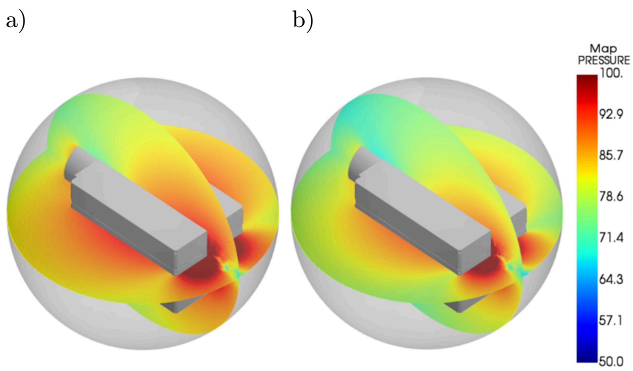

Fig. 15. Acoustic radiation slice images at fundamental frequency: a) original pump; b) improvement.

As demonstrated in Fig. 15, the sound pressure level in the external field of the modified high-speed gear pump is significantly lower than that of the original pump. The statistics in Table 5 show that the high-speed gear pump with the oil replenishment groove construction has a lower external sound pressure level than the original pump, and the total sound pressure level is reduced by 4–5 dB. As a result, the cavitation

Table 5. Comparison of total sound pressure level before and after improvement of external measuring points [dB].

Revolution speed [r/min]		3000	4000	5000
Measuring point 1	Original	88	91	94
	Improvement	83	86	90
Measuring point 2	Original	87	90	94
	Improvement	82	87	90
Measuring point 3	Original	89	91	95
	Improvement	84	86	91

and flow pulsation levels in the pump can be lowered by opening the oil replenishment groove on the floating side plate near the gear side. As a result, the level of flow-induced noise caused by oil flow is reduced.

7. Conclusions

In this research, the CFD method and the Lighthill sound analogies theory are utilized to develop a calculation model of flow-induced noise of an external gear pump that accurately reflects the internal flow noise characteristics of the external gear pump. The findings indicate:

1) The experimental and simulation results show that the spectrum curves of the external sound pressure level of the high-speed gear pump measured in the experiment agree well with the simulation, and the maximum error of the noise value at the characteristic frequency is less than 7 dB, confirming the accuracy of the numerical simulation.
2) The flow-induced noise of the high-speed gear pump is dominated by discrete noise at the fundamental frequency and its harmonics frequency. The intensity of cavitation and flow-induced noise in the internal fluid region decreases as oil suction pressure increases, and the noise reduction effect in the outlet zone is greater than that in the inlet region. The main source of vibration is the oil trapped zone. The overall sound pressure level of flow-induced noise decreases with distance from the oil trapped zone in the inlet and outflow regions, and the sound pressure level in the outlet region is higher than that in the inlet region.
3) The effect of minimizing the flow pulsation rate at the output of the high-speed gear pump is optimal when the planned oil replenishment tank angle is approximately 240° and the depth is about 2.8 mm. The simulation shows that increasing the size of the oil replenishment tank can significantly reduce the intensity and range of cavitation in the pump, improve oil absorption ability, and reduce total sound pressure level in the external field by about 4–5 dB, achieving the flow-induced noise suppression effect.

Acknowledgments

This work was supported by the National Key R&D Program of China (grant no. 2022YFB3403002).

References

1. Carletti E., Miccoli G., Pedrielli F. (2016), Vibroacoustic measurements and simulations applied to external gear pumps. An integrated simplified approach, *Archives of Acoustics*, **41**(2): 285–296, doi: 10.1515/aoa-2016-0028.
2. Chen Q.G., Xu Z., Zhang Y.J. (2003), Application of RNG K-ε model in numerical calculation of engineering turbulence [in Chinese], *Chinese Quarterly Journal of Mechanics*, **24**(1): 88–95.
3. Fiebig W., Wróbel J. (2022), Experimental and numerical investigation on the noise development in fluid power units: An overview, *Archives of Civil and Mechanical Engineering*, **22**(4): 1–18, doi: 10.1007/S43452-022-00481-X.
4. Guo S.X., Guan X.F. (2021), Simulation research on trapped oil pressure of involute internal gear pump, *Mathematical Problems in Engineering*, **2021**: 8834547, doi: 10.1155/2021/8834547.
5. Huang P.L. *et al.* (2019), A study on noise reduction of gear pumps of wheel loaders based on the ICA model, *International Journal of Environmental Research and Public Health*, **16**(6): 999, doi: 10.3390/ijerph16060999.
6. Liu Y.Y., Wang L.Q., Zhu Z.C. (2015), Numerical study on flow characteristics of rotor pumps including cavitation, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, **229**(14): 2626–2638, doi: 10.1177/0954406214562634.
7. Liu Y.Y., Wang L.Q., Zhu Z.C. (2016), Experimental and numerical studies on the effect of inlet pressure on cavitating flows in rotor pumps, *Journal of Engineering Research*, **4**(2): 19, doi: 10.7603/s40632-016-0019-x.
8. Marinaro G., Frosina E., Senatore A. (2021), A numerical analysis of an innovative flow ripple reduction method for external gear pumps, *Energies*, **14**(2): 471, doi: 10.3390/en14020471.
9. Paszkowski W. (2020), Modeling of vibroacoustic phenomena using the method of parameterizing the audio signal, *Eksploatacja i Niezawodność – Maintenance and Reliability*, **22**(3): 501–507, doi: 10.17531/ein.2020.3.13.
10. Tang C., Wang Y.S., Gan J.H., Guo H. (2014), Fluid-sound coupling simulation and experimental validation for noise characteristics of a variable displacement external gear pump, *Noise Control Engineering Journal*, **62**(3): 123–131, doi: 10.3397/1/376212.
11. Wang H., Deng G., Li Q., Kang Q. (2016a), Research on bispectrum analysis of secondary feature for vehicle exterior noise based on nonnegative Tucker3 decomposition, *Eksploatacja i Niezawodność – Main-*

tenance and Reliability, **18**(2): 291–298, doi: 10.17531/ein.2016.2.18.

12. Wang C.X., Wu C.J., Chen L.J., Qiu C.L., Xiong J.S. (2016b), Review on mechanism and prediction of flow-induced noise [in Chinese], *Research on Chinese Ships*, **11**(1): 57–71.

13. Woo S., Vacca A. (2020), Experimental characterization and evaluation of the vibroacoustic field of hydraulic pumps: The case of an external gear pump, *Energies*, **13**(24): 6639, doi: 10.3390/EN13246639.

14. Woo S., Vacca A. (2022), An investigation of the vibration modes of an external gear pump through experiments and numerical modeling, *Energies*, **15**(3): 796, doi: 10.3390/EN15030796.

15. Zhang D.S., Zhang N.S., Xu B., Zhao R.J., Gao X.F., Li N. (2021), Numerical simulation of the flow-induced noise in a water-jet pump based on the Lighthill acoustic analogy theory [in Chinese], *Journal of Vibration and Shock*, **40**(10): 278–287.

16. Zhou Y.K., Li W.W., Yang D.P. (2018), Noise simulation analysis of a variable displacement external gear pump and unloading groove optimization, *Open Access Library Journal*, **5**(11): 1–9, doi: 10.4236/oalib.1105030.

Archives of Acoustics Vol. 49, No. 1, pp. 61–71 (2024), doi: 10.24425/aoa.2024.148770

Research Paper

Field Experiment as a Tool to Verify The Effectiveness of Prototype Track Structure Components Aimed at Reducing Railway Noise Nuisance

Cezary KRAŚKIEWICZ[(1)*], Grzegorz KLEKOT[(2)], Piotr KSIĄŻKA[(3)], Artur ZBICIAK[(1)], Przemysław MOSSAKOWSKI[(1)], Patrycja CHACIŃSKA[(3)], Anna AL SABOUNI-ZAWADZKA[(1)]

[(1)] *Faculty of Civil Engineering, Warsaw University of Technology*
[(2)] *Faculty of Automotive and Construction Machinery Engineering, Warsaw University of Technology*
[(3)] *National Research Institute, Department of Environmental Acoustics, Institute of Environmental Protection*
Warsaw, Poland

*Corresponding Author e-mail: cezary.kraskiewicz@pw.edu.pl

(*received April 19, 2023; accepted September 19, 2023; published online January 8, 2024*)

The almost unlimited possibilities of modern computational tools create the temptation to study phenomena related to the operation of engineering objects exclusively using complex numerical simulations. However, the fascination with multi-parametric complex computational models, whose solutions are obtained using iterative techniques, may result in qualitative discrepancies between reality and virtual simulations. The need to verify on real objects the conclusions obtained from numerical calculations is therefore indisputable. The enormous cost and uniqueness of large-scale test stands significantly limit the possibility of conducting tests under real conditions. The solution may be an experiment focused on testing features relevant to the given task, while minimising the dimensions of the objects under consideration. Such conditions led to the concept of conducting a series of field experiments to verify the effectiveness of prototype track components, which were developed using numerical simulations to reduce the noise caused by passing trains. The main aim of this study is to examine the acoustic efficiency of prototype porous concrete sound absorbing panels, in relation to the ballasted and ballastless track structures. Presented results of the proposed unconventional experiments carried out on an improvised test stand using the recorded acoustic signals confirm the effectiveness of the developed vibroacoustic isolators.

Keywords: vibroacoustic isolator; ballasted track structure; ballastless track structure; noise reduction; field test.

1. Introduction

Current regulations impose high requirements with regard to the protection of people and the environment against noise emissions (World Health Organization, 2018). At the same time, efforts are made to increase the capacity of railway lines, e.g., by increasing train speeds, which leads to higher levels of noise emitted to the environment. According to the report of the European Environment Agency (2020), railways are the second most dominant noise source in Europe. There are several solutions that can be applied to reduce such negative effects (Scossa-Romano, Oertli, 2012; de Vos, 2016; Thompson, 2008), for example: traditional acoustic screens, low-height noise barriers or various vibro-acoustic isolators integrated with the track.

Acoustic screens (Thompson, 2008) are placed along communication routes and their aim is to reduce the level of noise which is transmitted from the source of their emission (i.e., a railway route) to the surrounding environment. However, these traditional methods of protection against noise are not always possible to use for technical (location, necessary dimensions), economic or aesthetic reasons, or they do not give satisfactory results. In many cases, there is a need to reduce the noise level also at its source, i.e., in the emission zone. Considering that the modernised railway lines usually connect largest cities, and some of the railway line sections are located within the cities, and often in their very centres, it is necessary to use other, alternative solutions.

A good alternative are vibro-acoustic isolators, which are an integrated part of the track structure. Various solutions have been investigated and described in the literature. Glickman *et al.* (2011) conducted research on porous concrete sound absorptive panels used on the concrete slab trackbed to reduce the level of noise emitted to the surroundings. A similar solution was investigated by Zhao *et al.* (2014), where the authors examined the effect of porous sound-absorbing concrete slabs on the reduction of railway noise. They measured absorption coefficients of various materials in the laboratory and then, they tested selected slabs in a test section. They proved that porous sound-absorbing concrete slabs can significantly reduce railway noise at different train speeds.

Hong *et al.* (2005) studied the sound absorbing characteristics and performance of parallel perforated plate systems. They used an equivalent electroacoustic circuit approach, which was validated by comparing the calculated absorption characteristics with the ones measured by the two-microphone impedance tube method. Li and Guo (2017) proposed a numerical optimization method for acoustic performance of a micro-perforated plate aimed at the application in high-speed trains. Yori (2020) proposed a mathematical method for calculating the sound absorption coefficient of various sound absorbing materials depending on the incidence angle.

A group of scientists from the Seoul National University of Science and Technology conducted research on acoustic characteristics of the track structure in the urban train tunnel. They proposed an optimal mix design of a porous sound absorbing block applied on a concrete ballast (Lee *et al.*, 2016). They also developed a monitoring system measuring the noise reduction characteristics and structural behaviour of sound absorbing panels applied on the concrete trackbed (Oh *et al.*, 2017).

Shimokura and Soeta (2011) determined acoustic characteristics of the train noise for different types of railway stations: above-ground and underground, with side and island platforms. Matej and Orliński (2023) investigated possible ways of reducing wheel and rail wear in the operation of underground wagons on a curved track with small curve radii. Groll *et al.* (2023) studied transitional phenomena in railway systems with a focus on rail joints.

Vogiatzis and Vanhonacker (2015) investigated three different solutions for the reduction of railway rolling noise in light rail transit: sound absorbing precast elements, noise barriers, and rail dampers. They tested the proposed elements using a detailed rolling noise calculation procedure, and then, implemented selected solutions on site. Lázaro *et al.* (2022) studied the performance of low-height railway noise barriers with the addition of porous granular material on the inner face of the barrier.

According to the experience of foreign railway infrastructure managers (e.g., Germany and China), vibroacoustic isolators in the form of rail dampers or sound absorbing precast elements (usually made of porous concrete) are commonly used in the ballastless track structures and are able to effectively reduce the level of noise emitted to the environment. However, currently, Poland lacks a proper test site with a ballastless track system that would make it possible to conduct large-scale acoustic tests, as the ballastless track structures are used in the railways of PKP PLK S.A. marginally. The authors of this study have investigated ballastless track systems equipped with vibration isolators both experimentally and analytically (Zbiciak *et al.*, 2021), however, the laboratory research has focused on testing the particular vibration isolators, not the whole large-scale track section.

The research presented in this paper is a part of the BRIK InRaNoS project, co-financed by the European Union and PKP PLK S.A., which is aimed at developing an efficient vibroacoustic isolator to be applied in Polish railways for the reduction of railway noise emitted to the environment. Several works containing results of this research have been published so far. Kraśkiewicz *et al.* (2021a) proposed an experimental methodology for the identification of dynamic characteristics of a track structure, based on the determination of the track decay rate (TDR). They conducted field tests on the railway line section in Warsaw, where they measured TDR with the use of impulse tests. In another work, Kraśkiewicz *et al.* (2021b) investigated possible applications of rubber granulate SBR (styrene-butadiene rubber) produced from recycled waste tires as an elastic cover for prototype rail dampers. The authors performed laboratory tests on seven different SBR materials, with a focus on their operational durability.

The present paper aims at examining the acoustic efficiency of prototype sound absorbing panels based on porous concrete, in relation to the ballasted and ballastless track structures. So far, most of the studies on vibroacoustic isolators have concerned only the ballastless systems, assuming in advance that the ballast as a granular layer (crushed aggregate – usually crushed stone with the granulation of 31.5/50 or 31.5/63) provides a sufficient level of noise reduction, by absorbing and dissipating the acoustic wave. However, as stated by Zhang *et al.* (2019), the ballasted track does not always suppress noise better than the slab track. In their research, at lower frequencies (below 200 Hz) the noise level from the ballasted track was greater than that from the slab track, while at higher frequencies (250 to 1000 Hz) the slab track was noisier due to its lower track decay rates. What is important, the authors emphasised that although the results confirmed that the ballastless track is typically noisier than the one with ballast, the differences in the radiated noise depend on the physical properties of the compared tracks and should not be seen as universal.

Taking into account the results discussed above and the researchers' own experience, the authors of this paper have decided to investigate the noise reduction effectiveness of prototype porous concrete panels applied both in the ballasted and ballastless track structures. An innovative approach used in this study consists in implementing the same set of tests for the same prototype elements installed in two different types of track structures: ballasted and ballastless systems, in order to examine the acoustic efficiency of the developed sound absorbing panels.

2. Test methodology

The tests were aimed at determining the acoustic characteristics of prototype sound absorbing porous concrete panels installed on a full-scale test section of the track structure. The field experiment was an original idea of the authors, resulting from the lack of the real railway test section and a necessity of replacing it with a newly designed and constructed test stand, where sounds recorded during the passage of trains were used instead of actual excitations. A comparative approach was applied, which consisted in measuring the sound pressure levels in 1/3 octave bands in the reference and isolated systems:

– reference system – track structure without any vibroacoustic isolators;

– isolated system – track structure equipped with the tested sound absorbing panels.

In both systems, identical (as to the level and spectrum of the acoustic signal) excitations were emitted, that is: pink noise and real train passages, in the form of audio files. Measurement microphones were used to record the response of the tested systems at points located in their vicinity. Then, differences in the sound pressure levels in 1/3 octave bands were determined. It should be emphasised that the emitted sound levels were fully repeatable – the repeatability of the sound levels for various configurations of the track structure was ensured by using a microphone located in the immediate vicinity of the sound source (measurement point P0 – see Subsec. 3.2).

The effectiveness D_B of the solution (vibroacoustic isolator) is calculated as a difference of the sound pressure levels determined for a given observation point before and after the installation of the vibroacoustic isolator, provided that the noise source, terrain profiles, potential interference and reflective surfaces, as well as ground properties and meteorological conditions have not changed. It is a value expressed in decibels, which is determined for individual distances from the noise source using the formula:

$$D_B = L_{\mathrm{ref},d} - L_{\mathrm{iso},d} \quad [\mathrm{dB}], \tag{1}$$

where $L_{\mathrm{ref},d}$ is the sound pressure level in the reference system, measured at a distance d from the track axis, and $L_{\mathrm{iso},d}$ is the sound pressure level in the isolated system, measured at a distance d from the track axis.

In the conducted tests, a procedure of continuous recording of the acoustic signal was used, from which, at the analysis stage, acoustic events related to individual excitations were selected (emission of pink noise, passage of particular types of trains) and for those events, sound pressure levels in 1/3 octave bands and the values of the A-weighted equivalent sound level were determined.

The testing procedure was prepared by the authors based on two ISO standards: 3095 (2013) and 10847 (1997). However, the guidelines of ISO 3095 with regard to the location of measurement points could not be followed due to the technical limitations of the test stand. The measurement points were located closer to the railway noise source than specified in the standard, which resulted from the limited dimensions of the track structure sections prepared for experimental tests and the emitted levels of acoustic signals.

The applied procedure, however, is consistent with the main purpose of the research and does not affect the obtained measurement results. In the classic-speed railway lines, rolling noise is the dominant source of noise, and in the context of the realised research project, which focuses on the development and testing of vibroacoustic isolators, only this type of noise is the subject of further consideration. Such noise is generated due to the geometrical irregularities in the rolling surface of the wheel and the rail head, which generates dynamic forces acting on their contact surface. This, on the other hand, leads to relative vibrations of the wheel and the rail, with the vibration amplitude of each element depending on its dynamic properties. The resulting vibrations are the main source of noise.

During the measurements, the tested prototype sound absorbing panels had a total area of about 4.5 m^2, and the noise was not caused by the actual passing trains, but by emitted acoustic signals. Therefore, in order to observe and record the effect of absorption and dispersion of sound waves, smaller distances of measurement points from the track structure sections were applied.

3. Samples and test stand

3.1. Tested samples

The tested sound absorbing panels were made of porous concrete, whose recipe was marked with the symbol 220/10 – number 220 refers to the volume of cement grout (220 dm^3), number 10 indicates the percentage of sand in the crumb pile. The concrete recipe was elaborated within laboratory tests, and the surface grooving was designed using numerical simulations. All tested elements were produced in the laboratory of the Faculty of Civil Engineering at the Warsaw University of Technology.

Two types of panels were considered:

- panel 1 – porous concrete panels of 500 × 500 × 100 mm, with trapezoidal grooves;
- panel 2 – porous concrete panels of 500 × 500 × 100 mm, with half-round grooves.

In the conducted tests, nine panels of type 1 and nine panels of type 2 were applied together, with the aim of keeping the symmetry of the system. In this way, a hybrid system with 18 sound absorbing panels made of the same material, but with two different grooving patterns, was obtained.

The created vibroacoustic isolation system was tested on two different types of track structures: ballasted and ballastless. In the ballasted section, the panels were laid on the upper surface of the sleepers, between and outside the rails. The cross-section of the test stand is presented in Fig. 1, and the photographs of the tested section – in Fig. 2.

In the ballastless track system, the vibroacoustic isolators were installed on the slabs simulating the concrete track slabs, between and outside the rails. The cross-section of the test stand is presented in Fig. 3, and the photographs of the tested section – in Fig. 4.

The arrangement of the sound absorbing panels on the ballasted and ballastless track section was identical, as shown in Fig. 5.

3.2. Test stand

An original test stand was designed by the authors and constructed on the premises of the Warsaw University of Technology, in front of the building of the Faculty of Civil Engineering. The scheme of the test stand with marked locations of tested samples and measurement points is presented in Fig. 6.

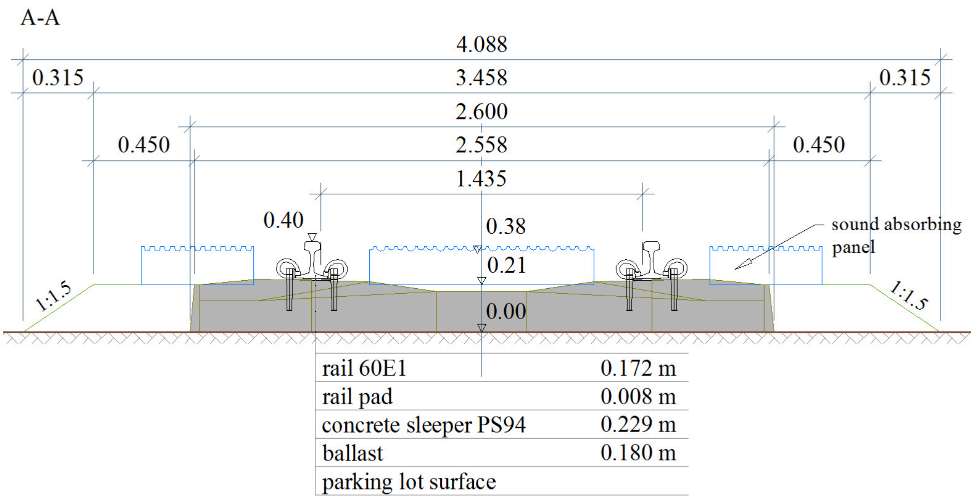

Fig. 1. Cross-section of the ballasted track structure with sound absorbing panels.

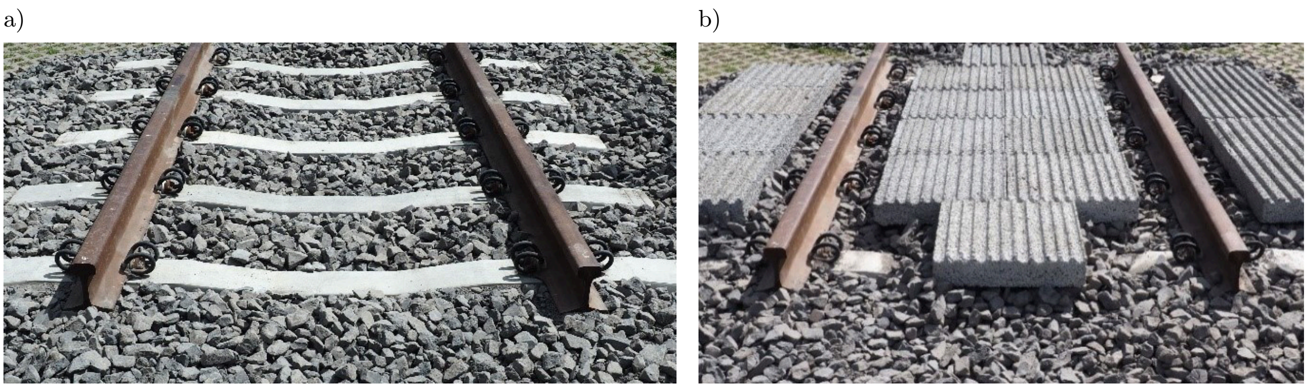

Fig. 2. View of the ballasted track structure: a) reference system; b) isolated system.

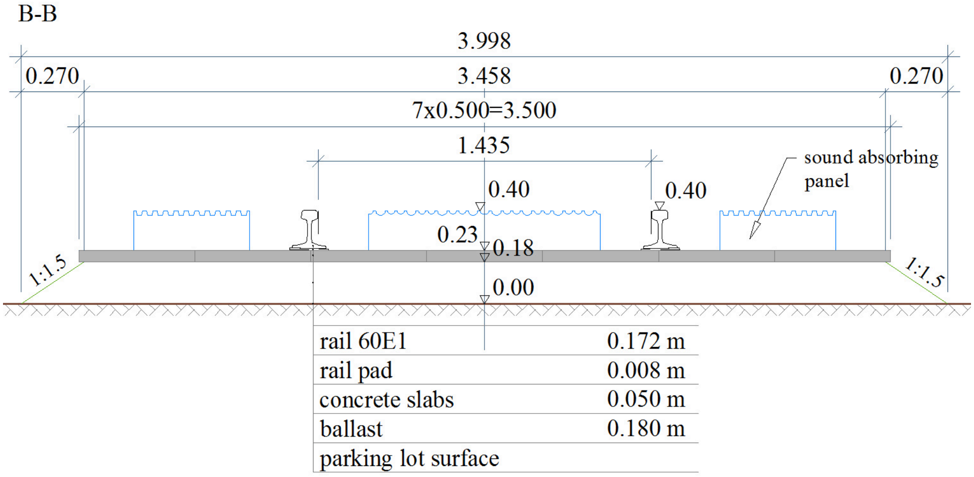

Fig. 3. Cross-section of the ballastless track structure with sound absorbing panels.

a) b)

Fig. 4. View of the ballastless track structure: a) reference system; b) isolated system.

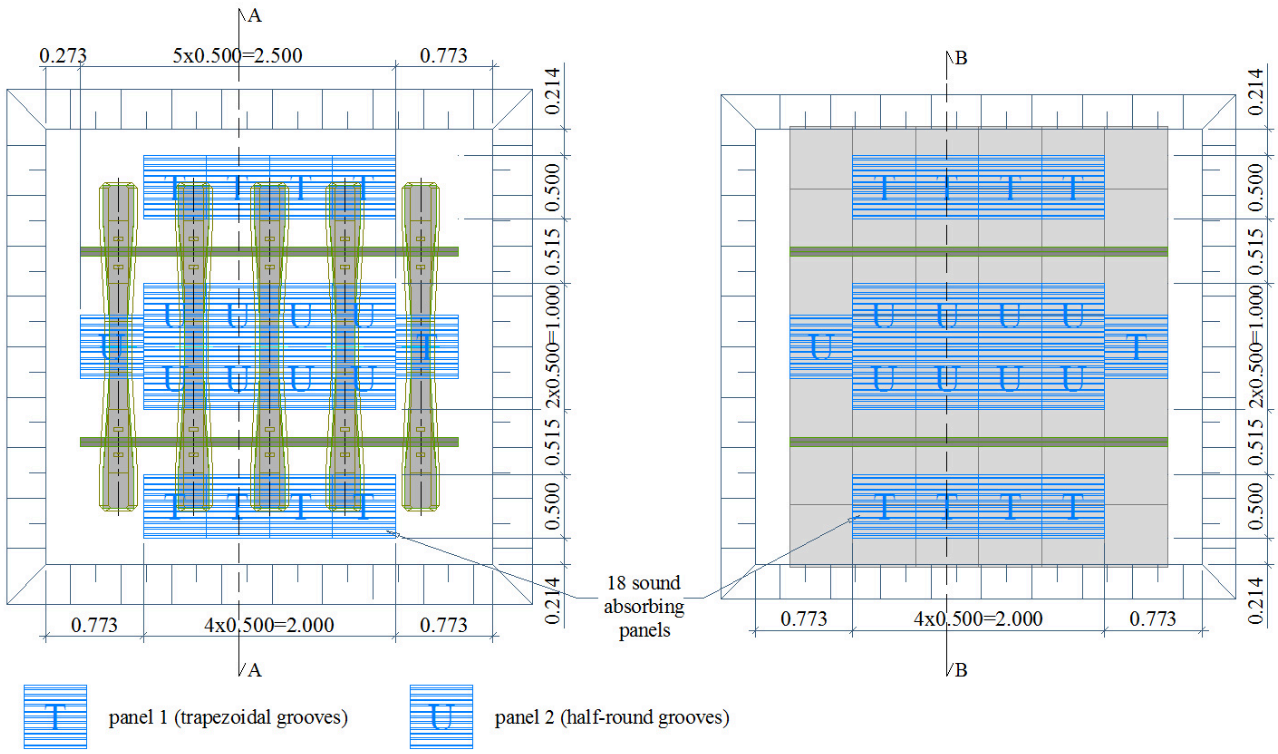

Fig. 5. Top view of the ballasted and ballasted track structure with the arrangement of sound absorbing panels.

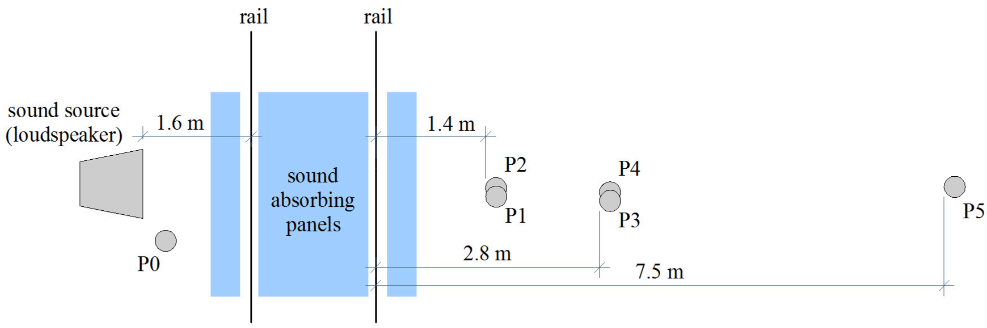

Fig. 6. Scheme of the test stand with marked location of measurement points.

Four configurations of the test stand were prepared:

- configuration I – ballasted track structure, isolated system;
- configuration II – ballasted track structure, reference system;
- configuration III – ballastless track structure, isolated system;
- configuration IV – ballastless track structure, reference system.

The ballasted track system consisted of: rail profiles 60E1, rail fastening system of SB type with the elastic rail pads PKV, and five sleepers PS-94 with 600 mm spacings. Such a track grid was placed directly on the parking lot slabs, and then covered with ballast to the level of the sleeper top in the zones between the sleepers and outside, at both sides of the track structure. Inclination of the ballast prism walls was 1:1.5. In this way, a total system width and length of around 4000 mm was obtained.

The ballastless track system was constructed on the newly laid paving slabs with the dimensions of 500×500 mm. The slabs were placed on the compacted ballast, in order to achieve a uniform level of the rails in both the ballasted and ballastless track system. PKV rail pads were laid on the slabs every 600 mm, and rail profiles 60E1 were placed on the pads with an axial spacing of 1500 mm (which corresponds to the track width of 1435 mm).

In each configuration, the measurements were made at four fixed measurement points (marked from P1 to P4 in Fig. 6) located at the constant distance from the tested samples and the constant height above the rail head/ground level, at the opposite side of the tested system in relation to the sound source (directional loudspeaker):

- P1 – 1.4 m from the closest rail head, at a height of 0.9 m;
- P2 – 1.4 m from the closest rail head, at a height of 1.8 m;
- P3 – 2.8 m from the closest rail head, at a height of 1.5 m;
- P4 – 2.8 m from the closest rail head, at a height of 3.5 m.

Moreover, in order to monitor the operation of the sound source and confirm the repeatability of the emitted acoustic signals, an additional point marked as P0 was used, located at a short distance of 0.4 m from the upper corner of the loudspeaker. Meteorological conditions during the measurements were monitored at the point marked as P5 – located 7.5 m from the rail head, at a height of 4.0 m

The signal source (loudspeaker) was located 1.8 m from the closest rail head, on a platform 0.85 m high. The loudspeaker diaphragm was directed at the tested sample of vibroacoustic isolators, at an angle of 25° to the ground plane. The chosen location of the sound source (loudspeaker) and microphones on the test stand resulted from the objective of the study, that is comparison of the acoustic signals reflected from the test samples. In the authors' opinion this location best reflected the adopted concept of outdoor testing, as the signals of passing trains used in the tests were recorded next to the track.

Fig. 7. View of the test stand with visible loudspeaker and microphones.

3.3. Measuring equipment

The measuring instruments were used in the conducted tests:

- SV 279 PRO Noise Monitoring Station (in points P1 to P5);
- SV 36 Acoustic Calibrator;
- Vaisala Weather Transmitter WXT530.

For the generation and emission of acoustic signals, the following measuring equipment was used:

– AMG mini amplifier/pink noise generator;
– laptop as a sound player;
– omnidirectional loudspeaker.

Six different types of acoustic signals (with a known spectrum measured in P0) were generated:

– no. 1 – pink noise;
– no. 2 – passage of the passenger train Pendolino ED250 (speed 152 km/h);
– no. 3 – passage of the freight train ET22 (speed 74 km/h);
– no. 4 – passage of the old-type (locomotive and carriages) passenger train EP09 – composition 1 (speed 106 km/h);
– no. 5 – passage of the old-type (locomotive and carriages) passenger train EP09 – composition 2 (speed 111 km/h);
– no. 6 – passage of the passenger train ED160 (electric multiple unit) (speed 105 km/h).

The train passages were recorded during acoustic tests carried out at the test section located in Nowy Dwór Mazowiecki, during real scheduled train passages on railway line no. 9 (LK-9), on the Legionowo–Nasielsk section. The audio signal (in WAVE format) was recorded using SVAN 979 sound analyser with GRAS 40AE microphone, with a sampling frequency of 48 kHz. For the purpose of the research objective, the microphone was located at a distance of 7.5 m from the track axis, at a height of approximately 2 m above the rails. The passing speeds of the individual trains whose acoustic signals were recorded are given in the brackets above.

Figures 8–13 present spectra of the emitted acoustic signals – sound pressure levels in 1/3 octave bands measured for individual signals within the frequency range of 20 Hz to 16000 Hz. The individual spectra were measured at a point marked as P0, located near the loudspeaker. All spectra are averaged for the measurement time covering: in the case of signal no. 1 (pink noise), the time is 60 s, and in the case of signals no. 2 to 6 – the duration of recorded train passages (from 33 to 107 s).

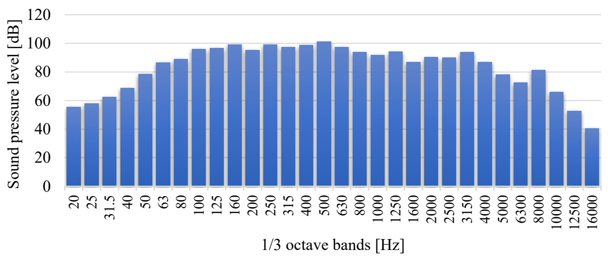

Fig. 8. Spectrum of the acoustic signal emitted by the sound source, no. 1 – pink noise.

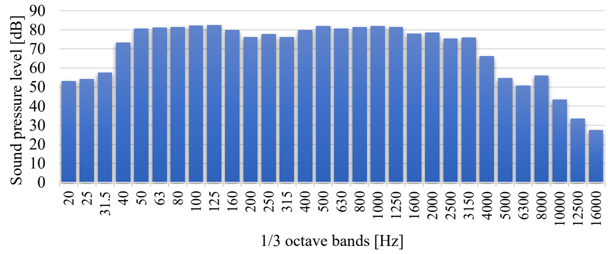

Fig. 9. Spectrum of the acoustic signal emitted by the sound source, no. 2 – Pendolino ED250.

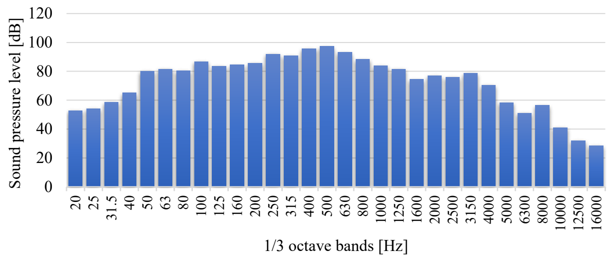

Fig. 10. Spectrum of the acoustic signal emitted by the sound source, no. 3 – ET22.

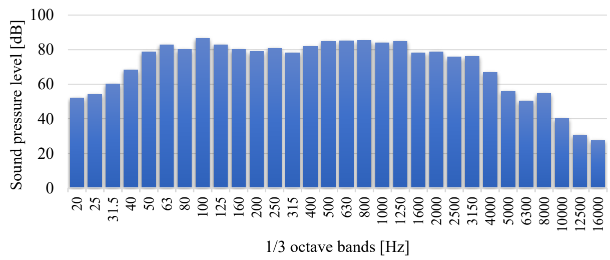

Fig. 11. Spectrum of the acoustic signal emitted by the sound source, no. 4 – EP09 composition 1.

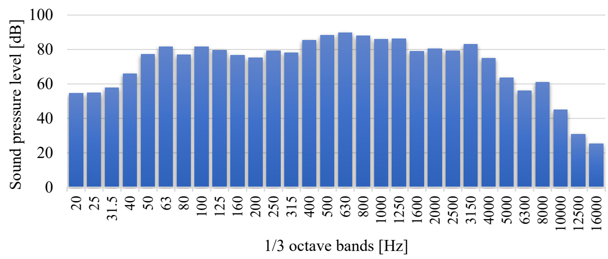

Fig. 12. Spectrum of the acoustic signal emitted by the sound source, no. 5 – EP09 composition 2.

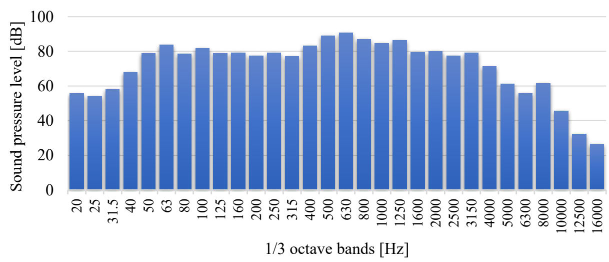

Fig. 13. Spectrum of the acoustic signal emitted by the sound source, no. 6 – ED160.

Environmental conditions during the measurements were monitored, controlled and recorded from

10 a.m. to 2:30 p.m. to prove that they had no negative impact on the reliability of the test results. Thanks to the SVAN PC++ software, the temperature, air humidity and wind speed were correlated with the measurement results obtained in the individual time intervals. The average wind speed was 1.65 m/s, the air temperature ranged from 15.6°C to 17.4°C, the atmospheric pressure was 1006 hPa, and the relative humidity ranged from 35% to 41%.

4. Results

4.1. Preliminary remarks

The tests were carried out on an open-air test stand, which does not provide a full reproduction of the conditions for real train passages. The technical limitations of the performed experiments meant that a number of simplifying assumptions had to be adopted which were relevant to the conclusions drawn from the study.

First of all, different propagation paths of the sound waves cause the phenomena of diffraction and interference of the sound waves generated by the loudspeaker, which, in practice, precludes the possibility of carrying out tests under free acoustic field conditions. Moreover, the sound signal generated by a sound source located at a fixed point (loudspeaker) does not make it possible to analyse the interaction between the moving vehicle and the environment: it is not possible to take into account the influence of phenomena caused by the air flow at the locomotive and carriages, or the dynamics of vibrating rails forced by the interaction of the wheel-rail pair (variable load on the rail, wheel passes over the joints, etc.).

Nevertheless, in the authors' opinion, the results of the research presented in this paper are of cognitive significance, and are important from the point of view of future application of the developed sound absorbing panels on various railway track systems.

4.2. Selection of the measurement point

When conducting acoustic tests under field conditions, the location of the microphone recording the signals used to interpret the studied parameters is crucial for the representativeness of the obtained results. As presented in Subsec. 3.2, the sound signals were registered synchronously with five microphones: the first one was located in the immediate vicinity of the source (in front of the test objects), the other four microphones were placed at different heights at two points: 1.4 m and 2.8 m behind the tested section of the track (Fig. 6).

As is known, the sound pressure in the free field is inversely proportional to the square of the distance from the sound source. This means that the decrease in the sound pressure between a point located in a distance of l_1 to a point located l_2 from the source is given by the formula:

$$\Delta L = 20 \log \left(\frac{l_1}{l_2} \right), \tag{2}$$

where ΔL – decrease in the sound level [dB], l_1 – distance between the first microphone and the source, l_2 – distance between the second microphone and the source.

If we take into account the distances as shown in Fig. 6, then for the free field the sound levels at points P1 and P2 will be about 20.9 dB lower than at point P0, while at points P3 and P4 they will be about 23.3 dB lower than at point P0. Thus, the difference between the sound levels at points P1, P2 and P3, P4 is about 2.4 dB.

Differences in sound levels of the signals registered during the experimental tests (averaged in 1/3 octave bands for all test signals) deviate from those calculated according to Eq. (2). The actual differences between the average sound levels in the immediate vicinity of the source (point P0) and the levels at points P1, P2, P3, P4 are 14.3, 15.0, 16.5, and 17.8 dB, respectively. The discrepancies between the values from theoretical calculations and those obtained from measurements reflect the influence of the test stands on the sound propagation between the source and the receivers (measurement microphones). This results in the limited usefulness of the signals registered at points P1 and P2 for comparing the effectiveness of solutions developed to reduce noise nuisance due to sound propagation disturbances in the near sound field.

It should be noted that the averaged (in 1/3 octave bands and for all test signals) differences in the sound levels of the signals recorded during the experimental tests at points P1 and P2 versus P3 and P4 (2.2 and 2.7 dB, respectively) only slightly deviate from the value calculated according to Eq. (2). Therefore, both points P3 and P4 located 2.8 m from the track can be considered useful for analysing the effectiveness of the tested vibroacoustic isolators, with the indication of point P3 (microphone 1.5 m above the ground surface).

4.3. Test results

Measurements of the sound absorption and dissipation characteristics of prototype vibroacoustic isolators were carried out in four configurations of the test stand, described in detail in Subsec. 3.2. In each configuration, simultaneous measurements of the sound pressure level were carried out in 1/3 octave bands in the mid-frequency range of 20 to 16000 Hz, at the same measurement points, for the same six excitation signals described in Subsec. 3.3.

As a result of the analysis, signal no. 1 – the pink noise emitted from the generator, which is commonly used in building acoustics and in the research on sound absorption properties of various types of materials – was treated as the leading signal. The results obtained for other acoustic signals were treated as supplementary data, because, according to the authors, the use of the train passes (recorded and subsequently emitted during the conducted measurements) as forcing signals is limited. However, in the authors' opinion, the obtained results are sufficient to demonstrate the technology readiness level specified in the realised research project.

Figure 14 presents the effectiveness of the solution D_B for the excitation with signal no. 1 (pink noise) determined from the results of the sound pressure level in 1/3 octave bands measured in the point P3, which was taken as the leading point.

Figure 15 shows the efficiency of the solution D_B for the average response of the vibroacoustic isolator system to excitation with signals no. 2 to 6 (acoustic signal emitted by passage of trains) determined from the results of the sound pressure level in 1/3 octave bands measured at point P3.

Analysing the test results in 1/3 octave bands for the mid-frequency range from 20 to 200 Hz, a reduction in sound pressure level values was found in each band after the application of the vibroacoustic isolators on both structures. In the ballastless system, for the frequencies from 20 to 40 Hz, very high values of the index determining the effectiveness of the solution from 4.8 to 7.7 dB were obtained. However, given the significant lengths of sound waves propagating in the air in the range of the analysed frequencies (from 8 to 17 m), it is unrealistic to obtain such values in real conditions due to the fundamental problem of limiting propagation of such long waves.

In the case of the mid-frequency range from 250 to 2500 Hz, a large variation was found in the obtained sound pressure level results, further varied by the structure under the isolators. For the ballasted system, the pressure level reductions were obtained for four bands with mid-frequencies of 250, 1000, 1250, and 2500 Hz. In contrast, for the ballastless system, the effectiveness of the panels was demonstrated for five bands with mid-frequencies of 250, 315, 400, 1250, and 1600 Hz. Only in two cases did the results coincide.

For the mid-frequency range from 3150 to 16 000 Hz, a reduction in the sound pressure level was found only for the panels located on the ballastless

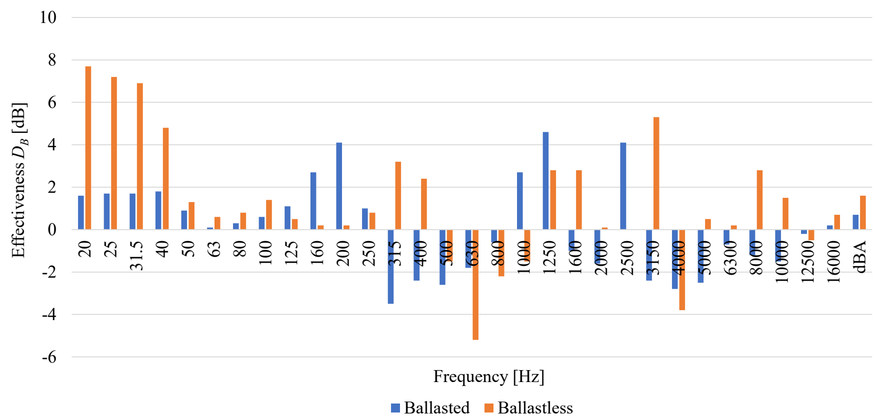

Fig. 14. Effectiveness D_B of the vibroacoustic isolators determined at point P3 for signal no. 1 (pink noise) in 1/3 octave bands with mid-frequencies: 20–16 000 Hz.

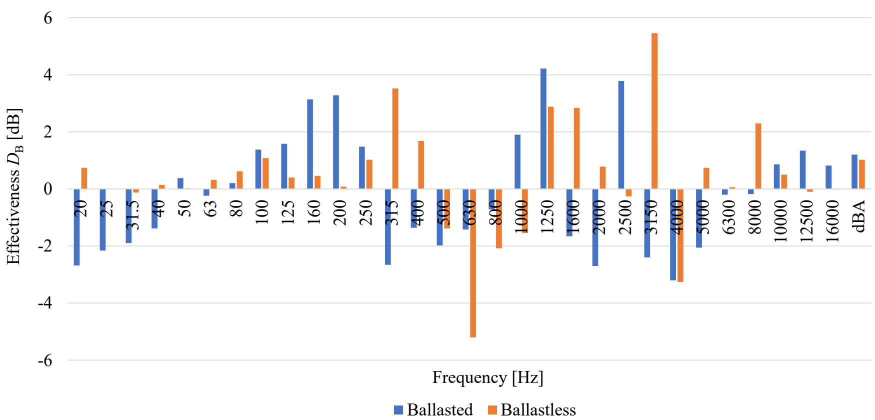

Fig. 15. Effectiveness D_B of the vibroacoustic isolators determined at point P3 for signals no. 2–6 (acoustic signal emitted by passage of trains) in 1/3 octave bands with mid-frequencies: 20–16 000 Hz.

structure. Figure 15 shows additionally the resultant effectiveness of the vibroacoustic isolators for the A-weighted sound level [dBA] – in both systems the D_B index has a positive value (0.7 dB for the ballasted system and 1.6 dB for the ballastless system).

4.4. Discussion of results

Due to the fact that the measurement point P3 was chosen as representative for the analysis and interpretation of the obtained results, the discussion of results is limited to the ones registered at P3.

Based on the test results (differences in the sound levels registered at point P3 before and after the application of prototype sound absorbing panels), it was found:

- ballasted track system with vibroacoustic isolators: in the case of the pink noise, a reduction in the sound pressure level was observed in 1/3 octave bands in the mid-frequency ranges of 20–250 Hz, 1000–1250 Hz, and in the bands of 2500 and 16 000 Hz. In the remaining frequency bands, no positive changes were observed – differences in the results were negative. For the A-weighted sound level, a reduction of 0.7 dB was found. In the case of other emitted signals, except for the passage of a freight train, a decrease in the A-weighted sound level was observed in the range of 1.1 dB (for signal no. 5 – passage of the old-type passenger train EP09) to 1.9 dB (for signal no. 6 – passage of the passenger train ED160, electric multiple unit);
- ballastless track system with vibroacoustic isolators: in the case of the pink noise, a reduction in the sound pressure level was observed in 1/3 octave bands in the mid-frequency ranges of 20–400 Hz, 1250–2000 Hz, 5000–10 000 Hz, and in the bands of 3150 and 16 000 Hz. In the remaining frequency bands, no positive changes were observed. For the A-weighted sound level, a reduction of 1.6 dB was found. In the case of all other emitted signals, a decrease in the A-weighted sound level was observed in the range of 0.2 dB (for signal no. 3 – passage of the freight train ET22) to 1.8 dB (for signal no. 6 – passage of the passenger train ED160, electric multiple unit).

It should be emphasized that the results discussed above were obtained in experiments that did not reproduce the influence of a number of elements (indicated in Subsec. 4.1), which are significant for the analysed phenomena. However, the authors' experience in the research on acoustic properties of railway systems allows them to state that, although the applied original methodology with consciously adopted simplifications initiates various doubts, it does not undermine the validity of this type of research or the reliability of its performance.

5. Conclusions

In the present study an original field experiment was proposed as a tool to verify the effectiveness of the developed prototype vibroacoustic isolators aimed at reducing railway noise emitted to the environment. A set of porous concrete sound absorbing panels was tested on two types of track structures: ballasted and ballastless systems.

Results of the measurements confirm the required noise attenuation and dispersion capacity of the tested vibroacoustic isolators (a system consisting of two types of panels). The condition "X–0.5 dB" for the designed solutions has been fulfilled (where X is the initial value of the noise level measured for the reference track structure). The results confirm the higher effectiveness in reducing railway noise emitted to the environment after the installation of the prototype sound absorbing panels on the ballastless structure. The findings of this study coincide with the experience of foreign railway infrastructure managers (e.g., from Germany and China), where vibroacoustic isolators (usually made of porous concrete) are used and show the highest effectiveness in reducing noise levels mainly when using ballastless systems. Currently, Poland lacks a proper testing site with the ballastless track structure, as such systems are used on the PKP PLK S.A. network to a marginal extent. Consequently, there was no technical possibility to carry out reliable tests of the noise level in the real environment.

However, the results of unconventional experiments carried out on an improvised test stand using the recorded acoustic signals confirmed the effectiveness of the developed solutions of the track structure components aimed at limiting the noise nuisance caused by railway traffic. Taking into account the characteristics of the acoustic field when selecting the spatial location of the sound measuring point, makes it possible to solve tasks aimed at optimising the developed structural solutions. The examples presented in this study demonstrate the versatility of the proposed concept of comparative research and show good prospects for their further use.

Funding

The research presented in the publication was carried out as part of the project "Innovative solutions of people and the environment protection against rail traffic noise". The project is co-financed by the European Union from the European Regional Development Fund under the Smart Growth Operational Programme and by PKP PLK S.A. within the framework of the BRIK.

References

1. de Vos P. (2016), *Railway noise in Europe, state of the art report*, International Union of Railways (UIC).

2. European Environment Agency (2020), *Environmental noise in Europe – 2020*, Technical Report, Copenhagen, Denmark.
3. Glickman G.M., Bensing S.J., Carman R. (2011), An examination of trackbed sound absorptive panels for minimizing wayside noise from rail transit, [in:] *Proceedings of Noise-Control 2011*, Portland, Oregon.
4. Groll W., Sowiński B., Konowrocki R. (2023), Study of transitional phenomena in rail vehicle dynamics in relation to the reliability and operational state of the continuous welded rail track in terms of rail joints, *Eksploatacja i Niezawodność – Maintenance and Reliability*, **25**(1): 7, doi: 10.17531/ein.2023.1.7.
5. Hong B.-K., Son H.-Y., Seo S.-W., Lee D.-H. (2005), A study on the sound absorptive characteristics and performance of parallel perforated plate systems, *Journal of Korean Society for Noise and Vibration Engineering*, **15**(9): 1003–1008, doi: 10.5050/ksnvn.2005.15.9.1003.
6. ISO 10847:1997 (1997), Acoustics – In-situ determination of insertion loss of outdoor noise barriers of all types.
7. ISO 3095:2013 (2013), Acoustics – Railway applications. Measurement of noise emitted by railbound vehicles.
8. Kraśkiewicz C., Chmielewska B., Zbiciak A., Al Sabouni-Zawadzka A. (2021b), Study on possible application of rubber granulate from the recycled tires as an elastic cover of prototype rail dampers, with a focus on their operational durability, *Materials*, **14**(19): 5711, doi: 10.3390/ma14195711.
9. Kraśkiewicz C., Mossakowski P., Zbiciak A., Al Sabouni-Zawadzka A. (2021a), Experimental identification of dynamic characteristics of a track structure influencing the level of noise emission, *Archives of Civil Engineering*, **67**(4): 543–557, doi: 10.24425/ace.2021.138517.
10. Lázaro J., Pereira M., Costa P.A., Godinho L. (2022), Performance of low-height railway noise barriers with porous materials, *Applied Sciences*, **12**(6): 2960, doi: 10.3390/app12062960.
11. Lee H.J., Oh S.T., Lee D.-J. (2016), An optimal mix design of sound absorbing block on concrete ballast in urban train tunnel, *Journal of Korean Tunnelling and Underground Space Association*, **18**(1): 75–82, doi: 10.9711/KTAJ.2016.18.1.075.
12. Li Y., Guo C. (2017), Numerical optimization for acoustic performance of the micro-perforated plate and its application in high-speed trains, *Journal of Vibroengineering*, **19**(6): 4724–4741, doi: 10.21595/jve.2017.17019.
13. Matej J., Orliński P. (2023), New possibilities to reduce wheel and rail wear in the operation of metro wagons on curved track with small curve radii without considering propulsion and braking systems, *Eksploatacja i Niezawodność – Maintenance and Reliability*, **25**(1): 15, doi: 10.17531/ein.2023.1.15.
14. Oh S.-T., Lee D.-J., Lee D.-H. (2017), Developments of monitoring system to measure sound absorbing coefficient and structural stability of sound absorbing panel on the concrete track in the urban train tunnel, *Journal of Korean Tunnelling and Underground Space Association*, **19**(1): 1–9, doi: 10.9711/KTAJ.2017.19.1.001.
15. Scossa-Romano E., Oertli J. (2012), *Rail dampers, acoustic rail grinding, low height noise barriers, a report on the state of the art*, International Union of Railways (UIC), Schweizerische Bundesbahnen SBB (SBB CFF FFS), https://uic.org/IMG/pdf/2012_dampers_grinding_lowbarriers.pdf (access: 16.03.2023).
16. Shimokura R., Soeta Y. (2011), Characteristics of train noise in above-ground and underground stations with side and island platforms, *Journal of Sound and Vibration*, **330**(8): 1621–1633, doi: 10.1016/j.jsv.2010.10.021.
17. Thompson D. (2008), *Railway Noise and Vibration: Mechanisms, Modelling and Means of Control*, Elsevier, Amsterdam, The Netherlands.
18. Vogiatzis K., Vanhonacker P. (2015), Noise reduction in urban LRT networks by combining track based solutions, *The Science of the total environment*, **568**: 1344–1354, doi: 10.1016/j.scitotenv.2015.05.060.
19. World Health Organization (2018), *Environmental noise guidelines for the European region*, Copenhagen, Denmark.
20. Yori A. (2020), Method for calculating the sound absorption coefficient for a variable range of incidence angles, *Archives of Acoustics*, **45**(1): 67–75, doi: 10.24425/aoa.2020.132482.
21. Zbiciak A., Kraśkiewicz C., Dudziak S., Al-Sabouni-Zawadzka A., Pełczyński J. (2021), An accurate method for fast assessment of under slab mats (USM) performance in ballastless track structures, *Construction and Building Materials*, **300**: 123953, doi: 10.1016/j.conbuildmat.2021.123953.
22. Zhang X., Jeong H., Thompson D., Squicciarini G. (2019), The noise radiated by ballasted and slab tracks, *Applied Acoustics*, **151**: 193–205, doi: 10.1016/j.apacoust.2019.03.012.
23. Zhao C., Wang P., Wang L., Liu D. (2014), Reducing railway noise with porous sound-absorbing concrete slabs, *Advances in Materials Science and Engineering*, **2014**: 206549, doi: 10.1155/2014/206549.

Archives of Acoustics
Vol. 49, No. 1, pp. 73–81 (2024), doi: 10.24425/aoa.2024.148769

Research Paper

Design and Verification of Sector Vortex Archimedean Spiral Phased Array Transducer for Improving Focus Acoustic Pressure

Xiaodan LU[(1)], Deping ZENG[(1), (2)*]

[(1)] *State Key Laboratory of Ultrasound in Medicine and Engineering, College of Biomedical Engineering Chongqing Medical University*
Chongqing, China

[(2)] *National Engineering Research Center of Ultrasound Medicine*
Chongqing, China

*Corresponding Author e-mail: 102300@cqmu.edu.cn

(*received April 8, 2023; accepted October 3, 2023; published online January 8, 2024*)

The emergence of high-intensity focused ultrasound applications brings great potential to establish non-invasive therapeutic treatment in place of conventional surgery. However, the development of ultrasonic technology also poses challenges to the design and manufacture of high-power ultrasound transducers with sufficient acoustic pressure. Here, the design of a sector vortex Archimedean spiral phased array transducer that is able to enhance focal acoustic pressure is proposed by maximizing the filling factor of the piezoelectric array. The transducer design was experimentally verified by hydrophone measurements and matched well with acoustic simulation studies. The focal deflection was shown to be feasible up to ±9 mm laterally and up to ±20 mm axially, where the effective focal acoustic pressure can be maintained above 50% and the level of the grating lobe below 30%. Furthermore, a homogeneous pressure distribution without secondary focus was observed in the pre-focal region of the transducer. The rational design of a high-intensity focused ultrasound transducer indicates promising development in the treatment of deep tissue thermal ablation for clinical applications.

Keywords: phased array transducer; Archimedean spiral; high-intensity focused ultrasound (HIFU); focal deflection.

1. Introduction

High-intensity focused ultrasound (HIFU) is rapidly developing as an ideal non-invasive alternative to conventional surgery, by using the good penetrability of ultrasound waves in human tissues. HIFU can be used for the treatment of benign and malignant solid tumors, providing a non-invasive and green treatment technology with broad development prospects. Clinically, this technology has been used to treat uterine fibroids (Liu *et al.*, 2017), liver cancer (Fukuda *et al.*, 2011), breast cancer (Feril *et al.*, 2021), etc. To enable effective HIFU treatment, the ultrasound transducer plays a pivotal role in precisely focusing the ultrasound on the tumor tissue in the human body and forming a high temperature above 65°C in the target area within a short time. The thermal effect, mechanical effect, and cavitation effect of ultrasound at the focus synergistically cause protein denaturation, coagulative necrosis of tissue cells, or irreversible serious damage. This achieves the objective of treating tumors without damaging the surrounding tissues (Shehata Elhelf *et al.*, 2018).

Among the currently available transducers, the phased array transducer is an ideal candidate for delivering focused ultrasonic waves in HIFU treatment. It consists of multiple piezoelectric elements, each with a specific excitation signal. By controlling the phase of the element excitation signals, single or muti-focus as well as axial or lateral focal deflection can be achieved without moving the transducer. Assisted by electronic focusing, the ultrasonic beam can move flexibly and quickly to track the movement of organs, for example, the kidney and liver during breathing. In addition, wave aberration can be compensated by heterogeneous media to avoid obstacles such as ribs and forming multiple lesions (Auboiroux *et al.*, 2011; Lean, Zhou, 2019).

In clinical applications, the critical requirements for HIFU usage are: 1) providing high focal acoustic pressure; 2) reducing levels of sidelobe and grating lobe; 3) increasing the focal deflection range to ensure sufficient focal intensity for deep tissue thermal ablation. However, problems in phased array focused ultrasonic devices include an acoustic pressure decreases and the occurrence of excessive grating lobe caused by focus deflection. Leading to thermal damage in the proximal region (such as the skin). Particularly, this effect becomes pronounced as the distance of focal deflection increases (Ebbini, Cain, 1991).

To improve focus quality, especially for deep tissue thermal ablation, the rational design of a phased array transducer should follow several principles, including a large aperture and close packing of single elements, where the distance between each element center should not exceed half of the wavelength (Hynynen, Jones, 2016; Ramaekers *et al.*, 2017a). In earlier studies, regular and periodic arrays were used, leading to an excessively high level of the grating lobe limiting focal deflection (Daum, Hynynen, 1999; Goss *et al.*, 1996). Later, Goss *et al.* (1996) introduced random arrays to optimize array performance and effectively reduce the level of the grating lobe. Gavrilov and Hand (2000) compared the acoustic field between regular and random arrays through a numerical simulation and element size modification. The authors confirmed improved focal quality in random arrays, while a lower focal acoustic pressure was observed attributed to the low filling factor of the elements. In 2015, Gavrilov *et al.* proposed the spiral array arrangement as a solution to increase the intensity of focus and suggested that the filling factor of elements directly affected the intensity of the focus acoustic pressure. Therefore, the spiral array is one of the solutions to the above problems. Notably when the piezoelectric material is constrained by power limitations, the radiation area of the transducer should be increased as much as possible (Rosnitskiy *et al.*, 2020), so that the active surface of the transducer and the close filling of elements contribute in achieving the desired focal acoustic pressure and spread the generated acoustic pressure uniformly in the pre-focal region to avoid damage caused by undesired hotspots (Payne *et al.*, 2011).

Furthermore, compared to other spiral structures (16-spirals (Rosnitskiy *et al.*, 2018), Fermat's spiral (Ramaekers *et al.*, 2017b), etc.), the Archimedean spiral phased array provides a simple structure, flexible element arrangement, and a wider deflection range (Morrison *et al.*, 2014). Wang *et al.* (2021) proposed an Archimedean spiral PMUT array, which can generate greater axial acoustic pressure, with acoustic emission pressure 18% higher than that of traditional array structure. However, in current studies, most elements of the Archimedean spiral array are in circular shape, resulting in relatively low filling factors of the array. Therefore, in this work, we sought a novel approach to build a high-power phased array transducer with a small number of transducer elements while allowing sufficient focal deflection and maintaining simplicity in design.

Our research approach is to restrict the number of array elements to 128 and improve the design of the Archimedean spiral array by using sector vortex elements to further increase the filling factor. These elements were distributed on a spherical crown with an opening diameter of 200 mm and a radius of curvature of 180 mm. In the authors' previous research (Lu, Zeng, 2023), the improved focal acoustic pressure was demonstrated through simulations and compared to a configuration with circular elements, the obtained focal acoustic pressure in the sector vortex elements was 32.28% higher compared to the circular configuration. Consequently, according to the simulated structural parameters, a sector vortex Archimedean spiral phased array transducer that met the requirements was fabricated, the acoustic performance of the transducer was validated in acoustic field scanning experiments and the results were compared with acoustic simulations. This research provides feasible promises for the design and optimization of focused ultrasonic transducers.

2. Materials and method

2.1. Array design

Figure 1 shows the design of the 128-sector vortex Archimedean spiral array studied in this work. The sector vortex array elements are closely distributed on a spherical crown with an opening diameter of 200 mm, a radius of curvature of 180 mm, and 7.5 turns of Archimedean spiral. A circular hole with an inner diameter of 60 mm at the center is cut out to allow the B-ultrasound probe for image monitoring. The number of array elements is set to 128. The design of the array

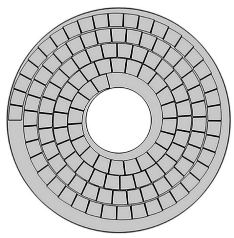

Fig. 1. Illustration of the 128-sector vortex Archimedean spiral phased array transducer design.

elements shape was first generated on a planar surface (Raju *et al.*, 2011) and then projected through the center onto a spherical crown with a radius of curvature 180 mm to ensure that the areas of elements were uniform. On the spherical crown, the projected area of a single sector vortex element was 173.35 mm^2, and a filling factor of the elements was about 70.55%. The filling factor can be calculated using Eq. (1) (Rosnitskiy *et al.*, 2020), where Σ_{act} is the sum of the effective areas of all array elements, and Σ is the total area of spherical crown radiation. Specific parameters of the 128-sector vortex Archimedean spiral phased array transducer are shown in Table 1:

$$\Psi = \left(\frac{\Sigma_{\text{act}}}{\Sigma}\right) \times 100\%. \tag{1}$$

Table 1. Parameters of Archimedean spiral phased array transducer.

Number of elements	Opening diameter [mm]	Inner diameter [mm]	Radius of curvature [mm]
128	200	60	180
F number	Shape of elements	Spiral turns	Filling factor
0.9	sector vortex	7.5	70.55%

2.2. Acoustic simulation

The acoustic fields of the 128-sector vortex Archimedean spiral phased array transducer were simulated by finite element analysis. The simulations only modeled linear acoustic effects, which were sufficient to evaluate the focal deflection capabilities of various array designs. The frequency of the acoustic simulations was taken as 1 MHz. Assuming that the propagation medium is homogeneous, the density of water was considered as 1000 kg/m^3, and the speed of sound as 1500 m/s. A pressure of 1 Pa was applied to each array element, and finally, the size of each grid was adjusted to a minimum of $\lambda/8$ and a maximum of $\lambda/6$, where λ is the wavelength.

2.3. Transducer fabrication

The single element of the array is made of P8-type piezoelectric ceramic, with the concave side being the negative electrode and the convex side being the positive electrode, and the frequency of the transducer is 1 MHz. The shell contains 128 mounting points to hold the element assembly in a suitable place, and the elements are connected by positive and negative electrode leads. The center hole of the shell and the array allow for the placement of a B-ultrasound probe for image monitoring, further enhancing the practicability of the transducer. An image of the as-fabricated 128-sector vortex Archimedean spiral phased array transducer is shown in Fig. 2.

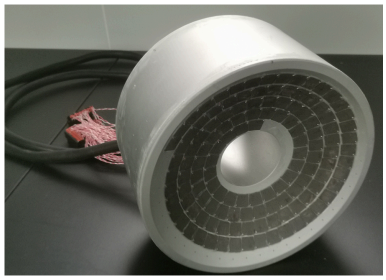

Fig. 2. Image of the as-fabricated 128-sector vortex Archimedean spiral phased array transducer.

2.4. Hydrophone measurements

Hydrophone measurements were performed in a degassed water tank with the as-fabricated transducer and compared with the acoustic simulation to verify the acoustic performances. The whole measurement system contained the phased array control system, the acoustic field scanning system, and the three-dimensional motion system, as shown in Fig. 3. In the experiment, the transducer was immersed in degassed water, the surface of the transducer was perpendicular to the hydrophone, and the upper computer software controlled the 128-channel digital generator to produce signals driving the transducer to transmit ultrasonic waves. The hydrophone was moved by a three-dimensional stepping motor to scan and measure in three orthogonal directions with a step length of 0.1 mm, acquiring and processing the signal. The measured signal voltage was converted into acoustic pressure through the sensitivity provided by the hydrophone manufacturer. To evaluate the focal deflection capability, the delay time was calculated based on known array element coordinates to control the focus deflection in three orthogonal directions. This process allows to determine changes in focus acoustic pressure, levels of the sidelobe and grating lobe, and focal plane acoustic pressure distributions. Furthermore, focal plane acoustic pressure distributions were obtained at 15, 25, and 50 mm in the pre-focal region to observe the pre-focal acoustic pressure distribution.

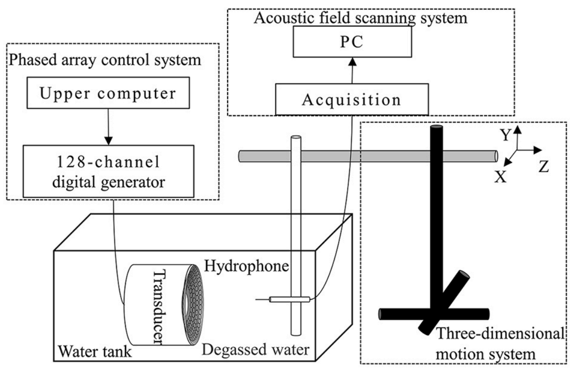

Fig. 3. Schematic diagram of the acoustic field measurement device.

3. Results and discussion

3.1. Focusing at the geometric focus

Figure 4 illustrates the one-dimensional normalized acoustic pressure distributions at the geometric focus obtained in three orthogonal directions, obtained through hydrophone measurements and acoustic simulations under free-field conditions. It can be seen that the results between the hydrophone measurements and acoustic simulations are in good agreement, demonstrating symmetry in positive and negative directions. Specifically, the experimental results exhibit sidelobe levels of 21% laterally (Figs. 4a and 4b) and 24% axially (Fig. 4c), while the simulation values of 25% and 19%, respectively, with a maximum error of 5%. No evidence of strong sidelobes and grating lobe approaching −10 dB of the main lobe is observed. Furthermore, Figs. 4a and 4b show that the sidelobe levels of the experimental results are lower than the simulation results, which is likely because of the inevitable imperfect manufacturing process of the actual transducer implementation, including the loss of regularity in element placement, thereby improving the performance of acoustic field distribution.

3.2. Focal deflection capability

According to the criteria for assessing the focal deflection capability of phased array transducers established in earlier studies (Ebbini, Cain, 1991; Ramaekers *et al.*, 2017a; Rosnitskiy *et al.*, 2018), focal deflection is effective when the decrease in acoustic pressure after focal deflection is less than 50% and it is safe if the level of the grating lobe is less than 30% when the focus is deflected. In this work, the measurements for each point are normalized and evaluated according to this criterion (the dashed lines in Figs. 5 and 6 are the standard thresholds).

Figure 5 illustrates the focal acoustic pressure obtained by the hydrophone measurements and acoustic simulations after controlling the focus deflected in three orthogonal directions under free-field conditions. The focus is deflected up to ±10 mm in steps of 2 mm

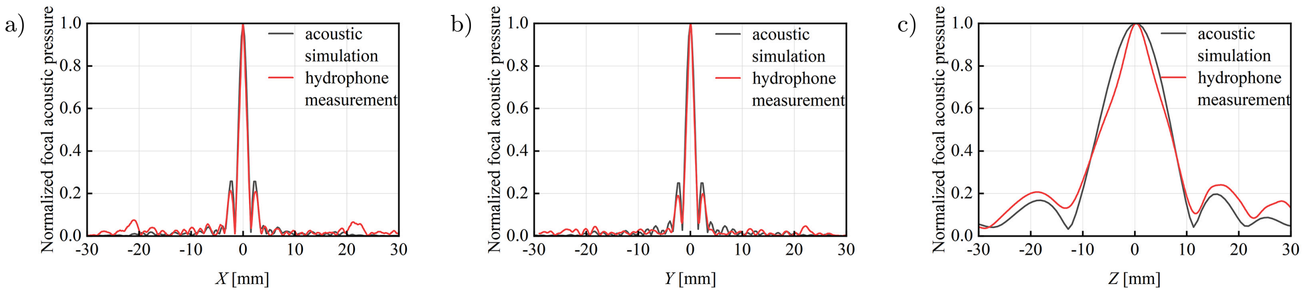

Fig. 4. Illustration of the one-dimensional normalized acoustic pressure distributions at the geometric focus on the X-axis (a), the Y-axis (b), and the Z-axis (c), obtained by the hydrophone measurements and the acoustic simulations.

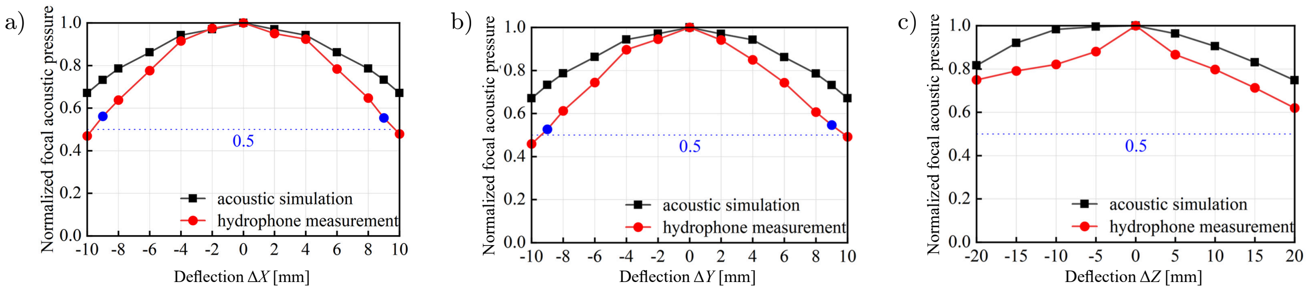

Fig. 5. Illustration of the normalized focus acoustic pressure levels deflected on the X-axis (a), the Y-axis (b), and the Z-axis (c), obtained by the hydrophone measurements and the acoustic simulations.

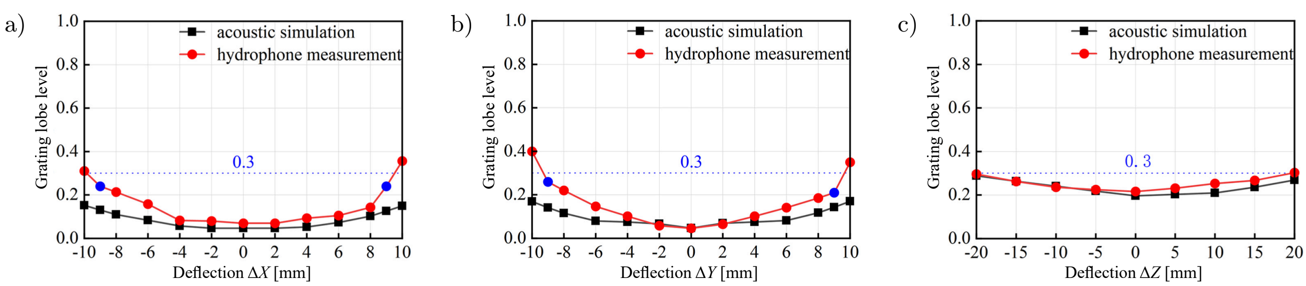

Fig. 6. Illustration of the grating lobe levels for focus deflected on the X-axis (a), the Y-axis (b), and sidelobe levels for focal deflected on the Z-axis (c), obtained by the hydrophone measurements and the acoustic simulations.

on the X-axis (Fig. 5a) and Y-axis (Fig. 5b), and up to ±20 mm in steps of 5 mm on the Z-axis (Fig. 5c). A good symmetry between the positive and negative deflections of the focus is observed, and as the focal deflection distance increases, the focal acoustic pressure decreases linearly. The blue dots in Figs. 5a and 5b indicate that focal deflection appears to be feasible for lateral deflections up to ±9 mm, where the focal acoustic pressure level is reduced by approximately 45%. At ±10 mm, the pressure level drops by more than 50%, however, in the acoustic simulations this drop is 33%. In Fig. 5c, the acoustic pressure reductions of 25% for the negative direction and 38% for the positive direction are observed in experimental results for the focal deflection up to ±20 mm on the Z-axis. These values are 5% and 10% higher than the acoustic simulations. Therefore, the focal pressure reductions and the margin of error are smaller compared to those observed for lateral deflections.

Figure 6 illustrates the levels of the grating lobe and sidelobe corresponding to Fig. 5. When the focus is deflected up to ±10 mm, the proportion of positive and negative grating lobe levels increases from 7 to 35 and 31%, respectively, on the X-axis (Fig. 6a), and increases from 4.6 to 35 and 40%, respectively, on the Y-axis (Fig. 6b), while the results of the acoustic simulations are less than 17%. Therefore, the blue dots in Figs. 6a and 6b show that the proportion of the grating lobe is less than 26% at ±9 mm on the lateral axis, which satisfies the safety standard. When the focus is deflected up to ±20 mm, the sidelobe levels are less than 29% on the Z-axis (Fig. 6c), and the margin of error is less than 3% compared to the acoustic simulations. Thus, the results of hydrophone measurements on the Z-axis are in good consistency with the acoustic simulations, while the grating lobe levels for the lateral deflections are generally higher than those simulated.

To better observe the acoustic field distributions when the focus is deflected laterally, Fig. 7 shows the normalized transversal acoustic pressure distributions with the focus deflected up to ±10 mm in steps of 5 mm on the X-axis, comparing hydrophone measurements and acoustic simulations. As the focal deflection distance increases, both methods show a continuous increase in the levels of the grating lobe and sidelobe. At ±5 mm, hydrophone measurements match well with acoustic simulations, the sidelobe levels are less than 30% and the grating lobe levels are smaller. However, for deflections up to ±10 mm, corresponding to the results in Figs. 6a and 6b, the two-dimensional acoustic field distributions (Figs. 7a and 7e) clearly illustrate that the levels of the grating lobe near the focus are higher and their number is bigger than in the acoustic simulations.

Considering the effectiveness and safety of focused ultrasound, the experimental results indicate that the focal deflection range of this transducer is 40 mm for the Z-axis, 18 mm for the X-axis, and 18 mm for the Y-axis, while lateral deflection is 2 mm smaller than in the acoustic simulations. Therefore, the experimental results of the transducer show good correspondence with the acoustic simulations, especially on the Z-axis. However, the focal deflection is a bit worse on the lateral axis. As shown in Fig. 7, the transversal acoustic pressure distributions at ±10 mm show that the levels of grating lobe near the focus are higher and more numerous compared to the acoustic simulations.

Several possible explanations for this observation can be considered. The presence of the grating lobe is not only related to the focal deflection distance but also to the array structure of the phased array. Theoretically, when the distance between each element center is less than half of the wavelength, there are no grating lobes (Ellens *et al.*, 2015; Hynynen, Jones, 2016). However, this rule leads to a very small element size and an excessive number of elements, and the small size is extremely challenging in terms of transducer fabrication cost and matching. In this study, to improve the focus acoustic pressure, the sector vortex elements are used to maximize the filling factor, and the size of the elements limits the deflection of the focus on the lateral axis. Meanwhile, individual elements must have the same area (Rosnitskiy *et al.*, 2020), and the single area of elements used in this study has a difference of approximately 0.75 mm^2 due to the simulation and fabrication, complicating the matching and having some effect on the acoustic field characteristic or even deteriorating them. Furthermore, among hydrophone measurements, the hydrophone size (Umchid *et al.*, 2009) and the crosstalk between the wires may also have some influence on the acoustic field.

In summary, an advantage of the sector vortex array designed in this work is that the increased filling factor produces higher focal acoustic pressure under the same excitation while maintaining the focal deflection range, especially with no effect on the focal deflection capability on the Z-axis. Even though the acoustic pressure decreases more in the lateral axis, the sector vortex array still maintains higher focal acoustic pressure with the focus deflected up to ±10 mm. Consequently, the design of the sector vortex array is suitable for deep tissue treatment requiring high output power, and the array is found to be able to adequately deflect the focus in three dimensions to cover the required treatment volume. Additionally, the use of only one type of element leads to simpler electrical matching. In this work, the array design is the result of increasing the filling factor while keeping the number of elements fixed, and if a larger focal deflection range and higher output power are required, it is possible to reduce the area of array elements and increase the number of elements.

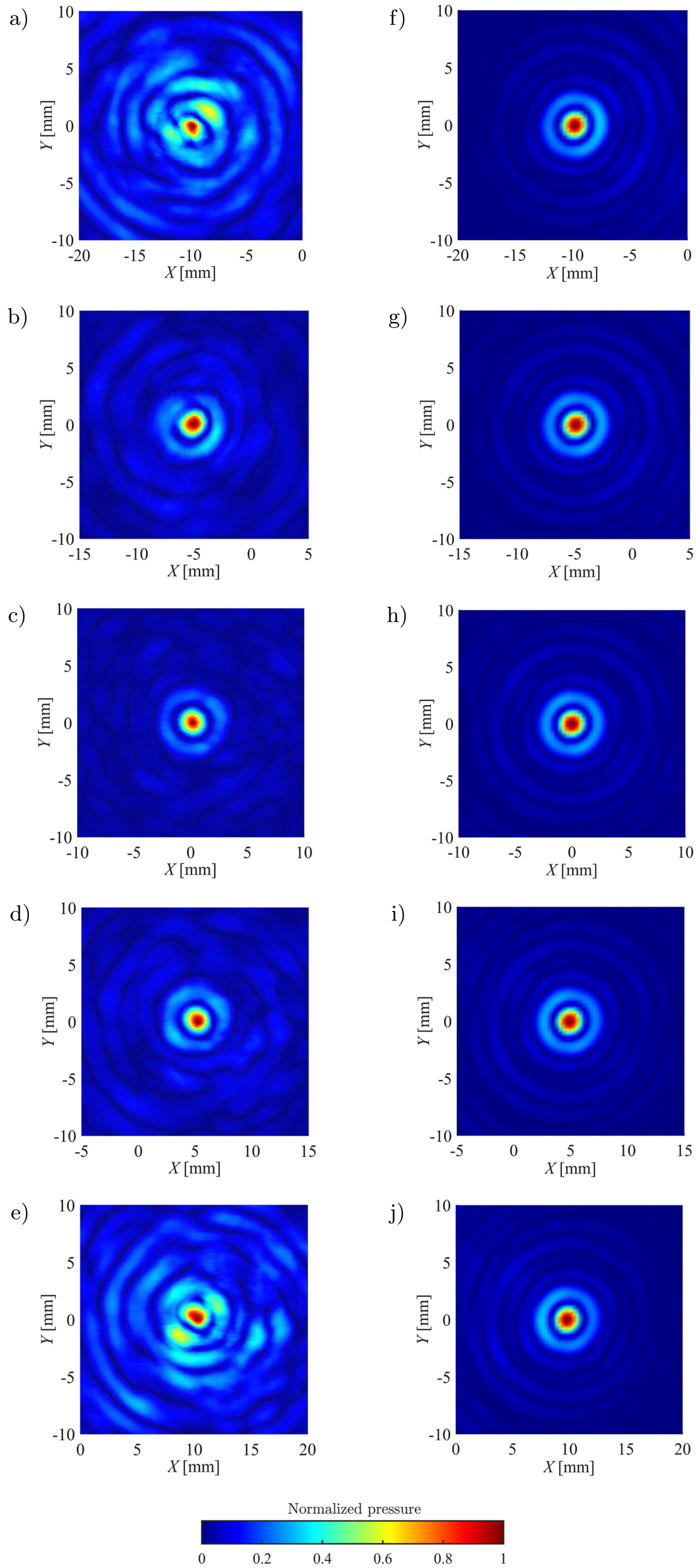

Fig. 7. Illustration of the normalized transversal acoustic pressure distribution with the focus deflected up to ±10 mm in steps of 5 mm and a field of view of 20 × 20 mm, obtained by the hydrophone measurements (a–e), and the acoustic simulations (f–j).

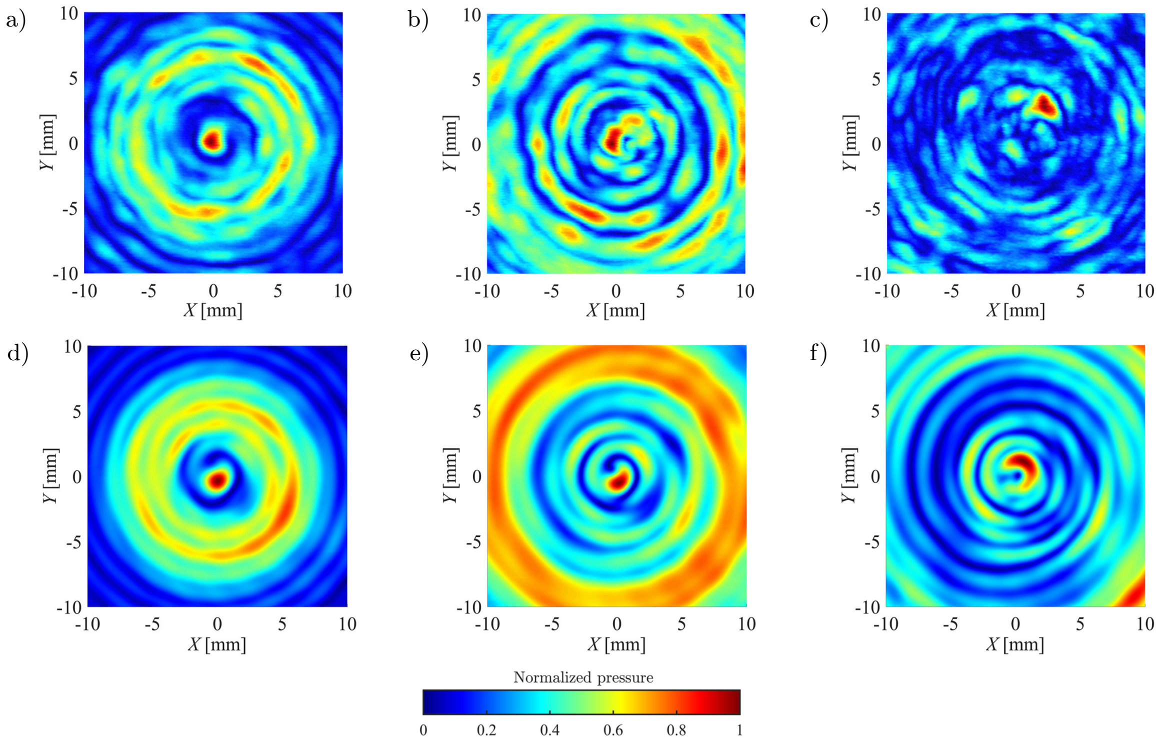

Fig. 8. Illustration of the normalized transversal acoustic pressure distributions in the pre-focal region of 15, 25, and 50 mm, and the field of view of 20 × 20 mm, obtained by the hydrophone measurements (a–c), and the acoustic simulations (d–f).

3.3. Pre-focal acoustic pressure distribution

To observe the pre-focal acoustic field distribution of the transducer, the normalized transversal acoustic pressure distributions were obtained respectively at 15, 25, and 50 mm in the pre-focal region by both hydrophone measurements and acoustic simulations, as shown in Fig. 8. The two methods show good correspondence at the three different distances, and more homogeneous acoustic pressure distributions are observed in the field of view without local hot spots.

In fact, the homogeneous propagation of acoustic pressure generated by the transducer in the pre-focal region, to avoid undesired local hotspots, is also one of the important factors to be considered for the phased array transducer design. Therefore, not only higher focal acoustic pressure and a large focal deflection range are required in evaluating the transducer design, but also avoiding secondary maximum and local hotspots in the pre-focal region (Köhler *et al.*, 2012; Ramaekers *et al.*, 2017a), which is conducive to the safety of clinical treatment. The hydrophone verification of the pre-focal acoustic pressure distributions of the transducer is given in Figs. 8a–c, and the results demonstrate good uniformity and consistency.

4. Conclusion

As a new generation of HIFU transducers – phased arrays offer the advantage of more flexible electronic focusing, higher focal acoustic pressure and a larger focal deflection range, which are critical in clinical applications. This study proposed the design of a sector vortex array based on the Archimedean spiral phased array transducer, which provided the maximum filling factor within the limits of the outer diameter and Archimedean spiral structure while ensuring non-periodic closely filled elements. Considering the safety and effectiveness, the transducer was evaluated through the comparison of hydrophone measurements and acoustic simulations to verify the feasibility of this array design, including the acoustic field distributions of geometric focus, focal deflection capability, and acoustic pressure distributions in the pre-focal region.

The results demonstrated that hydrophone measurements at the geometric focus were in good correspondence with acoustic simulations, and the margin of error was approximately 5%. The transducer's focal deflection capability was nearly identical, with a margin of error of 2%, and the acoustic pressure distributions in the pre-focal region were homogeneous. Therefore, it is indicated that a greater filling factor should be used for designing phased array transducer, enabling higher focal acoustic pressure. The irregular and non-periodic arrangement of the elements can reduce the level of the grating lobe, and the dense filling of elements contributes to the homogeneous distribution of acoustic pressure in the pre-focal region to avoid the local hotspots generated.

In conclusion, the design of the 128-sector vortex Archimedean spiral phased array transducer proposed in this work is suitable for clinical applications requiring high acoustic output power in deep tissues, the lateral focal deflection of ±9 mm and axis focal deflection of ±20 mm are also sufficient for clinical applications.

Declarations

The authors declare that they have no conflicts of interest. This article does not contain any studies with human or animal subjects performed by any of the authors.

References

1. Auboiroux V., Dumont E., Petrusca L., Viallon M., Salomir R. (2011), An MR-compliant phased-array HIFU transducer with augmented steering range, dedicated to abdominal thermotherapy, *Physics in Medicine & Biology*, **56**(12): 3563–3582, doi: 10.1088/0031-9155/56/12/008.

2. Daum D.R., Hynynen K. (1999), A 256-element ultrasonic phased array system for the treatment of large volumes of deep seated tissue, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **46**(5): 1254–1268, doi: 10.1109/58.796130.

3. Ebbini E.S., Cain C.A. (1991), A spherical-section ultrasound phased array applicator for deep localized hyperthermia, *IEEE Transactions on Biomedical Engineering*, **38**(7): 634–643, doi: 10.1109/10.83562.

4. Ellens N.P., Lucht B.B., Gunaseelan S.T., Hudson J.M., Hynynen K.H. (2015), A novel, flat, electronically-steered phased array transducer for tissue ablation: Preliminary results, *Physics in Medicine & Biology*, **60**(6): 2195–2215, doi: 10.1088/0031-9155/60/6/2195.

5. Feril L.B., Fernan R.L., Tachibana K. (2021), High-intensity focused ultrasound in the treatment of breast cancer, *Current Medicinal Chemistry*, **28**(25): 5179–5188, doi: 10.2174/0929867327666201111143206.

6. Fukuda H. *et al.* (2011), Treatment of small hepatocellular carcinomas with US-guided high-intensity focused ultrasound, *Ultrasound in Medicine & Biology*, **37**(8): 1222–1229, doi: 10.1016/j.ultrasmedbio.2011.04.020.

7. Gavrilov L.R., Hand J.W. (2000), A theoretical assessment of the relative performance of spherical phased arrays for ultrasound surgery, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **47**(1): 125–139, doi: 10.1109/58.818755.

8. Gavrilov L.R., Sapozhnikov O.A., Khokhlova V.A. (2015), Spiral arrangement of elements of two-dimensional ultrasonic therapeutic arrays as a way of increasing the intensity at the focus, *Bulletin of the Russian Academy of Sciences: Physics*, **79**(10): 1232–1237, doi: 10.3103/s106287381510010x.

9. Goss S.A., Frizzell L.A., Kouzmanoff J.T., Barich J.M., Yang J.M. (1996), Sparse random ultrasound phased array for focal surgery, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **43**(6): 1111–1121, doi: 10.1109/58.542054.

10. Hynynen K., Jones R.M. (2016), Image-guided ultrasound phased arrays are a disruptive technology for non-invasive therapy, *Physics in Medicine & Biology*, **61**(17): R206, doi: 10.1088/0031-9155/61/17/R206.

11. Köhler M., Mougenot C., Ylihautala M. (2012), Near-field heating of volumetric MR-HIFU hyperthermia, [in:] *11th International Symposium on Therapeutic Ultrasound. AIP Conference Proceedings*, **1481**(1): 180–184, doi: 10.1063/1.4757331.

12. Lean H.Q., Zhou Y. (2019), Acoustic field of phased-array ultrasound transducer with the focus/foci shifting, *Journal of Medical and Biological Engineering*, **39**(6): 919–931, doi: 10.1007/s40846-019-00464-z.

13. Liu Y., Ran W., Shen Y., Feng W., Yi J. (2017), High-intensity focused ultrasound and laparoscopic myomectomy in the treatment of uterine fibroids: A comparative study, *BJOG*, **124**(S3): 36–39, doi: 10.1111/1471-0528.14745.

14. Lu X., Zeng D. (2023), Simulation research on increasing the focus sound pressure of Archimedean spiral phased array transducer [in Chinese], *Technical Acoustic*, **42**(2): 263–268.

15. Morrison K.P., Keilman G.W., Kaczkowski P.J. (2014), Single archimedean spiral close packed phased array HIFU, [in:] *IEEE International Ultrasonics Symposium*, pp. 400–404, doi: 10.1109/ULTSYM.2014.0099.

16. Payne A., Vyas U., Todd N., de Bever J., Christensen D.A., Parker D.L. (2011), The effect of electronically steering a phased array ultrasound transducer on near-field tissue heating, *Medical Physics*, **38**(9): 4971–4981, doi: 10.1118/1.3618729.

17. Raju B.I., Hall C.S., Seip R. (2011), Ultrasound therapy transducers with space-filling non-periodic arrays, [in:] *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **58**(5): 944–954, doi: 10.1109/tuffc.2011.1895.

18. Ramaekers P., de Greef M., Berriet R., Moonen C.T.W., Ries M. (2017a), Evaluation of a novel therapeutic focused ultrasound transducer based on Fermat's spiral, *Physics in Medicine & Biology*, **62**(12): 5021–5045, doi: 10.1088/1361-6560/aa716c.

19. Ramaekers P., Ries M., Moonen C.T.W., de Greef M. (2017b), Improved intercostal HIFU ablation using a phased array transducer based on Fermat's spiral and Voronoi tessellation: A numerical evaluation, *Medical Physics*, **44**(3): 1071–1088, doi: 10.1002/mp.12082.

20. Rosnitskiy P.B., Sapozhnikov O.A., Gavrilov L.R., Khokhlova V.A. (2020), Designing fully populated

phased arrays for noninvasive ultrasound surgery with controlled degree of irregularity in the arrangement of elements, *Acoustical Physics*, **66**(4): 352–361, doi: 10.1134/s1063771020040090.

21. Rosnitskiy P.B., Vysokanov B.A., Gavrilov L.R., Sapozhnikov O.A., Khokhlova V.A. (2018), Method for designing multielement fully populated random phased arrays for ultrasound surgery applications, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **65**(4): 630–637, doi: 10.1109/TUFFC.2018.2800160.

22. Shehata Elhelf I.A., Albahar H., Shah U., Oto A., Cressman E., Almekkawy M. (2018), High intensity focused ultrasound: The fundamentals, clinical applications and research trends, *Diagnostic and Interventional Imaging*, **99**(6): 349–359, doi: 10.1016/j.diii.2018.03.001.

23. Umchid S. *et al.* (2009), Development of calibration techniques for ultrasonic hydrophone probes in the frequency range from 1 to 100 MHz, *Ultrasonics*, **49**(3): 306–311, doi: 10.1016/j.ultras.2008.09.011.

24. Wang J., Sun S., Ning Y., Gong Y., Zhang M., Pang W. (2021), A spiral Archimedean PMUT array with improved focusing performance, [in:] *2021 IEEE International Ultrasonics Symposium (IUS)*, pp. 1–3, doi: 10.1109/IUS52206.2021.9593339.

Archives of Acoustics Vol. 49, No. 1, pp. 83–93 (2024), doi: 10.24425/aoa.2023.146811

Research Paper

Normal Mode Solutions of Target Strengths of Solid-filled Spherical Shells and Discussion of Influence Parameters

Bing JIA[(1),(2)], Jun FAN[(1)*], Gui-Juan LI[(2)], Bin WANG[(1)], Yun-Fei CHEN[(2)]

[(1)] *Key Laboratory of Marine Intelligent Equipment and System Ministry of Education*
Shanghai Jiao Tong University
Shanghai, China

[(2)] *Science and Technology on Underwater Test and Control Laboratory*
Dalian, Liaoning, China

*Corresponding Author e-mail: fanjun@sjtu.edu.cn

(*received December 16, 2022; accepted April 5, 2023; published online October 2, 2023*)

The normal mode solution for the form function and target strength (TS) of a solid-filled spherical shell is derived. The calculation results of the spherical shell's acoustic TS are in good agreement with the results of the finite element method (FEM). Based on these normal mode solutions, the influences of parameters such as the material, radius, and thickness of the inner and outer shells on the TS of a solid-filled spherical shell are analyzed. An underwater spherical shell scatterer is designed, which uses room temperature vulcanized (RTV) silicone rubber as a solid filling material and does not contain a suspension structure inside. The scatterer has a good TS enhancement effect.

Keywords: solid-filled spherical shell; room temperature vulcanized silicone rubber; target strength enhancement.

1. Introduction

In underwater acoustic engineering, spherical acoustic scatterers have good directivity and are often used as standard test objects for the calibration of underwater active acoustic testing systems (Anstee, 2002; Foote *et al.*, 2007; Stanton, Chu, 2008; Islas-Cital, Atkins, 2012; Atkins *et al.*, 2017). The acoustic target strength (TS) represents the reflection ability of underwater targets with respect to acoustic waves. When the TS is larger, the target can be detected more easily by active sonar systems. To obtain a larger TS, spherical acoustic scatterers with geometric scattering as the primary mechanism are usually large in volume and mass, resulting in challenges for offshore deployment. Therefore, the current research focuses on the design of TS enhancement of spherical scatterers with constant volume, including the design of medium frequency thickness resonance (Xu *et al.*, 2020) and acoustic focusing. In recent years, the calibration of liquid-filled focusing spheres as standard scatterers has also been applied in some scenarios (Deveau, Lyons, 2009; Foote, 2018). Research on focusing spheres, which are thin spherical shells filled with liquids, has been carried out on the focusing effects of different filling liquids (Jia *et al.*, 2020). However, since the material CCL4 (carbon tetrachloride) commonly used in liquid-filled spheres is toxic and the focusing effect is affected by external water temperature (Deng *et al.*, 1982), studying solid-filled spherical shells with the TS enhancement effect becomes crucial and meaningful. Similar to the focusing effect of liquid, solid fillers that meet certain sound velocity and density conditions should also have a focusing effect to enhance TS, which needs further research.

In the construction process of underwater spherical acoustic scatterers, to meet certain TS design requirements, it is necessary to establish a calculation method and discuss the influence of the spherical shell diameter, thickness, material, and acoustic characteristics of filling materials on the TS. In the earliest research on spherical sound scatterers, a common assumption was that the scatterer was a rigid material, and an approximate solution was proposed in which the wavelength was larger than the spherical diameter scattering sound field. Then, the normal series solution of a thin shell filled with air was studied (Junger, 1952; Goodman, Stern, 1962). With the demand for spherical targets with small volume and large TS, researchers have studied liquid-filled spherical shells with a certain focusing effect, obtained the normal series solution of spherical acoustic scatterers placed in water, further analyzed the relationship between the parameters of the liquid as the filling material and TS, and carried out experimental verifications (Kaduchak, Loeffler, 1998; Fawcett, 2001). In addition, researchers have theoretically derived the analytical solution of the double-layer elastic spherical shell in water (Tang, Fan, 1999), and researched the echo characteristics of the double-layer elastic spherical shell in the case of equivalent laying layers (Fan, Tang, 2001; Tang *et al.*, 2018). The derivation process and condition settings provide important inspiration for the theoretical derivation of other spherical shells. In addition, the impact of an elastic sphere composed of new materials on TS was studied (Zhou *et al.*, 2019; 2020).

In this paper, the method of separating variables is used to derive the normal mode solutions of the TS of a solid-filled spherical shell. Based on the normal mode solution, the far field form function of the echo of the spherical shell is calculated. We also discuss the influence of the outer spherical shell material, thickness, acoustic characteristics of the inner solid-filled material, and radius on the TS of the spherical shell. Finally, the feasibility of the specific RTV material as the filling material is discussed, which provides theoretical support for the design of the solid-filled spherical shell.

2. Normal mode solution of the scattering sound field from a solid-filled spherical shell

The problem of sound scattering from a solid-filled spherical shell can be simplified as the solution of the sound field of sound propagating in a four-layered medium, as shown in Fig. 1a. The areas (1), (2), (3), (4) are marked. Each area represents one of the four-layered media. Area (1) is the surrounding medium of water in which the spherical shell is located. Area (2) is a spherical shell. Area (3) is the interlayer between the solid core and the spherical shell. Area (4) is a solid core and, in this study, RTV material is used as the solid filling material. Figure 1b illustrates acoustic focusing in multilayer targets, the scattered wave consists of an elastic wave and a focused wave, and the incident plane waves converge on the spherical shell and reflect.

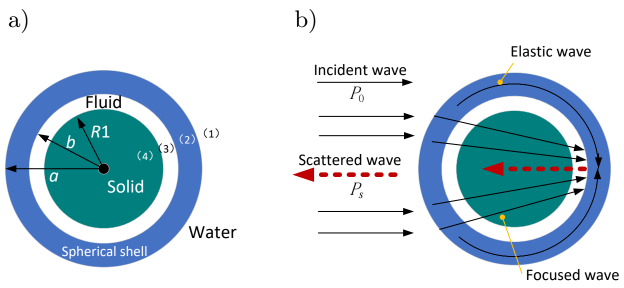

Fig. 1. Structure and focusing phenomenology of a solid-filled spherical shell: a) structure of a solid-filled spherical shell; b) acoustic focusing in multilayer targets.

In underwater acoustics, form function $f(x,\theta)$ is commonly used to describe the far-field scattering characteristics of targets:

$$f(x,\theta) = \frac{2r}{a}\frac{P_s(x,\theta)}{P_0(x)}e^{-ikr}, \tag{1}$$

where P_0 is the incident pressure, P_s is the scattered pressure, r is the distance, a is the size of the scatterer, $x = ka$ – is the dimensionless frequency, k is the wave number, and θ is the scattering angle.

The acoustic TS represents the backscattering ability of the target, and its relationship to the form function:

$$\mathrm{TS} = 10\log\left|\frac{a}{2}f(x,\pi)\right|^2, \tag{2}$$

where TS is the target strength.

When a steady plane wave is incident, the sound pressure can be expressed as:

$$P_0 = p_0\sum_{n=0}^{\infty} i^n(2n+1)j_n(k_1 r)P_n(\cos\theta). \tag{3}$$

In the formula, the amplitude of the incident sound pressure p_0 is 1. We omit the time factor $e^{-j\omega t}$ both here and below.

The outer layer is water, the range is $r > a$, and only the compressional wave propagates. The total sound pressure of the outer sound field is:

$$P_1 = p_0\sum_{n=0}^{\infty} i^n(2n+1) \cdot \left[j_n(k_1 r) + b_n h_n^{(1)}(k_1 r)\right]P_n(\cos\theta). \tag{4}$$

The scalar potential $\boldsymbol{\Phi}$ is used in fluid, and scalar potential $\boldsymbol{\Phi}$ and vector potential $\boldsymbol{\varphi}$ (considering symmetry, only the component in the P_0 direction) are used in the elastic body to express the sound field.

The second layer is a solid spherical shell, and the range is $b < r < a$, including the compressional wave and shear wave.

The compressional wave field in a spherical shell is expressed as follows:

$$\boldsymbol{\Phi}_2 = p_0\sum_{n=0}^{\infty} i^n(2n+1) \cdot \left[c_n j_n(k_{d2} r) + d_n y_n(k_{d2} r)\right]P_n(\cos\theta). \tag{5}$$

Similarly, the shear wave field in a spherical shell is expressed as:

$$\boldsymbol{\varphi}_2 = p_0 \sum_{n=0}^{\infty} i^n (2n+1) \cdot [e_n j_n(k_{s2}r) + f_n y_n(k_{s2}r)] P_n(\cos\theta). \quad (6)$$

The third layer (middle layer) is the fluid layer, with the range of $R1 < r < b$ and only compressional waves:

$$p_3 = p_0 \sum_{n=0}^{\infty} i^n (2n+1) \cdot [g_n j_n(k_3 r) + q_n y_n(k_3 r)] P_n(\cos\theta). \quad (7)$$

The fourth layer is a solid sphere with the range $r < R1$, including the compressional wave and shear wave.

The compressional wave field in the fourth layer of the solid sphere is as follows:

$$\boldsymbol{\Phi}_4 = p_0 \sum_{n=0}^{\infty} i^n (2n+1) s_n j_n(k_{d4}r) P_n(\cos\theta). \quad (8)$$

The shear wave field in the fourth layer of the solid sphere is:

$$\boldsymbol{\varphi}_4 = p_0 \sum_{n=0}^{\infty} i^n (2n+1) u_n j_n(k_{s4}r) P_n(\cos\theta). \quad (9)$$

In Eqs. (4)–(9), n is the order, $j_n(X)$ is the n-th-order Bessel function, $y_n(X)$ is the n-th-order Neumann function, $h_n^{(1)}(X)$ is the n-th-order Hankel function of the first kind, and $P_n(X)$ is the n-th-order associated Legendre function.

Mentioned k_1 is the wavenumber of the outer water, k_{d2} is the compressional wavenumber of the second layer, and k_{s2} is the shear wavenumber of the second layer, k_3 is the wavenumber of the fluid medium in the middle layer, k_{d4} is the compressional wavenumber of the fourth layer's solid sphere, and k_{s4} is the shear wavenumber of this solid sphere.

In addition, b_n, c_n, d_n, e_n, f_n, g_n, q_n, s_n, and u_n are the undetermined coefficients.

There are nine boundary conditions:

$$T_{rr}^{(2)}|_{r=a} = -p_1, \quad u_r^{(1)}|_{r=a} = u_r^{(2)}|_{r=a}, \quad T_{r\theta}^{(2)}|_{r=a} = 0,$$

$$T_{rr}^{(2)}|_{r=b} = -p_3, \quad u_r^{(2)}|_{r=b} = u_r^{(3)}|_{r=b}, \quad T_{r\theta}^{(2)}|_{r=b} = 0,$$

$$T_{rr}^{(4)}|_{r=R1} = -p_3, \quad u_r^{(3)}|_{r=R1} = u_r^{(4)}|_{r=R1}, \quad T_{r\theta}^{(4)}|_{r=R1} = 0,$$

where $T_{rr}^{(2)}$ is the normal stress of the (2) layer boundary, $u_r^{(1)}$ is the displacement of the (1) layer boundary, and $T_{r\theta}^{(2)}$ is the tangential stress of the (2) layer boundary.

The undetermined coefficient b_n can be solved according to the following formal solution and boundary conditions:

$$b_n = -B_n/D_n, \quad (10)$$

where B_n and D_n are presented in Appendix.

The expression of the far-field form function of the backscattering of a solid-filled spherical shell is:

$$|f(k_1 a, \pi)| = \frac{2}{k_1 a}\left|\sum_{n=0}^{\infty}(-1)^n(2n+1)b_n\right|. \quad (11)$$

Accordingly, the expression of the spherical shell TS is:

$$\mathrm{TS} = 10\log\left|\frac{a}{2} f(k_1 a, \pi)\right|^2. \quad (12)$$

3. Comparison between the normal mode and numerical solutions

To verify the normal mode solution derived above, the calculation results, using COMSOL Multiphysics version 5.5 – a common numerical calculation tool, are compared with the results of the normal mode solution. The numerical calculation tool uses the finite element method (FEM). Mesh size is selected as 1/6 of the wavelength corresponding to the highest frequency. The calculation conditions are as follows: the outer layer of the spherical shell is water, the spherical shell is steel with a radius of 1 m and a thickness of 0.1 m, the middle fluid layer is water, and the inner solid filler is a steel sphere with a radius of 0.8 m. See Table 1 for the specific parameter selection.

Table 1. Parameters of materials used in the paper.

Material	Density [kg/m^3]	Compressional wave velocity [m/s]	Shear wave velocity [m/s]
Water	1000	1482	–
Air	1.02	344	–
Aluminum	2700	6420	3040
Steel	7900	5940	3100
RTV (SDL-1-41)	1031	1030	202

Equation (11) involves the calculation of infinite series; according to (Tang *et al.*, 2018), the accuracy of the above infinite series can be guaranteed as long as the highest number of terms $n > ka+5$ is taken; ka is the non-dimensional frequency, k is the wavenumber, and a is the radius. In the calculation process, speed and accuracy are comprehensively considered. The order n is $ka + 20$, where ka is the non-dimensional frequency and the product of wavenumber k and target apparent scale a. Figure 2 shows the comparison results between the normal mode solution and the numerical solution. It can be seen that in the range of $ka < 30$, the normal mode solution and the finite element numerical solution are in good agreement overall. When ka is small

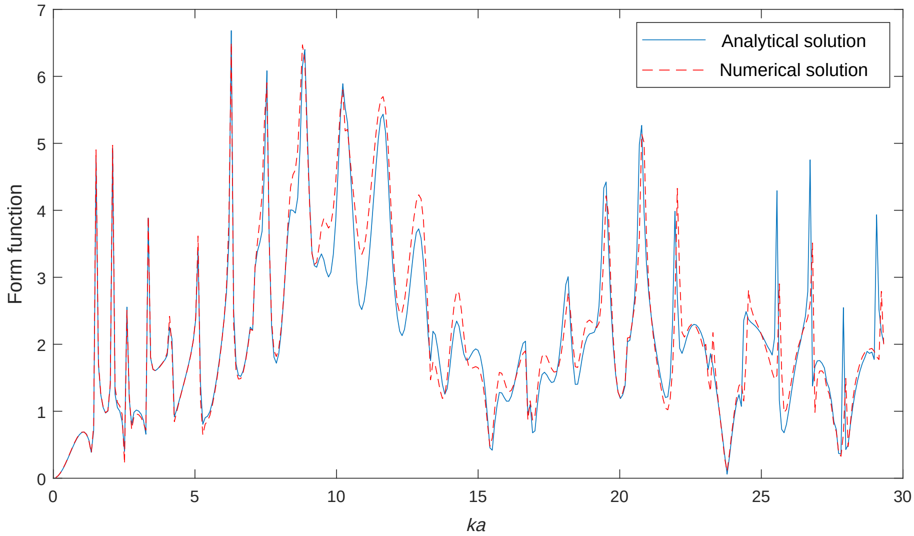

Fig. 2. Comparison of the normal mode solution and the numerical solution.

($ka < 10$ with the same size target and low frequency), the two solutions are completely consistent. In the case of large ka ($ka > 10$ with the same size target and high frequency), the peak position and trend of the morphological function calculated by the two methods are consistent. However, there are differences in the individual peak amplitudes, mainly because the grid size of the finite element calculation is limited after the frequency increases. Figure 2 can prove the accuracy of the normal mode solution derived in this paper, which provides a theoretical basis for the property analysis of the TS of the solid-filled spherical shell discussed next.

4. Analysis of the influencing factors on the TS of a solid-filled spherical shell in water

According to the structural characteristics of the solid-filled spherical shell, the main factors affecting the TS include the diameter, thickness and material of the spherical shell, the material of the middle layer, and the radius and material of the fourth layer of the solid sphere. The conclusion, "the larger the radius of the spherical shell, the stronger the TS", has been reached in classical underwater acoustics books and will not be analyzed here. In the following, we first discuss the material and thickness of the spherical shell, and then we select the material of the middle layer. Finally, we analyze and select the solid-filled material and the radius of the fourth layer solid sphere, and discuss the impact on the TS.

In the comparative analysis, the TS of a rigid sphere with the same radius is taken as the standard, and the calculation formula for the TS of a rigid sphere is:

$$\mathrm{TS} = 10 \log \frac{a^2}{4}, \tag{13}$$

where a is the radius of this sphere.

The applicable condition of Eq. (13) is $ka \gg 1$. In general, if the geometric optical zone of the rigid sphere is satisfied ($ka > 2\pi$), Eq. (13) can be used to estimate its TS.

4.1. Selection and analysis of the material and thickness of the spherical shell

In underwater acoustics, when spherical acoustic scatterers are used as underwater targets, common materials with reasonable costs include aluminum and steel. In addition, spherical acoustic scatterers made from the same material and diameter but with different shell thicknesses will also affect the TS of the spherical acoustic scatterers. This paper simulates the TS of an aluminum spherical shell and a steel spherical shell and obtains the comparison results (see Figs. 3–6). In our calculations, the shell material is aluminum or steel, the radius of the shell is 1 m, the ratio of the shell thick-

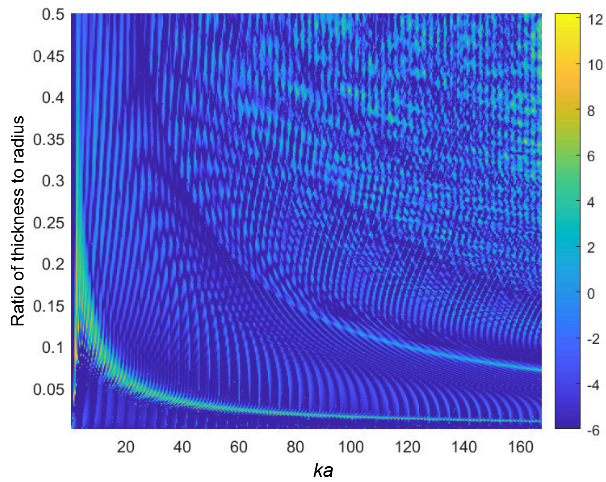

Fig. 3. TS of the aluminum spherical shell containing air varying with the thickness to radius ratio.

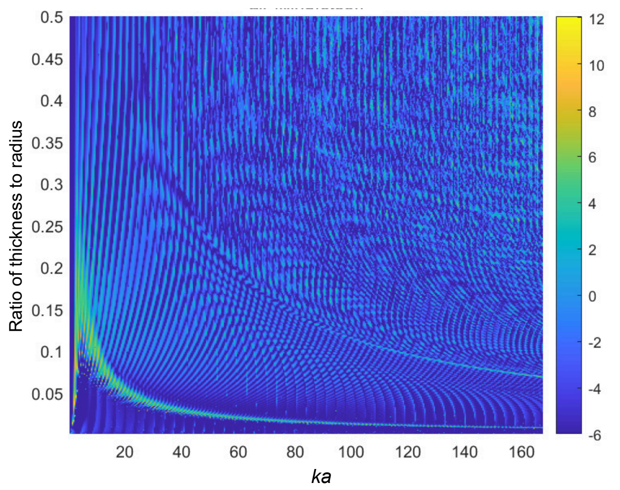

Fig. 4. TS of the steel spherical shell containing air varying with the thickness to radius ratio.

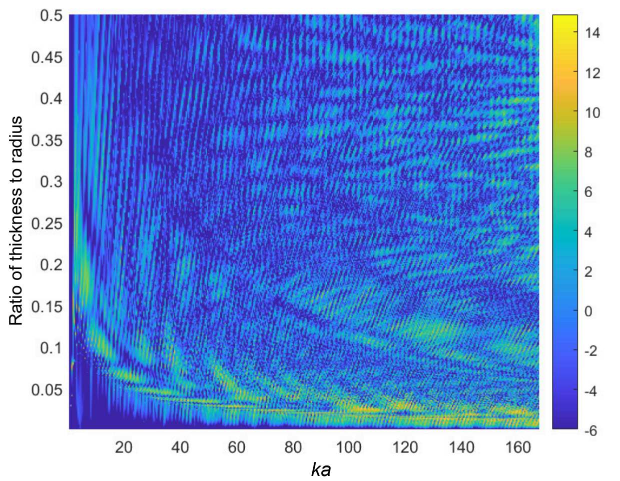

Fig. 5. TS of the aluminum spherical shell containing water varying with the thickness to radius ratio.

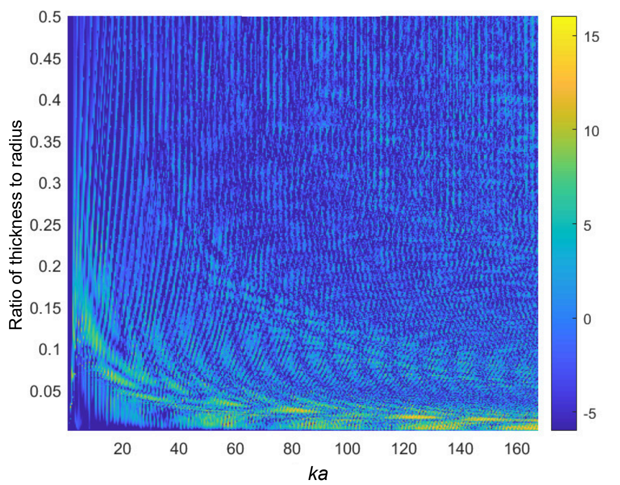

Fig. 6. TS of the steel spherical shell containing water varying with the thickness to radius ratio

ness to the radius is 0.1–50%, the step size is 0.1%, and the inner layer is air or water. Other material properties are shown in Table 1.

Figures 3 to 6 illustrate the four conditions of an air-filled aluminum spherical shell, an air-filled steel spherical shell, a water-filled aluminum spherical shell, and a water-filled steel spherical shell, respectively. Due to the existence of resonance under these conditions and with the gradual increase in the thickness to radius ratio, the maximum value of the TS moves to the small ka, that is, to the low frequency. However, when the thickness reaches a certain level (in this calculation, the ratio of thickness to radius of the aluminum shell is greater than 0.3, and the ratio of thickness to radius of the steel shell is greater than 0.25), the peak value of the TS disappears. It can be seen from the comparison between Fig. 4 and Fig. 6 that the main peak value of the air-filled structure is more obvious than that of the water-filled structure. Because the acoustic parameters of the air in the sphere and the water outside the spherical shell differ greatly, the sound field distribution is simpler than that of the water-filled structure. It can be seen from the TS color scale that the maximum TS of the water-filled structure is larger than that of the air-filled structure. The TS of the water-filled steel spherical shell is slightly higher than that of the water-filled aluminum spherical shell. Both steel and aluminum can be used as solid-filled spherical shells. In engineering design and processing, the shell material can be selected by comprehensively considering the processing difficulty, project cost, placement, and storage conditions.

To analyze the influence of the thickness of the solid spherical shell on the TS, the spherical shell is filled entirely with air, and the typical TS values of the spherical shell with different thickness to radius ratios are selected for comparison. The calculation results are shown in Fig. 7. The calculation conditions are as follows: the radius of the spherical shell is 1 m, the material is steel or aluminum, the spherical shell is filled with air, and the material outside the spherical shell is water. Other parameters are listed in Table 1.

When the ratio of shell thickness to radius is 1% (Figs. 7a and 7b) and ka is between 119 and 138, the steel spherical shell will exhibit resonance, and the TS can be increased by more than 6 dB. If ka is between 150 and 170, the aluminum spherical shell will resonate. With the increase in the thickness to radius ratio, the enhancement effect gradually moves to low frequency. When the ratio of thickness to radius is 2% (Figs. 7c and 7d) and ka is between 55 and 67, the steel spherical shell will resonate, and when ka is between 63 and 81, the aluminum spherical shell will resonate. When the thickness to radius ratio is 10% (Figs. 7e and 7f) and ka is between 5 and 11, the steel spherical shell will resonate, and if ka is between 6 and 14, the aluminum spherical shell will resonate.

By comparing the TS of the steel spherical shell and aluminum spherical shell with the ka change curve, it can be seen that the values of ka corresponding to the

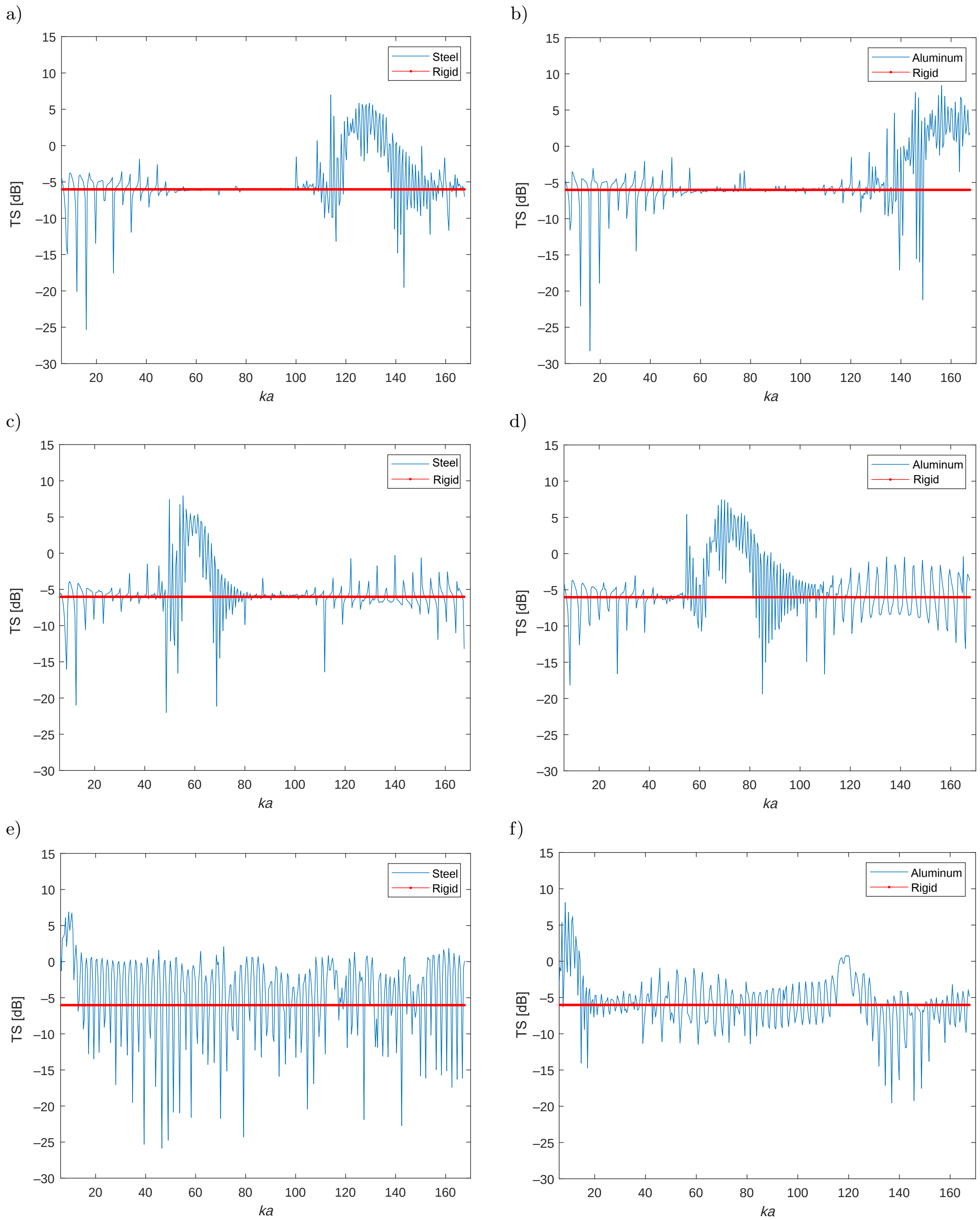

Fig. 7. Calculations of TS of spherical shells with different thickness to radius ratios: a) steel shell and the ratio is 1%; b) aluminum shell and the ratio is 1%; c) steel shell and the ratio is 2%; d) aluminum shell and the ratio is 2%; e) steel shell and the ratio is 10%; f) aluminum shell and the ratio is 10%.

middle-frequency enhancement of different materials are different under the same thickness of the spherical shell. When the steel spherical shell is strengthened, its *ka* value is slightly smaller than that of the aluminum spherical shell. Therefore, when designing a solid-filled spherical shell with specific requirements for frequency applicability, to reduce the volume of the spherical shell, steel materials can be used for low-frequency spherical shells, and a spherical shell suitable for high-frequency can be made of aluminum.

4.2. Analysis and selection of solid materials filled in the spherical shell

The fourth layer of the solid-filled spherical shell is a solid sphere. The acoustic properties of the solid sphere's materials have a strong influence on the TS of the overall spherical shell. Therefore, it is necessary to discuss the acoustic scattering capabilities of solid spheres made of different materials. We note that research on focusing spheres filled with liquid has been relatively mature (Jia *et al.*, 2020). One of the important conditions for selecting the fourth layer's solid sphere material is the slow propagation speed of sound waves in the solid material. In addition, to conveniently place solid-filled spherical shells in water, the density of the inner spherical solid filler should be as close as possible to the density of the water. According to the relevant literature, an RTV rubber material meets the above requirements (Niu, Zhang, 1982). The specific parameters are as follows: SDL-1-41 material, density 1031 kg/m^3, longitudinal wave velocity 1030 m/s, and transverse wave velocity 201.98 m/s. Of course, other alternative materials can also be used if the sound speed and density conditions are met.

In summary, the inner solid sphere materials discussed in this paper are steel, aluminum, and RTV materials with low sound velocity characteristics. The results are compared with the rigid sphere's TS of the same radius. The material parameter settings are shown in Table 1. The other calculation conditions are as follows: the radius of the spherical shell is 1 m and the medium outside the spherical shell is water.

Figure 8 shows that the TS of the steel sphere and aluminum sphere fluctuates near the theoretical value of the rigid sphere, while the TS of the RTV sphere

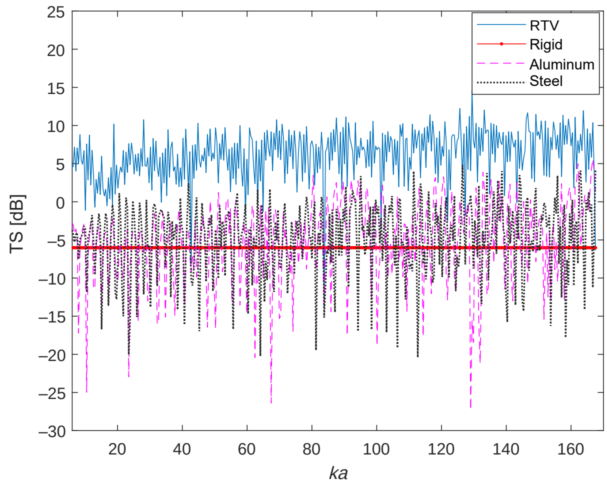

Fig. 8. TS of solid spheres with different materials.

is significantly higher than that of the remaining three spheres. Compared with the theoretical value of the rigid sphere, the TS of the RTV solid sphere is 6–16 dB higher, and the acoustic enhancement effect is obvious. Therefore, during the design of the spherical shell filled with a solid, if we want to improve its TS, the inner solid sphere can be made of RTV material.

4.3. Selection and analysis of the materials of the middle layer

The coupling (middle) layer is a layer of fluid (liquid or air) positioned between the inner solid sphere and the outer spherical shell. The selection of the coupling layer material and the design of its thickness will have a certain impact on the TS of the solid-filled spherical shell. In this paper, air or water is used as the coupling layer material, and the influence of the coupling layer thickness on the TS is discussed.

Calculation condition 1: the spherical shell is made of steel with a radius of 1 m, the thickness of the spherical shell is 1% of the radius (thin spherical shell), the solid sphere material is RTV, the middle coupling layer material is air or water, the radius of the inner solid sphere is 0.5–0.99 m, the corresponding coupling layer thickness is 0.49–0 m, and the other material properties are listed in Table 1.

Figure 9a shows that when the material of the middle coupling layer of the thin spherical shell is water, the thickness of the coupling layer gradually decreases with the increase in the radius of the inner solid sphere between 0.5 m and 0.99 m, and the TS of the solid-filled spherical shell fluctuates but increases overall. When the radius of the solid sphere is 0.99 m, the thickness of the coupling layer is close to 0, that is, the inner solid sphere is close to the spherical shell. When the value of ka is large, the TS has a maximum value of approximately 20 dB, which is 26 dB higher than the TS (−6 dB) of a rigid sphere with the same radius.

Figure 9b reveals that when the medium of the middle coupling layer of a thin spherical shell is air, as the acoustic impedance of air and water has a large difference and acoustic waves are approximately totally reflected, the TS, in this case, is similar to that of a rigid sphere (−6 dB under calculation conditions). However, with increasing frequency, the thin spherical shell in calculation condition 1 gradually gains the characteristics of a thick spherical shell. Therefore, at high frequency, its TS is higher than that of a rigid sphere with the same radius.

Calculation condition 2: the spherical shell is made of steel with a radius of 1 m, the thickness of the spherical shell is 10% of the radius (so we have a thick spherical shell), the inner solid sphere material is RTV, the middle coupling layer material is air or water, the inner solid sphere's radius is 0.5–0.9 m, and the cor-

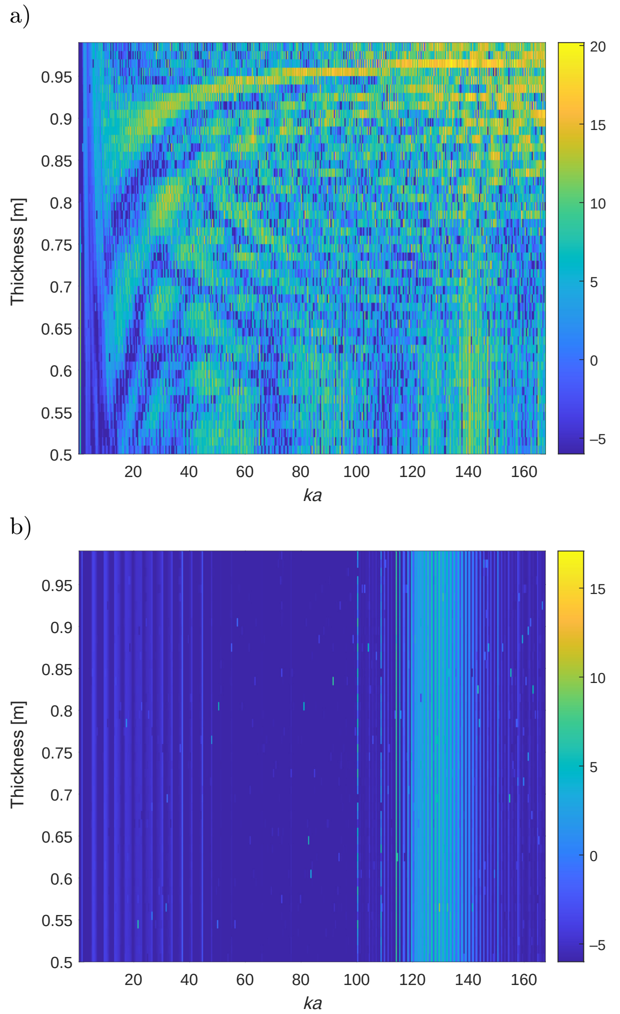

Fig. 9. Calculation of TS of a thin shell sphere changing with coupling layer thickness: a) water filling; b) air filling.

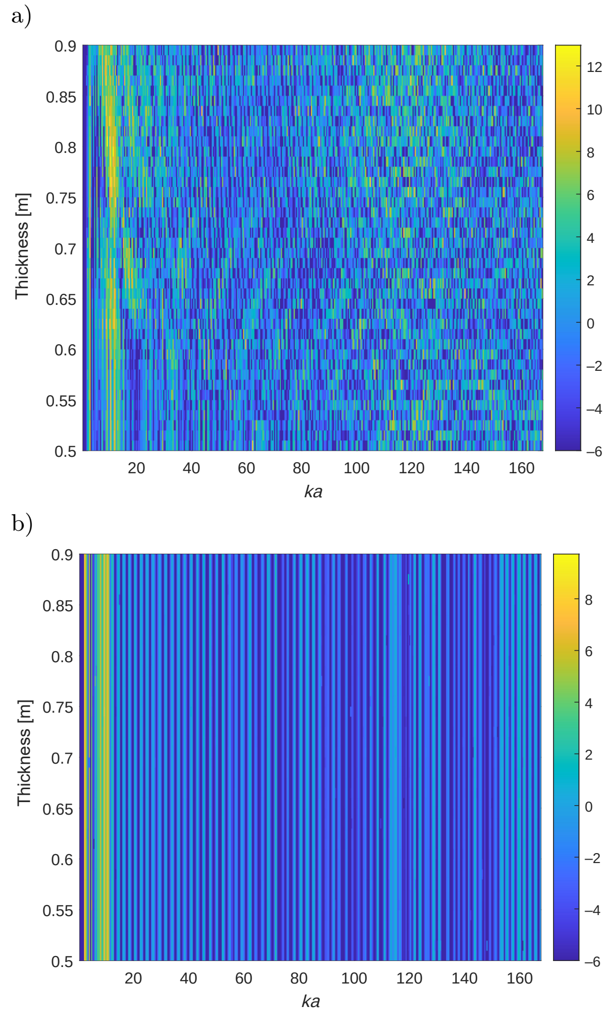

Fig. 10. Calculation of TS of a thick shell sphere changing with coupling layer thickness: a) water filling; b) air filling.

responding middle coupling layer thickness is 0.4–0 m. Other material properties are given in Table 1.

Figure 10a shows that when the medium of the middle coupling layer of the thick spherical shell is water, there is an obvious peak near the low-frequency band ($ka = 10$) due to the thickness of the outer shell, and the TS is greater than 12 dB. Due to the influence of the middle coupling layer and inner solid sphere, the TS of the solid-filled spherical shell has strong and weak periodic variations.

Figure 10b shows that when the medium of the middle coupling layer of the thick spherical shell is air, similar to Fig. 10a, there is a stable and obvious peak near the low-frequency band, and the TS is approximately 10 dB. Due to the poor coupling effect of air, the influence of solid-filled spheres in the inner layer is almost not reflected, resulting in the same change in the TS with ka for all thicknesses, as shown in the figure.

To better compare the influence of different materials as coupling layer materials on the TS of solid-filled spherical shells, the TS calculation results of these shells with different coupling layer materials under the same conditions are selected for comparison. In the calculation, the material of the spherical shell is steel, the radius is 1 m, the shell thickness is 0.05 m, the inner solid sphere's material is RTV, the middle coupling layer material is air or water, the inner solid sphere radius is 0.9 m, and other material properties are given in Table 1.

In summary, the fourth layer solid sphere plays a major role in enhancing the TS. The strengthening effect of the outer-thickness spherical shell is weaker than that of the inner-filled sphere. The TS strengthening effect of the outer-layer-thickness spherical shell has a certain frequency selectivity. When the middle layer is water, the coupling effect is good (Fig. 11), enabling the full utilization of the TS enhancement effect of the inner solid sphere. When the middle layer is gas, the coupling effect is poor, and the enhancement effect of the inner solid sphere cannot be realized. Therefore, when a high TS is needed, the middle coupling layer uses water. Of course, other liquids can also be used. However, the use of water as a coupling material

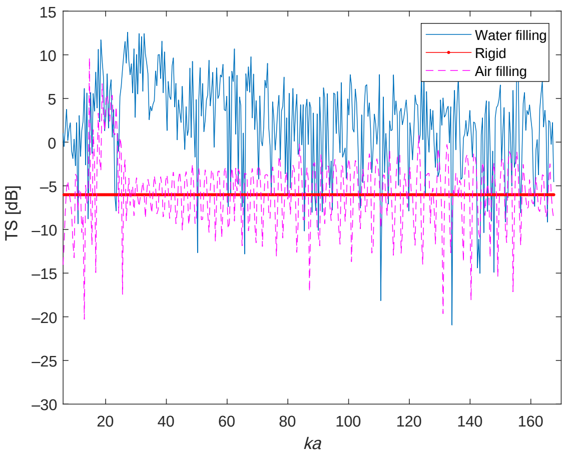

Fig. 11. Comparison of TS of solid-filled spheres with different coupling materials.

has another advantage: the shell structure can be designed to allow water penetration to reduce the weight of solid-filled spherical shells resulting from their closed design, and this, in turn, facilitates the implementation of offshore tests.

5. Conclusions

The normal mode solution of the scattering sound field from a solid-filled spherical shell derived in this paper is in good agreement with the finite element numerical solution, which proves that the solution can be used to calculate the sound scattering characteristics of solid-filled spherical shells.

The sound scatterer of the spherical shell-water-RTV structure has the underwater acoustic focusing ability, which can improve the TS.

Due to the frequency-selective enhancement characteristics of spherical shells with different thicknesses, when designing underwater acoustic standards and markers, the thicknesses of spherical shells can be designed to match the working frequency band of the available underwater acoustic equipment.

Appendix

Solution of the scattering coefficient b_n of the solid-filled spherical shell.

$$b_n = -\frac{B_n}{D_n},$$

$$B_n = \begin{vmatrix} A_1 & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} & d_{18} & d_{19} \\ A_2 & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} & d_{28} & d_{29} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} & d_{38} & d_{39} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} & d_{48} & d_{49} \\ d_{51} & d_{52} & d_{53} & d_{54} & d_{55} & d_{56} & d_{57} & d_{58} & d_{59} \\ d_{61} & d_{62} & d_{63} & d_{64} & d_{65} & d_{66} & d_{67} & d_{68} & d_{69} \\ d_{71} & d_{72} & d_{73} & d_{74} & d_{75} & d_{76} & d_{77} & d_{78} & d_{79} \\ d_{81} & d_{82} & d_{83} & d_{84} & d_{85} & d_{86} & d_{87} & d_{88} & d_{89} \\ d_{91} & d_{92} & d_{93} & d_{94} & d_{95} & d_{96} & d_{97} & d_{98} & d_{99} \end{vmatrix},$$

$$D_n = \begin{vmatrix} d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & d_{16} & d_{17} & d_{18} & d_{19} \\ d_{21} & d_{22} & d_{23} & d_{24} & d_{25} & d_{26} & d_{27} & d_{28} & d_{29} \\ d_{31} & d_{32} & d_{33} & d_{34} & d_{35} & d_{36} & d_{37} & d_{38} & d_{39} \\ d_{41} & d_{42} & d_{43} & d_{44} & d_{45} & d_{46} & d_{47} & d_{48} & d_{49} \\ d_{51} & d_{52} & d_{53} & d_{54} & d_{55} & d_{56} & d_{57} & d_{58} & d_{59} \\ d_{61} & d_{62} & d_{63} & d_{64} & d_{65} & d_{66} & d_{67} & d_{68} & d_{69} \\ d_{71} & d_{72} & d_{73} & d_{74} & d_{75} & d_{76} & d_{77} & d_{78} & d_{79} \\ d_{81} & d_{82} & d_{83} & d_{84} & d_{85} & d_{86} & d_{87} & d_{88} & d_{89} \\ d_{91} & d_{92} & d_{93} & d_{94} & d_{95} & d_{96} & d_{97} & d_{98} & d_{99} \end{vmatrix},$$

where

$$d_{11} = (\rho_1/\rho_2)k_{s2}^2a^2h_n^{(1)}(k_1a),$$

$$d_{12} = \left[2n\left(n+1\right) - k_{s2}^2a^2\right]j_n\left(k_{d2}a\right) - 4k_{d2}aj_n'(k_{d2}a),$$

$$d_{13} = \left[2n\left(n+1\right) - k_{s2}^2a^2\right]y_n\left(k_{d2}a\right) - 4k_{d2}ay_n'(k_{d2}a),$$

$$d_{14} = 2n(n+1)\left[k_{s2}aj_n'\left(k_{s2}a\right) - j_n(k_{s2}a)\right],$$

$$d_{15} = 2n(n+1)\left[k_{s2}ay_n'\left(k_{s2}a\right) - y_n(k_{s2}a)\right],$$

$$d_{16} = 0,$$

$$d_{17} = 0,$$

$$d_{18} = 0,$$

$$d_{19} = 0,$$

$$d_{21} = -k_1ah_n^{(1)\prime}(k_1a),$$

$$d_{22} = k_{d2}aj_n'\left(k_{d2}a\right),$$

$$d_{23} = k_{d2}ay_n'\left(k_{d2}a\right),$$

$$d_{24} = n(n+1)j_n\left(k_{s2}a\right),$$

$$d_{25} = n(n+1)y_n\left(k_{s2}a\right),$$

$$d_{26} = 0,$$

$$d_{27} = 0,$$

$$d_{28} = 0,$$

$$d_{29} = 0,$$

$$d_{31} = 0,$$

$$d_{32} = 2\left[j_n(k_{d2}a) - k_{d2}aj_n'(k_{d2}a)\right],$$

$$d_{33} = 2\left[y_n(k_{d2}a) - k_{d2}ay_n'(k_{d2}a)\right],$$

$$d_{34} = 2k_{s2}aj_n'(k_{s2}a)+\left[(k_{s2}a)^2 - 2n(n+1)+2\right]j_n(k_{s2}a),$$

$$d_{35} = 2k_{s2}ay_n'(k_{s2}a)+\left[(k_{s2}a)^2 - 2n(n+1)+2\right]y_n(k_{s2}a),$$

$$d_{36} = 0,$$

$$d_{37} = 0,$$

$$d_{38} = 0,$$

$d_{39} = 0,$

$d_{41} = 0,$

$d_{42} = \left[2n(n+1) - (k_{s2}b)^2\right] j_n(k_{d2}b) - 4k_{d2}bj_n'(k_{d2}b),$

$d_{43} = \left[2n(n+1) - (k_{s2}b)^2\right] y_n(k_{d2}b) - 4k_{d2}by_n'(k_{d2}b),$

$d_{44} = 2n(n+1)\left[k_{s2}bj_n'(k_{s2}b) - j_n(k_{s2}b)\right],$

$d_{45} = 2n(n+1)\left[k_{s2}by_n'(k_{s2}b) - y_n(k_{s2}b)\right],$

$d_{46} = (\rho_3/\rho_2)(k_{s2}b)^2 j_n(k_3b),$

$d_{47} = (\rho_3/\rho_2)(k_{s2}b)^2 y_n(k_3b),$

$d_{48} = 0,$

$d_{49} = 0,$

$d_{51} = 0,$

$d_{52} = k_{d2}bj_n'(k_{d2}b),$

$d_{53} = k_{d2}by_n'(k_{d2}b),$

$d_{54} = n(n+1)j_n(k_{s2}b),$

$d_{55} = n(n+1)y_n(k_{s2}b),$

$d_{56} = -k_3bj_n'(k_3b),$

$d_{57} = -k_3by_n'(k_3b),$

$d_{58} = 0,$

$d_{59} = 0,$

$d_{61} = 0,$

$d_{62} = 2\left[j_n(k_{d2}b) - k_{d2}bj_n'(k_{d2}b)\right],$

$d_{63} = 2\left[y_n(k_{d2}b) - k_{d2}by_n'(k_{d2}b)\right],$

$d_{64} = 2k_{s2}bj_n'(k_{s2}b) + \left[(k_{s2}b)^2 - 2n(n+1)+2\right] j_n(k_{s2}b),$

$d_{65} = 2k_{s2}by_n'(k_{s2}b) + \left[(k_{s2}b)^2 - 2n(n+1)+2\right] y_n(k_{s2}b),$

$d_{66} = 0,$

$d_{67} = 0,$

$d_{68} = 0,$

$d_{69} = 0,$

$d_{71} = 0,$

$d_{72} = 0,$

$d_{73} = 0,$

$d_{74} = 0,$

$d_{75} = 0,$

$d_{76} = (\rho_3/\rho_4)(k_{s4}R_1)^2 j_n(k_3R_1),$

$d_{77} = (\rho_3/\rho_4)(k_{s4}R_1)^2 y_n(k_3R_1),$

$$d_{78} = \left[2n\,(n+1) - (k_{s4}R_1)^2\right] j_n\,(k_{d4}R_1) - 4k_{d4}R_1j_n'(k_{d4}R_1),$$

$d_{79} = 2n(n+1)\left[k_{s4}R_1j_n'\,(k_{s4}R_1) - j_n(k_{s4}R_1)\right],$

$d_{81} = 0,$

$d_{82} = 0,$

$d_{83} = 0,$

$d_{84} = 0,$

$d_{85} = 0,$

$d_{86} = -k_3R_1j_n'(k_3R_1),$

$d_{87} = -k_3R_1y_n'(k_3R_1),$

$d_{88} = k_{d4}R_1j_n'(k_{d4}R_1),$

$d_{89} = n(n+1)j_n(k_{s4}R_1),$

$d_{91} = 0,$

$d_{92} = 0,$

$d_{93} = 0,$

$d_{94} = 0,$

$d_{95} = 0,$

$d_{96} = 0,$

$d_{97} = 0,$

$d_{98} = 2\left[j_n(k_{d4}R_1) - k_{d4}R_1j_n'(k_{d4}R_1)\right],$

$$d_{99} = 2k_{s4}R_1j_n'(k_{s4}R_1) + \left[(k_{s4}R_1)^2 - 2n(n+1) + 2\right] j_n(k_{s4}R_1),$$

$A_1 = (\rho_1/\rho_2)(k_{s2}a)^2 j_n(k_1a),$

$A_2 = -k_1aj_n'(k_1a),$

where ρ_1 is the density of (1) layer, $j_n'(X)$ is the derivative of the n-th-order Bessel function, $y_n'(X)$ is the derivative of the n-th-order Neumann function, and $h_n^{(1)'}(X)$ is the derivative of the n-th-order Hankel function of the first kind.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grant no. 61901079).

References

1. Anstee S. (2002), *Use of spherical objects as calibrated mine hunting sonar targets*, Maritime Operations Division Systems Sciences Laboratory, DSTO-TN-0425.

2. Atkins P.R., Yan T., Hayati F. (2017), Standard target calibration: practical intercomparison with planar surface targets and calibrated transducers, [in:] *Proceeding of Oceans '17 MTS/IEEE*, doi: 10.1109/OCEANSE.2017.8084801.

3. Deng W.-X., Yang T.-S., Yang J.-B. (1982), Experimental study on liquid filled focused spherical reflector [in Chinese], *Acta Acustica*, **7**(2): 88–93, doi: 10.15949/j.cnki.0371-0025.1982.02.003.

4. Deveau D.M., Lyons A.P. (2009), Fluid-filled passive sonar calibration spheres: Design, modeling, and measurement, *IEEE Journal of Oceanic Engineering*, **34**(1): 93–100, doi: 10.1109/JOE.2008.2010755.

5. Fan J., Tang W.L. (2001), Echoes from double elastic spherical shell covered with viscoelastic materials in water [in Chinese], *Acta Acustica*, **26**(4): 302–306, doi: 10.15949/j.cnki.0371-0025.2001.04.003.

6. Fawcett J.A. (2001), Scattering from a partially fluid-filled, elastic shelled sphere, *The Journal of the Acoustical Society of America*, **109**(2): 508–513, doi: 10.1121/1.1339827.

7. Foote K.G. (2018), Standard-target calibration of active sonars used to measure scattering: principles and illustrative protocols, *IEEE Journal of Oceanic Engineering*, **43**(3): 749–763, doi: 10.1109/JOE.2017.2713538.

8. Foote K.G., Francis D.T.I., Atkins P.R. (2007), Calibration sphere for low-frequency parametric sonars, *The Journal of the Acoustical Society of America*, **121**(3): 1482–1490, doi: 10.1121/1.2434244.

9. Goodman R.R., Stern R. (1962), Reflection and transmission of sound by elastic spherical shells, *Journal of the Acoustical Society of America*, **34**(3): 338–344, doi: 10.1121/1.1928120.

10. Islas-Cital A., Atkins P. (2012), Practical consideration in the amplitude and phase calibration of SONAR systems in laboratory water tanks using the standard-target method, [in:] *Proceedings of Meetings on Acoustics, ECUA 2012 11th European Conference on Underwater Acoustics*, **17**(1): 070013, doi: 10.1121/1.4767969.

11. Jia B., Wang J.H., Li G.J., Chen Y.F. (2020), Numerical analysis on target strength of the filled spherical elastic shell, *IOP Conference Series: Materials Science and Engineering*, **813**: 012004, doi: 10.1088/1757-899X/813/1/012004.

12. Junger M.C. (1952), Sound scattering by thin elastic shells, *The Journal of the Acoustical Society of America*, **24**(4): 336–373, doi: 10.1121/1.1906905.

13. Kaduchak G., Loeffler C.M. (1998), Relationship between material parameters and target strength of fluid-filled spherical shells in water: Calculations and observations, *IEEE Journal of Oceanic Engineering*, **23**(1): 26–30, doi: 10.1109/48.659447.

14. Niu F.Q., Zhang X.P. (1982), Acoustic properties of domestic RTV silicone rubber [in Chinese], *Applied Acoustics*, **1**(4): 39–42.

15. Stanton T.K., Chu D. (2008), Calibration of broadband active acoustic systems using a single standard spherical target, *The Journal of the Acoustical Society of America*, **124**(1): 128–136, doi: 10.1121/1.2917387.

16. Tang W.L., Fan J. (1999), Echoes from double elastic spherical shell in water [in Chinese], *Acta Acustica*, **24**(2): 174–182, doi: 10.15949/j.cnki.0371-0025.1999.02.009.

17. Tang W.L., Fan J., Ma Z.C. (2018), *Acoustic Scattering from Underwater Targets*, Science Press.

18. Xu S.Y., Fan J., Wang B. (2020), Low frequency target strength enhancement based on acoustic tunnel effect, *Technical Acoustics*, **39**(1): 34–39, doi: 10.16300/j.cnki.1000-3630.2020.01.006.

19. Zhou Y.L., Fan J., Wang B. (2019), Inversion for acoustic parameters of plastic polymer target in water [in Chinese], *Acta Physica Sinica*, **68**(21): 214301, doi: 10.7498/aps.68.20190991.

20. Zhou Y.L., Fan J., Wang B. (2020), Phase characteristics of acoustic-scattering field for underwater spherical targets [in Chinese], *Journal of Harbin Engineering University*, **41**(7): 945–950, doi: 10.11990/jheu.201905022.

Archives of Acoustics Vol. 49, No. 1, pp. 95–106 (2024), doi: 10.24425/aoa.2024.148767

Research Paper

The Influential Factors and Prediction of Kuroshio Extension Front on Acoustic Propagation-Tracked

Weishuai XU[(1)], Lei ZHANG[(2)*], Hua WANG[(2)]

[(1)] *No. 5 Student Team, Dalian Naval Academy*
Dalian, Liaoning, China; e-mail: xuweishuai2022@163.com

[(2)] *Department of Military Oceanography and Hydrography and Cartography*
Dalian Naval Academy
Dalian, Liaoning, China

*Corresponding Author e-mail: stone333@tom.com

(*received May 16, 2023; accepted September 14, 2023; published online December 19, 2023*)

The Kuroshio Extension front (KEF) considerably influences the underwater acoustic environment; however, a knowledge gap persists regarding the acoustic predictions under the ocean front environment. This study utilized the high-resolution ocean reanalysis data (JCOPE2M, 1993–2022) to assess the impact of the KEF on the underwater acoustic environment. Oceanographic factors were extracted from the database using the Douglas-Peucker algorithm, and acoustic propagation characteristics were obtained using the Bellhop ray-tracing model. This study employed a backpropagation neural network to predict the acoustic propagation affected by the KEF. The depth of the acoustic channel axis and the vertical gradient of the transition layer of sound speed were identified as the fundamental factors influencing the first area of convergence, with correlations between the former and the distance of the first convergence zone ranging from 0.52 to 0.82, and that for the latter ranging from −0.42 to −0.7. The proposed method demonstrated efficacy in forecasting first convergence zone distances, predicting distances with less than 3 km error in >90% of cases and less than 1 km error in 68.61% of cases. Thus, this study provides a valuable predictive tool for studying underwater acoustic propagation in ocean front environments and informs further research.

Keywords: Kuroshio Extension front; acoustic propagation; convergence zone prediction.

1. Introduction

The Kuroshio Current, a warm ocean current, flows eastward off the coast of Japan at around 35°N and 141°E and expands eastwards at ~165°E. The eastern segment of this current, termed the Kuroshio Extension (Yasuda, 2003), is characterized by a zonal jet with substantial amplitude bending, and its location is presented in Fig. 1. The warm and highly saline Kuroshio Current flowing from the south merges with the cold and slightly saline Oyashio Current coming from the north in the eastern region of Japan. The convergence of these two western boundary currents forms a distinct transition zone between subtropical and subpolar gyres, designated as the Kuroshio Extension

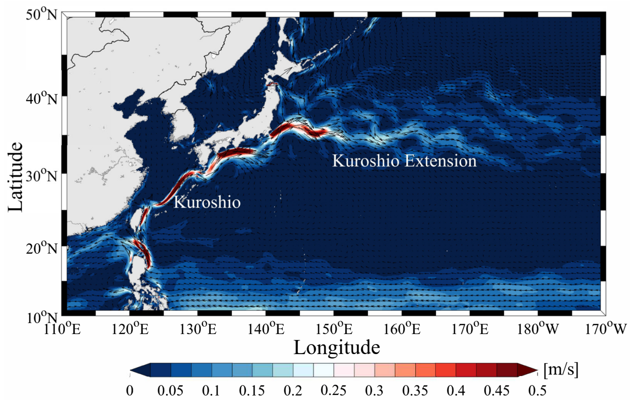

Fig. 1. Schematic diagram of the Kuroshio and Kuroshio Extension (base map: multiyear averaged flow field).

front (KEF) (Chen, 2008). As an exceptionally prominent mesoscale phenomena across global oceans, the KEF demonstrates unique physical and chemical attributes that greatly impact underwater acoustic propagation. By examining the acoustic propagation at the KEF, we can gain a deeper understanding of the influence of the marine environment on sound transmission. This study provides essential guidance for applications such as underwater acoustic communication and sonar detection. Furthermore, predicting acoustic propagation variables in the oceanic front environment contributes to technical assistance for ocean environmental monitoring and the development of marine sonar detection technology.

The spatial distribution of the sound speed field experiences rapid alterations in an ocean front with a narrow transition zone of seawater types characterized as an ocean front discontinuity (Cheney, Winfrey, 1976). Such a phenomenon considerably affects underwater acoustic propagation and associated transmission losses. As Etter (2013) initially posited, the impact of ocean fronts on acoustic propagation is exemplified through the surface sound speed and the structure of the acoustic channel axis. Prior research has comprehensively scrutinized the impacts of ocean fronts on acoustic propagation (Dreini, Jensen, 1990; Mellberg *et al.*, 1991; Rousseau *et al.*, 1982; Shapiro *et al.*, 2014), revealing that the morphology and position of ocean fronts have a significant impact on sound propagation, which results in an increase in transmission loss of 6–20 dB. Ozanich *et al.* (2022) investigated the underwater acoustic propagation of low-frequency sound waves with low grazing angles via the New England Shelf Front in spring, demonstrating the sensitivity of low-frequency propagation to the geometric configurations of ocean fronts. Being one of the major ocean fronts in the northwest Pacific, the acoustic characteristics of the KEF have also received considerable attention. Liu *et al.* (2015) utilized the absolute gradient method to examine the spatial information of the KEF front axis from WOA13 data and explored the acoustic propagation characteristics within the front area, identifying a strong influence of the sound source depth on the underwater axis of the acoustic channel. Additionally, Chen *et al.* (2017) developed a sound speed characteristic model using ARGO and WOA data, which demonstrated that the KEF modifies acoustic propagation by varying the depth of the acoustic channel axis. A collaborative experiment by Liu *et al.* (2021) in marine acoustics and physical oceanography in the northwest Pacific Ocean revealed a sharp rise in acoustic transmission loss toward the side of the cold-water mass when the sound source was positioned within the front area.

Artificial neural networks (ANNs) represent a fundamental machine learning model, with the advancement of machine learning; these neural network models have been progressively refined in recent years. The adaptability of ANNs has increasingly improved. Their ability to adapt to a variety of complex underwater environments and signal conditions has resulted in the extensive implementation of ANNs; additionally, ANNs have achieved widespread applications in the field of underwater acoustics domain. Doan *et al.* (2020) employed convolutional neural networks for target recognition and classification within a passive sonar dataset, ensuring the overall accuracy of 98.85% at 0 dB signal-to-noise ratio. Lee-Leon *et al.* (2021) designed a receiver system based on a deep belief network and, through simulation modeling and sea trials, demonstrated that the receiver system exhibited superior performance in channels affected by the Doppler effect and multipath propagation. Lee *et al.* (2022) introduced a supervised learning-based method for quantifying uncertainty in transmission loss and assessed its ability to simulate long-distance underwater propagation using different computational models, environmental scenarios, and sources and levels of uncertainties.

As the understanding of the KEF advances, along with the ascent of neural networks, the prediction of underwater acoustic propagation based on KEF oceanographic variables is emerging as an area of research deserving attention. However, the current dearth of studies on the impact of the KEF on the underwater acoustic environment and the considerable gaps in the neural network-based prediction of acoustic propagation in ocean front environments necessitate further investigation. Therefore, this study employs high-resolution ocean reanalysis data to analyze the vertical characteristics of the KEF and identify the frontal zone with the most potent KEF using statistical analysis methods. The factors affecting acoustic propagation in proximity to multiple factors near the frontal zone are investigated in Sec. 3. The study examines the primary oceanic structures influencing acoustic propagation in the KEF region. Based on backpropagation (BP) neural networks, a convergence zone prediction model is constructed for the KEF environment in Sec. 4, with the aim of providing benchmarks for future acoustic propagation forecasts in marine front environments.

2. Materials and methods

2.1. Materials

2.1.1. Reanalysis data

The temperature and salinity data used in this study were derived from the high-resolution and high-analysis product JCOPE2M (Japan Coastal Ocean Predictability Experiment 2 Modified) based on the Princeton Ocean Model and was provided by the Japan

Agency for Marine-Earth Science and Technology (Miyazawa *et al.*, 2017; 2019). The regional coverage includes the northwest Pacific Ocean. The dataset used in this research spans from January 1, 1993, to August 31, 2022, with a horizontal resolution of 1/12° and 46 σ-levels in the vertical dimension. This dataset has high resolution and high accuracy and has been widely applied in the study of mesoscale phenomena and flow fields in the Kuroshio region (Chang *et al.*, 2015; 2018; Liu *et al.*, 2019).

2.1.2. Multiyear average flow field data

The Navy Coupled Ocean Data Assimilation system, integrating the hybrid coordinate ocean model and multiple observational datasets, was harnessed by the United States Naval Research Laboratory to generate a 22-year average flow field from 1994 to 2015. This dataset was used to display the position of the Kuroshio Extension. The spatial resolution of this system is 1/12° horizontally and consists of 40 nonuniformly and vertically spaced layers (Chassignet *et al.*, 2007). Based on the hybrid isopycnal-sigma-pressure (generalized) coordinate ocean assimilation model, the average flow field serves as a valuable reference for reporting the velocity and location of Kuroshio.

2.1.3. Bathymetric data

The bathymetric data used for underwater acoustic propagation simulation in this study were obtained from the joint release of the ETOPO1 bathymetric model, bearing a grid size of $1' \times 1'$. This model integrates global land topography and ocean depth data based on various relevant models and measured data (Amante, Eakins, 2009). The majority of the ocean depth data in this model are derived from the bathymetric model released by the Scripps Institution of Oceanography, while the land topography data mainly come from the GTOPO30 global digital elevation model with a resolution of 30 arc-s. The study area based on ETOPO1, as illustrated in Fig. 2, covers the region from 142° to 162°E and 33° to 37°N.

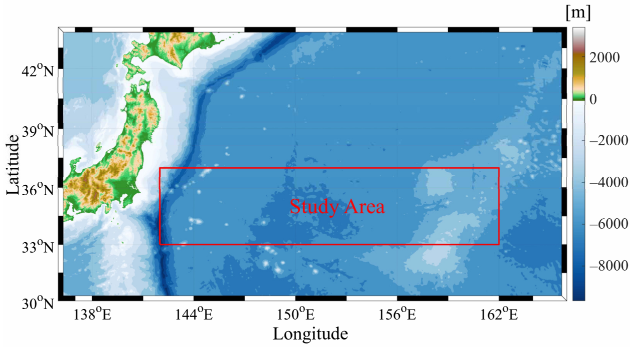

Fig. 2. Topography of the study area.

2.2. Methods

This study uses reanalyzed temperature and salinity data to explore the KEF environment. The Mackenzie empirical formula, which was introduced in 1981, was utilized to calculate the sound speed field in the KEF environment. Additionally, the absolute gradient method was employed to extract the temperature and salinity frontal information near the KE region. The variation of the KEF strength with depth is examined, and a cross-section with the highest horizontal temperature gradient (HTG) in the KEF area is selected as the focus of the study from the daily JCOPE2M dataset. Based on Etter's (2013) work on the impact of ocean fronts on acoustic propagation and fundamental propagation modes for both near and distant acoustic propagation, calculations were performed for five environmental and two acoustic propagation parameters (Table 1). A correlation heatmap was created to statistically analyze the significant impact of the environmental parameters on underwater acoustic propagation. Based on these findings, an acoustic propagation prediction model for the KEF environment was constructed using a BP neural network.

2.2.1. Calculation methods for environmental parameters

The KEF is formed by the interaction between the warm and salty Kuroshio Extension and the cold and fresh coastal currents. Therefore, this study uses the

Table 1. Factors considered in this study and their calculation methods.

	Parameters	Calculation method
Environmental parameters	Horizontal temperature gradient	Absolute gradient method
	Horizontal salinity gradient (HSG)	
	Acoustic channel axis depth (ACAD)	Douglas–Peucker (DP) algorithm
	Sonic layer depth (SLD)	
	Transition layer of sound speed (TLSS)	
Underwater acoustic propagation parameters	Short-range detection (SRD) distance	Propagation distance of underwater acoustic at the figure of merit factor level of 90 dB
	First convergence zone distance	Horizontal distance to the first minimum value of underwater acoustic transmission loss

absolute gradient method to determine the horizontal temperature and salinity gradients and the absolute gradient as a measure of the KEF strength. This method is widely used in oceanic frontal research (Dong *et al.*, 2006; Liu *et al.*, 2015; Wang *et al.*, 2020; Yu *et al.*, 2020). The equation is stated as:

$$\mathrm{Grad} = \sqrt{(\partial\phi_U/\partial x)^2 + (\partial\phi_V/\partial y)^2}, \qquad (1)$$

where $\partial\phi_U$ represents the meridional difference of the study variable (temperature and salinity), $\partial\phi_V$ represents the zonal difference of the study variable, ∂x represents the meridional distance, and ∂y represents the zonal distance.

Furthermore, this study employs the Douglas–Peucker algorithm to simplify the underwater acoustic field environment into a three-layer sound velocity structure, as proposed by Urick (1975). This includes extraction of the sonic layer depth and the transition layer of sound velocity. The transition layer of sound velocity refers to the vertical gradient within the sound speed interface. The Douglas–Peucker algorithm is a classic line simplification algorithm that can extract characteristic points in the sound velocity structure based on a fixed distance threshold. It uses the vertical distance as a simplification indicator, which ensures the shape characteristics of the sound velocity profile. Moreover, when the distance threshold is set, the extraction of characteristic points remains relatively consistent.

2.2.2. Calculation methods for acoustic propagation factors

This study considers both long-range and short-range sonar detection methods. The horizontal distance from the first convergence zone (FCZ) and the horizontal distance of a passive sonar's figure of merit (FOM) factor, which is 90 dB, are extracted as the acoustic propagation factors. The convergence zone is a major acoustic propagation mode for long-range sonar detection, and the horizontal detection distance at a 90 dB FOM factor denotes the detection distance for short-range detection. Both factors are computed using the Bellhop ray-tracing model (Porter, 2011), which is based on geometric and physical propagation laws and can accommodate various types of rays, including Gaussian beams. The parameters for the Bellhop model used herein were set according to Table 2, where the seafloor parameters were derived from the acoustic properties of sediment provided by Hamilton (1980).

Table 2. Parameter settings for the Bellhop ray-tracing model.

Sound source parameters	Source frequency		Grazing angle range
	1 kHz		−45°–45°
Seafloor parameters	Density	Compressional wave velocity	Attenuation coefficient
	1.421 g/cm^3	1520 m/s	0.12

2.2.3. Factor analysis and prediction methods

After determining the aforementioned seven factors (Table 1), we constructed a correlation heatmap for statistical analysis, representing the relations between variables. Different colors were used to encode the Pearson correlation coefficients between each pair of variables. This study aims to investigate the impact of five oceanographic environmental factors on acoustic propagation in the ocean frontal zone. Correlation analysis results help in selecting the highly correlated factors of acoustic propagation to construct a prediction model using the BP neural network. The model proposed herein is a typical supervised learning algorithm that employs the learning process of the error BP algorithm to autonomously learn the complex relation between the sound waves and environmental factors, thereby improving the accuracy and precision of the predictions. Moreover, the model provides valuable insights for the comprehensive analysis of acoustic propagation forecasts.

3. Results of the statistical analysis between KEF and acoustic propagation

3.1. Variation characteristics of KEF strength with depth

The HTG in the KE region serves as a principal method for measuring KEF. Extensive research has been conducted on the variation characteristics of the KE temperature front (Seo *et al.*, 2014; Wang *et al.*, 2016; 2020; Yu *et al.*, 2020). However, there is limited understanding of the evolving trends of subsurface KEF characteristics beneath the sea surface. In this study, we combine the KEF range offered by Kida *et al.* (2015) and adopt the method proposed by Sugimoto *et al.* (2014). We utilize two indicators, namely, average HTG and maximum HTG, within the region of 142°–162°E and 33°–37°N as the measures of the KEF strength. For a comprehensive representation of the KEF strength, we statistically analyze the depths at which these two indicators reach their maximum values, and their frequency and distribution are plotted in Fig. 3.

The findings indicate that the KEF strength increased and the strength exhibited an increasing trend followed by a decreasing trend with depth. The peak intensity of KEF frequently occurred at depths of 300 and 400 m, with strengths of 0.8395 and 0.7656°C/km, respectively. Below 800 m, the frequency approached 0, indicating that the peak KEF strength occurred most frequently at depths of 300 and 400 m.

This finding aligns with the conclusion of Liu *et al.* (2015), which demonstrates that the intensity of KEF

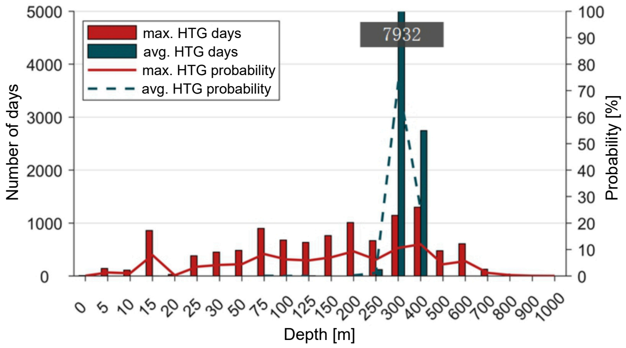

Fig. 3. Statistics and frequency distribution of the strongest depth frequency in KEF.

is considerably higher between 200 and 500 m compared to other water layers. Similarly, the average temperature gradient showcases a pronounced high-value zone at 300 m, indicating the largest range of the KEF at this depth. The maximum average temperature gradient in this area is 0.0646°C/km.

Thus, this study demonstrates that the utilization of the HTG at a depth of 300 m provides improved accuracy in determining the precise location of the KEF. This methodology is more effective compared to solely relying on surface intensity and position as it uncovers a comprehensive understanding of the variation characteristics exhibited by the KEF.

This study categorized the intensity levels of the KEF based on a maximum HTG of 300 m. Combining statistical data from JCOPE daily data, the KEF with a maximum HTG between 0 and 0.2°C/km were defined as weak fronts, accounting for 22.0% of the total recorded days. Further, the KEF with a maximum HTG between 0.2 and 0.3°C/km were classified as moderate intensity fronts (50.7%), whereas those with a gradient between 0.3 and 0.9°C/km were classified as strong fronts (27.3%). Three different months were selected to examine acoustic propagation under varying frontal conditions (Table 3), and a correlation heatmap was plotted to illustrate the impact of ocean fronts on the oceanic structure and acoustic propagation.

Table 3. Representative month for three intensity temperature gradients.

Intensity	Month	Temperature gradient [°C/km]	Monthly average [°C/km]
Strong	2012.07	0.304(7.14)–0.831(7.30)	0.5498
Middle	2002.12	0.220(12.1)–0.315(12.5)	0.2793
Weak	2022.08	0.130(8.16)–0.388(8.10)	0.1874

3.2. Statistical analysis of the impact of environmental parameters on underwater acoustic propagation in the KEF environment

To pinpoint the location of the most potent KEF, the grid where it transpires is identified using daily data, and a section spanning 12 nodes (~1°) from both the warm and cold sides of KEF is selected as the research scope for the given day. Subsequently, sound sources are arranged in the first 11 nodes, and information regarding ocean structure alterations and acoustic propagation distance between the outliers is documented (Fig. 4). Throughout the research process, only the alterations in the sound field environment of adjacent grids are considered. For instance, when the sound source is located at Station A1, the sound field environment is constructed using the sound speed profile between the A1 and A2 grids. Thereafter, 10 sets of sound speed profiles from A2 are replicated along the acoustic propagation direction to construct the environmental sound field file. This methodology aids in circumventing any interference caused by the changes in the ocean structure at other locations on acoustic propagation. Considering the practical scenarios in which sonar is employed, this study sets the deployment depth of the sound source at 30 m and the receiver depth at 150 m.

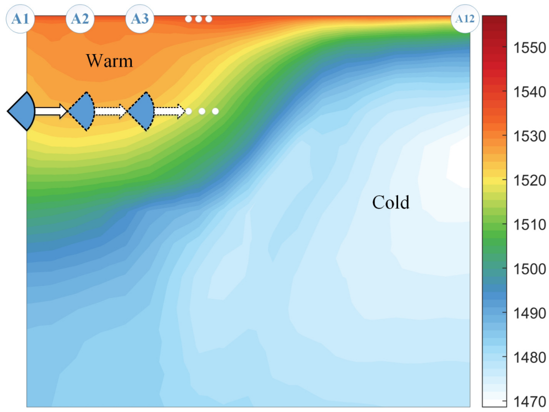

Fig. 4. Schematic of the research methodology (the base figure represents the sound velocity distribution of 2012.7.30 section [m/s]).

Subsequences of 5 environmental parameters and 2 underwater acoustic propagation parameters, with each subsequence containing 11 datasets, were constructed. The daily subsequences of the mentioned parameters are concatenated into monthly sequences for correlation analysis using the equations:

$$X = x_1 \oplus x_2 \oplus \ldots \oplus x_N, \tag{2}$$

$$r = \frac{\sum_{i=1}^{M}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{M}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{M}(Y_i - \overline{Y})^2}} \quad (M = 11N), \tag{3}$$

where x and y symbolize two factors, n signifies the number of days in the current month, and the $\oplus$ operator indicates the sequence concatenation. The correlation analysis results are displayed in Fig. 5, where

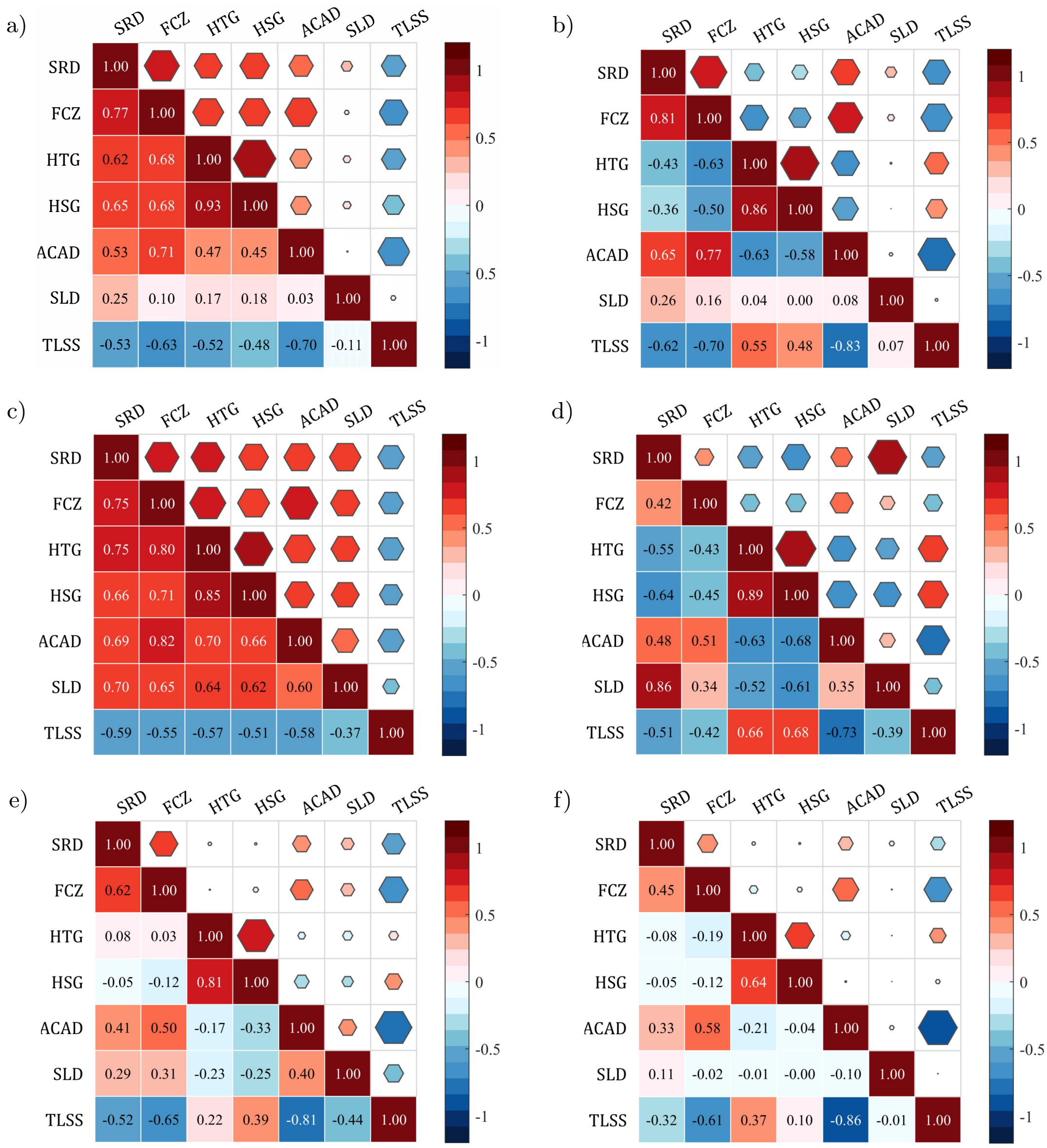

Fig. 5. Heatmap of correlation between sea structure and the underwater acoustic propagation parameters of Kuroshio Extension front presented at three different intensities. Panels (a) and (b) display data from July 2012 with warm-water side to cold-water side and cold-water side to warm-water side orientations, respectively. Panels (c) and (d) present data from December 2002, with warm-water side to cold-water side and cold-water side to warm-water side orientations, respectively. Panels (e) and (f) display data from August 2022, with warm-water side to cold-water side and cold-water side to warm-water side orientations, respectively.

the correlation heatmap translates the values in the correlation matrix into colors based on specific rules, thus visualizing the correlation through color variations. Warm colors suggest a positive correlation, and cold colors denote a negative correlation. The color and data of each cell in the heatmap represent the correlation between the row and column variables. The primary conclusions are:

(1) When the sound source emits sound waves from the warm-water side toward the cold-water side, the horizontal temperature and salinity gradients considerably influence the detection range at the strong KEF (July 2012) and the moderate KEF (December 2002). The correlation coefficients for the temperature gradient are 0.62 and 0.75, whereas for the salinity gradient, they are 0.65 and 0.66, respectively. Thereafter, we focused on the depth of the acoustic channel axis, exhibiting a positive correlation of 0.53 and 0.69. The vertical gradient of the transition layer of sound speed demonstrates a negative correlation with absolute correlation coefficients exceeding 0.5. The impact of the sound layer depth on underwater acoustic propagation displays noticeable seasonal characteristics. For the moderate KEF in winter, the surface sound layer

is thicker, resulting in a high correlation coefficient of 0.7 between the sound layer depth and the short-range detection distance in December. Conversely, in July, a weak correlation is seen. At the weak KEF (August 2022), the correlation between the horizontal temperature-salinity gradient and the short-range detection distance considerably decreases and becomes unrelated. The factors with higher correlation denote the vertical gradient of the transition layer of sound speed and the depth of the acoustic channel axis, exhibiting a negative correlation of −0.53 and a positive correlation of 0.41, respectively.

(2) When the sound source emits sound waves from the warm-water side toward the cold-water side, the depth of the acoustic channel axis correlates most strongly with the FCZ distance, with a maximum positive correlation coefficient of 0.82. Thereafter, the vertical gradient of the transition layer of sound speed, which displays a significant negative correlation, has a maximum value of −0.63. The horizontal temperature-salinity gradient highly correlates with the convergence zone than within the range of the source reception depth, indicating that as the sound source migrates from the KEF to the cold-water side with the reduced KEF strength, the horizontal distance between the FCZ distance and the sound source becomes shorter.

(3) When the sound source emits sound waves from the cold-water side toward the warm-water side, the vertical gradient of the transition layer of sound speed greatly impacts underwater acoustic propagation at short range, with negative correlation coefficients of −0.62, −0.51, and −0.32. The correlation coefficient decreases with the KEF strength. The depth of the acoustic channel axis displays the highest positive correlation of 0.65 with the short-range underwater acoustic propagation distance. The horizontal temperature-salinity gradient exhibits a negative correlation with the short-range underwater acoustic propagation distance when the sound source is located on the warm-water side, with correlation coefficients ranging from −0.3 to −0.6. The sound layer depth exhibits the strongest correlation in December 2002, reaching 0.86, while the remaining periods display weak or no correlation.

(4) The statistical results are analogous when the sound source is on the warm-water side. The vertical gradient of the transition layer of sound speed and the depth of the acoustic channel axis pose the most significant impact on the convergence zone, with absolute correlation coefficients ranging from 0.4 to 0.8. Subsequently, the horizontal temperature-salinity gradient, which shows different degrees of negative correlation, ranged between −0.4 and −0.6 at strong and moderate KEF intensities. The correlation of the temperature gradient is slightly higher (by 0.13) than that of the salinity gradient by suggesting that as the sound source migrates from the KEF to the warm-water side with a decrease in the KEF strength, the horizontal distance between the FCZ and the sound source lengthens.

Based on the above analysis, we concluded that the KEF has a more pronounced effect when the sound source is situated on the warm-water side. Moreover, the correlation between the HTG and HSG on the warm-water side and the underwater acoustic propagation parameters is 0–0.3 higher than on the cold-water side. Among the factors examined, the depth of acoustic channel axis and the vertical gradient of the transition layer of sound speed have the most significant influence on the convergence zone distance. The correlation between the former and the distance of the FCZ can reach 0.52–0.82, whereas the latter ranges from −0.42 to −0.7. Moreover, the FCZ distance has a slightly higher correlation of HTG (0–0.13) compared to HSG. Nevertheless, under weak front conditions, the temperature-salt gradient is either weakly correlated or uncorrelated with underwater acoustic propagation. Additionally, the sound layer exhibits noticeable seasonal variation characteristics, posing a substantial impact on the short-range detection distance (up to 0.86) in winter and a weak correlation in July and August.

4. Establishment and validation of the underwater acoustic propagation prediction model

Construction of the underwater acoustic propagation prediction model based on the BP neural network follows the same acoustic field construction method as described in Sec. 3. Over a 30-year period, the JCOPE2M dataset from February and August were utilized to construct the input sequences, encompassing four datasets with the sound source positioned on the warm-water and cold-water sides. Based on the statistical analysis of the correlations between the input parameters, the five environmental parameters were analyzed as input features for the month of February, while for the month of August, four environmental parameters, excluding the sound layer depth, were considered as the input features. The output features of the model include the distance to the FCZ distance and the short-range detection distance.

After performing normality tests on the data and eliminating outliers based on the 3σ rule, the selected sequences were randomly sorted and split into training and testing sets at a ratio of 9:1. The BP neural network was employed to estimate initial predictions on the first convergence distance of the ocean front zone. This is a standard algorithm used to train ANNs. By training with sample data and constant adjustments of network weights and thresholds, the error function can be minimized in the direction of a negative gradient, thereby approaching the desired output. In this

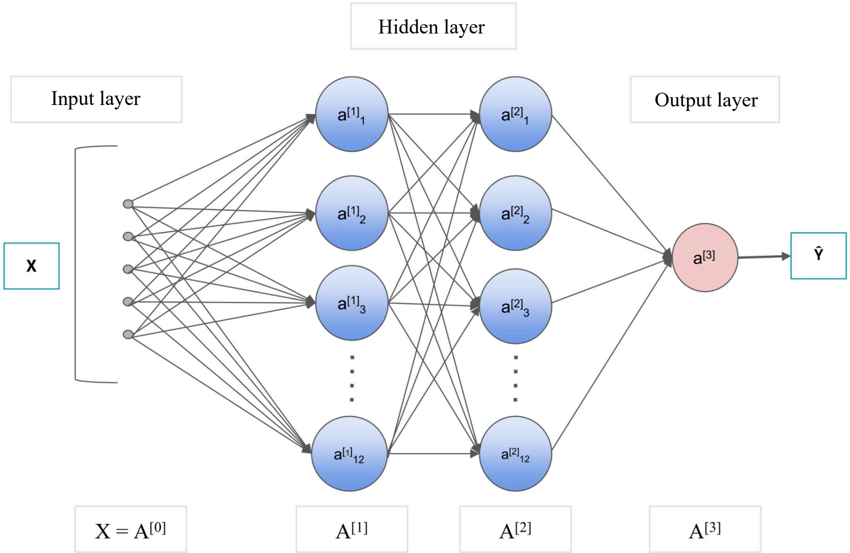

Fig. 6. Structural diagram of the backpropagation (BP) neural network.

study, the sigmoid function was utilized for the hidden and output layers. Two hidden layers were set, and a training algorithm was used for optimization iteration. The following empirical equation was used to determine the number of nodes in the hidden layer:

$$m = \sqrt{n + l} + \partial, \tag{4}$$

where m signifies the number of nodes in the hidden layer, n represents the number of nodes in the input layer, l represents the number of nodes in the output layer, and ∂ is a constant ranging from 1 to 10, which determines the range of hidden layer nodes. Multiple network structures are created within this range, and each network with varying nodes in the hidden layer is modeled and trained. The mean square error is computed for each network, and the optimal number of nodes in the hidden layer is determined to be 12 based on the lowest mean square error. The structure of the constructed neural network is illustrated in Fig. 6.

When the sound source is positioned on the warm side, a scatter plot is drawn between five parameters and the distance to the FCZ (Fig. 7). The correlation between the depth of the acoustic channel axis and the distance to the FCZ distance displayed the best continuity, suggesting that a deeper acoustic channel axis generally produces a convergence zone at a more distant location. The distribution of the sound layer depth is similar to the former but more spread out. The horizontal thermohaline gradient shows a trend of increasing and then decreasing with the distance to the convergence zone, with the maximum value appearing at ~60 km. The vertical gradient of the transition layer of sound speed exhibits a decreasing trend with the convergence zone, which is in accordance with the conclusions derived in the previous section.

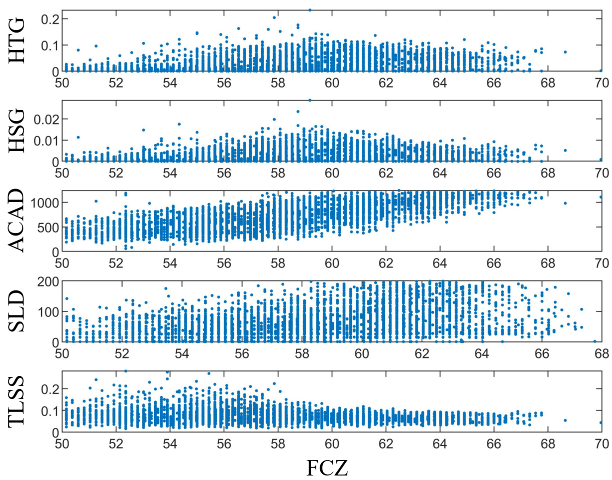

Fig. 7. Scatter plot between the convergence zone and the fitting factors.

Using the constructed BP neural network, we modeled and predicted the convergence zone distance in the KEF environment in February. The regression and fitting results of the prediction are presented in Fig. 8. We observed that the predicted outcomes were primarily clustered near the regression line, and the trend variation between the predicted values and the input values was generally consistent. The regression coefficient (R) was 0.84, indicating a good prediction effect.

The training outcomes of the remaining three datasets are similar to the first one. Table 4 presents the fitting performance of the four datasets evaluated in terms of the mean absolute error (MAE) and the coefficient of determination (R^2), both of which reflect the degree of agreement between the predicted and actual datasets; smaller values imply superior fit-

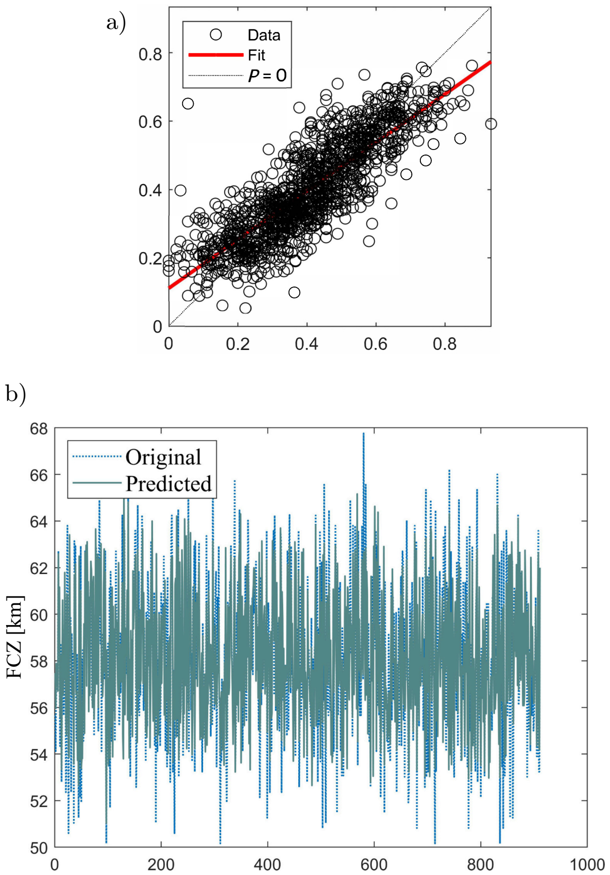

Fig. 8. Prediction effect of BP neural network: a) prediction regression scatter plot (abscissa denotes the normalized original distance and ordinate charts the normalized predicted distance); b) line graph comparing the predicted distance and the original distance of the FCZ distance (the horizontal coordinate is the number of data).

ting performance. The percentage deviation between the predicted and original datasets is also provided. Overall, the results observed that the predictive performance of the August dataset is superior, with an MAE of ~0.3–0.4 km lower than that of the February dataset. When the sound source is situated on the cold-water side, the frequency of test data with prediction errors under 1 km reaches 68.61% and that with errors under 2 km surpasses 90%, which is higher than that of the February dataset on the same cold-water side (49.24 and 76.52%, respectively). As there is a certain thickness of the sound layer in February, the distribution of convergence distance is more dispersed, resulting in an R^2 of >0.7 for the warm-water side of February, offering the best fitting effect among the four datasets. However, the proportion of test data with training errors of less than 1 km is less than half.

Similarly, we follow the same forecasting process for modelling and predicting underwater acoustic propagation distances at short range (Table 5). The occurrence of the sound layer is closely associated with the season, and the prediction accuracy of the short-range detection distance exhibits significant seasonal variations. In the summer months, the surface layer in the KE region is dominated by sound ducts, where sound rays rapidly bend downward, resulting in relatively shorter horizontal detection distances. Therefore, when the sound source is located on the warm-water and cold-water sides in August, the probability of the prediction error being less than 3 km is extremely high, exceeding 99%. On the contrary, during the winter season, the surface layer is typically dominated by the sound layer. Based on the BP neural network, the probability of the prediction error at less than 3 km was ~80%.

In conclusion, this research assembled four datasets with sound sources located on the cold-water and warm-water sides of the KEF for the months of February and August over a span of 30 years. The BP neural network was employed to establish a prediction model for the FCZ distance and the short-range detection distance. The results verified that the convergence zone distance and the short-range detection distance

Table 4. Predictive performance of the four datasets.

Month	Sound position	Train		Test		Predicted error [%]		
		MAE	R^2	MAE	R^2	<1 km	<2 km	<3 km
2	W	1.35	0.71	1.39	0.71	45.77	75.30	90.67
	C	1.26	0.61	1.25	0.58	49.24	76.52	90.65
8	W	0.99	0.70	0.98	0.65	59.73	87.77	96.13
	C	0.50	0.37	0.51	0.37	68.61	91.62	97.85

Table 5. Prediction performance of the source reception depth distance based on the backpropagation neural network.

Month	Sound position	Train		Test		Predicted error [%]		
		MAE	R^2	MAE	R^2	<1 km	<2 km	<3 km
2	W	1.71	0.47	1.80	0.43	39.45	68.72	83.36
	C	1.96	0.40	2.28	0.30	36.45	61.68	74.77
8	W	0.42	0.55	0.41	0.54	95.52	99.59	99.90
	C	0.50	0.40	0.51	0.35	90.80	98.98	99.69

can be feasibly predicted by fitting the model using five oceanic front structures and acoustic field environmental factors. Experimental results indicated that the predicted distances followed the same trend as the training set, with superior prediction performance in August. The proportion of data with a prediction error of less than 3 km for the convergence zone distance exceeded 90%, and for the short-range detection distance, it accounted for ~80% of the total data.

5. Conclusions

In this study, JCOPE2M, a high-resolution ocean reanalysis product, was utilized to determine the depth of the maximum intensity of the KEF based on the horizontal absolute gradient. Subsequently, three months with different KEF intensities were selected, and environmental parameters and underwater acoustic propagation parameters were constructed using the DP algorithm and the Bellhop underwater acoustic propagation model. Through analysis, the effects of the KEF on underwater acoustic propagation were revealed. Moreover, a BP neural network with 2 hidden layers, each containing 12 hidden layer nodes, was constructed to model and forecast the convergence zone distance based on five factors. The primary findings of this study are:

(1) Our analysis revealed that the strength of KEF follows an increasing-then-decreasing pattern as the depth increases. The highest frequency of the most intense KEF occurred in the water layer with a depth of 300–400 m, showcasing a maximum strength ranging from 0.7656 to 0.8395°C/km. Nevertheless, the frequency of the strongest KEF tended to be zero in the water layer with a depth exceeding 800 m.

(2) This study primarily focused on two acoustic propagation parameters: the short-range detection distance and the FCZ distance. The influence of the KEF is more pronounced when the sound source is located on the warm-water side. The correlation between the horizontal temperature and salinity gradient on the warm-water side and the acoustic propagation factors is superior compared to the cold-water side, ranging from 0 to 0.3. When examining the short-range detection distance, significant seasonal effects are observed. In winter, the impact on the short-range detection distance is significant, with a maximum correlation of 0.86, whereas in July and August, the correlation is weaker. The significant factors influencing the FCZ distance are the acoustic channel axis depth and the vertical gradient of the transition layer of sound speed discontinuity. The former demonstrated a correlation with the FCZ distance ranging from 0.52 to 0.82, whereas the latter ranged from −0.42 to −0.7. The horizontal temperature and salinity gradient are the subsequent influencing factors. The correlation of the temperature gradient is higher than that of the salinity gradient, with the difference ranging from 0 to 0.13. Under weak front conditions, the temperature and salinity gradient may exhibit weak or no correlation with underwater acoustic propagation.

(3) There is considerable potential for using neural networks in forecasting the convergence zone distance. Based on the constructed sequences of ocean and acoustic environmental parameters, the BP neural network can achieve good fitting and prediction of the convergence zone distance in ocean front environments. The prediction performance in August surpasses that in February, and the prediction accuracy is the highest when the sound source is located on the cold-water side. The frequency of the prediction distance error less than 3 km exceeds 90%, and the highest frequency with an error less than 1 km is 68.61%.

The correlation between the vertical temperature and salinity gradient and underwater acoustic propagation in this study depends on the modeling approach. To avoid the mutual influence among the analyzed environmental parameters, the KEF was segmented for analysis, which introduced some discrepancies from the actual KEF. However, these deviations did not impact the proposed approach for underwater acoustic propagation prediction based on the environmental parameters outlined in this study. Future research should focus on improving the selection and discrimination of the environmental parameters by utilizing more accurate modeling and feature extraction methods to enhance prediction accuracy.

Acknowledgments

This research was supported by the Dalian Naval Academy Research and Development Fund. The authors would like to express their sincere gratitude to the National Ocean Partnership Program (NOPP) and National Centers for Environmental Information for providing several years of current (https://data.hycom.org) and bathymetry data (https://www.ncei.noaa.gov). We are also grateful to the Japan Agency for Marine-Earth Science and Technology for their support in providing the JCOPE2M data (https://www.jamstec.go.jp), particularly to Mr. RuoChao Zhang for his assistance with processing the JCOPE2M data. Finally, we would like to acknowledge the assistance provided by other scholars and institutions who helped us during the course of this research.

References

1. Amante C., Eakins B.W. (2009), ETOPO1 1 Arc-Minute Global Relief Model: Procedures, data sources and analysis. NOAA technical memorandum NESDIS NGDC-24, *National Geophysical Data Center*, doi: 10.7289/V5C8276M.

2. Chang Y.-L., Miyazawa Y., Guo X. (2015), Effects of the STCC eddies on the Kuroshio based on the 20-year JCOPE2 reanalysis results, *Progress in Oceanography*, **135**: 64–76, doi: 10.1016/j.pocean.2015.04.006.

3. Chang Y.-L.K., Miyazawa Y., Béguer-Pon M., Han Y.-S., Ohashi K., Sheng J. (2018), Physical and biological roles of mesoscale eddies in Japanese eel larvae dispersal in the western North Pacific Ocean, *Scientific Reports*, **8**(1): 503, doi: 10.1038/s41598-018-23392-5.

4. Chassignet E.P. *et al.* (2007), The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system, *Journal of Marine Systems*, **65**(1–4): 60–83, doi: 10.1016/j.jmarsys.2005.09.016.

5. Chen C., Yang K., Duan R., Ma Y. (2017), Acoustic propagation analysis with a sound speed feature model in the front area of Kuroshio Extension, *Applied Ocean Research*, **68**: 1–10, doi: 10.1016/j.apor.2017.08.001.

6. Chen S. (2008), The Kuroshio Extension Front from satellite sea surface temperature measurements, *Journal of Oceanography*, **64**(6): 891–897, doi: 10.1007/s10872-008-0073-6.

7. Cheney R.E., Winfrey D. (1976), Distribution and classification of ocean fronts, *U.S. Naval Oceanographic Office*, Washington D.C.

8. Doan V.-S., Huynh-The T., Kim D.-S. (2020), Underwater acoustic target classification based on dense convolutional neural network, *IEEE Geoscience and Remote Sensing Letters*, **19**: 1–5, doi: 10.1109/LGRS.2020.3029584.

9. Dong S., Sprintall J., Gille S.T. (2006), Location of the Antarctic Polar Front from AMSR-E satellite sea surface temperature measurements, *Journal of Physical Oceanography*, **36**: 2075–2089, doi: 10.1175/JPO2973.1.

10. Dreini G., Jensen F.B. (1990), Sound propagation through an oceanic front, *Journal de Physique Colloques*, **51**(C2): 1025–1028, doi: 10.1051/jphyscol:19902240.

11. Hamilton E.L. (1980), Geoacoustic modeling of the sea floor, *The Journal of the Acoustical Society of America*, **68**(5): 1313–1340, doi: 10.1121/1.385100.

12. Etter P.C. (2013), *Underwater Acoustic Modeling and Simulation*, 4th ed., CRC Press, Taylor & Francis Group.

13. Kida S. *et al.* (2015), Oceanic fronts and jets around Japan: A review, *Journal of Oceanography*, **71**: 469–497, doi: 10.1007/s10872-015-0283-7.

14. Lee-Leon A., Yuen C., Herremans D. (2021), Underwater Acoustic Communication Receiver Using Deep Belief Network, *IEEE Transactions on Communications*, **69**(6): 3698–3708, doi: 10.1109/TCOMM.2021.3063353.

15. Lee B.M., Johnson J.R., Dowling D.R. (2022), Neural network predictions of acoustic transmission loss uncertainty, *The Journal of the Acoustical Society of America*, **152**(4): 152, doi: 10.1121/10.0015877.

16. Liu J., Piao S., Zhang, M., Zhang S., Guo J., Gong L. (2021), Characteristics of three-dimensional acoustic propagation in Western North Pacific Fronts, *Journal of Marine Science and Engineering*, **9**(9): 1035, doi: 10.3390/jmse9091035.

17. Liu J., Zhang Y., Zhang X. (2015), Analysis on the characteristics of the spatial and temporal variation of the Kuroshio Extension Front and the distribution of the sound field [in Chinese], *Journal of Ocean Technology*, **34**(02): 15–20, http://hyjsxb.cnjournals.org/ch/reader/view_abstract.aspx?file_no=20150203&flag=1.

18. Liu Z.-J., Nakamura H., Zhu X.-H., Nishina A., Guo X., Dong M. (2019), Tempo-spatial variations of the Kuroshio current in the Tokara Strait based on long-term ferryboat ADCP data, *Journal of Geophysical Research: Oceans*, **124**(8): 6030–6049, doi: 10.1029/2018JC014771.

19. Mellberg L.E., Johannessen O.M., Connors D.N., Botseas G., Browning D.G. (1991), Acoustic propagation in the western Greenland Sea frontal zone, *The Journal of the Acoustical Society of America*, **89**(5): 2144–2156, doi: 10.1121/1.400908.

20. Miyazawa Y. *et al.* (2017), Assimilation of high-resolution sea surface temperature data into an operational nowcast/forecast system around Japan using a multi-scale three-dimensional variational scheme, *Ocean Dynamics*, **67**(6): 713–728, doi: 10.1007/s10236-017-1056-1.

21. Miyazawa Y. *et al.* (2019), Temperature profiling measurements by sea turtles improve ocean state estimation in the Kuroshio-Oyashio Confluence region, *Ocean Dynamics*, **69**(2): 267–282, doi: 10.1007/s10236-018-1238-5.

22. Ozanich E., Gawarkiewicz G., Lin Y.-T. (2022), Study of acoustic propagation across an oceanic front at the edge of the New England shelf, *The Journal of the Acoustical Society of America*, **152**(6): 3756, doi: 10.1121/10.0016630.

23. Porter M.B. (2011), *The BELLHOP Manual and User's Guide: Preliminary Draft*, Heat, Light, and Sound Research, Inc., USA.

24. Rousseau T.H., Siegmann W.L., Jacobson M.J. (1982), Acoustic propagation through a model of shallow fronts in the deep ocean, *The Journal of the Acoustical Society of America*, **72**(3): 924–936, doi: 10.1121/1.388173.

25. Seo Y., Sugimoto S., Hanawa K. (2014), Long-Term variations of the Kuroshio Extension Path in winter: Meridional movement and path state change, *Journal of Climate*, **27**(15): 5929–5940, doi: 10.1175/JCLI-D-13-00641.1.

26. Shapiro G., Chen F., Thain R. (2014), The effect of ocean fronts on acoustic wave propagation in the Celtic Sea, *Journal of Marine Systems*, **139**: 217–226, doi: 10.1016/j.jmarsys.2014.06.007.

27. Sugimoto S., Kobayashi N., Hanawa K. (2014), Quasi-decadal variation in intensity of the western part of the winter subarctic SST Front in the Western North Pacific: The influence of Kuroshio extension path state, *Journal of Physical Oceanography*, **44**(10): 2753–2762, doi: 10.1175/JPO-D-13-0265.1.

28. Urick R.J. (1975), Principles of Underwater Sound, 3rd ed., McGraw-Hill, New York, London.

29. Wang J., Mao K., Chen X., Zhu K. (2020), Evolution and structure of the Kuroshio extension front in spring 2019, *Journal of Marine Science and Engineering*, **8**(7): 502, doi: 10.3390/jmse8070502.

30. Wang Y., Yang X., Hu J. (2016), Position variability of the Kuroshio Extension sea surface temperature front, *Acta Oceanologica Sinica*, **35**(07): 30–35, doi: 10.1007/s13131-016-0909-7.

31. Yasuda I. (2003), Hydrographic structure and variability in the Kuroshio-Oyashio transition area, *Journal of Oceanography*, **59**(4): 389–402, doi: 10.1023/A:1025580313836.

32. Yu P., Zhang L., Liu M., Zhong Q., Zhang Y., Li X. (2020), A comparison of the strength and position variability of the Kuroshio Extension SST front, *Acta Oceanologica Sinica*, **39**(5): 26–34, doi: 10.1007/s13131-020-1567-3.

Archives of Acoustics Vol. 49, No. 1, pp. 107–120 (2024), doi: 10.24425/aoa.2024.148773

Research Paper

DOA Estimation of Ultrasonic Signal by Indirect Phase Shift Determination

Bogdan KRECZMER

Department of Cybernetics and Robotics, Faculty of Electronics, Photonics and Microsystems
Wrocław University of Science and Technology
Wrocław, Poland; e-mail: bogdan.kreczmer@pwr.edu.pl

(*received March 25, 2023; accepted October 4, 2023; published online January 17, 2024*)

The paper presents the concept of the method of determining the direction of ultrasonic signal arrival, i.e., the azimuth and elevation angles. This method is an extension of the previous approach which was proposed to determine only the azimuth angle. The approach is based on the indirect phase determination. This makes it possible to tolerate spacing of receivers greater than half the wavelength of the received signal. At the same time, it provides increased measurement accuracy and reduced hardware requirements. To check the robustness of the method, simulations were carried out for the geometric arrangement of the receivers of the sonar module, for which the method was then implemented. This sonar module was used in the conducted experiments. The results of these simulations and experiments are included in the paper and discussed.

Keywords: direction of arrival; signal phase; sonar; ultrasonic range finder.

1. Introduction

Determining the direction of arrival (DOA) for an ultrasonic signal is an essential issue for the effective use of ultrasonic echolocation in mobile robot navigation. However, this is not the only area of application of this type of sensors. A very important field where it is also possible to use such a device is in supporting the activities of people who are blind or visually impaired. In this case in particular, a very important requirement is, on the one hand, to reduce energy consumption and simplify calculations, and on the other hand, to increase the reliability of detection and perception of the environment. This problem is addressed in this paper. The proposed approach makes it possible to estimate the DOA using the indirect determination of the signal phase shift. Because it does not rely on a signal amplitude measurement, no sampling is needed and computations are simplified. Experimental verification of the method was carried out using a sonar module equipped with four convectional piezoelectric receivers.

The problem of determining DOA is an area of intensive research. There are many methods to deal with this problem. A brief overview of them is given in the next section. Then, in Sec. 3, the basic concept of the indirect phase determination method is presented. The problem that arises in the proposed method is the ambiguity of the solutions. This is described in Sec. 4. How the ambiguities can be removed is the subject of Sec. 5. Then, Sec. 6 discusses the robustness of the method. Computations of the limit of the tolerable measurement errors for a specific arrangement of receivers are presented in Sec. 7. This layout of receivers was applied to the sonar module used in the experiments. Section 8 presents an approach to reducing the impact of noise on DOA determination. The results of the experiments are described and discussed in Sec. 9. The conclusions and scope of future work are presented in the final section.

2. Determining direction of signal arrival

The simplest approaches to DOA estimation are based on triangulation methods. They rely on time-of-flight (TOF) measurements (Peremans *et al.*, 1993; Kleeman, 1995; Choi *et al.*, 2014). However, much more effective approaches use array signal processing. These are the most common techniques for determining DOA that find applications in radar, sonar, medicine or communications. Two categories of such approaches can be distinguished, namely spectral-based and parametric (Krim, Viberg, 1996). Methods falling into the first of these categories include beamforming (Krishnaveni *et al.*, 2013) and the algorithm of multiple signal classification (MUSIC) (Schmidt, 1986), as well as its variants, i.e., the total spectral search MUSIC method (Zhou *et al.*, 2013), or the partial spectral search (Sun *et al.*, 2015). The second category includes methods such as deterministic maximum likelihood (DML) (Cadzow, 1990), root-MUSIC (Barabell, 1983) and the estimation of signal parameters via rotational invariance techniques (ESPRIT) (Roy *et al.*, 1986; Roy, Kailath, 1989). The latter method has several other variants and adaptations, e.g., Unitary ESPRIT (Haardt, Nossek, 1995), Conjugate ESPRIT (Tayem, Kwon, 2003).

The aforementioned methods require a large number of floating-point calculations. Beamforming allows for some simplifications. This technique acts as a spacial filter (Van Veen, Buckley, 1988). It includes delay-and-sum approaches and as well as other methods based on different weights of added signals that are delivered by individual receivers. Changes in the weight values make it possible to control sidelobes. However, even with uniform weights, good results can be obtained (Dokmanić, Tashev, 2014). The most important advantage of the delay-and-sum technique is that it can be implemented in hardware. Such implementations using MEMS microphones and FPGA are presented in (Kerstens *et al.*, 2017; 2019; Allevato *et al.*, 2020). Unfortunately, this does not allow for high angular resolution. Therefore, the improved version of the sensor presented in (Kerstens *et al.*, 2019), uses an implementation of the MUSIC algorithm (Verellen *et al.*, 2020). Such an approach provides much more accurate environmental data at the cost of increased computational complexity. To reduce this, methods based on co-prime arrays are proposed (Yang *et al.*, 2018; Li *et al.*, 2020). In this approach, a co-prime array is divided into two or more uniform sparse subarrays. In these subarrays, the distances between the receivers can be greater than half the wavelength of the received signal. This results in ambiguity in the determination of DOA for each subarray. However, they can be eliminated because the distances in each subarray are co-prime numbers. The similar idea can be found in (Gialich *et al.*, 2012). Instead of exploiting several subarrays with uniform spacing between receivers greater than half a wavelength, a single such receiving array is used in (Chen *et al.*, 2020). The problem of DOA ambiguity is solved by emitting pulses at different frequencies.

Despite described improvements, the discussed approaches involve methods that are relatively computationally expensive. Most of them are able to determine the DOA of signals that come from several sources. The approach presented in this paper is restricted to the problem of DOA determination for an echo coming from a single direction. This is the obvious drawback. However, when an emitted signal is short enough, it can be acceptable for many applications of mobile robot navigation. The important advantages are that the computational burden is reduced and the hardware requirements are very low. Despite this, the method allows for good DOA estimation accuracy. In this sense, the proposed approach can be exploited in order to build an inexpensive sonar that will be an attractive replacement of traditional ultrasonic range finders. Such an inexpensive device can also become a good support for the blind and visually impaired.

3. Indirect signal phase determination

To locate an object using ultrasonic sonar, in addition to measuring the distance to it, the key is to find the direction from which the echo of the reflected signal arrives. Using the horizontal coordinate system, DOA is determined by the azimuth and elevation angles denoted here as ϕ and θ, respectively (see Fig. 1). The problem of their determination can be simplified when considering a narrowband signal. This approach becomes natural when a piezoelectric transducer is used as an emitter. An important advantage of using piezoelectric over electrostatic transmitters is that their signal emission is actually limited to the ultrasonic band only. Electrostatic transducers during emission of ultrasonic wave, also emit quiet sizzles in the audible band. Despite the fact that they are not loud, when they are repeated very often they are annoying. This type of sizzling is also emitted by piezoelectric transducers. However, they are much quieter. In practice, they can be noted by a person when the transducer is no more than a few centimeters from the ear.

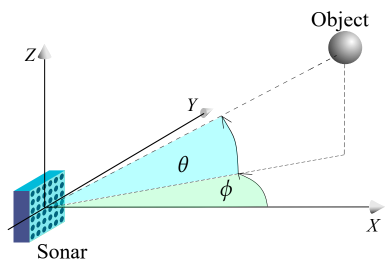

Fig. 1. Horizontal coordinates of an object in relation to a sonar.

Considering a narrowband signal, some simplifications can be made. It may be assumed that the signal carrier has a constant frequency. In fact, in the presented approach it is sufficient to assume that changes of signal carrier frequency are negligible during about three signal periods. Assuming that the signal source is in the far-field of the receivers, the signal wave can be treated as a plane wave. This is a common approach in signal array processing (Van Trees, 2004). Using these assumptions, a phase shift measurement can be applied to determine DOA. For the 1-D case, when only one angle is being determined, e.g., the azimuth angle, it is sufficient to use two receivers provided that their mutual distance is less than half the signal wavelength. For the 2-D case, when two angles are determined, i.e., the azimuth and elevation angles, three receivers are needed, which must be placed non-collinearly and must be two pairs whose mutual distances are lower than half the signal wavelength. Since a plane wave can be modeled as a set of planes relating to specific signal phases, wave points belonging to the same signal phase must satisfy the equation:

$$\mathsf{a}x + \mathsf{b}y + \mathsf{c}z + \mathsf{d} = 0, \tag{1}$$

where a, b, and c are coordinates of the vector perpendicular to the plane. Considering a plane wave, it is convenient to use a normalized vector pointed to the direction of wave propagation expressed by direction cosines, i.e.:

$$\boldsymbol{a} = \begin{bmatrix} -\cos\theta\cos\phi \\ -\cos\theta\sin\phi \\ -\sin\theta \end{bmatrix}. \tag{2}$$

Thus, Eq. (1) can be rewritten as

$$a_x x + a_y y + a_z z + \mathsf{d} = 0, \tag{3}$$

where $\boldsymbol{a} = (a_x, a_y, a_z)$. When a plane wave is propagated towards the origin of the coordinate system, then d is the distance of the plane wave from the origin. For a plane wave propagated in the opposite direction, d has the same absolute value, but with the minus sign. Equation (3) is a special form of a more general function:

$$D(\boldsymbol{p}) = a_x x_p + a_y y_p + a_z z_p + \mathsf{d}, \tag{4}$$

where $\boldsymbol{p} = (x_p, y_p, z_p)$. The absolute value of this function for a point $\boldsymbol{p}$ corresponds to its distance to the plane described by Eq. (3). The sign of this value depends on which side of the plane the point $\boldsymbol{p}$ is located.

Considering the set of three receivers, it is useful to choose such a coordinate system that its origin is at the center of one of them, e.g., the receiver R_0 (see Fig. 2). For the sake of simplicity, let us assume that the signal wavefront first reaches the receiver R_0 (this

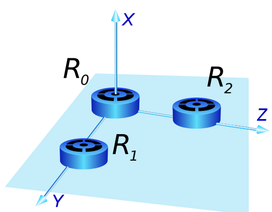

Fig. 2. Coordinate system for the set of three receivers.

assumption will be dropped later). Then, the wavefront is detected by receivers R_1 and R_2. The data obtained allow expressing the distances from receivers R_1 and R_2 to the plane wave (see Fig. 3a) as follows:

$$s_1 = v_a \tau_{01}, \qquad s_2 = v_a \tau_{02}, \tag{5}$$

where v_a is the speed of the acoustic wave, $\tau_{01} = t_1 - t_0$, $\tau_{02} = t_2 - t_0$ and t_0, t_1, t_2 are the wave detection times by receivers R_0, R_1, and R_2, respectively. On the other hand, these distances can be determined using the function $D(.)$ modeling a plane wave corresponding to the moment of arrival at the receiver R_0. It gives

$$D(R_1) = s_1, \qquad D(R_2) = s_2.$$

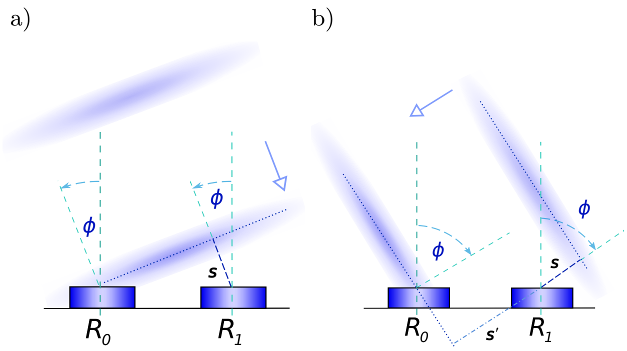

Fig. 3. Distance to the wavefront when the receiver R_0 is reached: a) before R_1; b) after R_1.

Since the plane wave at that moment contains the origin of the coordinate system, then $\mathsf{d} = 0$. Taking into account that $R_1 = (0, y_1, z_1)$ and $R_2 = (0, y_2, z_2)$ and the the vector $\boldsymbol{a}$ is normalized, a system of equations is obtained:

$$\begin{cases} a_y y_1 + a_z z_1 = s_1, \\ a_y y_2 + a_z z_2 = s_2, \\ a_x^2 + a_y^2 + a_z^2 = 1. \end{cases} \tag{6}$$

Its solution is

$$\begin{aligned} a_y &= \frac{z_2 s_1 - z_1 s_2}{y_1 z_2 - y_2 z_1}, \\ a_z &= \frac{y_1 s_2 - y_2 s_1}{y_1 z_2 - y_2 z_1}, \\ a_x &= \sqrt{1 - a_y^2 - a_z^2}. \end{aligned} \tag{7}$$

According to Eq. (2)

$$a_z = -\sin\theta \quad \wedge \quad a_y = -\cos\theta\sin\phi. \tag{8}$$

Due to the fact that $\theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, its cosine values are positive. As a result, the expression of a_y can be rewritten to the form:

$$a_y = -\left(\sqrt{1 - a_z^2}\right)\sin\phi.$$

This gives the formulas for calculating the azimuth and elevation angles:

$$\begin{aligned}\phi &= \arcsin - \frac{a_y}{\sqrt{1-a_z^2}},\\ \theta &= \arcsin - a_z.\end{aligned} \tag{9}$$

They show that a_x does not need to be computed to find ϕ and θ.

The presented calculation procedure can be directly applied to the obtained measurements when the distance between the receivers is not more than half the wavelength of the received signal, and the same wavefront is detected first by the receiver R_0 and then by R_1. In such a case, the signal detection times t_0 and t_1 satisfy the relationship:

$$0 \le t_1 - t_0 < \frac{T_a}{2}.$$

Otherwise, when after registering the first wavefront by R_0, R_1 registers the next wavefront (see Fig. 3b), the following condition is fulfilled:

$$\frac{T_a}{2} \le t_1 - t_0 < T_a.$$

These conditions are disjoint because the distance between the receivers is assumed to be less than half the wavelength. In the last case discussed, a correction is needed to calculate the correct distance s' (see Fig. 3b), namely:

$$\tau'_{01} = t_1 - t_0 - T_a. \tag{10}$$

In this case, $s' < 0$ which means that R_1 is behind the wavefront detected by R_0 (see Fig. 3b).

4. Ambiguities for larger mutual distances of receivers

Using popular piezoelectric transducers as receivers, it is impossible to meet the condition that their mutual distances are less than half the wavelength. This is because their smallest diameter is about 10 mm. Fortunately, there are available MEMS microphones on the market, which have much smaller housing and allow the mutual distances to be reduced. Sensor constructions which exploit them to create microphone arrays are presented in (Steckel *et al.*, 2013; Verellen *et al.*, 2020). Nevertheless, it is still useful to have longer mutual distances as this can reduce the error in determining the azimuth and elevation angles. In (Kreczmer, 2018), it was shown that such a result is obtained for 2-D sonar (i.e., determining distance and azimuth angle) for calculating the azimuth angle values. However, when mutual distances between receivers are greater than half the wavelength of the signal, ambiguities arise. This is due to the fact that the same distances s_1 and s_2 refer to several DOAs. As discussed before, when the plane wave detected by R_0 is not the same one detected by R_1 or R_2, a correction should be added to the measured time τ_{01} or τ_{02}, respectively. This correction is a multiple of the signal period T_a. It can be different for both receivers. Equation (10) is a special case of such a correction for which the multiplier equals −1.

The possible measurement values of τ_{0i} (where $i \in \{1,2\}$) are determined by the distance between receivers R_0 and R_i. Since for a given τ_{0i}, only the relative position of the two respective receivers need be considered, the range of possible multiples for the full range of arrival angles $\boldsymbol{\alpha} = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is limited to:

$$\mathcal{I}_{\boldsymbol{\alpha}}(b_{0i}) = \left\{k : k \in \boldsymbol{I} \ \wedge \ \left\lfloor -\frac{b_{0i}}{\lambda}\right\rfloor \le k \le \left\lfloor \frac{b_{0i}}{\lambda}\right\rfloor\right\},$$

where $i \in \{1,2\}$ and the range of variation of indices in further expressions, in this section, is the same, $\boldsymbol{I}$ is the set of integer numbers, b_{0i} is the distance between R_0 and R_i, λ is the wavelength of the signal, and $\lfloor . \rfloor$ is the floor function. Due to the directionality of the receivers and transmitter, the range of signal arrival angles that can be detected is much smaller, i.e., $\boldsymbol{\alpha} = [\alpha_{\min}, \alpha_{\max}] \subset \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. For simplicity, it can be assumed that the receivers have axial symmetry along their acoustic axes and have the same directivity pattern. Therefore, $\alpha_{\min} = -\alpha_{\max}$ and $\alpha_{\max} > 0$. Then, the range of possible multiples can be expressed as follows

$$\mathcal{I}_{\boldsymbol{\alpha}}(b_{0i}) = \left\{k : k \in \boldsymbol{I} \ \wedge \ \left\lfloor -\frac{b_{0i}\sin\alpha_{\max}}{\lambda}\right\rfloor \le k \le \left\lfloor \frac{b_{0i}\sin\alpha_{\max}}{\lambda}\right\rfloor\right\}.$$

Having a set of all multipliers, the set of possible values of time measurements can be constructed as follows:

$$\boldsymbol{\tau}_{\boldsymbol{\alpha}}(\tau_{0i}) = \{\tau_j : \tau_j = \tau_{0i} + kT_a, k \in \mathcal{I}_{\boldsymbol{\alpha}}(b_{0i}) \wedge |v_a \tau_j| < b_{0i}\}.$$

These times refer to the different possible directions of arrival of the wavefront detected by R_0, then R_1 or R_2. The times of signal detection by both receivers are related values. Their set can be defined as follows:

$$\boldsymbol{\tau}^2_{\boldsymbol{\alpha}}(\tau_{01}, \tau_{02}) = \{(\tau_i, \tau_j) : \tau_i \in \boldsymbol{\tau}_{\boldsymbol{\alpha}}(\tau_{01}) \wedge \tau_j \in \boldsymbol{\tau}_{\boldsymbol{\alpha}}(\tau_{02})\}.$$

Taking into account the velocity of the acoustic wave, a set of possible distances (s_i, s_j) to the wavefront can be created:

$$\begin{aligned}\boldsymbol{s}^2(\tau_{01}, \tau_{02}) = \{&(s_i, s_j) : s_i = v_a \tau_i, s_j = v_a \tau_j,\\ &(\tau_i, \tau_j) \in \boldsymbol{\tau}^2_{\boldsymbol{\alpha}}(\tau_{01}, \tau_{02})\}.\end{aligned}$$

Applying these distances to Eq. (6), the set of its solutions is obtained:

$$\mathfrak{a}^*(\tau_{01}, \tau_{02}) = \{(a_{x,1}, a_{y,1}, a_{z,1}), ..., (a_{x,k}, a_{y,k}, a_{z,k})\}.$$

Values of intervals τ_{01} and τ_{02} depend on geometry of the receiving system $\mathbb{R} = (R_0, R_1, R_2)$ and the received signal S. Therefore, it is more convenient to use $\mathbb{R}$ and S as arguments of $\mathfrak{a}^\star$. It is also worth to note that it is enough to know a_y and a_z to calculate angles ϕ and θ. This makes it possible to consider a bit more restricted set defined as follows:

$$\mathfrak{a}^\star(\mathbb{R}, S) = \{(a_{y,1}, a_{z,1}), ..., (a_{y,k}, a_{z,k})\}.$$

This set of solutions determines all possible values of ϕ and θ for given measurements.

5. Ambiguity removal

It can be easily shown for the 1-D case of DOA that when mutual distances between receivers are properly chosen then the only common direction in the set of determined possible arrival angles, is the proper one (Kreczmer, 2019) (see Fig. 4). The same idea can be applied to the 2-D case of DOA. The main difference is that an elementary receiving system in the 2-D case has to consist of three receivers instead of two. In Fig. 5a, there is an arrangement of two such elementary systems $\mathbb{R}_a = (R_0, R_1, R_2)$ and $\mathbb{R}_b = (R_3, R_4, R_5)$. Applying the aforementioned idea, the solution sought is

$$\mathfrak{a}^\star(\mathbb{R}_a, S) \cap \mathfrak{a}^\star(\mathbb{R}_b, S).$$

These two systems can be integrated into one set in such a way that both have two common receivers

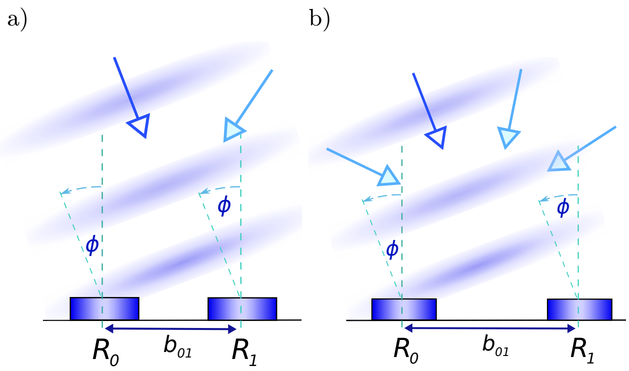

Fig. 4. Examples of possible incident angles for measurement data obtained while the real incident angle is equal to 20°. Directions were determined for the gap size b equal to: a) 11 mm; b) 15 mm.

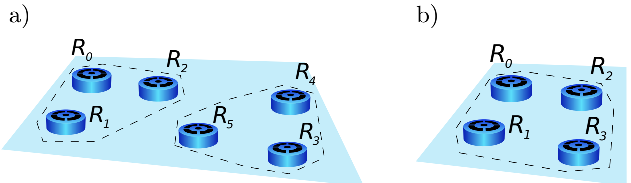

Fig. 5. Receiving system for the 2-D case that makes possible to solve the problem of ambiguity: a) two separate elementary receiving systems; b) a single integrated receiving system.

(see Fig. 5b). In this way, having the set of transducers $\boldsymbol{R} = (R_0, R_1, R_2, R_3)$, two receiving systems can be distinguished, i.e. $\mathbb{R}_1 = (R_0, R_1, R_2)$ and $\mathbb{R}_2 = (R_1, R_2, R_3)$. However, it is simple to notice, that also two additional systems can be found. In this case, they are $\mathbb{R}_3 = (R_0, R_1, R_3)$ and $\mathbb{R}_4 = (R_0, R_2, R_3)$. Taking into account all elementary systems, the solution sought is

$$\mathfrak{a}^\star(\mathbb{R}_1, S) \cap \mathfrak{a}^\star(\mathbb{R}_2, S) \cap \mathfrak{a}^\star(\mathbb{R}_3, S) \cap \mathfrak{a}^\star(\mathbb{R}_4, S).$$

6. Robustness

Due to the time measurement errors, the determined values of a_y and a_z are also burdened with them. It is convenient to assume that measurement errors of τ intervals are approximated by the same value $\Delta\tau$. Then, taking into account Eq. (7), errors in determining a_y and a_z can be estimated by formulas:

$$\begin{aligned} \Delta a_y &= \frac{|z_1| + |z_2|}{|y_1 z_2 - y_2 z_1|} v_a \Delta\tau, \\ \Delta a_z &= \frac{|y_2| + |y_1|}{|y_1 z_2 - y_2 z_1|} v_a \Delta\tau. \end{aligned} \tag{11}$$

They show that for each receiving system $\mathbb{R}_q$ the errors of determining the coordinates a_y and a_z are constant and strongly depend on the location of its receivers. Moreover, it is worth noting that to minimize Δa_y and Δa_z, the receivers R_1 and R_2 should be located along perpendicular lines crossing at the location of the receiver R_0. Following this idea and considering the arrangement of the receivers $R_1 = (0, y_1, 0)$ and $R_2 = (0, 0, z_2)$, Eq. (11) reduces to the form:

$$\Delta a_y = \frac{1}{|y_1|} v_a \Delta\tau, \qquad \Delta a_z = \frac{1}{|z_2|} v_a \Delta\tau. \tag{12}$$

6.1. Influence of measurement errors on the ambiguity of solutions

In order to take into account measurement errors, instead of a point set $\mathfrak{a}^2(\mathbb{R}_k, S)$, a set of rectangles is obtained. It can be defined as follows:

$$\begin{aligned} \mathfrak{a}^2_\Delta(\mathbb{R}_k, S) = \{ A^i_{S,k} &= [a_{y,i} - \Delta_k a_y, a_{y,i} + \Delta_k a_y] \\ &\times [a_{z,i} - \Delta_k a_z, a_{z,i} + \Delta_k a_z] : (a_{y,i}, a_{z,i}) \\ &\in \mathfrak{a}^2(\mathbb{R}_k, S)\}, \end{aligned}$$

where the index k identifies an elementary receiving system.

Assuming that the direction vector of propagation of the signal S is $\boldsymbol{a}_S = (a_{y,S}, a_{z,S})$, the uncertainty rectangle corresponding to the geometry of the system $\mathbb{R}_k$ and relating to the actual signal S is

$$\begin{aligned} A_{S,k} = &[a_{y,S} - \Delta_k a_y, a_{y,S} + \Delta_k a_y] \\ &\times [a_{z,S} - \Delta_k a_z, a_{z,S} + \Delta_k a_z]. \end{aligned}$$

According to the idea shown in Fig. 4, each receiving system determines the correct direction as one of possible solutions. Therefore, considering the set of receiving systems $\mathbb{R}^* = \{\mathbb{R}_1, \ldots, \mathbb{R}_m\}$, the uncertainty rectangles $A_{S,k}$ referring to subsequent receiving systems $\mathbb{R}_k$ must meet the condition:

$$\boldsymbol{a}_S \in \bigcap_{k=1}^{m} A_{S,k}. \tag{13}$$

The ambiguity is removed when there is only one sequence of uncertainty rectangles $A_{S,1}, ..., A_{S,m}$, denoted as $\boldsymbol{A}_S^*$, that creates a non-empty common part, i.e.:

$$\boldsymbol{a}_S \in \bigcap_{A_{S,k} \in \boldsymbol{A}_S^*} A_{S,k} \quad \wedge \quad \forall (A_{S,1}^j, ..., A_{S,m}^j) \neq \boldsymbol{A}_S^*, \qquad \bigcap_{k=1}^{m} A_{S,k}^j = \varnothing. \tag{14}$$

In order to be able in a simple way to select the common part of all uncertainty rectangles, it is essential to choose the same orientation for the local coordinate frames of all elementary receiving systems.

6.2. *Computing the angles of azimuth and elevation*

When the results of measurements are not corrupted by any errors, the uncertainty rectangles meeting Eq. (13) have their centers in the common point $\boldsymbol{a}_S$ (see Fig. 6a). In the case of actual measurement values, the centers of the uncertainty rectangles are in different places due to errors. For correctly estimated values of measurement errors, all rectangles should contain a point corresponding to the coordinates of the actual signal direction (see Fig. 6b). There are several ways to estimate the coordinates of the true signal direction $\boldsymbol{a}_S$. To describe them a bit more formally, a function which extracts the coordinates of the middle point of the rectangle, denoted as $M(.)$, is used.

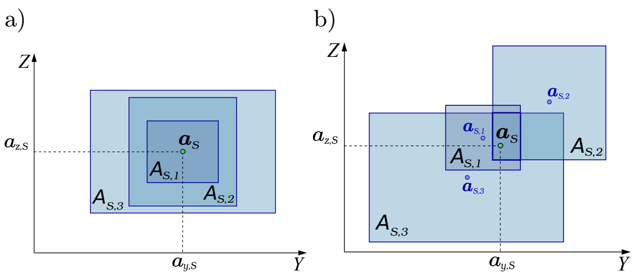

Fig. 6. Uncertainty rectangles satisfying Eq. (13) for the set of receiving systems $\mathbb{R}^* = \{\mathbb{R}_1, \mathbb{R}_2, \mathbb{R}_3\}$ and their measurements that: a) are not affected by errors; b) are affected by errors.

The simplest way is to calculate the mean values of the coordinates of the centers of the rectangles:

$$(\overline{a}_y, \overline{a}_z)_1 = \frac{1}{m} \sum_{i=1}^{m} M(A_{S,i}).$$

In order to distinguish the discussed approaches, a numerical index was provided along with the coordinates of the signal direction.

Using the more sophisticated approach, the uncertainty of the measured values are taken into account. The sizes of rectangles reflect the uncertainty of these values. The smaller the size, the more accurate the determination of the coordinates of the signal direction vector.

Denoting $L_y(.)$ and $L_z(.)$ as functions extracting the length of the uncertainty rectangle along the coordinate axes OY and OZ, respectively, the certainty coefficient for each coordinate can be expressed as follows:

$$C_y(A_{S,l}) = \frac{1}{L_y(A_{S,l})} \left(\sum_{k=1}^{m} \frac{1}{L_y(A_{S,k})} \right)^{-1},$$

$$C_z(A_{S,l}) = \frac{1}{L_z(A_{S,l})} \left(\sum_{k=1}^{m} \frac{1}{L_z(A_{S,k})} \right)^{-1}.$$

Then, the estimated coordinates can be computed using the formula

$$(\overline{a}_y, \overline{a}_z)_2 = \sum_{k=1}^{m} \Big(C_y(A_{S,k}) M_y(A_{S,k}),\, C_z(A_{S,k}) M_z(A_{S,k}) \Big),$$

where $M_y(.)$ and $M_z(.)$ extract the y and z coordinates of the center of the uncertainty rectangles $A_{S,k}$, respectively.

However, the most simple and very natural way is to calculate the coordinates of the center of the rectangle, that is the common part of them all:

$$(\overline{a}_y, \overline{a}_z)_3 = M\left(\bigcap_{i=k}^{m} A_{S,k} \right).$$

Having the coordinates $(\overline{a}_y, \overline{a}_z)$, the azimuth and elevation angles can be computed using Eq. (9). The effectiveness of the presented three approaches to DOA estimation is discussed in the next section.

7. Simulations

To find the range of tolerable measurement errors, computations were made for the azimuth and elevation angles using the signal detection times burdened with all possible combinations of measurement errors in a given range. The step of changing the time measurement error was $2^{-5} = 0.03125$ µs. The power of 2 was applied to eliminate the problem that results in the accumulation of calculation errors due to the finite representation of numbers. The value of time error Δt was increased until the value of the azimuth or elevation angle determined by the method no. 2 discussed in Subsec. 6.2 was greater than 10°. This procedure was also stopped when a combination of errors made it impossible to remove the ambiguity of the DOA solution. In this way, the maximum value of the simulated time error for which the two conditions aforementioned were met was taken as the limit of the tolerable time measurement errors. The computations were performed for the geometrical arrangement of receivers presented in Fig. 7. It corresponds to the geometry of the sonar module used in experiments presented in Sec. 9. It was assumed that the sensitivity range of the receivers was $[-40°, 40°]$ in each direction. The results of the computations are presented in Fig. 8. The biggest found value of Δt is 1.91 µs. The diagram presented in Fig. 8 does not have any plane symmetry, due to unsymmetrical arrangement of receivers. However, it has the axial symmetry induced by the axis of Δt. This is due to the symmetry of the sine function used in Eq. (9).

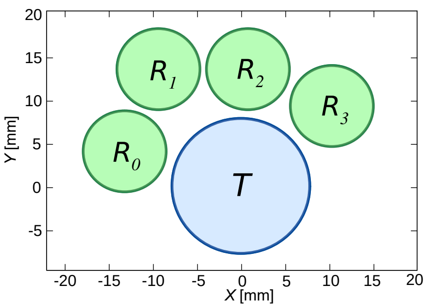

Fig. 7. Geometrical arrangement of sonar receivers and a transmitter.

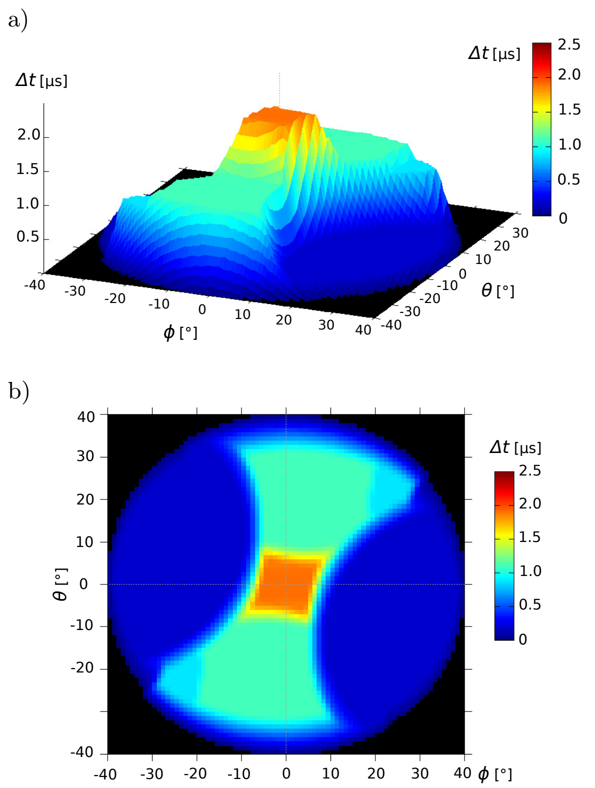

Fig. 8. Distribution of the maximal tolerable error of time measurement: a) side and b) top view.

In Subsec. 6.2, it was discussed how $\boldsymbol{a}_S$ can be estimated. The question then arises, using the same timing errors, which of the three approaches presented gives the best estimate of the angles ϕ and θ. Figure 9a shows a diagram comparing the maximum errors of azimuth angle estimation using the discussed approaches at $\theta = 0°$.

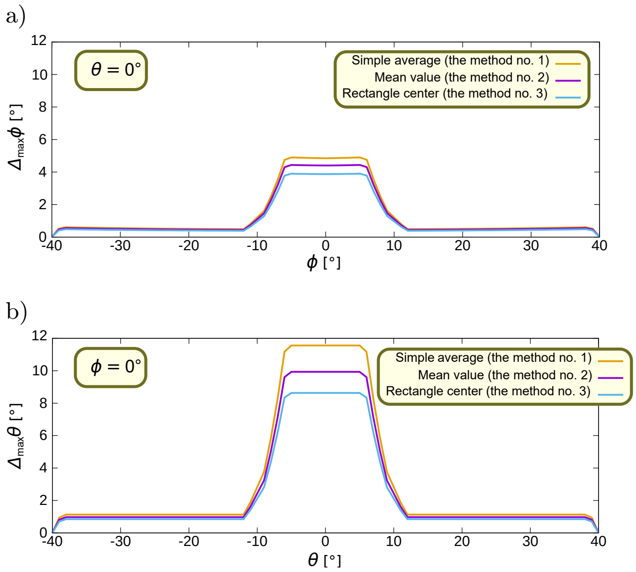

Fig. 9. The maximal errors for angles determined by the method no. 1, 2, and 3: a) for the azimuth angle at $\theta = 0°$; b) for the elevation angle at $\phi = 0°$.

An analogous diagram for the elevation angle at $\phi = 0°$, is presented in Fig. 9b. The mutual relation between errors in determining the angles ϕ and θ, visible in the diagrams, is also preserved for other directions. It was expected that the method no. 2 which involves some measure of uncertainty should be the best, but it is not. The best one is the method no. 3. An additional advantage of this situation is that this method is the simplest to compute. The same relation is also observed when comparing RMSD values.

It is worth noting that the error values for determining the elevation angle are much larger compared to the same type of errors for the azimuth angle (see Fig. 9). The increase in the error value is a direct consequence of the fact that the distance between the extreme receivers in the vertical plane is much smaller than their counterparts in the horizontal plane.

8. Reducing the impact of noise

The procedure described in the previous sections is applied to a single wavefront. However, it can also be applied to each successive one in a given wave packet, and then average values can be calculated. To explain this, the measurement procedure used in this method is presented in more detail. In order to detect a wave packet, the threshold method is used. It is worth emphasizing that this method is used only for detection, not for actual measurements, the results of which are used to determine DOA. When the signal in the receiver channels exceeds the detection threshold, in the next moment in each of these channels, the times of crossing the zero level are recorded (see Fig. 10a). Signal transitions through the zero level are detected only for raising edges. The results of these measurements do not depend on the value of the signal amplitudes. The measured times are used to determine the azimuth and elevation angles of DOA. Once a wave packet is detected, the same time measurements can be also performed for successive signal transitions through the zero level (see Fig. 10b). When the signal is reflected from a single object and there is no noise, the same angle values are obtained for all subsequent measurements. In the proposed approach, in order to reduce the influence of noise, averages of azimuth and elevation angles are calculated based on four consecutive wavefronts, whereby a measurement that deviates most from all others is discarded. This procedure starts with the third detected wavefront.

a)

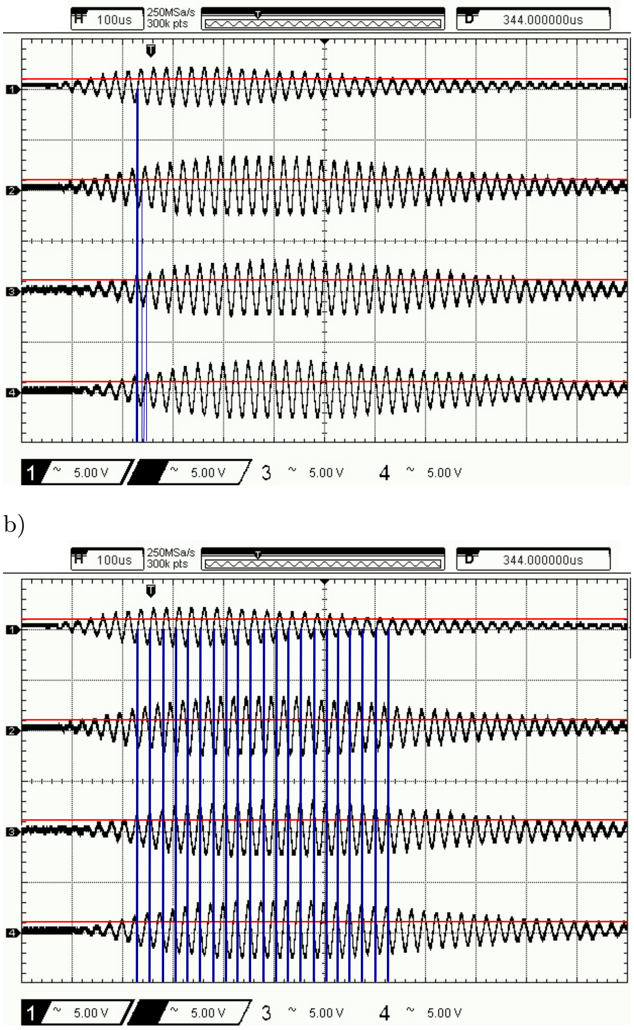

Fig. 10. Measurements of times of signal zero crossings: a) for a single signal pulse; b) for subsequent pulses for the entire wave packet, for clarity, moments of zero-crossing detection for the first receiver was marked only.

9. Experiments

In order to verify the effectiveness of the proposed method, a series of experiments was carried out using the sonar module shown in Fig. 11. This module consists of one BPU-1640IOAH12 transmitter and four MA40S4R receivers. All ultrasonic transducers are controlled by the mini-module (Cholewiński *et al.*, 2013), which exploits the microcontroller MK40DN512VLK10. To verify DOA determination for different orientations of the ultrasonic transducers, the sonar module was mounted on a rotating base. This made it possible to perform a scan in the horizontal plane with a step equal to 1°. Each scan was made in the angular range from −40° to 40°. It was taken 20 measurements for each orientation of the sonar module. This allowed the calculation of mean and RMSD values. The sonar module was placed 1.4 m above the floor. This made it possible to avoid the ground effect (Kapoor *et al.*, 2018). Measurements were made for two different cases. The first case was a series of measurements to a wall (see Fig. 13a). The second case was a series of measurements to a small glass ball suspended above the floor at different heights. For each case, the azimuth and elevation angles, i.e., ϕ and θ, were determined in the local coordinate system of the sonar module. To check the correctness of the calculation of the azimuth angle ϕ, its value was compared with the value of the orientation angle of the sonar module α, which was measured in the global coordinate system of the sonar stand. Considering the case of a wall, it can be noticed that echoes always arrive from the direction perpendicular to the wall surface and it does not depend on the sonar module orientation, if that direction is within the range of the sonar beam (see Fig. 12). Therefore, for the ideal case when there are no measurement errors, the relation between ϕ and α is as follows:

$$\phi = -\alpha. \tag{15}$$

A more detailed explanation of this feature can be found in (Kreczmer, 2019). The same relation is also

Fig. 11. Sonar module used in experiments.

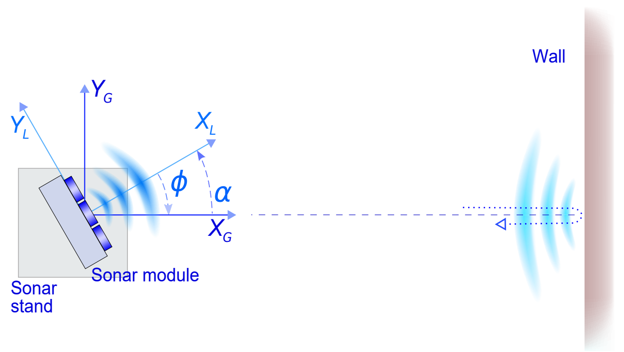

Fig. 12. Coordinate systems chosen for the experiment. The azimuth angle ϕ measured in the local coordinate system of the sonar module has an opposite sign to the rotation angle α measured in the global coordinate system.

valid for such an object as a glass ball used in these experiments. Performing measurements for the case of a wall, the sonar module was placed at a distance of about 2.4 m from the wall (see Fig. 13a). Over the entire range of sonar module orientations, the wall was detected, as shown in the diagram of distance measurements (see Fig. 13b). In the discussed experiments, the distance is measured by using the threshold detection method. Therefore, this type of measurements cannot be expected to be very accurate. Nevertheless, they show very well the orientation ranges of the sonar module for which an object was observed. Despite object detection over the entire range, DOA angles were determined properly in a smaller region, which is about $[-28°, 33°]$ (see Figs. 13c and 13d). This limitation results mainly from the directional characteristics of the transmitter, which emits a significantly weaker signal for directions well away from its acoustic axis.

According to the results of the simulations, it would be expected that the accuracy and precision of determining the azimuth angle would be higher than that of the elevation angle. However, this is not the case here. The diagram in Fig 14a shows the absolute values of the differences between the actual and determined values of the elevation and azimuth angles denoted as $\Delta\phi$ and $\Delta\theta$, respectively. It can be noticed that the discrepancies for both angles are similar. An analogous relationship is also observed for RMSD (see Fig. 14b). The reason is that the wall is a very good reflector and the returning echoes are strong. This results in small measurement errors. Their consequence is the small differences mentioned earlier.

The second series of experiments was done for the glass ball, which scatters the signal much more than the wall. In the first experiment of this series, the glass ball was placed at a distance of about 1.6 m from the sonar module and 1.8 m above the floor (see Fig. 15a). Due to the much weaker echo, SNR is decreased, and thus azimuth and elevation angles were determined correctly over a much smaller range of sonar module orientation (see Figs. 15b–d). This also results in

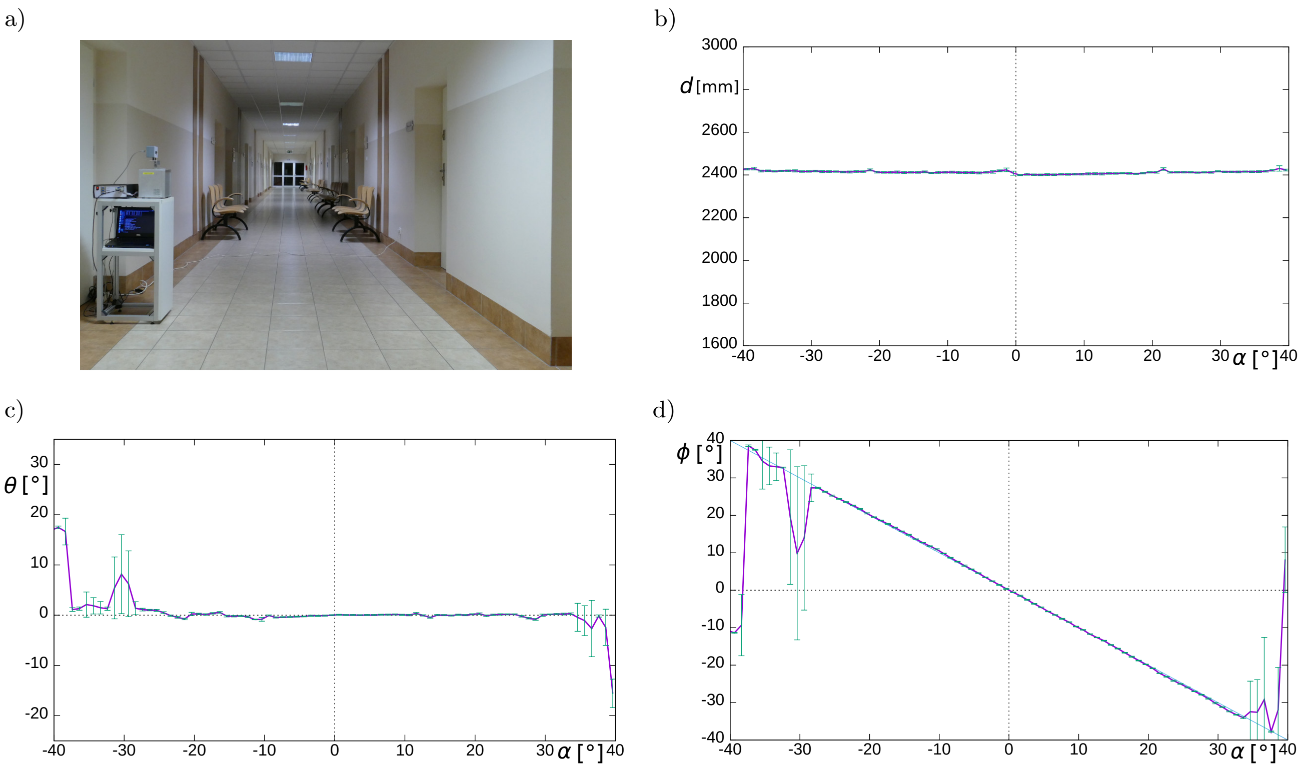

Fig. 13. Results of measurements performed for the wall depending on the orientation angle α of the sonar module, which was located at a distance of 2.4 m from the wall: a) arrangement of the sonar stand and the wall; b) values of the measured distance d; c) determined values of the elevation angle θ; d) determined values of the azimuth angle ϕ.

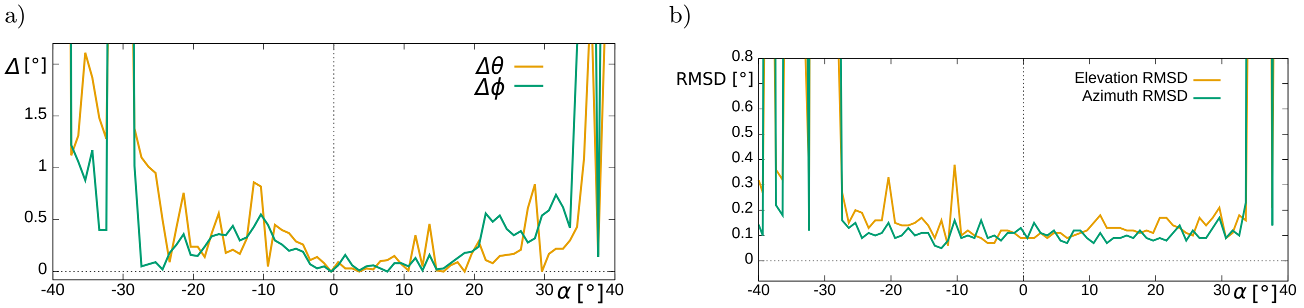

Fig. 14. Accuracy and precision of the azimuth and evaluation angles depending on the orientation angle α of the sonar module: a) accuracy expressed as the absolute value of the difference between the actual and determined values of the angles; b) precision expressed by RMSD of the determined values of the angles.

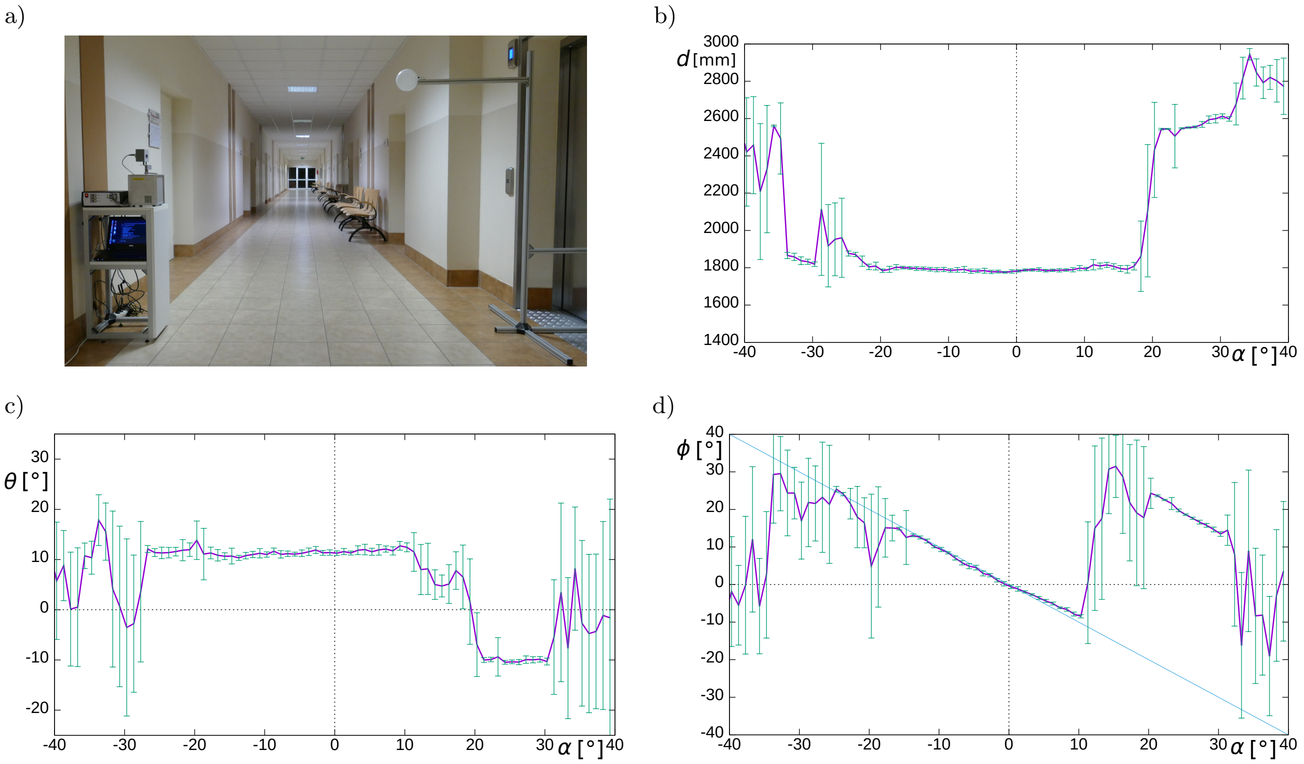

Fig. 15. Results of measurements performed for the glass ball depending on the orientation angle α of the sonar module. The glass ball was placed above the horizontal plane of the sonar module: a) arrangement of the sonar stand and the glass ball; b) values of the measured distance d; c) determined values of the elevation angle θ; d) determined values of the azimuth angle ϕ.

higher RMSD values. In this case, the differences between the RMSD for the azimuth and elevation angles become much more apparent and significant. In the range of α $[-15°, 10°]$, the azimuth angle values are as expected. In the same range, the results of determining the elevation angle change very little. This is also as expected. Unfortunately, its values are below the desired one which is $16°$ while the obtained results are in the range $[11°, 13°]$. The main reason for this discrepancy is the small distance of the receivers in the vertical direction compared to their size. Due to the large diameter of the receivers in relation to their mutual distances, it can be assumed that treating them as fixed points is not sufficient for different angles of incidence of the acoustic wave. With small mutual vertical distances, even small changes in the location of these points translate into significant discrepancies in the determined angle. This makes it difficult to obtain uniform calibration parameters for the entire range of angles considered in this experiment.

In the next series of measurements, the glass ball was placed near the horizontal plane of the sonar module (see Fig. 16a). The elevation angle was about $1°$. The object was detected in almost the entire range of orientation angles of the sonar module (see Fig. 16b). However, azimuth and elevation angle values, that were close to expected ones, were obtained only for the orientation range $[-20°, 19°]$ (see Figs. 16c and 16d).

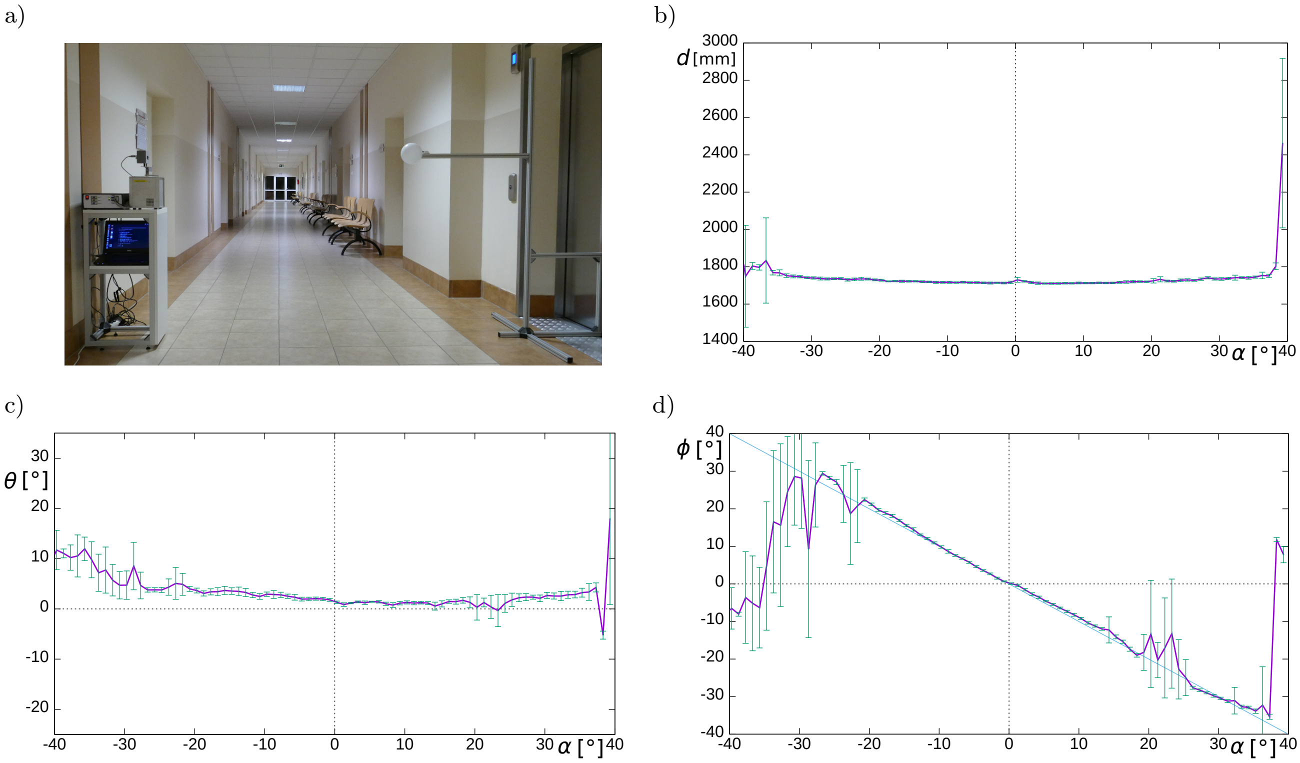

Fig. 16. Results of measurements performed for the glass ball depending on the orientation angle α of the sonar module. The glass ball was placed near the horizontal plane of the sonar module: a) arrangement of the sonar stand and the glass ball; b) values of the measured distance d; c) determined values of the elevation angle θ; d) determined values of the azimuth angle ϕ.

Compared to the previous case, a shift to the left of such an interval can be observed. This can be said even considering the much more limited visibility of the glass ball in the previous case. This is consistent with simulation results presented in Sec. 7 (see Fig. 8). In order to make it clearer and easier to interpret, the diagram shown in Fig. 8b should be presented in the space of the rotation angle α (orientation of the sonar module), instead of the azimuth angle ϕ. To do this, the relationship expressed by Eq. (15) must be taken into account. It flips the original diagram along the vertical axis (see Fig. 17a). The final form obtained after sonar calibration with marked lines relating to the results of the three experiments presented is shown in Fig. 17b. This offset is more evident in the results of the third experiment, when the glass ball was placed below the horizontal plane of the sonar module (see Fig. 18). The range of module orientations, for which

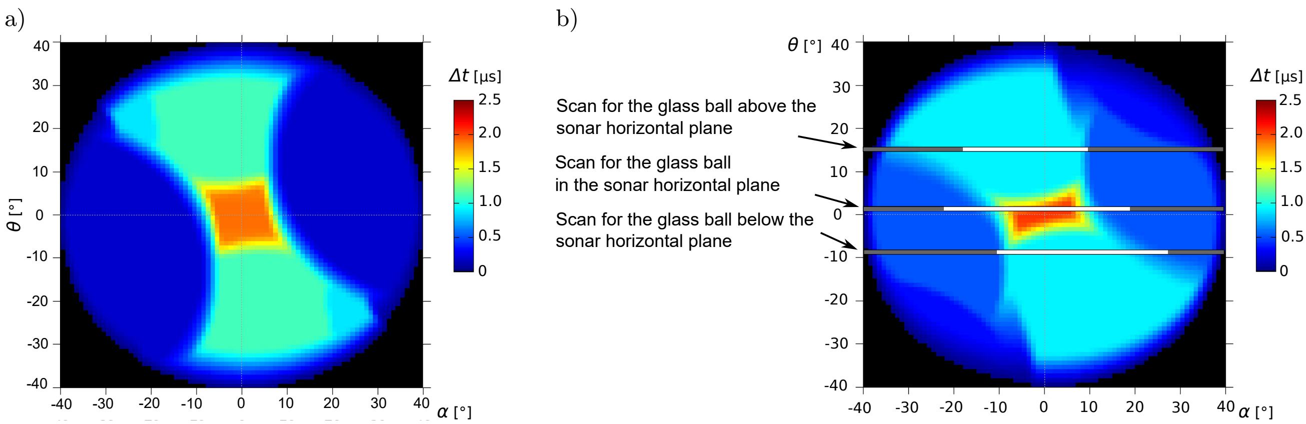

Fig. 17. Distribution of the maximal tolerable error of time measurement (top view) in space of the sonar module rotation angle α: a) for the arrangement of receivers in accordance with the original design; b) for the acoustic center coordinates of the receivers obtained after sonar calibration. The diagram also shows the lines corresponding to DOAs of the signals reflected by the glass sphere during the three consecutive experiments conducted. The white stripes represent regions where the determined DOAs were close to the expected values.

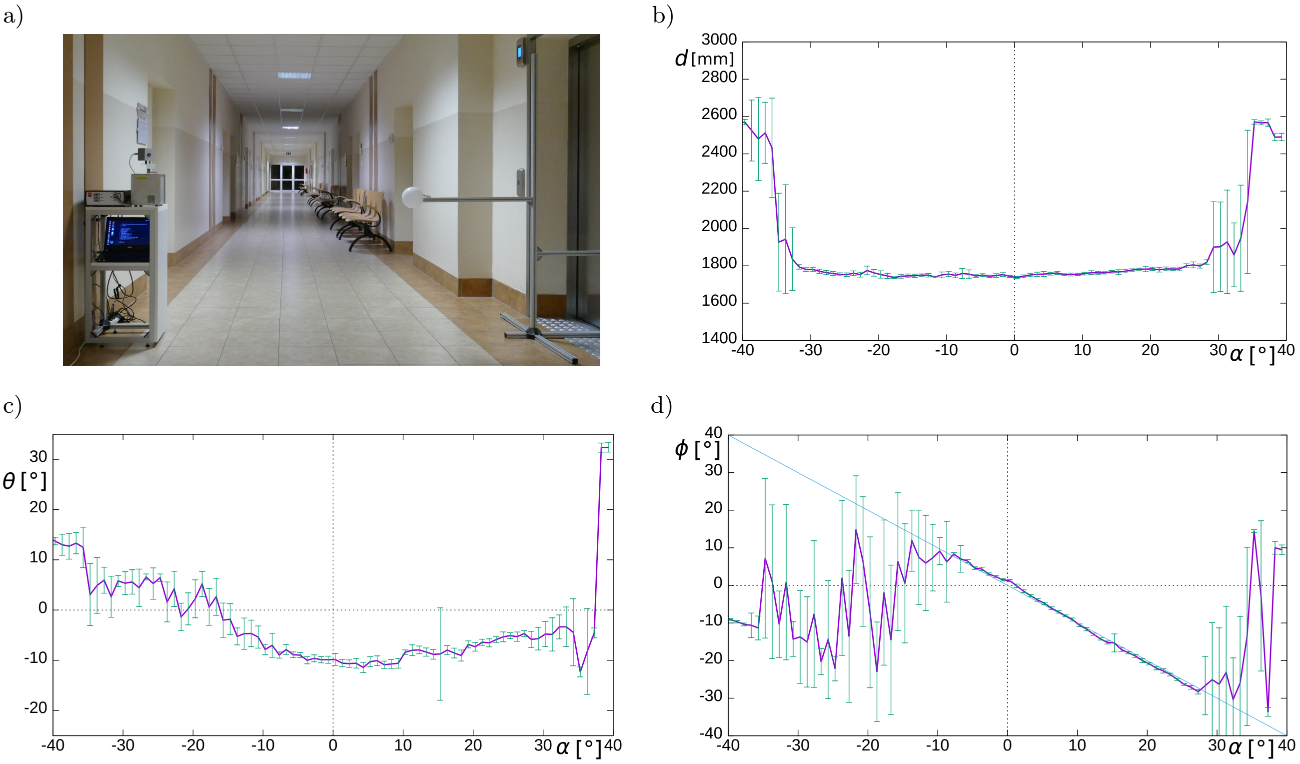

Fig. 18. Results of measurements performed for the glass ball depending on the orientation angle α of the sonar module. The glass ball was placed below the horizontal plane of the sonar module: a) arrangement of the sonar stand and the glass ball; b) values of the measured distance d; c) determined values of the elevation angle θ; d) determined values of the azimuth angle ϕ.

the azimuth angle was correctly determined, is shifted to $[-10°, 28°]$. Considering the elevation angle, larger deviations from the expected value (about 9°) are noticeable. It can be assumed that these disturbances result from the scattering of the returning ultrasonic wave at the upper edge of the base of the rotating sonar column (see Fig. 18a).

The simulation results presented in Fig. 17b take into account the correction of the position of the acoustic centers of the receivers obtained as a result of the calibration procedure. However, they do not take into account the directional patterns of the transmitter and receivers. Therefore, full agreement with experimental results cannot be expected. Nevertheless, the main trend of change is clearly visible. For the case of the wall shown in Fig. 13, this feature was not observed, as the received echoes were strong enough to cause very small measurement errors.

10. Conclusion and further research

The main advantage of the proposed method is its simplicity. However, the cost of simplicity is sensitivity to signal interference. Fortunately, the method makes it possible to distinguish the echo coming from a single object from overlapping echoes, and to assess the reliability of the results obtained (Kreczmer, 2021). The sonar module used in the experiments to verify the proposed method uses popular piezoelectric ultrasonic receivers. Due to their large size relative to their mutual distances, high accuracy DOA determination cannot be expected for a wide range of sonar module orientations. Nevertheless, in the case of a relatively strong echo received after reflecting the signal from objects such as a wall, the sonar module used works very well. The same cannot be said of the glass ball case. It is worth noting, however, that the observed limits of the sonar module orientation, for which acceptable results are obtained, are consistent with the simulation results. These limits are determined by the geometry of the receiver arrangement. The results of simulations and experiments also indicate that it is desirable to enlarge the distance between some receivers in order to increase the precision of angle determination. This can be clearly seen in the difference in the horizontal and vertical geometric arrangement of the receivers of the sonar module used. Due to this arrangement, the accuracy of the determined azimuth angle was greater than that of the elevation angle. To be able to determine azimuth and elevation angles with the same high accuracy, it is necessary to maintain an increased distance between the selected microphones in both horizontal and vertical directions. Another important consideration is to maximize in terms of angle and distance the area visible by the sonar, for which the greatest possible robustness to measurement errors will be obtained. It seems that this feature is rather difficult to achieve with piezoelectric microphones. MEMS microphones are much more useful for this case. Furthermore, 4 microphones seem insufficient to achieve improvements in DOA determination accuracy. How much their number should be changed and how to arrange them is a topic for further research.

References

1. Allevato G., Rutsch M., Hinrichs J., Pesavento M., Kupnik M. (2020), Embedded air-coupled ultrasonic 3D sonar system with GPU acceleration, [in:] *2020 IEEE SENSORS*, pp. 1–4, doi: 10.1109/SENSORS47125.2020.9278601.
2. Barabell A. (1983), Improving the resolution performance of eigenstructure-based direction-finding algorithms, [in:] *ICASSP'83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 336–339, doi: 10.1109/ICASSP.1983.1172124.
3. Cadzow J.A. (1990), Multiple source location the signal subspace approach, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **38**(7): 1110–1125, doi: 10.1109/29.57540.
4. Chen H., Ballal T., Muquaibel A.H., Zhang X., Al-Naffouri T.Y. (2020), Air writing via receiver array-based ultrasonic source localization, *IEEE Transactions on Instrumentation and Measurement*, **69**(10): 8088–8101, doi: 10.1109/TIM.2020.2991573.
5. Choi K.H., Ra W.-S., Park S.-Y., Park J.B. (2014), Robust least squares approach to passive target localization using ultrasonic receiver array, *IEEE Transactions on Industrial Electronics*, **61**(4): 1993–2002, doi: 10.1109/TIE.2013.2266076.
6. Cholewiński M., Dziergwa M., Kaczmarek P.M., Kędzierski J., Wnuk M. (2013), *Malaga minimodule* [in Polish], Technical report SPR no 13/2013, Wrocław University of Science and Technology.
7. Dokmanić I., Tashev I. (2014), Hardware and algorithms for ultrasonic depth imaging, [in:] *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6702–6706, doi: 10.1109/ICASSP.2014.6854897.
8. Gialich M., Im A., Lee T., Aliyazicioglu Z., Hwang H.K. (2012), DOA estimation using array antenna with large inter-element spacing, [in:] *2012 11th International Conference on Signal Processing (ICSP)*, pp. 1701–1704, doi: 10.1109/ICoSP.2012.6491908.
9. Haardt M., Nossek J.A. (1995), Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden, *IEEE Transactions on Signal Processing*, **43**(5): 1232–1242, doi: 10.1109/78.382406.
10. Kapoor R., Ramasamy S., Gardi A., Schyndel R.V., Sabatini R. (2018), Acoustic sensors for air and surface navigation applications, *Sensors*, **18**(2): 499, doi: 10.3390/s18020499.
11. Kerstens R., Laurijssen D., Steckel J. (2017), Low-cost one-bit MEMS microphone arrays for in-air acoustic imaging using FPGA's, [in:] *2017 IEEE SENSORS*, pp. 1–3, doi: 10.1109/ICSENS.2017.8234087.
12. Kerstens R., Laurijssen D., Steckel J. (2019), eRTIS: A fully embedded real time 3D imaging sonar sensor for robotic applications, [in:] *2019 International Conference on Robotics and Automation (ICRA)*, pp. 1438–1443, doi: 10.1109/ICRA.2019.8794419.
13. Kleeman L. (1995), A three dimensional localiser for autonomous robot vehicles, *Robotica*, **13**(1): 87–94, doi: 10.1017/S0263574700017513.
14. Kreczmer B. (2018), Azimuth angle estimation of ultrasonic signal arrival by using multi-pair receiver system, [in:] *Automation 2018. Advances in Intelligent Systems and Computing*, Szewczyk R., Zieliński C., Kaliczyńska M. [Eds.], **743**: 672–681, doi: 10.1007/978-3-319-77179-3_65.
15. Kreczmer B. (2019), Estimation of the azimuth angle of the arrival direction for an ultrasonic signal by using indirect determination of the phase shift, *Archives of Acoustics*, **44**(3): 585–601, doi: 10.24425/aoa.2019.129273.
16. Kreczmer B. (2021), Influence of signal interference on determining direction of arrival by using the indirect phase determination method, [in:] *Automation 2021: Recent Achievements in Automation, Robotics and Measurement Techniques. AUTOMATION 2021. Advances in Intelligent Systems and Computing*, Szewczyk R., Zieliński C., Kaliczyńska M. [Eds.], Springer International Publishing, pp. 319–328, doi: 10.1007/978-3-030-74893-7_30.
17. Krim H., Viberg M. (1996), Two decades of array signal processing research: The parametric approach, *IEEE Signal Processing Magazine*, **13**(4): 67–94, doi: 10.1109/79.526899.
18. Krishnaveni V., Kesavamurthy T., Aparna B. (2013), Beamforming for direction-of-arrival (DOA) estimation – A survey, *International Journal of Computer Applications*, **61**(11): 4–11, doi: 10.5120/9970-4758.
19. Li P., Li C., Ye C., Zhang X. (2020), Low complexity DOA estimation using coprime circular array, [in:] *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*, pp. 592–597, doi: 10.1109/ICSIP49896.2020.9339298.
20. Peremans H., Audenaert K., Van Campenhout J.M. (1993), A high-resolution sensor based on tri-aural perception, *IEEE Transactions on Robotics and Automation*, **9**(1): 36–48, doi: 10.1109/70.210793.
21. Roy R., Kailath T. (1989), ESPRIT estimation of signal parameters via rotational invariance techniques, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **37**(7): 984–995, doi: 10.1109/29.32276.
22. Roy R., Paulraj A., Kailath T. (1986), Direction-of-arrival estimation by subspace rotation methods – ESPRIT, [in:] *ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2495–2498, doi: 10.1109/ICASSP.1986.1168673.

23. Schmidt R. (1986), Multiple emitter location and signal parameter estimation, *IEEE Transactions on Antennas and Propagation*, **34**(3): 276–280, doi: 10.1109/TAP.1986.1143830.

24. Steckel J., Boen A., Peremans H. (2013), Broadband 3-D sonar system using a sparse array for indoor navigation, robotics, *IEEE Transactions on Robotics*, **29**(1): 161–171, doi: 10.1109/TRO.2012.2221313.

25. Sun F., Lan P., Gao B. (2015), Partial spectral search-based DOA estimation method for co-prime linear arrays, *Electronics Letters*, **51**(24): 2053–2055, doi: 10.1049/el.2015.2261.

26. Tayem N., Kwon H.M. (2003), Conjugate ESPRIT (C-SPRIT), [in:] *Proceedings of IEEE Military Communications Conference, 2003, MILCOM 2003*, pp. 1155–1160, doi: 10.1109/MILCOM.2003.1290358.

27. Van Trees H.L. (2004), Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory Detection, Estimation, and Modulation Theory, John Wiley & Sons, Inc.

28. Van Veen B.D., Buckley K.M. (1988), Beamforming: A versatile approach to spatial filtering, *IEEE ASSP Magazine*, **5**(2): 4–24, doi: 10.1109/53.665.

29. Verellen T., Kerstens R., Steckel J. (2020), High-resolution ultrasound sensing for robotics using dense microphone arrays, [in:] *IEEE Access*, **8**: 190083–190093, doi: 10.1109/ACCESS.2020.3032177.

30. Yang X., Wu X., Li S., Sarkar T.K. (2018), A fast and robust DOA estimation method based on JSVD for co-prime array, [in:] *IEEE Access*, **6**: 41697–41705, doi: 10.1109/ACCESS.2018.2860680.

31. Zhou C., Shi Z., Gu Y., Shen X. (2013), DECOM: DOA estimation with combined MUSIC for coprime array, [in:] *2013 International Conference on Wireless Communications and Signal Processing*, pp. 1–5, doi: 10.1109/WCSP.2013.6677080.

Archives of Acoustics Vol. 49, No. 1, pp. 121–128 (2024), doi: 10.24425/aoa.2024.148766

Technical Note

Low-Frequency Sound Absorption Potential of Subwavelength Absorbers Based on Coupled Micro-Slit Panels

Yujie QIAN*, Zhengyuan GAO, Jie ZHANG

College of Information Science and Engineering, Hohai University
Changzhou, China

*Corresponding Author e-mail: 20141923@hhu.edu.cn

(*received February 21, 2023; accepted August 26, 2023; published online December 15, 2023*)

Due to space limitations during installation, reducing low-frequency noise has always been a challenging area. Sub-wavelength structures are typically favored in such scenarios for noise reduction. This paper explores the potential of micro-slit panels (MSP) for low-frequency sound absorption. To further optimize the panel thickness, coupled MSPs (CMSP) with a distance between two MSPs of less than 1 mm are proposed. Firstly, the low-frequency absorption performances of a single MSP based on two optimized schemes – the cavity-depth optimal scheme (COS) and the panel thickness optimal scheme (TOS) – are examined and compared with those of existing ultrathin metamaterials. The results demonstrate that MSP has significant potential for low frequency sound absorption, and COS allows for a smaller overall structural thickness but a larger panel thickness than TOS. Secondly, to reduce the panel thickness, the CMSP is developed and the theoretical model of its acoustic impedance is established and validated by experiments. Then, based on the theoretical model, the low-frequency absorption potential of CMSP is optimized using COS. The results show that both the overall thickness and the panel thickness of the CMSP absorber are reduced while maintaining better performance. Furthermore, the proposed absorber achieves a subwavelength scale since its total thickness can be as small as 0.138λ.

Keywords: coupled MSP (CMSP); cavity-depth optimal scheme (COS); panel thickness optimal scheme (TOS); low frequency; absorption performance.

1. Introduction

The absorption of low-frequency (i.e., 100 to 500 Hz) noise has persistently posed a challenge due to the inherent weak dissipation of classic sound-absorbing materials (Ma, Sheng, 2016; Allard, Atalla, 2009). Despite the progress made by active noise reduction approaches in reducing low-frequency noise, the complexity of the devices may prevent from finding their practical use. In practical applications, porous/fibrous materials, resonance-type structures such as micro-perforated panels (MPP) (Maa, 1998; Park, 2013; Wu *et al.*, 2019; Xu *et al.*, 2020; Liu *et al.*, 2021a), and micro-slit panels (MSP) (Maa, 2000; Randeberg, 2000), are typically preferred. While sound absorbing structures made of porous/fibrous materials are an effective noise absorbing structure, they often require a body thickness comparable to the operating wavelength, which seriously hinders their application in the low-frequency range. Resonance-type structures (MPP and MSP) are regarded as the most promising sound-absorbing materials due to their simple structure, variety of material options and environmental friendliness.

Recently, it has become a balanced goal in low-frequency noise control to reduce the dimension of sound-absorbing structures (i.e., their total thickness) to well below subwavelength. This not only contributes to space savings but also responds to the trend of device miniaturization (Chong *et al.*, 2010; Mei *et al.*, 2012). Many resonance-type metamaterials targeting low-frequency noise have been studied and developed in this direction, including metamaterial panels based on thin closed slits (Jiménez *et al.*, 2016; 2017a; 2017b), subwavelength systems with space-coiling structures (Liang, Li, 2012; Cai *et al.*, 2014; Zhu *et al.*, 2019; Almeida *et al.*, 2021a; Wang *et al.*, 2018; Liu *et al.*, 2021b; Wu *et al.*, 2021; Shen *et al.*, 2019; Ryoo, Jeon, 2018), a slim subwavelength absorber (Zhao *et al.*, 2018; Donda *et al.*, 2019; Almeida *et al.*, 2021b), a multicoiled acoustic metasurface (Cheng *et al.*, 2015; Liu *et al.*, 2017) and a micro-perforated panel and coiled-up channel-based hybrid absorber (Li, Assouar, 2016).

The majority of recent research, however, has focused on creating coiled-up space inside cavities in order to extend the acoustic wave's effective propagation distance, thereby allowing for the thinning of absorbing structure in order to absorb low-frequency noise. Studies have shown that when the other structural parameters of MPP or MSP remain the same, the increase in panel thickness causes the sound resonance frequency to shift to lower frequencies. This is because the increase in panel thickness may not only result in an increase in the amount of air mass in the perforation/slit of the panel (i.e., an increase in reactance of the acoustic mass) but also significantly contributes to greater energy dissipation due to greater friction. Consequently, the absorption peak is shifted to lower frequencies (Vigran, 2014). However, the increase in panel thickness will, on one hand, increase the weight of the acoustic structure and, on the other hand, cause an increase in processing costs. Studies have shown that when other structural parameters of MPP or MSP remain the same, an increase in panel thickness causes the sound resonance frequency to shift to lower frequencies. This is because the increase in panel thickness may not only result in an increase in the amount of air mass in the perforation/slit of the panel (i.e., an increase in reactance of the acoustic mass), but also contributes significantly to a greater dissipation of the energy due to the greater friction, and thus shifting the absorption peak to lower frequencies (Vigran, 2014). But an increase in panel thickness will, on one hand, increase the weight of the acoustic structure and, on the other hand, cause an increase in processing costs.

It is widely recognized that the acoustic impedance of the micro-slits is proportional to the panel thickness (Maa, 2000; Randeberg, 2000). And vice versa, an increase in acoustic impedance corresponds to an increase in effective panel thickness. Based on this, this paper proposes a coupled MSPs, which consists of two MSPs with a distance of less than 1 mm between them, forming inter-panel micro-slits. As a result, the acoustic impedance of this proposed structure is not only provided by the micro-slits on the panels but also by the micro-slits between the panels. In this way, the equivalent panel thickness can be increased by adjusting the acoustic impedance provided by the micro-slits formed between the panels. Consequently, this approach can be utilized to improve low-frequency sound absorption performance.

The remainder of this paper is organized as follows. The mathematical expression of the normal absorption coefficient and an analytical model of the acoustic impedance for a single MSP absorber are provided in Sec. 2. Based on this, the exhaustive method is used to investigate the minimum cavity depth and the minimum panel thickness required for an MSP absorber at a given resonant frequency and maximum absorption coefficient. Next, the results obtained are then compared with those in existing literature for metamaterials. In Sec. 3, the coupled MSP absorber is proposed and its theoretical model of impedance is developed and validated by experiments. The coupled MSP structure is then used to further optimize the panel thickness, resulting in the reduction of the acoustic structure's total thickness. Finally, the conclusions are drawn in Sec. 4.

2. MSP absorber and its low-frequency absorption performance

2.1. MSP absorber

The basic structure of a traditional MSP absorber consists of a micro-slit panel, a rigid backing wall and the air cavity between them. This is illustrated in Fig. 1a, where d is the slit width, t is the panel thickness, b is the distance between centers of adjacent slits, and D is the depth of the air cavity. Based on the acoustoelectric analogy, the equivalent circuit of an MSP absorber is depicted in Fig. 1b. Here, Z_{MSP} is the specific acoustic impedance of the MSP and Z_D is the specific acoustic impedance of the air cavity. The sound wave impinging on the structure is equivalent to a source of sound pressure $2p$, as produced on the rigid wall with the time factor $\exp(-j\omega t)$ suppressed throughout (analogous to the open-circuit voltage) and internal resistance $\rho_0 c$ as that of air, where ρ_0 is the air density and c is the sound speed in air. The acoustic impedance of a micro-slit with end correction can be given as (Maa, 2000; Randeberg, 2000):

$$Z_{\text{MSP}} = \frac{12\eta t}{d^2}\sqrt{1+\frac{k^2}{18}} + i\rho_0\omega t\left(1+\frac{1}{\sqrt{25+2k^2}}\right) + R_s + iX_m \quad (1)$$

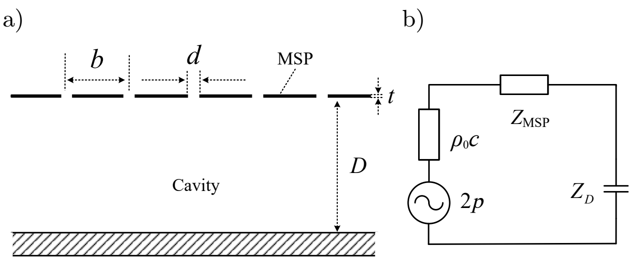

Fig. 1. Schematic diagrams of the composite sound absorber structure and its equivalent circuit.

with

$$k = d\sqrt{\rho_0\omega/\eta}/2, \tag{2}$$

$$R_s = \frac{\sqrt{2\eta\omega\rho_0}}{\sigma}, \tag{3}$$

$$X_m = -\frac{2}{\pi\sigma}\rho_0\omega d \ln\left[\sin\left(\frac{\sigma\pi}{2}\right)\right], \tag{4}$$

where $\omega = 2\pi f$ is the angular frequency with f being the frequency of the incident acoustic wave, σ is the perforation rate of the panel (the ratio of surface area of the slits to the total surface area of the panel), $\sigma = d/b$, i is an imaginary unit, R_s is the resistance end correction, and X_m is the reactance end correction.

The acoustic impedance of the air cavity behind the MPP with a depth of D is given by:

$$Z_D = -j\rho_0 c \cot(\omega D/c). \tag{5}$$

The overall acoustic impedance of an MSP absorber is given by:

$$Z = Z_{\text{MSP}} + Z_D. \tag{6}$$

The normal sound absorption coefficient is calculated using the equation:

$$\alpha = 1 - \left|\frac{Z - \rho_0 c}{Z + \rho_0 c}\right|^2. \tag{7}$$

2.2. Design optimization via exhaustive search

Due to the continuous improvement of computer computing power, the structural parameters of the MSP absorber can be globally designed and optimized by the exhaustive search (Lara-Valencia *et al.*, 2020; Pierro *et al.*, 2021), according to the noise reduction performance requirements. In this study, two parameters related to the sound absorption performance requirements are considered: the resonant frequency f_r and its corresponding absorption coefficient a_r. Once sound absorption requirements are given, an exhaustive search is used to select the combination of structural parameters that meet these requirements. If there are multiple sets of structural parameters that meet the requirements, the unique combination of parameters is determined by two selection schemes: one is to select the one with the smallest depth, smallest panel thickness and the largest resonance absorption coefficient in order of preference, i.e., the cavity-depth optimal scheme (COS), and the other is to give preference to the set with the smallest panel thickness, smallest depth, and the largest resonance absorption coefficient, i.e., the panel thickness optimal scheme (TOS). The combination of structural parameters used for optimization is: the micro-slit width d, panel thickness t, center distance between two adjacent slits b, and cavity depth D. The parameter range for the exhaustive search is:

$$\begin{aligned} &0.1\text{ mm} \le d \le 1\text{ mm}; \qquad 0.1\text{ mm} \le t \le 10\text{ mm}; \\ &0.3\text{ mm} \le b \le 170\text{ mm}; \qquad 5\text{ mm} \le D \le 30\text{ mm}. \end{aligned} \tag{8}$$

It should be noted that, firstly, the upper limit of the panel thickness is set to less than 10 mm, since the excessive panel thickness will increase the weight and manufacturing cost of the MSP. Secondly, the upper limit for the distance between the centers of adjacent micro-slits is set to 170 mm. This is because the low-frequency range studied in this paper is 100~500 Hz. To ensure the applicability of the above derived analytical equation for the acoustic impedance of the MSP (considering the micro-slits as parallel), the distance between two adjacent micro-slits should be less than 1/4 of the wavelength corresponding to the maximum frequency (Liu *et al.*, 2021).

Moreover, the lower and upper limits of the cavity depth are set to 5 and 30 mm, respectively. This is because, on one hand, this study is dedicated to finding structural parameters that not only meet the sound absorption requirements but also minimize the total thickness, so the upper limit should not be set too large, and, on the other hand, according to previous design experience, the cavity depth is too small to find a suitable combination of structural parameters, so the lower limit of the cavity depth should not be too small either. The search steps for each parameter are 0.01 mm for d and 0.1 mm for t, b, and D, respectively. Furthermore, combinations of parameters where the micro-slit width d is less than the micro-slit spacing b will be discarded.

2.3. Low-frequency absorption potential

Assuming that resonant frequencies studied are 254, 338.5, and 391 Hz, the optimal combination of parameters for different absorption coefficient requirements is determined via exhaustive search based on the specified above selection rules. The results are shown in Table 1, where T indicates the total thickness of the absorber. Note that all length-related variables in the tables of this paper are in millimeters. As can be seen in Table 1, as the required resonant frequency increases, both the required minimum cavity depth and minimum panel thickness decrease accordingly. This is reasonable, because the smaller the frequency the larger the wavelength, and thus the size of the required absorber increases.

Table 1. Performance requirements and the optimal parameter combination for MSP absorbers.

f_r	a_r	Group	D	t	d	b	T
254	≥ 0.98	MSP1-COS	21.7	8.4	0.99	169.8	30.1
		MSP1-TOS	30	3.7	0.73	170	33.7
338.5	≥ 0.98	MSP2-COS	13.4	7.5	0.99	170	20.9
		MSP2-TOS	29.3	0.5	0.38	170	29.8
391	≥ 0.98	MSP3-COS	10.5	7.1	0.99	170	17.6
		MSP3-TOS	25.8	0.1	0.24	170	25.9

To examine the low-frequency sound absorption potential of the MSP with the limited cavity depth, its absorption coefficients, based on Eq. (7), are compared with those of ultra-thin metamaterials (with the same theoretical resonant frequency) from existing literature (Jiménez *et al.*, 2016; Cai *et al.*, 2014). The schemes of the metamaterials used for comparison are shown in Fig. 2, and their relevant parameters are shown in Table 2. The comparison of the theoretical results for the normal absorption coefficient is presented in Fig. 3.

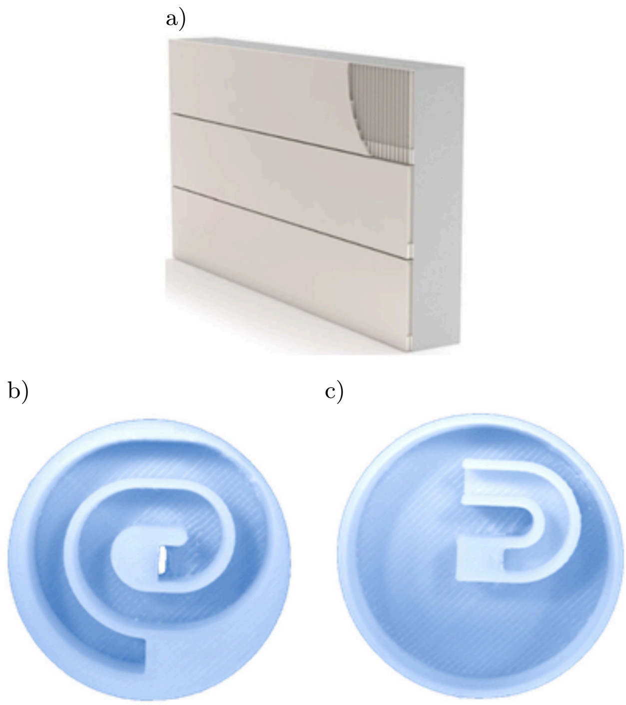

Fig. 2. Schemes of the ultra-thin metamaterial in existing literature: (a) Helmholtz resonator arrays (Jiménez *et al.*, 2016), circular absorbers with embedded (b) coplanar spiral tube, and (c) coplanar resonant chamber (Cai *et al.*, 2014).

Combining the data from Table 1, Table 2, and Fig. 3, several findings can be derived:

a) MSPs have the ability to achieve a maximum absorption value comparable to that of ultrathin metamaterials while maintaining a wider absorption bandwidth. Additionally, the panel-thickness optimal scheme-based MSP absorbers perform even better than the cavity-depth optimal scheme-based MSP absorbers in terms of sound absorption bandwidth. This is attributed to the fact that the bandwidth is proportional to the ratio of acoustic resistance to acoustic mass $\frac{\mathrm{real}(Z_{\mathrm{MSP}})}{\mathrm{imag}(Z_{\mathrm{MSP}})}$, in which real (Z_{MSP}) denotes the real part of the acoustic impedance of the MSP, i.e., acoustic resistance, and imag (Z_{MSP}) represents the imaginary part, i.e., acoustic mass. According to Table 1, the panel thickness optimal scheme-based MSP has a smaller panel thickness and slit width, resulting in a higher ratio of acoustic resistance to acoustic mass, thus yielding a better bandwidth (Maa, 2000; Randeberg, 2000).

b) At the resonant frequency f_r = 391 Hz, the total thickness of the MSP3-COS is almost equal to that of the metamaterial. In other conditions, the total thickness of the MSP absorber is about 1.5 to 2.7 times that of the metamaterial. Meanwhile, the total thickness of the panel thickness optimal scheme-based MSP absorbers is greater than that of the cavity-depth optimal scheme-based MSP absorbers.

c) For cavity-depth optimal scheme-based MSP absorbers, in addition to the cavity depth, the panel thickness significantly contributes to the overall thickness of the absorber.

Table 2. Performance and structural parameters of the ultrathin metamaterials.

Members	Metamaterial	f_r [Hz]	a_r	T
Meta-1	Circular absorber with embedded coplanar resonant chamber (Cai *et al.*, 2014)	254	≥ 0.98	13.3
Meta-2	Helmholtz resonator arrays (Jiménez *et al.*, 2016)	338.5	≥ 0.98	11
Meta-3	Circular absorber with embedded coplanar spiral tube (Cai *et al.*, 2014)	391	≥ 0.97	17

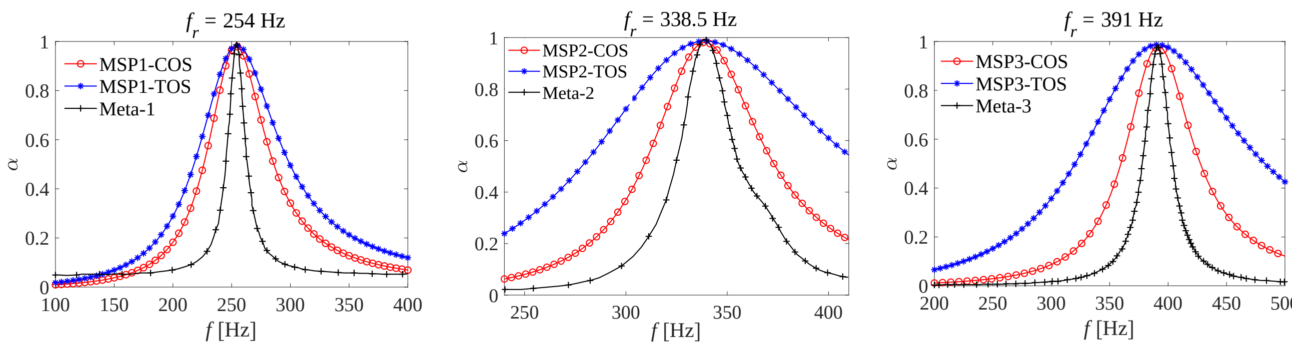

Fig. 3. Comparison of the theoretical normal absorption coefficients between MSP and ultra-thin metamaterials in existing literature.

From the above conclusions, it can be concluded that MSP absorbers exhibit great potential for low-frequency sound absorption in both maximum absorption coefficient and absorption bandwidth. Specifically, the cavity-depth optimal scheme-based MSP absorbers show promise in achieving a balance between better sound absorption performance and a smaller total structure thickness. But, their comparatively larger panel thickness may contribute to an increase in manufacturing cost and structure weight.

3. Optimized panel thickness by using coupled micro-slit panels (CMSP)

3.1. CMSP absorber and theoretical model of its acoustic impedance

In order to further optimize the panel thickness, an absorber based on CMSP is proposed. The CMSP absorber consists of two MSPs with a distance between them being less than 1 mm, fixed before a solid surface with a cavity of depth D, as depicted in Fig. 4a. Here, d_g represents the gap thickness. The micro-slits on the top and bottom panels need to be staggered so that the airflow from the micro-slits of the top panel flows equally into the two adjacent micro-slits of the bottom panel. As shown in Fig. 4b, the micro air gap between MSP1 and MSP2 increases the length of the airflow path, thus can be equivalently regarded as increasing the thickness of the panel. Based on the acoustic-electric analogy, the equivalent circuit of the CMSP can be derived, as shown in Fig. 4c, where the MSP1 is coupled with MSP2 through the acoustic impedance of the micro-slits between the two panels.

It can be seen in Fig. 4b that the role of the air gap between two MSPs is the same as that of the MSP1 and MSP2. Based on the equivalent circuit, the total acoustic impedance of the entire structure can be expressed as

$$Z_{\mathrm{CMSP}} = Z_{\mathrm{MSP1}} + Z_{\mathrm{MSP1\_2}}/2 + Z_{\mathrm{MSP2}}, \tag{9}$$

where Z_{MSP1}, $Z_{\mathrm{MSP1\_2}}$ and Z_{MSP2} represents the acoustic impedance of MSP1, micro-slits between two MSPs, and MSP2, respectively. It is important to note that the micro-slit width, micro-slit thickness and micro-slit rate of the micro-slits between two MSPs are d_g, $b/2 - d$, and d_g/b, respectively. When the parameters of MSP1 and MSP2 are determined, the acoustic impedance of the CMSP can therefore be adjusted by the micro-slit width d_g alone. The normal sound absorption coefficient can also be calculated according to Eq. (7).

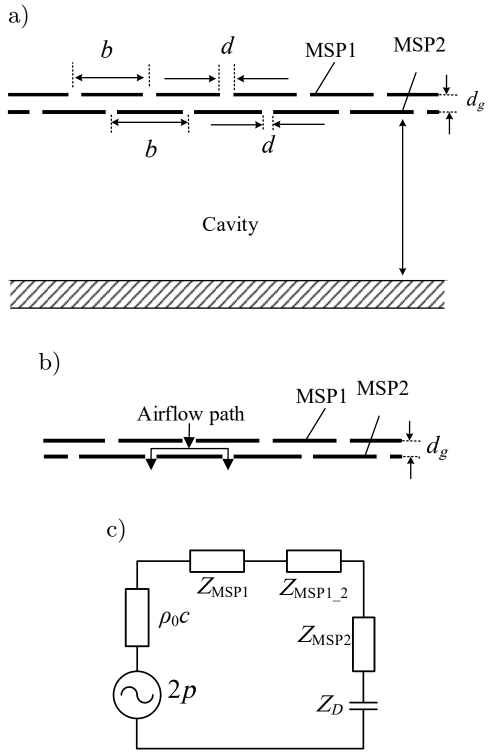

Fig. 4. Schematic diagrams of the CMSP structure and its equivalent circuit.

3.2. Experimental validation

In this section, the experiment is conducted to verify the theoretical results of CMSP. An impedance tube with diameter of 10 cm is used for experimental test, as shown in Fig. 5, with a working frequency range of 90~1800 Hz. The measured frequencies are 1/3 octave center frequencies from 100 to 1600 Hz. The test sample is made from epoxy resign. In order to form a CMSP absorber, a special design of the experimental sample is required in the design phase, and the design diagrams of the top MSP and bottom MSP are shown

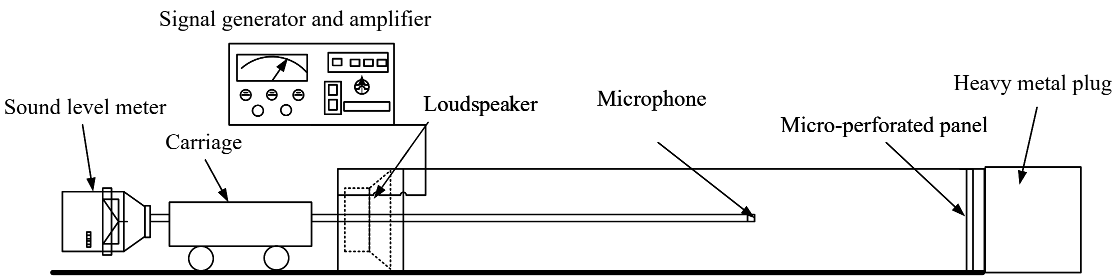

Fig. 5. Layout of the impedance tube.

in Fig. 6. The structural parameters of the samples are shown in Table 3, where the subscripts 1 and 2 denote MSP1 and MSP2, respectively. The experimental samples of CMSP#3 are presented in Fig. 7. The comparison between experiments and theoretical results is depicted in Fig. 8. It is evident from Fig. 8 that the theoretical prediction agrees well with the experimental data, which proves that the theoretical model is reliable.

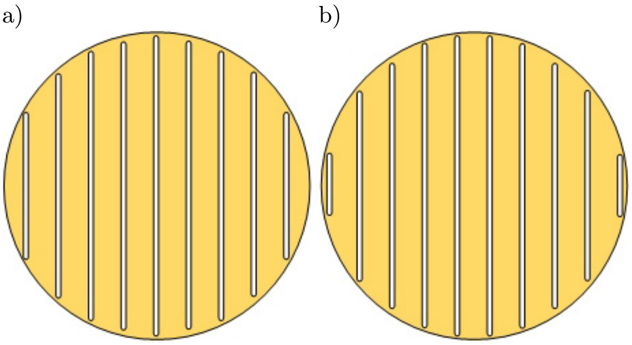

Fig. 6. Design schematic of MSPs: a) top MSP1; b) bottom MSP2.

Table 3. Structure parameters of CMSP absorbers for experiments.

Member	D	$t_{1,2}$	$d_{1,2}$	d_g	b
#1	30	1	0.3	0.2	6
#2	30	1.5	0.5	0.2	8
#3	30	1	0.6	0.3	8
#4	30	1.5	0.6	0.3	9

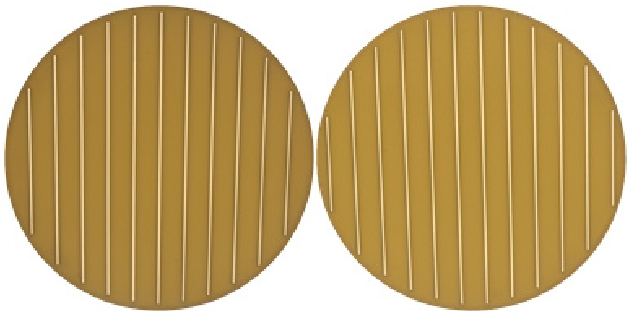

Fig. 7. Experimental samples of CMSP#3 for comparison.

3.3. Panel thickness optimization

The parameter range for the exhaustive search for the CMSP absorber is defined as:

$$0.1\ \text{mm} \le d \le 1\ \text{mm};\quad 0.1\ \text{mm} \le t \le 1\ \text{mm};$$
$$0.3\ \text{mm} \le b \le 170\ \text{mm};\quad 0.1\ \text{mm} \le d_g \le 1\ \text{mm};\quad (10)$$
$$5\ \text{mm} \le D \le 30\ \text{mm}.$$

The selection of the optimal combination of parameters for a CMSP absorber follows the cavity depth optimal scheme (COS). The optimal combinations of parameters selected for the CMSP absorber are shown in Table 4. It shows that the optimal panel thickness can be reduced significantly by the CMSP, with the panel thickness reaching the lower limit of the parameter range as thin as 0.1 mm. Although a panel thickness of 0.1 mm may be impractical in practical applications, it foreshadows the great potential of

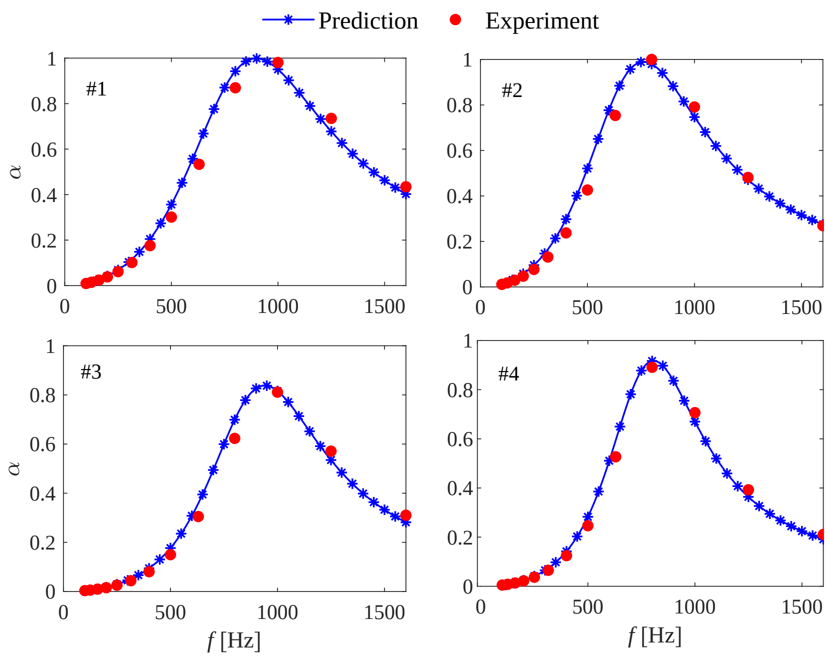

Fig. 8. Comparison between experimental data and theoretical prediction of normal sound absorption coefficient for CMSP absorbers.

Table 4. Theoretically optimal combination of parameters for CMSP absorbers.

Member	f_r	a_r	D	$t_{1,2}$	$d_{1,2}$	d_g	$b_{1,2}$	T
CMSP1	254	≥ 0.98	24.1	0.1	0.94	0.99	77	25.29
CMSP2	338.5	≥ 0.98	15.1	0.1	0.83	0.99	73	16.29
CMSP3	391	≥ 0.98	12	0.1	0.81	0.99	71	12.19

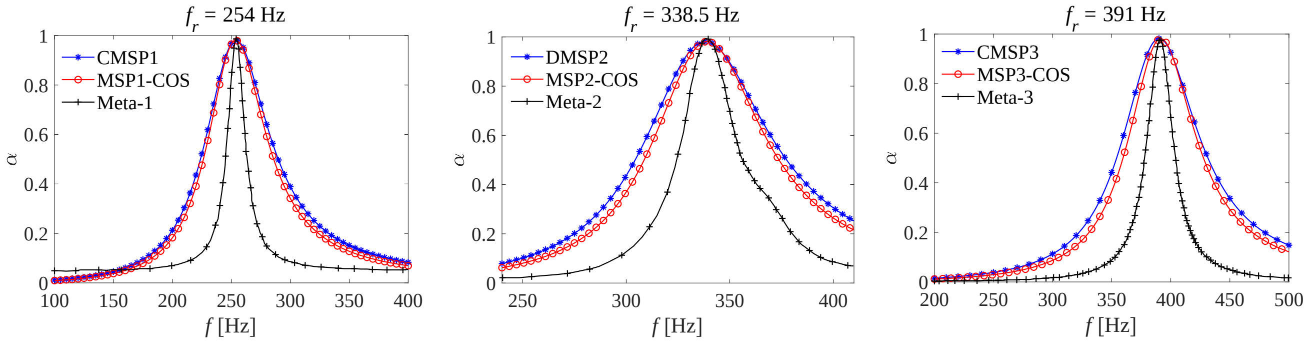

Fig. 9. Comparison of normal absorption coefficients between CMSP, MSP, and ultra-thin metamaterials.

CMSPs in reducing the panel thickness. In practical applications, the structural parameter range for optimization can be flexibly modified based on actual processing conditions. The comparison results of normal absorption coefficients between CMSP absorbers, COS-based MSP absorbers and metamaterials are illustrated in Fig. 9. As can be seen in Table 4 and Fig. 9, for different resonance frequency requirements (254, 338.5, and 391 Hz), although the overall thickness of CMSP absorbers decreases by 16, 22, and 27%, and the panel thickness is reduced by 98.8, 98.6, and 98.5%, respectively, compared to single MSP-COS absorbers, CMSP absorbers still outperform both MSP-COS absorbers and metamaterials in terms of sound absorption bandwidth. In particular, at f_r = 391 Hz, the CMSP achieves better sound absorption performance with less total thickness compared to Meta-3. Moreover, at resonant frequencies of 254, 338.5, and 391 Hz, the total thickness of the CSMPs is 0.187, 0.16, and 0.138 times the resonant wavelength, respectively, demonstarting sub-wavelength dimensions.

4. Conclusion

A subwavelength absorber based on CMSP is proposed and investigated to achieve high absorption in the low-frequency range at a smaller thickness. Firstly, COS and TOS, respectively, are employed to maximize the low-frequency absorption potential of a MSP), demonstrating that COS enables MSP to achieve a smaller total thickness but a higher panel thickness. The CSMPs' acoustic impedance theoretical model is subsequently developed and experimentally verified. Based on the theoretical model, COS is used to optimize the CSMP's low-frequency absorption capability in order to further reduce panel thickness. It is shown that the CSMPs can significantly reduce the panel thickness while maintaining relatively better sound absorption properties compared to both MSP and metamaterials in existing literature. Notably, at a resonance frequency of 391 Hz, the total thickness of the CMSP can reach subwavelength dimension of 0.138λ (λ denotes the resonance wavelength).

Acknowledgments

This work was supported by Aviation Science Foundation (No. 2023Z074108001).

References

1. Allard J.F., Atalla N. (2009), *Propagation of Sound in Porous Media: Modelling Sound Absorbing Materials*, Wiley, Chichester, doi: 10.1002/9780470747339.
2. Almeida G.N., Erasmo F.V., Barbosa L.R., Lenzi A., Birch R.S. (2021a), Sound absorption metasurface with symmetrical coiled spaces and micro slit of variable depth, *Applied Acoustics*, **183**: 108312, doi: 10.1016/j.apacoust.2021.108312.
3. Almeida G.N., Vergara E.F., Barbosa L.R., Brum R. (2021b), Low-frequency sound absorption of a metamaterial with symmetrical-coiled-up spaces, *Applied Acoustics*, **172**: 107593, doi: 10.1016/j.apacoust.2020.107593.
4. Cai X.B., Guo Q.Q., Hu G.K., Yang J. (2014), Ultrathin low-frequency sound absorbing panels based on coplanar spiral tubes or coplanar Helmholtz resonator, *Applied Physics Letters*, **105**: 121901, doi: 10.1063/1.4895617.
5. Cheng Y., Zhou C., Yuan B.G., Wu D.J., Wei Q., Liu X.J. (2015), Ultra-sparse metasurface for high reflection of low-frequency sound based on artificial Mie resonances, *Nature Mater*, **14**: 1013–1019, doi: 10.1038/nmat4393.
6. Chong Y., Ge L., Cao H., Stone A.D. (2010), Coherent perfect absorbers: Time-reversed lasers, *Physi-*

cal Review Letters, **105**(5): 053901, doi: 10.1103/PhysRevLett.105.053901.

7. Donda K., Zhu Y., Fan S.-W., Cao L., Li Y., Assouar B. (2019), Extreme low-frequency ultrathin acoustic absorbing metasurface, *Applied Physics Letter*, **115**(17): 173506, doi: 10.1063/1.5122704.
8. Jiménez N., Romero-García V., Pagneux V., Groby J.-P. (2017a), Quasi-perfect absorption by subwavelength acoustic panels in transmission using accumulation of resonances due to slow sound, *Physical Review Journals*, **957**: 014205, doi: 10.1103/PhysRevB.95.014205.
9. Jiménez N., Romero-García V., Pagneux V., Groby J.-P. (2017b), Rainbow-trapping absorbers: Broadband, perfect and asymmetric sound absorption by subwavelength panels for transmission problems, *Scientific Reports*, **7**: 13595, doi: 10.1038/s41598-017-13706-4.
10. Jiménez N., Huang W., Romero-García V., Pagneux V., Groby J.-P. (2016), Ultra-thin metamaterial for perfect and quasi-omnidirectional sound absorption, *Applied Physics Letters*, **109**: 121902, doi: 10.1063/1.4962328.
11. Lara-Valencia L.A., Farbiarz-Farbiarz Y., Valencia-González Y. (2020), Design of a tuned mass damper inerter (TMDI) based on an exhaustive search optimization for structural control of buildings under seismic excitations, *Shock and Vibration*, **2020**: 8875268, doi: 10.1155/2020/8875268.
12. Li Y., Assouar B.M. (2016), Acoustic metasurface-based perfect absorber with deep subwavelength thickness, *Applied Physics Letter*, **108**(6): 063502, doi: 10.1063/1.4941338.
13. Liang Z.X., Li J. (2012), Extreme acoustic metamaterial by coiling up space, *Physical Review Journals*, **108**: 114301, doi: 10.1103/PhysRevLett.108.114301.
14. Liu C.M., Xia B.Z., Yu D.J. (2017), The spiral-labyrinthine acoustic metamaterial by coiling up space, *Physics Letters A*, **381**(36): 3112–3118, doi: 10.1016/j.physleta.2017.07.041.
15. Liu X., Wang C.Q., Zhang Y.M., Huang L.X. (2021a), Investigation of broadband sound absorption of smart micro-perforated panel (MPP) absorber, *International Journal of Mechanical Sciences*, **199**: 106426, doi: 10.1016/j.ijmecsci.2021.106426.
16. Liu Y.Y., Ren S.W., Sun W., Lei Y., Wang H.T., Zeng X.Y. (2021b), Broadband low-frequency sound absorbing metastructures based on impedance matching coiled-up cavity, *Applied Physics Letter*, **119**(10): 101901, doi: 10.1063/5.0061012.
17. Ma G., Sheng P. (2016), Acoustic metamaterials: From local resonances to broad horizons, *Science Advances*, **2**(2): e1501595, doi: 10.1126/sciadv.1501595.
18. Maa D.Y. (1998), Potential of microperforated panel absorber, *The Journal of the Acoustical Society of America*, **104**(5): 2861–2866, doi: 10.1121/1.423870.
19. Maa D.Y. (2000), Theory of microslit absorbers [in Chinese], *Acta Acustica*, **25**(6): 481–485, doi: 10.15949/j.cnki.0371-0025.2000.06.001.
20. Mei J., Ma G., Yang M. (2012), Dark acoustic metamaterials as super absorbers for low-frequency sound, *Nature Communications*, **3**: 756, doi: 10.1038/ncomms1758.
21. Park S.H. (2013), Acoustic properties of micro-perforated panel absorbers backed by Helmholtz resonators for the improvement of low-frequency sound absorption, *Journal of Sound and Vibration*, **332**(20): 4895–4911, doi: 10.1016/j.jsv.2013.04.029.
22. Pierro V. *et al.* (2021), Ternary quarter wavelength coatings for gravitational wave detector mirrors: Design optimization via exhaustive search, *Physics Review Research*, **3**(2): 023172, doi: 10.1103/PhysRevResearch.3.023172.
23. Randeberg R.T. (2000), *Perforated panel absorbers with viscous energy dissipation enhanced by orifice design*, Ph.D. Thesis, Norwegian University of Science and Technology, Trondheim.
24. Ryoo H., Jeon W. (2018), Perfect sound absorption of ultra-thin metasurface based on hybrid resonance and space-coiling, *Applied Physics Letter*, **113**(12): 121903, doi: 10.1063/1.5049696.
25. Shen Y., Yang Y., Guo X., Shen Y., Zhang D. (2019), Low-frequency anechoic metasurface based on coiled channel of gradient cross-section, *Applied Physics Letter*, **114**(8): 083501, doi: 10.1063/1.5081926.
26. Vigran T.E. (2014), The acoustic properties of panels with rectangular apertures, *The Journal of the Acoustical Society of America*, **135**(5): 2777–2784, doi: 10.1121/1.4871363.
27. Wang Y., Zhao H.G., Yang H.B., Zhong J., Zhao D., Lu Z.L., Wen J.H. (2018), A tunable sound-absorbing metamaterial based on coiled-up space, *Journal of Applied Physics*, **123**(18): 185109, doi: 10.1063/1.5026022.
28. Wu F., Xiao Y., Yu D.L., Zhao H.G., Wang Y., Wen J.H. (2019), Low-frequency sound absorption of hybrid absorber based on micro-perforated panel and coiled-up channels, *Applied Physics Letters*, **114**: 151901, doi: 10.1063/1.5090355.
29. Wu Y., Liang Q., He J., Feng J., Chen T. (2021), Deep-subwavelength broadband sound absorbing metasurface based on the update finger coiling-up method, *Applied Acoustics*, **195**: 108846, doi: 10.1016/j.apacoust.2022.108846.
30. Xu Z.M., He W., Peng X.G., Xin F.X., Lu T.J. (2020), Sound absorption theory for micro-perforated panel with petal-shaped perforations, *The Journal of the Acoustical Society of America*, **148**(18): 18–24, doi: 10.1121/10.0001462.
31. Zhao H., Wang Y., Wen J., Lam Y.W., Umnova O. (2018), A slim subwavelength absorber based on coupled microslits, *Applied Acoustics*, **142**: 11–17, doi: 10.1016/j.apacoust.2018.08.004.
32. Zhu Y.F., Donda K., Fan S.W., Cao L.Y., Assouar B. (2019), Broadband ultra-thin acoustic metasurface absorber with coiled structure, *Applied Physics Express*, **12**: 114002, doi: 10.7567/1882-0786/ab494a.

Archives of Acoustics Vol. **49**, No. 1, pp. 129–140 (2024), doi: 10.24425/aoa.2024.148771

Technical Note

Performance of the Direct Sequence Spread Spectrum Underwater Acoustic Communication System with Differential Detection in Strong Multipath Propagation Conditions

Jan H. SCHMIDT, Iwona KOCHAŃSKA*, Aleksander M. SCHMIDT

Faculty of Electronics, Telecommunication and Informatics, Department of Signals and Systems
Gdańsk University of Technology
Gdańsk, Poland

*Corresponding Author e-mail: iwona.kochanska@pg.edu.pl

(*received October 24, 2020; accepted April 2, 2020; published online January 8, 2024*)

The underwater acoustic communication (UAC) operating in very shallow-water should ensure reliable transmission in conditions of strong multipath propagation, significantly disturbing the received signal. One of the techniques to achieve this goal is the direct sequence spread spectrum (DSSS) technique, which consists in binary phase shift keying (BPSK) according to a pseudo-random spreading sequence.

This paper describes the DSSS data transmission tests in the simulation and experimental environment, using different types of pseudo-noise sequences: m-sequences and Kasami codes of the order 6 and 8. The transmitted signals are of different bandwidth and the detection at the receiver side was performed using two detection methods: non-differential and differential.

The performed experiments allowed to draw important conclusions for the designing of a physical layer of the shallow-water UAC system. Both, m-sequences and Kasami codes allow to achieve a similar bit error rate, which at best was less than 10^{-3}. At the same time, the 6th order sequences are not long enough to achieve an acceptable BER under strong multipath conditions. In the case of transmission of wideband signals the differential detection algorithm allows to achieve a significantly better BER (less than 10^{-2}) than non-differential one (BER not less than 10^{-1}). In the case of narrowband signals the simulation tests have shown that the non-differential algorithm gives a better BER, but experimental tests under conditions of strong multipath propagation did not confirm it. The differential algorithm allowed to achieve a BER less than 10^{-2} in experimental tests, while the second algorithm allowed to obtain, at best, a BER less than 10^{-1}. In addition, two indicators have been proposed for a rough assessment which of the detection algorithms under current propagation conditions in the channel will allow to obtain a better BER.

Keywords: direct sequence spread spectrum; DSSS, m-sequences; Kasami codes; shallow-water channel; multipath propagation; underwater acoustic communications; UAC.

1. Introduction

The underwater acoustic communication (UAC) system should efficiently use the UAC channel bandwidth to perform the reliable data transmission. To achieve this goal spread spectrum techniques can be used in the physical layer of data transmission (Schmidt, 2020). One of them is the direct sequence spread spectrum (DSSS) technique. Its advantage is the fact that the knowledge of the pseudo-random sequence on the receiving side is required for correct receipt of information. Keeping this sequence secret may be part of the protection against unauthorised access to transmitted information. Pseudorandom spreading sequences used in DSSS systems should be characterized by a normalized autocorrelation function as close as possible to the Kronecker delta. For this reason, in telecommunications systems with the spread spectrum technique, the maximum length sequences (m-sequences) are used (Zepernick, Finger, 2005). In systems using the code multiple access technique, the spreading sequences used should also have as little cross-correlation as possible. Both these requirements are met by Kasami sequences (Sarwate, Pursley, 1980).

Descriptions of numerous implementations of the DSSS technique in underwater telecommunications systems can be found in the literature. Some of them, similar to radiocommunication systems, implement the RAKE algorithm in the receiver. However, this is not the only way to receive DSSS signals in UAC systems. Many of them implement matched filtration instead of the RAKE technique (Freitag, Strojanovic, 2004; Freitag *et al.*, 2001; Mironov *et al.*, 2018). Among these systems, few operate in shallow waters (Pelekanakis, Cazzanti, 2018; Qu *et al.*, 2018; Ra *et al.*, 2021; Freitag *et al.*, 2001; Sozer *et al.*, 1999). These are systems operating in different frequency bands, using m-sequences, Kasami sequences or Gold sequences – also of different ranks. Moreover, each of these systems was tested in different conditions.

In (Kochanska *et al.*, 2021) the results of DSSS signal transmission experiments aimed at comparing the possible data transmission rate with the use of different pseudorandom spreading sequences and depending on the bandwidth of the UAC system were described. A filtering method has been implemented in the receiver to detect information. The analysis of the results has shown that in conditions of strong multipath propagation, the BER of the transmission of signals of narrower bandwidth is better, than in the case of wider transmission bandwidth. Moreover, using m-sequences allowed to obtain better BER than the use of other PN sequences.

In (Kochańska, 2021) a new algorithm of differential detection of the DSSS signal, constructed of m-sequences of rank 8, has been described. The simulation and experimental tests have shown that this detection technique allows to obtain better values of BER than the algorithm applied in the system using this particular DSSS signals, described in (Kochanska *et al.*, 2021), but only for the transmission of wideband signals.

As mentioned in (Kochanska *et al.*, 2021), the DSSS-based UAC systems described in the literature use specific PN sequences and the fixed transmission bandwidth. To the best of the authors' knowledge, there are no publications presenting an analysis of the underwater DSSS system's performance depending on its bandwidth or the PN sequence used.

In this paper a comparison of the performance of two detection techniques is presented, applied in a laboratory model of the UAC system using a wide set of DSSS signals: constructed using m-sequences and Kasami codes of rank 6 and 8, occupying five different bandwidths from 1 to 5 kHz. The shallow-water channel was simulated using the Watermark simulator (van Walree, 2011), and the experimental tests were performed in the model pool.

2. Structure of communication signal

The process of the DSS signal generation is shown in Fig. 1. The input data stream $d[n]$ is converted into binary phase-shift keying (BPSK) symbols which are multiplied by the PN sequence $m[n]$ and the result is upsampled by the factor $R = \frac{f_s}{B}$ equal to the quotient of sampling frequency f_s (which is 200 kHz) and the system bandwidth B. The upsampled signal $x[n]$ is then used for phase modulation of the digital carrier waveform of the frequency f_c.

Two types of PN sequences are used for DSSS signal generation, namely m-sequences and Kasami codes of rank 6 and 8. Its autocorrelation functions are shown in Figs. 2 and 3.

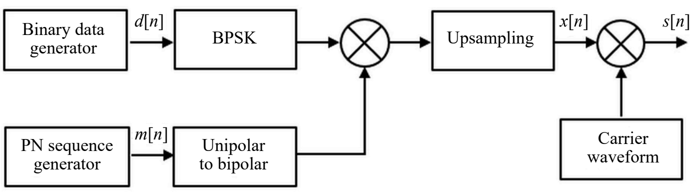

Fig. 1. Generation of the DSSS signal.

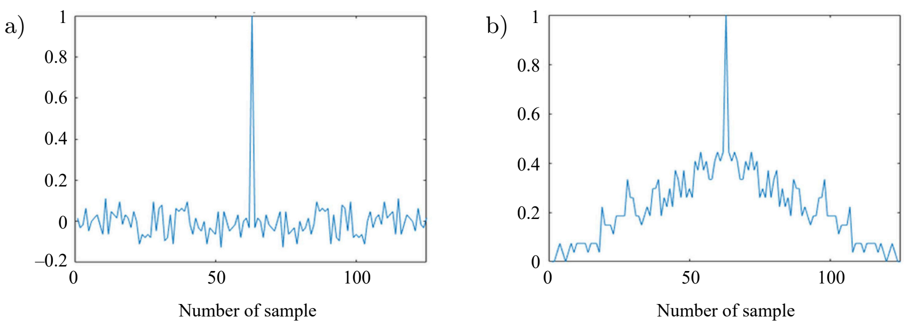

Fig. 2. Autocorrelation functions of PN sequences of rank 6: m-sequence (a) and Kasami code (b).

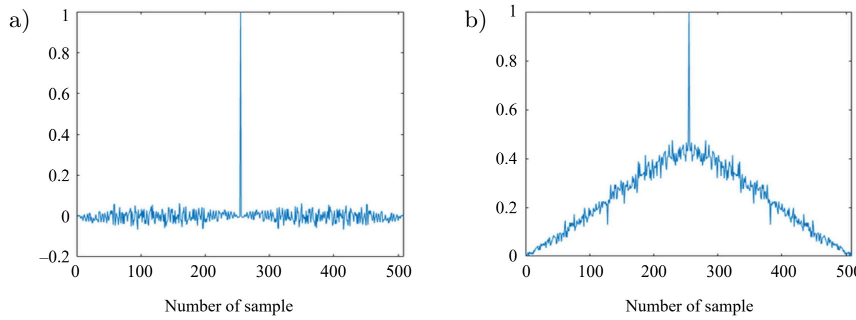

Fig. 3. Autocorrelation functions of PN sequences of rank 8: m-sequence (a) and Kasami code (b).

Table 1. DSSS signal parameters.

Bandwidth B [kHz]	Upsampling factor R	PN sequences rank	Symbol length N_s	Symbol duration T_s [ms]	Transmission rate [bps]
1	200	6	12600	63.00	15.87
1	200	8	51000	255.00	3.92
2	100	6	6300	31.50	31.75
2	100	8	25500	127.50	7.84
4	50	6	3150	15.75	63.49
4	50	8	12750	63.75	15.69
5	40	6	2520	12.60	79.37
5	40	8	10200	51.00	19.61

The generated signals occupy different bandwidths B, as shown in Table 1. The bandwidth B determines the upsampling factor R. The upsampling factor and the number of samples of PN sequence (which depends on its rank) determine the length of a single DSSS symbol N_s, as well as its duration T_s for a given sampling frequency $f_c = 200$ kHz The transmission rate has been calculated as a number of symbols per second.

3. The DSSS demodulator

At the receiver side (Fig. 4), the baseband equivalent signal $y_b[n]$ is calculated based on $y[n]$ and filtered by a matched filter described by complex-value coefficients $m_c[n] = m[n] + j\widehat{m}[n]$, where $m[n]$ is the PN sequence used for construction of a transmitted signal, and $\widehat{m}[n]$ is equal to the Hilbert transform of $m[n]$. Next the output of the matched filter is processed by the detection algorithms.

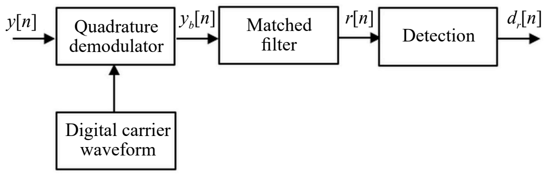

Fig. 4. Digital processing path of DSSS signal at the receiver side.

3.1. Detection algorithm A

The differential algorithm A has been described previously in (Kochańska, 2021). It is based on the assumption that the influence of the propagation conditions on the transmitted adjacent modulation symbols is similar. Therefore, in quasi-stationary conditions the response of the matched filter to subsequent modulation symbols is similar. If adjacent modulation symbols carry different information, then the modulus of the matched filter responses for these symbols is almost the same, while the phase will be opposite.

The output $r[n]$ of the matched filter is divided to segments of the length N_s (Table 1). For every two subsequent segments r_{k-1} and r_k of the signal $r[n]$ the Pearson correlation coefficient $C[k]$ of its arguments is calculated as:

$$C[k] = \frac{1}{L-1} \sum_{i=1}^{L} \frac{(\arg r_k[i] - \mu_k)}{\sigma_k} \frac{(\arg r_{k-1}[i] - \mu_{k-1})}{\sigma_{k-1}}, \tag{1}$$

where μ_k and μ_{k-1} are the mean values, and σ_k and σ_{k-1} are standard deviations of arguments of segments r_k and r_{k-1}, respectively.

If the value of the correlation coefficient $C[k]$ is positive or equal to 0, it indicates that a given segment of $r[n]$ represents the same information as the previous one. A negative value of $C[k]$ means that the information bit is the negation of the previous one. It can be expressed as:

$$d_r[k] = \begin{cases} d_r[k-1] & \text{if} \quad C[k] \ge 0, \\ \sim d_r[k-1] & \text{if} \quad C[k] < 0, \end{cases} \tag{2}$$

where k is the index of information bit number, corresponding to a given segment of $r[n]$.

The construction of this algorithm is such that its performance should increase when the similarity between amplitudes of consecutive DSSS symbols at the matched filter output increases. With a large similarity of amplitude, the only significant difference is the phase of these symbols, which is the transmitted information. To confirm this thesis for all received symbols a mean value C_A of the absolute values of the correlation coefficient $C[k]$ has been calculated:

$$C_A = \frac{1}{K}\sum_{k=1}^{K}|C[k]|, \tag{3}$$

which can be related to BER achieved with the algorithm A during simulation and experimental tests.

3.2. Detection algorithm B

The second detection algorithm B was applied during the simulation and experimental tests described in (Kochanska *et al.*, 2021). It compares the minimum $A_{\min}[k]$ and maximum $A_{\max}[k]$ values of the real part of the matched filter output $r[n]$. If the maximum value $A_{\max}[k]$ is greater than the absolute value $A_{\min}[k]$ then the modulation symbol carries a bit information equal to 1, otherwise the information is equal to 0. It can be denoted as:

$$d_r[k] = \begin{cases} 1 & \text{if} \quad A_{\max}[k] \geq |A_{\min}[k]|, \\ 0 & \text{if} \quad A_{\max}[k] < |A_{\min}[k]|. \end{cases} \tag{4}$$

The algorithm can perform effectively if the communication channel is characterized by a stable dominant propagation path well separated from the other paths. Then, also in the waveform at the output of the matched filter, a stable, dominant extreme "peak" will be observed, which is detected by the algorithm B as the $A_{\max}$ or $A_{\min}$ value. In order to make it possible to confirm this thesis, the coefficient C_B was defined:

$$C_B = \frac{C_{\text{ex}}}{C_{\text{all}}}, \tag{5}$$

where C_{ex} is the number of received DSSS symbols, in which at the output of the matched filter the extreme appears for the most frequant delay among all symbols, and C_{all} is the number of all received symbols. It is expected, that the higher C_{ex} to C_{all} ratio is, the better BER should be obtained with the use of the algorithm B.

4. Simulation tests

The performance of the DSSS-based UAC system has been tested using the Watermark simulator for MATLAB software environment. It is a replay channel using at-sea measurements of time-varying impulse responses of UAC channels (van Walree, 2011), valued in the hydroacoustic environment. Three channels available at Watermark and representing different propagation conditions were selected, namely: Norway-Oslofjord (NOF1), Norway-Continental Shelf (NCS1), and Brest Commercial Harbor (BCH1).

NOF1 is a channel measured in a shallow stretch of Oslofjorden between a stationary source and a stationary single-hydrophone receiver. It represents a relatively smooth communication channel. The first arrival path, as shown in Fig. 5 presenting the scattering function of the channel, has no frequency spread, whereas later arrivals are Doppler spreads due to sea surface interactions. Most energy of the received signal is concentrated in a narrow delay-Doppler window.

NCS1 was measured similarly as NOF1, between a stationary source and a stationary single-hydrophone receiver. The measurements were conducted on Norway's continental shelf. As shown in Fig. 6, most energy is also concentrated at the start of the impulse response, but considering the next arrival paths, the differences from NOF1 are significant. There are no stable paths, thus it is more challenging than NOF1, in particular for coherent communication schemes such as DSSS, which need to detect the phase of a signal.

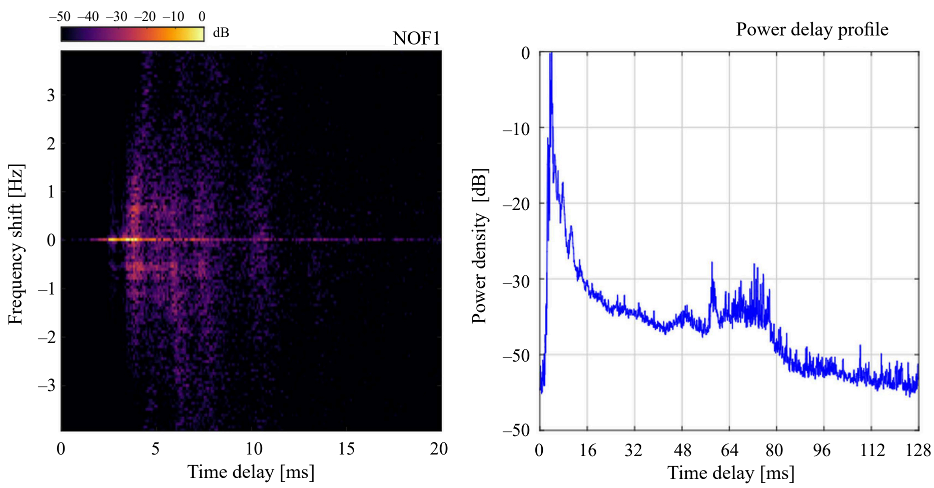

Fig. 5. Scattering function and power delay profile of NOF1 channel (van Walree, 2011).

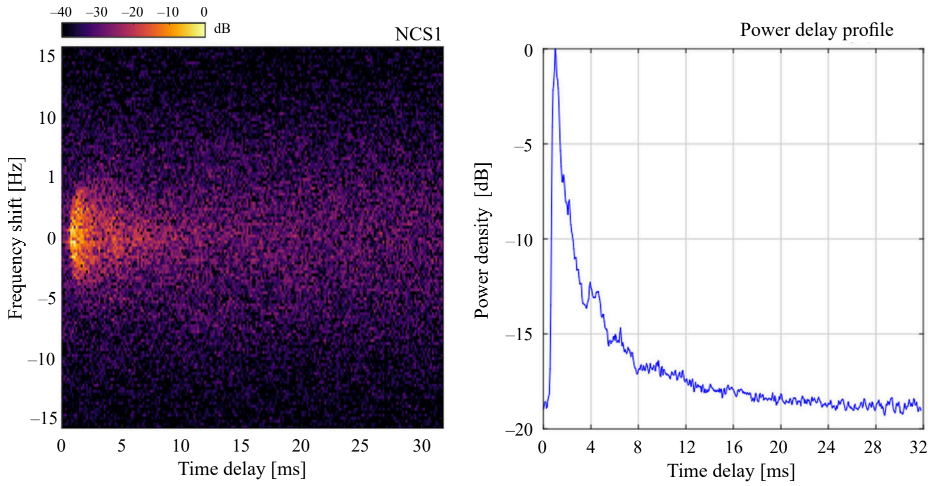

Fig. 6. Scattering function (a) and power delay profile (b) of NCS1 channel (van Walree, 2011).

BCH1 channel was measured in the harbour of Brest, France. A source and a receiver were not stationary mounted at the bottom, as in case of NOF1 and NCS1, but lowered into the water column from two docks. Similarly to NOF1, the channel is a mixture of stable and fluctuating arrivals, but with a larger number of distinct trailing paths (Fig. 7) (van Walree, 2011).

These three channels, representing different propagation conditions, were used to simulate the DSSS signal transmission.

During the simulation tests, different carrier frequencies were used to fit the DSSS signal to the frequency band of a given Watermark channel. It was equal to 14 kHz in case of NOF1 and NCS1 channels and 35 kHz in case of BCH1 channel.

As described in Sec. 2, 16 kinds of DSSS signals were transmitted in each of Watermark channels. They differed in kind of the PN spreading sequence, its length, and the bandwidth. Using each of the three Watermark channels, 180 transmission tests were performed using every DSSS signal of time duration equal to 6 s. Every DSSS signal was carrying different information bits. The detection of information was performed using the detection algorithm A and B.

Figure 8 shows the exemplary real part of the filter output of the matched DSSS receiver in the case of receiving a DSSS signal (constructed of m-sequence of rank 8) with a bandwidth of 1 kHz, representing the "0" bit value and the "1" bit value on the NOF1 channel. The extreme of the waveform occurs in the same place, with a different sign for the "0" and "1" bits. Figure 9 shows the real part of the response of the matched filter to a signal with a bandwidth of 4 kHz, carrying the value of the "0" bit and the value of the "1" bit. It can be seen that in case of wideband signal transmitted in stationary NOF1 channel the dominant path might not occur in the output waveform of the filter with the same delay.

The results of a bit error rate (BER) achieved in all 180 transmission tests of each DSSS signal are shown in Figs. 10, 12, 14, and 16. In Figs. 11, 13, 15, and 17 the coefficients C_A and C_B also calculated for all transmission tests are presented.

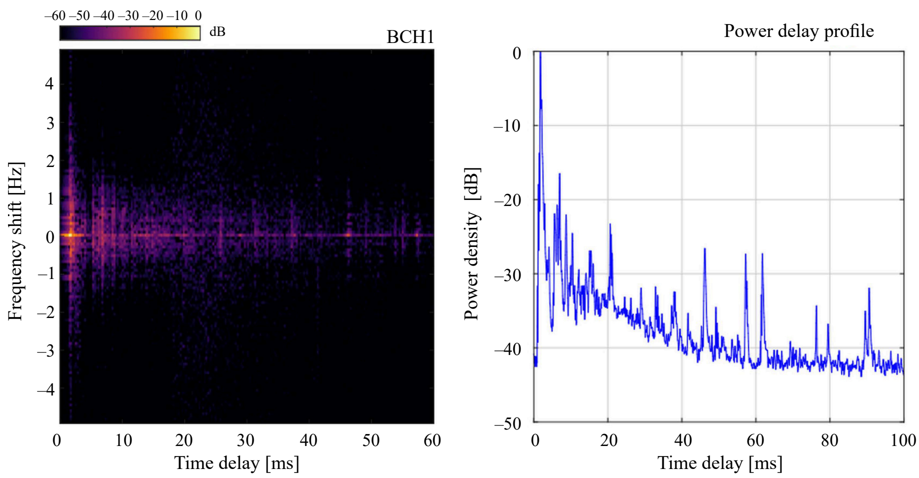

Fig. 7. Scattering function and power delay profile of BCH1 channel (van Walree, 2011).

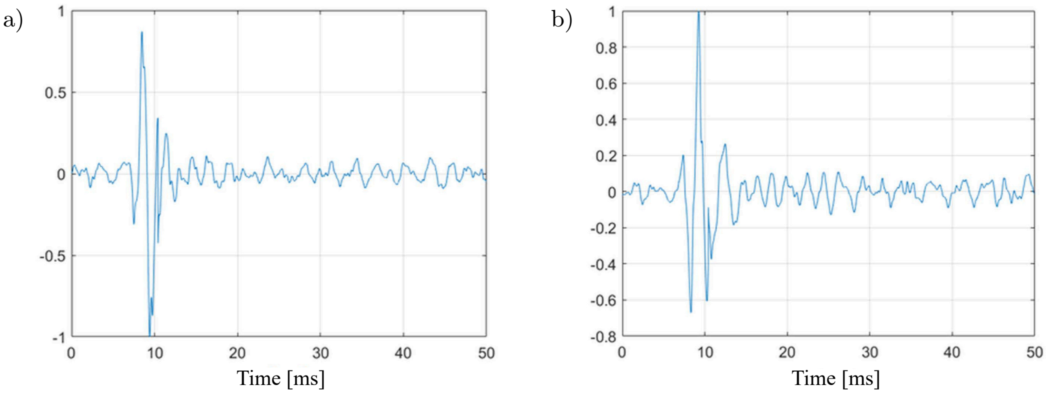

Fig. 8. Real part of the response of matched filter to DSSS signal constructed of m-sequence of rank 8 and bandwidth equal to 1 kHz, received in NOF1 channel, representing the "0" bit value (a) and the "1" bit value (b).

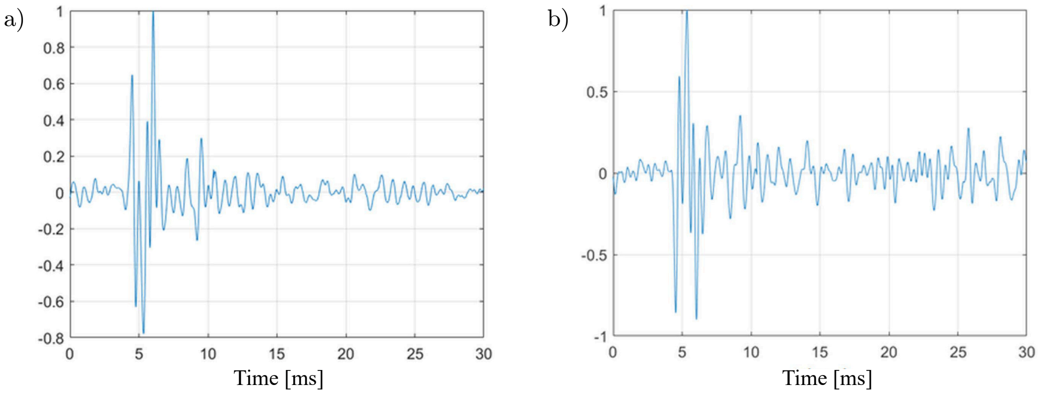

Fig. 9. Real part of the response of matched filter to DSSS signal constructed of m-sequence of rank 8 and bandwidth equal to 4 kHz, received in NOF1 channel, representing the "0" bit value (a) and the "1" bit value (b).

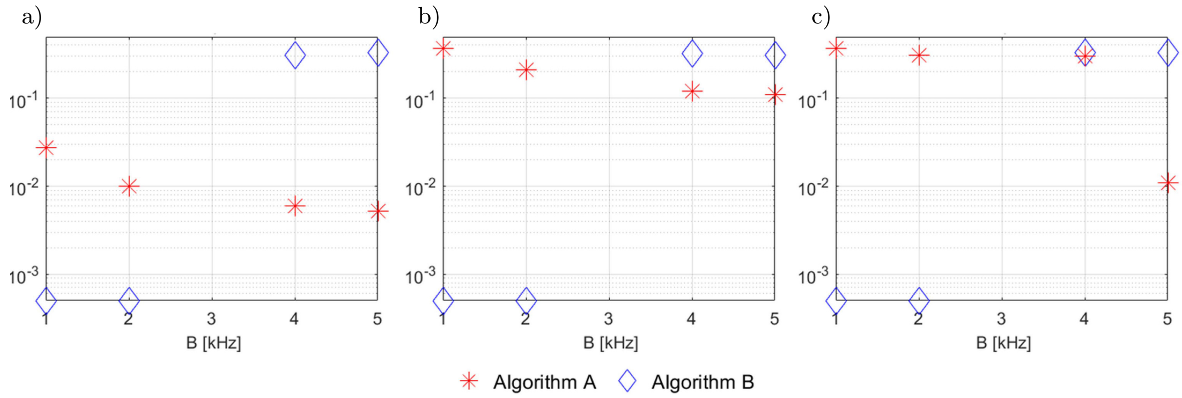

Fig. 10. The BER of the simulated transmission using DSSS signals constructed of m-sequences of rank 8 in NOF1 (a), NCS1 (b), and BCH1 (c) watermark channel.

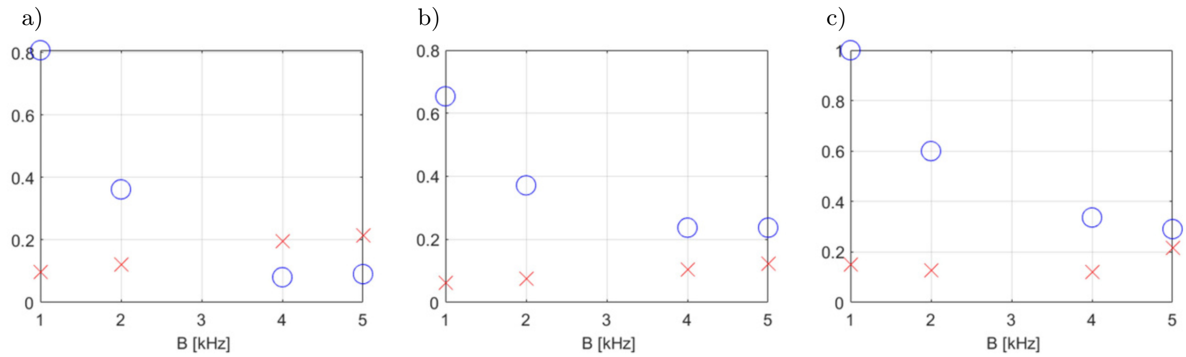

Fig. 11. The coefficients: C_A (×) and C_B (○) for transmission of DSSS signals constructed of m-sequnces of rank 8 in NOF1 (a), NCS1 (b), and BCH1 (c) watermark channel.

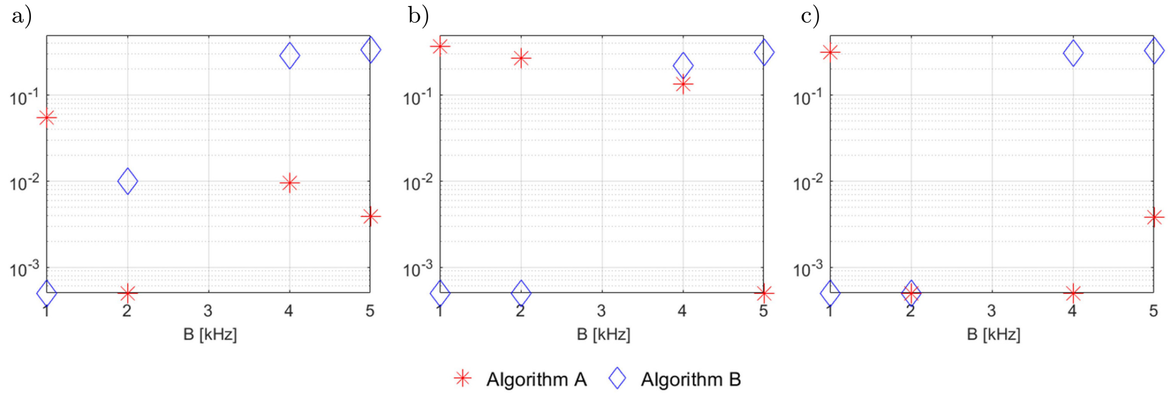

Fig. 12. The BER of the simulated transmission using DSSS signals constructed of Kasami of rank 8 in NOF1 (a), NCS1 (b), and BCH1 (c) watermark channel.

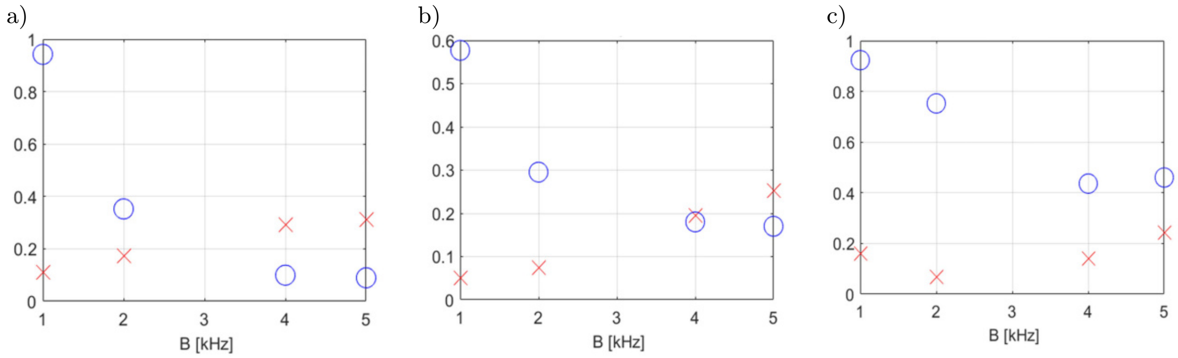

Fig. 13. The coefficients: C_A ($\times$) and C_B ($\bigcirc$) for transmission of DSSS signals constructed of Kasami codes of rank 8 in NOF1 (a), NCS1 (b), and BCH1 (c) watermark channel.

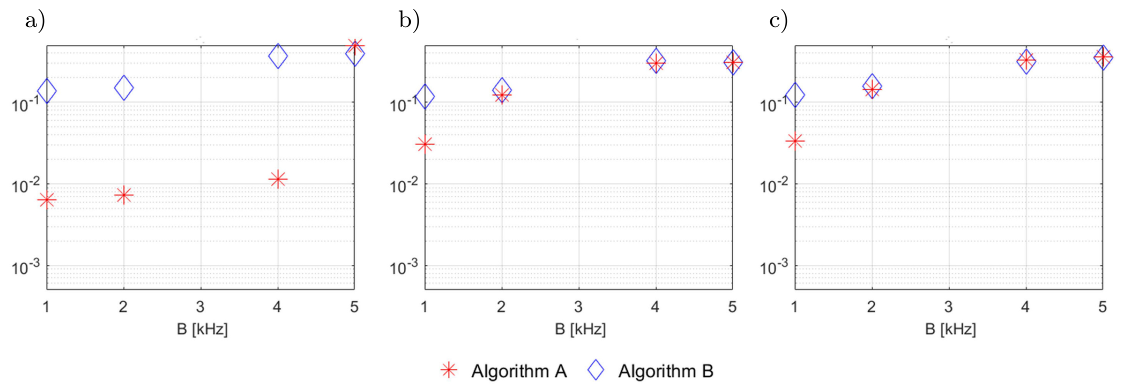

Fig. 14. The BER of the simulated transmission using DSSS signals constructed of m-sequences of rank 6 in NOF1 (a), NCS1 (b), and BCH1 (c) Watermark channel.

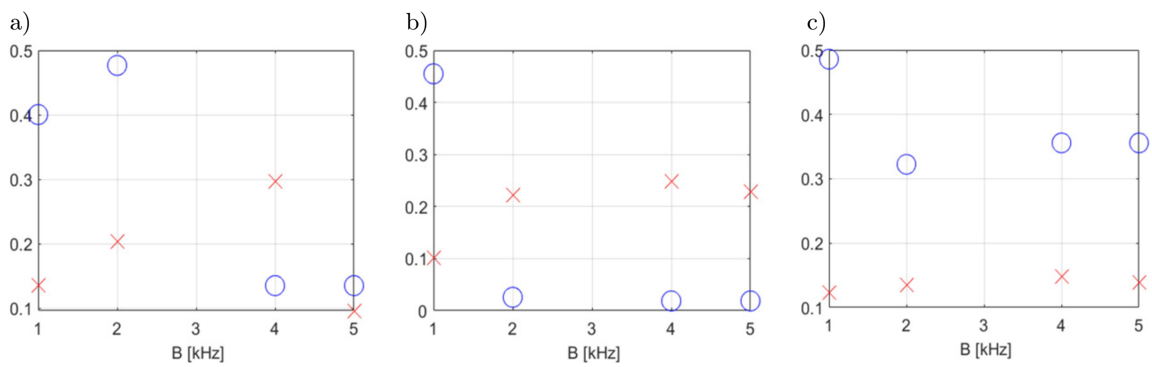

Fig. 15. The coefficients: C_A ($\times$) and C_B ($\bigcirc$) for transmission of DSSS signals constructed of m-sequnces of rank 6 in NOF1 (a), NCS1 (b), and BCH1 (c) Watermark channel.

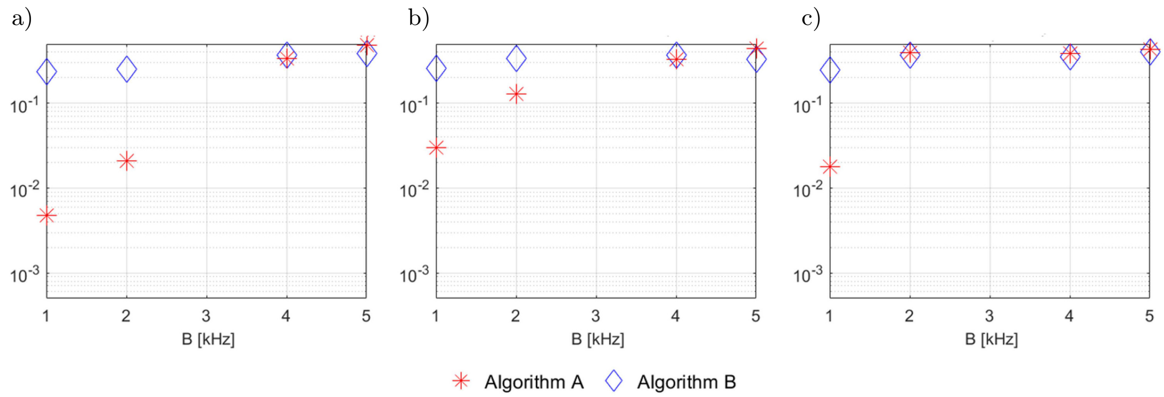

Fig. 16. The BER of the simulated transmission using DSSS signals constructed of Kasami codes of rank 6 in NOF1 (a), NCS1 (b), and BCH1 (c) Watermark channel.

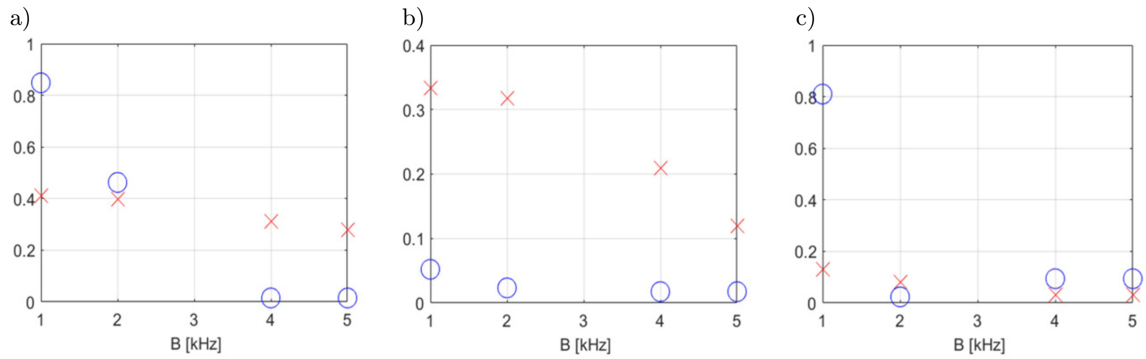

Fig. 17. The coefficients: C_A (×) and C_B (○) for transmission of DSSS signals constructed of Kasami codes of rank 6 in NOF1 (a), NCS1 (b), and BCH1 (c) Watermark channel.

During simulation tests with the use of m-sequences of the order 8, the use of the algorithm B allowed to obtain a BER less than 10^{-3} in each of the tested channels in the case of a bandwidth equal to 1 and 2 kHz. The use of the algorithm A gave much worse results (BER $<10^{-1}$ in the NOF1 channel and BER $>10^{-1}$ in the NCS1 and BCH1 channels). However, in wider frequency bands (4 and 5 kHz) it was the algorithm A that produced a lower BER than the algorithm B.

Using Kasami codes of the rank 8, a BER less than 10^{-3} was obtained with the detection algorithm B for a bandwidth of 1 kHz (a BER was less than 10^{-1} with the algorithm A) and the width detection algorithm A for a bandwidth of 2 kHz (a BER was of the rank 10^{-1} with the algorithm B). For a bandwidth equal to 4 and 5 kHz, the algorithm A made it possible to obtain a BER of the rank 10^{-2}, while the algorithm B produced a BER greater than 10^{-1}. In NCS1 and BCH1 channels the algorithm B made it possible to obtain a BER less than 10^{-3} for bandwidths of 1 and 2 kHz, while the algorithm A gave a BER greater than 10^{-1} in both bands in the NCS1 channel, and a BER greater than 10^{-1} for the 1 kHz bandwidth and a BER less than 10^{-3} for a bandwidth equal to 2 kHz. When detecting a signal with a bandwidth of 4 and 5 kHz the algorithm A turned out to give better results than the algorithm B.

Thus, using signals constructed of PN sequences of the rank 8 in each of the tested channels, it was observed that the algorithm A gives better results in case of 4 and 5 kHz bandwidth signals, while for signals of a bandwidth equal to 1 and 2 kHz, the algorithm B makes possible to achieve a better BER than the algorithm A.

Analysis of the value of the C_B coefficient (Fig. 11), determining "how often" the extremum at the output of the matched filter has the same delay, allows to confirm that with an increase of the signal bandwidth, the value of this coefficient decreases, and thus decreases the stability of the "peak" delay on the output of the matched filter, which is recognized as an extremum. In turn, the comparison of the BER graphs obtained for the B algorithm and the C_B coefficient values confirms the thesis that a decrease in the C_B coefficient is accompanied by the deterioration of the BER obtained using the B algorithm.

The relationship between the BER obtained by the algorithm A and the coefficient C_A is less clear than in the case of the algorithm B and the coefficient C_B. The value of the C_A coefficient obtained during all tests using sequences of the order 8 oscillates around the value of 0.2. However, as can be seen in the BER plots in Fig. 10 and 12, such a small average value of the absolute value of the correlation coefficient between successive symbols at the output of the matched filter is sufficient for the detection algorithm A to work with a BER of even 10^{-3}, so it made an incorrect decision less than once every 1000 symbols based on the value of the correlation coefficient between adjacent symbols at the matched filter output.

During tests conducted using DSSS signals built from m-sequences of the rank 6 (Fig. 14), it was possible to obtain a BER of the rank of 10^{-2} only in NOF1 using the detection algorithm A. For a bandwidth equal to 5 kHz band, the BER obtained with both detection methods was greater than 10^{-1}. The use of Kasami codes of a length 6 gave a BER of the rank 10^{-2} for the 1 and 2 kHz bandwidths using the algorithm A, also in the NOF1 channel. In case of other two channels detection algorithms failed to achieve a BER less than 10^{-2}.

In contrast to the results of tests carried out for signals constructed from sequences of the rank 8, in the case of signals built on sequences of the 6th order, i.e., 4 times shorter, it is difficult to observe the relationship between a BER and the C_A and C_B coefficients. Although the C_B coefficient in most tests has much higher values in the 1 kHz bandwidth than in the other bands, the BER obtained in this band using the B algorithm is not lower than 10^{-1} and at the same time close to the BER obtained with the same algorithm in the other bands.

5. Experimental tests

Experimental tests of DSSS communication were carried out in the model pool of the Gdańsk University of Technology. The pool is 40 m long, 4 m wide, and 3 m deep. During the experiment the transmitting and receiving transducers were immersed to a depth of 1.5 m (Schmidt, Schmidt, 2023). On both, transmitter and receiver sides, the laboratory model of the UAC system consisted of laptop computers with the MATLAB environment, underwater HTL-10 telephones (Schmidt, 2016), NI-USB6363 recording and generating devices and omnidirectional transducers with a resonant frequency of 34 kHz. A more detailed description of the experiment setup can be found in (Kochanska *et al.*, 2021).

5.1. Channel characteristics

The communication tests were preceded by measurements of the time-varying impulse responses (TVIR) of the UAC using correlation method and pseudo-random binary sequence (PRBS) probe signals, constructed of m-sequence of the rank 8. The probe signals were of four different bandwidths: 1, 2, 4, and 5 kHz, and the carrier frequency was equal to 30 kHz. The sampling frequency on the receiving side was equal to 200 kHz. Figure 18 shows modules of TVIRs measured using signals of a bandwidth equal to 1 and 5 kHz.

5.2. Communication tests

Similarly as in simulation tests, 16 kinds of DSSS signals were transmitted in the communication channel in a model pool. The carrier frequency of each signal was equal to 30 kHz. Every DSSS signal carried different information bits. The detection of information was performed using both detection algorithms – A and B. The results of achieved BER are shown in Figs. 19 and 21. Figures 20 and 22 show corresponding values of the coefficients C_A and C_B.

During transmission tests of signals constructed of PN sequences of the rank 6, a BER less than 10^{-3} was achieved in the 1 kHz bandwidth using the detection algorithm A. For the 2 kHz bandwidth, such a BER was obtained only in the case of the Kasami codes. The detection algorithm B did not allow to obtain a BER less than 10^{-1} in any of the transmission bands, even though the C_B coefficient was relatively high in some bands. Similarly, in the case of the use of pseudorandom sequences of the rank 8, the algorithm B did not allow to obtain a BER less than 10^{-1} (except for the 2 kHz bandwidth in the case of Kasami code). On the other hand, using the algorithm A allowed to obtain BER less than 10^{-3} in most transmission tests. The worst result was a BER less 10^{-2}

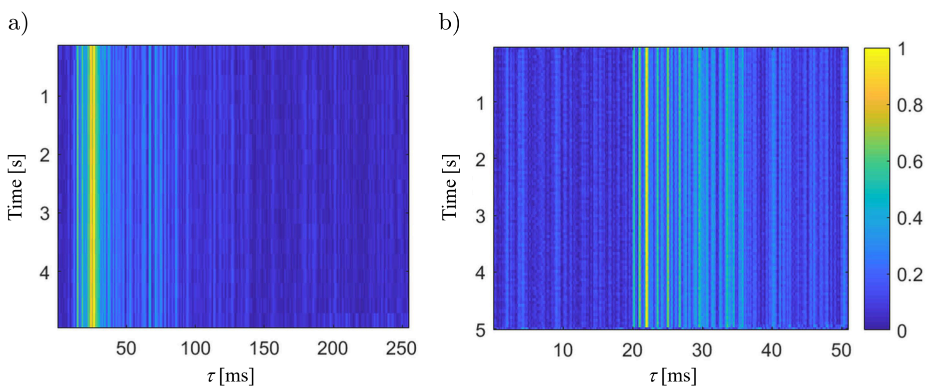

Fig. 18. Modules of TVIRs measured in the model pool using signals of bandwidth equal to 1 kHz (a) and 5 kHz (b).

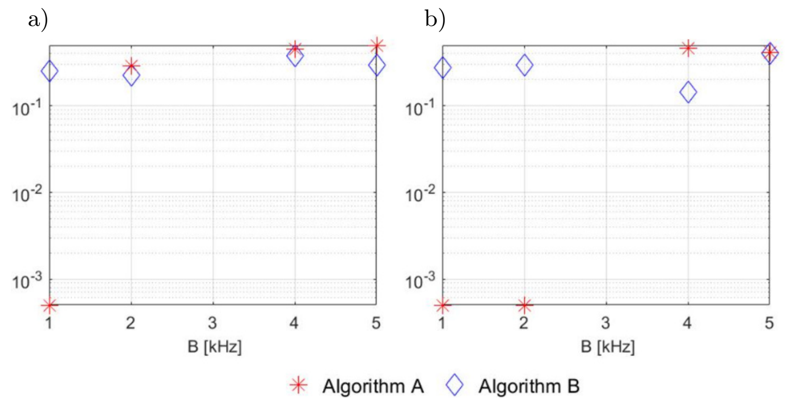

Fig. 19. The BER of the data transmission in model pool using DSSS signals constructed of m-sequences (a) and Kasami codes (b) of rank 6.

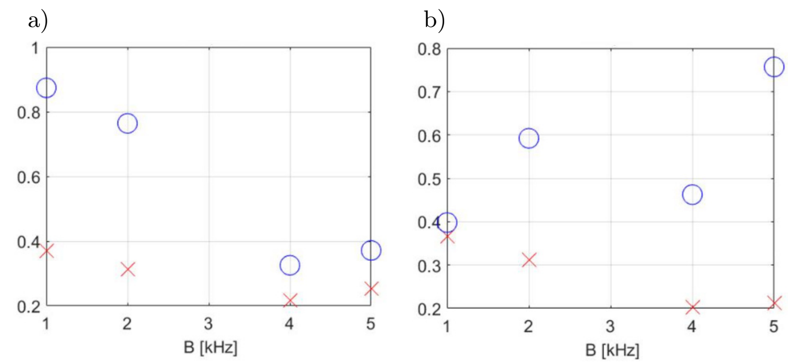

Fig. 20. The coefficients: C_A (×) and C_B (○) for transmission in model pool of DSSS signals constructed of m-sequences (a) and Kasami codes (b) of rank 6.

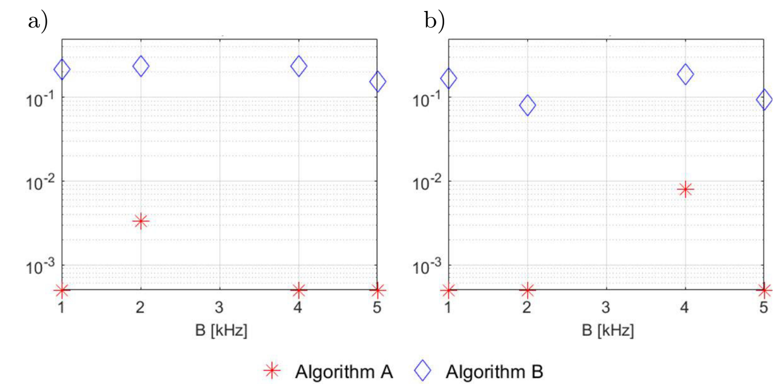

Fig. 21. The BER of the data transmission in model pool using DSSS signals constructed of m-sequences (a) and Kasami codes (b) of rank 8.

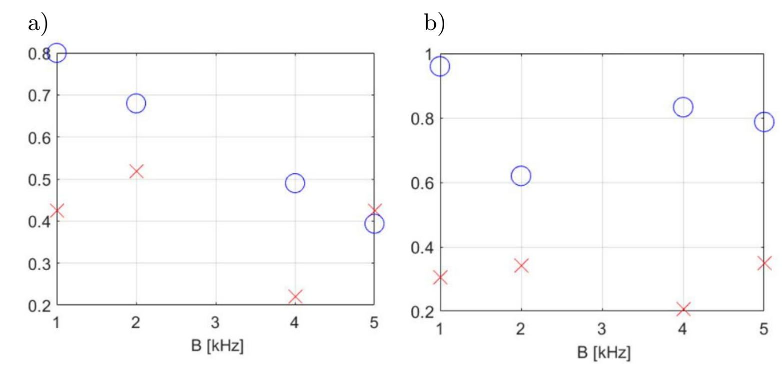

Fig. 22. The coefficients: C_A (×) and C_B (○) for transmission in model pool of DSSS signals constructed of m-sequnces (a) and Kasami codes (b) of rank 8.

for the 2 kHz bandwidth in the case of m-sequences and for the 4 kHz bandwidth in the case of Kasami sequences. Both for Kasami codes of the rank 6 and for the rank 8, a relationship was observed between an increase in the value of the C_A coefficient and the improvement in a BER obtained using the A algorithm. In the case of m-sequences, no such relationship was observed.

6. Conclusions

The simulation tests conducted with the use of three different simulation models of the UAC channel in a multipath environment and experimental tests conducted in a model pool allowed the formulation of the following conclusions, which are important guidelines for the design of the physical layer of the UAC DSSS system operating in very shallow waters.

The use of m-sequences and Kasami codes as spreading sequences allows to obtain a similar BER, so there are no contraindications to use Kasami sequences in systems with the code multiple access technique due to their good cross-correlation properties.

The order of 8 sequences allowed to obtain a much better BER than the order of 6 sequences. In a UAC system operating in very shallow waters, the longer ones should be used, at the expense of the achievable transmission rate.

The simulation studies carried out confirmed the thesis mentioned earlier in the article (Kochańska, 2021) that the detection algorithm A will allow to obtain a better BER than the detection algorithm B in the case of transmission in a wider band, i.e., 4 and 5 kHz. The algorithm B, on the other hand, performs better in narrower bands: 1 and 2 kHz. However, during experimental tests in the model pool, where multipath propagation was particularly strong, only the algorithm A (in each transmission band) allowed to obtain an acceptable BER.

Two indicators presented in the article: the C_A coefficient, assessing the degree of similarity of successive DSSS symbols at the output of the matched receiver, and the C_B coefficient, evaluating the delay stability of the dominant propagation path in the channel, are helpful in interpreting the differences in the transmission BER obtained by the two tested detection algorithms and can be used to predict which of the detection algorithms, under current propagation conditions in the channel, will allow to obtain a better BER. However, the relationship between the BER obtained by the B algorithm and the C_B index was much more clearly visible in the studies using the 8th order sequence than the relationship between the index C_A and a BER obtained with the algorithm A. These indicators can be used in the adaptive UAC system, enabling a rough assessment of the current conditions in the channel and, depending on them, the selection of a detection algorithm for the current operation in the receiver.

Acknowledgments

The paper was written as a result of the research project no. DOB-SZAFIR/01/B/017/04/2021 financed by the National Centre for Research and Development.

References

1. Freitag L., Stojanovic M. (2004), MMSE acquisition of DSSS acoustic communications signals, [in:] *Oceans '04 MTS/IEEE Techno-Ocean '04*, pp. 14–19, doi: 10.1109/OCEANS.2004.1402888.
2. Freitag L., Stojanovic M., Singh S., Johnson M. (2001), Analysis of channel effects on direct-sequence and frequency-hopped spread-spectrum acoustic communication, *IEEE Journal of Oceanic Engineering*, **26**: 586–593, doi: 10.1109/48.972098.
3. Kochańska I. (2021), A new direct-sequence spread spectrum signal detection method for underwater acoustic communications in shallow-water channel, *Vibrations in Physical Systems,* **32**(1): 2021106, doi: 10.21008/j.0860-6897.2021.1.06.
4. Kochanska I., Salamon R., Schmidt J., Schmidt A. (2021), Study of the performance of DSSS UAC system depending on the system bandwidth and the spreading sequence, *Sensors*, **21**: 2484, doi: 10.3390/s21072484.
5. Mironov A.S., Burdinskiy I.N., Karabanov I.V. (2018), The method of defining the threshold value of the symbolic correlation function for detecting DSSS hydroacoustic signal, *2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*, pp. 1–6, doi: 10.1109/FarEastCon.2018.8602588.
6. Pelekanakis K., Cazzanti L. (2018), On adaptive modulation for low SNR underwater acoustic communications, *OCEANS 2018 MTS/IEEE*, pp. 1–6, doi: 10.1109/OCEANS.2018.8604521.
7. Qu F., Qin X., Yang L., Yang T.C., (2018), Spread-spectrum method using multiple sequences for underwater acoustic communications, *IEEE Journal of Oceanic Engineering*, **43**(4): 1215–1226, doi: 10.1109/JOE.2017.2750298.
8. Ra H.-I., An J.-H., Yoon C.-H., Kim K.-M. (2021), Superimposed DSSS transmission based on cyclic shift keying in underwater acoustic communication, *OCEANS 2021 MTS/IEEE*, pp. 1–4, doi: 10.23919/OCEANS44145.2021.9706130.
9. Sarwate D.V., Pursley M.B. (1980), Cross-correlation properties of pseudorandom and related sequences, *Proceedings of the IEEE*, **68**(5): 593–619, doi: 10.1109/PROC.1980.11697.
10. Schmidt J.H. (2016), The development of an underwater telephone for digital communication purposes, *Hydroacoustics*, **19**: 341–352.

11. Schmidt J.H. (2020), Using fast frequency hopping technique to improve reliability of underwater communication system, *Applied Sciences*, **10**(3): 1172, doi: 10.3390/app10031172.

12. Schmidt J.H., Schmidt A.M. (2023) Wake-up receiver for underwater acoustic communication using in shallow water, *Sensors*, **23**(4): 2088, doi: 10.3390/s23042088.

13. Sozer E.M., Proakis J.G., Stojanovic R., Rice J.A., Benson A., Hatch M. (1999), Direct sequence spread spectrum based modem for under water acoustic communication and channel measurements, *Oceans '99. MTS/IEEE. Riding the Crest into the 21st Century. Conference and Exhibition. Conference Proceedings*, pp. 228–233, doi: 10.1109/OCEANS.1999.799743.

14. van Walree P. (2011), *Channel sounding for acoustic communications: Techniques and shallow-water examples*, FFI-Rapport 2011/00007, Forsvarets Forskningsinstitutt.

15. Zepernick H.J., Finger A. (2005), *Pseudo Random Signal Processing: Theory and Application*, John Wiley & Sons Ltd.

Archives of Acoustics Vol. 49, No. 1, pp. 141–147 (2024), doi: 10.24425/aoa.2023.146823

Technical Note

Impact of the Passage of Time on the Correct Identification of the Speaker Using the Auditory Method

Stefan BRACHMAŃSKI*, Bartosz HUS, Piotr STARONIEWICZ

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

*Corresponding Author e-mail: stefan.brachmanski@pwr.edu.pl

(*received January 25, 2023; accepted May 31, 2023; published online December 5, 2023*)

Courts in Poland, as well as in most countries in the world, allow for the identification of a person on the basis of his/her voice using the so-called voice presentation method, i.e., the auditory method. This method is used in situations where there is no sound recording and the perpetrator of the criminal act was masked and the victim heard only his or her voice. However, psychologists, forensic acousticians, as well as researchers in the field of auditory perception and forensic science more broadly describe many cases in which such testimony resulted in misjudgement. This paper presents the results of an experiment designed to investigate, in a Polish language setting, the extent to which the passage of time impairs the correct identification of a person. The study showed that 31 days after the speaker's voice was first heard, the correct identification for a female voice was 30% and for a male voice 40%.

Keywords: speaker recognition; crime acoustics; aural identification.

1. Introduction

Most courts around the world allow auditory identification of a person, i.e., testimony in which the witness is able to identify the speaker or other auditory impression. Auditory identification is one of the oldest methods of identifying a speaker from the voice. It was the first method accepted by courts of various countries, e.g., in the USA, in the state of Florida it has been used since 1907 (Hollien, 1990; 2002). Auditory identification is a complex technique due, among other things, to the temporal fluctuation of speech features and parameters caused by the psychophysical state of people involved in an event, as well as external acoustic conditions (Hollien, 1990; 2002; Hollien *et al.*, 2016). In Poland, this method is used when there is no sound recording and the identification of a person can only be made on the basis of an auditory assessment by the injured person or by witnesses to the event (The Code of Criminal Procedure [Kodeks Postępowania Karnego], 2016). Auditory assessment is also used in linguistic-measurement (Błasikiewicz, 1971; Dolecki, Rzeszotarski, 2002) and auditory-spectral (Alexander *et al.*, 2005; Begault, Poza, 2005; McDermott, Owen, 1996; Rose, 2002) methods. In these methods, at the auditory assessment stage, attention is paid, among other things, to the sound of voices being compared, the manner of accentuation, the rate of speech, pronunciation defects and the manner of utterance are analysed. Psychologists, phonoscopy specialists, and researchers in the field of auditory perception and forensic science more broadly describe many cases where this testimony has resulted in a wrongful conviction (Elmore, 2020; Possley, 2018).

Humans are able to recognise speakers based on their voice with varying degrees of effectiveness. Many factors affect the reliability of this method, namely familiarity with the speaker, duration of the speech sample, context, emotion, pronunciation defects, etc. (Deffenbacher, 1989; Hollien, Schwartz, 2000; Yarmey *et al.*, 2001). Professor Frances McGehee, through the events of 1935 (Conviction of Bruno Hauptmann in the case of the kidnapping and murder of Charles Lindbergh Junior (Hollien, 1990; The State of New Jersey v. Bruno Richard Hauptmann, 1935; Van Wyk, 1953)), conducted an experiment in which she wanted to prove the thesis that a person is unable to recognise an unfamiliar voice after a considerable lapse of time from its first hearing.

To this day, there are still expert claims that after 29 months, as in the case of Charles Lindbergh's identification of Hauptmann, auditory identification of the other person is impossible. McGehee, with her research, showed that the speaker's voice identification on the next day was quite high (83%), but that the effectiveness of the identification gradually declines over time, reaching only 13% after five months. Many people including Harry Hollien point out that Frances McGehee in her experiment did not take into account several important aspects such as the fact that the accuracy of identification can be affected by the appearance of an additional stimulus when hearing the speaker – emotional involvement or the ability of different people to remember the voice (Elmore, 2020; Hollien, 1990; 2002).

Later studies (Bricker, Pruzansky 1966; Hollien, 2002; Pruzansky, 1963) confirmed the conclusions of McGehee's research. However, these studies did not faithfully reproduce the experiment performed by McGehee.

The aim of the experiment presented in this paper was to investigate, under the Polish language conditions, how the passage of time affects the correct identification of men and women when the recogniser does not know the speaker's voice and when the speaker's voice is well known.

2. Study by Frances McGehee

The conviction of Bruno Hauptmann for the kidnapping and murder of little Charles Lindbergh Junior, and more specifically the fact that it was based on evidence of voice identification, initiated a series of experiments aimed at confirming the ability of humans to remember the voice of a speaker over the long term. One of the most famous experiments was an experiment performed by Frances McGehee in 1937 (McGehee, 1937; Hollien, 1990). The study involved 740 students (554 men, 186 women), while there were 49 speakers (31 men, 18 women). The study participants were divided into 15 groups. Each group was assigned a number of days until the next listening day, i.e., a lapse of time from 1 day to 5 months. The listeners' task was to recognise and indicate which of the five voices they had heard previously. The speakers presented to the listeners were selected from 49 people.

The thesis she put forward in her research can be formulated as follows: "Humans are unable to recognise an unfamiliar voice after a significant lapse of time from when they first hear it". To confirm this thesis, McGehee performed an experiment consisting of two parts. In the first part of the experiment, listeners heard a 56-word text read by a single speaker sitting behind an opaque screen. After a set amount of time, the group members heard the same sequence read in random order by five speakers (one identified and four whom the listeners had never heard). The listeners' task was to write down the number of the speaker they thought they had originally heard. McGehee repeated this experiment with the difference that the speakers presented to the listeners were prerecorded on tape. In both experiments, the results were very similar, with the effectiveness of identification decreasing as time passed. The correct identification after 1 and 2 days was 83% and after 7 days 81%. A noticeable decrease was found after 2 weeks, when the correct identification was 69%, dropping to 51% after 3 weeks and 35% after 3 months. The last period studied was after 5 months when the correct identification dropped to 13% (McGehee, 1937; Hollien, 1990).

McGehee's research showed that the correct speaker identification depends on the time which elapsed between hearing the voice of the person being identified and attempting to recognise the speaker, as well as the listener's ability to remember the voice pattern.

3. Experiment I – Recognition of an unknown speaker

The test material was a passage from the book "Norse Mythology" by Neil Gaiman (2017) read by five female and five male speakers. The speakers were selected based on a subjective assessment of the similarity of the voice tone, as well as the value of the laryngeal tone and the first four formants. The utterances were recorded on a digital recorder at the sampling rate of 44.100 samples/s and a resolution of 16 bits in PCM (wav) format. Before recording, each speaker practised reading the text to ensure fluency. The text was read out in an even, calm voice. The recordings were made in a home environment, in a quiet room isolated from external distractions. The utterances, i.e., the text being read out, were recorded five times. From all the recordings, one with the best-sounding utterances, without stammers, repetitions or uncontrolled artefacts, was selected for each reader (Hus, 2022).

Due to the state of epidemic emergency prevailing both nationally and internationally caused by the COVID-19 coronavirus (SARS-CoV-2 virus), the entire research process took place online from sending the message about the start of listening to sharing the evidence and comparison recording and receiving feedback with the listener's response.

The female voice recognition study involved 100 participants (65 male, 35 female) who were randomly divided into 11 smaller study groups. In contrast, 150 participants (88 male, 62 female) took part in the male voice recognition study, who were randomly divided into 11 smaller study groups of 10 participants each. The ages of the study participants ranged from 21 to 26 years for the female voice identification and 20–30 years for the male voice identification. The speaker identification group was formed from the candidates who passed the so-called zero test. This consisted of each candidate being presented with the statements of all five speakers (the male voice to the candidates in the male identification group and the female voice to the female identification group) 30 seconds after hearing the speech of the person being identified. Each candidate had to correctly identify the speaker being recognised. If the speaker identification was not correct then the candidate was not included in the study group.

The recording of the identified speaker was presented on the same day and at the same time to all study participants, with a female and a male identification group separately. After listening to the recording, it was deleted from the folder provided to the listeners. The participants in the experiment were therefore not given the opportunity to listen again to the recording read by the identified speaker. After the time set for the group had elapsed, the group members were informed that the listening window had been opened and the speaker's identification should be made. Each group member listened to a prerecorded text read by five speakers. After listening to all the voices, he or she indicated the number of the speaker whose voice, in his or her opinion, corresponded to the voice heard for the first time. In addition, each listener provided a degree of confidence in the identification. The information provided by the listener was automatically entered into the measurement form. All the information entered automatically updated the results table, which included:

- information about the correct identification of the suspect for a given listener;
- the total number of listeners who correctly identified the suspect;
- the number of people who correctly identified the suspect after a certain time;
- information about which speaker the people tested pointed to most often;
- information specifying the number of people not tested.

The recognition of the female voice was performed after: 1, 2, 3, and 7 days, 2 and 3 weeks, and 1 month after the first hearing of the recording of the recognised speaker's utterances, and the male voice additionally after 2, 3, and 4 months.

Figure 1 presents the results obtained from the experiment presented and reported by McGehee. The recognition performance of the female voice is more dependent on the passage of time than that of the male voice. After one day of hearing the female voice, 90% of the listeners made the correct identification, whereas for the male voice, all the listeners correctly recognised the person being identified. As the time passed, the recognition success rate decreased, so that after seven days the female voice was 60% and the male voice 90%; McGehee's figures were 81%. On the other hand, after

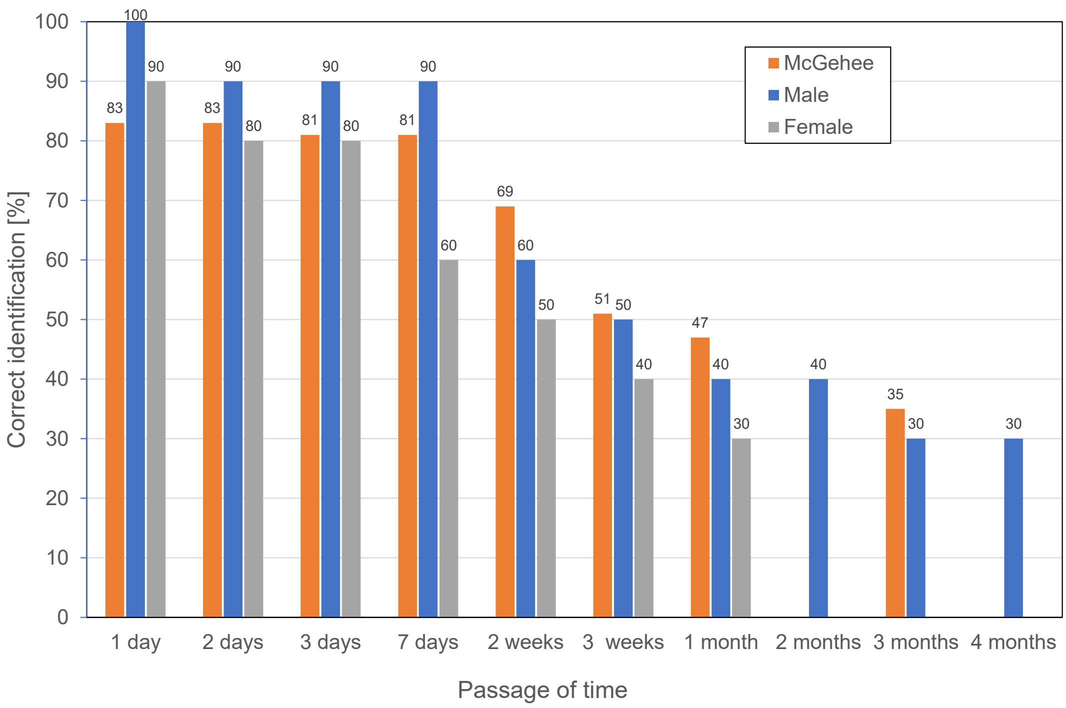

Fig. 1. Effectiveness of identifying an unknown person as a function of the passage of time.

one month, the recognition rate decreased very significantly, with the female voice at 30% and the male voice at 40%. The speaker recognition performance 1 month after the first hearing of the male voice was similar to McGehee's result (40% and 47%, respectively).

Analysing the summary results of the speaker identification by all the listeners, it was found that the correct identification of the male speaker was marginally better than that of the female speaker (Fig. 2). The correct identification of the female was 58% and that of the male was 62%, a difference of 4% in favour of the male voice.

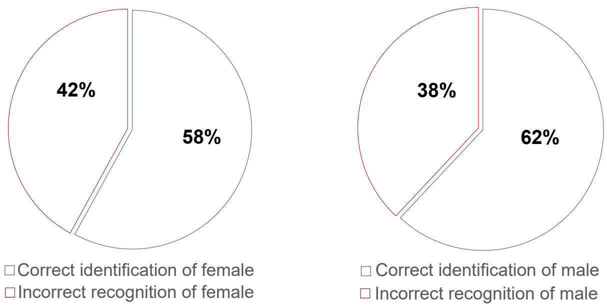

Fig. 2. Percentage of correct and incorrect identifications of female (a) and male (b) voices.

The results were analysed to see what effect the gender of the person identifying the speaker has on the correct identification. As a result of this analysis, it was found that there is no significant difference whether the speaker is recognised by a woman or a man (Fig. 3). A difference of 2% in the effectiveness of the person identification is within the statistical error range.

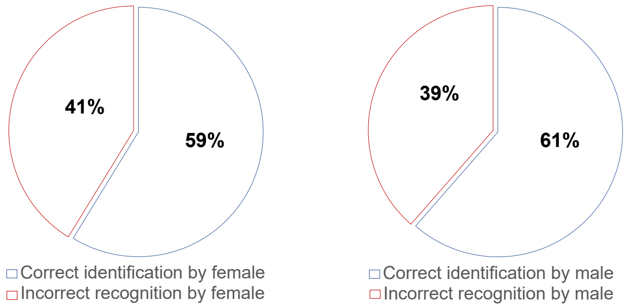

Fig. 3. Percentage of correct and incorrect voice identifications by female (a) and male (b).

4. Experiment II – Recognition of a known speaker

The second part of the experiment concerned the recognition of a speaker whose voice was previously known to the listeners. Hundred participants (54 men, 46 women), the age range 15 to 55 years, took part in the recognition of the female voice. Each person in this group had contact with an identified female at least once every fortnight. Ninety people (63 men, 27 women) participated in the male recognition, the age range 20 to 45 years. As in the female recognition group, also in this case each person in this group had contact with the male suspect at least once every fortnight.

This part of the experiment used the same test material as in the first part of the study, i.e., the recognition of the unknown speaker. The speaker recognition procedure was the same as in experiment I.

The results of this part were statistically analysed and the results are shown in Fig. 4. This part of the

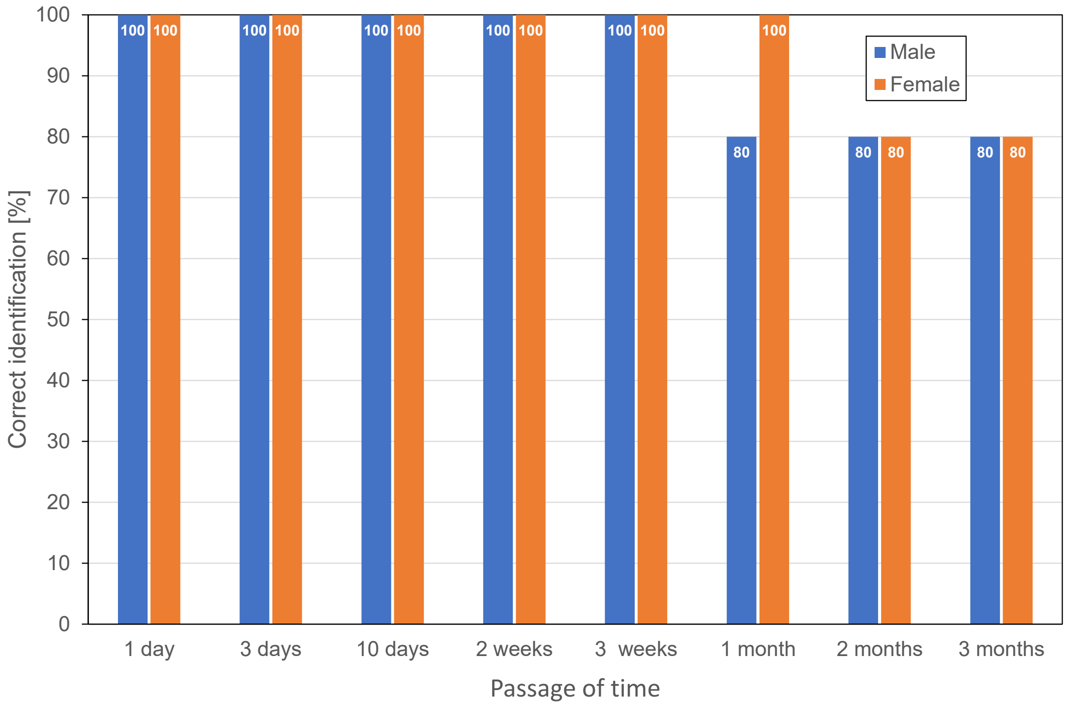

Fig. 4. Effectiveness of identifying a known person as a function of the passage of time.

experiment showed that if the voice of a well-known speaker is identified, even after 1 month it is possible to correctly identify both male and female speakers. It is only after 2 months that the identification efficiency of the familiar speaker drops to 80%. Analysing the results of the speaker identification carried out by all the listeners, it was found that the correct identification is not influenced by the gender of the person being identified (Fig. 5). When the speaker is known, the difference between female and male identification is only 1%, which is within the statistical error range.

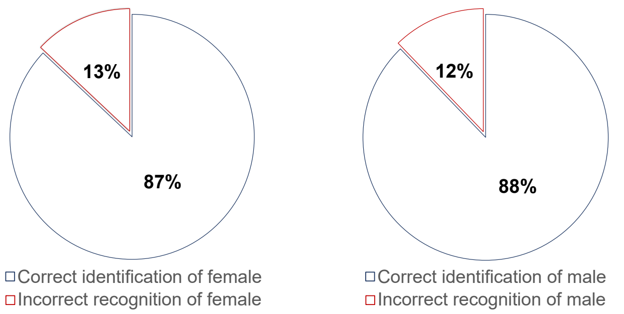

Fig. 4. Percentage of correct and incorrect identifications of voice female (a) and male (b).

In experiment II, taking into account the results of experiment I, the results were not analysed for the effect of the gender of the person identifying the known speaker on the correctness of the identification.

5. Conclusions

The research confirmed the conclusions of McGehee's experiment that as the time passes after hearing a speaker's voice, the speaker's recognition efficiency decreases rapidly. Comparing the results obtained in experiment I with those of Frances McGehee's study, it can be concluded that under the Polish language conditions the identification efficiency of the female voice decreases faster, while that of the male voice at a comparable rate. In the first three days, the correct identification of both female and male voices exceeded 80%. The very high efficiency of male voice identification persisted for one week (90%), but after two weeks there was a decrease to 60% (69% in McGehee's study), and after one month it decreased to 50% (47% in the McGehee study).

The resulting convergence of the results was expected in the light of Ebbinghaus' research on memory. In his classic work, he presented quantitative data on the decay of stored material over time (Ebbinghaus, 1885). The conclusion that the number of remembered items decreases with time has been confirmed by other researchers (Falkowski, 2004; Iwanicka, 2020). According to the Ebbinghaus curve, also known as the forgetting curve, which shows the relationship between the amount of information stored in memory and the time elapsed since hearing it, a person is able to reconstruct a limited number of units heard, e.g., after 5 days only 25% of the units heard, and after 30 days 20% (Ebbinghaus, 1885). Stressful circumstances can affect learning and memory processes. However, the nature of the effect of stress on memory is not fully understood, as both memory-enhancing and memory-impairing effects have been reported (Schwabe *et al.*, 2012). The memory curve is language-independent and can be adapted to many branches of learning related to perception (Ebbinghaus, 1885; Falkowski, 2004). Thus, it can be assumed that the Ebbinghaus curve also applies to the ability to remember the sound of the voice, including the auditory identification of the speaker. It should be noted, however, that in the case of remembering the sound of the voice, the curve falls much more slowly than the Ebbinghaus curve. Humans are able to remember the artefacts of the speaker's voice for longer than learned speech units, and especially in situations of emotional involvement. It can be assumed that auditory identification of a speaker does not depend on a language when it is made by speakers of the same language as the person being identified.

The results of experiment II showed that if the recognised voice was previously known to the identifier then 100% correct identification of the female voice was maintained over a period of 1 month. The high value of correct identification was still maintained after 3 months and was 80% for both female and male voices.

The results obtained in both experiments I and II presented here, as well as in the McGehee study, cast considerable doubt on the validity of identifying a person solely on the basis of a voice two years after hearing the voice for the first time (as in the case of Charles Lindbergh). However, retention, i.e., the ability to remember, especially under conditions of threat or personal involvement, cannot be overlooked. In Lindbergh's case, the production of a significant dose of adrenaline, which sharpened the hearing and also enhanced the ability to remember the voice for a long time, may have played an indiscernible role in remembering the voice of the abductor. Hollien's research shows that in a stressful situation, a person can remember the sound of a voice for a very long time. Therefore, it cannot be ruled out that in individual cases it is possible to recognise a voice even after 2 years, as was the case with Lindbergh.

In the process of auditory speaker identification, the technique of remembering the speaker's voice is also important. In the experiment in question, one listener used the technique of associating the voice he heard with the voice of a person he knew well. When listening to recordings of five people, including the identified person, the listener looked for the speaker whose voice best reproduced that of the person close to him. This listener correctly identified the suspect after 40 days and rated his confidence of identification at 10.

In summary, it can be said that speaker identification deteriorates very quickly when it is made by people who do not know the speaker and are not emotionally involved in the event, whereas it persists for a longer period of time in people who know the speaker.

In addition to further research into the effect of the passage of time on the effectiveness of speaker recognition by listeners, future work will focus on factors studied so far for automatic methods. These are factors masking the personal parameters of the speaker's voice, such as the influence of the speech coding and transmission techniques used (Jarina *et al.*, 2017), voice disguise techniques or the speaker's state or condition (Staroniewicz, 2021).

References

1. Alexander A., Dessimoz D., Botti F., Drygajlo A. (2005), Aural and automatic speaker recognition in mismatched conditions, *International Journal of Speech Language and the Law*, **12**(2): 214–234, doi: 10.1558/sll.2005.12.2.214.

2. Begault D.R., Poza F. (2005), Voice identification and elimination using aural-spectrographic protocols, [in:] *Audio Engineering Society Conference: 26th International Conference: Audio Forensics in the Digital Age*, www.aes.org/tmpFiles/elib/20220803/13256.pdf (access 03.08.2022).

3. Błasikiewicz S. (1971), The listening method for whispering and intensely distorted speech [in Polish: Metoda odsłuchu szeptu i mowy intensywnie zakłóconej], *Problemy Kryminalistyki*, **90**: 159–183.

4. Bricker P., Pruzansky S. (1966), Effect of stimulus content and duration on talker identification, *Journal of the Acoustical Society of America*, **40**(6): 1441–1450, doi: 10.1121/1.1910246.

5. Deffenbacher K.A. *et al.* (1989), Relevance of voice identification research to criteria for evaluating reliability of an identification, *Journal of Psychology*, **123**(2): 109–119, doi: 10.1080/00223980.1989.10542967.

6. Dolecki J., Rzeszotarski J. (2002), Use of the language-measurements method for the identification of people in phonoscopic examinations, *Problems of Forensic Sciences,* **LII**: 108–123, www.forensicscience.pl/pfs/52_dolecki.pdf (access 03.08.2022)

7. Ebbinghaus H. (1885), *Memory: A contribution to Experimental Psychology* [in German: *Uber das gedachtnis: Untersuchungen zur experimentallen psychologie*], Duncker and Humboltd.

8. Elmore T.L. (2020), *The effect of cued memory recognition strategies on word and speaker identification*, Ph.D. Thesis, University of London.

9. Falkowski A. (2004), Memory and knowledge in the context of scientific cognition development [in Polish: Pamięć i wiedza w kontekście rozwoju poznania naukowego], *Nauka*, **2**: 105–124.

10. Gaiman N. (2017), *Norse Mythology* [in Polish: *Mitologia Nordycka*], MAG.

11. Hollien H. (1990), *The Acoustics of Crime: The New Science of Forensic*, Kluwer Academic – Plenum Publishers, New York.

12. Hollien H. (2002), *Forensic Voice Identification*, Academic Press, London – San Diego.

13. Hollien H., Didla G., Harnsberger J.D., Hollien K.A. (2016), The case for aural perceptual speaker identification, *Forensic Science International*, **269**: 8–20, doi: 10.1016/j.forsciint.2016.08.007.

14. Hollien H., Schwartz R. (2000), Aural-perceptual speaker identification: Problems with noncontemporary samples, *Forensic Linguistics*, **7**(2): 199–211, doi: 10.1558/sll.2000.7.2.199.

15. Hus B. (2022), *The impact of the passage of time on auditory identification of a person* [in Polish: *Analiza wpływu czasu na audytywną identyfikację osoby w akustyce kryminalistycznej*], MA Thesis, Wroclaw University of Science and Technology.

16. Iwanicka K. (2020), Psychological basis of working with a group, with a particular emphasis on the issue of learning styles [in Polish: Psychologiczne podstawy pracy z grupą ze szczególnym uwzględnieniem zagadnienia stylów uczenia się] [in:] *Pedagogiczne aspekty edukacji filozoficznej z uwzględnieniem systemów penitencjarnych*, Lipski D., Malinowski M. [Eds.], pp. 105–113, Wydawnictwo DiG, https://swws.edu.pl/wp-content/uploads/2021/05/3.-Pedagogiczne-aspekty-edukacji-filozoficznej.pdf#page=105 (access 13.08.2022).

17. Jarina R., Polacky J., Pocta P., Chmulik M. (2017), Automatic speaker verification on narrowband and wideband lossy coded clean speech, *IET Biometrics*, **6**(4): 276–281, doi: 10.1049/iet-bmt.2016.0119.

18. McDermott M.C., Owen T. (1996), *Voice identification: The aural/spectrographic method*, Frank M. McDermott, Ltd., Owl Investigations, Inc. (access 03.08.2022).

19. McGehee F. (1937), The reliability of the identification of the human voice, *Journal of General Psychology*, **17**(2): 249–271, doi: 10.1080/00221309.1937.9917999.

20. Possley M. (2018), *Wilber Jones*, The National Registry of Exonerations, www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=5392 (access 01.08.2022).

21. Pruzansky S. (1963), Pattern-matching procedure for automatic talker recognition, *The Journal of the Acoustical Society of America*, **35**(3): 354–358, doi: 10.1121/1.1918467.

22. Rose P. (2002), *Forensic Speaker Identification*, Taylor & Francis Inc, London – New York, doi: 10.1201/9780203166369.

23. Schwabe L., Joëls M., Roozendaal B., Wolf O.T., Oitzl M.S. (2012), Stress effects on memory: An update and integration, *Neuroscience & Biobehavioral Reviews*, **36**(7): 1740–1749, doi: 10.1016/j.neubiorev.2011.07.002.

24. Staroniewicz P. (2021), Effect of deliberate and non-deliberate natural voice disguise on speaker recognition performance, *Acoustics, Acoustoelectronics and Electrical Engineering*, **888**: 312–325.

25. *The Code of Criminal Procedure* [in Polish: *Kodeks Postępowania Karnego*] (2016), Dziennik Ustaw, 1749.

26. *The State of New Jersey v. Bruno Richard Hauptmann* (1935), Famous Trials by Professor Douglas O. Linder, https://famous-trials.com/hauptmann/1393-hauptmann appeal (access 28.07.2022).

27. Van Wyk H.S.J. (1953), State of New Jersey v. Bruno Richard Hauptmann, *The South African Law Journal*, **70**: 182, https://heinonline.org/HOL/LandingPage?handle=hein.journals/soaf70&div=30&id=&page= (access 16.07.2022).

28. Yarmey A.D., Yarmey A.L., Yarmey M.J., Parliament L. (2001), Commonsense beliefs and the identification of familiar voices, *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, **15**(3): 283–299, doi: 10.1002/acp.702.